UNDERSTANDING THE POPULATION GENETIC STRUCTURE OF BOVINE MASTITIS-CAUSING *STAPHYLOCOCCUS AUREUS* IN NEW ZEALAND TO IDENTIFY POTENTIAL VACCINE CANDIDATES USING REVERSE VACCINOLOGY

A THESIS PRESENTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN
VETERINARY SCIENCE
AT MASSEY UNIVERSITY, PALMERSTON NORTH,
NEW ZEALAND.

JABIN NESARAJ

2021

Contents

List of Tables

List of Figures

ABSTRACT

Staphylococcus aureus is one of the major causes of bovine mastitis in New Zealand and worldwide, causing severe economic loss to the dairy industry. With significant advances in whole-genome sequencing (WGS) and associated bioinformatics, *S. aureus* from bovine mastitis has been studied worldwide, but the understanding of the organism's genomics is still incomplete. To my knowledge, there have been no published WGS studies of *S. aureus* in dairy cattle in New Zealand to date. As WGS has become more affordable, and the concomitant bioinformatic analysis offers high resolution in discriminating between *S. aureus* lineages, WGS analysis was applied to a sample of *S. aureus* isolates obtained from dairy cattle in New Zealand. The work undertaken in this thesis utilised advanced WGS analyses to study the genomic epidemiology of *S. aureus* from dairy cattle over a period of 15 years. The findings from the analysis enabled a subsequent reverse vaccinology analysis that identified a number of conserved peptides potentially useful for incorporation in subunit vaccines for cattle.

The first study analysed the population genetic structure of the pathogen using 188 bovine *S. aureus* isolates collected from dairy farms across New Zealand in 2002-03, 2013-14, and 2017-18. Ruminant-specific and non-specific clonal complexes (CCs) were identified. CC1 was the dominant CC, a unique feature not observed in dairy cattle in other countries. CC1 was predominant in cattle in the three periods of isolate collection, suggesting a stable and successful clonal lineage. Interestingly, CC1 is also the predominant CC in humans in New Zealand, and is mainly associated with humans in other countries and not commonly reported in cattle.

The second study compared the genomes of the bovine *S. aureus* isolates from the first study with genomes of quasi-contemporaneous human, canine and feline (n=59), and small ruminant (n=30) isolates also collected in New Zealand. Comparative genomic analyses of the core and accessory genomes were used to assess the

effect of the host species of origin on the phylogenetic clustering of the isolates, and to identify host-specific/host-adaptive genomic signatures. Comparative analysis of CC1 isolates identified marked phylogenetic segregations of both the core and accessory genomes among cattle and humans, and the presence of previously described ruminant-adaptive genes in bovine isolates, but not in human isolates. To my knowledge, this is the first report of ruminant host adaptation within CC1.

The third study compared the antimicrobial resistance results obtained using the Disk Diffusion test (DD) with a resistome analysis, to assess the potential usefulness of WGS analysis to predict the antimicrobial resistance phenotype of bovine *S. aureus*. The antimicrobial resistance genes identified were: the β -lactamase gene *blaZ*, the erythromycin resistance gene *ermC*, the streptomycin resistance gene *str*, and the fusidic acid resistance gene *fusC*. WGS identification of the *blaZ* gene had a sensitivity of 71% and a specificity of 100% in predicting resistance when the DD test was considered the reference standard. Similar analyses could not be performed with other antimicrobials due to the low frequency of other resistance genes in the sample.

Finally, the information obtained from the first two studies was used to construct a reverse vaccinology bioinformatic pipeline to identify potential vaccine candidate (PVC) proteins for *S. aureus* mastitis. Eighteen PVC proteins were identified using a range of bioinformatics tools. Some of these proteins have previously been shown to be immunogenic through *in vitro* and animal studies, providing cross-validation to the pipeline, while others have not yet been tested.

In summary, this thesis presents a detailed description of the population genetic structure of bovine mastitis-causing *S. aureus* in New Zealand over 15 years. The work provides new insights into the complex mechanisms of *S. aureus* host-adaptation to ruminants, and identifies some obstacles for the successful application of genomic analysis for the prediction of the antimicrobial resistance phenotype of clinical *S. aureus* isolates. Potentially useful proteins to be included in subunit vaccines for cattle are also reported.

ABBREVIATIONS

AMR ..... Antimicrobial Resistance Genes

ARGANNOT ..... Antibiotic Resistance Gene ANNOTation

ARIBA ..... Antibiotic Resistance Identification By Assembly

AST ..... Antimicrobial Susceptibility Test

AMOVA ..... Analysis of MOlecular VAriance

BURP ..... Based Upon Repeat Pattern

BURST ..... Based Upon Repeat Sequence Types

bp ..... base pair

CARD ..... Comprehensive Antibiotic Resistance Database

CC ..... Clonal Complex

CDS ..... CoDing Sequences

CI ..... Confidence Interval

CLSI ..... Clinical and Laboratory Standards Institute

CP ..... Capsular Polysaccharide

DLV ..... Double Variant Locus

DNA ..... DeoxyriboNucleic Acid

EUCAST ..... European Committee on Antimicrobial Susceptibility Testing

ELISA ..... Enzyme Linked ImmunoSorbent Assay

IMI ..... IntraMammary Infection

iTOL ..... Interactive Tree Of Life

MALDI-TOF ..... Matrix-Assisted Laser Desorption Ionization-Time Of Flight

MGE ..... Mobile Genetic Element

MIC ..... Minimum Inhibitory Concentration

MLST ..... MultiLocus Sequence Typing

MPI ..... Ministry of Primary Industries

MRSA ..... Methicillin Resistant *Staphylococcus aureus*

MSCRAMM ..... Microbial Surface Components Recognizing Matrix Molecule

MSSA ..... Methicillin Susceptible *Staphylococcus aureus*

NCBI ..... National Center for Biotechnology Information

NGS ..... Next Generation Sequencing

ORF ..... Open Reading Frame

PCR ..... Polymerase Chain Reaction

PVL ..... Panton Valentine Leukocidin

PVC ..... Potential Vaccine Candidate

RR ..... Relative Risk

SaPI ..... *Staphylococcus aureus* Pathogenicity Islands

SCC ..... Somatic Cell Count

SLV ..... Single Locus Variant

SNP ..... Single Nucleotide Polymorphism

SNV ..... Single Nucleotide Variant

ST ..... Sequence Type

WGS ..... Whole Genome Sequencing

WHO ..... World Health Organisation

GLOSSARY

Contig A set of overlapping DNA segments that together represent a consensus region of DNA.

Clonal complex A group of bacteria showing a high degree of genetic similarity that is conventionally based on near-identical multi-locus sequences types.

***De novo* assembly** Assembly of short sequence data into large contigs without using a reference genome.

de Bruijn graph Sequences are split into shorter lengths (k-mers), which are assembled based on k-1 common nucleotides to give a topology which can be traversed to deduce the full sequence.

Genome The complete set of genes or genetic material present in a cell or organism.

Genotype The genetic composition of an organism but also used to refer to single gene or set of genes.

Isolate A population of bacterial cells in pure culture derived for a single colony.

Jukes Cantor model Simplest substitution model which assumes equal base frequencies and equal mutation rates (e.g. base A is equally likely, following a mutation event, to be replaced by T, C, G or even A).

Lineage Groups of isolates sharing characteristics due to common descent.

Neighbour-joining (NJ) tree Evolutionary tree build based on matrix of pairwise evolutionary distances between the given sequences.

Phenotype Observable characteristics of an organism.

Reads Sequence of base pairs corresponding to all or part of a single DNA fragment.

Single nucleotide polymorphism (SNP) Nucleotide variation at a given genetic locus found to be frequent within a population.

Strain The descendants of a single isolate in pure culture.

ACKNOWLEDGEMENTS

In the eye of the storm, You remain in control
In the middle of the war, You guard my soul
You alone are the anchor, when my sails are torn
Your love surrounds me
In the eye of the storm

-Eye of the storm
(song by Bryan Fowler and Ryan Stevenson)

I would like to thank my primary supervisor, Dr Alex Grinberg, for offering me this research opportunity. I am grateful to him for sharing his immense knowledge, specifically on *Staphylococcus aureus* and appreciate our meetings where I could glean from his expertise. I would also like to acknowledge my supervisors, Professor Richard Laven for his pastoral care and help in statistical analysis in this project and Associate Professor Patrick Biggs, for his support with the bioinformatics part of my project given I was a fledgling in this field when I started.

I am grateful for the scholarship established by Dr Alex Grinberg and Professor Richard Laven. I am thankful for the funding I received from the School of Veterinary Sciences postgraduate fund and the Joan Berry Fellowships in Veterinary Science (Postgraduate) for their support of my research.

I am grateful to the many collaborators that helped me by donating their time and experience in this project – Dr Kristene Gedye, Thank you for your guidance with

molecular techniques starting from when I was pursuing my Postgraduate Diploma. I also cannot forget our many conversations about Marvel movies and the various theories we postulated about Infinity War and Endgame. Xiaoxiao Lin, Thank you for your support with whole-genome sequencing. Niluka Velathanthiri, Thank you for helping me organise and revive the isolates for this project. I am grateful to researchers, Associate Professor Anne Ridler, Dr Ali Karkaba and Dr Mick Clews, for providing the isolates that were the backbone for my study.

Through my journey towards the completion of my PhD, my family and friends have continued to be a pillar of overwhelming support. Su, my beautiful wife, I started my PhD around the same time I met you; I had my PhD confirmation by the time we got married; I am submitting my thesis as we step into our new journey of becoming parents. Thank you for putting up with me as I yammered on about my research and for the last minute proofreading. You are my strong support, pushing me to my potential, encouraging me when I am down, holding my hand when I am shaken and teaching me to ask for help. With you by my side, I am sure I will continue to reach greater heights.

I would not be who I am today without the people who brought me into this world, my Appa and Amma. You sacrificed years of your lives to always provide the best for me. I can never repay this debt of love. Thank you to my loving Anna, Kabha for always being there when I need you. You paved the way for me to move to New Zealand, by paying for my postgraduate diploma. Thank you Wesley, for being more than my brother by challenging me to succeed in everything I begin. Thank you to my Mummy and Daddy through marriage for your continual support through your encouraging words and prayer. Thank you Dharhas, my bava, for being my outsourced Indian tech support. Thank you Jada, my sister through marriage, for helping proofread my thesis in its final stages. Thank you to my parents in New Zealand, Don and Val, for always being available to listen to my concerns and offer words of wisdom, delicious food, and prayers. Thank you Aunty Noveline and Uncle Samuel, for opening your home to me for great conversations, food and gardening tips.

A special mention to my friend Gokul for directing me to Massey University when I was considering options for postgraduate studies. I thank my dear friends, Areesh, Roslyn and Roshni, for their continued support. I would like to thank my bosses and

work colleagues from Southern Rangitikei Veterinary Services, Bulls and Central Vets and Pets, Palmerston North, who supported me by providing the necessary time-off from work to focus on my PhD.

A special mention to my fur-babies, Feegle (6-year old Labrador x Mastiff) and Hiccup (2-year old Huntaway) for being considerate during my writing and only barking when I needed a break or during my zoom meetings with my supervisors. Shifu (3-year old Domestic Short Hair), thank you for being a typical cat (for ignoring me and lounging all day).

All this would not be possible if not for my loving Saviour, Jesus Christ. He has truly been with me through it all. I was most aware of His presence when the Virtual box in my computer crashed, and I lost 8 months of my research data. After multiple devastating failed attempts at retrieving my data, I was contemplating quitting my PhD when He reminded me of the song 'Eye of the storm' quoted above. He reminded me that He is my anchor when my sails are torn. Encouraged, I restarted my bioinformatics work and He directed me to a newly released genome assembler. This software fast-tracked the assembly process and I was able to restore my data along with correcting previous fallacies within weeks instead of spending another 8 months on this. This is only one of the many instances when God has been my anchor, my hope and the source of my strength and knowledge.

1

LITERATURE REVIEW

1.1 *STAPHYLOCOCCUS AUREUS*

1.1.1 THE PATHOGEN

The genus *Staphylococcus* comprises spherical Gram-positive bacteria that appear as single, pairs or grape-like clusters under the microscope. The bacteria are catalase positive, non-motile and non-spore-forming, with optimal growth at 30-37 °C in both aerobic and aerophilic incubation conditions. On agar, they grow as round, smooth colonies varying from yellow to white (Zangerl and Asperger, 2003; Smyth and Kahlmeter, 2005; Davis et al., 2006).

One species in the genus – *Staphylococcus aureus* – is one of the most important bacterial pathogens in humans. It causes a wide range of diseases ranging from skin diseases such as impetigo, folliculitis, and cellulitis, musculoskeletal infections such as osteomyelitis and pyomyositis, as well as bacteraemia, sepsis and endocarditis (Bergin et al., 2015; Idelevich et al., 2016; Olaniyi et al., 2016).*S. aureus* is also a common commensal organism in humans, with about 20% of the population carrying it in their nasal cavity (Kluytmans et al., 1997). The bacterium is found on other sites of the body, including the skin, axillae, perineum, vagina, pharynx, and the gastrointestinal tract (Wertheim et al., 2005).

S. aureus is also found as a commensal in livestock, being isolated from nasal and skin swabs of cattle, sheep and goats (Mork et al., 2012; Rahimi et al., 2015). *S. aureus* is also a common cause of mastitis in ruminants (Martins et al., 2017; Adkins et al., 2018), bumblefoot in birds (Heidemann Olsen et al., 2018), femoral head necrosis in chickens (McNamee et al., 2000), pododermatitis and abscess in rabbits (Viana et al., 2007), metritis and dermatitis in horses (Devriese et al., 1985; Shimizu et al., 1991), opportunistic infections such as osteomyelitis, pyoderma and septic arthritis in dogs (Tomlin et al., 1999), and urinary tract infections, upper respiratory tract infections, and wound infections in all species (Peton and Le Loir, 2014).

1.1.2 IDENTIFICATION OF *S. AUREUS*

S. aureus can be differentiated from other staphylococci by the presence of the enzyme coagulase, and the clumping factor, or Protein A. Along with a number of biochemical tests used for phenotypic identification (Table1.1a), its ability to grow in the presence of high salt concentrations is widely exploited for *S. aureus* isolation in culture. For example, Mannitol Salt agar is widely used as a selective growth medium for *S. aureus* that inhibits the growth of other organisms to a considerable extent (Chapman, 1945).

The molecular identification of *S. aureus* may involve DNA based tests, or proteome analysis. DNA-based tests for the identification of *S. aureus* involve the PCR-detection of species-specific genes such as the *S. aureus* nuclease gene (*nuc*) (Sasaki et al., 2010), coagulase (*coa*), clumping factor (*clfA*) (Mason et al., 2001) or the protein A (*spa*) genes (Haghkhah and Lotfi, 2016). Modern proteome analysis include MALDI-TOF (Matrix-Assisted Laser Desorption/ Ionisation-Time Of Flight) mass spectroscopy, that utilises molecular mass profiling of specific proteins and are increasingly being used for rapid and cost-effective identification of clinical isolates (Wieser et al., 2012).

1.1.3 THE *S. AUREUS* GENOME AND POPULATION GENETIC STRUCTURE

S. aureus contains a circular genome consisting of approximately 2.8 million base-pairs. Since *S. aureus* is found in various host species and across various environments, not all strains have an identical gene complement. Hence, in order to describe the population genetic structure, it is necessary to firstly understand the difference between bacterial ‘core’ and ‘accessory’ genomes. The core genome is often described as the set of genes that is present in >95% of the isolates, and the accessory genome is the set of genes present in a variable proportion (1-95%) of the isolates (Lindsay et al., 2006; Fuchs et al., 2018).

The core genome makes up approximately 75% of the typical *S. aureus* genome (Fuchs et al., 2018). Core genes code for proteins involved in functions that are fundamental for survival, such as metabolism, growth, replication, and also include

Table 1.1: Biochemical tests for the identification of *Staphylococcus aureus* to differentiate from *Staphylococcus epidermidis* and *Staphylococcus intermedius* (Vos et al., 2011)

Characteristics	***S. aureus***	***S. epidermidis***	***S. intermedius***
Acetoin production	Positive	Positive	Negative
Arginine dihydrolase	Positive	Weak positive	Mostly positive
Lactic acid production	Positive	Produces +isomer	Produces +isomer
Clumping factor	Positive	Negative	Mostly positive
Coagulase	Positive	Negative	Positive
Deoxyribonuclease	Positive	Weak negative	Positive
Fibrinolysin	Positive	Mostly positive	Negative
β-Glucosidase	Positive	Delayed reaction	Mostly positive
β-Galactosidase	Negative	Negative	Mostly positive
Heat stable nuclease	Positive	Weak negative	Positive
Haemolysis	Positive	Weak negative	Mostly positive
Hyaluronidase	Positive	Mostly positive	Not determined
Nitrate reduction	Positive	Weak positive	Positive
Urease	Weak positive	Positive	Positive
Galactose	Positive	Mostly positive	Positive
Lactose	Positive	Mostly positive	Mostly positive
Maltose	Positive	Positive	Weak reaction
Mannitol	Positive	Negative	Delayed reaction
Mannose	Positive	Positive	Positive
Melezitose	Negative	Delayed reaction	Negative
Ribose	Positive	Mostly positive	Positive
Trehalose	Positive	Negative	Positive
Turanose	Weak positive	Mostly positive	Mostly positive

Positive: 90% or more strains positive, Negative: 90% or more strains negative

species-specific virulence genes present only in *S. aureus* that code for toxins, cell surface-binding proteins, capsule biosynthetic proteins and exoenzymes (Lindsay and Holden, 2004).

A species' accessory genome consists of genes that encode for various non-essential functions, including virulence, antimicrobial resistance, and metabolic activity related to the host niche (Shittu et al., 2007). The accessory genome contains mobile genetic elements (MGE), which the organism has obtained through lateral (synonym: horizontal) gene transfer (HGT) (Hacker and Kaper, 2000). MGE makes up 10%-25% of the genome and consists of plasmids, integrated bacteriophages, transposons, insertion sequences, and pathogenicity islands (Chua et al., 2013).

The *S. aureus* population is highly clonal (Feil et al., 2003; Hanage et al., 2006; Méric et al., 2015), with low homologous recombination rates (Feil et al., 2003; Vos and Didelot, 2009), a conserved gene order (synteny), where the genes are localized in the same loci in the chromosome across strains (Takuno et al., 2012), and a 98-100% amino acid similarity of individual core genes (Lindsay and Holden, 2004).

The highly clonal aspect of *S. aureus* is assumed to be due to the existence of a lineage-specific Type1 restriction-modification enzyme system (*Sau*1) composed of the *hsdR*, *hsdM* and *hsdS* genes, that has been found to block HGT from other species and other clonal lineages of *S. aureus* (Waldron and Lindsay, 2006). A mutation in this enzyme system in an animal-associated *S. aureus* lineage has made it more susceptible to the transfer of antimicrobial resistance genes (Sung and Lindsay, 2007).

The *S. aureus* population shows some features of host specificity. Studies have found that certain *S. aureus* clonal lineages are either exclusively or most-often associated with one host species (Cuny et al., 2015). Based on a study of 161 human and 56 animal *S. aureus* isolates, Sung et al. (2008) concluded that the patterns of occurrence of genes in *S. aureus* were mainly lineage-specific, rather than host-specific, and that only a small number of genes or gene combinations were responsible for host specificity. In fact, the authors reported a set of well-characterized lineage-specific genes such as *fnbA*, *fnbB* and *coa*, which were present in all human isolates, but were variably present or missing in some animal lineages. They also reported that some genes carried on MGEs such as *chp*, *scn*, and *sak* were less common in animal *S. aureus* isolates than in human isolates. For example, the *scn* gene encodes for the staphylococcal complement inhibitor (SCIN) protein, protecting bacteria from

phagocytosis by neutrophils. The *scn* seemed to be a host-specific gene, as the SCIN protein isolated from human *S. aureus* appeared to be specific for the human complement system, having no effect on the complement system in other mammals (Rooijakkers et al., 2005). More recently, however, a SCIN described in a *S. aureus* isolated from a horse was found to have a broad host range, affecting the complement systems of horses, humans and pigs (De Jong et al., 2018).

1.1.4 TYPING OF *S. AUREUS*

The classification of *S. aureus* isolates into closely related groups, or 'strains' helps in outbreak identification, the study of transmission pathways, and understanding bacterial phylogeny and evolution. In the past such grouping was obtained using phenotypic methods such as phage typing and capsular typing, and later using molecular (genotypic) typing methods such as pulsed-field gel electrophoresis, multilocus sequencing typing, and spa-typing. Here below are brief descriptions of some of the commonest typing methods used for *S. aureus* (Table 1.2).

1.1.4.1 PHENOTYPIC METHODS

There are multiple phenotypic methods to type *S. aureus,* but the two methods that will be mentioned in other chapters are phage typing and capsular typing.

1.1.4.1.1 PHAGE TYPING This is based on the principle that the bacteriophages which kill *S. aureus* are strain-specific. The method involves inoculation of the *S. aureus* to be typed onto an agar plate and, once the plate is dry, the deposit of a small drop of diluted phages approved by the International Subcommittee for Staphylococcus Phage-Typing. The plates are then incubated at 30°C for 18 hours, or at 37°C for 4-6 hours, and then held at room temperature overnight. The plates are examined for lysis of *S. aureus* against a dark background using indirectly transmitted light (Blair and Williams, 1961). Phage typing was not used in this project.

Table 1.2: Common typing methods that have been used to classify *S. aureus* isolates.

Typing method	**Basis**	**Reference**
Phage typing	Infection with phage	(Blair and Williams, 1961; Martín-Bourgon, 1985)
Capsular typing	Reaction to a monoclonal antibody	(Karakawa et al., 1985)
Pulsed-field gel electrophoresis	Digestion of DNA, visualize on gel	(Prévost et al., 1991)
Multi-locus sequence typing	DNA sequence analysis of seven housekeeping loci	(Enright and Spratt, 1999)
Spa typing	DNA sequence analysis of the Protein A gene variable number tandem repeat region	(Shopsin et al., 1999)
Coa typing	DNA sequence analysis of the coagulase gene variable-number tandem repeat region	(Shopsin et al., 2000)

1.1.4.1.2 CAPSULAR TYPING This method is based on direct bacterial cell agglutination and immunoprecipitation of cell extracts by monospecific antisera against the capsular polysaccharides of *S. aureus*. One drop of specific antiserum is added to one drop of the bacterial suspension and observed for bacterial clumping 10 seconds later. Positive clumping identifies the capsular type for the organism (Karakawa et al., 1985). Capsular typing was not used in this project.

1.1.4.2 MOLECULAR METHODS

The most common molecular typing methods for *S. aureus* are explained in detail below.

1.1.4.2.1 SPA-TYPING This is a molecular typing method that utilises a single gene locus, the highly polymorphic X or Short Sequence Repeat (SSR) region of the *S. aureus* Protein A gene (*spa*) (Shopsin et al., 1999). This polymorphic region

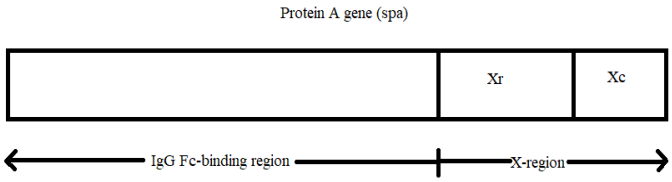

Figure 1.1: A diagrammatic representation of the Protein A(spa) gene. The spa gene consists of an IgG-binding region and an X region, which lacks IgG binding ability. The X region consists of Xr (repetitive region) and Xc (C-terminal region)

consists of the SSR (Xr) region, and a region encoding a C-terminal cell wall attachment sequence (Xc) (Uhlen et al., 1984) (Figure1.1). The SSR region contains a hypervariable, 21-30 base pair (bp)-long repeat that evolves continuously via duplications, deletions and point mutations and can, therefore, be used to distinguish *S. aureus* strains. Each unique repeat variant is allocated a numerical code using a software called RidomStaphType (Ridom GmbH, Würzburg, Germany), and the combination of these repeat units is used to define unique spa-type (t-) identifiers. The identifier of the spa types are numeric categories and do not indicate genetic closeness or relatedness. Relatedness of spa-types can be addressed using the Based Upon Repeat Pattern (BURP) algorithm, which groups spa-types into spa 'clonal clusters' (Mellmann et al., 2007). This clustering is achieved by assigning certain phylogenetic 'costs' to the gains and losses of repeats, and to point mutations. The Ridom software calculates the 'cost', and spa-types that differ by a cost of ≤ 4 are clustered into spa clonal clusters. Spa-typing along with BURP analysis has proved to be as effective for clustering *S. aureus* strains in epidemiological investigations, as other conventional molecular typing methods (Strommenger et al., 2006, 2008; O'Hara et al., 2016).

Nevertheless, spa-typing has its own limitations. Firstly, since spa-typing is based on a single locus, it does not represent an entire genome. Secondly, it lacks complete congruence with other typing methods. In fact, it is often unable to differentiate

closely related isolates within the same lineages defined by other methods, while at the same time it tends to segregate isolates defined by other methods as highly genetically related. This excessive segregation of spa types can be limited by the use of the BURP algorithm, that clusters spa-types into spa clonal clusters. Finally, the use of the RidomStaph Type software requires a paid license costing ~$NZ 4000 as of January 2020, which limits its use. Due to these limitations, particularly its inability to differentiate closely related isolates, spa typing was not used as the primary typing method in the present project.

1.1.4.2.2 PULSED-FIELD GEL ELECTROPHORESIS (PFGE) PFGE has been one of the most common molecular typing methods used in the study of disease outbreaks and the epidemiology of bacterial diseases in the 1990s and early 2000s. It was first developed by Schwartz and Cantor (1984) for yeast DNA.

The protocol (Figure 1.2) begins with the extraction of bacterial DNA in agarose plugs, and its digestion by a rare-cutter restriction endonuclease enzyme such as *SmaI*. The digested DNA fragments are then run in a pulsed-field gel electrophoresis chamber where the fragments are separated in an electric field which continuously changes direction, so that the big DNA chunks can migrate through the agar. This allows the separation of fragments through the pores of the gel based on both the pore size and molecular weight. The voltage used is usually 6 V/cm to separate molecules up to 1.5 Mb in size. At a higher concentration of the agarose gel, the DNA bands become sharper with only the smaller DNA molecules being resolved.

The temperature of the electrophoresis affects the rate of migration and resolution of the DNA fragments and thus the temperature is usually kept to between 12-16°C, which allows a balance between the migration rate and resolution of the DNA bands. The gel electrophoresis is usually run for 21 to 24 hours. At the end of the run, the DNA is labelled, the gel image is captured and the banding pattern analysed. Software such as BioNumerics (`http://www.applied-maths.com`) is then used to analyse and identify PFGE patterns. Isolates which produce the same pattern in the gel are considered to be the same Pulsed Field Type (PFT). Established interpretative criteria are used to infer the genetic relatedness of the strains (Tenover et al., 1995). Although there is a high degree of concordance between PFGE, MLST/eBURST and spa/BURP, PFGE has higher discriminatory power than the other methods

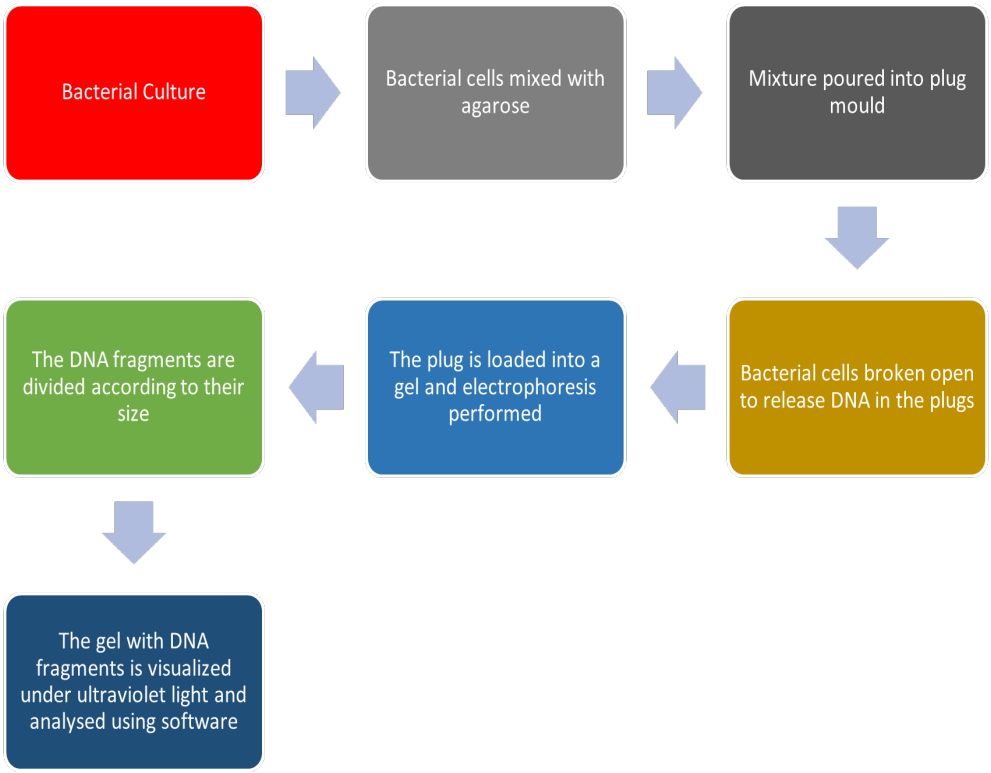

Figure 1.2: Flowchart of the process of the Pulsed-Field Gel Electrophoresis

(Strommenger et al., 2006), and seems to perform better in outbreak investigation scenarios. For example, in one study isolates were clustered into one sequence type (ST8) by MLST while PFGE segregated them into two different PFTs USA300 and USA500 (McDougal et al., 2003). For this reason, PFGE was often considered to be the reference standard method for short-term investigations such as outbreak investigations. However, PFGE appears to be less effective than other molecular typing methods for long-term epidemiological surveillance or for identifying phylogenetic relationships due to the technical difficulty to generate and visualize small bands with good resolution (Blanc et al., 2002).

Furthermore, the lack of a standard database for interlaboratory comparisons, lack of interpretative criteria and need for technical skills has made PGFE of limited value for national or international phylogenetic studies. PFGE was not used in this project.

1.1.4.2.3 MULTI-LOCUS SEQUENCE TYPING (MLST) MLST assigns *S. aureus* strains into sequence types (STs) based on the sequence variations in ~500 bp fragments of seven housekeeping genes. Housekeeping genes are found in all

Table 1.3: Primers used to amplify the seven housekeeping genes for multi-locus sequencing typing of *S. aureus*adapted from https://pubmlst.org/organisms/staphylococcus-aureus/primers

Gene	Primer	Sequence (5'-3')
Carbamate kinase *(arcC)*	*arcC*-up	TTGATTCACCAGCGCGTATTGTC
Carbamate kinase *(arcC)*	*arcC*-dn	AGGTATCTGCTTCAATCAGCG
Shikimate dehydrogenase *(aroE)*	*aroE*-up	ATCGGAAATCCTATTTCACATTC
Shikimate dehydrogenase *(aroE)*	*aroE*-dn	GGTGTTGTATTAATAACGATATC
Glycerol kinase *(glpF)*	*glpF*-up	CTAGGAACTGCAATCTTAATCC
Glycerol kinase *(glpF)*	*glpF*-dn	TGGTAAAATCGCATGTCCAATTC
Guanylate kinase *(gmk)*	*gmk*-up	ATCGTTTTATCGGGACCATC
Guanylate kinase *(gmk)*	*gmk*-dn	TCATTAACTACAACGTAATCGTA
Phosphate acetyltransferase *(pta)*	*pta*-up	GTTAAAATCGTATTACCTGAAGG
Phosphate acetyltransferase *(pta)*	*pta*-dn	GACCCTTTTGTTGAAAAGCTTAA
Triosephosphate isomerase *(tpi)*	*tpi*-up	TCGTTCATTCTGAACGTCGTGAA
Triosephosphate isomerase *(tpi)*	*tpi*-dn	TTTGCACCTTCTAACAATTGTAC
Acetyl coenzyme A acetyltransferase *(yqiL)*	*yqiL*-up	CAGCATACAGGACACCTATTGGC
Acetyl coenzyme A acetyltransferase *(yqiL)*	*yqiL*-dn	CGTTGAGGAATCGATACTGGAAC

prokaryotes and are involved in basic cellular metabolic functions. Therefore, they are part of the core genome. The housekeeping genes chosen for typing bacteria are genes that display neutral mutations, i.e., changes not driven by selective pressure, but by the accumulation of unbiased or nearly unbiased mutations (Spratt, 1999). Another important requisite for their selection is their spread throughout the whole bacterial genome, which reduces the chances of multiple loci co-segregation during recombination. The seven housekeeping genes used for *S. aureus* MLST are carbamate kinase (*arcC*), shikimate dehydrogenase (*aroE*), glycerol kinase (*glp*), guanylate kinase (*gmk*), phosphate acetyltransferase (*pta*), triosephosphate isomerase (*tpi*), and acetyl coenzyme A acetyltransferase (*yqiL*) (Enright et al., 2000). In the MLST method, the housekeeping genes are amplified through PCR using specific primers (Table 1.3), amplifying ~500 bp of the genes and the protocols listed in Table 1.4 and Table 1.5.

Once confirmed, the PCR products are sequenced to identify the nucleotides of the amplified fragment. A unique numerical identifier is provided for each allele (single or multiple nucleotide changes define different alleles), and the combination of the seven allelic numbers forms the allelic profile, or ST or the isolate. MLST was initially used for *Neisseria meningitidis* using 11 housekeeping genes (Maiden et al.,

Table 1.4: Polymerase chain reaction set up for the amplification of the seven housekeeping genes of *S. aureus* MLST (Ji, 2020).

Component	Volume	Final concentration
Forward primer	0.5 μL	0.2 μM
Reverse primer	0.5 μL	0.2 μM
Taq DNA polymerase	0.125 μL	1.25 units/50 μL PCR
10mM dNTPs	0.5 μL	200 μM
10X PCR buffer	2.5 μL	1x
Distilled water	Up to 25 μL	

Table 1.5: Polymerase chain reaction settings for the amplification of the seven housekeeping genes for MLST (Ji, 2020)

No of Cycle	Process	Temperature	Time
1 cycle	Initial denaturation	94°C	2 minutes
30 cycles	Denaturation	94°C	30 seconds
	Annealing	54-55°C	30 seconds
	Elongation	72°C	30 seconds
1 cycle	Final elongation	72°C	5 minutes

1998). Since then, MLST schemes have been designed for many bacterial species, *Streptococcus pneumonia* (Enright and Spratt, 1998), *Strep. pyogenes* (Enright et al., 2001), *Campylobacter* spp.(Dingle et al., 2001), *Cronobacter* (Baldwin et al., 2009), *S. aureus* (Enright et al., 2000). The organisms for which STs are defined are listed in the the PubMLST database (`https://pubmlst.org/organisms`) (Jolley et al., 2018).

The STs can be grouped into clonal complexes. A clonal complex (CC), a term first used in 2000 by Enright et al. (2000), is a group of STs showing a high degree of similarity based on near-identity of STs. CCs are important operational taxonomic units within the *S. aureus* population (Kuhn et al., 2006), as within each CC, strains share recent common descent. CCs are defined as *S. aureus* STs that differ at only one or two alleles across the seven housekeeping genes (Enright et al., 2000). However, PubMLST defines CCs as STs that match the central genotype at four or more loci unless they more closely match another central genotype (`https://pubmlst.org/organisms/staphylococcus-aureus/clonal-complexes`).

CCs can be defined using the algorithm Based Upon Repeat Sequence Types (BURST) (Feil et al., 2004). This algorithm groups all the non intersecting clusters of STs deposited in the PubMLST public repository into CCs, along with their founding

clone, which is usually the clone with the greatest number of single locus variants in the cluster (Feil et al., 2004). The BURST algorithm establishes each STs as nodes and combines them with other STs if they are a single locus variant (SLV) (varying at only one of the seven housekeeping loci) or double locus variants (varying at two loci). The STs are linked around the ST with the maximum SLVs and if there is a tie, then the double locus variant (DLV), triple locus variant (TLV) and finally the frequency of the ST in the sample are used (Francisco et al., 2009). For example, ST1 with an allelic profile of *arcC*(1), *aroE*(1), *glpF*(1), *gmk*(1), *pta*(1), *tpi*(1), *yqiL*(1) belongs to CC1, which includes approximately 372 STs (as of June 2020). The BURST algorithm can be run by packages such as eBURST (Feil et al., 2004), available as a plugin in the PubMLST database (Jolley et al., 2018), or goeBURST (Francisco et al., 2009). The BURST visualization in Figure 1.3 showcases one group of STs (ST1), defined as CC1, containing SLVs of ST1 encircled in red, and ST1 as the centre, and DLVs in the outer blue circle. The ST1 is also linked to its SLVs such as ST188, ST4768, ST852, ST81, ST3, ST2125, ST772, ST610 and their SLVs encircled in red. MLST was used in this project but by retrieval of sequences by *in silico* process from the whole genome sequences.

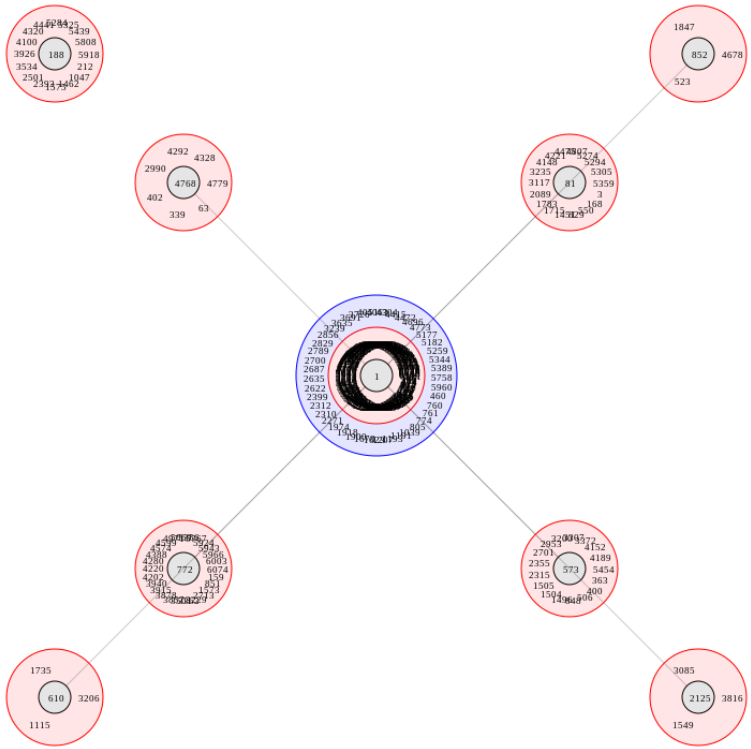

Figure 1.3: Visualization of BURST pattern of group 1 containing ST1 and the SLVs, DLVs of *S. aureus* at the centre with other major STs (ST188, ST4768, ST852, ST81, ST3, ST2125, ST772, ST610) with their SLVs. The STs within the red circle are the SLVs of the ST in the centre circle. The STs in the blue circle are the DLVs of the ST in the centre circle. BURST: Based Upon Repeat Sequence Types, ST: Sequence Types, SLV: Single locus variant, DLV: Double locus variant.

The availability of an accessible free MLST database for *S. aureus* (https://pubmlst.org/saureus/) where researchers can deposit metadata on their isolates has made MLST an important tool for the molecular epidemiological surveillance of *S. aureus.* As an average, there are 840 alleles per locus (as of October 2021), which can theoretically resolve ~ 2.9 x 10^{21} possible STs (7162 STs as of October 2021), which means that there is an extremely low probability of two phylogenetically unrelated strains being classified by chance within a single ST (Enright et al., 2002).

Conserved housekeeping genes evolve slowly compared to other genes of the bacterial genome, thus the degree of polymorphism is low in the housekeeping genes used for MLST when compared to variably present (accessory) genes, such as adhesion genes (Kuhn et al., 2006) or the surface protein (*sas*) gene (Robinson and Enright, 2003). MLST variation thus tends to represent distant evolutionary events, while changes in accessory genes may reflect more recent events; nevertheless, gene trees constructed using concatenated sequences of core and accessory genes are often highly congruent (Enright et al., 2002; Kuhn et al., 2006). Hence, the CCs, as defined by MLST, are highly predictive of both the core and accessory genome composition of the strain even though they are based only on seven housekeeping genes.

CCs *of S. aureus* are monophyletic. This has been cross-validated using different typing methods and whole-genome analysis (Aanensen et al., 2016; Boss et al., 2016). CCs are therefore useful descriptors of the genetic structure of the *S. aureus* population.

The *S. aureus* population appears to be discontinuous, with much less genetic diversity within CCs than between them (Planet et al., 2017). Some of this discontinuity can be explained by sampling biases, as sampling often focuses on medically or epidemiologically related strains, which has led to the available *S. aureus* genomes in databases being dominated by a small number of CCs (Planet et al., 2017), the majority of which originate from industrialised western countries. Nevertheless, Planet et al. (2017) concluded that the pattern of diversity in the *S. aureus* genome does reflect natural subdivisions into clonal lineages within the species.

Isolates of the same CC may vary in their ability to cause disease due to differences in the distribution of variably present virulence genes among the strains (Feil et al.,

2003). Thus, a CC or an ST cannot be categorised as 'pathogenic' or 'non-pathogenic'. This suggestion is supported by experimental evidence indicating that the presence or absence of single specific genes (for example the leukotoxin genes), or point mutations and truncations in central regulatory genes can result in a dramatic change in the virulence of a strain (Robinson et al., 2005; Cheung et al., 1994; Benoit et al., 2018).

1.1.4.3 LIMITATIONS OF MOLECULAR TYPING METHODS

The information obtained from the molecular methods used for *S. aureus* typing is limited and depends on the observed variations in either a single gene (spa-typing), multiple genes (MLST), or on mere DNA banding patterns (PFGE). The amplified DNA obtained in spa-typing or MLST represents only a minuscule fraction of the ~2.8 Mb genome of *S. aureus* (Kuroda et al., 2001), and while PFGE reflects changes accumulating across the whole genome, the method does not report any informative sequence information.

MLST is resource and time consuming as it needs multiple PCR and sequencing reactions that can be prohibitive for large sample sizes. However, among the molecular typing methods MLST is highly discriminatory for clonal organisms such as *S. aureus* and is commonly employed for population genetic investigation of *S. aureus* worldwide.

1.1.5 WHOLE GENOME SEQUENCING

Whole-genome sequencing (WGS) determines the complete sequence of nucleotide bases in the genome of an organism in a single process. This was first used in 1977 to sequence the bacteriophage ϕX174 (Sanger et al., 1977). The method was based on Sanger sequencing and was used commonly from 1977 to 2005 and is still used today, though less frequently. The Sanger sequencing method produces 500-1000 base-pairs(bp) high-quality reads of DNA sequence. The first complete genome of a free-living organism (*Haemophilus influenza)* was sequenced using this method in 1995 (Fleischmann et al., 1995). However, Sanger sequencing is time-consuming, complicated and very expensive.

In 2005, a new sequencing method using a pyrosequencing protocol (dubbed Next Generation Sequencing (NGS)) was developed, which reduced the run time and, cost of sequencing and improved the overall quality of the sequencing because of the huge volume of reads generated (Margulies et al., 2005). Today, NGS is available on several platforms including NextSeq2000, NextSeq550 and MiSeq (Illumina, San Diego, USA), Gene Studio S5 (Life Technologies, Massachusetts, USA), Sequel and Sequel II (Pacific Biosciences, California, USA).

Long reads (1-100 kbp) and short reads (50-400 bp) can be generated by different NGS platforms. The attributes of various next-generation sequencing platforms are listed in Table1.6

Table 1.6: Various next-generation sequencing platforms with their attributes

Attributes	Illumina			IonTorrent	PacBio	
	MiSeq	NextSeq550	NextSeq2000	Gene Studio S5	Sequel	Sequel II
Run time (hours)	4 to 55	12 to 30	11 to 48	2.5	Up to 20	Up to 30
Maximum output (Gb)	15	120	360	50	50	160
Maximum reads per run	25 million	400 million	1.2 billion	260 million	5 million	2 million
Maximum read length (base pairs)	2x300	2x150	2x150	600	190000	50000

NGS platforms have enabled the use of WGS in the study of disease outbreaks; for example, it has been used intensively by PulseNet (Centres for Disease Control and Prevention, USA) to identify and control foodborne disease outbreaks caused by a variety of species (https://www.cdc.gov/pulsenet/). WGS is increasingly also used for the study of the population genetic structure of organisms.

1.1.5.1 SEQUENCING PROCESS

Most current NGS platforms use similar sequencing processes, with minor changes in either the preparation or base calling phases. This discussion will focus on the process of 'Illumina sequencing by synthesis', the platform used in this PhD project.

Illumina sequencing starts with the preparation of the target DNA by its fragmentation into fragments of average sizes ranging from 250 to 1500 bp in length. PCR is then used to amplify the fragments, followed by adaptor ligation to the ends of the DNA fragments creating a DNA 'library'. The adaptors contain three sites: a sequence complementary to the oligonucleotides anchored on the flow cell of the sequencing machine, indices, or barcode sequences which are used to identify the sample from which the fragment was derived, and the primer binding site to allow attachment of the sequencing primer. The flow cell is a glass slide with lanes containing channels. The channels contain two types of oligonucleotide sequences. The adaptors bind to one type of complementary oligonucleotides on the surfaces of a flow cell. The flow cell is then washed to remove all unattached DNA fragments. Once the DNA fragments are attached to the surface of the flow cell, cluster generation commences. Cluster generation is the process of creating hundreds of identical strands of a single fragment of DNA. This is achieved by DNA polymerase creating a complementary strand of DNA besides the attached DNA fragments. The original DNA fragment is washed away while the complementary DNA fragment attaches its top adapter sequence to the second type of oligonucleotides on the flow cell surface forming a bridge. These bridges are then amplified by DNA polymerase producing double-stranded fragments and then denatured to form two separate DNA fragments attached to two different oligonucleotide sequences. One strand would be forward while the other would be a reverse strand. This process is called bridge amplification. This process is repeated over and over again until there are hundreds of identical strands of both forward and reverse sequences for a single DNA fragment. These identical strands are called clusters. This is simultaneously done across the flow cell for thousands of DNA fragments in the cell, forming millions of clusters by a process dubbed 'clonal amplification'.

After the bridge amplification, the reverse strands are removed and washed off from the flow cell leaving only the forward strands. The sequencing is done to produce reads by adding complementary fluorescent-tagged nucleotides to the forward strands

after the attachment of the read-1 primer. As the nucleotides are added, they emit fluorescence based on the nucleotide base when excited by a light source which is recorded by the machine. This process is called sequencing by synthesis. The number of cycles determines the length of the reads. This process is undertaken for all the clusters in the lane simultaneously. After the first read is sequenced, the read is washed away and then the first index read is also sequenced similarly using the index primer.

The forward strand then forms a bridge with the second type of oligonucleotides on the flow cell where the second index read is sequenced. Then the bridge is amplified forming a double stranded fragment which is then denatured. This forms two strands: forward and reverse strand. This time the forward strand is removed and washed away. The sequencing by synthesis process is repeated for the reverse strand using the read-2 primer.

These data are stored as 'reads', in which strings of identified base pairs that constitute either part or all of the DNA fragment, and in this study are approximately 150bp in length (the lengths of the reads depend on the sequencing machine and number of cycles used). In paired-end sequencing, the DNA fragments are sequenced from both ends and the reads are stored in paired files.

1.1.5.2 PRE-ASSEMBLY PROCESS

Once the millions of DNA reads are obtained at the end of the sequencing process, the quality of the data is checked by a per-base sequence quality checker, and a per-sequence GC content to check uniformity of GC content across all the reads and proportion of duplicated reads, using tools such as FastQC ((http://www.bioinformatics.babraham.ac.uk/projects/fastqc/)). The statistics provided by these tools are then used to trim low-quality data using software such as SolexaQA++ (Cox et al., 2010) and Trimmomatic (Bolger et al., 2014). This step is essential when assembling software packages that do not have inbuilt trimming and error correction processes. Certain softwares such as ALLPATHS-LG (Gnerre et al., 2011), can assemble sequences from raw reads.

Foreign sequences from primers and vectors must be removed. This can be done using software tools such as Trimmomatic or SolexaQA++. The adapters tagged to

the fragments are usually removed at the end of the sequencing process, but any adapter sequences left behind can be removed with scripts such as cutadapt (Martin, 2011), or programs such as Skewer (Jiang et al., 2014). PhiX phage DNA is added to the sequencing reaction by Illumina platforms to calibrate sequence quality scores. This PhiX phage DNA can affect the assembly process by producing contaminated contiguous sequences. These known vector sequences can be removed from the raw data after mapping to the phage DNA sequence using short read aligners such as Burrow-Wheeler Aligner (BWA) (Li and Durbin, 2009) or Bowtie (Langmead et al., 2009). This whole pre-assembly process can be completed with scripts such as QCtool (`https://github.com/mtruglio/QCtool`), which was used in this study. The tool incorporates and automates the process of quality reading, trimming and adapter removal. The resulting processed reads can then be used for assembling the whole genome.

1.1.5.3 GENOME ASSEMBLY

Short read sequencing machines such as Illumina Hi-Seq generate millions of reads which can be between 50-300 bp in length. The reads are a continuous nucleotide sequence that is generated from the target DNA. Each read is a representation of only a minuscle fraction of the genome. The reads overlap each other so that each genome position is included in numerous reads ('read depth'). To make sense of the reads, they must be assembled into longer strings (dubbed 'contigs'), or mapped via alignment to an existing reference genome sequence.

Assembly of the reads can be done in one of two ways: mapping-based assembly, or *de novo* assembly. Mapping-based assembly is done by mapping the reads to a previously sequenced reference genome sequence. *De novo* assembly joins the reads together based on either read overlap (overlap method) or graph methods such as string or de Bruijn graphs (Sohn and Nam, 2018). Mapping-based assembly produces genome sequences that solely depend on the sequence of the reference genome, and reads that don't map to it are discarded. Hence, mapping-based assembly is not useful for organisms or strains that have no previously sequenced genomes, or for highly genetically variable organisms, such as bacteria. In theory, the *de novo* assembly method can assemble reads into a whole genome without the need for a reference genome, but in practice, it produces several contiguous consensus sequences known

as contigs, which can be further assembled into supercontigs by scaffolding multiple contigs. Scaffolding is a process by which the contigs are oriented and the whole genome with known gaps is reconstructed. However, errors in sequencing and the presence of repeat sequences means that a genome assembled using *de novo* assembly will still contains gaps between contigs (Saraswathy and Ramalingam, 2011). For this reason, genome sequences composed of contigs are dubbed 'draft genomes', as opposed to complete genomes. Increasing read lengths and/or sequencing depth, and reducing sequencing error could in the future make assembly more accurate and reduce the length of the gaps (Ekblom and Wolf, 2014). Multiple software packages are available to automate the assembly process and they vary in their programming language, algorithm, and the input read types that they can process. Some of these packages are listed in Table 1.7.

Several parameters can be used to assess the quality of assembly (Table1.8), using programs such as QUAST (Gurevich et al., 2013). If a complete reference genome is present for the target organism, the completed assembly can be mapped to it, to check for the quality of the assembly process. As mentioned above, the use of a reference genome is not always useful for bacterial genomes.

1.1.5.4 GENOMIC ANALYSIS

The first complete genome of *S. aureus* was generated using Sanger sequencing and was published in 2001 (Kuroda et al., 2001). As of October 2021, there were 655 fully annotated *S. aureus* whole genome sequences in the National Centre for Biotechnology Information (NCBI) Reference Sequences (RefSeq) database (O'Leary et al., 2016). There were also 772 *S. aureus* complete genome sequences currently uploaded (as of October 2021) to the Pathosystems Resource Integration Center (PATRIC) database (Wattam et al., 2017). The advent of NGS has increased the throughput of sequencing and the number of deposited genomes has increased in a shorter time frame. Due to the development of efficient bioinformatics tools the focus has shifted from fully annotated complete genomes to draft genomes or incomplete assembled sequences. This has led to the deposition of around 13406 and 17018 *S. aureus* genomes (as of October 2021) in the NCBI and PATRIC databases, respectively, totals which include complete genomes, chromosomes, contigs, and scaffolds (O'Leary et al., 2016; Wattam et al., 2017).

WGS allows the comparison of bacterial genomes down to the level of a single base pair. Species identification (Cineros and Lund, 2017), prediction of antimicrobial resistance (Bortolaia et al., 2020), spa-typing (Jünemann et al., 2013), virulence gene identification (Chen et al., 2005), and MLST (Seemann, 2020) can all be accomplished with a single WGS by utilizing various open-source or commercial bioinformatic tools. WGS therefore enables extensive comparisons and avoids the need to run single PCR reactions for every individual gene required for the analysis.

WGS can also be used to study the evolution and phylogenetic relatedness of various lineages and the predominance of certain strains in a host, disease process, or any ecological niche (Price et al., 2013). In particular, genome-wide association studies using WGS have helped in identifying genetic determinants for host specificity and virulence (Falush and Bowden, 2006), and studying the adaptive evolution of *S. aureus* in chronic infections (McAdam et al., 2011).

Yet, WGS has a number of limitations. Firstly, despite the significant reduction in the cost of sequencing and reduction of the turn around times, WGS is still too expensive for use in point of care or diagnostic laboratory routines. Additionally, the sequencing platforms produce short reads that have to be assembled, requiring bioinformatic and programming skills which are not always available. The assembled genome is often incomplete due to mismatching with reference sequences or because repeat regions are not resolved by *de novo* assembly methods (Price et al., 2013). However, there are increasingly user-friendly programs available for analysis.

The availability of WGS provides many research options and opportunities to perform in depth analyses of the population genetic structure of pathogenic bacteria and investigate their phenotype-genotype associations of interest.

Table 1.7: Bioinformatics software packages available for genome assembly with their programming language, algorithm, and the input reads

Assembler	Programming Language	Algorithm	Input reads	Reference
ABySS	C++	De Bruijn graph	Paired-end and single-end	(Simpson et al., 2009; Jackman et al., 2017)
Velvet	C	De Bruijn graph	Paired-end and single-end	(Zerbino, 2010)
SPAdes	Python	De Bruijn graph	Paired-end and single-end	(Nurk et al., 2013)
Edena		Overlap layout consensus	Paired-end and single-end	(Hernandez et al., 2008, 2014)
SGA	Python	String graphs	Paired-end and single-end	(Simpson and Durbin, 2012)
SOAPdenova		De Bruijn graph	Paired-end and single-end	(Luo et al., 2012)
SKESA	C++	De Bruijn graph	Paired-end and single-end	(Souvorov et al., 2018)
MEGAHIT	C++, Python	De Bruijn graph	Paired-end and single-end	(Li et al., 2015a, 2016)
Miniasm	C	Overlap layout consensus	Long reads	(Li, 2016)

Table 1.8: Quality parameters of the whole-genome assembly with their descriptions provided in the QUAST manual (Gurevich et al., 2013).

Parameter	Description
Number of contigs	The total number of contigs in the assembly
Largest contig	Length of the longest contig in the assembly (in base pairs)
Total length	Total number of the bases in the assembly including all contigs
GC (%)	It is the total number of G and C nucleotides present in the assembly divided by the total length of the assembly
N50	It is the length of all the contigs of that length or longer covers at least half of the assembly length.
NG50	Length for which the collection of all contigs of that length or longer covers at least half the reference genome
N75	It is the length of all the contigs of that length or longer covers at least 75% of the assembly length
NG75	Length for which the collection of all contigs of that length or longer covers at least 75% of the reference genome
L50	Minimal number of contigs that cover half the assembly
LG50	Minimal number of contigs that cover half the reference genome
L75	Minimal number of contigs that cover 75% of the assembly
LG75	Minimal number of contigs that cover 75% of the reference genome
Misassembled contigs length	Total number of bases in the misassembled contigs
Genome fraction (%)	Percentage of aligned bases in the reference genome
#N's per 100kbp	The average number of uncalled bases (N's) per 100000 assembly bases
Genomic features	Number of genomic features such as genes, CDS in the assembly based on the reference genome and annotated list of genomic features

1.2 *S. AUREUS* MASTITIS

1.2.1 BOVINE MASTITIS

Mastitis, the inflammation of the mammary gland, is the most economically important disease of dairy cattle (Hogeveen et al., 2011, 2019). The losses due to mastitis are direct (e.g. reduced milk production and treatment costs including milk withholding) and indirect (e.g. reduced fertility and increased culling). A recent analysis to estimate the total cost of mastitis reported an average cost of 11-18% of the gross margin per cow per year, with milk production loss and culling being the most important cost factors (Hogeveen et al., 2019). In New Zealand, it is estimated, mastitis costs $280 million a year to the dairy industry, with an average estimated cost of $180 per cow (NZHerald, 2012). Although bovine mastitis can be caused by non-bacterial causes and injury, the main cause is bacterial infections.

The most prevalent bacteria causing bovine mastitis are *S. aureus*, *Streptococcus uberis*, *Streptococcus agalactiae*, *Streptococcus dysgalactiae*, *Escherichia coli*, coagulase-negative *Staphylococci* and *Klebsiella pneumonia* (Dufour et al., 2019). In New Zealand, *S. aureus* is the second most common causative agent of bovine mastitis identified by diagnostic laboratories after *Streptococcus uberis* (McDougall et al., 2007; Petrovski et al., 2009, 2011).

S. aureus can cause clinical mastitis in dairy herds, but most commonly causes subclinical mastitis (Wellnitz and Bruckmaier, 2012; Petersson-Wolfe et al., 2010). Clinical mastitis manifests with visible changes in the milk, like changes in colour and the presence of clots. There may also be observable udder changes of the affected quarter, such as swelling, pain and redness. If the onset of the mastitis is rapid, then it is termed acute mastitis. In contrast, in subclinical mastitis, there are no apparent signs of inflammation with no visible changes in the udder and the milk, except for a possible decrease in milk production and an increase in somatic cell count (SCC). If the subclinical mastitis persists for longer than two months, which is common with *S. aureus* intramammary infections (IMIs) it is termed as chronic mastitis (Erskine, 2020). Once *S. aureus* establishes the infection in the udder, it may persist for an entire lactation or even for the lifetime of the cow.

1.2.2 PATHOGENESIS OF *S. AUREUS* MASTITIS

Most of the knowledge about the pathogenesis of *S. aureus* IMIs are extrapolated from *in vitro* studies. Moreover, many of the studied virulence factors are coded in mobile genetic elements. Thus their expression varies across strains.

S. aureus is principally a contagious mastitis pathogen; i.e. the udder is the main reservoir of the bacteria, and transmission between quarters and between cows generally occurs during milking, and infection of the udder from the environment is considered rare (Contreras and Rodríguez, 2011; Rainard et al., 2018).

There have been several detailed reports of the pathogenesis of *S. aureus* in bovine mastitis over the years (Sutra and Poutrel, 1994; Middleton, 2008; Rainard et al., 2018; Cote-Gravel and Malouin, 2019). There are recognised major phases involved in the pathogenesis of *S. aureus* IMIs: 1) adhesion to the host mammary gland cells; 2) invasion into the mammary gland tissue and its cells; and 3) evasion of the host immune system (Shinefield and Black, 2005).

1.2.2.1 ADHESION PHASE

The adhesion phase starts when *S. aureus* enters the teat canal from the teat orifice. Contamination of the teat orifice generally occurs at milking, either from a contaminated milking machine (usually a contaminated liner) or from the hands of the milker. Once in the teat canal, *S. aureus* probably spreads into the mammary gland either by progressive colonization, or more quickly, when it is forced up into the gland during milking as a consequence of changes in intramammary pressure (Dego et al., 2002). Once in the mammary gland, *S. aureus* adheres to receptors on the surface of the epithelial cell using one or more of the group of surface proteins collectively known as Microbial Surface Components Recognizing Matrix Molecules (MSCRAMMs) through a complex process (Castilho et al., 2017).

There are several barriers which *S. aureus* needs to overcome to enter the mammary gland (Figure 1.4). The teat orifice which is closed except at milking forms the first line of defence against mastitis pathogens. However, damage to the teat canal either due to improper manual milking or faulty milking machines can lead to

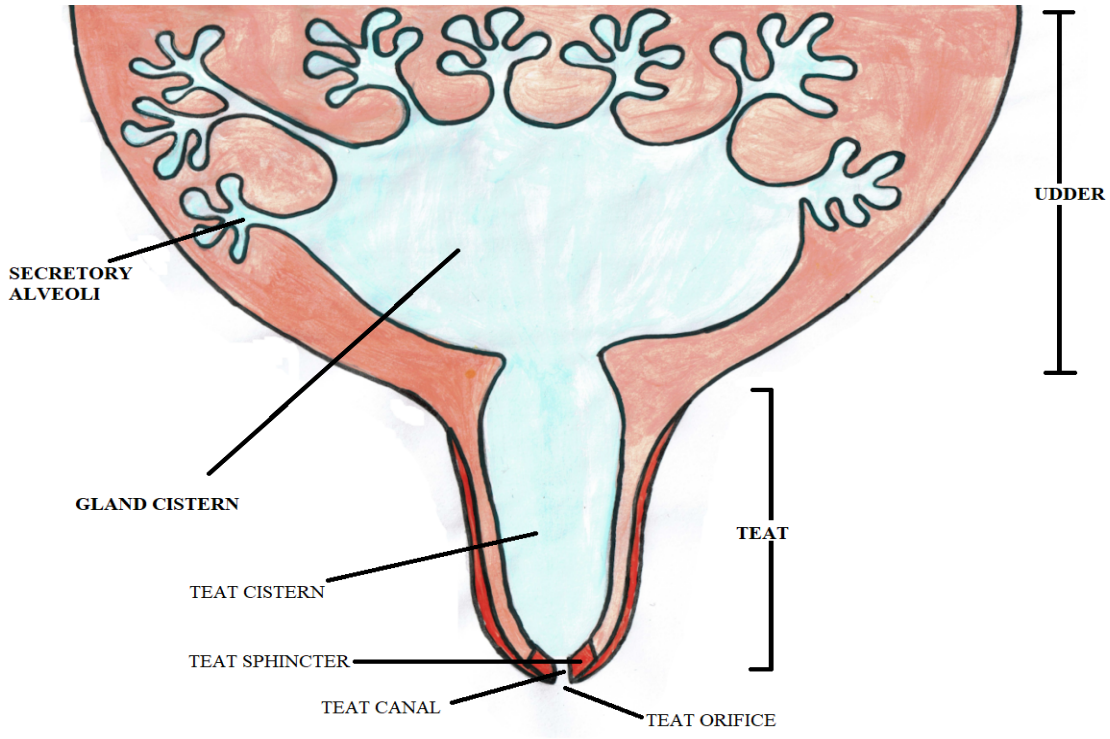

Figure 1.4: Schematic representation of the mammary gland of cattle. Adapted from Blowey (2010)

hyperkeratosis of the teat canal. This in turn allows the easier entry of bacteria and it also becomes a reservoir (Seykora and McDaniel, 1985). The next line of defence is the teat canal which is lined with folds of keratinized epidermis just as the teat skin with a thin lining of lipid film has bacteriostatic activity. When the sphincter muscles of the teat canal close, the folds of the canal form a tight seal to prevent the entry of bacteria. However, the folds trap some milk droplets which may contain bacteria. If the milk is not flushed out before the next milking, the bacteria gain entry deeper into the mammary gland. Damage to the teat canal and decreased frequency of milking may facilitate the entry of *S. aureus* into the mammary gland (Blowey, 2010).

Once *S. aureus* gains access into the mammary gland, the ability of the strains to adhere to the intact mammary epithelium may play a part in establishing an IMI (Baselga et al., 1994). The adhesion is believed to be facilitated by the expression of virulence factors such as fibronectin-binding protein, fibrinogen binding protein, clumping factors (Clf) A and B, teichoic acids and collagen-binding proteins (Cna) (Brouillette et al., 2003; Foster et al., 2014; Ashraf et al., 2017). These factors belong to a set of proteins known as MSCRAMMs. The fibronectin-binding proteins (FnBP),

FnBP-A and FnBP-B encoded by *fnb-A* and *fnb-B* genes, respectively, play a vital role in the adherence of *S. aureus* to bovine epithelial cells as demonstrated through *in vitro* studies (Dziewanowska et al., 1999; Hensen et al., 2000). The FnBPs are glycoproteins that act as bridging molecules, binding *S. aureus* to the fibronectin receptor, the integrin-$\alpha 5 \beta 1$ of mammalian cells (Patti et al., 1994; Sinha et al., 1999). Clumping factors (ClfA and ClfB) and fibrinogen binding protein (Efb) and collagen binding protein (Cna) enable *S. aureus* to adhere to fibrinogen and collagen in the tissue matrix (Mamo et al., 1988). Recently, ClfA has been identified to bind to the AnnexinA2 protein on the bovine mammary epithelial cells just prior to the invasion of *S. aureus* into those cells (Ashraf et al., 2017). AnnexinA2 belongs to the annexin family, calcium-dependent phospholipid proteins that are involved in various cellular activities such as cell proliferation, metabolism, apoptosis, exocytosis, endocytosis and invasion (Gerke and Moss, 2002).

Another important factor involved in the adhesion and colonization of the bacteria in the mammary gland is biofilm production (Otto, 2013). Biofilms are microbial communities adhering to either a biological or non-biological surface through the production of an extracellular polymeric matrix along with various host factors (Hall-Stoodley et al., 2004; Lister and Horswill, 2014). The biofilms in *S. aureus* are encoded by the intercellular adhesion (ica) gene cluster that contains *icaA*, *icaB*, *icaC*, and *icaD* genes, and the biofilm-associated protein (*bap*) gene. *S. aureus* can produce biofilms *in vitro* even in the absence of one of the above-mentioned genes (Notcovich et al., 2018). The bacteria that are present within the biofilms are better able to evade host immune response and antimicrobial penetration (Stewart and William Costerton, 2001; Hathroubi et al., 2017). The biofilms, and the other adhesion factors mentioned above are believed to prevent the flushing of *S. aureus* by milk flow (Dego et al., 2002).

1.2.2.2 INVASION PHASE

Once adhered to the bovine mammary epithelial cells and the tissue matrix, *S. aureus* invades into the mammary tissue through the use of exotoxins (e.g. haemolysins and leukocidins) and exoenzymes (e.g. proteases, hyaluronidases, coagulases, and lipases) (Dego et al., 2002). These degrade the mammary epithelial cells exposing their basal membranes and extracellular matrix. *S. aureus* then adheres to these using adhesin

factors (Anderson, 1976). Some of the adhesin factors such as FnBP is also involved in the invasion phase of *S. aureus* (Lammers et al., 1999).

Haemolysins such as α-haemolysin and β-haemolysin are pore-forming toxins encoded by the *hla* and *hlb* genes, respectively. The α-haemolysin attaches to the host cell membrane through ADAM10, a transmembrane protein. By attaching to this protein, α-haemolysin forms a transmembrane channel leading to increased cell permeability and, subsequently cell death (Ezekwe et al., 2016; Oliveira et al., 2018). The role of β-haemolysin in the pathogenesis of mastitis is not yet fully understood.

Leukocidins are bicomponent pore-forming toxins. They are secreted as two monomers (intially nominated as the F and S-subunits) which could be separated *in vitro* based on their liquid chromatography elution speed (Woodin, 1960).

Multiple different leukocidins are produced by some *S. aureus* strains, including the Panton-Valentine leukocidins (LukS-PV and LukF-PV in human strains), LukMF' (LukM and LukF'-PV), Luk E/D (LukE and LukD) and γ-hemolysin (HlgA and HlgB, HlgC and HlgB) (Alonzo and Torres, 2014; Reyes-Robles et al., 2016). Among these, LukM and LukF'-PV are found in some *S. aureus* strains isolated from bovine mastitis (Vrieling et al., 2016) while Luk E/D have been found in more than 90% of bovine *S. aureus* isolates (Fueyo et al., 2005; Yamada et al., 2005; Haveri et al., 2007). The principal role of leukocidins is to destroy leukocytes, but their host range varies. For example, LukM has a high level of activity against bovine neutrophils and macrophages (Vrieling et al., 2015b), but not for human neutrophils, while in human and mice models, LukE/D has been shown to have lytic activity against neutrophils, macrophages, T-cells, dendritic cells, and natural killer cells (Alonzo et al., 2013; Reyes-Robles et al., 2013). While HlgAB and HlgCB have lytic activity against host immune cells and also somatic cells in animal models (Siqueira et al., 1997; Spaan et al., 2014), its effect has not been studied on bovine mammary cells.

The role of *S. aureus* exoenzymes in the bovine mammary gland has not yet been fully elucidated. However, the genes for serine proteases, serine protease-like proteases, staphylococcal enterotoxins have been identified to a greater extent in strains causing bovine mastitis than in human strains, and it is thought that their role is to effect the breakdown of host cells to promote invasion and destruction of the bovine mammary epithelium (Magro et al., 2017).

1.2.2.3 EVASION PHASE

The evasion of host immunity plays an important role in the survival and persistence of *S. aureus* in bovine mammary glands leading to chronic and subclinical forms of mastitis. The bovine immune response is characterised initially by a neutrophil response, followed by the production of opsonizing antibodies and the activation of the complement system (Burton and Erskine, 2003). Part of the resistance against the host defences is determined by bacterial hemolysins and leukocidins that destroy neutrophils by altering their cell permeability (Vandenesch et al., 2012). However, in addition, *S. aureus* also produces surface exopolymers, such as the capsular polysaccharides and poly-N-acetyl-β-1,6 glucosamine (PNAG) found in biofilms, that resist phagocytosis (Kampen et al., 2005; Vandenesch et al., 2012; Grunert et al., 2018). Even when ingested by phagocytic cells, *S. aureus* strains can survive in the host cell cytoplasm and replicate by forming small colony variants within the phagocytic cells (Gresham et al., 2000; Hébert et al., 2000). The small colony variants are slow-growing bacterial subpopulations with distinguishing phenotypic and pathogenic properties (Proctor et al., 2006). The small colony variants evolve by mutations in the metabolic genes during host immune response and antimicrobial stress and even though they are less virulent, they can survive and replicate within host cells (Melter and Radojevič, 2010). The small colony variants have been attributed to persistence of *S. aureus* even with antimicrobial treatments leading to chronic IMIs (Atalla et al., 2008; Brouillette et al., 2004).

1.2.2.4 REGULATION OF VIRULENCE FACTORS

The expression of many virulence factors is regulated by global regulatory genes, such as the accessory gene regulator (*agr*), transcriptional regulator *sarA* and transcriptional sigma factor (*sigB*) (Le and Otto, 2015).

The *agr* operon appears to play a central role in gene regulation during the transition from the adhesion and colonization phase, to the invasion phase (Guldimann et al., 2016). As bacteria colonise the udder and the density of bacteria increases, Agr downregulates the production of adhesions proteins and upregulates the expression of genes coding for hydrolytic enzymes, exotoxins, and detachment from the biofilms (Otto, 2013). Agr has been identified to regulate the internalisation of the bacteria

into the host cells, especially bovine mammary epithelial cells, and regulates the induction of apoptosis of those cells (Wesson et al., 1998). The *agr* operon is formed by four genes *agrA*, *agrB*, *agrC* and *agrD* which encode for the AgrA, AgrB, AgrC and AgrD proteins, respectively. *S. aureus* isolates are classified into four Agr-groups (I, II, III and IV) based on the sequences of the *agrC* and *agrD* genes. Association between agr type and the manifestation of diseases by *S. aureus* has been reported in humans, with type III involved in non-invasive diseases, type I in invasive diseases and type IV identified in exfoliatin-producing strains (Jarraud et al., 2000; Ben Ayed et al., 2006). The CCs of human associated strains often predicts its agr-type, but such an association has yet to be established for bovine strains (Monecke et al., 2008).

The functioning of the Agr proteins is a complex process. It starts the activation of AgrD and AgrB, which combine to induce the secretion of auto-inducing peptide (AIP). Once it reaches a critical concentration, AIP activates the two-component system AgrC and AgrA, which in turn upregulates the expression of a regulatory RNA (RNAIII). RNAIII reduces the expression of surface proteins, represses adhesins, and promotes the production of alpha-haemolysin (HLA) (Morfeldt et al., 1995; Novick and Geisinger, 2008).

The *sarA* gene is involved in the upregulation of fibronectin-binding proteins, haemolysins, proteases and the downregulation of protein A, lipase, collagen adhesin, aureolysin. It is also involved in the regulation of biofilm formation by enhancing the transcription of the *ica* genes and *bap* genes (Blevins et al., 1999; Valle et al., 2003; Shaw et al., 2004; Liu et al., 2006). Along with Agr, SarA is necessary for the activation of RNAIII (Chien and Cheung, 1998).

SigB regulates the expression of virulence factors in response to stressors such as heat shock, osmotic shock, and starvation (Kullik and Giachino, 1997; Kullik et al., 1998). SigB, along with Agr also regulates the activation of SarA (Deora et al., 1997). SigB also protects *S. aureus* against oxidative damage by the neutrophils by upregulating the expression of catalase (Kullik et al., 1998).

As described above, the regulation of virulence factors during the pathogenesis of *S. aureus* requires complex, poorly understood multi-level interactions.

1.2.3 DIAGNOSIS OF *S. AUREUS* MASTITIS

Due to the non-pathognomonic clinical signs of *S. aureus* mastitis, the aetiological diagnosis is performed through the detection of the organism in milk samples collected from the mammary gland. The detection of the organism can be achieved either by bacterial isolation by culture, by the use of molecular methods such as PCR (Khan et al., 1998; Oliver and National Mastitis, 2004), or commercial antigen detection systems. Isolation by culture remains the reference standard against which the other techniques are often compared.

Various culture media such as Baird Parker agar, Mannitol salt agar, specialized chromogenic agar have been and are used to isolate and identify *S. aureus*. However, Columbia blood agar (CBA) containing 5% sheep blood appears to be the commonest bacteriological medium used in New Zealand diagnostic laboratories to culture bovine milk samples for common pathogens. *S. aureus* present in milk grows readily on CBA and its colonies can usually be visualized after 24 hours of incubation aerobically at 35-37°C, as round, convex, golden-white colonies, often surrounded by a clear or double-zone haemolysis observed after two days of incubation. However, some variants of *S. aureus* produce white colonies and some do not produce a zone of clear haemolysis (Zhang et al., 2016).

Even though standard aerobic culture is the reference standard for aetiological diagnosis, the cost of testing and shipment, and the extended turnaround time remain major limitations of culture in the field. These limitations have been overcome with the advent of on-farm culture methods (Mansion-de Vries et al., 2014; Royster et al., 2014; Ferreira et al., 2018; Malcata et al., 2020), which provide results often reported to be comparable (accuracy 94%) to those obtained by standard aerobic culture methods (Ganda et al., 2016). A major limitation in isolation of *S. aureus* by culture is the intermittent shedding of the bacteria in milk, which might lead to false-negative results (Middleton et al., 2017). However, this limitation applies to any bacterial detection method.

The final identification of *S. aureus* usually requires the use of biochemical tests (see section 1.1.2) . Bacterial identification to the species level is being revolutionised by the introduction of Matrix-Assisted Laser Desorption/Ionization Time-Of-Flight Mass Spectrometry (MALDI-TOF MS) into diagnostic routines. This approach reduces the

cost of consumables and time spent on diagnostics. Direct MALDI-TOF MS (without prior isolation of bacteria) identified only 14.2% of S. aureus when compared to the standard microbiological culture methods (Barreiro et al., 2018). Hence, the use of MALDI-TOF MS requires prior isolation of the pathogen.

Another method of identification of *S. aureus* is via PCR amplification of *S. aureus*-specific genes. Three genes are widely used for this purpose: the thermonuclease (*nuc*) gene, the coagulase (*coa*) gene, and the adhesion-like bovine protein (*adlb*) gene (Kim et al., 2001; Ahmadi et al., 2010; Sartori et al., 2017). There are commercially available PCR-based kits which identify *S. aureus:* VetMax MastiType multiplex qPCR kit (ThermoFisher Scientific, Massachusetts, USA), and Mastit 4 (DNA Diagnostic, Risskov, Denmark). Compared to culture-based methods these tests have a much faster turnaround time (2-3 hours vs 24-48 hours), but they have a higher cost, do not differentiate between viable and non-viable organisms, and can be inhibited by compounds present in the milk or colostrum (Steele, 2015; Nyman et al., 2016).

Immunoassays such as the Enzyme-Linked Immunosorbent Assay (ELISA) using *S. aureus* specific antigens, antibodies or biomarkers have been used in the diagnosis of *S. aureus* mastitis (Yazdankhah et al., 1998; Fox and Adams, 2000; Jaeger et al., 2017). The immunoassays that identify antigens detect the presence of *S. aureus* in the milk while the immunoassays that detect antibodies or biomarkers indicate the *S. aureus* infection but don't guarantee the presence of bacteria in milk or in the mammary gland. There are commercially available ELISA kits for the detection of *S. aureus* in milk, such as the *S. aureus* antibody test kit (SAATK, Veterinary Medical Research and Development, Washington, USA) and the Koru StaphGold ELISA test (Koru Diagnostics Ltd, Palmerston North, NZ). Immunoassays are useful in virtue of their speed, low cost and simplicity. However, immunoassays have limitations such as a perceived relatively low specificity due to cross-reactions leading to false positives, and the false negatives due to a delayed antibody response in milk during the infection (Fox and Adams, 2000; Duarte et al., 2015). The big advantage of ELISA antibody tests is that they don't require aseptic collection of milk and the testing can therefore be integrated in herd testing protocols. It should be emphasized that the culture-free methods such as ELISAs and DNA-based methods do not conduce to the creation of a clinical isolate, and as such, do not consent antimicrobial susceptibility testing.

DNA-based typing methods such as MLST, spa-typing and PFGE (Section 1.1.4) are not usually implemented in diagnostic laboratories but may sometimes be required, for instance, to identify the endemic strains and to differentiate sporadic strains from the outbreak strains in big herds. Lastly, whole-genome sequencing (WGS) described in Section 1.1.5 has been increasingly used in understanding the relationship between the strain of bacteria and the pathogenicity and the geographical distribution of strains of *S. aureus* causing bovine mastitis. WGS is not yet used in the detection of milk from *S. aureus* but has been used as a frontline tool to identify an outbreak of *S. aureus* especially MRSA, new antimicrobial-resistant strains in human medicine (Price et al., 2013).

1.2.4 PRINCIPLES OF TREATMENT AND PREVENTION OF *S. AUREUS* MASTITIS

1.2.4.1 TREATMENT METHODS

Antimicrobial therapy is the common treatment for *S. aureus* mastitis (Burgess and French, 2017; Cheng and Han, 2020). Intramammary infusions and systemic antimicrobials are used to treat *S. aureus* (Hossain et al., 2017). The use of antimicrobials for the treatment of bovine mastitis is low in New Zealand compared to other countries, and penicillin is the most common drug used (Bryan and Hea, 2017). Antimicrobial therapy for *S. aureus* mastitis produces highly variable cure rates reportedly ranging from 4% to 92%. This variability is believed to be due to various factors including host and pathogen factors, herd transmission rates, and duration of treatment (Ziv and Storper, 1985; Owens et al., 1988; Sol et al., 1994; Wilson et al., 1995; Sol et al., 1997; Deluyker et al., 2005). The host factors affecting treatment outcomes include the age of the cow, the levels of SCC, infection duration, the bacterial load in the milk before treatment, and the number of infected quarters (Sol et al., 1997; Barkema et al., 2006). The pathogen factor that has been studied in detail is the antimicrobial resistance of the infecting strain (Barkema et al., 2006). Studies done on the effect of strain variation suggested that isolates belonging to CC97 and other bovine-associated CCs had a lower response to treatment compared to non-bovine associated strains (Van Den Borne et al., 2010) which might be explained by the host adaptation of those strains (Budd et al., 2015).

The success rate of antimicrobial treatment against bovine mastitis does not completely correlate with *in vitro* susceptibility of the strains. Clinical failure of antimicrobial treatment was found to be significantly higher in *S. aureus* infected quarters when compared to other mastitis-causing pathogen-infected quarters (McDougall et al., 2007, 2019). The ability of *S. aureus* to survive in host cells such as mammary epithelial cells and neutrophils, the formation of small colony variants in the host cells, the production of biofilms, and the formation of microabscesses in the mammary gland reduces the efficacy to antimicrobial treatment (Owens, 1987; Proctor et al., 1995; Kahl et al., 1998; Valle et al., 2003; Melchior et al., 2006).

Dry cow therapy (DCT) can also be considered a form of prevention as it eliminates persisting IMI from previous lactations but also prevents any new IMI (Berry and Hillerton, 2002; Green et al., 2002; Huxley et al., 2002). Infusion of long-acting antimicrobial into the teats at the end of lactation is a cost-effective method (Scherpenzeel et al., 2018). Limitations of DCT are the presence of antimicrobial residues in the milk and the selective pressure exerted, with potential of a consequent development of antimicrobial resistance (Bachmann et al., 2018). The limitations of antimicrobial residues can partially be avoided by applying strict withholding periods, and development of robust antimicrobial stewardship, such as selective application of DCT for infected quarters (Rindsig et al., 1978).

Newer modalities being studied for the treatment of *S. aureus* mastitis are the use of antimicrobial peptides (AMPs) (Li et al., 2017) and bacteriophages (Breyne et al., 2017). Two AMPs used by Li et al. (2017) were found to decrease the *S. aureus* concentration both in *in vitro* and *in vivo* animal models, with an activity comparable to tetracycline antimicrobials. However, the AMPs cleared the *S. aureus* from bovine mammary epithelial cells more efficiently compared to tetracycline in the *in vitro* studies. The study by Breyne et al. (2017), used a combination of bacteriophages against *S. aureus* which was found to reduce the *S. aureus* numbers *in vitro,* but cleared less bacteria than antimicrobial treatment in a mouse mastitis model.

1.2.4.2 PREVENTATIVE METHODS

Preventative methods have been formulated to prevent the spread of infection and to remove the infection from the herd. Hygienic procedures include frequent wearing

and changing gloves by milkers, wiping off dirt from teats, dipping teats with disinfectant solutions with recommended contact time pre- and post-milking and checking of teats for cracks or lesions which may harbour bacteria (Keefe, 2012). Separating, milking last, or culling clinically infected cows reduced the prevalence of *S. aureus* mastitis in the herd (Wilson et al., 1995). However, separation or culling would not be possible in unidentified subclinical mastitis.

Problems with milking machines such as vacuum fluctuations, liner slips, flooded lines can cause backflow of milk against the teat which can cause the entry of bacteria into the teats. Liner slip alone has been found to facilitate 10-15% of new mastitis cases (O'Shea et al., 1984; Baxter et al., 1992). Hence, regular maintenance of milking machines is recommended to prevent the spread of *S. aureus* bovine mastitis in the herd.

Vaccination against *S. aureus* may represent a viable solution due to the undesirable public health impact of the use of antimicrobials. With the ever increasing trend of antimicrobial resistance in the world, there have been calls for reducing antimicrobials in the food and dairy industry. Antimicrobial resistance in *S. aureus* will be detailed in section 1.3 of this thesis, and the available vaccines for bovine mastitis (*S. aureus* in particular), their efficacy and drawbacks will be discussed in section 1.4.3.

1.3 ANTIMICROBIAL RESISTANCE IN *S. AUREUS*

1.3.1 INTRODUCTION

Antimicrobials are compounds that kill or inhibit the growth of microorganisms. The classification includes antiseptics, disinfectants, antibacterials, antifungals, and antivirals. Antibacterials are natural or synthetic compounds that act selectively to destroy or inhibit the growth of bacteria with little or no effect on the host (Varley et al., 2009).

The first antibacterial compound to be identified and developed for that purpose was sulfamidochrysoïdine (Prontosil) which was first used in the mid-1930s. This was followed by the development of penicillin (the first antimicrobial) in the 1940s

(Aminov, 2010). These discoveries have been followed by the development of a large number of antibacterials some of which were adaptations of the original products (e.g. sulphonamides and semi-synthetic penicillins) and some were new groups altogether (e.g. macrolides and fluoroquinolones). Antibacterials are usually classified based on their chemical class and they can also be grouped according to their mode of action. This grouping (Table1.9) is useful for understanding both how antimicrobials work and also how resistance develops.

Table 1.9: Classification of antibacterials based on their mode of action along with relevant veterinary examples

Mode of action	Chemical classes	Examples
Act on the bacterial cell wall	β-lactams, Cephems, carbapenems, monobactams, clavams, Glycopeptides	Penicillin, Amoxicillin, Ceftiofur, Cloxacillin
Inhibit protein synthesis at ribosomal 30S subunit	Tetracyclines, Aminoglycosides, Aminocyclitols	Tetracycline, Oxytetracycline, Neomycin, Streptomycin
Inhibit protein synthesis at ribosomal 50S subunit	Lincosamides, Macrolides, Pleuromutilins, Streptogramins, Chloramphenicol	Clindamycin, Tylosin, Lincomycin
Inhibit protein synthesis by other means	Pseudomonic acid	Mupirocin, Fusidic acid
Target nucleic acid biosynthesis	Quinolones, Nitroimidazole	Enrofloxacin, Marbifloxacin, Metronidazole
Target other metabolic pathways	Folate pathway antagonists, RNA polymerase antagonists, Fatty acid synthesis antagonists	Sulphamethoxazole, Trimethoprim, Rifampicin, Triclosan

When antimicrobials were first introduced, *S. aureus* was found to be very susceptible to these compounds. However, within 3 years of the first use of penicillin against *S. aureus* in 1941 (Abraham et al., 1941), studies were reporting that more than 10% of the clinical strains of *S. aureus* were penicillin-resistant (Barber and Rozwadowska-Dowzenko, 1948).

1.3.2 *S. AUREUS* AND ANTIMICROBIALS: MECHANISMS OF ACTION AND MECHANISMS OF RESISTANCE

Since these first penicillin-resistant strains of *S. aureus* were reported, the organism has continued to develop resistance against most of the antibacterial groups in use to treat *S. aureus* infections. In order to provide sufficient background to the study presented in Chapter 5, this section aims to provide an overview of the mode of action of each antibacterial group (Table 1), and the antimicrobial resistance mechanisms developed by *S. aureus*, with particular focus on antibacterials used in veterinary medicine.

1.3.2.1 ANTIMICROBIALS THAT TARGET THE CELL ENVELOPE

1.3.2.1.1 βLACTAMS β-lactams are a class of antibacterial drugs that include a β-lactam ring in their structure. This class includes penicillin and penicillin derivatives (penams), cephalosporins (cephems), carbapenems, carbacephems, and monobactams. Penicillin was first used in cattle to treat mastitis in the UK in 1943 (Woods, 2014). Penicillin and other β-lactams remain the most commonly used antimicrobials in cattle in New Zealand, especially for the treatment of mastitis in lactating cows and dry cow therapy (MPI, 2019).

The main target of β-lactam antimicrobials is the final step in the biosynthesis of cell wall peptidoglycans. This step is the transpeptidation of the peptidoglycan, which is catalysed by transpeptidases such as the D-alanyl-D-alanine (D-Ala-D-Ala) carboxypeptidase. Penicillin and other β-lactams bind to these transpeptidases because of the similarity of the β-lactam ring to the backbone of the D-Ala-D-Ala (Scheffers and Pinho, 2005; Kong et al., 2010). This binding inhibits the cross-linking of the peptidoglycan disrupting the normal formation and maintenance of the bacterial cell wall.

The transpeptidases which catalyse transpeptidation are not the only bacterial proteins that have an affinity towards β-lactams. There is a whole range of penicillin-binding proteins (PBPs) in the bacterial cell, which vary in their molecular weight, and function, as well as their affinity towards β-lactams (Kong et al., 2010). The number of distinct PBPs varies across bacterial species. In *S. aureus,* four different intrinsic

PBPs have been identified: PBP1, PBP2, PBP3, and PBP4 (Georgopapadakou and Liu, 1980). Each of these PBPs has a different function related to cell wall formation, therefore, the inhibition of each PBP will produce a different effect on the bacterial cell. For example, inhibition of PBP2 reduces the elongation of the peptidoglycan leading to leakage of the cytoplasmic contents and cell lysis, while inactivation of PBP3 results in abnormally sized and shaped bacteria (Georgopapadakou et al., 1986).

PBP2 is the major target for the β-lactam antimicrobials (Pinho et al., 2001a,b). It is the only PBP in *S. aureus* which functions as both a transpeptidase and a transglycosylase, and is essential for growth, survival, and cell wall biosynthesis.

One of the mechanisms by which β-lactam-resistant organisms develop resistance to β-lactam antimicrobials is the production of β-lactamase. These enzymes hydrolyse the β-lactam ring thereby preventing the antimicrobial from binding to PBPs (Hartman and Tomasz, 1984; Lingzhi et al., 2018). β-lactamase is a lipoprotein located on the outer surface of the cytoplasmic membrane and also secreted into the surrounding environment (Nielsen et al., 1981).

Bacterial β-lactamases can be classified into two groups based on the mode of hydrolysation of the β-lactam ring: i) using an acyl-enzyme with an active-site serine; or ii) using zinc ions as a co-factor (Metallo-β-lactamases (MBLs)). Active serine β-lactamases can be further classified based on their amino acid profiles into class A, C, and D β-lactamases, whereas class B β-lactamases are those that use Zn ions. More recently, 17 functional groups of β-lactamases have been identified based on molecular and functional characteristics. These groups differ in their substrates (e.g. penicillins vs. cephalosporins) and the molecules that can inhibit them (clavulanic acid, avibactam) (Bush, 2018).

In *S. aureus,* the *blaZ* gene encodes for a serine-based β-lactamase that forms an acyl-enzyme complex which is similar to the acyl-PBP2 complex formed by the transpeptidase of the PBP2 during cell wall synthesis (Lowy, 2003). The *blaZ* gene is carried by transposon Tn552 or Tn552-like elements and is present on large plasmids such as pl258, as well as integrated into the chromosome (Rowland and Dyke, 1989, 1990). The regulatory genes for *blaZ* (*blaR1* and *blaI*) are located upstream (towards the 5' end) of *blaZ* and are transcribed in the opposite direction. The *blaR1* and *blaI* upregulate and repress the expression of the *blaZ* gene, respectively (Hackbarth and Chambers, 1993).

The importance of β-lactamase production as a means of bacterial resistance to β-lactams resulted in the development of β-lactamase inhibitors which inactivate the β-lactamases by competitive and irreversible inhibition. Two groups of β-lactamase inhibitors have been developed: 1) those that contain a β-lactam ring (clavulanic acid, sulbactam, tazobactam) and 2) those which do not (avibactam) (Essack, 2001; Stachyra et al., 2009; Castanheira et al., 2012). These inhibitors have weak or no antibacterial activity but act by protecting the β-lactam antimicrobials against hydrolysis.

1.3.2.1.2 ISOXAZOLYL PENICILLINS Penicillinase-resistant β-lactam antimicrobials such as isoxazolyl penicillins, including methicillin, cloxacillin, oxacillin, and flucloxacillin, were introduced to counter penicillinase-producing *S. aureus*. Adding the isoxazolyl side chain prevents the acid hydrolysis by the β-lactamases produced by Staphylococcal species. Therefore, this group of antimicrobials is also referred to as anti-staphylococcal penicillins (Marcy and Klein, 1970). Methicillin, the first isoxazolyl penicillin to be developed, was firstly used against penicillin-resistant *S. aureus* in 1959 (Enright et al., 2002). However, resistance against the drug was reported as early as 1961 in the UK (Barber, 1961), and the first outbreak of methicillin-resistant *S. aureus* (MRSA) was documented in the USA in 1968 (Barrett et al., 1968). Strains of *S. aureus* resistant to isoxazolyl penicillins and cephalosporins continue to be referred to as MRSA, even though cloxacillin, oxacillin and flucloxacillin are now used much more commonly in clinical practice than methicillin.

The first documented MRSA in livestock was reported in cattle by Devriese et al. (1972) and potential transmission of MRSA from livestock to humans has since been reported (Lee, 2003; Voss et al., 2005).

S. aureus resistance to methicillin and other isoxazolyl penicillins is due to an acquired gene that codes for a penicillin-binding protein, PBP2A (Hartman and Tomasz, 1984). This PBP is not susceptible to the isoxazolyl penicillins because its active-site serine is positioned deep within the protein (Lim and Strynadka, 2002). PBP2A can take over the transpeptidase activity of other PBPs, in the presence of β-lactam antimicrobials, but does not have the transglycosylase activity. Therefore, in MRSA, peptidoglycan biosynthesis for the bacterial cell wall takes place through the

cooperative action of PBP2A and PBP2, creating a high-level of resistance to β-lactam antimicrobials.

The acquired gene *mecA* encodes for PBP2A and is located in the Staphylococcal Chromosomal Cassette mec (SCCmec) along with a family of related elements (Jensen and Lyon, 2009; Malachowa and DeLeo, 2010; Liu et al., 2016). SCCmec may also contains genetic elements that encode resistance against non-β-lactam antimicrobials (Ito and Hiramatsu, 1998). In 2011, a homologue of *mecA*, initially named $mecA_{LGA251}$ but, now known as *mecC*, was found in dairy cattle and shown to cause methicillin resistance (Garcia-Alvarez et al., 2011). The *mecC* gene was only ~69% identical to the conventional *mecA* gene and the protein encoded by the *mecC* gene was ~63% identical to PBP2A (Garcia-Alvarez et al., 2011).

The expression of *mecA* is regulated by two regulatory genes *mecRI* and *mecI* located upstream to the *mecA* gene in the SCCmec complex and the expression and location of *mecA* is similar to the β-lactamase genes *blaZ* and its regulatory genes (Hiramatsu et al., 1992). The expression of methicillin-resistance in *S. aureus* is found to be usually heterogeneous where a small population shows high-levels of resistance to the drug while the larger population displays low to no resistance (Finan et al., 2002). In *S. aureus* with chromosomal mutations in genes such as *relA*, *rpoB*, *rpoC* which are involved in the bacterial stress response, the resistance to methicillin is homogenous (Dordel et al., 2014).

Recent research has found more novel β-lactams such as ceftaroline have alternate binding sites such as the allosteric site of PBP2A in *S. aureus* which helps in countering the resistance caused by MRSA (Otero et al., 2013; Gonzales et al., 2015).

To date, there has been only one case of MRSA reported from cattle in New Zealand, but the result was clinically uncertain due to the mixed nature of the cultured isolate (Grinberg et al., 2005). Around 20% of MRSA has been reported in humans in New Zealand (Heffernan and Bakker, 2018)

1.3.2.1.3 GLYCOPEPTIDES Glycopeptides are another group of antimicrobials that act on the bacterial cell wall. The most commonly used glycopeptide antimicrobial is vancomycin, which is used widely to treat MRSA infections in humans (Deresinski, 2009). Semisynthetic lipoglycopeptides related to vancomycin,

such as oritavancin and telavancin, are approved by the FDA (United States Food and Drug Administration) for use against *S. aureus* infections (Crotty et al., 2016).

The mode of action of vancomycin is similar to the β-lactams in that it binds to the D-Ala-D-Ala dipeptide, thereby preventing the cross-linking of peptidoglycans by the PBPs. The semisynthetic lipoglycopeptides also act on the lipid bilayer of the cell membrane depolarizing the cells causing membrane damage (Damodaran and Madhan, 2011).

In *S. aureus*, high-level resistance to vancomycin (MIC $\geq$ 16 μg/ml) is mediated by the alteration of the terminal peptide of the peptidoglycan from the D-Ala-D-Ala dipeptide to D-alanyl-D-lactate (D-Ala-D-Lac) depsipeptide. This modification can occur when the bacteria are exposed even to low concentrations of the drug (MIC=8 μg/ml) (Gonzalez-Zorn and Courvalin, 2003). This alteration is encoded by the *vanA* operon, which was acquired through conjugation from vancomycin-resistant *Enterococcus faecalis* (Noble, 1992; Showsh et al., 2001). Another potential mechanism of low levels of resistance (MIC 8-16 μg/ml) or reduced susceptibility to vancomycin in *S. aureus* (Hiramatsu, 1997) is by trapping or altering the passage of the drug through the cell wall. This type of resistance is seen in heteroresistant strains known as hetero-VRSA (hetero-vancomycin resistant *S. aureus*) or VISA (Vancomycin intermediate *S. aureus*). These strains are predominantly susceptible to vancomycin but contain resistant subpopulations (Hiramatsu, 2001). This resistance is mediated by increasing the cell wall thickness, reduction in peptidoglycan cross-linking and the overproduction of D-Ala-D-Ala residues (the substrate for vancomycin) during cell wall formation. This prevents vancomycin from entering the cell by diffusion by increasing the concentration of the drug at the cell wall (Cui et al., 2003, 2006), thereby safeguarding the bacterial cell against a low concentration of the drug (Hanaki et al., 1998a,b). This mechanism of resistance is encoded by the two-component regulatory system, such as *walKR* expression when the bacteria has prolonged exposure to vancomycin (McEvoy et al., 2013).

Glycopeptides products are not approved for use in livestock in New Zealand with Avoparcin banned in the use of poultry as a prophylactic growth promoter in 2000 (MPI, 2016; Manson et al., 2004).

1.3.2.2 ANTIMICROBIALS THAT INHIBIT PROTEIN SYNTHESIS

Antimicrobials that inhibit bacterial protein synthesis act by either binding to the ribosomes, or by interfering with cytoplasmic proteins involved in protein synthesis. The ribosomal binding sites of antimicrobials are on one of two major sites: 1)the 16S rRNA of the 30S subunit at the A site, where the aminoacyl-tRNA binds and 2) the 23S rRNA of the 50S subunit around the peptidyl transferase and the polypeptide exit tunnel end (Chukwudi, 2016). Antimicrobials acting on the 30S subunit, such as tetracyclines and aminoglycosides (e.g. streptomycin, neomycin and gentamycin), and antimicrobials that act on the 50S subunit, such as macrolides, lincosamides, amphenicols and streptogramins are approved for use in veterinary practice in New Zealand, and will be discussed in detail in the following sections.

1.3.2.2.1 TETRACYCLINES Tetracyclines, discovered in the 1940s, are a class of broad-spectrum antimicrobials that contain a primary structure of four (tetra) hydrocarbon rings, either isolated from cultures of *Streptomyces* species or modified from these natural compounds (semi-synthetic). Tetracyclines have also been found to have immunomodulatory, anti-inflammatory and neuroprotective effects (Bahrami et al., 2012; Di Caprio et al., 2015). Tetracyclines were one of the top five antimicrobial classes sold for animal use in New Zealand in 2017 (MPI, 2019).

Tetracyclines inhibit protein synthesis by binding to the 16S RNA and S7 protein of the 30S ribosome, thereby preventing the binding of aminoacylated tRNA to its target site (A-site) on the ribosome. This mode of action means that tetracyclines are also effective on non-bacterial parasites that possess similar ribosomal subunits such as *Babesia* and *Theileria* species (del Castillo, 2013).

In *S. aureus*, resistance against tetracyclines occurs via two pathways: 1) via efflux pumps or 2) by ribosomal protection. Two active efflux pumps (TetK and TetL) have been identified in tetracycline resistant *S. aureus* strains (Chopra and Roberts, 2001). TetK is encoded by the *tetK* gene situated on plasmid pT181 but is also found in the chromosomal SCCmec of some MRSA strains (Li et al., 2011). The plasmid pT181 is a multicopy plasmid, and copies have been found integrated into other larger plasmids such as pJ3358, which contains mupirocin resistance genes (Roberts, 1996). The TetK and TetL proteins pump out the tetracycline molecule from within the cell at the

expense of a proton against a concentration gradient (Guay and Rothstein, 1993). The expression of *tetK* and *tetL* genes, unlike other *tet* efflux genes, are not regulated by a repressor protein, but by a process known as translational attenuation, where the process of translating the protein is disrupted by the attachment of the molecule (tetracycline) to the ribosome (Roberts, 1996).

S. aureus also protects the ribosomes by dislodging the tetracycline molecule from the A site of the ribosome (Roberts, 1996). This process is mediated by the binding of the TetO or TetM protein on the Elongation Factor-G (EF-G) binding site of the ribosome as these proteins have sequence similarity to the EF-G. This produces a structural change in the ribosome where the tetracycline molecule cannot remain bound (Li et al., 2013). The proteins TetO and TetM are encoded by *tetO* and *tetM* genes respectively which are located on the chromosomal conjugative transposons such as Tn916 and Tn1545 (Spahn et al., 2001). This allows protein synthesis in the bacterial cell even in the presence of tetracyclines.

1.3.2.2.2 AMINOGLYCOSIDES Aminoglycosides are potent, broad-spectrum antimicrobials particularly effective against *Enterobacteriaceae*, which have also been used against *S. aureus*. Streptomycin, the first aminoglycoside, was initially isolated in 1943 from the soil bacterium, *Streptomyces griseus*. Aminoglycosides include both natural products such as streptomycin, neomycin, gentamicin and semi-synthetic products such as amikacin.

In cattle, streptomycin combined with its derivative dihydrostreptomycin is still used to treat bacterial infections, but historically, the most common use of aminoglycosides in cattle has been in combination with other antimicrobials, especially penicillins. In New Zealand, aminoglycosides are available only as combination products for intramammary use (MPI, 2019).

Aminoglycosides inhibit protein synthesis by binding to the A-site of the 16S ribosomal subunit, causing misreading during translation. This produces faulty proteins, including membrane proteins, resulting in cell membrane damage (Davis et al., 1986). During bacterial protein synthesis, the ribosome matches aminoacyl-tRNAs anticodons with the correct mRNA codon present in the A-site of the small ribosomal subunit. This bound tRNA is stabilized by two 16S rRNA bases A1492 and A1493 which rotate out from the DNA helix structure (a process known as

flipping out of bases) and interact only with the correct codon-anticodon pairing (Laursen et al., 2005). In contrast, when an aminoglycoside binds with the 16S rRNA, the A1492 and A1493 bases stabilize incorrect codon-anticodon pairings (Demirci et al., 2013; Wilson, 2014).

Even though all aminoglycosides inhibit protein synthesis through this process, the mechanisms used can vary between compounds. For example, gentamicin stabilizes the incorrect codon-anticodon pairing by displacing the two 16S rRNA bases while streptomycin stabilizes the incorrect pairing by modifying the ribosomal conformation (Demirci et al., 2013).

Aminoglycoside resistance in bacteria is mediated by 1) ribosomal mutation, 2) ribosomal modification by methyltransferases and 3) aminoglycoside modifying enzymes (AMEs). *S. aureus* resistance to aminoglycoside is mainly mediated by AMEs. These AMEs can be classified into three categories, based on the chemical modification of the aminoglycosides: 1) aminoglycoside acetyltransferases (AACs); 2) aminoglycoside phosphotransferases (APHs); and 3) aminoglycoside nucleotidyltransferases (ANTs). Some AMEs are bifunctional, possessing activity in two of the three categories. One bifunctional enzyme AAC(6′)-Ie/APH(2″)-Ia is found in *S. aureus* and confers resistance against gentamicin and neomycin. The gene encoding this enzyme is found in the transposon Tn4001 (Daigle et al., 1999; Frase et al., 2012). Other enzymes such as *aphA* (APH(3′)III) are found in Tn5405, *aadD* (ANT(4′)-Ia) on the pUB110 plasmid for neomycin resistance (Derbise et al., 1997) and *aad*(6) or *str* gene on the plasmid pS194 (Projan et al., 1988).

1.3.2.2.3 MACROLIDES Macrolides include a large macrocyclic lactone ring to which deoxy sugars are attached (Mazzei et al., 1993). This chemical class include antibacterials, antifungals, prokinetics and immunomodulators. Erythromycin was the first macrolide antibacterial discovered in 1952, and was used principally as an alternative to penicillin in penicillin-allergic individuals and against penicillin-resistant infections. Commonly used macrolide antibacterials in veterinary medicine registered in NZ include erythromycin, oleandomycin, tylosin, spiramycin, tulathromycin, and tilmicosin with oleandomycin, tilmicosin, tulathromycin, and tylosin approved for use in cattle in New Zealand. Macrolides bind near the PTC at the entrance of the exit tunnel, where they block the progression of the nascent

peptide chain when it becomes around 3-9 amino acids long, thereby blocking translation (Tenson et al., 2003).

1.3.2.2.4 LINCOSAMIDES Lincosamides are a small class of antimicrobials containing lincomycin, clindamycin, and pirlimycin, which are produced by *Streptomyces* species. In NZ, lincomycin is available in an intramammary preparation for use against mastitis caused by Gram-positive organisms. The α-methylthiolincosamine (α-MTL) part of the lincosamide binds to the 23S rRNA of the large ribosomal subunit and dissociates peptidyl-tRNA chains when they contain 2-4 amino acids (Tenson et al., 2003; Matzov et al., 2017).

1.3.2.2.5 AMPHENICOLS Amphenicols form a class of broad-spectrum antimicrobials that have a phenylpropanoid chemical structure. In animals, the most commonly used amphenicol is florfenicol. Florfenicol is currently authorised for use in animals in NZ. It is registered for use in cattle in New Zealand but no commercial product is currently available. The amphenicols act by binding to the A2451 and A2452 residues of the 23S rRNA subunit, thereby directly inhibiting the peptidyl transferases enzyme of the 50S subunit ribosome (Schifano et al., 2013).The efficacy of florfenicol is increased by the presence of fluorine as it prevents acetylation of the drug by the bacteria (Papich, 2016).

1.3.2.2.6 STREPTOGRAMINS Streptogramins are a family of antimicrobials consisting of two distinct groups. Group A streptogramins contain a 23-membered unsaturated ring with lactone and peptide bonds while group B streptogramins are depsipeptides. The most commonly used streptogramin is virginiamycin, a mixture of a group A and a group B streptogramin. The principal use of virginiamycin in cattle has been in the feed as a growth promoter but in New Zealand the use has been limited only to horses and poultry (MPI, 2019). Streptogramins are used in humans to treat Vancomycin-Resistant *Staphylococcus aureus* (VRSA) and Vancomycin-Resistant *Enterococcus* (VRE). Group A streptogramins act by binding to peptidyl transferase, while Group B streptogramins cause premature dissociation of the polypeptide chains, like macrolides and lincosamides. Group A and B streptogramins are bacteriostatic separately but bactericidal together (Vannuffel and Cocito, 1996).

There are three mechanisms of resistance reported in *S. aureus* against macrolides, lincosamides and group B streptogramins (MLS_B antimicrobials): 1) modification by methylation or mutation of the target site; 2) drug inactivation; and 3) efflux of the drug. Resistances against the three groups of antimicrobials are often conferred together, while resistance against group A streptogramins is determined by a different mechanism. Modification of the target site by either methylation or mutation confers general resistance to MLS_B antimicrobials, while the other two mechanisms are more antimicrobial-specific (Leclercq, 2002).

1.3.2.2.6.1 MODIFICATION OF TARGET SITE Target site modification is the most common pathway of resistance to the MLS_B antimicrobials used by *S. aureus*. The ribosomal modification is mediated by Erythromycin ribosome methylases (Erm) encoded by the *erm* genes (Weisblum, 1995). Approximately 42 types of *erm* genes have been identified to date (Alcock et al., 2019). These genes are carried on either plasmids or transposons. The most common erm genes found in *S. aureus* are ermA and ermC (Papagiannitsis and Constantinos, 2018). These genes have been found to transfer between closely related species and to phylogenetically distant bacteria (Park et al., 2010). The Erm enzymes add one or two methyl groups to the A2058 base in the 23S rRNA (Roberts et al., 1999). This addition leads to a conformational change in the target site for the antimicrobials, and since MLS_B antimicrobials share their target site, there is a cross-resistance to these antimicrobials (Fernandez-Munoz et al., 1971; Leclercq, 2002).

1.3.2.2.6.2 DRUG INACTIVATION Esterases and phosphotransferases encoded by the *ere* and *mphC* genes inactivate 14 and 15-membered macrolides. In contrast, lincosamide nucleotidyl transferases encoded by *lnu(A)* and *lnu(B)* inactivate lincosamides in *S. aureus* (Ross et al., 1990; Wondrack et al., 1996; Roberts et al., 1999; Chesneau et al., 2007). Virginiamycin is inactivated by enzymes such as virginiamycin B hydrolase encoded by the vgbA gene, rarely found in staphylococci (Roberts et al., 1999). Acetyltransferases encoded by *vat*(A), *vat*(B), or *vat*(C) mediates resistance to streptogramin group A antimicrobials in *S. aureus* (Schwarz et al., 2011, 2014).

1.3.2.2.6.3 EFFLUX PUMPS The efflux pump MsrA that confers resistance to macrolides and Group B streptogramins is part of the ATPase binding cassette (ABC) superfamily and is encoded by the *msr(A)* gene, which is plasmid-borne (Leclercq, 2002). The efflux pump requires ATP to pump the antimicrobials out (Ross et al., 1990). Other efflux pumps such as Vga, Lsa, and Sal (also part of the ABC superfamily) are encoded by the *vga*, *lsa* and *sal* genes, respectively. These pumps confer resistance to lincosamides and streptogramin group A antimicrobials (Hauschild et al., 2011; Lozano et al., 2012; Wendlandt et al., 2012; Hot et al., 2014). The regulation of the expression of these pumps is either inducible or constitutive, similar to the *erm* gene regulation.

1.3.2.3 ANTIMICROBIALS THAT TARGET PROTEIN SYNTHESIS BY OTHER MECHANISMS

Fusidic acid is a steroidal antimicrobial synthesised by the fungus, *Fusidium coccineum*, which is commonly used in humans to treat skin infections caused by *S. aureus* including MRSA. The compound targets bacterial protein synthesis by binding to the Elongation Factor G (EF-G). In the presence of fusidic acid, EF-G binds with the ribosome which interferes in the functions of EF-G (van Bambeke et al., 2017). Two major mechanisms mediate the resistance of *S. aureus* to fusidic acid: modification of the target site due to the mutations in the *fusA* gene encoding EF-G (Besier et al., 2003); and protection of EF-G from fusidic acid via FusE proteins encoded by the *fusB*, *fusC* and *fusD* genes (O'Neill et al., 2007; O'Neill and Chopra, 2006; Lannergård et al., 2009) carried in mobile genetic elements (O'Brien et al., 2002).

Mupirocin, isolated from the bacterium *Pseudomonas fluorescens* in 1971, is a topical agent for skin infections and is also used to decrease the nasal carriage load of MRSA. Mupirocin, targets isoleucyl-tRNA synthetase (IleRS). This enzyme catalyzes the binding of tRNA to its equivalent amino acid. Hence, by binding to this enzyme, mupirocin inhibits protein synthesis in the bacterial cell (Parenti et al., 1987). Low levels of resistance (MIC = 8-64 μg/ml) are caused by mutations in IleRS while high levels of resistance (MIC >512 μg/ml) are caused by plasmid-borne mupA or mupB genes which encode for a new IleRS (Udo et al., 2001; Antonio et al., 2002; Seah et al., 2012).Resistance to fusidic acid and mupirocin in New Zealand, in particular in

ST5 and ST1, is increasing due to the increased usage of these topical antimicrobials (Williamson et al., 2014b; Carter et al., 2018).

1.3.2.4 ANTIMICROBIALS THAT TARGET NUCLEIC ACID BIOSYNTHESIS

Quinolones are a completely synthetic class of antibacterial drugs that target the nucleic acid biosynthesis of bacteria. The most commonly used quinolones are the fluoroquinolones, with enrofloxacin, marbofloxacin and pradofloxacin approved for veterinary medicine in NZ (MPI, 2019). In cattle, enrofloxacin is indicated for use against pneumonia and acute mastitis.

There are two target sites for fluoroquinolones: 1) the DNA gyrase, and 2) the topoisomerase IV. These enzymes are both classified as type II topoisomerases. DNA gyrase helps relax the supercoil of chromosomal DNA by introducing opposite coiling (negative supercoil) to the positive supercoil of the DNA. This process is essential during replication and transcription when the DNA is unwound by DNA helicase (Reece and Maxwell, 1991). DNA gyrase is composed of two subunits GyrA and GyrB encoded by the *gyrA* and *gyrB* genes. The fluoroquinolone-gyrase-DNA complex prevents negative supercoiling of the DNA which inhibits DNA synthesis.

Topoisomerase IV like DNA gyrase has two subunits ParC and ParB, which are homologous to GyrA and GyrB, respectively. In *S. aureus*, the subunits of topoisomerase IV have been referred to as GrlA and GrlB. The topoisomerase IV decatenates or unlinks the overwinding of the daughter DNA duplexes after DNA replication, which is necessary for cell division. It also relaxes the supercoil of the DNA, like DNA gyrase but does not introduce negative supercoil (Champoux, 2001). The fluoroquinolones act on the topoisomerase IV similarly to DNA gyrase by forming a complex with topoisomerase IV and DNA. This complex causes damage to the DNA leading to cell death (Khodursky et al., 1995; Khodursky and Cozzarelli, 1998).

Whether gyrase or topoisomerase IV is the preferred target depends on the type of the quinolone. For example, norfloxacin preferentially targets topoisomerase IV (and blocks DNA synthesis slowly), while nalidixic acid targets gyrase (and rapidly inhibits replication) and ciprofloxacin targets both. The quinolone structure likely

determines which enzymes are the preferred target (Fournier et al., 2000).

Resistance to fluoroquinolones in *S. aureus* is conferred by either mutation of the topoisomerases or by efflux pumps. The mutation commonly occurs in the GyrA or ParC subunits of the DNA gyrase or topoisomerase IV, respectively. *S. aureus* is susceptible to fluoroquinolones, and hence mutations in both GyrA and ParC are essential to cause high-level resistance (ciprofloxacin MIC_{50} =50 μg/ml); mutations in ParC alone cause low-level resistance (ciprofloxacin MIC_{50} = 3.13-12.5 μg/ml) (Tanaka et al., 2000). The mutations are usually amino acid substitutions that occur at a subinhibitory concentration of the drug.

High resistance to fluoroquinolones in *S. aureus* is also mediated by overexpression of Nor efflux pumps. Nor belongs to the Major Facilitator Superfamily (MFS); these are membrane proteins that transport solutes which are encoded in the chromosomal DNA. There are several types of Nor efflux pumps reported in *S. aureus*. NorA pumps out hydrophilic fluoroquinolones such as ciprofloxacin and norfloxacin and NorB and NorC pump out hydrophobic fluoroquinolones such as moxifloxacin and sparfloxacin (Neyfakh et al., 1993; Truong-Bolduc et al., 2006; Ding et al., 2008). Other Nor pumps reported in *S. aureus* include NorD and NorK (Ding et al., 2012; Briaud et al., 2019). The Nor efflux pumps are multidrug pumps and are indicated in the efflux of tetracyclines, quaternary ammonium compounds, and dyes such as ethidium bromide (Costa et al., 2013).

The expression of NorA and NorB efflux pumps is regulated by global regulator MgrA where increased expression of MgrA upregulates one (NorB) while downregulating the other (NorA) (Kaatz et al., 2005; Ding et al., 2008).They are also regulated by the two-component system ArlRS (Fournier et al., 2000). Overexpression of the efflux proteins is also controlled by environmental factors such as pH, aeration and availability of iron and it has been noted that the NorB and NorD efflux proteins are overexpressed in abscess models aiding bacterial survival (Ding et al., 2008, 2012).

1.3.2.5 ANTIMICROBIALS THAT TARGET OTHER METABOLIC PATHWAYS

1.3.2.5.1 ANTIMICROBIALS THAT TARGET FOLIC ACID METABOLISM

Some antibacterial agents inhibit bacterial growth by interfering with folic acid metabolism. Folic acid is essential for protein, and nucleic acid synthesis in bacteria, and the bacteria are able to synthesize folic acid from pteroate and para-aminobenzoic acid (PABA).

Sulphonamides were the first synthetic drugs that were used selectively against bacterial infections. Prontosil the first sulphonamide introduced in the 1930s was a pro-drug converted to the active compound sulphanilamide *in vivo* (Gaynes, 2017). The use of sulphonamides has been dramatically reduced due to their side effects, ineffectiveness where bacteria have the availability of excess PABA such as in the presence of pus, and the discovery of more effective antimicrobials (Jick, 1982; Macleod, 1940). Sulphonamides are commonly used in combination with other antibacterials, especially with diaminopyrimidines such as trimethoprim. The common sulphonamides used clinically are sulphamethoxazole and sulphadiazine (Foster, 2017).

Sulphonamides inhibit dihydropteroate synthase (DHPS), which combines pteroate and PABA in the folic acid synthesis pathway, as shown in Figure 1.5. The sulphonamides compete with PABA to bind with DHPS and form an inactive pteroate-sulphonamide product (Skold, 2000).

Resistance is caused by mutations in the chromosomally located *folP* gene, which encodes DHPS. This mutation alters DHPS preventing the binding of sulphonamide (Hampele et al., 1997; Skold, 2000).

The diaminopyrimidines target dihydrofolate reductase (DHFR), one of the enzymes involved in the folic acid metabolism, which converts dihydrofolic acid to tetrahydrofolic acid (Gleckman et al., 1981) as shown in Figure 1.5. Although DHFR enzymes are found in mammals, diaminopyrimidines have a high affinity towards bacterial DHFR (Hitchings and Burchall, 1965). The DHFR enzyme is also used to reduce tetrahydrofolic acid to dihydrofolic acid when methylene tetrahydrofolate donates carbon and hydrogen atoms in the synthesis of deoxythymidylate

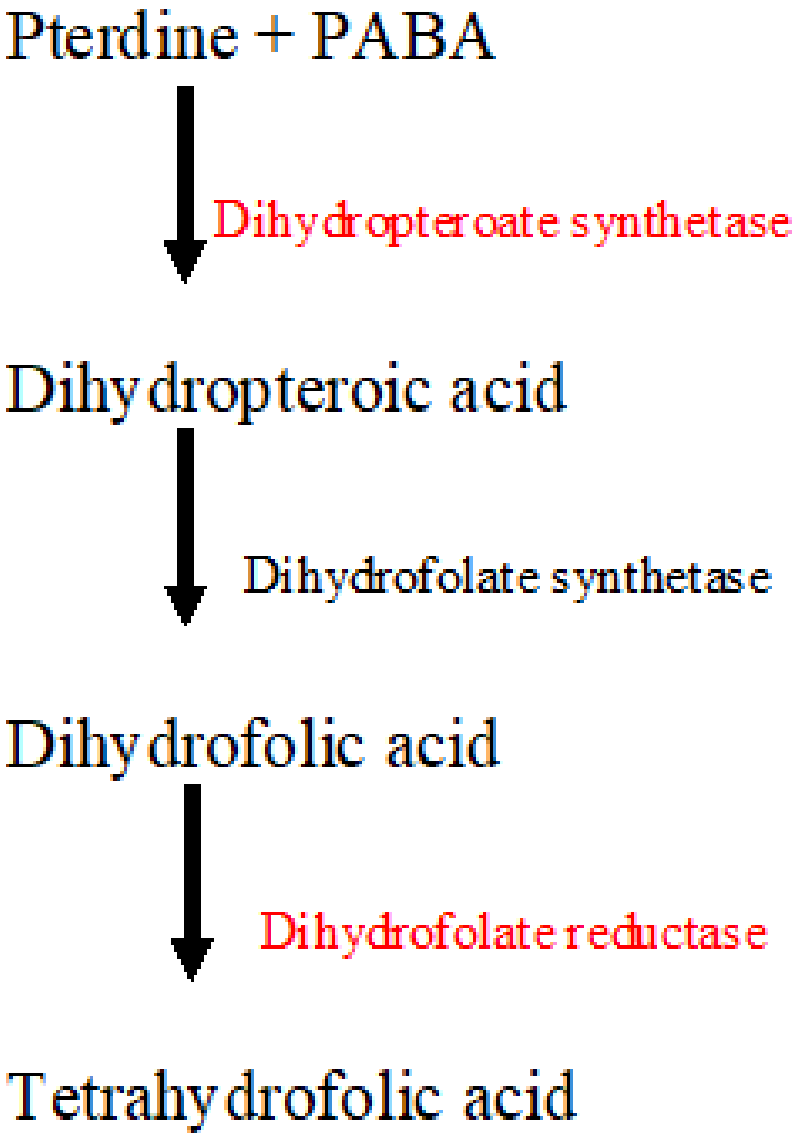

Figure 1.5: Folic acid *de-novo* synthesis in the bacterial cell. The enzymes marked in red are targets for antibacterials such as sulphonamides and diaminopyrimidines

monophosphate, which is one of the nucleotides of DNA (Salter, 1982). Thereby the diaminopyrimidines affect DNA synthesis as well as folate production.

Resistance to diaminopyrimidines in *S. aureus* is mediated either by a single amino acid mutation of the chromosomally located *dfrB* gene that codes for DHFR, or through acquired *dfr* genes located on plasmids. These plasmid DHFRs provide high-level resistance to trimethoprim (MIC $\geq$512 mg/L), while the chromosomal *dfrB* mutation confers resistance to trimethoprim at MIC $\leq$256 mg/L (Dale et al., 1997; Sekiguchi et al., 2005; Nurjadi et al., 2014).

1.3.3 DIAGNOSIS OF ANTIMICROBIAL RESISTANCE

Diagnosis of antmicrobial resistance has been increasingly recommended as a component of responsible use and for antimicrobial stewardship programs (NZVA, 2019; Watts et al., 2018; Coombe et al., 2019) and helps in monitoring the changes and trends in resistance across various species at regional and national levels. Diagnosis of antimicrobial resistance at the farm level can be helpful in the investigation of persistent non-antimicrobial-responsive *S. aureus* infection. However, routine susceptibility tests at the farm level are not always feasible due to the need to initiate empirical treatment as soon as possible. Moreover, there is a lack of standard interpretative criteria for veterinary pathogens for all the routinely used antimicrobials and in many instances, the associations between antimicrobial susceptibility test results and treatment outcome are weak (Constable and Morin, 2003; Barlow, 2011).

At the industry level, monitoring of antimicrobial resistance can be useful in the early identification of emerging resistance (Filioussis et al., 2020). This is especially important in organisms such as *S. aureus*, which can develop multi-drug resistance and pan-resistance and has been identified as zoonotic, especially in the case of CC398 MRSA (Goerge et al., 2017).

Identification of antimicrobial resistance is based on *in vitro* phenotypic detection using different antimicrobial susceptibility testing methods, including disk diffusion method, dilution methods (broth and agar dilution), and the Epsilometer test (E-test). Alternatively, genotypic detection of resistance involves the identification of resistance genes or mutations by means of PCR, DNA hybridization (array

technologies) or, more recently, whole-genome sequencing analysis.

1.3.3.1 PHENOTYPIC METHODS

1.3.3.1.1 DISK DIFFUSION METHOD Disk diffusion, also known as the Kirby-Bauer disk diffusion method, is a standardised method of antimicrobial susceptibility testing widely used in clinical veterinary laboratories in New Zealand and abroad. The method is primarily qualitative, and categorises a bacterial isolate as susceptible, resistant or intermediate, based on the inhibition of the isolate's growth around a disk containing the antimicrobial. The test involves the inoculation of bacteria at a specific concentration ($\sim$1-2 X 10^8 colony forming units (CFU) per ml) on an agarised medium (usually Muller-Hinton (MH) agar) plate. The concentration of 1-2 X10^8 CFU is achieved by comparing the turbidity of the bacterial suspension to a 0.5 McFarland standard. McFarland standard is a reference suspension for adjusting the turbidity of bacterial inoculum to achieve specific bacterial concentrations, and the standard is prepared by mixing a specific quantity of barium chloride dihydrate to sulphuric acid (0.5 McFarland standard = 0.05 mL of 1.175% barium chloride dihydrate with 9.95 mL of 1% sulphuric acid). Commercial McFarland standards are prepared using suspensions of latex particles which increases their shelf-life. Antimicrobial infused disks are placed on the plate and incubated 16-24 hours at 37°C. Plates are incubated and at the end of the incubation, they are checked for zones of no bacterial growth known as zones of inhibition. The diameter of the zones of inhibition is then measured and checked against standard interpretative criteria (Reller et al., 2009), and the isolate is designated as susceptible, intermediate, or resistant. The intermediate classification is used when the concentration of the antimicrobial is inhibitory *in vitro* but has uncertain therapeutic effect (Rodloff et al., 2008). The interpretative criteria are published by organizations such as the Clinical and Laboratory Standards Institute (CLSI) (CLSI, 2018,a) or the European Committee on Antimicrobial Susceptibility Testing (EUCAST) (Matuschek et al., 2014). These institutes also provide a detailed stepwise procedure for the disk diffusion method and interpretative criteria for each antimicrobial commonly used, when available.

The main advantages of the disk diffusion method is that the procedure is simple and does not require specialized instrumentation, with a comparatively low cost per test (Reller et al., 2009). Its main disadvantage is the lack of interpretative criteria for

most veterinary pathogen/ antimicrobial/ target organ combinations. This issue has been recognised by CLSI and EUCAST, and both organisations have subcommittees (VAST and VEtCast, respectively) responsible for increasing the information and interpretative criteria for veterinary pathogens and antimicrobials. Other disadvantages of the disk diffusion test include the lack of automated systems. Only a certain number of antimicrobial strips can be used per plate, limiting the number of antimicrobials that can be tested simultaneously. In spite of these disadvantages, the disk diffusion test remains the main test used by clinical veterinary laboratories in New Zealand.

1.3.3.1.2 DILUTION METHODS Dilution methods are quantitative antimicrobial susceptibility testing methods. The outcome of such testing is the minimum inhibitory concentration (MIC) of the antimicrobial for a specific isolate. The MIC is the lowest concentration of the antimicrobial at which there is no visible growth of the bacteria under defined conditions (Wiegand et al., 2008). There are two types of dilution based on the growth media: 1) agar dilution and 2) broth dilution.

In agar dilution, the bacteria are inoculated onto multiple MH agar plates in which different dilutions of the test antimicrobial have been incorporated. The plates are incubated at established temperatures, usually 35-37°C for 16-20 hours, and checked for visible bacterial growth. The MIC is represented by the antimicrobial concentration present in the first plate that does not display bacterial growth, and this MIC is checked against the interpretative threshold of CLSI or EUCAST, and the isolate is then designated as susceptible, resistant, or intermediate.

Broth dilution is more commonly used than agar dilution. The volume of liquid media used defines a method as macrodilution (bacteria are inoculated into test tubes) or microdilution (bacteria are inoculated in microtitre plates with total volumes ≤ 1 mL). Two-fold dilutions (1:2, 1:4, 1:8, and so on) of the antimicrobial in MH are inoculated with the bacterial isolate and incubated. A turbidity, for some bacteria sediment identifies the growth of the bacteria, and the MIC is recorded as the lowest concentration of antimicrobial that does not show any bacterial growth. The MIC is then interpreted using the same interpretative criteria indicated for the agar dilution method.

The microdilution method is more commonly used as it is more efficient and less

prone to error than than macrodilution method (Reller et al., 2009).

The advantage of the dilution method is finding the MIC of the antimicrobial for precise determination of the concentration of antimicrobial required to inhibit the growth of the bacteria. Moreover, as a quantitative measurement, its diffused use enables more powerful monitoring of resistance at the population level than the qualitative disk diffusion test. However, when performed manually, dilution methods are more labor intensive and costlier than the disk diffusion test. High throughput clinical laboratories can use commercially available microdilution plates that enable automatisation of the testing.

1.3.3.1.3 E-TEST The E-test is an antimicrobial susceptibility method that combines the principle of both dilution and diffusion method, to determine the MIC of the drug. It was manufactured first in 1991 by AB BIODISK as an E-test plastic strip to detect resistance against multiple antimicrobials (Sanchez and Jones, 1992). E-test are plastic strips, each coated with a gradient concentration of an antimicrobial with corresponding MIC ranges marked as a scale on the back of the strip. The strips are placed on an agar plate pre-inoculated with the organism of interest. Following overnight incubation, the plates are checked for zones of inhibition. The MIC of the drug for the tested organism corresponds to the intersection point between the zone of inhibition and the edge of the strip (Sanchez and Jones, 1992). The advantages of the E-test are simplicity and accuracy of MIC results without the need for dilutions. Although the E-test was generally found to have a good overall agreement (95%-97%) with dilution methods (Skov et al., 2006; Cantón et al., 2016), the level of agreement between E-test and dilution methods are dependent on the antimicrobial used. Varela et al. (2008) reported a poor agreement for ampicillin and nalidixic acid, moderate agreement for azithromycin and erythromycin and reasonable agreement for streptomycin and tetracycline. Compared to the disk diffusion method, the cost is much higher if multiple antimicrobials are tested.

Automated systems such as Vitek 2 (BioMerieux, Marcy-l'Étoile, France) or Sensititre ARIS 2X (ThermoFisher Scientific, Massachusetts, USA) that can process a vast number of antimicrobial susceptibility tests in much less time than manual methods are available. However, they were found to be less reproducible compared to conventional manual methods (d'Azevedo et al., 2009; Ferreira et al., 2018).

1.3.3.2 GENOTYPING METHODS

Genotyping methods identify the genomic basis of antimicrobial resistance and are particularly useful for organisms that are slow-growing.

1.3.3.2.1 POLYMERASE CHAIN REACTION (PCR) The most common method of genotypic detection of genomic markers of antimicrobial resistance is PCR amplification of pre-selected markers. This method involves the amplification of the target (resistance) gene from the organism using specific primers. In *S. aureus*, PCR-based resistance detection is used frequently to detect the *mecA* gene in MRSA strains (Pournajaf et al., 2014), and has also been used for other resistance genes, including *aacA-aphD*, *tetK*, and *erm(C)* (Strommenger et al., 2003).

1.3.3.2.2 DNA HYBRIDIZATION In DNA hybridization, nucleotides pair with their complementary nucleotides (cytosine with guanine and adenine with thymidine). In this technique, the DNA of the test sample is separated into single strands, and single-stranded probes usually anchored on a solid surface combine with its complementary region, if it is present in the analysed DNA. For antimicrobial susceptibility testing, the probes used are complementary to known resistance genes. The probes are tagged or labelled with radioactive isotopes, enzymes, chemiluminescent or fluorescent compounds, or antigenic substrates so that if hybridisation occurs, it can be easily demonstrated. If the hybridization occurs, then it is labelled with one of the compounds mentioned above. If there is no target sequence present in the sample DNA, there is no hybridization and detectable labelling (Fluit et al., 2001). DNA hybridization is amenable for introduction into commercial arrays or microarrays.

PCR and hybridisation are both rapid, reliable tests. These techniques are useful in the identification of resistance traits for which there are well characterised genetic components. However, there is an increased time and cost associated with identification of multiple resistance genes over large samples.

1.3.3.2.3 WHOLE-GENOME SEQUENCE ANALYSIS An alternative to the amplification of individual genes is to sequence the whole-genome and assess the

presence of resistance genes or genetic determinants of resistance (collectively dubbed 'resistome') in the genome. Advancements in whole genome sequencing (WGS) have resulted in increasingly high throughput technologies and reduced costs. As a result, WGS analysis is increasingly being used for the identification of antimicrobial resistance in epidemiological research. The key advantages of WGS are the identification of multiple resistance determinants from a single WGS reaction, and the possibility of studying the effect of associations between different genomic determinants of resistance, and the expression of phenotypic resistance.

There are two approaches to identification of genetic determinants of antimicrobial resistance using WGS (Su et al., 2019). The first is a rules-based approach that uses computerised algorithms that rely on searches of current, curated resistance genes stored in various databases. This approach has the same limitations as PCR and hybridisation; i.e. only properly characterised genes can be identified. However, an advantage of genomic analysis is that once the genome is sequenced, it is permanently available and can be further analysed when new antimicrobial resistance determinants are discovered, whereas *in vitro* testing using PCR or hybridisation will require additional testing.

The rules-based approach is based on two assumptions in addition to the requirement for the determinants to be properly characterised and stored in the databases (Su et al., 2019): i) the phenotypic resistance derives from the mere presence of a single determinant in the genome, or if multiple determinants are responsible for the resistance, they interact in a simple, rather than a complex manner and ii) there is a strong positive relationship between the presence of a determinant and the resistance phenotype, which is not affected by the genetic background of the strain.

The limitations of the rules-based approach have led to the development of the machine learning (ML)-based approach that uses models that are constructed based on training two sets of isolates of known phenotype, resistant and susceptible. The trained sets are then tested for accuracy by predicting resistance in a different set of isolates (Su et al., 2019).

Irrespective of the approach, the starting point of any WGS-AST are the raw reads of the genomic sequence. The rules-based approach maps these sequences to the resistance determinants available in the databases. AMR databases can either cover multiple bacterial species, for example, the Comprehensive Antibiotic Resistance

Database (CARD) (Alcock et al., 2019), Resfinder (Zankari et al., 2012), (Antibiotic Resistance Gene-Annotation) ARG-ANNOT (Gupta et al., 2014), ARDB (Liu and Pop, 2009), MEGARes (Lakin et al., 2017), RAST (Davis et al., 2016), Bacterial Antimicrobial Resistance Reference Gene Database (BARRGD) (https://www.ncbi.nlm.nih.gov/bioproject/313047) or be specific for a single species such as TBDReaMDB (Sandgren et al., 2009) and MUBII-TB-DB (Flandrois et al., 2014) for *Mycobacterium tuberculosis*.

Various software tools have automated the process of cross-referencing various AMR gene databases. These tools work with either raw sequence data or with sequence data after genome assembly. Using raw reads reduces the time to produce results, but there is a trade-off in result quality because of sequencing errors present in the raw data or contamination with sequences from non-target organisms. Genome assembly can be done either *de novo* or by mapping to a reference genome. *De novo* assembly detects the single nucleotide polymorphism (SNP) in resistance genes with better accuracy than the mapping-based assembly. However, *de novo* assembly can split genes across multiple contigs, resulting in an AMR gene being missed. ARIBA, the software used for this study, does partial *de novo* assembly of the reads that may map to the target genes from the AMR database of choice (Hunt et al., 2017). Other programs use different approaches; e.g. SRST2 (Inouye et al., 2014) uses a read alignment program called Bowtie2 (Langmead and Salzberg, 2012) to map the target genes to the reads and Point-Finder identifies known point mutations in the sample by mapping reads against the reference genomes from databases (Zankari et al., 2017).

Studies have shown that rules-based WGS-AST generally has high sensitivity (>90%) and high specificity (>95%) for multiple antimicrobial resistance phenotypes across multiple pathogens (Gordon et al., 2014; Mason et al., 2018; Coll et al., 2015).

The ML approach of WGS-AST uses training sets of genomes with known phenotypes and is fed from an existing database and focussed on genetic features (e.g. SNPs or substitutions, insertions and deletions). These genetic features are given scores or weights based on their contribution to antimicrobial resistance. This approach can also model interactions between resistance genes and between resistance genes and the genetic background of the strain (Su et al., 2019). The major disadvantage of this learning approach is the dependency on the quality of the data used to develop and test models. The model will fail to be accurate if the data is not

accurate. (Boolchandani et al., 2019).

Even though WGS-AST has significant advantages over PCR and hybridisation in determining WGS-AST, the complexity of the genetic basis of resistance mechanisms may deny the routine use of WGS-AST instead of the established *in vitro* phenotypic susceptibility testing.

1.4 REVERSE VACCINOLOGY

Vaccines are one of the most important inventions of humankind. Currently, it is estimated that vaccination prevents more than 2.5 million human deaths per year and that 83% of the entire world's infant population has had at least one vaccine such as the diptheria or measles vaccines (WorldHealthOrganization, 2020). Vaccination is also of considerable importance in animals, where in addition to improving animal health and welfare, it has been used to improve the productivity of livestock and reduce/prevent the spread of zoonotic diseases (Meeusen et al., 2007). The latter two roles of vaccines are becoming increasingly important with the increasing focus on reducing the use of antimicrobials in food producing animals (Laxminarayan et al., 2016; Martin et al., 2015). Veterinary vaccines are an expanding market; in 2015, the veterinary vaccines market was worth US$ 12.10 billion, and it is expected to grow to US$20.6 billion by 2021 (ZionMarketResearch, 2016). Livestock vaccines are a large part of this market, accounting for 62% of the market as of 2015.

Vaccines can be categorised in different ways. One categorisation which is growing in importance is based on whether the physical organism (conventional vaccinology) or the genome (reverse vaccinology) is used as the starting point for vaccine design (Figure 1.6)

1.4.1 CONVENTIONAL VACCINES AND THEIR TYPES

The broad classification of vaccines as conventional and reverse vaccinology vaccines began with the development of the reverse vaccinology method by Rappuoli (2000). This development of reverse vaccinology classified vaccines that started with the

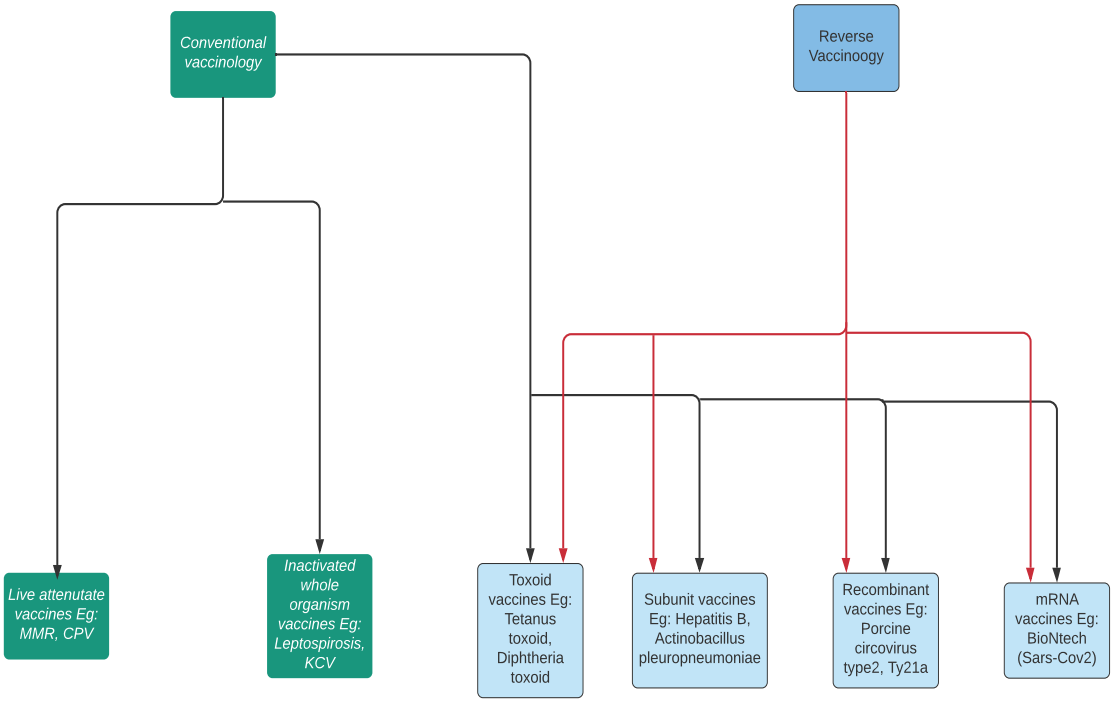

Figure 1.6: Types of vaccines based on whether the physical organism or the genome of the organisms is used as the starting point for vaccine design. MMR- Measles, Mumps and Rubella, CPV- Canine Parvovirus, KCV- Kennel Cough Vaccine (Modified from WorldHealthOrganization (2018); Iwasaki and Omer (2020)

physical organism as conventional vaccines.

The practice of vaccination started in Asia, by utilizing lesions from smallpox to induce mild infection to protect against deadly disease, a practice later dubbed 'variolation' (Fenner et al., 1989). But the term 'vaccination' was coined by Edward Jenner in 1796 when he used pustules from cowpox to immunize against smallpox (Baxby, 1981). Robert Koch and Louis Pasteur later independently identified the causal relationship between microbes and communicable diseases which opened the way for a rational approach to treatment and prevention of infectious diseases (Ullmann, 2007).

Among conventional vaccines, there are various types of vaccine design. The earliest method of vaccine development was the preparation of live, attenuated vaccines. These may be produced by passaging the organism (either bacteria or virus) in a culture media for a specific number of times so that the virulence of the organism is significantly reduced ('attenuation'), but its antigenicity is maintained. Another method is to use strains of reduced pathogenicity found in nature (i.e., naturally attenuated). Jenner's vaccine against smallpox used poxvirus from a cow, because it was much less pathogenic than the smallpox virus, but the fowl cholera vaccine developed by Louis Pasteur was the first vaccine to use a laboratory attenuated organism (an old culture of *Pasteurella multocida*) (Pasteur, 1880). Pasteur and his team also developed vaccines against anthrax and rabies through this method of attenuation (Lombard et al., 2007). Other examples of laboratory-attenuated vaccines are the vaccines against smallpox, poliovirus, rinderpest, influenza virus, and oral typhoid vaccine. One of the advantages of this type of vaccine is that the live organism elicits an immune response that is very similar to that produced by the wild fully virulent pathogen, as the live organism reproduces in the host. However, the organisms in these vaccines can still cause severe disease-like reactions, or mutate and revert to a more virulent phenotype, for example, vaccine-associated paralytic poliomyelitis (Dowdle et al., 2003). In immunocompromised individuals, live vaccines can produce clinical disease and can even be fatal. However, these events are rare and there are guidelines established to ensure that the benefits of live vaccines outweigh the risks (Kengsakul et al., 2002; Papadopoulou and Sipsas, 2014; Huber et al., 2018).

Inactivated, or 'killed' vaccines containing whole organisms are developed by

inactivating the organisms using either physical (e.g. heat) or chemical methods (e.g. formaldehyde). These types of vaccines were first proposed by Salmon and Smith (1886), and were initially applied to immunise against cholera, typhoid, and plague. Initially, the inactivated vaccines consisted of heat-killed organisms, but Madsen (1933) developed a pertussis vaccine with *Bordetella pertussis* which had been inactivated using formaldehyde, a process which was later standardized by Kendrick and Eldering (Shapiro-Shapin, 2010). Other inactivated vaccines include vaccines for influenza, typhoid and rabies. Inactivated vaccines do not replicate within the host, so may not stimulate immune function in the same way as a live vaccine and may therefore produce immunity of shorter duration and which may not be as effective and therefore, require the use of adjuvants to enhance immunogenicity. However, adjuvants can cause local and systemic reactions and have been linked to narcolepsy, autoimmune syndromes, and Alzheimer's disease (Petrovsky, 2015).

Another form of inactivation is the production of toxoid (or other acellular fractions), where the toxins of infective organisms are inactivated and used as a vaccine. Formalin treatment of diphtheria toxin to reduce toxicity was first used to produce a vaccine by Glenny and Hopkins (1923). Other examples of toxoid vaccines include clostridial vaccines (e.g. tetanus and black leg). The toxoid vaccines have advantages such as no reversion to virulence and high stability. However, not all toxins produced by pathogens can be used as toxoids as the host immune response is not the same for every toxin.

Subunit vaccines are similar to inactivated vaccines in that they do not contain live organisms. However, in contrast to inactivated vaccines they do not contain the whole organism, but just the parts of the organisms that have been identified as antigenic (in this sense, toxoid vaccines can be considered subunit vaccines). Subunit vaccines can be further divided into protein-based, polysaccharide-based, and conjugate subunit vaccines. An example of a protein-based subunit vaccine is the monovalent influenza A virus vaccine, which contains purified viral hemagglutinin protein (Gross and Ennis, 1977). The virus is grown in embryonated eggs and then broken into various components using detergents, and the most antigenic parts are fractionated and used in the vaccine. The polysaccharide-based subunit vaccines contain polysaccharide component of the organisms, often the bacterial capsule. The first use of the capsular component in a subunit vaccine was in the development of a meningococcal vaccine against *Neisseria meningitidis* (Gotschlich et al., 1969). Polysaccharide-based subunit

vaccines are not immunogenic in infants as they are unable to produce a B-cell response to the capsular polysaccharide alone. Hence, Schneerson et al. (1980) designed a conjugated vaccine against *Hemophilus influenzae* type b using proteins such as BSA, diphtheria toxin, hemocyanin, etc, which were conjugated to the polysaccharide of the bacteria's capsule to enhance its immunogenicity.

The genetic engineering revolution has had a great impact on vaccine development, resulting in the conception of recombinant vaccines. Hilleman et al. (1983) purified surface antigens of hepatitis B virus from infected patients, to use as vaccine against the virus. However, large quantities of these antigens were produced by genetic engineering by implanting the coding sequences for these surface antigens into yeast cells (Valenzuela et al., 1982). Other examples of recombinant vaccines are the Ty21a vaccine against typhoid, which was invented by mutating an enzyme essential for virulence (Levine et al., 1989), and the CVD 103-HgR cholera vaccine, where the organism was mutated to be unable to produce cholera toxin (Levine et al., 1988). Further developments of recombinant vaccines are ongoing, with several viral or bacterial vectors being tested to carry the antigens. Even though there are some advantages with the use of recombinant vaccines over live and inactivated vaccines, such as lower production cost and improved safety, recombinant vaccines often elicit a weak immune response, which can be overcome by the use of adjuvants as with all inactivated, non-live vaccines (Pérez et al., 2012).

In the last two decades, messenger RNA (mRNA) vaccines have been developed. The mRNA is the intermediate product during protein synthesis. The coding region of the DNA is transcribed to mRNA which is then translated to form a protein. The use of mRNA as a vaccine candidate has received a lot of attention with the advent of the Sars-Cov2 pandemic and the subsequent vaccine development. The mRNA vaccines are non-infectious and there is no risk of infection or insertional mutagenesis. The inoculated mRNA is degraded by normal cellular processes and the immunogenicity can be modulated to make it efficacious and safe. The production of mRNA vaccines can be rapid and high yielding (Pardi et al., 2018) which is being recognised during the Sars-Cov2 pandemic .

1.4.2 OVERVIEW OF REVERSE VACCINOLOGY

The conventional methods for the identification of antigens relies on the study of those antigens that are expressed during the disease process or *in vitro*. This process requires the identification of individual antigens as potential vaccine candidates (PVCs), and is slow and resource intensive. Many potential immunogenic antigens may be missed when conventional techniques are used. This drawback was the main issue in the development of a vaccine against Neisseria menigitidis, which necessitated a new approach to the identification of vaccine candidates. The identification of PVCs by analysing the whole genome of the target pathogen is called reverse vaccinology (Rappuoli, 2000). The term reverse vaccinology (RV) implies that the antigens are identified from the genome, then produced and tested *in vitro* and *in vivo*, in contrast to the conventional vaccinology approach of starting with the target organism. As such, the first phase of reverse vaccinology vaccine development is always implemented *in silico*. This may accelerate and reduce the costs associated with PVC discovery compared to conventional vaccine development, which initiates by the identification of PVCs in the laboratory. The reverse vaccinology method can be used to design subunit, recombinant, mRNA or toxoid vaccines.

1.4.2.1 HISTORY OF REVERSE VACCINOLOGY

One of the big promises in the vaccine design that arose in the 20th century was the use of genomic technology. The first complete whole-genome sequence of any living organism (*Haemophilus influenza*) was published in the year 1995 by Fleischmann et al. (1995). This paved the way for the rapid sequencing of entire genomes of prokaryotes. This in turn led to the use of genomic information to analyse and discover new antigens that had not been discovered by conventional vaccinology: the reverse vaccinology approach.

The first disease that was addressed by this new method of genome analysis was meningococcal meningitis. *Neisseria meningitidis*, serogroup B (MenB) which, worldwide, was the cause of 50% of all meningococcal meningitis. Conventional vaccines using capsular polysaccharide of MenB, as used for other meningococci pathogens, was not feasible because the capsular polysaccharide of MenB was chemically identical to α2–8 linked polysialic acid that is found in the host (human)

tissues. When the previously studied protein antigens were analysed for vaccine candidature, the conformation of the protein antigens prevented the recombinant proteins to be immunogenic. This was due to the folding of the protein within the outer membrane which was necessary for the presentation of the epitope of the protein for a protective immune response. Additionally, the surface proteins present in the outer membrane vesicles of the meningococci, which showed promise in studies as vaccine candidates, contained external loops of high sequence variability. This meant that when used in vaccines the protection was only against the homologous strain. As the attempts to develop a conventional vaccine failed, a team at The Institute of Genome Research (TIGR), Maryland, USA began sequencing the MenB genome and analysed the nucleotides using bioinformatic tools to predict the features of the proteins for their localization in the bacteria. The sera from mice immunized with these proteins were collected and tested for antigenicity and the protective response of the proteins. The proteins that elicited immune response were studied for their presence in other serovars of *Neisseria meningitidis.* From 570 proteins identified in the genome, 7 proteins were selected as PVCs. After safety and immunogenicity tests (Kimura et al., 2011) and clinical trials (Vesikari et al., 2013), the subunit vaccine 4CMenB, which contained the selected antigens, was licensed in Europe in 2012 and marketed worldwide by GlaxoSmithKline as Bexsero (Serruto et al., 2012). Through this attempt to find and identify PVCs, 90 new surface proteins were identified. This finding was surprising as only 12 surface antigens were previously known, despite over three decades of studies of surface proteins (Tettelin et al., 2000; Pizza et al., 2000; Giuliani et al., 2006).

After this success, reverse vaccinology was used for group B *Streptococcus*. The process identified four PVC proteins by comparing genome sequences of eight group B *Streptococcus* strains belonging to the five most common disease-causing serotypes. These four proteins were combined to create a universal vaccine against the common diseases caused by group B *Streptococcus* (Maione et al., 2005).

Other pathogens for which this genome-based approach has been or is being used to identify vaccine candidates include *Chlamydia* and *Leptospira* (Dellagostin et al., 2017; Thorpe et al., 2007). The features, benefits and drawbacks of conventional and reverse vaccinology are detailed in Table 1.10.

As can be seen from the table, reverse vaccinology has the potential to overcome the

key problems and challenges faced by conventional approaches for identifying antigens that are PVCs. With the reduction in costs and increasing ease of genome sequencing and analysis, reverse vaccinology is likely to be increasingly used to help identify and design vaccines, especially against pathogens where conventional vaccine technology has not been effective to discover novel antigens. The advent of the mRNA vaccines will only accelerate this trend.

1.4.2.2 PROCESS OF REVERSE VACCINOLOGY

The RV approach can be classified into classical RV, or pan-genomic RV, based on the number of genomes explored. The classical RV process involves designing a vaccine from a single representative genome of an organism. This approach led to the production of an effective vaccine against serogroup B *Neisseria meningitidis*. This approach limits the number of possible PVCs to those that are known from previous studies to be core, and of conserved amino acid sequence. A pan-genomic RV differs as it compares whole-genome sequences of a representative sample of strains of an organism, like group B *Streptococcus*, to produce a set of PVC that are core and of a conserved sequence without previous knowledge. The process of RV (either classical or pan-genomic) starts with whole-genome sequencing (section 1.15). The coding sequences (CDS) are annotated from the genome sequence using publicly available programs such as Prokka (Seemann, 2014), PGAP (Tatusova et al., 2016), or DFAST (Tanizawa et al., 2018). The CDS are then processed by a reverse vaccinology pipeline, to predict PVCs. Reverse vaccinology pipelines classify CDS according to a pre-established set of bioinformatic criteria, implemented in a predetermined order (Roy et al., 2018). The process results in a shortlist of CDS that code the selected PVCs.

Reverse vaccinology pipelines can be classified into two types according to their algorithmic approach: 1) Decision-tree methods and 2) Classifying methods.

The decision-tree method is also known as the filtering method, as the pipeline uses a series of filtering criteria, such as the localization of the proteins, the probability of the protein being an adhesin, its antigenicity prediction, the number of transmembrane helices. These filters eliminate protein sequences that do not fulfil the defined thresholds (Dalsass et al., 2019). These pipelines usually utilise programs to

Table 1.10: Comparison of the features, benefits and disadvantages of conventional vaccinology and reverse vaccinology (modified from Rappuoli (2000)).

Attributes	Conventional vaccinology	Reverse Vaccinology
Features	1. Identifies common and most prevalent antigens during the disease process. 2. Identifies antigens known to trigger an immune response during disease. 3. Organisms need to be culturable in the laboratory.	1. Identifies all protein antigens feasible from the organism. 2. Identifies antigens that are non-immunogenic during the disease process. 3. Identifies antigens even from organisms that cannot be cultivated easily. 4. Analysis and expression of these antigens are done under a high-throughput system
Benefits	1. Non-protein-based antigens such as polysaccharides, lipopolysaccharide and glycolipids can be used.	1. Less time consuming as all the available antigens can be analysed rapidly *in silico*. 2. Non-immunogenic antigens/transient antigens during the disease process can be identified. 3. Novel antigens can be identified and thereby newer immune responses can be explored. 4. Rare antigens can be discovered. 5. Antigens that are not expressed in cultures can be examined. 6. Vaccines can be designed for organisms that are difficult to culture.
Drawbacks	1. Antigen identification consumes a long time duration. 2. Antigens that are not expressed in cultures cannot be discovered. 3. Non-structural proteins are not considered.	1. Non-protein antigens are not used.

eliminate proteins, as shown in Table 1.11.

The choice of filters is of paramount importance, as decision-tree method may eliminate possible useful antigens which fail to pass one of the many filters used. For example, non-adhesion proteins such as porin and flagellin interact with the host and are highly antigenic. However, adhesion-based filtering would eliminate such non-adhesion proteins. Examples of decision-tree/filtering pipelines are NERVE (Vivona et al., 2006), Jenner-Predict (Jaiswal et al., 2013), Vaxign (He et al., 2010), and VacSol (Rizwan et al., 2017). Vaxign and VacSol are highly configurable pipelines that allow users to define the filters, based on their knowledge of the target organism.

The classifying method is also known as a machine-learning (ML) method as an ML algorithm is utilized to classify proteins as PVCs and non-PVCs. Two training sets of proteins are initially provided to the ML algorithm. One is a set of known PVCs and another a set of known non-PVCs. These training sets are provided from previously available or user-created databases. Once the ML is trained with known PVCs and non-PVCs, the proteins of interest are inputted and the pipeline ranks the entire set of input proteins based on their probability of being a PVC (Dalsass et al., 2019). The major limitation of the ML method is that the understanding of protein antigens is not yet complete and therefore training sets may not accurately predict which proteins are truly PVCs. Nevertheless, using ML method can identify PVCs which might be discarded by conventional RV methods. Some examples of RV pipelines that utilize the ML method are VaxiJen (Doytchinova and Flower, 2007), Vaxign-ML (Ong et al., 2020) and Vacceed (Goodswen et al., 2014).

Vaxign and VaxiJen were found to be the most user-friendly RV pipelines. The newest version of Vaxign, Vaxign2, allows the user to filter proteins based on their subcellular localization, adhesion probability and number of transmembrane helices, but also includes Vaxign-ML which can score the proteins for their likelihood of being a PVC. VaxiJen utilizes the ML method to classify proteins as antigens and non-antigens based on their physico-chemical properties. Hence, both Vaxign2 and VaxiJen were utilised in this study.

Table 1.11: Commonly used bioinformatics tools and reverse vaccinology pipelines that are used in identifying potential vaccine candidates.

Bioinformatic software/tool	Features	References
Prokka	Annotation of the genome	(Seemann, 2014)
PGAP	Annotation of the genome	(Tatusova et al., 2016)
DFAST	Annotation of genome	(Tanizawa et al., 2018)
BLAST	Searching sequences of DNA, RNA and protein for similarity	(Ye et al., 2006)
PGAT	Identifies genes present and maps orthologues between genomes	(Brittnacher et al., 2011)
Pfam	Database of protein family information	(El-Gebali et al., 2019)
ProDom	Database of protein domain families	(Bru et al., 2005)
ProtParam	Provides basic statistics of the protein	(Gasteiger et al., 2005)
PSORTb	Predicts subcellular localization of the protein	(Yu et al., 2010)
TargetP	Predicts subcellular localization of the protein	(Emanuelsson et al., 2007)
Blocks database	Database of multiple alignments of conserved regions in proteins	(Pietrokovski, 1996)
SPAAN	Predicts adhesin and adhesin-like proteins	(Sachdeva et al., 2005)
LipoP	Predicts lipoproteins and signal peptides	(Rahman et al., 2008)
SignalP	Predicts signal peptides	(Nielsen, 2017)
DictyOGlyc	Predicts glycosylation of the proteins	(Gupta et al., 1999)
NetGlycate	Predicts glycation of the proteins	(Johansen et al., 2006)
NetAcet	Predicts acetylation of the proteins	(Kiemer et al., 2005)

Continued on the next page

Bioinformatic software/tool	Features	References
NetNGlyc	Predicts N-glycosylation of the proteins	(Gupta et al., 2004)
NetOGlyc	Predicts O-glycosylation of the proteins	(Steentoft et al., 2013)
NetPhosBac	Predicts phosphorylation in bacterial proteins	(Miller et al., 2009)
ProP	Predicts cleavage sites	(Duckert et al., 2004)
HMMTOP 2.0	Predicts the number of transmembrane helices of the proteins	(Tusnady and Simon, 2001b)
TMHMM	Predicts transmembrane helices of the proteins	(Krogh et al., 2001)
SVMHC	Server for predicting MHC class I and II binding peptides	(Dönnes and Kohlbacher, 2006)
Vaxitop	Predicts MHC class I and II peptides	(He et al., 2010)
BepiPred	Predicts linear B-cell epitopes of the proteins	(Jespersen et al., 2017)
BCPred	Predicts linear B-cell epitopes of the proteins	El-Manzalawy et al. (2008)
AAPPred	Predicts linear B-cell epitopes of the proteins	IaI and Tonevitskiĭ (2009)
Jenner-Predict server	Antigen filtering pipeline for RV	(Jaiswal et al., 2013)
NERVE	First automated RV pipeline	(Vivona et al., 2006)
Vaxign	First online interface tool for RV	(He et al., 2010)
Vaxign-ML	Ranks proteins based on their probability to be potential vaccine candidates using machine learning	(Ong et al., 2020)
VacSol	Recent high throughput pipeline for RV, specifically for prokaryotic pathogens	(Rizwan et al., 2017)

Continued on the next page

Bioinformatic software/tool	Features	References
VaxiJen	Predicts antigenicity based on physico-chemical properties of the proteins using machine learning	(Doytchinova and Flower, 2007)

1.4.3 CURRENTLY AVAILABLE VACCINES AGAINST *S. AUREUS* BOVINE MASTITIS

Few commercially available conventional vaccines are available to reduce the risk of bovine *S. aureus* mastitis. Two commercial vaccines are registered in New Zealand: Lysigin (Boehringer Ingelheim Vetmedica Inc, USA) and StartVac (Hipra, Spain).

Lysigin contains a lysed culture of five different phage types (for phage typing, see section 1.1.4 of *S. aureus* (Williams et al., 1975). According to the capsular typing scheme (see Section 1.4), the vaccine contains one capsular polysaccharide (CP) serotype 5 strain, two CP8 strains, and two surface polysaccharide (SP) serotype 336 strains (Ma et al., 2004). Lysigin has been evaluated for efficiency in heifers in three different studies (Nickerson et al., 1999; Middleton et al., 2006, 2009). In general, the results of these studies were ambiguous.

Nickerson et al. (1999) followed heifers vaccinated with Lysigin at 6 months of age, with booster vaccination 2 weeks later and every 6 months until calving. The 35 vaccinated animals had a ~45% reduction in new *S. aureus* mastitis, both during pregnancy and at calving when compared to the 35 unvaccinated animals. The vaccine also reduced the incidence of coagulase-negative staphylococci (CNS) mastitis by 31%. The animals were not experimentally challenged with either *S. aureus* or CNS organisms and the study concluded that vaccination may help reduce the risk of new *S. aureus* IMI.

In another study, Middleton et al. (2006) compared Lysigin with two experimental vaccine formulations and unvaccinated controls. Heifers were vaccinated 28 days apart in late gestation. Fourteen heifers were vaccinated with Lysigin, two groups of eleven were vaccinated with one of two experimental bacterins,and 11 heifers were used as unvaccinated control. All the groups were challenged with strain ATCC29740

of *S. aureus* through intramammary infusion during 6^{th} to 8_{th} day of lactation in one infection-free mammary quarter. The bacterial clearance rates were calculated by culturing the milk every day from day 7 to day 35. Other parameters measured were the clinical mastitis score, milk yield, somatic cell count and the duration of clinical mastitis. All the animals had mastitis after the challenge. Only one cow from the Lysigin group and three from one of the experimental vaccine groups had cleared *S. aureus* from the mammary gland by the end of the study. The heifers vaccinated with Lysigin had a lower mean duration of clinical mastitis (3 days) compared to the controls (10 days). However, no significant difference was found between vaccinated and control heifers in somatic cell count, milk yield or bacterial clearance rates. The animals vaccinated with Lysigin had higher serum IgG against all the serotypes contained in the vaccine compared to the controls. However, in milk, there was no apparent IgG response against CP5 in vaccinated animals when compared to the controls (Luby et al., 2007).

The different results observed in the two studies could be due to different study designs (experimental vs natural challenge) and the definition of cure, as well as due to herd and cow factors, and the infecting strains. The apparent lack of increase in milk IgG against CP5 is interesting because the challenge strain used by Middleton et al. (2006) (ATCC 29740) is a CP5 strain (Bouchard et al., 2012).The CP5 was reported only in 18% of the bovine mastitis cases while SP336 was found in 59% and CP8 in 23% of the cases in the USA (Guidry et al., 1998). The low prevalence of CP5 strain in the field in the USA might explain the better results observed by Nickerson et al. (1999).

Middleton et al. (2009) evaluated the efficacy of Lysigin vaccination in lactating animals. The vaccinated group of 44 cows received two doses of Lysigin 14 days apart subcutaneously. The control group contained 46 unvaccinated cows. All these cows were at least 30 days in lactation and were identified to be free of *S. aureus* through three consecutive negative *S. aureus* cultures from the milk before the vaccination. After vaccination, five milk samples (Day 0, 14, 28, 49 and 70) were collected from each cow. Day 0 was the day of the first dose of vaccination. The definition of new *S. aureus* IMI was at least two positive *S. aureus* cultures from three consecutive milk samples. The milk was also checked for antibodies against *S. aureus*. The authors reported no new *S. aureus* IMI in any cow during the study period (preventing any measure of efficacy). However, in contrast to Luby et al. (2007), they

reported no difference between vaccinated and unvaccinated cows in milk antibody titres. There is no data on the use of Lysigin in dairy cattle in New Zealand.

The second vaccine available in New Zealand, StartVac (HIPRA, Spain), contains a combined lysed culture of *S. aureus* and *E.coli* J5 antigen (Freick et al., 2016). The inactivated *S. aureus* SP140 strain used in the vaccine expresses the slime-associated antigenic complex. This complex is the biofilm produced by some *S. aureus* to evade the host immune system Arciola et al. (2015).

Landin et al. (2015) vaccinated 365 cows in two herds with the StartVac vaccine, leaving 391 unvaccinated controls. The vaccination programme was based on the manufacturer's instructions: i.e., first dose 45 days before the expected calving day, followed by second and third doses 35 and 62 days later, respectively. Milk samples were collected from cows with clinical or subclinical mastitis for the first 4 months after calving. Approximately 30% of both vaccinated and controls had clinical or subclinical mastitis. There was no statistically significant effect of vaccination on the odds of mastitis due to *S. aureus* (odds ratio for vaccinated vs unvaccinated was 1.03 with p = 0.93). There was no significant differences found between vaccinated and control groups in terms of milk production (p = 0.69), new subclinical mastitis (p = 0.75), or culling (p = 0.62).

In contrast, the study by Schukken et al. (2014) reported a protective effect of StartVac vaccination using the manufacturer's protocol. They reported data from 343 lactations, from cows which were fully vaccinated, and 658 lactations from unvaccinated cows. These authors reported that the basic reproduction ratio, R_0, for vaccinated animals was 0.89 (95% CI = 0.44 to 1.57) and that for vaccinated animals was 1.72 (95% CI = 1.06 to 3.17). This risk reduction was principally due to the cure rates in vaccinated cattle being 41% higher than in controls.

To my knowledge, the studies by Schukken et al. (2014) and Landin et al. (2015) are the only peer-reviewed efficacy studies of StartVac published in the literature for *S. aureus* mastitis. There are other published studies of StartVac (Bradley et al., 2015a; Kawai et al., 2021), but these have focused on the *E. coli* component of the vaccine, and thus are not relevant to this review.

Even in the studies reporting favourable results, the impact of *S. aureus* vaccination on *S aureus* IMI can be considered, at best moderate. Schukken et al. (2014)

suggested that farm management would be crucial for vaccination success with farms following hygienic milking, segregation and culling of known chronic infected animals, having improved outcomes compared with farms with poor management.

In addition, the protection against *S. aureus* by Lysigin and StartVac is not very broad. Lysigin protection is restricted to the CP or SP strains in the vaccine, whereas StartVac is restricted to *S. aureus* that are producing biofilms. This means any variation in the strain or failure of *S. aureus* to produce biofilms can result in low to no vaccine-related protection against *S. aureus* mastitis (Luby et al., 2007; Darwish and Asfour, 2013; Lee et al., 2014; Raafat et al., 2019).

None of the cited studies evaluated cell-mediated immunity, which has been shown to be crucial for clearing *S. aureus* from the udder (Broker et al., 2016). In addition, Th17 cells (a sub-population of T-cells) have been shown to be crucial in triggering the inflammatory response through the mobilisation of neutrophils and stimulating the production of the cytokine IL-17A (Kanevsky-Mullarky et al., 2013; Murphy et al., 2019). More research is needed to identify antigens that can stimulate both humoral and cell-mediated immunity against *S. aureus* mastitis, eliciting a protective immune response. Reverse vaccinology can be integral to this process, as it can provide a more comprehensive and faster alternative than conventional vaccinology.

1.5 STRUCTURE OF THE THESIS

The objectives of the PhD project were:

1. To analyse the population genetic structure of *S. aureus* in dairy cattle in New Zealand. This is addressed in Chapter 3 using whole-genome sequencing and bioinformatics methods.

2. The evidence obtained in the study presented in Chapter 3 was so compelling, that it was necessary to perform comparative genomic analyses of *S. aureus* in various hosts, in particular, comparisons between human and bovine *S. aureus* in New Zealand. Hence, Chapter 4 presents a comparative genomics analysis of *S. aureus* across cattle, humans, pets, and small-ruminants.

3. To analyse the concordance between resistome analysis and phenotypic antimicrobial susceptibility testing of bovine *S. aureus*. Chapter 5 presents the comparison of disk diffusion testing of the *S. aureus* isolates from Chapter 3 with the antimicrobial resistomes identified from their genomes.

4. To utilise the knowledge on the population genetic structure obtained in Chapter 3 to identify potential vaccine candidate proteins using an *in silico* reverse vaccinology pipeline. This is presented in Chapter 6 where established bioinformatics packages were used to shortlist potential vaccine candidate proteins.

This thesis is presented in a traditional thesis format, with the intention to submit each chapter as a standalone publication. As a consequence, there is some unavoidable repetition. In particular, some of the material discussed in the literature review, and material and methods may be repeated in other chapters.

All the bioinformatics code used in this project have been compiled and stored in the Github repository. The tables and figures that were too big to be included in the thesis document were also stored in the Github repository. All these can be accessed at `https://github.com/jabinnes/Thesis-files`.

2

GENERAL MATERIALS AND METHODS

2.1 ORIGIN OF *STAPHYLOCOCCUS AUREUS* ISOLATES

The New Zealand bovine *S. aureus* isolates used in this study (n=188) were selected from a bank of isolates collected over a period of about 17 years by Dr Alex Grinberg of the School of Veterinary Science (SoVS), Massey University, New Zealand. They were donated by various collaborators and commercial diagnostic laboratories and were stored at the Microbiology laboratory of the School. The isolates were obtained during three periods: 2002-03; 2012-13; and 2018-2019. When selecting isolates for this project, efforts were made to include isolates originating from a large number of farms and a wide geographical spread. However, it was not possible to ascertain whether farms had contributed isolates in more than one collection period.

Most small ruminant *S. aureus* isolates (n=30) were collected during a survey of the prevalence of udder defects in non-dairy sheep, conducted in the Lower North Island in 2017-18 and two small ruminant isolates were obtained from veterinary diagnostic laboratories (in 2012 and 2018). The details of the bovine and small ruminant isolates are reported in Tables 2.1 and 2.2 respectively. Bovine and small ruminant *S. aureus* isolates were stored in glycerol at -80°C.

For the comparative genomic analysis, previously published genome sequences (n=59) of New Zealand human, canine, and feline *S. aureus* (Grinberg et al., 2017) were used as listed in Table 2.3. These sequences were publicly available as reads in the Sequence Read Archive (SRA) of the US National Center for Biotechnology Information (NCBI) (Leinonen et al., 2011), as BioProject number PRJNA391123. They were retrieved using the fastq-dump utility in the sra-tool kit (Leinonen et al., 2011) with the ‘ --split-files’ option to convert the downloaded sra file into two paired-end fastq files.

The available details and metadata of the isolates and genomes used in this project are reported in the following sections.

2.1.1 BOVINE ISOLATES FROM 2002-2003

The bovine isolates from 2002-2003 represented the almost complete set of isolates obtained from bovine milk specimens submitted to veterinary diagnostic laboratories from Hamilton and Palmerston North, New Zealand over a period of about 6 months, ending in February 2003. These isolates had been used for a previously published study (Grinberg et al., 2005). Laboratory staff were requested to submit multiple samples from the same farms but avoid duplicates from the same cow. Unique confidential identifiers for the farm of origin were supplied, without the geographical location of the farm. Other data supplied were the cow's tag number and the laboratory location (Palmerston North or Hamilton). As previously described, the isolates were phenotypically re-identified as *S. aureus* (Grinberg et al., 2005) and stored in glycerol at -80°C after identification. A total of 116 isolates were available. Two isolates originated from two-quarters of the same cow, so one was arbitrarily eliminated. Fifty-nine out of 115 isolates were selected to represent one isolate per farm.

2.1.2 BOVINE ISOLATES FROM 2012-13

A total of 65 bovine isolates from 2012-13 were selected from a frozen collection of 134 isolates obtained from milk samples submitted to veterinary laboratories from Auckland, Hamilton, Palmerston North, Christchurch and Dunedin, New Zealand. Data available included the name of the sourcing laboratory, the laboratory case number, date of collection and a descriptor of the geographical area of origin of the sample. No farm identifier was supplied. However, since the last three digits of the accession number represented the cow number, and the rest of the digits indicated the farm number, unique farm identifier could be assigned based on the accession number. For this study, one isolate was selected at random from each farm, for a total of 53 isolates. Subsequently, 12 additional isolates (one per cow) were selected from farms supplying multiple isolates, allowing also an analysis of multiple isolates per farm. Hence, a total of 65 isolates were used from this collection.

2.1.3 BOVINE ISOLATES FROM 2017-18

A total of 64 isolates from 2017-2018 were used. These isolates originated from two different sources. Fifty-one isolates were selected from a set of 79 isolates collected during a clinical trial of mastitis therapeutics performed on 11 dairy farms in central North Island. The isolates were donated for this study by Dr Mick Clews, Vetora Limited. Isolates from this collection were included in the study based on the following criteria: Firstly, from each farm, one isolate was selected at random from each treatment group (treated and control groups), for a total of 17 isolates (five farms had only isolates from either treatment or control groups). Secondly, 16 additional isolates from each farm, irrespective of groups, were selected randomly. The remaining 18 isolates were selected randomly from the post-treatment isolates obtained from non-cured cases. The remaining 13 isolates from 2017-18 were obtained from veterinary diagnostic laboratories in Hamilton and Palmerston North, New Zealand. The isolates arrived accompanied by the laboratory accession number. The last three digits of the accession number represented the cow number, and the rest of the digits indicated the farm number.

2.1.4 SMALL RUMINANT ISOLATES

Thirty sheep isolates originated from a field study assessing the prevalence of udder defects and bacterial pathogens in non-dairy-breed ewes performed on 11 sheep farms in the lower North Island in 2018-2019 (Ridler et al., 2021). Twelve of these isolates were obtained from sheep showing a variety of udder defects, 7 from sheep showing hard mastitis lesions, and six from sheep showing lumpy udder lesions. Five isolates originated from sheep with no lesions. The isolates came with unique identifiers for the farm and animal of origin. Two additional isolates were provided by a diagnostic laboratory, and were isolated in 2012 (from a goat) and 2018 (from a sheep).

Table 2.1: *S. aureus* isolates from cattle used in this study.

Original unique identifier	Geographical origin of farm (if available), or sourcing lab	Date of isolation (if available), or collection period	Assigned farm identifier
BM33	Te Kauwhata	16/11/2012	F01
BM34	Te Kauwhata	17/11/2012	F01
BM35	Returua	3/12/2012	F02
BM36	Reporoa	12/12/2012	F03
BM37	Franklin	12/12/2012	F04
BM38	Kamu	12/12/2012	F05
BM39	Returua	12/12/2012	F02
BM40	Kamu	12/12/2012	F05
BM41	Morrinsville	12/12/2012	F07
BM42	Morrinsville	12/12/2012	F07
BM43	Reporoa	12/12/2012	F03
BM45	Eltham	7/01/2013	F08
BM46	Dannevirke	5/01/2013	F09
BM49	Edgecumbe	7/01/2013	F10
BM51	Inglewood	7/01/2013	F12
BM52	Eltham	7/01/2013	F134
BM53	Rotorua	7/01/2013	F13
BM54	Bulls	5/01/2013	F14
BM55	Tepuke	11/01/2013	F15
BM57	Okato	14/01/2013	F16
BM58	Belmont	9/01/2013	F17
BM59	Wairoa	14/01/2013	F18
BM60	Te Aroha	14/01/2013	F19
BM61	Rotorua	14/01/2013	F13
BM63	Cambridge	17/01/2013	F20
BM64	Tirau	16/01/2013	F21
BM65	Helensville	7/01/2013	F22
BM66	Hastings	9/01/2013	F23
BM67	Hawera	10/01/2013	F24

Continued on next page

Original unique identifier	Geographical origin of farm (if available), or sourcing lab	Date of isolation (if available), or collection period	Assigned farm identifier
BM69	Palmerston North	14/01/2013	F25
BM71	Palmerston North	14/01/2013	F26
BM72	Eltham	16/01/2013	F08
BM73	Te Aroha	21/01/2013	F19
BM75	Tokoroa	19/01/2013	F27
BM129	Ashburton	5/04/2013	F27
BM130	Ashburton	5/04/2013	F28
BM134	Ashburton	4/05/2013	F29
BM136	Ashburton	4/05/2013	F29
BM128	Canterbury	3/04/2013	F30
BM132	Christchurch	22/04/2013	F31
BM87	Dannevirke	25/01/2013	F32
BM124	Dannevirke	22/02/2013	F33
BM126	Dannevirke	22/02/2013	F33
BM118	Dunsandel	22/04/2013	F34
BM137	Edendale	18/04/2013	F35
BM123	Edgecumbe	28/02/2013	F36
BM88	Eltham	28/01/2013	F37
BM89	Eltham	28/01/2013	F37
BM114	Eltham	22/02/2013	F38
BM81	Gordonton Road	24/01/2013	F39
BM133	Greymouth	22/04/2013	F40
BM113	Hawera	23/02/2013	F41
BM116	Murchison	22/02/2013	F06
BM117	Murchison	22/02/2013	F42
BM77	Murupara	21/01/2013	F43
BM90	Raetihi	9/02/2013	F44
BM112	Rakaia	22/04/2013	F45
BM76	Rotorua	21/01/2013	F46
BM119	Rotorua	28/02/2013	F47

Continued on next page

Original unique identifier	Geographical origin of farm (if available), or sourcing lab	Date of isolation (if available), or collection period	Assigned farm identifier
BM110	Upper Hutt	25/02/2013	F48
BM111	Upper Hutt	25/02/2013	F48
BM85	Vetfocus Potaruru	24/01/2013	F49
BM131	Waimate	5/03/2013	F50
BM109	Waipapa	18/01/2013	F51
BM92	Whangarei	18/01/2013	F52
H01	Palmerston North Lab	2017-18 collection	F54
H02	Palmerston North Lab	2017-18 collection	F55
H03	Palmerston North Lab	2017-18 collection	F56
H04	Palmerston North Lab	2017-18 collection	F55
H05	Palmerston North Lab	2017-18 collection	F54
H06	Palmerston North Lab	2017-18 collection	F57
H07	Palmerston North Lab	2017-18 collection	F58
H09	Palmerston North Lab	2017-18 collection	F55
H11	Palmerston North Lab	2017-18 collection	F55
H12	Palmerston North Lab	2017-18 collection	F58
H08	Palmerston North Lab	2017-18 collection	F59
H10	Palmerston North Lab	2017-18 collection	F60
H13	Palmerston North Lab	2017-18 collection	F61
S2	Central North Island	2017-18 study	F62
S6	Central North Island	2017-18 study	F62
S7	Central North Island	2017-18 study	F62
S9	Central North Island	2017-18 study	F63
S10	Central North Island	2017-18 study	F63
S15	Central North Island	2017-18 study	F64
S21	Central North Island	2017-18 study	F65
S24	Central North Island	2017-18 study	F66
S27	Central North Island	2017-18 study	F65
S28	Central North Island	2017-18 study	F65
S31	Central North Island	2017-18 study	F65

Continued on next page

Original unique identifier	Geographical origin of farm (if available), or sourcing lab	Date of isolation (if available), or collection period	Assigned farm identifier
S34	Central North Island	2017-18 study	F67
S35	Central North Island	2017-18 study	F67
S39	Central North Island	2017-18 study	F65
S40	Central North Island	2017-18 study	F65
S41	Central North Island	2017-18 study	F65
S43	Central North Island	2017-18 study	F68
S44	Central North Island	2017-18 study	F68
S48	Central North Island	2017-18 study	F63
S65	Central North Island	2017-18 study	F69
S66	Central North Island	2017-18 study	F69
S79	Central North Island	2017-18 study	F64
S80	Central North Island	2017-18 study	F64
S84	Central North Island	2017-18 study	F65
S86	Central North Island	2017-18 study	F65
S87	Central North Island	2017-18 study	F65
S88	Central North Island	2017-18 study	F65
S89	Central North Island	2017-18 study	F65
S96	Central North Island	2017-18 study	F70
S97	Central North Island	2017-18 study	F62
S98	Central North Island	2017-18 study	F62
S105	Central North Island	2017-18 study	F62
S153	Central North Island	2017-18 study	F70
S154	Central North Island	2017-18 study	F70
S155	Central North Island	2017-18 study	F70
S159	Central North Island	2017-18 study	F66
S171	Central North Island	2017-18 study	F65
S107	Central North Island	2017-18 study	F62
S108	Central North Island	2017-18 study	F62
S109	Central North Island	2017-18 study	F62
S156	Central North Island	2017-18 study	F62

Continued on next page

Original unique identifier	Geographical origin of farm (if available), or sourcing lab	Date of isolation (if available), or collection period	Assigned farm identifier
S110	Central North Island	2017-18 study	F62
S166	Central North Island	2017-18 study	F62
S138	Central North Island	2017-18 study	F71
S144	Central North Island	2017-18 study	F69
S142	Central North Island	2017-18 study	F69
S118	Central North Island	2017-18 study	F72
S117	Central North Island	2017-18 study	F73
S181	Central North Island	2017-18 study	F63
S182	Central North Island	2017-18 study	F63
S183	Central North Island	2017-18 study	F63
ST99	Hamilton Lab	2002-03 collection	F74
ST136	Hamilton Lab	2002-03 collection	F75
ST117	Hamilton Lab	2002-03 collection	F76
ST120	Hamilton Lab	2002-03 collection	F77
ST130	Hamilton Lab	2002-03 collection	F78
ST132	Hamilton Lab	2002-03 collection	F79
ST116	Hamilton Lab	2002-03 collection	F80
ST119	Hamilton Lab	2002-03 collection	F81
ST137	Hamilton Lab	2002-03 collection	F82
ST143	Hamilton Lab	2002-03 collection	F83
ST206	Palmerston North Lab	2002-03 collection	F84
ST199	Palmerston North Lab	2002-03 collection	F85
ST147	Palmerston North Lab	2002-03 collection	F86
ST157	Palmerston North Lab	2002-03 collection	F87
ST173	Palmerston North Lab	2002-03 collection	F88
ST202	Palmerston North Lab	2002-03 collection	F89
ST220	Palmerston North Lab	2002-03 collection	F90
ST153	Palmerston North Lab	2002-03 collection	F91
ST135	Palmerston North Lab	2002-03 collection	F92
ST150	Palmerston North Lab	2002-03 collection	F93

Continued on next page

Original unique identifier	Geographical origin of farm (if available), or sourcing lab	Date of isolation (if available), or collection period	Assigned farm identifier
ST164	Palmerston North Lab	2002-03 collection	F94
ST166	Palmerston North Lab	2002-03 collection	F95
ST185	Palmerston North Lab	2002-03 collection	F96
ST180	Palmerston North Lab	2002-03 collection	F97
ST168	Palmerston North Lab	2002-03 collection	F98
ST177	Palmerston North Lab	2002-03 collection	F99
ST213	Palmerston North Lab	2002-03 collection	F100
ST205	Palmerston North Lab	2002-03 collection	F101
ST144	Palmerston North Lab	2002-03 collection	F102
ST154	Palmerston North Lab	2002-03 collection	F103
ST155	Palmerston North Lab	2002-03 collection	F104
ST156	Palmerston North Lab	2002-03 collection	F015
ST159	Palmerston North Lab	2002-03 collection	F016
ST160	Palmerston North Lab	2002-03 collection	F107
ST163	Palmerston North Lab	2002-03 collection	F108
ST167	Palmerston North Lab	2002-03 collection	F109
ST169	Palmerston North Lab	2002-03 collection	F111
ST174	Palmerston North Lab	2002-03 collection	F112
ST176	Palmerston North Lab	2002-03 collection	F113
ST181	Palmerston North Lab	2002-03 collection	F114
ST182	Palmerston North Lab	2002-03 collection	F115
ST187	Palmerston North Lab	2002-03 collection	F116
ST188	Palmerston North Lab	2002-03 collection	F117
ST189	Palmerston North Lab	2002-03 collection	F118
ST196	Palmerston North Lab	2002-03 collection	F119
ST211	Palmerston North Lab	2002-03 collection	F120
ST212	Palmerston North Lab	2002-03 collection	F121
ST198	Palmerston North Lab	2002-03 collection	F122
ST200	Palmerston North Lab	2002-03 collection	F123
ST203	Palmerston North Lab	2002-03 collection	F124

Continued on next page

Original unique identifier	Geographical origin of farm (if available), or sourcing lab	Date of isolation (if available), or collection period	Assigned farm identifier
ST204	Palmerston North Lab	2002-03 collection	F125
ST208	Palmerston North Lab	2002-03 collection	F126
ST209	Palmerston North Lab	2002-03 collection	F127
ST215	Palmerston North Lab	2002-03 collection	F128
ST216	Palmerston North Lab	2002-03 collection	F129
ST217	Palmerston North Lab	2002-03 collection	F130
ST218	Palmerston North Lab	2002-03 collection	F131
ST219	Palmerston North Lab	2002-03 collection	F132
ST222	Palmerston North Lab	2002-03 collection	F133

Table 2.2: *S.aureus* isolates from small ruminants used in this study.

Original unique identifier	Geographical origin (if available)	Date of isolation (if available)	Assigned farm number
BM50	Waitara	2013-14	F11
GM1		2017-18	F53
420R	Central North Island	07/11/2017	F01
2JR	Central North Island		F02
5JR	Central North Island		F02
6JR	Central North Island		F02
4101R	Central North Island	17/10/2017	F03
17	Central North Island	28/11/2017	F03
11	Central North Island	28/11/2017	F03
4305R	Central North Island	20/11/2017	F03
16BT-A	Central North Island		F04
15PW	Central North Island		F05
17BT	Central North Island		F04
11MR	Central North Island		F06
3JR	Central North Island		F02
12GS	Central North Island		F07
9IH	Central North Island		F08

Continued on next page

Original unique identifier	Geographical origin (if available)	Date of isolation (if available)	Assigned farm number
15IH-B	Central North Island		F08
1IH	Central North Island		F08
NP2	Central North Island		F09
NP1	Central North Island		F09
15IH-A	Central North Island		F08
NP15	Central North Island		F09
2IH	Central North Island		F08
5	Central North Island	28/11/2017	F03
2	Central North Island	28/11/2017	F03
3	Central North Island	28/11/2017	F03
NP3	Central North Island		F09
3JB	Central North Island		F10
5JB	Central North Island		F10

Table 2.3: *S. aureus* isolates from human, feline and canine hosts used in this study.

SRA Accession number	Original unique identifier	Source of isolate
SRR5714651	H44	Canine clinical
SRR5714659	H31	Canine clinical
SRR5714675	H57	Canine clinical
SRR5714690	H3	Canine clinical
SRR5714693	H8	Canine clinical
SRR5714696	H9	Canine clinical
SRR5714699	H21	Canine clinical
SRR5714704	H26	Canine clinical
SRR5714706	H30	Canine clinical
SRR5714653	H46	Canine colonisation
SRR5714654	H47	Canine colonisation
SRR5714655	H48	Canine colonisation
SRR5714658	H32	Canine colonisation
SRR5714664	H33	Canine colonisation
SRR5714684	H34	Canine colonisation
SRR5714694	H7	Canine colonisation

Continued on next page

SRA Accession number	Original unique identifier	Source of isolate
SRR5714700	H22	Canine colonisation
SRR5714702	H28	Canine colonisation
SRR5714705	H29	Canine colonisation
SRR5714676	H56	Feline clinical
SRR5714689	H4	Feline clinical
SRR5714691	H6	Feline clinical
SRR5714692	H5	Feline clinical
SRR5714652	H45	Feline colonisation
SRR5714695	H10	Feline colonisation
SRR5714648	H41	Human clinical
SRR5714650	H43	Human clinical
SRR5714660	H20	Human clinical
SRR5714661	H19	Human clinical
SRR5714670	H18	Human clinical
SRR5714671	H17	Human clinical
SRR5714677	H55	Human clinical
SRR5714678	H54	Human clinical
SRR5714679	H53	Human clinical
SRR5714680	H52	Human clinical
SRR5714681	H51	Human clinical
SRR5714683	H59	Human clinical
SRR5714687	H2	Human clinical
SRR5714703	H25	Human clinical
SRR5714649	H42	Human colonisation
SRR5714656	H49	Human colonisation
SRR5714657	H50	Human colonisation
SRR5714662	H36	Human colonisation
SRR5714663	H35	Human colonisation
SRR5714665	H37	Human colonisation
SRR5714666	H14	Human colonisation
SRR5714667	H13	Human colonisation
SRR5714668	H12	Human colonisation
SRR5714669	H11	Human colonisation

Continued on next page

SRA Accession number	Original unique identifier	Source of isolate
SRR5714672	H16	Human colonisation
SRR5714673	H15	Human colonisation
SRR5714674	H58	Human colonisation
SRR5714682	H38	Human colonisation
SRR5714685	H40	Human colonisation
SRR5714686	H39	Human colonisation
SRR5714688	H1	Human colonisation
SRR5714697	H23	Human colonisation
SRR5714698	H24	Human colonisation
SRR5714701	H27	Human colonisation

2.2 RESUSCITATION OF FROZEN ISOLATES AND DNA EXTRACTION

A scrape of frozen bacteria was extracted from the vial, deposited and streaked on the surface of a 5% sheep Columbia Blood Agar plate (Fort Richard Laboratories, Auckland, NZ), and incubated aerobically overnight at 35-37°C.

Bacterial growth was visually assessed for the presence of a pure culture. *S. aureus* was re-identified based on the presence of round, golden-white colonies consisting of Gram-positive, catalase-positive cocci which were positive by latex agglutination for clumping factor and/or protein A (Staphaurex, Thermo Fisher Scientific, Auckland, NZ). Eleven isolates selected at random were submitted for Matrix-Assisted Laser Desorption Ionisation-Time Of Flight Mass Spectrometry (MALDI-TOF MS) identification at Massey University. All these isolates were identified as *S. aureus* by MALDI-TOF MS with strong hits to *S. aureus* (log score of 2.2 to 2.5).

DNA was extracted from the isolates using the DNeasy UltraClean Microbial Kit (Qiagen, Auckland, NZ) following the manufacturer's protocol except for the use of a mini bead-beater instead of the Vortex adapter. Briefly, six individual colonies from each plate were picked using a sterile bacterial loop and suspended in 1.8 mL phosphate buffer solution (PBS). This suspension was centrifuged at 10000 x g for 30

seconds and the cell pellet was resuspended in the solutions (PowerBead solution and Solution SL) provided in the kit, in a PowerBead tube, to lyse the bacterial cells using a mini bead-beater (Biospec, Oklahoma, USA) for 2 minutes. This suspension was centrifuged at 10000 x g for 30 seconds and the supernatant was transferred to a clean collection tube provided in the kit. The supernatant was mixed with Solution IRS (provided in the kit) to precipitate the proteins. This suspension was centrifuged at 10000 x g for one minute after incubation at 4°C for 5 minutes. The supernatant was then transferred to a clean collection tube and mixed with Solution SB (provided in the kit) to bind the DNA and then filtered and washed using the spin column and Solution CB (provided in the kit) and eluted with Solution EB (provided in the kit). The extracted DNA was analysed for quality and quantity using a NanoDrop Micro volume Spectrophotometer (Thermo Fisher Scientific, Auckland, NZ) and a Qubit 2.0 Fluorometer (Thermo Fisher Scientific, Auckland, NZ), respectively. The DNA (suspended in Solution EB provided in the kit) samples which had satisfactory quantity (> 100 ng/mL) and quality (A_{260}/A_{280} ratio between 1.8 and 2.0) were stored at -80°C and were later used for whole-genome sequencing.

2.3 WHOLE-GENOME SEQUENCING

Library preparation of the 188 bovine and two small ruminant isolates was performed by the Massey Genome Service, Palmerston North, New Zealand, using the methods described below. The remaining thirty small ruminant *S. aureus* DNA samples were transported in person in transport kits (DNAstable LD, Sigma-Aldrich, Auckland, NZ) by Dr Alex Grinberg to the Centro di Ricerca Pediatrica Romeo ed Enrica Invernizzi, University of Milan, Italy, where they were processed.

2.3.1 LIBRARY PREPARATION

At the Massey Genome Service, library preparation was done using the Nextera XT library preparation kit (Illumina™ #FC-131-1096) (Illumina, San Diego, USA). Extracted DNA suspended in Solution EB from the DNA extraction process was used for library preparation. The frozen DNA was thawed and rechecked for quantity and quality of DNA using NanoDrop Micro volume Spectrophotometer and Qubit 2.0

Fluorometer respectively. About 5 μL of the DNA extracts were diluted to 0.2 ng/μL using lab-grade water adequate for library preparation. Library preparation involved five major steps; 1) Tagmentation, 2) Amplification of libraries, 3) Cleaning up of libraries, 4) Normalisation of libraries, and 5) Pooling of libraries.

2.3.1.1 TAGMENTATION

In each well of a 96-well plate, 2.5 μL of the diluted DNA (described in section 1.3.1), 5 μL of Tagment DNA Buffer (provided in the kit), and 2.5 μL of Amplicon Tagment Mix (provided in the kit) were mixed with a pipette and centrifuged at 280 x g at 20°C for 1 minute. This mixture was incubated at 55°C for 5 minutes in a thermal cycler and cooled to 10°C. A 5 μL Neutralising Tagment Buffer (provided in the kit) was immediately added to each well to stop the tagmentation process, centrifuged at 280 x g for 1 minute and incubated for further 5 minutes at room temperature.

2.3.1.2 AMPLIFICATION OF LIBRARIES

The tagmented DNA samples in each well were mixed with 15 μL of Nextera PCR Master Mix (provided in the kit) and 5 μL each of index adapters i7 and i5 from the TG Nextera XT Index Kit v2 Set A (Illumina™ #FC-131-2001) or TG Nextera XT Index Kit v2 Set B (Illumina™ #FC-131-2004), and the mixture centrifuged at 280 x g for 1 minute. Unique index adapter combinations were used for the two 96-well plates used in this study and the index adapters for each isolate were recorded. This mixture was then amplified using a thermal cycler under the following conditions: 72°C for 3 minutes, 95°C for 30 seconds, 12 cycles of: i) 95°C for 10 seconds, ii) 55°C for 30 seconds, iii) 72°C for 30 seconds, iv) 72°C for 5 minutes, and then maintained at 10°C.

2.3.1.3 CLEANING OF LIBRARIES

This step was used to purify the libraries using AMPure XP beads and remove any short library fragments from the previous step. About 50 μL of PCR product from each well from the previous step was transferred to a new plate and 30 μL of AMPure

XP beads were added to each well (ratio of 3:2 of PCR product: AMPure XP beads) and incubated at room temperature for 5 minutes to let the DNA fragments bind to the beads. The plate was then placed on a magnetic stand until the supernatant was clear (~2 minutes). The supernatant was carefully removed without disrupting the beads which were aggregated to the side of the well. Each well was then washed twice with 200 μL of freshly prepared 80% ethanol with the plate incubated on the magnetic stand for 30 seconds and the supernatant discarded after each wash. After removal of any residual ethanol, the plate was air-dried on the magnetic stand for 15 minutes at room temperature. The plate was then removed from the magnetic stand and 52.5 μL of Resuspension Buffer (provided in the kit) was added to each well to resuspend the contents. The plate was placed back on a magnetic stand after 2 minutes of incubation at room temperature. During this time, the DNA detached from the AMPure XP beads and became suspended in the solution while the magnetic beads were aggregated at the side of each well. About 50 μL of the solution was transferred from each well to another plate. The DNA concentration of each library was measured using a Qubit 2.0 Fluorometer (Thermo Fisher Scientific, Auckland, NZ) and their qualities were checked using DNA High Sensitivity LabChip assay (average library size) and Quant-iT dsDNA High Sensitivity assay (DNA concentration) through the Massey Genome Service, Massey University. The libraries were stored at -20°C for a maximum of seven days, before further manipulation.

2.3.1.4 NORMALISATION AND POOLING

A normalisation step was undertaken to equalise the representation of the library fragments from each isolate, as they were pooled together before sequencing. A 20 μL supernatant from each well of the PCR plate from the previous step was transferred to a new PCR plate. A library normalisation (LN) master mix was prepared by mixing 46 μL Library Normalisation Additive 1 (LNA1) and 8 μL Library Normalisation Bead 1 (LNB1) (both provided in the kit) per isolate. Of the LN master mix 45 μL was added to each well in the PCR plate and the plate was sealed and mixed using a plate shaker at 1800 rpm for 30 minutes. The plate was then placed on a magnetic stand and left undisturbed until the supernatant was clear for approximately 2 minutes. The supernatant was discarded without disturbing the beads and each well was washed twice with LN wash 1 reagent, mixed using the plate shaker at 1800 rpm for 5 minutes, incubated on a magnetic stand for 2 minutes and

the supernatant discarded. Thirty μL of 0.1 N NaOH was added to each well and mixed using plate shaker at 1800 rpm for 5 minutes and placed on a magnetic stand for approximately 2 minutes. From this, 30 μL supernatant from each well was transferred to a new plate containing 30 μL of LN Storage 1 Buffer in each well. The plate was sealed and centrifuged at 1000 x g for 1 minute.

The libraries of 190 isolates were pooled after diluting them to equal molarity using resuspension buffer (provided in the kit).

2.3.2 SEQUENCING

Two hundred ng of the pooled library was sent to Novogene AIT (Hong Kong) in a DNAstable (Sigma-Aldrich, Auckland, NZ) tube using FedEx International Priority service through the Massey Genome Service. The libraries were quality checked and then loaded across one lane of HiSeq X sequencer and demultiplexed at the end of sequencing. The paired raw reads which contained low-quality reads and adapters were analysed and trimmed to provide clean reads. The Phred scores of the reads and GC content were measured. Thousands of random reads were BLASTed against the nucleotide NCBI database (Altschul et al., 1990) to check if the reads mapped against *S. aureus*. The data was then loaded on to a 1 TB external hard disk and was couriered to New Zealand for further downstream analysis.

2.4 WHOLE-GENOME SEQUENCE DATA PROCESSING AND ANALYSIS

The clean reads were analysed and trimmed using QCtool (`https://github.com/mtruglio/QCtool`) which utilises FastQC (Andrews, 2010), FastQscreen (Wingett and Andrews, 2018), SolexaQA++ (Cox et al., 2010), Bowtie2 (Langmead and Salzberg, 2012) and samtools (Li et al., 2009) for PhiX and adapter removal. The FastQC program utilised in the QCtool analyses the quality of the reads by measuring the Phred scores for each base in all the sequences, the no-base call (N) content of the reads, GC content, sequence duplication, overrepresented sequences, and adapter content. The FastQscreen program maps the

reads against standard databases to identify the origin/species of the target sequence (*S.aureus* in this case) and also identifies PhiX, vectors or any contamination from other sources. SolexaQA++ analyses the reads for quality and dynamically trims off the bases which do not have a good Phred score (user-defined cut-off or based on the read length returned by BWA algorithm) and sorts the reads into high quality and low quality reads. PhiX and adapter removal are done by mapping the reads against the PhiX and adapter sequences databases using Bowtie and removing any PhiX and adapter sequences left behind from the Illumina sequencing. The output of the QC-tool processing was termed as 'processed reads' for genome assembly. The processed reads were used for assembling the whole-genome sequence and initiate the bioinformatics analyses, as shown Figure 2.1.

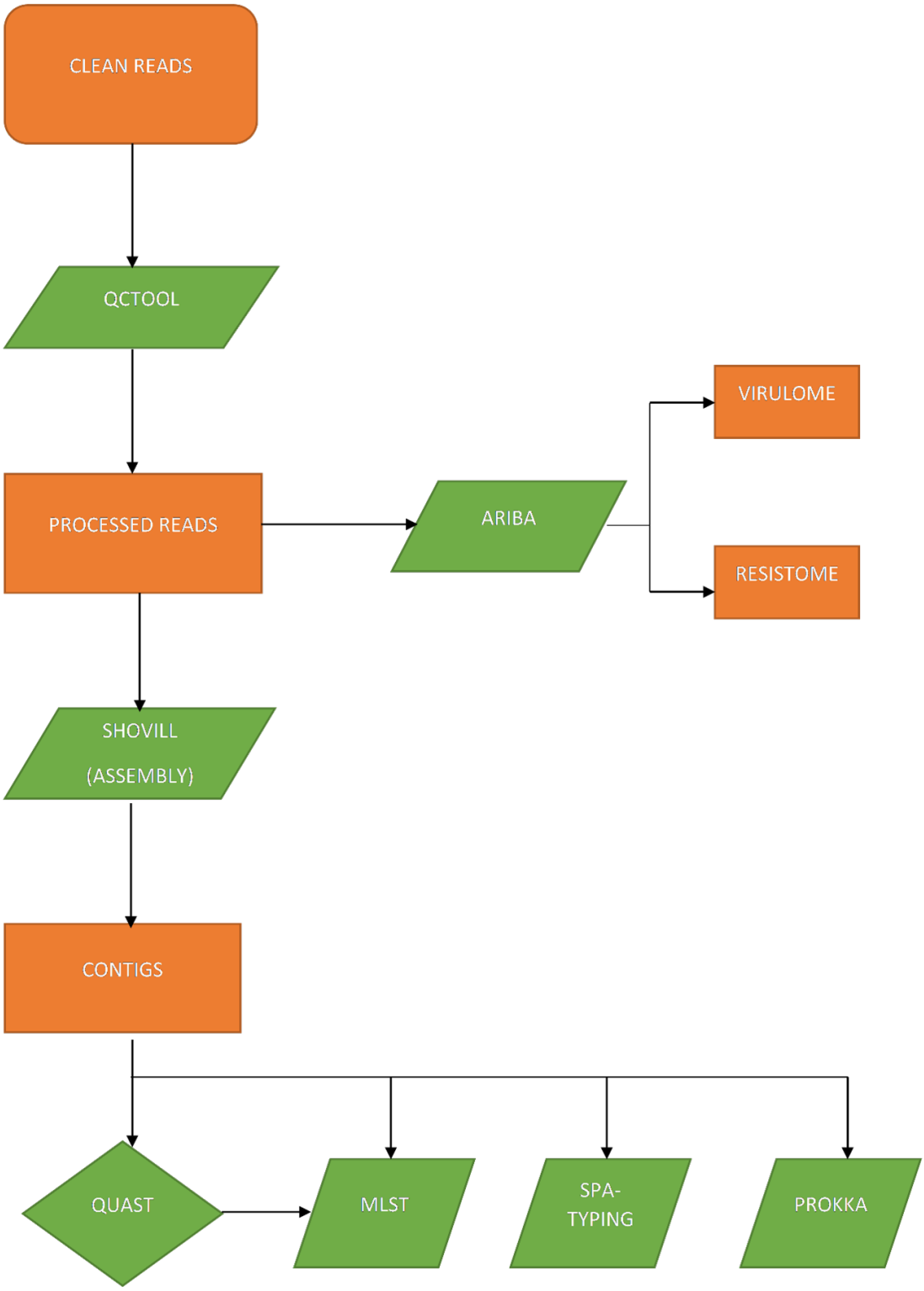

Figure 2.1: Flowchart of the whole-genome sequence data analysis workflow from clean paired-end reads. The orange boxes indicate input/output data, while the green boxes indicate bioinformatic processes. QUAST is a decision-making process as contigs with questionable data are discarded

2.4.1 ASSEMBLY

The processed paired reads were assembled into contigs using an open-source program called shovill (Seemann, 2018b) using SPAdes 3.13.1 assembler (Nurk et al., 2013) with all the other parameters set at default. An in-house python script was used to automate the process of feeding the paired-end reads to shovill (--R1 and --R2) and changing the output folder name for each isolate to a user-defined name. The shovill program estimates the genome size (--gsize; default: AUTODETECT) using KMC software (Kokot et al., 2017) and the read length using seqtk (Li, 2018). Shovill can trim adapters using trimmomatic (Bolger et al., 2014) but is turned off in the default mode (--trim; default: OFF). Shovill reduces the FASTQ files to a sensible depth (--depth; default: 100X) using seqtk, corrects sequencing errors in reads (disabled by --noreadcorr; default: OFF) using Lighter (Song et al., 2014), pre-overlaps or stitches paired-end reads (disabled by --nostitch; default: OFF) using FLASH (Magoč and Salzberg, 2011) before assembling the reads using either SPAdes, SKESA (Souvorov et al., 2018) or Megahit (Li et al., 2016) programs (--assembler; default: SPAdes). Once the assembly is done, shovill corrects minor assembly errors by mapping the reads back to their contigs (disabled by --nocorr; default: OFF) using BWA MEM (Li, 2013) and samtools (Li et al., 2009). The contigs that are too short (--minlen; default: 0) or with too low coverage (--mincov; default: 2) are removed by shovill using pilon (Walker et al., 2014). The output files are stored in the output directory (--outdir). Shovill produces output folders for each isolate containing the assembled contigs as a fasta file. This contig file was used for various downstream analysis.

The command line for shovill was as follows:

```
$ shovill --R1 read1.fastq --R2 reads2.fastq --depth 100 --gsize AUTODETECT
--outdir user-defined names --force OFF --minlen 0 --mincov 2 --keepfiles OFF
--tmpdir /tmp/jabin --cpus 4 --ram 23.14 --assembler spades --kmers AUTO
--trim OFF --noreadcorr OFF --nostitch OFF --nocorr OFF
```

2.4.2 QUALITY ANALYSIS

The assembled contigs were analysed for quality using QUAST 5.02 (Mikheenko et al., 2018) using the default settings with the following command line:

```
$ quast.py -o output directory input_contig.fasta
```

The percentage of Guanine and Cytosine (GC%) bases in the assembled genomes and the total length of the assembled genomes were used as qualifiers to identify and discard bad quality assemblies. The thresholds considered were based on a summary analysis of other studies and were: 32.6% to 32.9% GC content, and 2.6 to 2.9 Mb of total length. Isolates above or below this threshold were discarded.

2.4.3 MULTILOCUS SEQUENCE TYPING (MLST)

The STs of the isolates were identified using a bioinformatics tool called 'mlst' (Seemann, 2020), which scans the contigs files against the PubMLST database (Jolley et al., 2018) to define the allelic profiles and their corresponding sequence types (ST). Any novel allelic sequences (i.e., sequences not occurring in the PubMLST) identified using this tool were mapped back to their corresponding isolate sequence reads using SAMtools (Sequence Alignment/Map) and viewed using Geneious 6.1.8 (Biomatters, Auckland, NZ) to manually check if any single nucleotide polymorphism (SNP) was not the result of an assembly error. The SNPs that had a read depth of more than 50x were considered true SNPs and the allele was assigned as 'novel'. Identifiers for the novel alleles and the resulting novel sequence types were subsequently assigned by the curators at PubMLST after the submission of the allelic profile and the isolate profile. The STs were assigned to clonal complexes (CCs) based on the CC nomenclature available in the PubMLST database (Jolley et al., 2018), which uses the BURST algorithm (Feil et al., 2004). The database reported only 9 CCs as of February 2021. Hence, groups of single-locus variant (SLV) or double locus variant (DLV) isolates that could not be identified as one of the nine defined CCs were assigned a CC which took the number of the ST with the most SLVs found in the group. The human, canine and feline *S. aureus* isolates were assigned the same CCs assigned in the original study by Grinberg et al. (2017). Once the STs and CCs were

identified, the sample genomes could be suffixed to include the ST and CC.

2.4.4 SPA-TYPING

The spa-typing identification was performed using a tool called 'spa-typing' (https://github.com/mjsull/spa_typing). This tool identifies the repeat sequences from the assembled contigs using an *in silico* PCR with primers commonly used to identify spa regions from bacterial DNA. The identified repeat sequences are checked against the sequences stored in the Ridom database, either manually provided as a fasta file or downloaded from the Ridom spa server (http://www.ridom.de/) by the program. These sequences are provided with spa-types respective to the repeats as assigned in the Ridom database. Any novel spa-type was assigned with 't-new' due to the lack of Ridom StaphType software, for this project, to submit to the Ridom database.

2.4.5 ANNOTATION

The annotation of the assembled contigs was done using the bioinformatics tool 'Prokka' (Seemann, 2014) with default values except the locus tag flag to be assigned with the sample name. This tool uses external tools such as the Bioperl tool kit (Stajich, 2002), BLAST+(Camacho et al., 2009), Prodigal (Hyatt et al., 2010), ARAGORN (Laslett, 2004), HMMER (Finn et al., 2011), Barrnap (Seemann, 2018a) and other optional tools. The output files of Prokka used for downstream analysis were the GFF file containing annotations of the genes along with their sequences, the FAA file that was the protein fasta file containing translated sequences of the genes, and the FFN file that contained the gene sequences.

Other bioinformatics analyses varied for different studies, and are discussed in relevant chapters.

3

GENOMIC EPIDEMIOLOGY OF BOVINE MASTITIS-CAUSING *STAPHYLOCOCCUS AUREUS* IN NEW ZEALAND

3.1 INTRODUCTION

Worldwide (including in New Zealand), mastitis is considered one of the most economically important diseases of dairy cattle (Hogeveen et al., 2011; Kerslake et al., 2018). The ability of *S. aureus* to cause chronic intramammary infections (IMIs), along with its highly contagious nature, has established this bacterium as one of the major pathogens associated with bovine mastitis, even in countries where robust mastitis prevention programs are implemented (Perez et al., 2020; Rainard et al., 2018; Smith et al., 1998). In New Zealand, *S. aureus* is the second most common species isolated at clinical veterinary laboratories both from clinical and subclinical mastitis cases, after *Streptococcus uberis* (Bates et al., 2020; McDougall, 2010). *S. aureus* IMIs are of concern due to their contagiousness, and the organism's ability to survive in the phagocytic cells and mammary epithelial cells causing chronic inflammation associated with poor therapeutic cure rates. *S. aureus* IMIs lead to economic loss principally by reducing milk quality due to increased somatic cell counts, but also by reducing milk quantity, and increasing the risk of culling (Stott et al., 2002). An association between the persistence of *S. aureus* IMI and the type of infecting strains has been recently observed (Pichette-Jolette et al., 2019; Veh et al., 2015).

At the individual cow level, *S. aureus* IMI is usually caused by a single strain, with the definition of 'strain' varying according to the bacterial typing method used. Within herds, there is often a single dominant strain, but two or more strains may be present (Middleton and Fox, 2002).

S. aureus has a highly clonal population structure due to the existence of chromosomal restriction-modification systems (Lindsay, 2010), with a preserved ability to exchange genetic material by horizontal transfer. The species displays a variable presence of virulence factors and resistance genes carried by mobile genetic elements, with frequent reassortments and emergence of virulent strains.

Molecular epidemiological studies performed in several countries revealed the majority of infections in dairy cattle are caused by a limited number of clonal lineages, defined by multilocus sequence typing (MLST) as sequence types (STs), or more broadly as clonal complexes (CCs). Some of these lineages, such as CC97, CC151 and CC133, seem to be confined to cattle or ruminants (Fitzgerald, 2012;

Shepheard et al., 2013; Smith et al., 2005). Detailed genomic studies of representative 'ruminant' isolates often revealed genomic features of adaptation to the bovine mammary environment by acquisition of genes that encode ruminant-specific virulence factors, and/or loss of genes that code for human-specific factors (Guinane et al., 2010; Sung et al., 2008; Matuszewska et al., 2020). A detailed discussion of the virulence factors associated with the pathogenesis of bovine mastitis can be found in section 1.2.2.

Several studies have shown associations between the mastitis severity and the type of infecting strain (Middleton and Fox, 2002; Fox et al., 2005; Dingwell et al., 2006; Haveri et al., 2007) although this was not always the case (Middleton and Fox, 2002). It is thus likely that some of the differences in severity are due to the expression of certain virulence factors by the infecting strain (Fournier et al., 2008).

The expression of these factors is controlled by a complex system of regulatory genes which is not fully understood (Novick, 2003). Additionally, *in vitro* expression of virulence factors is not the same as their expression *in vivo*. Thus, not all potential virulence factors have yet been identified, or their role completely understood. Further investigations of the whole genomes are likely to improve our understanding of the virulence variability of *S. aureus* and enable the development of prognostic markers and vaccines.

In addition to genes coding for virulence factors, antimicrobial resistance (AMR) in *S. aureus* is a major concern, as resistance to most known antimicrobials has been identified in this species. A detailed discussion of the antimicrobial resistance mechanisms observed in *S. aureus* can be found in section 1.3.

AMR is subject to genetic control. *S. aureus* can acquire AMR genes through horizontal gene transfer from other strains or other bacterial species and/or by mutation of specific genes such as efflux pumps or their regulators. The increase of AMR has been in part attributed to the use of antimicrobials both for prophylactic and therapeutic reasons in the livestock industry (Khachatourians, 1998; Mehndiratta and Bhalla, 2014). However, in cattle in New Zealand, despite the consistent use of penicillin to treat and prevent mastitis caused by *S. aureus*, the reported resistance of *S. aureus* to penicillin has declined from 70% in 1977 to 30% in 2003-2006, and there was no evidence of resistance of *S. aureus* to cloxacillin or cefuroxime despite those antimicrobial types being the most commonly used antimicrobials to treat or prevent

mastitis (McDougall et al., 2021). These conclusions were based on a phenotypic *in vitro* assessment of resistance, and there have been no genomic studies on the prevalence of AMR genes in IMI-causing *S. aureus* in New Zealand.

Knowledge of the genetic makeup of the *S. aureus* strains circulating in New Zealand could enhance the control of IMI with this pathogen. Traditional genotyping methods for *S.aureus* (detailed in section 1.1.4), such as MLST and the identification of virulence factors and AMR genes require the use of multiple polymerase chain reaction (PCR) amplifications, which are time-consuming and labour-intensive,in particular when it is necessary to analyse large samples. These can be done more efficiently by employing whole-genome sequencing (WGS). Moreover, WGS is advantageous as it enables an evaluation of the genetic structure of bacterial species at the highest possible resolution of the single nucleotide. Thus, with the cost of sequencing decreasing, WGS is increasingly becoming a common tool for bacterial genetic studies.

Hence, the aims of this study were to: 1) use WGS to study the population genetic structure and the clonal lineages of bovine-mastitis-causing *S. aureus* circulating in New Zealand over a period of 15 years; 2) identify virulence factor genes; and 3) identify the antimicrobial resistance genes present in bovine *S. aureus* in New Zealand.

3.2 MATERIALS AND METHODS

3.2.1 ISOLATES

All the 188 bovine *S. aureus* isolates described in General Materials and Methods (section 2.1.1, 2.1.2, 2.1.3) were used in this study. These isolates were isolated from bovine milk samples submitted for bacterial culture to New Zealand veterinary laboratories (NZVP or Gribbles) during 2002-03, 2012-13 and 2017-18, as well as from milk samples collected from cases of bovine mastitis during an antimicrobial efficacy study undertaken in 2017-18.

3.2.2 ISOLATE STORAGE AND BACTERIAL CULTURE

Prior to use in this study, after identification, the isolates were stored in glycerol at -80°C. A scrape of frozen bacteria was extracted from the vial, deposited and streaked on the surface of a 5% sheep Columbia Blood Agar plate (Fort Richard Laboratories, Auckland, NZ), and incubated aerobically overnight at 35-37°C.

3.2.3 RE-IDENTIFICATION OF ISOLATES

The stored isolates were re-identified prior to this study as detailed in section 2.1.2. In brief, the incubated agar plates were inspected for the growth of a pure culture of round, golden-white colonies, and colonies meeting this criterion were selected using a sterile bacterial loop and tested using Gram staining, a catalase test and a latex agglutination test (Staphaurex, Thermo Fisher Scientific, Auckland, NZ). The Staphaureux latex agglutination test detects agglutination in the presence of the *S. aureus* protein A and/or the Clumping Factor. Gram-positive, catalase-positive and latex agglutination-positive isolates were confirmed as *S. aureus* and selected for DNA extraction. Eleven of the 188 confirmed isolates were selected at random and tested using Matrix-Assisted Laser Desorption Ionization-Time Of Flight Mass Spectrometry (MALDI-TOF MS) at Massey University. All had strong hits to *S. aureus* (a log score of 2.2 to 2.5).

3.2.3.1 DNA EXTRACTION

The DNA extraction method is detailed in section 2.1.2. In brief, a suspension of *S. aureus* in phosphate buffer solution was used for DNA extraction. The DNA of *S. aureus* was then extracted using a DNeasy UltraClean Microbial Kit (Qiagen, Maryland, USA) following the manufacturer's protocol except that a mini bead-beater (Biospec, Oklahoma, USA) was used for 2 minutes instead of the Vortex adapter. The extracted DNA was analysed for quality using a NanoDrop Microvolume Spectrophotometer (Thermo Fisher Scientific, Auckland, NZ), and for quantity using a Qubit 2.0 fluorometer (Thermo Fisher Scientific, Auckland, NZ). Extracted DNA of satisfactory quality (A_{260}/A_{280} ratio between 1.8 and 2.0) and quantity (> 100ng/ml)

was stored at -80°C until sequenced.

3.2.4 WHOLE-GENOME SEQUENCING

The whole-genome sequencing was undertaken as detailed in section 2.1.3. The library preparation was done by Massey Genome Service, Palmerston North, New Zealand. In brief, the libraries of the extracted DNA diluted to a concentration of 0.2 ng/μL were prepared using the Nextera XT library preparation kit (Illumina #FC-131-1096) (Illumina, San Diego, USA). The library preparation involved five main steps: 1) tagmentation, 2) amplification of libraries, 3) clean-up of libraries, 4) normalisation of libraries, and 5) pooling of libraries. The DNA fragments were tagged with an index using the TG Nextera XT Index Kit v2 Set A (Illumina#FC-131-2001) or the TG Nextera XT Index Kit v2 Set B (Illumina #FC-131-2004).

Two hundred ng of the library pool were sent to a commercial sequencing provider (Novogene AIT, Hong Kong) in a transport kit at ambient temperature (DNAstable tube, Sigma-Aldrich, Auckland, New Zealand) using the priority service of an international courier through Massey Genome Service. The pooled library was quality checked at the commercial facility and then loaded across one lane of a HiSeq X (150bp PE) sequencer and demultiplexed at the end of sequencing. Low-quality reads and adapters were removed from the paired raw reads to provide clean reads. The Phred scores of the reads and GC content were measured. Thousands of random reads were blasted against the nucleotide NCBI database (Altschul et al., 1990) to confirm whether the reads mapped against *S. aureus*. The resultant sequence reads were stored on a 1TB hard disk (Seagate Technology, California, USA) for further downstream analysis.

3.2.5 READ QUALITY CHECK, GENOME ASSEMBLY AND ASSEMBLY QUALITY ANALYSIS

The data obtained from the sequencing provider included reads containing adapter sequences, PhiX vectors, low-quality reads and contamination sequences that may

affect the reliability of the assembly and further downstream analysis (Del Fabbro et al., 2013). Hence, the reads were processed using the QCtool (https://github.com/mtruglio/QCtool) which assesses the quality of the reads and the presence of adapters of each isolate using FastQC, SolexaQA++, Bowtie. This process was done to remove low-quality reads or adapters left behind after the quality check by the sequencing providers. The QCtool was run with default parameters with FastQC, FastQscreen, SolexaQA++, Phix vector removal and adapter trimming tools turned on.

The output of the QCtool was considered as the processed reads ready for bioinformatic analysis. The processed reads were *de novo* assembled by an assembly optimiser, Shovill (Seemann, 2018b), using the SPAdes assembler (Bankevich et al., 2012). The assembly quality was then assessed using QUAST (Gurevich et al., 2013) for various parameters as detailed in section 2.1.4.2. The percentage of guanine and cytosine (GC%) bases in the assembled genomes and the total length of the assembled genomes were used as qualifiers to identify and discard bad quality assemblies. The thresholds considered were based on a summary analysis of other studies, and were: 32.6% to 32.9% GC content, and genome length between 2.6 and 2.9 Mb. Genomes above or below this threshold were discarded.

3.2.6 MLST AND SPA TYPING

MLST and Spa-typing, two highly discriminatory and widely used typing methods for *S. aureus* (Koreen et al., 2004; Saunders and Holmes, 2007), were used in this study to infer phylogenies. An MLST scheme that classifies isolates into Sequence Types (ST) based on the sequence variation within seven housekeeping genes, was used to establish the clonal descent of the isolates (Enright et al., 2000).

The contigs produced in the assembly process were used to identify the STs of the isolates using the mlst tool described in section 2.1.4.3 (Seemann, 2020). The STs were assigned to clonal complexes (CCs) based on the CC nomenclature available in the PubMLST database (Jolley et al., 2018), which uses the BURST algorithm (Feil et al., 2004). The database reported only 9 CCs as of February 2021. Hence, groups of single-locus variant (SLV) or double locus variant (DLV) isolates that were not identified as one of the nine defined CCs were assigned a CC which took the name of

the ST with the most SLVs found in the group. The variant loci of any novel ST (i.e., any ST not reported in the PubMLST) were mapped against the reads of the isolates using the samtools (Li et al., 2009) software, to verify that it was not generated by assembly errors. Novel STs were submitted to the PubMLST database, where unique new identifiers were assigned, and those identifiers were used in this study.

Spa-typing, another typing method used for the discrimination of *S. aureus* isolates, was also used. The spa-typing method analyses variation in the repeat region of the Protein A gene (Frénay et al., 1996). The spa-types were identified using spa-typing software (https://github.com/mjsull/spa_typing). Assignment of any new spa-type was not attempted due to the unavailability of the RidomStaphType software for this study.

3.2.7 PANGENOME, CORE GENOME AND ACCESSORY GENOME ANALYSIS

3.2.7.1 GENERAL OVERVIEW OF ANALYSIS

S. aureus displays an open pangenome, with the number of detected genes increasing as sample size increases until reaching the plateau (Medini et al., 2005). A species pan-genome includes genes that form the core and accessory genome. The core genome is composed of genes present in the vast majority of the strains and mostly comprises genes encoding for basic, essential functions, but also pathogenicity and virulence genes (Medini et al., 2005). The accessory genome is composed of genes observed only in a certain proportion of strains and usually encoding for factors related to virulence and antimicrobial resistance. Accessory genes are predominantly mobile genetic elements carried in transposons, pathogenicity islands and bacteriophages. Since this study used draft genomes, the core genome was defined as the whole set of genes that were retrieved in all the isolates, whereas the rest of the genes composed the accessory genome.

The delineation and annotation of the coding sequences (CDS) of the assembled contigs were performed using Prokka (Seemann, 2014) (see section 2.1.4.5).

Core and accessory genome analyses were undertaken using Roary (Page et al.,

2015), which utilizes the annotated assembly GFF3 files produced by Prokka. Roary aligns the core genes (nucleotides) using the MAFFT program (Nakamura et al., 2018). The genes were clustered based on 95% amino-acid sequence similarity (default parameter) using blastp. Roary was run with default parameters except for core gene definition (-cd 100, i.e., present in all isolates) and the number of threads (-p4; assigning all the 4 CPU cores). The parameter 'cd' is the core gene definition that allows the user to assign their own core gene definition where the default setting is 99% of the genomes. A 'strict' definition of core gene (-cd 100) was used to include only genes that were present in all the isolates to avoid bias in downstream analyses with regards to CCs. The parameter 'p' stands for the number of threads or the number of threads of execution where a program splits its process into simultaneously running tasks. The default threads parameter was 1, which meant the process does not split to run more than one task at a time. The number of threads was increased to the maximum capacity of the computer system, to increase the speed of processing.

The Roary command line used in this study was as follows:

```
roary -f outputdirectory -e --mafft -r -p 4 -cd 100 inputdirectory/*.gff
```

Roary generates reports of summary statistics, core gene nucleotide alignment and a gene presence/absence spreadsheet. The summary statistics file lists the number of genes in the pangenome divided into four categories: 1) core genes (present in 100% of the isolates); 2) soft-core genes (present in 95% to 99% of the isolates); 3) shell genes (present in 15% to 94% of isolates); and 4) cloud genes (present in <15% of isolates). As previously mentioned, categories 2-4 were considered in this study as the accessory genome.

3.2.7.2 ANALYSIS OF THE CORE GENOMES

To compare the core genomes of all the isolates, the core gene nucleotide alignment output of Roary was used. A neighbour-joining (NJ) tree was constructed in Geneious v.6.1 (Biomatters, Auckland, New Zealand) using the nucleotide pairwise distance matrix, implementing the Jukes-Cantor model (Jukes and Cantor, 1969).

Subsequently, a separate NJ tree was constructed for each major CC with the Roary core gene alignment. Separate NJ trees were generated, as reducing the number of

genomes analysed increases the number of core genes, improving the resolution of the tree within each CC, as compared to the tree containing all the CCs (Grinberg et al., 2017).

3.2.7.3 ANALYSIS OF THE ACCESSORY GENOMES

An accessory genome analysis was essential for identifying the distribution of virulence factors and antimicrobial resistance genes within the different *S. aureus* CCs. The accessory genome analysis was performed using the gene presence-absence (binary) matrix generated by Roary. Using this binary spreadsheet, Principal Coordinate Analysis (PCoA) was implemented in PAST software (Hammer et al., 2001) to identify segregations in accessory gene composition between the CCs and the periods of collection.

3.2.8 ANALYSIS OF INDIVIDUAL GENES, VIRULENCE FACTORS AND ANTIMICROBIAL RESISTANCE GENES

It has been proposed that the presence or absence of single genes, and single mutations or truncations in influential genes might profoundly affect the phenotype of *S. aureus* (Mairpady Shambat et al., 2016). Hence, analysis of a number of potentially influential individual genes was undertaken. One of the major regulatory genes in *S. aureus* is the accessory gene regulator (*agr*) operon. The operon consists of the *agrA*, *agrB*, *agrC* and *agrD* genes. *S. aureus* can be grouped into four distinct groups based on their *agrC* and *agrD* sequences (*agr*-I, *agr*-II, *agr*-III, and *agr*-IV). In this study, the *agr* group of isolates was identified using *Agr*VATE (`https://github.com/VishnuRaghuram94/AgrVATE`), which maps the assembled contigs to an *agr* allele database and also extracts the *agr* operon using *in silico* PCR if they are intact in contigs.

Virulence genes and antimicrobial resistance genes were identified using the program ARIBA (Hunt et al., 2017), which maps the reads to reference sequences downloaded from a number of online databases, namely, the Virulence Factors Database (VFDB)

(Chen et al., 2005), VirulenceFinder (Joensen et al., 2014), Resfinder (Bortolaia et al., 2020), CARD (Alcock et al., 2019), NCBI (https://www.ncbi.nlm.nih.gov/pathogens/isolates#/refgene/), and PATRIC (Wattam et al., 2014). The virulence gene presence/absence matrix generated by ARIBA from the VFDB was analysed to identify associations of virulence factors with certain CCs.

3.3 RESULTS

3.3.1 READS AND ASSEMBLY QUALITY

The reads of the bovine isolates had Nextera adapters and sequencer contaminant adapters identified and removed using QCtool. The number of raw reads in each isolate ranged from 798,967 to 4,935,109 (mean 2,195,049). After removing the contaminants and trimming poor quality bases, the number of clean reads in each isolate ranged from 763,813 to 4,934,592 (mean 2,101,027).

QUAST metrics of the assemblies of the 188 bovine isolates are detailed in Appendix A.1. The number of contigs in the 188 *S. aureus* assembled genomes ranged from 16 to 190 (mean = 55.2). The size of the largest contig for each isolate ranged from 101,688 to 1,223,753 bp (mean = 331,794 bp), and the total assembled genome lengths ranged from 2,642,588 to 2,858,814 bp (mean = 2,754,658 bp). The GC% ranged from 32.62 to 32.85 (mean = 32.73). The N_{50}, which is the length of all the contigs of that length or longer covers at least half of the assembly length, ranged from 30,859 to 527,369 bp (mean = 140,133 bp), and the L_{50}, which is the minimal number of contigs that cover half the assembly, from 2 to 28 (mean = 8.85). Hence, all 188 isolates were used for the analysis.

3.3.2 CLONAL COMPLEXES AND SEQUENCE TYPES

A total of 10 CCs and

The dominant CC was CC1, accounting for 140/188 (74.5%) of the isolates, followed

Table 3.1: The distribution of clonal complexes identified in 188 bovine *S. aureus* isolates according to the years of isolation. The prevalence percentages are rounded up to the nearest decimal and provided in parantheses

Clonal complex	Number of isolates			
	All years	2002-03	2012-13	2017-18
CC1	140 (74.5)	44 (74.6)	49 (75.4)	47 (73.4)
CC97	27 (14.4)	5 (8.5)	9 (13.8)	13 (20.3)
CC151	7 (3.7)	5 (8.5)	0	2 (3.1)
CC8	6 (3.2)	2 (3.4)	4 (6.2)	0
CC133	2 (1.1)	1 (1.7)	1 (1.5)	0
CC5367	2 (1.1)	0	0	2 (3.1)
CC5	1 (0.5)	1 (1.7)	0	0
CC45	1 (0.5)	1 (1.7)	0	0
CC15	1 (0.5)	0	1 (1.5)	0
CC78	1 (0.5)	0	1 (1.5)	0
Total	188	59	65	64

by CC97 (27/188; 14.4%). These CCs were also the two most common CCs in each one of the three collection periods. Other relatively common CCs that appeared in multiple periods were CC151, CC8 and CC133. Other CCs were observed sporadically. One novel ST identified in this study was submitted to the PubMLST, and was named ST5367 by the database curators. Two isolates from the same farm from the 2017-18 collection belonged to this ST, and there were no SLVs or DLVs associated with it. Hence, it was assigned as CC5367 for this study. The distribution of CCs assigned using PubMLST or based on the predominant ST by collection period are reported in Table 3.1.

The 140 CC1 isolates included 133 (95%) ST1 isolates, two (1.4%) ST6140 and two ST6141 isolates, and one (0.7%) each of ST6161, ST6163 and ST4551. Four new STs were identified within CC1, and were assigned the following ST denominations by the PuMLST curators: ST6140 (two isolates), ST6141(two isolates), ST6161 (one isolate), and ST6163 (one isolate). The 27 CC97 isolates included 16 (59%) ST97, 8 (30%) ST6160, and one (4%) each of ST71, ST6162, and ST6164, out of which ST6160, ST6162 and ST6164 were newly identified STs. Of the six CC8 isolates, three were identified as ST8, while the rest were identified as a new ST designated ST6143.

3.3.3 SPA TYPES

The identified spa-types are detailed in Table 3.2. For 30 isolates the spa-type sequences could not be identified by the spa_typing tool. The paired reads and the assembled contigs of these isolates were subsequently run through another web-based Spa-Typer (Spa-Typer 1.0; (Bartels et al., 2014) (https://cge.cbs.dtu.dk/services/spatyper/; accessed on August 2020), with the sequencing platform set to the option 'Illumina paired-end reads' and 'assembled genome/contigs', respectively. This Spa-Typer tool produced spa-types for these 30 isolates (To assess the validity of the results returned by the Spa-Typer, 10 random isolates that had their spa-type initially identified by the spa_typing tool were processed through the Spa-Typer, with identical results).

The dominant spa-type identified was t114 (106/188 isolates; 56.9%) and was associated with CC1, mainly ST1. Nine (4.8%) isolates were identified as t524, and seven (3.7%) as t922. All seven CC151 isolates had spa type t529. The two isolates forming the newly identified CC5367 had a single, new (unassigned) spa-type.

3.3.4 CORE GENOME ANALYSIS

The samples' pangenome consisted of 6003 genes, of which 1685 were classified as core genes. When just one isolate per farm-CC-year combination was used (133 isolates in total), the total number of genes identified was 5827, of which 1710 were core genes (usually, the number of core genes detected is inversely correlated to the number of isolates analysed). As expected, the percentages of each CCs' pangenome represented by core genes were higher in each CC analysis than in the aggregate analysis of all the isolates. The 140 CC1 isolates displayed 3,987 genes, of which 2,104 were CC1 core genes. When one CC1 isolate per farm was considered, a total of 98 isolates displayed 3866 genes, 2133 of which were core genes. The total number of genes identified for the major CCs (one isolate per farm-year combination), and the total number of genes are reported in Table 3.3.

The NJ tree constructed using the alignment of all the core genes is shown in Figure 3.1. The tree shows a marked segregation of the core genomes based on CCs, with each CCs forming a monophyletic group. Within each CCs, isolates segregated

Table 3.2: Distribution of spa types among 188 bovine *S. aureus*. The new spa types are identified as 'tnew' followed by a serial number and the repeat number. The number of isolates (N) belonging to each spa type are provided along with the prevalence percentages in parantheses

Spa-type	Associated CCs	N(%)
t114	CC1(ST1, ST4551, ST6161, ST6140)	106 (56.4)
t008	CC8 (ST8)	2 (1.1)
t015	CC4(ST508)	1 (0.5)
t062	CC5(ST5)	1 (0.5)
t084	CC15(ST199)	1 (0.5)
t1166	CC133(ST1247)	1 (0.5)
t127	CC1(ST1)	6 (3.2)
t1407	CC1(ST1)	1 (0.5)
t16350	CC97(ST97)	2 (1.1)
t17282	CC97(ST97)	2 (1.1)
t1784	CC1(ST1)	4 (2.1)
t17906	CC1(ST6165)	1 (0.5)
t186	CC78(ST78)	1 (0.5)
t1931	CC1(ST1)	1 (0.5)
t211	CC8 (ST6143)	3 (1.6)
t2246	CC1(ST1)	1 (0.5)
t2421	CC97(ST97)	1 (0.5)
t267	CC97(ST97)	2 (1.1)
t2699	CC1(ST1)	3 (1.6)
t2802	CC97(ST97)	1 (0.5)
t3380	CC97(ST97)	2 (1.1)
t386	CC1(ST1)	1 (0.5)
t4540	CC1(ST1)	1 (0.5)
t4682	CC97(ST6164)	1 (0.5)
t521	CC97(ST97)	1 (0.5)
t524	CC97(ST6160, ST71)	9 (4.8)
t529	CC151(ST151, ST705)	7 (3.7)
t693	CC1(ST1)	4 (2.1)
t6980	CC1(ST1, ST6140)	2 (1.1)
t865	CC97(ST6162)	1 (0.5)
t922	CC1(ST1, ST6163)	7 (3.7)
tnew1 (07-16-02-13)	CC1(ST6141)	2 (1.1)
tnew2 (07-23-12-05-17-02-17-34-34-34-34-33-34)	CC97(ST97)	1 (0.5)
tnew3 (07-23-12-05-17-34-34-13-33-34)	CC97 (ST97)	1 (0.5)
tnew4 (07-23-12-05-22-34-13-34-13-33-34)	CC97(ST97)	1 (0.5)
tnew5 (11-19-12-21-17-34-24-34-24-34-22-33-25)	CC8 (ST8)	1 (0.5)
tnew6 (14-12-05-17-34-34-13-33-34)	CC97 (ST97)	1 (0.5)
tnew7 (14-12-05-17-34-34-34-13-33-34)	CC97 (ST97)	1 (0.5)
tnew8 (14-23-17-362-13-22-22)	CC5367 (ST5367)	2 (1.1)
tnew9 (16-34-33-13)	CC1 (ST1)	1 (0.5)

according to the STs. Isolates from the same farm always segregated together in a monophyletic group if they belonged to the same CC. However, there was no marked segregation within each CC according to the year of collection.

Table 3.3: Distribution of genes as core genes (present in 100% of the isolates), soft-core genes (present in 95% to 99% of isolates), shell genes (present in 15% to 94% of isolates) and cloud genes (present in 0% to 14% isolates) across various clonal complexes among the 188 bovine-associated *S. aureus* isolates. The prevalence percentages are provided in parantheses

Clonal complex	Number of isolates	Total genes	Core genes (% of the total)	Soft-core genes	Shell genes	Cloud genes
All	188	6003	1685 (28.1)	341	781	3196
All (one isolate per farm-CC selected)	133	5827	1710 (29.3)	310	784	3023
CC1 only	140	3987	2104 (52.8)	188	396	1299
CC1 only (one isolate per farm- selected)	98	3866	2133 (55.2)	153	421	1159
CC97 only	27	3427	2141 (62.5)	86	655	545
CC151 only	7	2673	2420 (90.5)	0	198	55
CC8 only	6	2890	2388 (82.6)	0	502	0

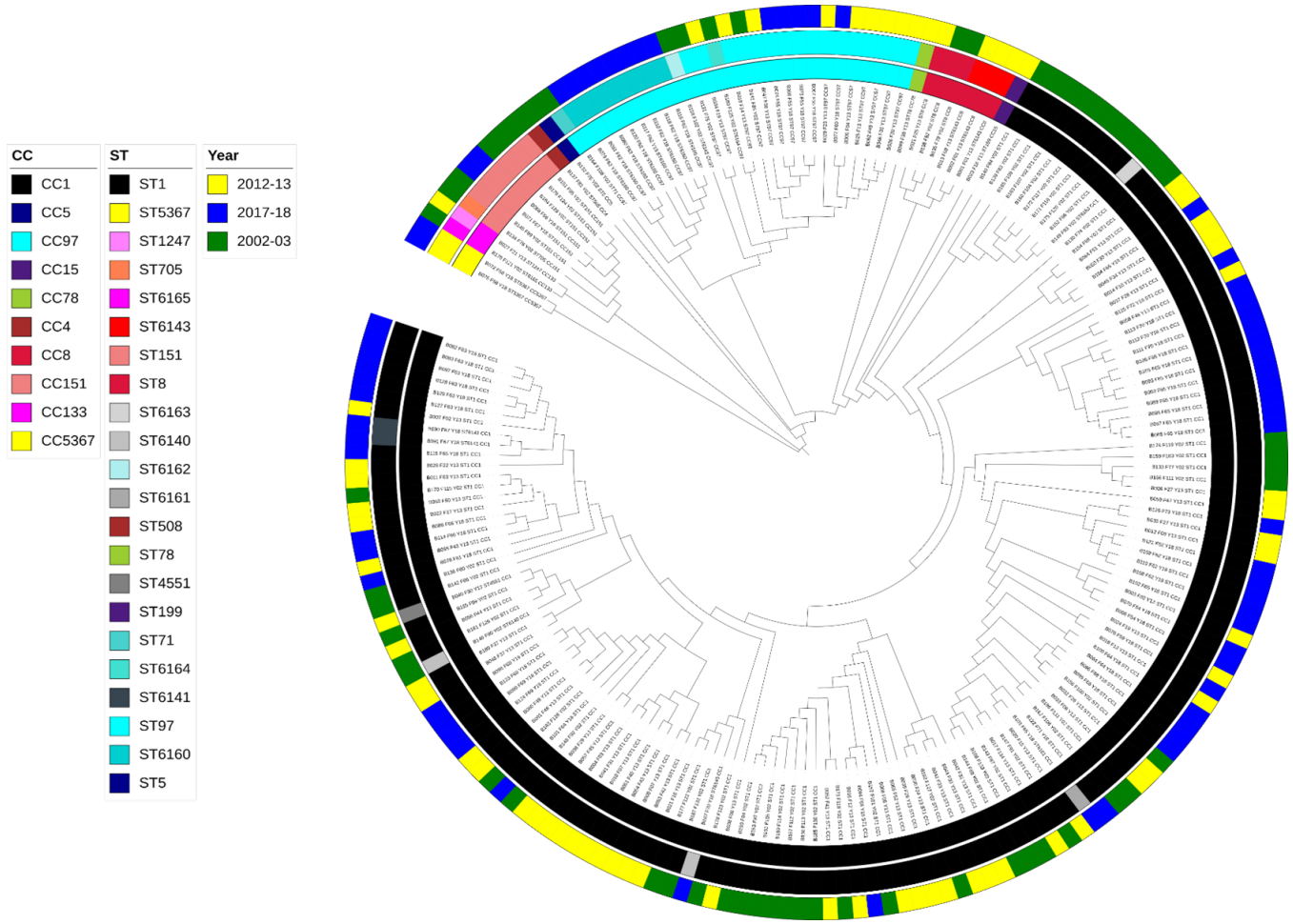

Figure 3.1: Neighbour-joining tree obtained using the core gene alignment distance matrix of 188 bovine *S. aureus* isolates, calculated using the Jukes-Cantor method. The inner colour band represents the clonal complex (CCs), the intermediate band identifies the sequence types (STs) and the external band the year of collection of the isolates.

3.3.5 ACCESSORY GENOME ANALYSIS

The spreadsheet of the gene presence/absence binary matrix for this study can be found in https://github.com/jabinnes/Thesis-files/blob/main/gene_presence_absence_roarybovine.Rtab. As with the core genome analysis, the PCoA of the gene presence/absence binary matrix showed marked segregation of the 188 isolates based on their CCs (Figure 3.2). Within CC1, the isolates did not seem to further segregate according to their STs but with CC97 there was segregation based on STs (https://github.com/jabinnes/Thesis-files/tree/main/PCoA%20plots).

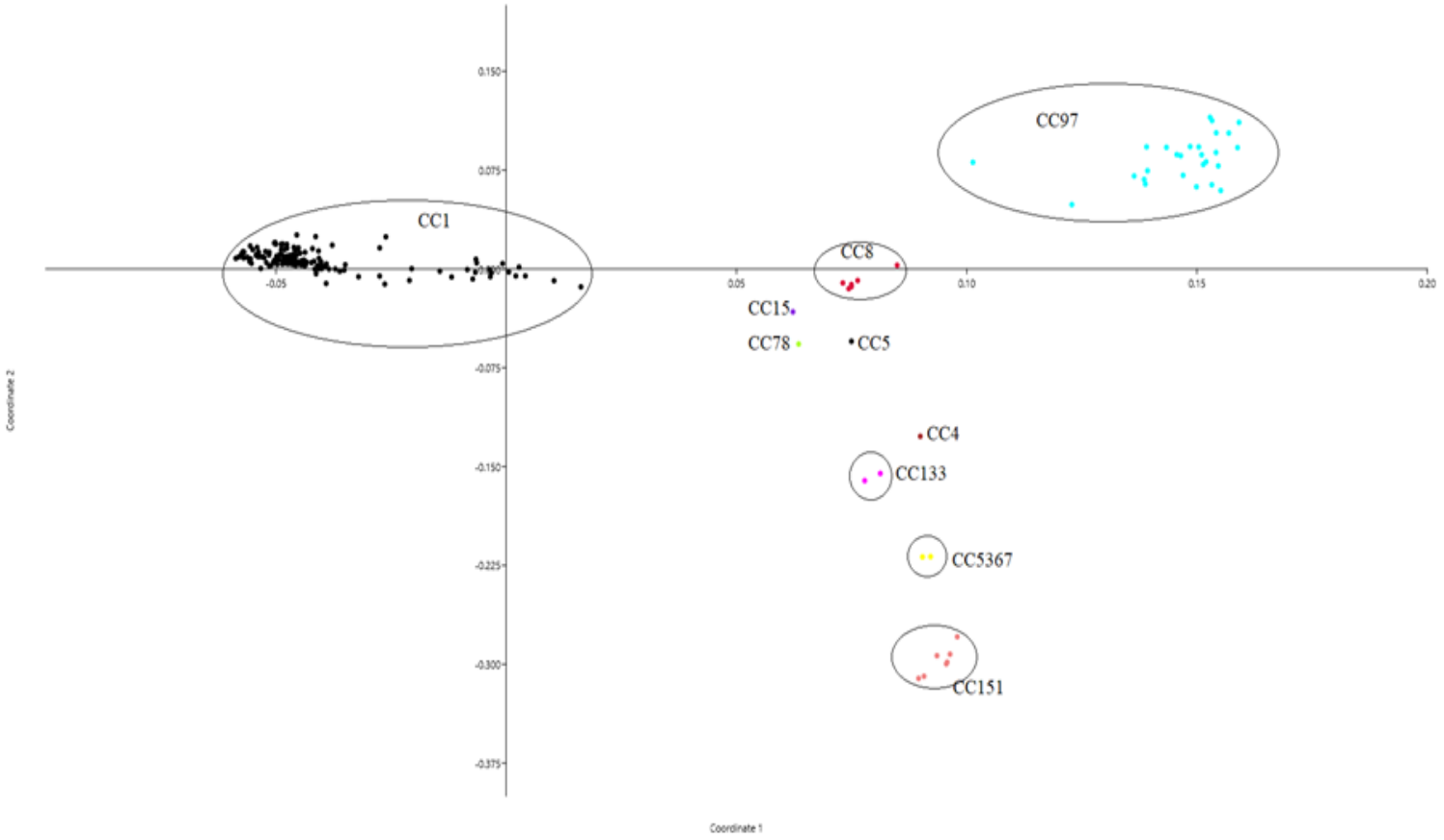

Figure 3.2: Principal Coordinate Analysis (PCoA) of the Roary gene presence-absence matrix for 188 bovine *S. aureus* isolates. CC1-Black dots, CC97-Aqua dots, CC151-Light coral dots, CC5367-Yellow dots, CC133- Fuchsia dots, CC8- Crimson dots, CC15Blue-violet dot, CC78- Yellow-green dot, CC5-Dark blue dot, CC4- Brown dot.

Table 3.4: Distribution of agr groups in 188 bovine *S.aureus* isolates across clonal complexes and year of collection. For each year of collection, the number of isolates belonging to the agr groups along with the total number of isolates from the year along with the prevalence percentages are provided in the parantheses.

Agr group	Number of isolates	Clonal complexes	Year of collection
agr-I	36 (19.1%)	CC97, CC8, CC133	2002-03 (9/59; 15.3%), 2012-13 (14/65; 21.5%), 2017-18 (13/64; 20.3%)
agr-II	11 (5.9%)	CC151, CC5367, CC5, CC15	2002-03 (6/59; 10.2%), 2012-13(1/65; 1.5%), 2017-18 (4/64; 6.3%)
agr-III	141 (75%)	CC1, CC78	2002-03 (44/59; 74.6%), 2012-13 (50/65; 76.9%), 2017-18 (47/64; 73.4%)

3.3.6 AGR OPERON, VIRULENCE FACTORS AND ANTIMICROBIAL RESISTANCE GENES

3.3.6.1 *AGR* OPERON

The *agr* groups identified by the AgrVATE software are presented in Table 3.4. As expected, all the isolates belonging to the same CC were of the same *agr* group (Robinson et al., 2005).

The most common group was *agr*-III, found in 141 (75%) isolates belonging to the 140 (100%) CC1 isolates and one (100%) CC78 isolate. Due to the high prevalence of CC1, ~75% of isolates from each collection year belonged to *agr*-III.

3.3.6.2 VIRULENCE FACTORS

A total of 191 virulence factors were identified by ARIBA. Thirty-two (16.8%) factors were found in all the isolates, and 92 (48.2%) in at least 90% of the isolates. Seventeen virulence factors were found in single isolates The presence and absence binary spreadsheet of virulence genes identified by ARIBA using VFDB for all the 188 isolates is provided in an online repository (https://github.com/jabinnes/Thesis-files/blob/main/Bovine_VFDB_presence_absence.csv).

Seven staphylococcal enterotoxin (se) genes were identified: *sea*, *sec*, *sed*, *seg*, *seh*, *sei*, and *sej*. Interestingly, the *seh* gene was found in all 140 CC1 isolates. The *sea* gene was found in only one isolate (B139; CC1), *sed* and *sej* were found in 4 (15%) CC97 isolates from the same farm. The *seg* and *sei* enterotoxin genes were found together in all seven CC151 isolates and the CC45 isolate, while the *seg* gene was found as the only enterotoxin gene in the single CC5 isolate.

An association between the CC and the staphylococcal 'enterotoxin-like' (sel) genes *selk, sell, selm, seln, selo, selq, selr, selu* was also observed. In fact, all CC1 isolates lacked any enterotoxin-like gene, except isolate B139, which contained the *selk* gene. Isolate B139 was also the only isolate that possessed the *selq* gene. The *selr* gene was found in the four CC97 isolates that also harboured the *sed* and *sej* genes. The *sell* gene was associated with CC78, and the *selm* gene was found in the CC5 and CC45 isolates. The *selo* and *selu* genes were found in all the CC151 isolates along with the CC5 and CC45 isolates, while the *seln* gene was found in 6 (86%) of the 7 CC151 isolates and the sole CC45 isolate.

The genes of the staphylococcal exotoxin-like protein family (*set*) were distributed across all the isolates. Two genes of this family, *set30* and *set40,* were found only in CC8 isolates (all six CC8 isolates had both genes).

The human Panton-Valentine bicomponent leukocidin (PVL) gene *lukF-PV* and *lukS-PV* was found in only one CC1 isolate (B139). Conversely, the bovine-associated bicomponent leukocidin gene *lukF'-PV* and *lukM* (Vrieling et al., 2016) was found in 128 (68.1%) isolates. Both components' genes were found in 127 (67.6%) and one isolate harboured the *lukM* gene only. The *lukF'-PV* was found in the majority (118/140; 84.3%) of CC1 isolates, in all the CC151, 2/27 (7%) of CC97, and in the sole CC15 isolate.

Four haemolysin toxins (α, β, γ, and δ) were identified in ~89% of the isolates. There seemed to be no association between the presence of haemolysin genes and the CCs.

The intercellular adhesion genes *icaADBC* play a role in biofilm production (Cue et al., 2012; Dhanawade et al., 2010), and presumably also in mastitis pathogenesis. The *icaA* gene was found in 57 (30.3%), *icaB* in 171 (91%), *icaC* in 177 (94.1%), and *icaD* in 55 (29.3%) isolates. Interestingly, the *icaR* gene, involved in the negative regulation of the *ica* operon, was found in 179 (95.2%) isolates but was not observed

in 9 (33%) of the CC97 isolates. These nine CC97 isolates did not possess any *ica* genes, but eight of them belonged to cows from the same farm.

3.3.6.3 ANTIMICROBIAL RESISTANCE GENES

The presence and absence binary spreadsheet of the antimicrobial resistance genes for all the 188 isolates are provided in an online repository (https://github.com/jabinnes/Thesis-files/blob/main/Bovine_AMR_Resfinder_presence_absence.csv). Four acquired antimicrobial resistance genes were identified. The *blaZ* gene (coding for β-lactam resistance) was the most common gene, observed in 48 (25.5%) isolates. Other resistance genes were the *fusC* (fusidic acid resistance), *str* (aminoglycoside resistance), and *ermC* (erythromycin resistance), and were found in 4 (2.1%), 3 (1.6%), and 2 (1.1%) isolates, respectively. The *blaZ* gene was broadly distributed, and was found in 16/140 (11.4%) CC1 isolates, 24/27 (89%) CC97 isolates and 5/6 (83%) CC8 isolates. The *blaZ* gene was evenly distributed among the different periods of collection, with ~25% of the isolates harbouring this gene in each period (2002-03: 23.7%, 2012-13: 27.7%, 2017-18: 25%). The three other antimicrobial resistance genes were found sporadically in isolates from 2002-03 and 2012-13. The *str* and *ermC* genes were found in CC1 isolates, while the *fusC* gene was found in 3/4 (75%) CC8 isolates and 1/4 (25%) CC1 isolates.

3.4 DISCUSSION

This Chapter reports a genomic epidemiology study of 188 bovine mastitis-causing *S. aureus* isolates obtained from different sources over a 15 year period. It represents the first comprehensive population genetic study of its kind in New Zealand, and probably one of the largest genomic epidemiology studies of bovine *S. aureus*, worldwide.

A total of 10 CCs comprising 22 STs were identified. Interestingly, the dominant clonal lineages remained unchanged over the 15 years of observation, with ~75% of the isolates belonging to CC1, in particular, ST1 which constituted ~70% of all the isolates. The reason for this clonal stability could be due to the lack of live dairy

cattle importation into New Zealand since the 1990s except from Australia (which has been interrupted since 2017 (MPI, 2017).

A limitation of this study was the unavailability of farm addresses, which precluded the ruling out of the presence of isolates from the same farm in different years. Nevertheless, a cross-validation for the dominance of CC1 was provided by the subset of isolates from the South Island, which were only available for the 2012-2013 collection period, as 6/7 isolates from different farms, with the exception of two isolates arising from one farm, belonged to CC1.

The finding of the predominance of CC1/ST1 in dairy cattle is unique. The CC1/ST1 is considered a human-adapted clone, uncommonly seen in animals (Fitzgerald, 2012). In fact, many molecular epidemiological studies worldwide have indicated that CC97, CC151, CC8, CC133 are the most common CCs in dairy cattle (Boss et al., 2016; Fursova et al., 2020; Hoekstra et al., 2020; Li et al., 2017; O'Dea et al., 2020; Ren et al., 2020; Thomas et al., 2021). CC1 was only the 3rd most common CC associated with bovine mastitis in Australia in 2013 (O'Dea et al., 2020). Interestingly, during the study period, CC1/ST1 was also the dominant Methicillin-susceptible *S. aureus* (MSSA) clonal lineage in humans in New Zealand (Grinberg et al., 2017; Heffernan et al., 2015). Thus, a detailed comparative genomic analysis of New Zealand bovine and human CC1 isolates will be presented in Chapter 4.

Other common CCs found in this study were CC97, CC8, CC151, and CC133 which, as stated above, are very common CCs in cattle worldwide. A novel ST (ST5367) identified in two isolates from 2017-18 was not an SLV or DLV of any of the STs found in the PubMLST database, suggesting the possibility that this novel CC emerged in New Zealand.

The most common spa-type was t114 (56%), and was associated with CC1. This spa type was reported in very few CC1 isolates from dairy cattle overseas (Mora-Hernández et al., 2021; O'Dea et al., 2020; Wang et al., 2018a). The t127 spa-type was found in six CC1 isolates. As of June 2021, t127 constituted 2.6% of the spa type frequency reported in the Ridom Spa server while t114 represented only 0.03%. It should be noted that ST1/t127 is a common lineage in humans in New Zealand (Heffernan et al., 2015). Other common spa types identified were t524 and t529, associated with CC97 and CC151, respectively. The t524 is commonly

associated with CC97 in other countries (Boss et al., 2016; Hasman et al., 2009; Li et al., 2015b), and the t529 is commonly associated with CC151 (Boss et al., 2016; Hasman et al., 2009). The common occurrence of the same ST/spa-type combinations in ruminants worldwide, and their infrequent occurrence in other host types indicates a high degree of host-adaptation of these clonal lineages.

Consistent with a previous study (Grinberg et al., 2017), *S. aureus* clustered according to their CCs, both at the core and accessory genome levels. The year of collection did not seem to impact the clustering. This indicates a stable population genetic structure over the 15 years of observation. As expected from a contagious organism that is transmitted between cows, isolates belonging to the same CC clustered based on their farm of origin. Yet, more than one CC were found in three farms, consistent with previous studies (Kadlec et al., 2019).

The study also included an analysis of the virulome and the resistome. Genes from the enterotoxin gene cluster (*egc*) comprising of enterotoxins (*se*) and enterotoxin-like (*sel*) genes are reported in strains from dairy products responsible for food poisoning in humans. This is due to the high prevalence of various *se* and *sel* in bovine mastitis causing strains present in milk (Jarraud et al., 2001; Kuzma et al., 2003; Omoe et al., 2013). Differences in the prevalence of the various *se* and *sel* genes have been reported in bovine *S. aureus* in various countries (Karahan et al., 2009). Staphylococcal enterotoxin C encoded by the *sec* gene has been indicated as an important virulence factor for bovine mastitis (Fang et al., 2019), and has been found more commonly in bovine mastitis strains along with the *sea* gene (Rall et al., 2014). Importantly, the *sec* gene has been associated with human mastitis postpartum and has the potential to be transferred from animal strains to human strains and vice versa (Franck et al., 2017). Interestingly, the *sec* gene was found only in 2 isolates in this study. The *sea* gene was found only in one isolate (B139) belonging to CC1. Other studies that analysed the occurrence of the *sec* and *sea* genes did not identify the clonal lineage of the isolates (Jarraud et al., 2001; Karahan et al., 2009; Rall et al., 2014). However, an association between the *egc* variants and the CCs has been established (van Belkum et al., 2006). The association between *egc* and CCs could explain the very low prevalence of *sea* and *sec* genes. The most common enterotoxin in the study was *seh* which was present in all the CC1 isolates.

The *sel* genes were found only in 8% of the isolates in the study, and three variants

(*seln*, *selo*, *selu*) were identified only in isolates of CC151, CC45 and CC5. The isolates harbouring *seln*, *selo* and *selu* also harboured *seg* and *sei*, and this was expected as these genes are part of the same type II pathogenicity island vSaß, that carries *seg-sei-selm-seln-selo-selu* genes (Wang et al., 2009). Among CC1, only one isolate (B139) had the *selk* and *selq* genes. The *selk* and *selq* genes have been found to have a higher prevalence in human than in bovine *S. aureus* (Omoe et al., 2005; Wang et al., 2009).

Interestingly, the bovine-adaptive leukocidin genes *lukM* and *lukF'-PV* were found in 128/188 (68%) isolates, and the leukocidin genes *lukD* and *lukE* were found in 176/188 (93.6%) isolates. The CC5367 isolates did not carry any leukocidin genes. The high prevalence of *lukM* and *lukF'-PV* in this study was comparable to that reported in other countries, which ranged from 50% to 86% (Monecke et al., 2007; Schlotter et al., 2012; Yamada et al., 2005). The *lukM* and *lukF'* together forms an effective cytotoxic factor against bovine neutrophils (Vrieling et al., 2015a).

One of the important leukocidins in human *S. aureus*, the Panton-Valentine leukocidin (PVL) with its bicomponent, *lukF-PV* and *lukS-PV* was found in one isolate (B139) belonging to CC1. This was the same isolate that harboured the human-associated *selk* and *selq* genes. The *lukF-PV* and *luKS-PV* together form an effective cytotoxic factor against human neutrophils (Löffler et al., 2010). These features of isolate B139 were intriguing and will be discussed in detail in the comparative genomic study presented in Chapter 4.

Biofilm formation is considered an important immune evasive mechanism of *S. aureus*. The biofilm-associated protein (*bap*) gene has also been associated with biofilm formation in some cases of bovine mastitis (Cucarella et al., 2004). The process is regulated by the intercellular adhesion genes *icaA*, *icaB*, *icaC*, and *icaD,* and the negative regulator *icaR*. More than 90% of the isolates in this study were positive for *icaB*, *icaC* and *icaR,* and 30% had *icaA* and *icaD* genes. No isolate carrying the *bap* gene was identified in this study. This result was similar to the lack of *bap* gene observed in previous studies of bovine *S. aureus* (Chen et al., 2020; Notcovich et al., 2018).

The *blaZ* gene was found in ~25% of the isolates. This was consistent with the prevalence of the *blaZ* gene found in European bovine *S. aureus* (15%), but lower than the prevalence of the *blaZ* found in Asia (86%) (Jamali et al., 2014) and South

America (82%) (Pérez et al., 2020). The prevalence of the *blaZ* gene in bovine *S. aureus* in New Zealand was stable over the 15 years (2002-03: 14/59; 24%, 2012-13: 18/65; 28%, 2017-18: 16/64; 25%). The prevalence of *blaZ* approximates the estimated prevalence of phenotypic penicillin resistance of 28% reported in bovine *S. aureus* in NZ in the same period (McDougall et al., 2014). None of the isolates carried the *mecA* or *mecC* genes, which encode methicillin resistance in *S. aureus*. To date, there has been only one report of methicillin resistant *S. aureus* (MRSA) in cattle in New Zealand, but the result was clinically uncertain due to the mixed nature of the cultured isolate (Grinberg et al., 2005). The prevalence of the MRSA genes reported in bovine *S. aureus* across the world ranged from 0% to 45% (Aklilu and Hui Ying, 2020; Fursova et al., 2020; Hoekstra et al., 2020; Liu et al., 2017; O'Dea et al., 2020; Obaidat et al., 2018; Schnitt and Tenhagen, 2020). The study presented in Chapter 5 will elucidate the correlation between the presence of resistance genes and expression of phenotypic resistance in bovine *S.aureus*.

More information regarding the clinical status of the cows would help illuminate the association between CCs and clinical outcomes. More research is needed to identify whether the distribution of CCs in South Island is similar to North Island as only seven isolates from South Island from the 2012-13 collection period were included in this study.

In conclusion, the predominance of CC1, an uncommon lineage in ruminants, worldwide was intriguing. The other common clonal lineages observed, in particular CC97 and CC151 have been associated with ruminants overseas. The virulome and resistome content of the isolates analysed here (including in CC1), such as the high prevalence of the bovine adaptive LukM/LukF'-PV, and the 25% prevalence of the *blaZ* gene are consistent with host-adaptation to cattle. Host adaptation will be further explored in the comparative genomics study presented in Chapter 4. The knowledge acquired in this study on the population genetic structure of bovine *S. aureus* will form the basis for the the reverse vaccinology method developed in the study presented in Chapter 6.

4

COMPARATIVE GENOMICS OF *STAPHYLOCOCCUS AUREUS* ACROSS HUMANS, COMPANION ANIMALS AND FARMED RUMINANTS

4.1 INTRODUCTION

Staphylococcus aureus is a worldwide distributed commensal bacterium and opportunistic pathogen of humans and animals (Wertheim et al., 2005; Haag et al., 2019). The bacterium is associated with a range of skin and internal organ infections and bacteriaemia in humans and many animal species. *S. aureus* is also among the most economically important agents of mastitis in farmed ruminants.

S. aureus has a clonal population structure and clonal complexes (CCs) can be defined using multi-locus sequence typing (MLST) (see section 1.1.4.2.2) (Feil et al., 2003; Grinberg et al., 2017).Most studies characterising ruminant-associated *S. aureus* have indicated the existence of strains that are distinct from human-associated strains. While the vast majority of *S. aureus* infecting or colonising humans is represented by 11 CCs: CC1, CC5, CC8, CC12, CC15, CC22, CC25, CC30, CC45, CC51 and CC121, the most commonly observed CCs in ruminants have been CC97, CC133, CC151 and CC522 (Cuny et al., 2015). However, there have been host jumps of CCs from humans to animals and vice versa (Akkou et al., 2018).Studies analysing *S.aureus* obtained from bovine mastitis cases have shown that bovine-associated strains often belong to these specialised 'ruminant' CCs (Sung et al., 2008; Guinane et al., 2010; Delgado et al., 2011).

There has been an increased interest in the study of livestock-associated *S. aureus* since the discovery, more than a decade ago, of CC398 methicillin resistant *S. aureus* (MRSA) in pigs and ruminants, with its subsequent spread to humans (Armand-Lefevre et al., 2005; Graveland et al., 2010). Despite the emergence of CC398 MRSA and its zoonotic spread, the mechanisms of *S.aureus* host adaptation are not well understood.

Three genomic criteria have been proposed as the basis of the definition of *S. aureus* host adaptation : 1) phylogenetic clustering of *S. aureus* isolates based on the host of origin, 2) the presence of host-specific/host-adaptive genes, and 3) the loss of irrelevant, non-host adaptive genes (Matuszewska et al., 2020).

Host-specific/adaptive genes are genes that code for structures or functions conferring a competitive advantage to the strain in a specific host. These genes are usually gained or lost prior to, or after host jumping (Guinane et al., 2010). The term

host-specific is often used to define genes that are associated with a specific host type, but their role in host-adaptation is not necessarily known. Conversely, the term 'host adaptive' is reserved for those genes that code for a specific structure or function unique to a host type. Host-specific/adaptive genes are usually accessory genes found in mobile genetic elements (MGEs) such as pathogenicity islands, bacteriophages, transposons, insertion sequences and plasmids. One of the identified MGEs in *S. aureus* is the ϕSa3 prophage, which contains the human-adaptive immune evasion cluster genes such as the Staphylococcal complement inhibitor (*scn*), chemotaxis inhibitory protein (*chp*), Staphylokinase (*sak*), Staphylococcal enterotoxin A (*sea*) (van Wamel et al., 2006; Richardson et al., 2018). Similarly, the pathogenicity island SaPIbov is a bovine pathogenicity island containing the bovine-specific staphylococcal enterotoxin C (*sec*) and toxic shock syndrome toxin (*tsst*) genes (Fitzgerald et al., 2001). Other pathogenicity islands, such as SaPIov1 containing *tsst-1*, *sec*, staphylococcal enterotoxin L (*sel*), *tsst-O* and SaPIov2 containing von Willebrand-binding protein (*vWbpSov2*), *scn* are found in ruminant isolates, predominantly in small-ruminants (Lee et al., 1992; Murray et al., 1994; Guinane et al., 2010). Other host-adaptive genes are the Panton-Valentine leukocidin (PVL) factors *lukS-PV* and *lukF-PV,* found in human isolates and their equivalent the *lukM-PV* and *lukF'-PV,* found in bovine isolates. These are leukocidins with mode of action specific to human or bovine neutrophils, respectively.

A previous comparative genomic study of *S. aureus* in humans and pets undertaken in New Zealand (Grinberg et al., 2017) found that *S. aureus* clustered predominantly by their CCs, and within each CC, human and pet isolates formed a homogeneous, unsegregated bacterial population . However, to the author's knowledge, there are no comparative genomic studies of human and cattle *S. aureus* from New Zealand.

The results of the study presented in Chapter 3, indicates that, unlike other countries where ruminants and humans are infected with different *S. aureus* CCs, in New Zealand, CC1 is the predominant CC, both in cattle and in humans. This unique epidemiological situation enabled a deeper genomic comparison, to include both the core, and accessory genome of *S. aureus* across ruminants and humans.

4.2 MATERIALS AND METHODS

4.2.1 BOVINE ISOLATES

All the 188 bovine isolates described in Chapters 2 and 3 were used in this study. Briefly, the isolates were isolated from bovine milk samples submitted to two veterinary laboratory networks (NZVP and Gribbles) for culture during 2002-03, 2012-13, and 2018-19 (n=137), and from samples collected from cases of mastitis during an antimicrobial efficacy study undertaken in 2018-19 in central North Island (n=51). All these isolates were subjected to multilocus sequence typing (MLST) and spa-typing as described below, but to avoid autocorrelation due to the presence of duplicates, one isolate per farm/CC combination was arbitrarily selected for the core and accessory genomic comparisons.

4.2.2 SMALL RUMINANT ISOLATES

Thirty two small ruminant *S.aureus* isolates were used. Thirty isolates originated from an equal number of sheep sampled on 11 sheep farms during a study of prevalence of udder defects in non-dairy-breed ewes performed in the lower North Island in 2018-2019. The sampling methods have been previously described (Ridler et al., 2021). The isolates were accompanied by unique sheep and farm identifiers. Two isolates (from a goat and a sheep) were isolated in 2012 and 2018, respectively, and were provided by two New Zealand clinical veterinary diagnostic laboratories without clinical details. All these isolates were subjected to MLST and spa-typing as described below, but to avoid autocorrelation due to the presence of duplicates, one isolate per farm/CC combination was arbitrarily selected for the core and accessory genomic comparisons.

4.2.3 HUMAN, CANINE AND FELINE ISOLATES

Pre-existing genomic sequences of New Zealand human, canine and feline *S. aureus* isolates were used for comparative analyses with the ruminant isolates. These genomes have been previously analysed as part of a different project (Grinberg et al.,

2017), and were publicly available as sequence reads in the Sequence Read Archive (SRA) of the US National Center for Biotechnology Information (NCBI) (Leinonen et al., 2011), under bio project PRJNA391123. Among the 59 genomes available in the SRA, 34 were from humans, 19 from dogs and 6 from cats. These isolates were collected from both infection sites and subclinical colonisation of nares, as reported in General Material and Methods Chapter. The isolates from humans were isolated in 2014, while those from dogs and cats were isolated between June 2012 and June 2013, and February 2016 and July 2016. The genomes were retrieved using the fastq-dump utility of the SRA-tool kit (Leinonen et al., 2011) with the ' --split-files' option to convert the SRA file to fastq file split into two paired-end reads.

4.2.4 RE-IDENTIFICATION OF *S. AUREUS*

The bovine and small ruminant isolates stored at -80°C in glycerol solution were resuscitated by streaking an aliquot of frozen bacteria on a 5% Columbia sheep blood agar plate (Fort Richard Laboratories, Auckland, New Zealand) and incubating overnight at 35-37°C in aerobic conditions. Isolates were re-identified phenotypically as *S.aureus* based on the presence of a pure culture of round, golden-white colonies consisting of Gram-positive, catalase-positive cocci, positive for the clumping factor and/or protein A by latex agglutination (Staphaurex, Thermo Fisher Scientific, Auckland, NZ). Eleven confirmed isolates were selected at random and tested using Matrix-Assisted Laser Desorption Ionization-Time Of Flight Mass Spectrometry (MALDI-TOF MS) at Massey University. All these isolates returned strong hits to *S. aureus* (log score range: 2.2 - 2.5; results not shown).

4.2.5 DNA EXTRACTION, LIBRARY PREP AND WHOLE-GENOME SEQUENCING

The DNA extraction method used in this study is detailed in section 2.2. In brief, a suspension of *S. aureus* in phosphate buffer solution was used for DNA extraction. The DNA was extracted using an extraction kit (DNeasy UltraClean Microbial Kit, Qiagen, USA) following the manufacturer's protocol, except that a mini bead-beater (Biospec, Oklahoma, USA) was used for 2 minutes instead of the Vortex adapter due

to the unavailability of a Vortex adapter. The extracted DNA was analysed for quality and quantity using two spectrophotometers: NanoDrop Microvolume Spectrophotometer (Thermo Fisher Scientific, Auckland, NZ) and Qubit 2.0 Fluorometer (Thermo Fisher Scientific, Auckland, NZ). Extracted DNA of satisfactory quantity (> 100ng/ml measured using Qubit 2.0 Fluorometer) and quality (A_{260}/A_{280} ratio between 1.8 and 2.0 measured using the NanoDrop Microvolume Spectrophotometer) was stored at -80°C until sequencing.

Whole-genome sequencing (WGS) of the bovine and the two small ruminant isolates obtained from diagnostic laboratories was undertaken as detailed in section 2.1.3. The extracted DNA was diluted to a concentration of 0.2 ng/μL and the libraries were prepared using Nextera XT library preparation kit (Illumina #FC-131-1096; Illumina, San Diego, USA) according to the following five main steps: 1) tagmentation, 2) amplification of libraries, 3) clean-up of libraries, 4) normalisation of libraries, and 5) pooling of libraries. The DNA fragments were tagged with index using TG Nextera XT Index Kit v2 Set A (Illumina#FC-131-2001) or TG Nextera XT Index Kit v2 Set B (Illumina #FC-131-2004). The library preparation was carried out by Massey Genome Service, Palmerston North, New Zealand.

A 200 ng aliquot of the pooled library was sent to a commercial sequencing provider (Novogene AIT, Hong Kong) at ambient temperature in a transport kit tube (DNAstable, Sigma-Aldrich, Auckland, NZ) using a priority courier service through Massey Genome Service. After quality check, the libraries were loaded across one lane of a HiSeq X (150bp PE) sequencer and demultiplexed at the end of sequencing. Low-quality reads and adapters were removed from the paired raw reads. The Phred scores of the reads and GC content were measured. Thousands of random reads were blasted against the nucleotide NCBI database (Altschul et al., 1990) to check if they mapped against *S. aureus*. The above-mentioned processing of the libraries were undertaken by the sequencing provider and the resultant sequence-reads were sent back in a 1TB hard disk (Seagate Technology, California, USA).

The other thirty sheep *S. aureus* DNA samples were transported in transport kits (DNAstable LD, Sigma-Aldrich, Auckland, NZ) at room temperature by Dr. Alex Grinberg to the Centro di Ricerca Pediatrica Romeo ed Enrica Invernizzi, University of Milan, Italy, where they were processed and sequenced. Briefly, library preparation was done using the Nextera XT library preparation kit (Illumina, San Diego, USA)

and WGS was performed using MiSeq (Illumina, San Diego, USA) technology. The resultant sequence reads were transferred to a web-based drive (Google Drive, Google, California, USA) for downstream analysis.

4.2.6 READ QUALITY CHECK, GENOME ASSEMBLY AND ASSEMBLY QUALITY ANALYSIS

After the sequencing, to remove any remaining low quality reads or adapters left behind after the quality check by the sequencing providers, that may affect the reliability of the assembly and downstream analyses (Del Fabbro et al., 2013), the reads were processed using the QCtool (https://github.com/mtruglio/QCtool), which assessed the reads using the tools FastQC, SolexaQA++, Bowtie. The QCtool was run with default parameters with FastQC, FastQscreen, SolexaQA++, Phix vector and adapter, trimming tools turned on.

The output of the QCtool was considered as the processed reads. The processed reads were assembled by an assembly optimiser, Shovill (Seemann, 2018b), using the SPAdes assembler (Bankevich et al., 2012), and the assembly quality was then assessed using QUAST (Gurevich et al., 2013) for various parameters as detailed in section 2.4.2. The percentage of guanine and cytosine (GC%) bases in the assembled genome and total length of the assembled genome were used to eliminate bad quality assemblies. The thresholds considered after a summary analysis of other studies were: 32.6% to 32.9% GC content, and 2.6 to 2.9 Mb of total length. Genomes of values above or below the thresholds were not included in the bioinformatics analyses.

4.2.7 MLST AND SPA-TYPING

MLST and Spa-typing, two highly discriminatory and widely used typing methods for *S. aureus* (Koreen et al., 2004; Saunders and Holmes, 2007), were used to infer phylogenies. Firstly, an established MLST scheme that classifies isolates into Sequence Types (ST) based on the sequence variation within seven housekeeping genes, was used to establish the clonal descent of the isolates (Enright et al., 2000).

The contigs produced by the assembly were used to identify the STs of the isolates

using the mlst tool (Seemann) (section 2.4.3). The STs were assigned to CCs based on the CCs defined in the PubMLST database (Jolley et al., 2018), which uses the BURST algorithm (Feil et al., 2004). There were only 9 defined CCs as of February 2021. Hence, groups of single-locus variant (SLV) or double locus variants (DLV) that could not be assigned to one of these nine CCs, were assigned to a CC which took the number of the ST with the most SLVs found in the group. The human, canine and feline *S. aureus* isolates were assigned the same CCs as in the study by Grinberg et al. (2017). All new loci of any novel STs not reported in the PubMLST were mapped against all the reads of the isolates using samtools (Li et al., 2009), to verify that the new variant was not due to assembly errors. Novel STs identified from bovine and small-ruminant isolates were submitted to the PubMLST database, where a unique ST identifier was assigned by the database curators.

A second typing method used to infer phylogenies was Spa-typing. Spa-typing is an established method that analyses the sequence variation in the hypervariable repeat region of the Protein A gene (Frénay et al., 1996). Spa types are reported in a publicly available online database (Ridom spa server: `https://spaserver.ridom.de/`). The spa types were identified using spa-typing software (`https://github.com/mjsull/spa_typing`). New spa-types were not submitted to the Ridom database due to the unavailability of the RidomStaphType software for this study. Thus, any new spa-type will be defined in this chapter as 'tnew'.

4.2.8 CORE GENOME COMPARISONS

A species' core genome mostly comprises genes encoding for basic cellular functions essential for survival, but also genes that encode for pathogenicity and virulence (Medini et al., 2005). In this study, the core genome was defined as the group of all the genes observed in all the isolates, and the accessory genome comprised the variably present genes. Thus, comparative analyses of the core and accessory genomes was performed to reveal any phylogenetic clustering.

The accessory genome comprises genes that encode for many functions, including virulence factors and antimicrobial resistance. Accessory genes are present in a variable proportion of the species' isolates, and are predominantly mobile genetic elements carried in transposons, pathogenicity islands and bacteriophages. In this

study, comparative analyses of the accessory genome was undertaken to ascertain the presence or absence of host-specific/adaptive genes in the groups of isolates obtained from the different host types.

The delineation of coding sequences (CDS) and the annotations of the assembled contigs was performed using Prokka (Seemann, 2014) (see section 2.1.4.5).

The analysis of both core and accessory genomes was undertaken using Roary (Page et al., 2015), which utilises the annotated assembly files produced by Prokka. Roary was used to align the concatenated nucleotide sequences of the core genome using the MAFFT program (Nakamura et al., 2018). The genes in the core genome were clustered based on a minimum of 95% amino-acid sequence similarity (default parameter) using blastp.

Roary was run with default parameters except for core gene definition (-cd 100; present in all isolates) and the number of threads (-p4; assigning all the 4 CPU cores). The parameter 'cd' is the core gene definition which enables the user to assign their own core gene definition where the default setting is 99%. A 'strict' definition of core gene of 100% (-cd 100) was used to include only genes that were present in all the isolates, as to avoid bias in downstream analyses due to the presence of different CCs. The parameter 'p' stands for the number of threads or the number of threads of execution where a program splits its process into simultaneously running tasks. The default threads parameter was 1, which meant the process does not split to run more than one task at a time. The number of threads was increased to the maximum capacity of the computer system to increase the speed of processing.

The command line used for Roary in this study was as follows:

```
roary -f outputdirectory -e --mafft -r -p 4 -cd 100 inputdirectory/*.gff
```

The main output files produced by Roary are summary statistics, core gene nucleotide alignment, and a gene presence/absence spreadsheet. The summary statistics file lists the number of genes in the pangenome divided into four categories; 1) core: present in 100% of the isolates, 2) soft-core: present in 95% to 99% of the isolates, 3) shell genes: present in 15% to 94% of isolates and 4) cloud genes: present in 0% to 14% of isolates. According to the strict definition in this study, the genes from the first category were used as core genome while the rest (2-4) were considered as accessory

genome.

A core gene nucleotide alignment was obtained using the concatenation of all the core genes of all the isolates and this alignment was used to analyse the core genomes. The core genomes were compared between host species using two methods: 1, a Neighbor-joining (NJ) tree; and 2, an analysis of molecular variance (AMOVA).

The NJ tree was constructed in Geneious v6.1 (Biomatters, Auckland, New Zealand), based on the pairwise Hamming distance matrix of the nucleotide alignment and using the Jukes-Cantor model (Jukes and Cantor, 1969).

Single nucleotide polymorphism (SNP) is the alteration of a single nucleotide in a specific position in the genome. SNP analysis can be utilised to identify the genetic association between *S. aureus* from various outbreaks, regions and hosts (Jian and Li, 2021). Core gene SNP analysis was implemented using Snippy (Seemann, 2018c) by mapping the reads of the isolates against the reference genome 1280.21671, a methicillin-susceptible *S. aureus* (MSSA) belonging to ST1 isolated from bovine mastitis, which was downloaded from the PATRIC database (Wattam et al., 2014). Snippy identifies SNP against the reference for each isolate and then generates a list of core SNPs at various positions for each isolate. Using this output, the genetic differentiation between various hosts at the core genome level was estimated statistically by AMOVA using GenAlex v6.5 program (Peakall and Smouse, 2012).

In addition, separate NJ trees were constructed for Roary core gene alignment for each CCs that were found in multiple isolates, in more than one host species. This was undertaken to increase the number of core genes aligned as a result of the alignment of a smaller number of genomes, and improve tree resolution within each CC, as previously described (Grinberg et al., 2017).

4.2.9 ACCESSORY GENOME COMPARISONS

The accessory genome predominantly comprises genes that encode for virulence factors and antimicrobial resistance. They are mobile genetic elements carried in transposons, pathogenicity islands and bacteriophages, hence, they are present in variable proportions of isolates. Analysis of the accessory genome is essential for identifying the presence of host-specific/host-adaptive genes.

In this study, the accessory genome was defined as the set of genes that were not observed in all the isolates.

Accessory genome comparisons between the host species were performed using the Roary gene presence-absence (binary) spreadsheet. Using this spreadsheet, Principal Coordinate Analysis (PCoA) was undertaken to visualise patterns of accessory genome segregation by host species and/or by CCs. PCoA was implemented using PAST software (Hammer et al., 2001). The gene presence-absence binary matrix was also used to statistically estimate the genetic segregation of the accessory genomes between host species using AMOVA, as implemented in GenAlex v6.5, as with the snippy core gene list output.

4.2.10 INDIVIDUAL-GENE COMPARATIVE ANALYSIS

The presence or absence, or mutations and truncations of single genes can profoundly affect the *S. aureus* phenotype (Mairpady Shambat et al., 2016). Hence, comparative analyses of several regulatory and virulence genes was undertaken. The accessory gene regulator (*agr*) system is one of the central regulatory operons in *S. aureus*. The agr operon consists of *agrA*, *agrB*, *agrC* and *agrD* genes. *S. aureus* are grouped into four distinct groups based on the *agrC* and *agrD* sequences as agr-I, agr-II, agr-III, and agr-IV (Jarraud et al., 2000). In this study, the agr operons were identified in the genomes using AgrVATE (`https://github.com/VishnuRaghuram94/AgrVATE`), which maps the assembled contigs to an agr database containing *agr* alleles, and also extracts the agr operon using *in silico* polymerase chain reaction if they are intact in contigs.

In addition, virulence genes and antimicrobial resistance genes were extracted using ARIBA (Hunt et al., 2017), which maps the reads to reference sequences downloaded from databases, namely, the Virulence Factors Database (VFDB) (Chen et al., 2005) and Resfinder (Bortolaia et al., 2020). The ARIBA virulence gene presence/absence matrix extracted from the VFDB was analysed for the presence of known host-specific/adaptive virulence genes in the genomes. In particular, the genes (Table 4.1) previously reported to be host specific/adaptive were searched (Matuszewska et al., 2020).

Table 4.1: List of host specific/adaptive genes of *S. aureus* in humans and ruminants along with their encoded protein and function (adapted from (Matuszewska et al., 2020).

Host	Genes	Protein	Function
Human	*scn*	Staphylococcal complement inhibitor (SCIN)	Blocks the activation of the human complement system (unknown functionality in bovine)
	sak	Staphylokinase (SAK)	Converts human plasminogen to plasmin (unknown functionality in bovine)
	chp	Chemotaxis inhibitory proteins of staphylococci (CHIPS)	Inhibits the infiltration of human neutrophils (unknown functionality in bovine)
	sea	Staphylococcal enterotoxin A (SEA)	Superantigen
	lukS-PV and *lukF-PV*	Panton Valentine leukocidin (PVL)	Lysis of human neutrophils (host adaptive: does not produce lysis of bovine neutrophils)
Ruminants	*tsst*	Toxic shock syndrome toxin (TSST)	Activates bovine T cells
	vWbp	Von Willebrand binding protein (VWBP)	Coagulation of ruminant plasma
	sec	Staphylococcal enterotoxin C (SEC)	Superantigen
	sel	Staphylococcoal enterotoxin L (SEL)	Superantigen
	lukM and *lukF'-PV*	Leukocidins	Lysis of ruminant neutrophils (host adaptive: does not produce lysis of human neutrophils)

4.3 RESULTS

4.3.1 READS AND ASSEMBLY QUALITY

The reads of the bovine isolates had the Nextera adapters and sequencer contaminant adapters identified and removed using the QCtool. The reads of the small-ruminant isolates sequenced at the 'Centro di Ricerca Pediatrica Romeo ed Enrica Invernizzi', University of Milan, Italy, and the human/canine/feline isolates did not have any adapters or contamination and did not require trimming.

The genomes of two small-ruminant-associated *S. aureus* isolates were discarded

from further analysis due to poor assembly quality. The GC% of the isolates were 34.25% and 34.01% (acceptable GC% range - 32.6% to 32.9%) and the total lengths were 3.3 Mbp and 2.4 Mbp, respectively (acceptable range: 2.6 Mbp to 2.9 Mbp). Selected QUAST metrics of the assemblies of the 279 *S. aureus* isolates from all host species are detailed in Table A.2.

4.3.2 CLONAL COMPLEXES (CCs) AND MULTILOCUS SEQUENCE TYPES (MLST)

The distribution of CCs across the three host species are shown in Table 4.2. The most common CC observed was CC1 (157/277 isolates; 56.7%) which was the dominant CC in cattle (140/188). Interestingly, in 2014, CC1 was the most common CC identified in humans in New Zealand (Heffernan et al., 2015), and it was the second most common CC (16/59) among human/canine/feline genomes analysed in this study. The second most common CC identified in this study was CC97 (27/277 isolates; 9.7%), which was only represented in bovine isolates (27/188; 14.4%). The third was followed by CC8, represented in 6/188 bovine (3.2%) and 8/30 small ruminant (26.7%) isolates, but not observed in human, canine and feline isolates. A total of 15 novel STs were identified among 30/277 (10.8%) isolates. These were submitted to PubMLST and assigned new ST identifiers by the database curators. One novel ST, ST5367 (5/277 isolates; 1.8%) did not have any associated SLV or DLVs. Hence, in this study it is considered as CC5367. The CC5367 was identified in bovine (2/188; 1.1%) and small ruminant (3/30; 10%) isolates. One canine and one feline isolate belonging to CC188, which were identified as ST188 by Grinberg et al. (2017) were noted to have the same SNP at the *tpi* locus and were both assigned as ST188v. A human isolate belonging to CC30 identified as ST30 by Grinberg et al. (2017) was noted to have a SNP at the *arcC* locus, and was assigned as ST30v. These minor discrepancies between the two studies might be attributed to the different genome assembly or different MLST identification methods used.

Table 4.2: Distribution of clonal complexes and sequence types (STs) for the 277 *S. aureus* isolates used in this study across various host species (bovine, small ruminants and human/canine/feline). The prevalence percentages are given in parentheses.

Clonal complex	Sequence types	All hosts (n=277)	Bovine isolates (n=188)	Small ruminant isolates (n=30)	Human, canine and feline isolates (n=59)
CC1	ST1	149 (53.8)	133 (70.7)	0	16 (27.1)
	ST3615	1 (0.4)	0	1 (3.3)	0
	ST4551	1 (0.4)	1 (0.5)	0	0
	ST6140	2 (0.7)	2 (1.1)	0	0
	ST6141	2 (0.7)	2 (1.1)	0	0
	ST6161	1 (0.4)	1 (0.5)	0	0
	ST6163	1 (0.4)	1 (0.5)	0	0
	Total	157 (56.7)	140 (74.5)	1 (3.3)	16 (27.1)
CC97	ST97	16 (5.8)	16 (8.5)	0	0
	ST6160	8 (2.9)	8 (4.3)	0	0
	ST6162	1 (0.4)	1 (0.5)	0	0
	ST6164	1 (0.4)	1 (0.5)	0	0
	ST71	1 (0.4)	1 (0.5)	0	0
	Total	27 (9.7)	27 (14.4)	0	0
CC151	ST151	6 (2.2)	6 (3.2)	0	0
	ST705	1 (0.4)	1 (0.5)	0	0
	Total	7 (2.5)	7 (3.7)	0	0
CC8	ST8	11 (4.0)	3 (1.6)	8 (26.7)	0
	ST6143	3 (1.1)	3 (1.6)	0	0
	Total	14 (5.1)	6 (3.2)	8 (26.7)	0
CC133	ST133	8 (2.9)	0	8 (26.7)	0
	ST701	1 (0.4)	0	1 (3.3)	0
	ST1247	1 (0.4)	1 (0.5)	0	0
	ST6137	1 (0.4)	0	1 (3.3)	0
	ST6138	1 (0.4)	0	1 (3.3)	0
	ST6139	1 (0.4)	0	1 (3.3)	0

Continued on next page

Clonal complex	Sequence types	All hosts (n=277)	Bovine isolates (n=188)	Small ruminant isolates (n=30)	Human, canine and feline isolates (n=59)
	ST6157	1 (0.4)	0	1 (3.3)	0
	ST6165	1 (0.4)	1 (0.5)	0	0
	ST6166	1 (0.4)	0	1 (3.3)	0
	Total	16 (5.8)	2 (1.1)	14 (46.7)	0
CC5367	ST5367	5 (1.8)	2 (1.1)	3 (5.1)	0
CC5	ST5	14 (5.8)	1 (0.5)	0	13 (25.4)
	ST835	1 (0.4)	0	0	1 (1.7)
	ST1259	1 (0.4)	0	0	1 (1.7)
	Total	16 (5.8)	1 (0.5)	0	15 (25.4)
CC45	ST508	1 (0.4)	1 (0.5)	0	0
CC15	ST15	2 (0.7)	0	0	2 (3.4)
	ST199	1 (0.4)	1 (0.5)	0	0
	ST582	1 (0.4)	0	0	1 (1.7)
	Total	4 (1.4)	1 (0.5)	0	3 (5.1)
CC78	ST78	1 (0.4)	1 (0.5)	0	0
CC30	ST30	5 (2.8)	0	0	5 (13.5)
	ST34	1 (0.4)	0	0	1 (1.7)
	ST39	1 (0.4)	0	0	1 (1.7)
	ST30v	1 (0.4)	0	0	1 (1.7)
	Total	8 (2.8)	0	0	8 (13.5)
CC692	ST692	1 (0.4)	0	1 (3.3)	0
CC1640	ST1640	3 (1.1)	0	3 (10.0)	0
CC188	ST188	15 (6.1)	0	0	15 (28.8)
	ST188v	2 (0.7)	0	0	2
	Total	17 (6.1)	0	0	17 (28.8)

4.3.3 SPA-TYPES

The spa-types of 242/277 (87.4%) isolates could be identified by the spa-typing tool. The paired reads and assembled contigs of the 35 genomes for which the spa types could not be initially identified were subsequently run successfully through another web-based Spa-Typer tool (Spa-Typer 1.0; (Bartels et al., 2014) https://cge.cbs.dtu.dk/services/spatyper/) with the sequencing platform set to Illumina paired-end reads and assembled genome/contigs, respectively. The comparison between the spa-types obtained by the two programs using 10 random isolates returned identical results.

The most common spa-type identified in this study was t114 (107/277, 38.6%), which was only found in bovine CC1 *S. aureus*. Nineteen CC1/ST1 (6.9%) isolates were spa-type t127, and were distributed across bovine (6/19; 31.6%), human (8/19; 42.1%), and companion-animals (5/19; 26.3%). Interestingly, t127 was the most common spa-type in humans in New Zealand in 2014 (Heffernan et al., 2015). Spa-type t084 was observed in CC15 and was distributed among bovine (1/3; 33%) and humans (2/3; 67%) . There were 16 new spa sequences, but these could not be assigned spa-type identifiers as the RidomStaphType software was not available for this study. One of the newly identified spa sequences defined in this study as tnew8 (Table 4.3) was observed in the bovine (2/5) and small ruminant (3/5) isolates belonging to the novel CC5367.

Table 4.3: Distribution of spa-types identified among the 277 *S. aureus*, along with the number of isolates (N) and percentage of total in parantheses. The clonal complex (CC) associated with the spa-type is provided along with the sequence types (STs). The new spa-types are defined by the observed repeat-type sequence numbers.

Spa-type	Host species	Associated CCs	N (%)
t114	Bovine	CC1(ST1, ST6140, ST6161, ST4551)	107 (38.6)
t127	Bovine, Canine, Feline, Human	CC1 (ST1)	19 (6.9)
t189	Canine, Feline, Human	CC188(ST188, ST188v)	15 (5.4)
t524	Bovine	CC97(ST6160, ST71)	9 (3.2)
t922	Bovine	CC1(ST1, ST6163)	7 (2.5)

Continued on next page

Spa-type	Host species	Associated CCs	N (%)
t529	Bovine	CC151(ST151, ST705)	7 (2.5)
t059	Sheep	CC8 (ST8)	5 (1.8)
t1265	Canine, Human	CC5 (ST5)	5 (1.8)
tnew8 (14-23-17-362-13-22-22)	Bovine, Sheep	CC5367 (ST5367)	5 (1.8)
tnew10 (03-12-21-17-23-13-17-17-23-24)	Sheep	CC133(ST6157, ST6138, ST133)	5 (1.8)
t1784	Bovine	CC1 (ST1)	4 (1.4)
t693	Bovine	CC1(ST1)	4 (1.4)
t118	Sheep	CC8 (ST8)	3 (1.1)
t179	Canine, Human	CC5(ST5)	3 (1.1)
t084	Bovine, Human	CC15 (ST199, ST15)	3 (1.1)
t2699	Bovine	CC1(ST1)	3 (1.1)
t211	Bovine	CC8 (ST6143)	3 (1.1)
t008	Bovine	CC8 (ST8)	2 (0.7)
t019	Human	CC30 (ST30)	2 (0.7)
t16350	Bovine	CC97 (ST97)	2 (0.7)
t17282	Bovine	CC97 (ST97)	2 (0.7)
t2207	Canine	CC1(ST1)	2 (0.7)
t267	Bovine	CC97 (ST97)	2 (0.7)
t3047	Sheep	CC133(ST133, ST6139)	2 (0.7)
t3380	Bovine	CC97(ST97)	2 (0.7)
t4735	Sheep	CC133(ST133, ST701)	2 (0.7)
t6980	Bovine	CC1(ST1, ST6140)	2 (0.7)
tnew1 (07-16-02-13)	Bovine	CC1 (ST6141)	2 (0.7)
t002	Human	CC5 (ST5, ST835)	2 (0.7)
t010	Human	CC5 (ST5)	1 (0.4)
t012	Human	CC30 (ST30v)	1 (0.4)
t015	Bovine	CC45 (ST508)	1 (0.4)
t062	Bovine	CC5(ST5)	1 (0.4)
t089	Canine	CC30 (ST34)	1 (0.4)

Continued on next page

Spa-type	Host species	Associated CCs	N (%)
t1166	Bovine	CC133 (ST1247)	1 (0.4)
t1407	Bovine	CC1(ST1)	1 (0.4)
t1509	Canine	CC15(ST582)	1 (0.4)
t15807	Sheep	CC133(ST133)	1 (0.4)
t16141	Human	CC1 (ST1)	1 (0.4)
t177	Sheep	CC1 (ST3615)	1 (0.4)
t186	Bovine	CC78(ST78)	1 (0.4)
t1931	Bovine	CC1(ST1)	1 (0.4)
t2246	Bovine	CC1(ST1)	1 (0.4)
t2421	Bovine	CC97 (ST97)	1 (0.4)
t2802	Bovine	CC97(ST97)	1 (0.4)
t2868	Human	CC30(ST30)	1 (0.4)
t386	Bovine	CC1(ST1)	1 (0.4)
t4453	Sheep	CC1640(ST1640)	1 (0.4)
t4540	Bovine	CC1(ST1)	1 (0.4)
t4558	Feline	CC188(ST188)	1 (0.4)
t4601	Canine	CC1(ST1)	1 (0.4)
t4682	Bovine	CC97(ST6164)	1 (0.4)
t5150	Human	CC5(ST5)	1 (0.4)
t5213	Canine	CC5(ST5)	1 (0.4)
t5402	Canine	CC5(ST1259)	1 (0.4)
t6690	Human	CC5(ST5)	1 (0.4)
t7310	Sheep	CC133(ST133)	1 (0.4)
t7623	Canine	CC30(ST30)	1 (0.4)
t865	Bovine	CC97(ST6162)	1 (0.4)
t8914	Canine	CC188(ST188)	1 (0.4)
t1414	Canine	CC30(ST30)	1 (0.4)
t17906	Bovine	CC133(ST6165)	1 (0.4)
t2271	Canine	CC30 (ST39)	1 (0.4)
tnew11 (03-16-05-17-23-13-17-17-17-17-17-23-24)	Sheep	CC133 (ST6137)	1 (0.4)

Continued on next page

Spa-type	Host species	Associated CCs	N (%)
tnew12 (03-16-05-17-23-13-17-17-17-17-23-24)	Sheep	CC133(ST133)	1 (0.4)
tnew13 (03-16-12-21-17-23-13-17-17-23-31-24)	Sheep	CC133(ST6166)	1 (0.4)
tnew2 (07-23-12-05-17-02-17-34-34-34-34-33-34)	Bovine	CC97(ST97)	1 (0.4)
tnew3 (07-23-12-05-17-34-34-13-33-34)	Bovine	CC97(ST97)	1 (0.4)
tnew4 (07-23-12-05-22-34-13-34-13-33-34)	Bovine	CC97 (ST97)	1 (0.4)
tnew5 (11-19-12-21-17-34-24-34-24-34-22-33-25)	Bovine	CC8 (ST8)	1 (0.4)
tnew14 (121-21-16-82-24-17-17-17-17-17)	Sheep	CC1640 (ST1640)	1 (0.4)
tnew6 (14-12-05-17-34-34-13-33-34)	Bovine	CC97 (ST97)	1 (0.4)
tnew7 (14-12-05-17-34-34-34-13-33-34)	Bovine	CC97 (ST97)	1 (0.4)
tnew9 (16-34-33-13)	Bovine	CC1 (ST1)	1 (0.4)
tnew15 (26-23-12-21-17-34-34-25-17)	Sheep	CC692 (ST692)	1 (0.4)

Continued on next page

Spa-type	Host species	Associated CCs	N (%)
tnew16 (591-16-34-17-82-24-17-17-17-17)	Sheep	CC1640 (ST1640)	1 (0.4)

4.3.3.1 CORE GENOME COMPARISONS

After the removal of duplication per CC/farm combination, there were 212 isolates (Bovine: 136, Small ruminants: 17, Human, canine and feline: 59).

A total of 7276 genes (pangenome) were identified across the 212 *S. aureus* isolates, of which 1448 (19.9%) were core genes (the pangenome size of the bovine isolates without removing duplicates is reported in Chapter3).

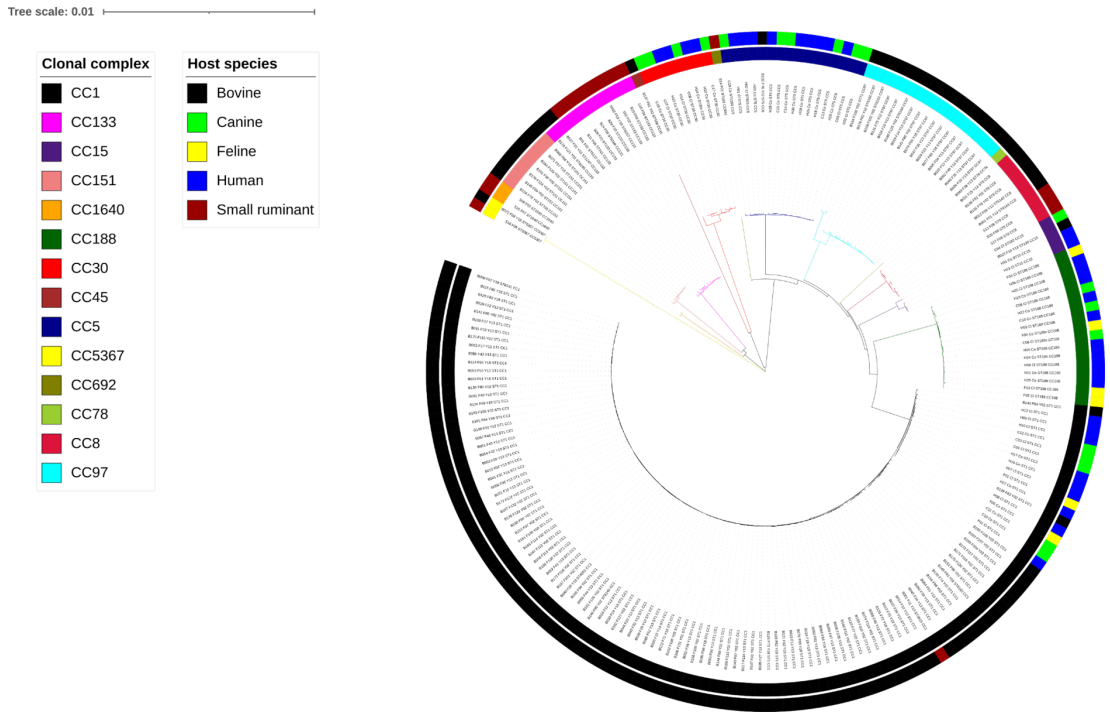

Figure 4.1: Neighbour-joining tree of the core gene alignment distance matrices (calculated by the Jukes-Cantor method) of the 212 *S. aureus* isolates isolated from bovine (n=136), small ruminants (n=17), canine, feline, and human (n=59) hosts. The outer colour band represents the host species from which *S. aureus* was isolated and the inner colour band indicates the clonal complexes (CCs).

The NJ tree constructed from the core genome alignment of the 1448 genes of the 212 *S. aureus* isolates is shown in 4.1 (for better image resolution, the image can be accessed from https://github.com/jabinnes/Thesis-files/blob/main/Core_gene_analysis/Core_gene_onefarm_all_NJtree.png). As expected from a previous study (Grinberg et al., 2017), the segregation of the isolates was broadly based on the CCs, with each CCs forming a monophyletic group. Noteworthily, within the CCs represented in more than one host species, such as CC8, CC133, CC5367, which were present in bovine and small ruminants, there was segregation based on host species, with each host species forming a monophyletic group. Also, CC1 which was the predominant CC in cattle and humans and was found in all the host species, segregated based on host (human/canine/feline identified as a set), except for two bovine isolates (B139 and B140) that grouped with the human isolates, and one human isolate (H11) that grouped with the bovine isolates.

To increase the discriminatory power of the trees, NJ trees were also constructed using the core genomes of individual CCs that were found in more than one host species and were represented by more than one isolate per species. The 117 CC1 isolates had 2009 core genes (48.5%) out of 4146 total genes. As in the previous tree, the NJ tree constructed from the core gene alignment of the 2009 core genes of CC1 (https://github.com/jabinnes/Thesis-files/blob/main/Core_gene_analysis/Core_gene_onefarm_CC1_NJtree.png) showed a marked clustering based on host species, except for one bovine isolate (B139) segregating with human isolates and one human isolate (H11) segregating with bovine isolates.

The eight CC8 isolates represented in bovine (5) and small ruminants (3) had a core genome composed of 2334/3050 (76.5%). Also, the NJ tree constructed from the core genome alignment of the eight isolates showed a partition with each host forming a monophyletic group (https://github.com/jabinnes/Thesis-files/blob/main/Core_gene_analysis/core_gene_onefarm_CC8_NJtree.png).

The 11 CC133 isolates (Bovine: 2, Small ruminants: 9) had core genome composed of 2324/3091 (75.2%) genes. The NJ tree constructed from the core genome alignment of the 11 isolates also showed segregation, with each host forming a monophyletic group (https://github.com/jabinnes/Thesis-files/blob/main/Core_gene_analysis/core_gene_onefarm_CC133_NJtree.png).

The core gene SNPs AMOVA was unable to be estimated using GenAlex as the

number of SNPs calculated against the reference genome for the 212 isolates were 90993, exceeding the 16384 columns that can be loaded in Microsoft Excel. The high number of SNPs was due to the high number of CCs represented in the dataset.

The core gene SNPs of the 116 CC1 isolates were 7102. The lone small-ruminant CC1 isolate was not included for AMOVA estimation as a population represented by one isolate cannot be used for calculating the AMOVA statistics. The AMOVA results indicated a significant partition of CC1 among humans/feline/canine, and bovine isolates (Phi_{PT}: 0.429, $p < 0.001$), consistent with the partition observed in the respective NJ tree.

4.3.4 ACCESSORY GENOME COMPARISONS

The accessory genome analysis of the gene presence-absence matrix based on PCoA showed a segregation of the 212 *S. aureus* isolates based on their clonal complexes (Figure 4.2). When only the matrix of CC1 was plotted, the visual resolution increased, to put in evidence of a marked segregation of the isolates based on their host species, with bovine isolates separated from the group of human/canine/feline isolates (Figure 4.3). Interestingly, the bovine isolate (B139), which segregated with human isolates in the core genome NJ tree, grouped near the human cluster of isolates and the human isolate (H11), which segregated with bovine isolates, grouped near the bovine cluster.

The results of the AMOVA using the gene presence-absence matrix of all the 212 isolates indicated that the within-host variation accounted for 77% and the between-host variation for 23% of the total variation. The pairwise population Phi_{PT} value between bovine and human/canine/feline isolates was statistically significant (0.208; $p > 0.0001$). The AMOVA between human/canine/feline and small-ruminant isolates was statistically significant (0.238; $p > 0.0001$). Also, the AMOVA result between bovine and small ruminants was significant(0.298; $p > 0.0001$).

To remove the bias determined by the CC structure of the sample, the AMOVA was estimated for the CC1 isolates belonging to bovine and human/canine/feline isolates. The within-host variation accounted for 62%, while the between-host variation accounted for 38% of the total variation. The Phi_{PT} value was 0.385, and it was

highly significant ($p > 0.0001$), indicating a significant partition of the accessory genome of CC1 among the bovine and the human/canine/feline isolates, consistent with the results of the PCoA.

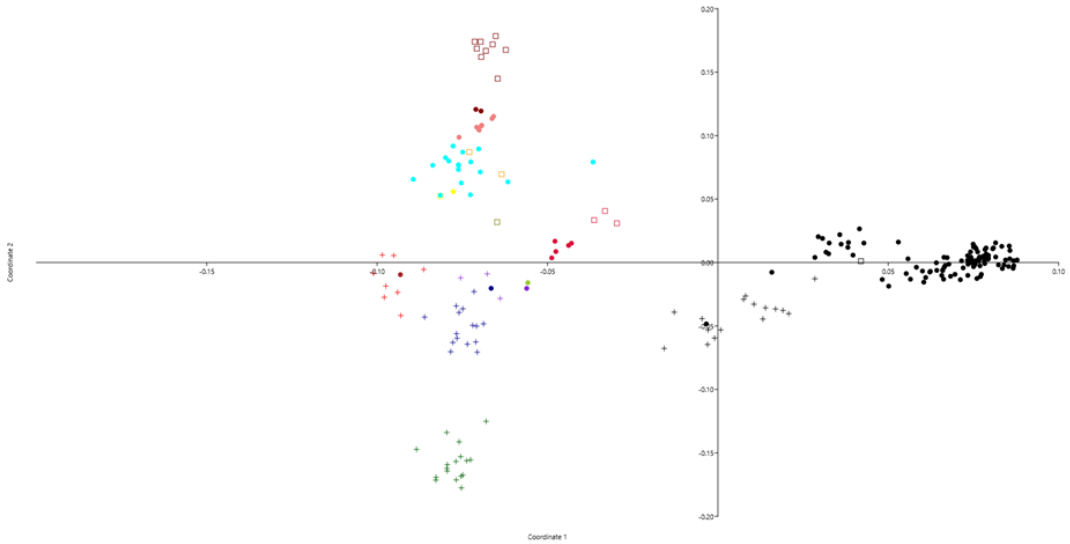

Figure 4.2: Principal coordinate analysis (PCoA) plot obtained using the binary gene/absence matrix of the 212 *S. aureus* isolates identified according to their clonal complexes(CCs) and host species. The isolates are represented by data points positioned in the bi-dimensional space delineated by the two main coordinates. Host species: Bovine – Dots, Human, canine and feline – Plus and Small ruminants – Squares. CCs: Black - CC1, Brown - CC45, Dark blue- CC5, Crimson - CC8, Blueviolet - CC15, Red - CC30, Yellow-green - CC78, Aqua - CC97, Maroon - CC133, Lightcoral - CC151, Darkgreen - CC188, Olive - CC692, Orange - CC1640, Yellow- CC5367.

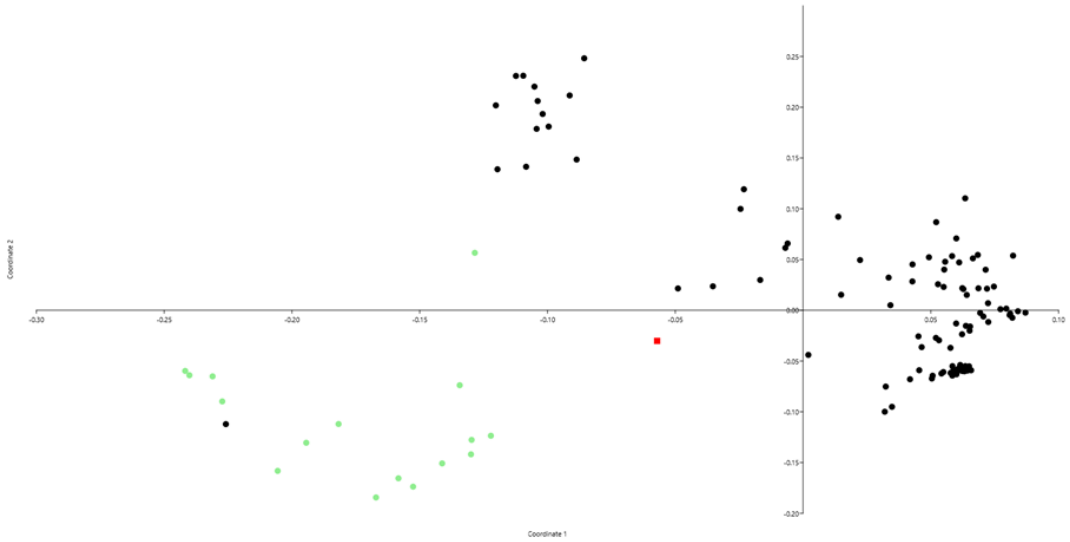

Figure 4.3: Principal coordinate analysis (PCoA) plot obtained using the binary gene/absence matrix of the 117 *S. aureus* isolates belonging to CC1. The isolates are represented by data points positioned in the bi-dimensional space in the two main coordinates. Black dots- Bovine, light green dots- human/canine/feline, and red square- Small ruminant isolate. Notice the segregation of the isolates based on their hosts except for one human isolate amongst the bovine isolates and one bovine isolate amongst the human/canine/feline isolates. The solitary small ruminant isolate did not cluster with human/canine/feline or bovine isolates

4.3.5 INDIVIDUAL GENE ANALYSIS

The agr groups identified by AgrVATE software in the various hosts are shown in Table 4.4. The major agr group identified was agr-III (126/212; 59.4%) which was associated with CC1, CC78 and CC30. The agr-III group was distributed across bovine, human, canine, feline and small ruminant hosts as CC1 was distributed across these hosts. The agr-I and agr-II were also identified across all the hosts and consisted of isolates belonging to various CCs as shown in Table 4.4. As expected, only one agr group was associated with each CC (Monecke et al., 2008). Interestingly, the agr-IV was identified only in small ruminant isolates belonging to CC1640. Agr groups of two isolates could not be identified because the contigs from these two isolates did not map to any *agrD* gene allele in the database. These isolates belonged to CC5 and CC188 from feline and human hosts.

A total of 197 virulence genes were identified among the 211 isolates (one human isolate, H14 had failed to run in the ARIBA process). The PCoA of the presence/absence matrix of the 197 virulence genes across 211 isolates showed clustering broadly based on their clonal complexes except for one canine isolate (C01) belonging to CC1, which segregated together with CC133 and CC15. Within each CC, the isolates segregated based on their host species (Figure 4.4). To improve the graphic resolution, a new PCoA of the presence/absence matrix was calculated using the 145 virulence genes found in the 117 CC1 isolates (Figure 4.5). Also, in this case, the isolates segregated based on the host species, except for one bovine isolate (B139) that segregated together with the human cluster. This was the same isolate that segregated with human isolates in both core and accessory genome analysis.

Table 4.4: Distribution of agr groups in the 212 *S.aureus* isolates across various hosts and clonal complexes. The prevalence percentages are given in parantheses.

Agr Group	Number of isolates (%)	Hosts (Number of isolates)	Clonal complexes
agr-I	54 (25.5)	Bovine (25), Human (10), Canine (3), Feline (3), Small ruminants (13)	CC8, CC133, CC188, CC97, CC45, CC692
agr-II	28 (13.2)	Bovine (10), Human (10), Canine (7), Small ruminants (1)	CC15, CC151, CC5, CC5367
agr-III	126 (59.4)	Bovine (101), Human (13), Canine (9), Feline (2), Small ruminants (1)	CC1, CC78, CC30
agr-IV	2 (0.9)	Small ruminants (2)	CC1640
Unknown	2 (0.9)	Feline (1), Human (1)	CC5, CC188

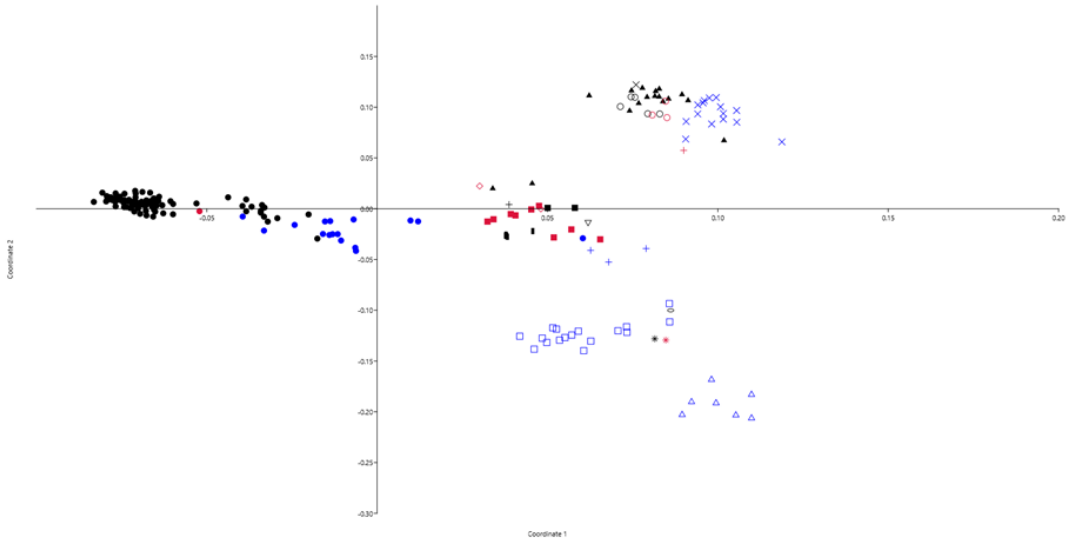

Figure 4.4: Principal Coordinate Analysis (PCoA) plots obtained using the binary matrix of the virulence genes identified by ARIBA against Virulence Factors Database (VFDB) of the 211 *S. aureus* isolates. The isolates are represented by data points positioned in the bi-dimensional space delineated by the two main coordinates. The isolates are identified by their host species: Bovine-black, human/canine/feline- blue, and small ruminants- red. The isolates are identified by their clonal complexes (CCs): CC1 – dots, CC8 – 'O', CC97 – solid triangle, CC15 and CC692 – plus, CC133 – solid square, CC78 – inverted triangle, CC151 – bar, CC5367 – star, CC5 – X, CC45 – oval, CC188 – square, CC30 – triangle, and CC1640 – diamond.

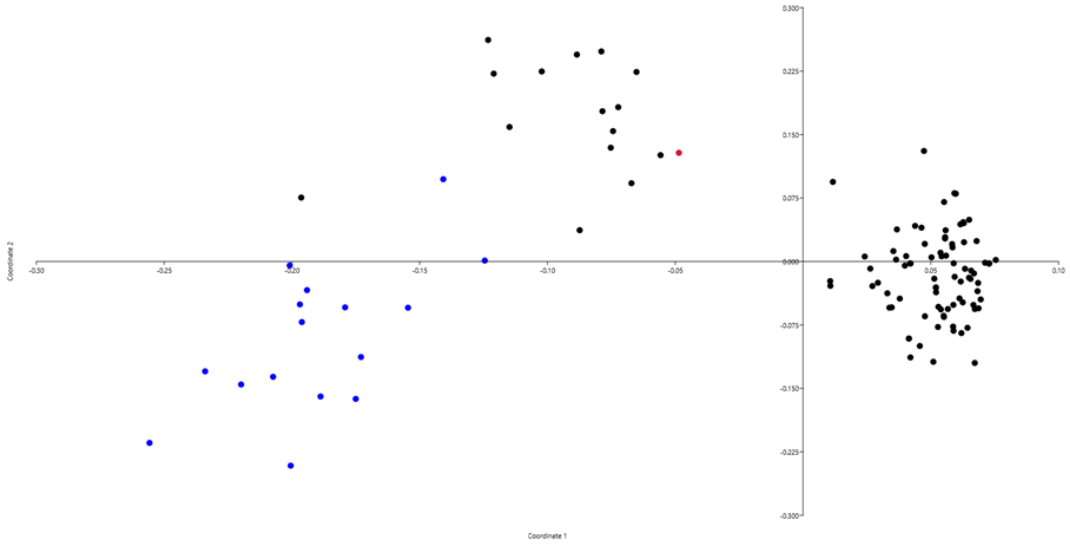

Figure 4.5: Principal Coordinate Analysis (PCoA) plots obtained using the binary matrix of the virulence genes identified by ARIBA against Virulence Factors Database (VFDB) of the 117 CC1 *S. aureus* isolates. The isolates are represented by data points positioned delineated by the bi-dimensional space in the two main coordinates. The isolates are identified by their host species: Bovine – black, human/canine/feline – blue and Small ruminants – red.

A gene-by-gene inspection was performed to assess the presence of putative host-specific/adaptive genes in the virulence gene absence/presence matrix (Table 4.5). The human-adaptive virulence factors carried in prophage ϕSa*int*3: staphylococcal complement inhibitor (*scn*), Chemotaxis inhibitory protein (*chp*), Staphylokinase (*sak*), and Staphylococcal enterotoxin variant A (*sea*), which form the immune-evasion cluster (IEC) (van Wamel et al., 2006; Richardson et al., 2018), were found in the majority of human/canine/feline isolates (*scn*: 39/58; *chp*: 32/58; *sak*: 49/58; *sea*:13/58;), only in a few bovine isolates (*scn*: 4/136; *chp*: 2/136; *sak*: 4/136; *sea*: 1/136;), and were absent in small-ruminant isolates. Interestingly, two bovine isolates (B023 of CC15, and B137 of CC45) contained all these four virulence factors and the bovine isolate that segregated with the human cluster in the core and accessory gene analyses (B139 of CC1) contained the *scn* and *sak* genes.

The human-adaptive bi-component Panton-Valentine Leukocidin (PVL), factors *lukF-PV* and *lukS-PV* (Löffler et al., 2010), were found in 6/58 (10.3%) human/canine/feline isolates, and in isolate B139, the bovine CC1/ST1 isolate that segregated with the human isolates in the previous analyses. The equivalent ruminant-adaptive bi-component factor, *lukF'-PV* (lukF-P83) and *lukM* (Vrieling et al., 2016) was found only in ruminant *S. aureus* isolates. In particular, the *lukF'-PV* factor was found in 90/136 (66.2%) bovine and 11/17 (64.7%) small-ruminant isolates, while the *lukM* factor was found in 89/136 (65.4%) bovine and 12/17 (70.6%) small-ruminant isolates. These ruminant adapted factors were not found in human/canine/feline isolates.

The von-Willebrand binding protein (*vWbp*) was found predominantly in isolates belonging to CC1 (111/117; 94.9%), followed by CC1640 (2/2; 100%) and CC8 (3/8; 37.5%) and, as expected due to the distribution of CC1, mostly in bovine (100/136; 73.5%), followed by human/canine/feline (13/58; 22.4%) and small-ruminant (3/17; 17.6%) isolates.

Among the staphylococcal enterotoxins, the *sec* variant was found largely among small-ruminant isolates (9/17; 52.9%) belonging to CC133. Two (1.5%) of the bovine isolates had the *sec* variant (CC78 and CC151), and none of the human isolates had this enterotoxin. The toxic shock syndrome toxin (*tsst*) was found in one bovine isolate (B134) belonging to CC151, the same isolate which contained the *sec* enterotoxin. In small ruminants, the *tsst* was found in 9/17 (52.9%) isolates, all

Table 4.5: Distribution of the virulence factors analysed using Virulence Factors Database (VFDB). The number of isolates containing each gene is listed for bovine, human/canine/feline and small ruminants. The prevalence percentage of each gene for each group is given in parantheses

Virulence Factors	Bovine		human/ canine/ feline (n=58)	Small ruminants	
	All (n=188)	One per farm-CC (n=136)		All (n=30)	One per farm-CC (n=17)
scn	4 (2.1)	4 (2.9)	39 (67.2)	0	0
sak	4 (2.1)	4 (2.9)	49 (84.5)	0	0
chp	2 (1.1)	2 (1.5)	32 (55.2)	0	0
sea	1 (0.5)	1 (0.7)	13 (22.4)	0	0
lukF-PV	1 (0.5)	1 (0.7)	6 (10.3)	0	0
lukS-PV	1 (0.5)	1 (0.7)	6 (10.3)	0	0
lukF'-PV	128 (68.1)	90 (66.2)	0	19 (63.3)	11 (64.7)
lukM	127 (67.6)	89 (65.4)	0	20 (66.7)	12 (70.6)
vWbp	140 (74.5)	100 (73.5)	13 (22.4)	4 (13.3)	3 (17.7)
tsst	1 (0.5)	1 (0.7)	4 (6.9)	13 (43.3)	9 (52.9)
icaA	57 (30.3)	51 (37.5)	56 (96.6)	30 (100)	17 (100)
icaB	171 (91.0)	126 (92.6)	48 (82.8)	29 (96.7)	16 (94.1)
icaC	177 (94.1)	133 (97.8)	49 (84.5)	30 (100)	17 (100)
icaD	55 (29.3)	49 (36.0)	57 (98.3)	30 (100)	17 (100)
icaR	179 (95.2)	134 (98.5)	41 (70.7)	29 (96.7)	16 (94.1)

scn - Staphylococcal complement inihibitor, *sak* - Staphylokinase, *chp* - chemotaxis, *sea* - staphylococcal enterotoxin A, *lukF-PV* - F subunit of the Panton Valentine leucocidin, *lukS-PV* - S subunit of the Panton Valentine leucocidin, *lukF'-PV* - F subunit of the leukocidin lukMF', *lukM) - S subunit of the leukocidin lukMF',* vWbp - *vonWillebrand binding protein,* tsst - *Toxic shock syndrome toxin,* ica - *intercellular adhesion operon containing A, B, C, D and R genes.*

Table 4.6: Antimicrobial resistance (AMR) genes analysed using Resfinder database. The number of isolates containing each gene is listed for bovine, human/canine/feline and small ruminants. The prevalence percentages are given in parentheses.

AMR genes	Bovine		Human/ Canine/ Feline (59)	Small ruminants	
	All (188)	One per farm-CC (136)		All (30)	One per farm-CC (17)
ant	0	0	1 (1.7)	0	0
aph	0	0	2 (3.4)	0	0
blaZ	48 (25.5)	34 (25)	40 (67.8)	2 (6.7)	2 (11.8)
ermA	0	0	1 (1.7)	0	0
ermB	0	0	1 (1.7)	0	0
ermC	2 (1.1)	2 (1.5)	0	0	0
fuC	4 (2.1)	3 (2.2)	5 (8.5)	0	0
str	3 (1.6)	2 (1.5)	0	0	0

ant-aminoglycoside aminotransferase , *aph*-aminoglycoside phosphotransferase , *blaZ*-β-lactamase , *ermA*, *ermB*, and *ermC*-erythryomycin ribosomal methylases A, B and C, *fusC*- fusidic acid resistance, *str*- streptomycin resistance gene.

belonging to CC133, and in the human isolates, it was found in 4/59 (6.9%) isolates belonging to CC30.

Among intracellular adhesion genes (icaABCD), *icaA* and *icaD* were found in more than 96% of the isolates from human/canine/feline (*icaA*: 96.5%, *icaD:* 98.3%) and in small ruminants (*icaA*: 100%, *icaD*: 100%), while these genes were found only in ~36% of the bovine isolates (*icaA*: 37.5%, *icaD*: 36%). However, the *icaB* and *icaC* genes were found commonly in isolates from all the host species (*icaB*: bovine -93%, human/canine/feline- 83%, and small ruminants- 94%; *icaC*: bovine- 98%, human/canine/feline- 85%, and small ruminants – 100%). The host variation was more evident in CC1 isolates, as *icaA* was found only in 17/100 (17%) bovine CC1 isolates but 16/16 (100%) human/canine/feline CC1 isolates and *icaD* was found only in 15/100 (15%) bovine and in all 16 (100%) human/canine/feline CC1 isolates.

Antimicrobial resistance (AMR) genes identified by ARIBA in the Resfinder database included the aminoglycoside aminotransferase (*ant*), aminoglycoside phosphotransferase (*aph*), the β-lactamase (*blaZ*), three erythromycin ribosomal methylases (*ermA*, *ermB*, *ermC*), the fusidic acid resistance (*fusC*) and streptomycin resistance (*str*) genes. The distribution of the AMR genes is reported in Table 4.6.

Interestingly, the vast majority of bovine (100/136; 73.5%) and small-ruminant (15/17; 88.2%) isolates did not possess any AMR gene, compared to 17/59 (28.8%) of human/canine/feline isolates. The human/canine/feline isolates were 2.7 times (95% CI: 1.9 to 3.7) more likely to possess an AMR gene when compared to the bovine isolates. The relative risk of human/canine/feline isolates to possess an AMR gene increased to 6 times (95% CI: 1.6 to 22.5) when compared to small ruminant isolates. The bovine isolates were 2.2 times (95% CI: 0.6 to 8.5) more likely to possess an AMR gene when compared to the small ruminant isolates. Moreover, a significantly higher proportion of bovine CC1 than human CC1 isolates did not carry any AMR gene (bovine CC1: 85/100; human/feline/canine CC1: 4/16; Relative risk: 5 (95%CI: 2.9 to 8.6)). The ruminant associated CC151, CC1640, and CC5367 did not display any resistance genes.

The *blaZ* gene was the predominant resistance gene, present in 76/212 (35.8%) isolates. There were 26 (22.2%) CC1 isolates carrying the *blaZ* gene. Interestingly, 14/17 (82.5%) CC97 isolates possessed the *blaZ* gene. The *blaZ* gene was identified in a significantly greater proportion of human/canine/feline isolates (40/59; 67.8%) compared to bovine (34/136; 25%) and small ruminant (2/17; 11.8%) isolates. Within the CC1 isolates, the *blaZ* gene was found in 12/16 (75%) human/canine/feline isolates and 13/100 (13%) bovine isolates. The human/canine/feline CC1 isolates were 5.8 times (95%CI: 3.2 to 10.3) more likely to carry the *blaZ* gene compared to the bovine CC1 isolates.

The *fusC* gene was found in 8/212 (3.8%) isolates (bovine: 3/136, 2.2%; human/canine/feline: 5/59, 8.5%). As expected from previous studies in New Zealand (Williamson et al., 2014b; Heffernan et al., 2015), the *fusC* gene was present in CC5 (3/16; 18.8%). However, it was also found in CC8 (2/8; 25%). Interestingly, the *fusC* gene was found only in two CC1 isolates (2/117; 1.7%) which is in contrast to the studies by Williamson et al. (2014b) and Heffernan et al. (2015) where the *fusC* gene was found commonly in CC1 and CC5 isolates. One of the CC1 isolate possessing the *fusC* gene was a human isolate and the other one was a bovine isolate (B139).

4.4 DISCUSSION

This chapter reports a comparative genomic study of *S. aureus* across ruminants and humans in New Zealand. The results support the evidence of a partition of *S. aureus* between ruminants and humans, with very little strain-sharing between the two host types. The evidence spans all the proposed genomic domains: 1) the phylogenetic domain contemplating a clustering of *S. aureus* according to the host type; 2) the accessory genome domain, with an abundance of host-adaptive genes in isolates obtained from the corresponding host type.

Ruminant associated *S. aureus* strains have been described earlier with CC97, CC151, CC133 being found predominantly in ruminants (Zadoks et al., 2011; Bar-Gal et al., 2015; Leijon et al., 2021). These lineages appear to be extremely rare in humans. The results of the study presented in Chapter 3 show a very different picture for New Zealand: here the dominant lineage is CC1/ST1, a major lineage of humans worldwide, and the most common lineage in humans in New Zealand. We also found host-specific strains: CC133 and CC5367 isolates in ruminants, CC97 and CC151 from cattle, CC1640 from small ruminants, CC30 and CC188 identified in human/canine/feline isolates.

The host specificity of *S. aureus* when looked at with regards to phylogenetic clustering of *S. aureus* based on core gene alignment and accessory genome analysis revealed that *S. aureus* isolates clustered based on host species within CCs. Previous studies have tried to identify clustering of *S. aureus* irrespective of their CCs or tried to identify CCs that are host-specific. However, *S.aureus* is highly clonal and clusters predominantly based on their CCs, which creates a bias to identify clustering based on host species. Hence, the study of CC1 isolates that are shared between bovine and human (including canine and feline) helped in identifying the clustering of *S. aureus* isolates based on host species. Unfortunately, other CCs did not have enough representation from more than one host species to make a strong conclusion on host-based clustering within CCs. Regardless, CC8, CC133 had isolates clustering based on host species (bovine and small ruminants).

Similar to previous reports of host-specific/host-adaptive genes (Sung et al., 2008; Matuszewska et al., 2020; Mrochen et al., 2020), this study identified the immune evasion cluster genes, human-specific PVL genes (*lukF-PV* and *lukS-PV*) present

predominantly in the human isolates. However, there were few isolates from bovine isolates that contained these genes. This could be explained either by human contamination during sample collection or a recent host spillover event. The B139 isolate is mostly likely human contamination during sample collection as the core and accessory gene analysis clustered this isolate with the human isolate.

Ruminant adaptation genes were seen with *lukM* and *lukF'*-PV leuckocidins that were noticed only in ruminant isolates. One of the enterotoxin gene, *sec* was associated only with ruminant isolates and predominantly identified in CC133 isolates. The *sec* gene has been identified to play an important role in the severity of mastitis in ruminants and in food poisoning in humans from dairy products (Matsunaga et al., 1993; Argudín et al., 2010).

Variation in *ica* genes between hosts was identified with only 36% bovine isolates harbouring *icaA* and *icaD* genes compared to more than 96% in humans, pets and small ruminants. These genes play an important role in biofilm production during pathogenesis (Gad et al., 2009; Marques et al., 2021). However, these genes does not seem to host-specific/host-adaptive as these genes are found predominantly in biofilm producing bovine *S. aureus* isolates worldwide (Marques et al., 2021; Szweda et al., 2012) but are associated with biofilm producing strains.

Further studies are needed to understand the host adaptation to niches as *S. aureus* from bovine mastitis has demonstrated increased lactose utilisation, the major carbohydrate available in cow's milk (Richardson et al., 2018). Isolates from the same niche or disease process (mastitis) from bovine and humans are required to study the genomic host adaptation without the bias of niche.

The unique predominance of CC1 in both bovine and humans in New Zealand allowed to study the host adaptation without the bias of CCs. Phenotypic studies and gene expression studies on these isolates would provide further insights into the host adaptation of *S. aureus*.

In conclusion, the results of this analysis demonstrate a host-based segregation of the *S. aureus* population between cattle and humans in New Zealand. The segregation did not manifest broadly at the CC or ST/spa-type levels, but only after the implementation of genomic explorations of the core and accessory genomes. Host-specific/adaptive genes were also identified as reported in previous studies.

5

COMPARISON OF ANTIMICROBIAL SUSCEPTIBILITY TESTING BY DISK DIFFUSION AND RESISTANCE GENE ANALYSIS THROUGH WHOLE-GENOME SEQUENCING

5.1 INTRODUCTION

Antimicrobial susceptibility testing (AST) is essential in identifying trends in the resistance patterns of bacteria against clinically important antimicrobials and, at the individual animal or farm level, guiding the choice of antimicrobial therapy (Toutain et al., 2017).

A range of phenotypic laboratory tests is available for antimicrobial susceptibility testing, including disk diffusion, dilution, and Epsiolomer tests (E-test). The disk diffusion test is a standardised qualitative test that involves placing antimicrobial-impregnated disks on appropriate culture media such as Muller-Hinton agar (MHA), then inoculated with a suspension of the organism of interest. Following overnight incubation, the diameter of the zone of growth inhibition (ZOI) is measured and checked against standard zone diameter interpretive criteria provided by international bodies such as the Clinical and Laboratory Standards Institute (CLSI) (CLSI, 2018a) or the European Committee on Antimicrobial Susceptibility Testing (EUCAST) (Matuschek et al., 2014). These interpretive criteria are used to establish if the organism can be considered 'susceptible' or 'resistant' to the said antimicrobial. Dilution tests (agar or broth dilution) are quantitative tests that measure the minimum concentration of the antimicrobial that inhibits the growth of the bacterial isolate of interest, known as minimum inhibitory concentration (MIC). Dilution tests involve the preparation of serial dilutions of an antimicrobial solution and the incorporation of the diluted drug in agarised or liquid media that are then inoculated with a standard number of bacterial colony-forming units (CFU). The minimum concentration of the antimicrobial at which there is no visible growth is recorded as the MIC.

The E-test represents an amalgamation of the disk diffusion and dilution methods. A plastic strip impregnated with a predefined and labelled gradient of the antimicrobial, which covers a continuous concentration range, is placed on an agar plate inoculated with a standard number of CFUs and incubated. The MIC is then extrapolated at the point where the ZOI meets the scaled trip.

Of these tests, disk diffusion is the most commonly used in veterinary diagnostic laboratories because of its simplicity and low cost per test, and the availability of standardised disks for most commonly used antimicrobials.

An emerging strategy that may become a valid alternative to phenotypic antimicrobial susceptibility testing, in particular for large scale surveys of antimicrobial resistance, is the use of genotyping to identify resistance genes or genetic determinants such as mutations. These methods can involve PCR amplification of target resistance loci in bacterial DNA or DNA hybridisation where the target gene is tagged with enzymes, chemiluminescent compounds or antigenic substrates. A more holistic method than this gene by gene approach is represented by whole-genome sequencing (WGS). This is a rapidly developing technology, with processing costs becoming increasingly affordable, and it is quite likely that benchtop sequencers could be in use in many clinical laboratories in the near future. Alongside the development of the technology, the use of whole-genome sequencing to identify antimicrobial resistance (AMR) genes has been made increasingly simple through the development of bioinformatics programs such as ARIBA (Hunt et al., 2017) and ARG-ANNOT (Gupta et al., 2014), that streamline the screening of whole genomes for the presence of multiple antimicrobial resistance genes.

A few studies have been published correlating the results of the phenotypic assessment of antimicrobial resistance of *S. aureus*, with the genotypic analysis for AMR genes obtained either through PCR or WGS. Four studies (Gordon et al., 2014; Bradley et al., 2015b; Aanensen et al., 2016; Mason et al., 2018) used a rules-based approach (see section 1.3.3.2.3) where reads or assembled genomes were searched against resistance gene databases (either standard or custom-built). These studies compared WGS-AMR to AST for multiple antimicrobials in multiple organisms, including *S. aureus*. Considering AST as the gold standard, the reported sensitivity and specificity of genotyping ranged from 95% to 100% and 88% to 100%, respectively, depending on the antimicrobial and the study.

Two additional studies (Alam et al., 2014; Davis et al., 2016) used statistical modelling or machine learning (see section 1.3.3.2.3) to account for the fact that interactions of genes are responsible for resistance ('genome-wide association' studies(GWAS)). The two studies only studied methicillin and vancomycin resistance in *S. aureus,* reporting accuracies of 99.5% and 81%, respectively. This modelling approach may enhance the predictive power or resistance compared with the rules-based WGS-AMR in the future, but the technique is still in its nascent stage. All the above studies were done on human *S. aureus* isolates. To the author's knowledge, there have been no comparative studies between WGS-AMR and AST test using

bovine *S.aureus* isolates.

The study presented in this Chapter aimed to compare the results of the disk diffusion tests with WGS genotyping using 174 bovine *S. aureus* isolates, to assess whether the presence or absence of resistance genes could predict the disk diffusion test results with sufficient accuracy to potentially replace it in routine clinical or surveillance testing in the future.

5.2 MATERIALS AND METHODS

5.2.1 ISOLATES

This study tested one hundred and seventy-four isolates out of the 188 of bovine *S. aureus* from Chapter 3. In addition, *S. aureus* strain American Type Culture Collection (ATCC) 25923 was used as a quality control strain. Thirteen isolates obtained from veterinary laboratories in New Zealand during 2017-18 were returned to the laboratory after their DNA was extracted for WGS, and were not available to be used for this study. One frozen isolate from 2002-03 did not grow at resuscitation and could not be used. Complete details of the samples and further details on the sampling process can be found in Chapter 2. In summary, 123 isolates were obtained from New Zealand clinical veterinary laboratories during 2002-03, and 2012-13. Fifty-one isolates were collected during a clinical trial of mastitis therapeutics performed on 11 dairy farms located in the central North Island during 2017-18. Except for the isolates from 2002-03 collection, other collection years had several sets of multiple isolates obtained from the same farm, as reported in General Materials and Methods, and many isolates from the same farm also belonged to the same clonal complex (CC) (see Chapter 3).

5.2.2 DISK DIFFUSION ANTIMICROBIAL SUSCEPTIBILITY TESTING

The disk diffusion methodology used in the study, including the use of quality control strains, followed the CLSI standard VET01 (CLSI, 2018).

5.2.2.1 INOCULUM PREPARATION

The 175 frozen *S. aureus* isolates, including the control organism (ATCC 29523), were revived on 5% Columbia Sheep blood agar (Fort Richard Laboratories, Auckland, NZ), using the quadrant streak technique to achieve individual colonies. The plates were incubated aerobically overnight at 37°C, and an inoculum suspension was created by suspending two to three morphologically similar colonies in 3 ml sterile saline in a clear test tube. The suspension was mixed using a vortex machine (Chiltern Scientific, Auckland, NZ) to produce an evenly turbid suspension. This suspension was compared to the a 0.5 McFarland standard suspension (Remel, Thermo Fisher Scientific, Auckland, NZ) by eye using the provided visual comparison card, and adjusted to that turbidity, by adding bacteria or sterile saline solution as necessary.

5.2.2.2 PLATE INOCULATION

Within 15 minutes of the preparation of the inoculum suspension, a sterile cotton swab was dipped into the suspension and pressed against the inner walls of the test tubes to remove the excess liquid. The swab was then spread on MHA by swabbing across the whole surface of the plate three times by rotating the plate. The inner edge of the plate alongside the agar was finally swabbed. The MHA plate was left on the bench partly open for approximately 3 minutes to evaporate the excess moisture.

5.2.2.3 ANTIMICROBIAL DISK APPLICATION

The antimicrobial disks were placed on the inoculated MHA plates using a disk dispenser (Oxoid, Auckland, NZ). Three antimicrobial disks were used per plate to avoid overlapping of the ZOIs. Fifteen antimicrobial disks were used in this study (Table 5.1). The antimicrobials were selected based on their use in dairy cattle in New Zealand. The antimicrobial contents of the disks were those recommended by the CLSI supplement VET08 (CLSI, 2018a). The disks were gently pressed onto the agar using flame-sterilised steel forceps. The MHA plates were inverted and incubated aerobically at 35-37°C for 16-20 hours within 15 minutes of applying the disks.

Table 5.1: Antimicrobial disks and the disk content used for the disk diffusion test. The available interpretive criteria for both resistance and susceptibility with the zone of inhibition diameter (mm) along with the reference are provided for each antimicrobial disk used. The zone of inhibition diameter for intermediate susceptibility is provided in brackets along with the susceptibility criteria. All the disks were sourced from Thermo Fisher Scientific (Auckland, NZ)

Antimicrobial disk	Disk content	Interpretive criteria		Reference
		Resistant (mm)	Susceptible (mm)	
Penicillin G	6 μg[a]	≤ 28	≥ 29	CLSI (2018a)
Amoxicillin	25 μg	≤ 28	≥ 29	Oxoid (2011)
Amox-clav	30 μg[b]	≤ 19	≥ 20	Oxoid (2011)
Lincomycin	15μg	NA	NA	NA
Neomycin	30 μg	≤ 13	≥ 16 (14-15)	Oxoid (2011)
Cloxacillin	5 μg	NA	NA	NA
Streptomycin	10 μg	NA	NA	NA
Tetracycline	30 μg	≤ 17	≥ 23 (18-22)	CLSI (2018a)
Enrofloxacin	5 μg	≤ 16	≥ 23 (17-22)	CLSI (2018a)
TM/S	25 μg[c]	≤ 10	≥ 16 (11-15)	CLSI (2018a)
Ceftiofur	30 μg	≤ 17	≥ 21 (18-20)	CLSI (2018a)
Cefoxitin	30 μg	≤ 21	≥ 22	CLSI (2018a)
Tylosin	30 μg	NA	NA	NA
Oleandomycin	15 μg	NA	NA	NA
Erythromycin	15 μg	≤ 13	≥ 23 (14-22)	CLSI (2018a)

Amox-clav: Amoxicillin-clavulanic acid, TM/S: Trimethoprim/Sulphamethoxazole, [a]: contains 10 Units, [b]: contains 20 μg of Amoxicillin and 10 μg of clavulanic acid, [c]: contains 1.25 μg of trimethoprim and 23.75 μg of sulphamethoxazole, NA:Not available

5.2.2.4 DISK PLATE READING FOR ZOI MEASUREMENT

The MHA plates were evaluated for growth of the organism and ZOI diameters after 16-20 hours of incubation. The zones of inhibition were read from the back of the MHA plate as per the standard consulted, against a dark background using reflected light. The zones of inhibition were measured by eye using a ruler with one millimetre being the smallest graduation. The readings were recorded to the nearest millimetre.

5.2.3 IDENTIFICATION OF RESISTANCE GENES

In order to identify antimicrobial resistance genes in the whole genomes, paired whole-genome sequence reads of the 174 *S. aureus* isolates were created using the Hi-Seq X (150PE) (Illumina) paired sequencing platform. More details of this process are available in section 2.1.3. These reads were then processed through the bioinformatics tool ARIBA (Antimicrobial Resistance Identification by Assembly) (Hunt et al., 2017). ARIBA identifies resistance genes by mapping the assembly of paired sequencing reads to the reference resistance genes recorded on publicly available AMR gene databases. The AMR databases that were used with ARIBA are listed in Table 5.2. ARIBA uses a fermi-lite assembler `https://github.com/lh3/fermi-lite` as the default assembler. An in-house script `https://github.com/jabinnes/Thesis-files/blob/main/Scripts/VAR_ARIBA.ipynb` was used to automate the ARIBA process to download reference AMR genes from the databases, run a local assembly on the reads and map them to identify the presence of AMR genes in the genome sequences. The ARIBA output is provided as a report table for each individual isolate, then compiled using the 'summary' task of ARIBA.

5.2.4 DATA ANALYSIS

The diameters of the ZOI were entered into a Microsoft Excel spreadsheet and were compared with the interpretive criteria in the CLSI standards and classified into Resistant, Susceptible, or Intermediate for the drugs mentioned in the standard VET08 (CLSI, 2018a). If a CLSI interpretive criteria was not found for an

Table 5.2: List of reference databases used by ARIBA to identify the AMR genes of the *S. aureus* paired sequencing reads.

Database	Reference
The Comprehensive antimicrobial Resistance Database (CARD)	Alcock et al. (2019)
Antimicrobial Resistance Gene-ANNOTation (ARG-ANNOT)	Gupta et al. (2014)
National Centre for Biotechnology Information (NCBI) Reference Catalog	https://www.ncbi.nlm.nih.gov/pathogens/isolates$\sharp$/refgene/
Resfinder	Zankari et al. (2012)

antimicrobial, the interpretive criteria document provided by the disk manufacturer, Oxoid, was utilised for classification (Oxoid, 2011). All the results categorised as intermediate were considered resistant in this study, for ease of calculations. If interpretive criteria for a given drug were not reported in the either CLSI standard or Oxoid document, a cut-off point was determined arbitrarily, corresponding to the smallest diameter between the two distributions in the bimodal kernel density estimation curve, if the histogram of the ZOI distribution showed a bimodal distribution. If a bimodal distribution was not observed, no cut-off points were determined. The frequency distribution graph with kernel density estimation curve was plotted using the statistical visualisation package Seaborn (Waskom et al., 2017) with an in-house script https://github.com/jabinnes/Thesis-files/blob/main/Scripts/VAR_ARIBA.ipynb for each antimicrobial used in the disk diffusion test. The sensitivity and specificity of the WGS result (i.e., presence of the respective genetic determinant of resistance) and their positive and negative predictive values (along with 95% confidence intervals (CI)) were calculated considering the disk diffusion test result as the reference standard. The Relative Risk (RR) of a CC97 isolate having the *blaZ* gene compared to a CC1 isolate was then calculated.

The distributions of ZOI of CC1 and CC97 isolates containing *blaZ* were compared using the Mann Whitney U (MWU) test with its P value. The distributions of ZOI of CC1 and CC97 for other resistance genes were not compared because of the very few isolates containing those resistance determinants.

The means of ZOI of CC1 and CC97 isolates were compared using an independent sample t-test as follows:

1. For β-lactam antimicrobials, means were compared between CC1 and CC97 isolates with the *blaZ* gene and between CC1 and CC97 isolates that lacked the *blaZ* gene.

2. For other antimicrobials (i.e., bell-shaped curve), comparison of means was done between all phenotypically susceptible CC1 and CC97 isolates as only a very few isolates were phenotypically resistant to non β-lactam antimicrobials.

The independent sample t-test results were provided with mean difference along with the 95% CI and the p-value of 2-tailed test.

Levene's test was used to determine the equality of variances. If p was ≤ 0.05, results of t-test with unequal variances were reported, if >0.05, then results of t-test with unequal variances were reported.

The specificity, sensitivity and relative risk calculations were undertaken using the online MedCalc software (https://www.medcalc.org/calc/relative_risk.php), accessed on August 2020). All other analyses used SPSS 27 (IBM, Seattle, USA).

The rates of phenotypic resistance observed in this study were compared with the results of the studies on antimicrobial resistance in *S. aureus* from bovine mastitis in New Zealand done by Petrovski et al. (2011, 2015).

5.3 RESULTS

The control organism *S. aureus* American Type Culture Collection (ATCC) 25923 had a ZOI within the CLSI standards for all the antimicrobials used in the study. The saline used in the study was incubated in an MHA plate at 37°C, and no bacterial growth was noticed after 24 hours.

5.3.1 β-LACTAMS

None of the isolates in this study had the *mec* gene, and hence all were categorised genotypically as methicillin-susceptible *S. aureus* (MSSA).

5.3.1.1 PENICILLIN G

A total of 173/174 isolates were successfully tested for resistance to Penicillin G using the disk diffusion test. The result could not be recorded for one isolate for which the disk dropped from the agar during incubation. The ZOI ranged from 10 to 50 mm. According to the CLSI veterinary standard, *S. aureus* isolates with ZOI $\geq$ 29 mm are considered susceptible, and isolates with ZOI <29 mm are resistant, with no intermediate category. Based on these criteria, 59/173 (34.1%) isolates were classified as resistant and 114/173 (65.9%) as susceptible.

As expected, the frequency distributions of the ZOI for penicillin and the other β-lactamase-sensitive β-lactams were bimodal (Figure 5.1). The nearest millimetre of the cut-off between penicillin-resistant and wild (susceptible) populations (using the kernel density estimation curve) was 29 mm, matching the CLSI cut-off. To remove any autocorrelation in the data, the cut-off of the kernel density function was re-calculated including only the first listed isolate for each farm-CC combination (for the farm-CCs combinations see Chapter 3). This cut-off was 29 mm, the same as the above-estimated cut-off with all isolates.

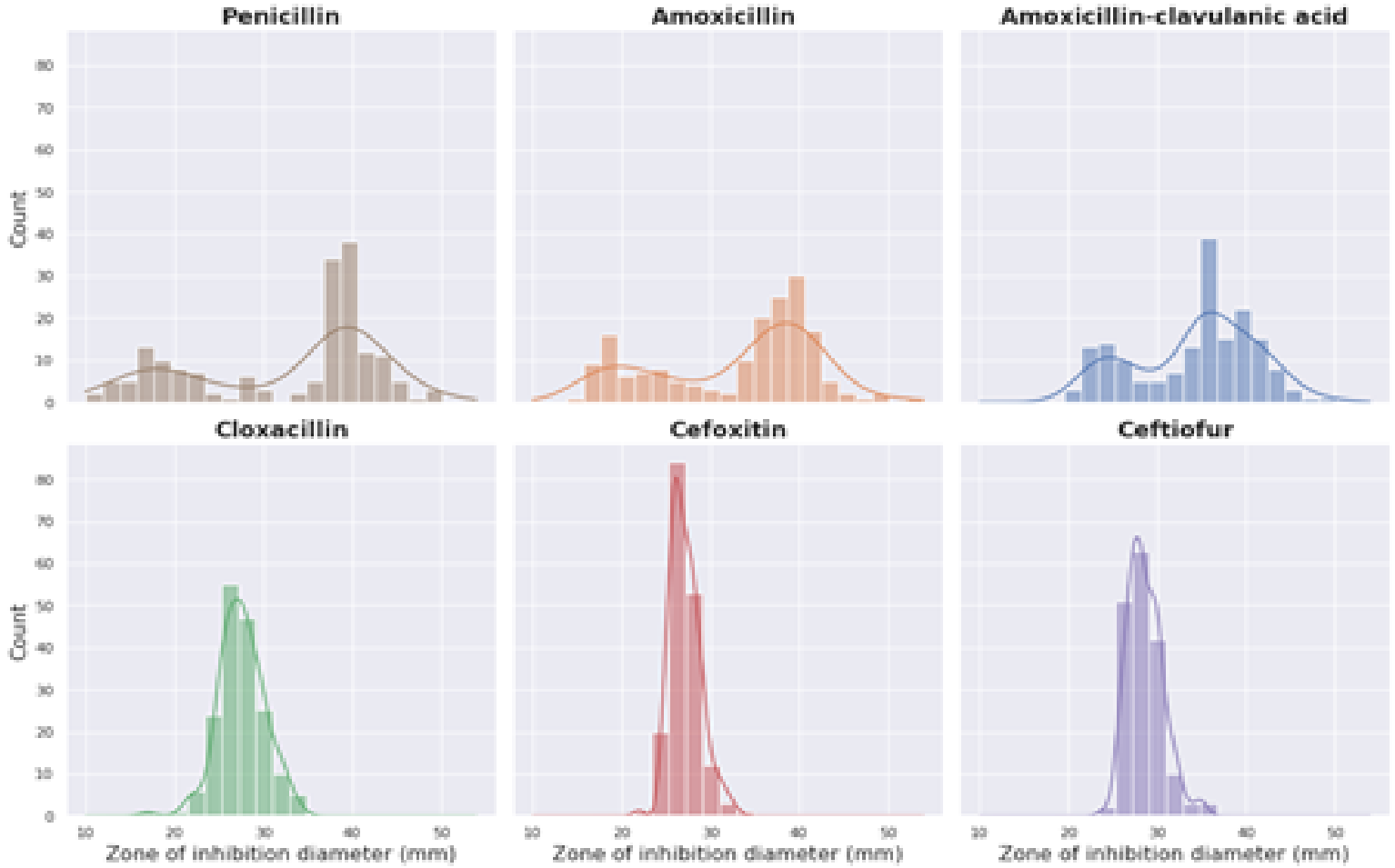

Figure 5.1: Frequency distribution histogram of the disk diffusion ZOI diameters (in mm) for the β-lactam antimicrobials. The kernel density estimation curves are laid over the histograms. Note the bimodal distribution of penicillin, amoxicillin, and amoxicillin-clavulanic acid. Cloxacillin, cefoxitin and ceftiofur are β-lactamase resistant, so the curves are not bimodal

Of the 173 genomes from isolates that had a disk diffusion result for penicillin, 42 (24.3%) were found to have the *blaZ* gene when checked against the Resfinder database. The NCBI database, CARD and Argannot found this gene in 40, 40, and 35 isolates, respectively. Only one *blaZ* gene negative isolate in Resfinder was positive in all the other databases. As it found more *blaZ* genes, the results from Resfinder were used for all comparisons of of WGS and disk diffusion results where *blaZ* gene status was included in the comparison. All 42 isolates identified by Resfinder as harbouring the *blaZ* gene were phenotypically resistant to penicillin G, with ZOI ranging between 10 and 23 mm (mean=16.8 mm). However, 17 isolates that did not harbour the *blaZ* gene were phenotypically resistant, with ZOI ranging from 20 to 28 mm (mean=24.1 mm). Interestingly, isolates that were phenotypically resistant but did not harbour the *blaZ* gene had larger ZOI than those which harboured the gene (mean difference =7.3 mm; 95% CI 5.5 to 9.1; P<0.0001).

The sensitivity and specificity of WGS-AMR by Resfinder for identifying phenotypically resistant isolates based on the CLSI criteria were 0.71 (95% CI 0.58 to 0.82) and 1 (95% 0.97 to 1.0), respectively. For the test population (where the prevalence of phenotypic resistance was 34%), the negative predictive value (i.e., the probability that a *blaZ* gene-negative isolate was phenotypically susceptible) was 0.87, and the positive predictive value (i.e. the probability that *a blaZ* gene-positive isolate was phenotypically resistant) was 1.

The predominant CC in this study was CC1, and the second most common was CC97 (see Chapter 3). Compared to the CC1 isolates, CC97 isolates were 7.7 (95% CI 4.8 to 12.4) times more likely to have the *blaZ* gene and 4.3 (95% CI 3.1 to 6.0) times more likely to be phenotypically resistant to penicillin. The MWU test showed that there was no statistical difference between the distributions of the ZOI for penicillin of CC1 and CC97 isolates containing the *blaZ* gene (P = 0.781) (Figure 5.2). There was no significant difference between the means of the ZOI of CC1 and CC97 isolates harbouring the *blaZ* gene for penicillin (P=0.96) while there was significant difference noticed between the means of ZOI of CC1 and CC97 isolates lacking the *blaZ* gene for penicillin (P=0.02), see Table 5.3.

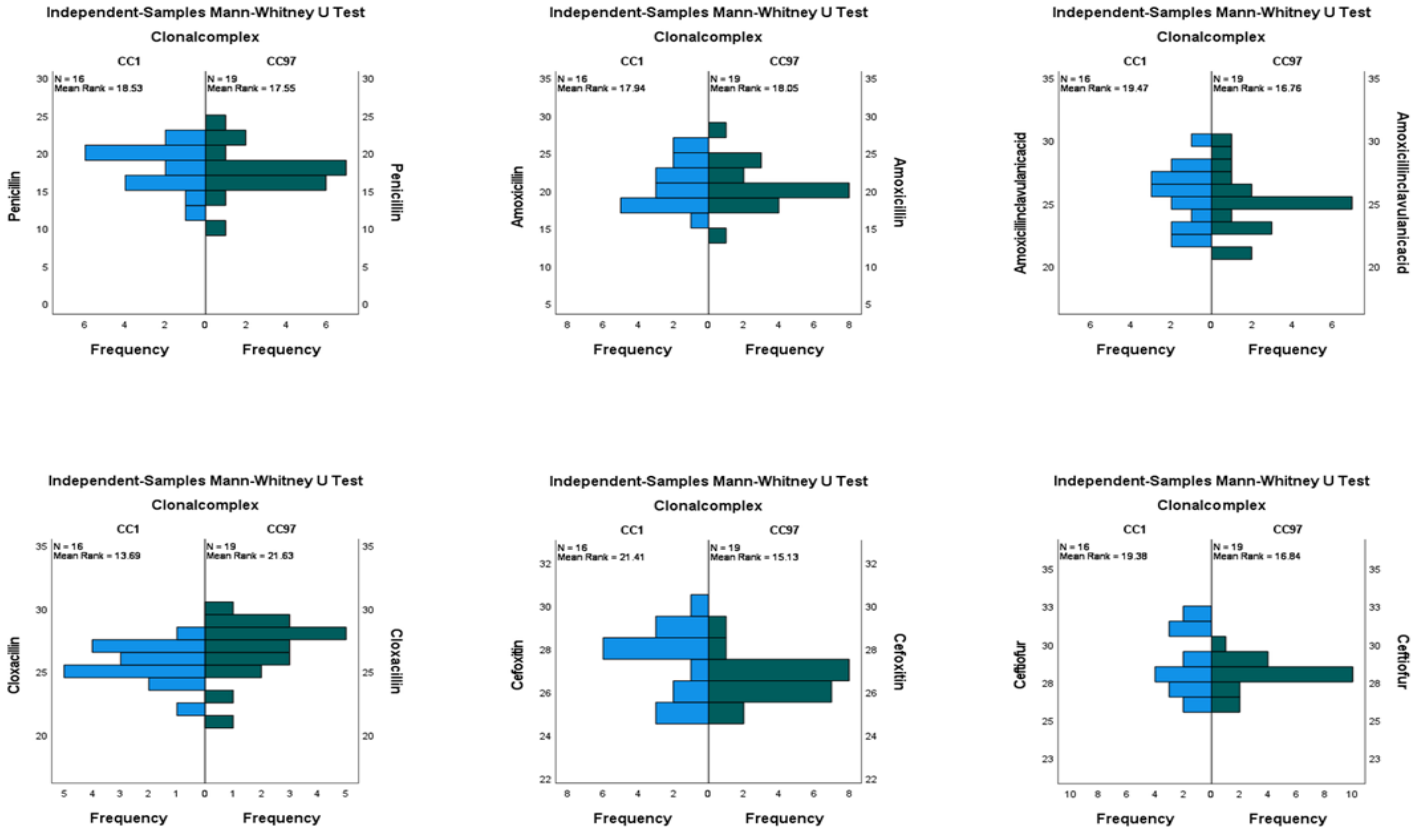

Figure 5.2: Distribution histograms of the Mann-Whitney U test for the zone of inhibition diameters of isolates belonging to CC1 and CC97 for the β-lactams antimicrobials with the mean rank for each distribution. (From left to right: Top row: Penicillin, Amoxicillin and Amoxicillin-clavulanic acid Bottom row: Cloxacillin, Cefoxitin and Ceftiofur).

Table 5.3: Mean and range of zone of inhibition (ZOI) diameters of the β-lactam antimicrobials used in this study against the *S. aureus* isolates from bovine mastitis, classified by clonal complex and the presence or absence of the *blaZ* gene.

Antimicrobials	Resistance gene	Clonal complex	No of isolates	Range of ZOI diameters (mm)	Mean ZOI (95% CI) (mm)	Mean difference (95% CI)	p-value*
Penicillin	*blaZ* +	CC1	16	11 to 21	17.3 (15.8 to 18.9)	0.05 (-2.00 to 2.10)	0.96
		CC97	19	10 to 23	17.3 (15.8 to 18.8)		
	blaZ -	CC1	120	20 to 50	37.9 (36.8 to 39.0)	9.90 (1.40 to 18.40)	0.02
		CC97	2 [a]	20 and 36	28.0 (12.3 to 43.7)		
Amoxicillin	*blaZ* +	CC1	16	16 to 26	20.3 (18.7 to 22.0)	0.15 (-1.90 to 2.30)	0.89
		CC97	19	14 to 28	20.2 (18.6 to 21.7)		
	blaZ -	CC1	120	22 to 54	37.3 (36.4 to 38.3)	5.00 (-1.30 to 11.30)	0.12
		CC97	3	22 to 40	32.3 (9.3 to 55.4)		
Amox-clav	*blaZ* +	CC1	16	22 to 30	25.6 (24.3 to 26.8)	0.50 (-1.10 to 2.10)	0.52
		CC97	19	21 to 30	25.1 (24.0 to 26.2)		
	blaZ -	CC1	120	24 to 50	37.4 (36.6 to 38.1)	4.70 (-16.80 to 26.20)	0.45 [b]
		CC97	3	23 to 40	32.7 (11.0 to 54.4)		
Cloxacillin	*blaZ* +	CC1	16	22 to 28	25.6 (24.8 to 26.4)	0.65 (-2.60 to 0.05)	0.06
		CC97	19	21 to 30	26.8 (25.8 to 28.0)		
	blaZ -	CC1	120	17 to 34	28.0 (27.6 to 28.4)	-1.30 (-4.20 to 1.50)	0.40
		CC97	3	27 to 33	29.3 (21.4 to 37.3)		
Cefoxitin	*blaZ* +	CC1	16	25 to 30	27.4 (26.6 to 28.3)	0.86 (-0.80 to 1.80)	0.07 [b]
		CC97	19	25 to 29	26.6 (26.1 to 27.0)		
	blaZ -	CC1	120	22 to 33	27.3 (27.0 to 27.7)	-0.70 (-2.80 to 1.50)	0.50
		CC97	3	26 to 31	28.0 (21.4 to 34.6)		
Ceftiofur	*blaZ* +	CC1	16	26 to 32	28.8 (27.7 to 29.8)	0.75 (-0.40 to 1.90)	0.20 [b]
		CC97	19	26 to 30	28.0 (27.5 to 28.5)		
	blaZ -	CC1	120	24 to 34	28.7 (28.3 to 29.0)	-0.30 (-13.20 to 12.50)	0.90 [b]
		CC97	3	26 to 35	29.0 (16.1 to 42.0)		

ZOI: zone of inhibition, *blaZ* +: Isolates that harbour *blaZ* gene, *blaZ* -: Isolates that lack *blaZ* gene, Amox-clav: Amoxicillin-clavulanic acid, [a]: One isolate did not produce a result, [*] Equal variances assumed except for p-values with [b] (Levene's test p <0.05), [†]:CC1 and CC97 were the two predominant clonal complexes in the dataset.

5.3.1.2 AMOXICILLIN

Amoxicillin does not have a CLSI or EUCAST interpretive criterion for disk diffusion test for veterinary medicine, and these standards do not report human interpretive criteria. According to the disk manufacturer (Oxoid, 2011), ampicillin can be used as a class representative for ampicillin and amoxicillin in humans. According to the same document, *Staphylococcus* spp. with a ZOI for ampicillin (10μg) < 29 mm are considered resistant. The cut-off point of the kernel density estimation curve of amoxicillin was also 29 mm (Figure 5.1).

Of the 174 isolates, 55 were found to be phenotypically resistant to amoxicillin based on the above criterion, including all 42 isolates identified by Resfinder as possessing the *blaZ* gene. These 42 isolates had ZOI ranging from 14 to 28 mm, with a mean of 19.8 mm. Thirteen isolates that did not contain the *blaZ* gene were phenotypically resistant, with ZOI ranging from 16 to 28 mm (with a mean of 24.2 mm). Also for amoxicillin, isolates that were phenotypically resistant but not genotypically resistant had larger ZOI than those which were also genotypically resistant (mean difference 4.4 mm; 95% CI 2.47 to 6.34; P<0.0001).

The sensitivity and specificity of WGS-AMR for identifying phenotypic resistance to amoxicillin were 0.76 (95% CI 0.36 to 0.87) and 1 (95% CI 0.97 to 1.0), respectively. For the test population (where the prevalence of phenotypic resistance was 32%), the negative predictive value was 0.90, and the positive predictive value was 1.

Compared to the CC1 isolates, CC97 isolates were 4.8 (95% CI 3.3 to 6.9) times more likely to be phenotypically resistant to amoxicillin. The MWU test showed that the distributions of the ZOI of CC1 and CC97 isolates containing the *blaZ* gene for amoxicillin did not differ statistically (Figure 5.2) with P = 1.0. The difference between means of the ZOI of CC1 and CC97 isolates harbouring the *blaZ* gene for amoxicillin was statistically insignificant (P=0.89).The difference between means of the ZOI of CC1 and CC97 isolates lacking the *blaZ* gene for amoxicillin was statistically insignificant (P=0.12). see Table 5.3

5.3.1.3 AMOXICILLIN-CLAVULANIC ACID

Amoxicillin-clavulanic acid does not have a disk diffusion interpretive criterion for *S. aureus* in either CLSI or EUCAST. However, interpretive criterion for amoxicillin-clavulanic acid disk diffusion for *Staphylococcus* species provided by Oxoid classified the isolates with ZOI diameter $\leq$ 19mm as resistant. With this criterion, all the isolates in the study were considered phenotypically susceptible. This was consistent with the finding that none of the isolates harboured *mec* genes. Interestingly, the frequency distribution graph for amoxicillin-clavulanic acid was bimodal, with a cut-off of 29 mm (Figure 5.1).

The ZOI of the isolates that contained the *blaZ* gene ranged from 21 to 30 mm (with a mean of 24.9 mm), while the ZOI of the isolates that lacked the *blaZ* gene ranged from 22 to 50 mm (with a mean of 37.3 mm). The difference in the means of the ZOI of isolates with and without *blaZ* was 12.5 mm (95% CI 11.4 to 13.5, P < 0.001; equal variances assumed).

The MWU test showed that the distributions of the ZOI of CC1 and CC97 isolates containing the *blaZ* gene for amoxicillin-clavulanic acid were statistically same (Figure 5.2) with P=0.441. There was no statistical significance in the mean difference of the ZOI of CC1 and CC97 isolates harbouring the *blaZ* gene (P=0.52) and lacking the *blaZ* gene (P=0.45) for amoxicillin-clavulanic acid (see Table 5.3).

5.3.1.4 CLOXACILLIN

Cloxacillin does not have human or veterinary disk diffusion interpretive criteria. The frequency distribution graph had an apparent normal distribution (a bell-shaped curve) (Figure 5.1), and there was no evidence of distinct wild type and resistant populations, consistent both with β-lactamase resistant β-lactam and the lack of *mec* genes in the sample. The ZOI ranged from 17 to 34 mm, with a mean of 27.6 mm. The ZOI of the isolates that contained the *blaZ* gene ranged from 21 to 30 mm (with a mean of 25.9 mm), and that of the isolates that lacked the *blaZ* gene ranged from 17 to 34 mm (with a mean of 28.2 mm). The difference in the means of the ZOI diameters of isolates with and without *blaZ* was 2.28 mm (95% CI 1.43 to 3.13, P < 0.05; equal variances assumed).

The MWU test showed that the distributions of the ZOI of CC1 and CC97 isolates containing the *blaZ* gene for cloxacillin differed statistically (Figure 5.2), with P = 0.02. There was no statistical significance in the difference between the means of the ZOI of CC1 and CC97 isolates harbouring the *blaZ* gene (P=0.06) and those lacking the *blaZ* gene (P=0.40) for cloxacillin (see Table 5.3).

5.3.1.5 CEFOXITIN

The ZOI for cefoxitin in this study ranged from 22 to 33 mm, so all isolates were classified as susceptible according to the CLSI veterinary standard interpretive criterion for *S. aureus* ($\geq$ 22mm). The kernel density estimation curve on the frequency distribution graph was bell-shaped (Figure5.1). The ZOI of 42 isolates that contained the *blaZ* gene ranged from 25 to 30 mm (with a mean of 26.9 mm), while the 132 isolates that lacked the *blaZ* gene had a ZOI diameter of 22 to 33 mm (with a mean of 27.4 mm). The difference in the means of these two groups was statistically significant with mean difference of 0.50 mm (95% CI -0.01 to 1.01, P=0.05).

The MWU test showed that the distributions of the ZOI of CC1 and CC97 isolates containing the *blaZ* gene for cefoxitin did not differ significantly (Figure 5.2) with P = 0.71. There were no statistical significance in the mean difference of the ZOI between CC1 and CC97 isolates harbouring the *blaZ* gene (P=0.07) and those lacking the *blaZ* gene (P=0.50) for cefoxitin (see Table 5.3). The susceptibility of all isolates to cefoxitin is consistent with none of the isolates in this study having the *mec* gene.

All the *S. aureus* isolates used in this study were susceptible to the third-generation cephalosporin, ceftiofur, with ZOI ranging from 24 to 35 mm (CLSI susceptibility criteria of $\geq$ 21 mm). The kernel density estimation graph was bell-shaped (Figure 5.1). The ZOI of 42 isolates that contained the *blaZ* gene ranged from 26 to 32 mm (with a mean of 28.3 mm), while the 132 isolates that lacked the *blaZ* gene had a ZOI diameter of 24 to 34 mm (with a mean of 28.8 mm). The difference in the means of these two groups was 0.53 mm (95% CI -0.06 to 1.11, P = 0.07; equal variances assumed).

The MWU test showed that the distributions of the ZOI of CC1 and CC97 isolates containing the *blaZ* gene for ceftiofur were not statistically different (Figure 5.2) with P = 0.481. There was no statistical significance in the mean difference of the ZOI

between CC1 and CC97 isolates harbouring the *blaZ* gene (P=0.20) and those lacking the *blaZ* gene (P=0.90) for ceftiofur (see Table 5.3).

5.3.2 MACROLIDES AND LINCOSAMIDES

The frequency distributions of the ZOI for these antimicrobials are presented in Figure 5.3. Of these antimicrobials, only erythromycin had an interpretive criterion in the CLSI standard for *S. aureus,* with ZOI $\geq$ 23 mm considered susceptible, 14 to 22 mm considered intermediate, and <22 mm, resistant. For this analysis, the intermediate isolates were considered resistant. According to these criteria, 5/174 (2.9%) isolates were phenotypically resistant to erythromycin and 169/174 (97.1%) were phenotypically susceptible. Two resistant isolates (from separate farms), had no visible growth around the antimicrobial disks (ZOI 0 mm). Both of these isolates had the *ermC* gene (found in all databases), and both belonged to CC1.

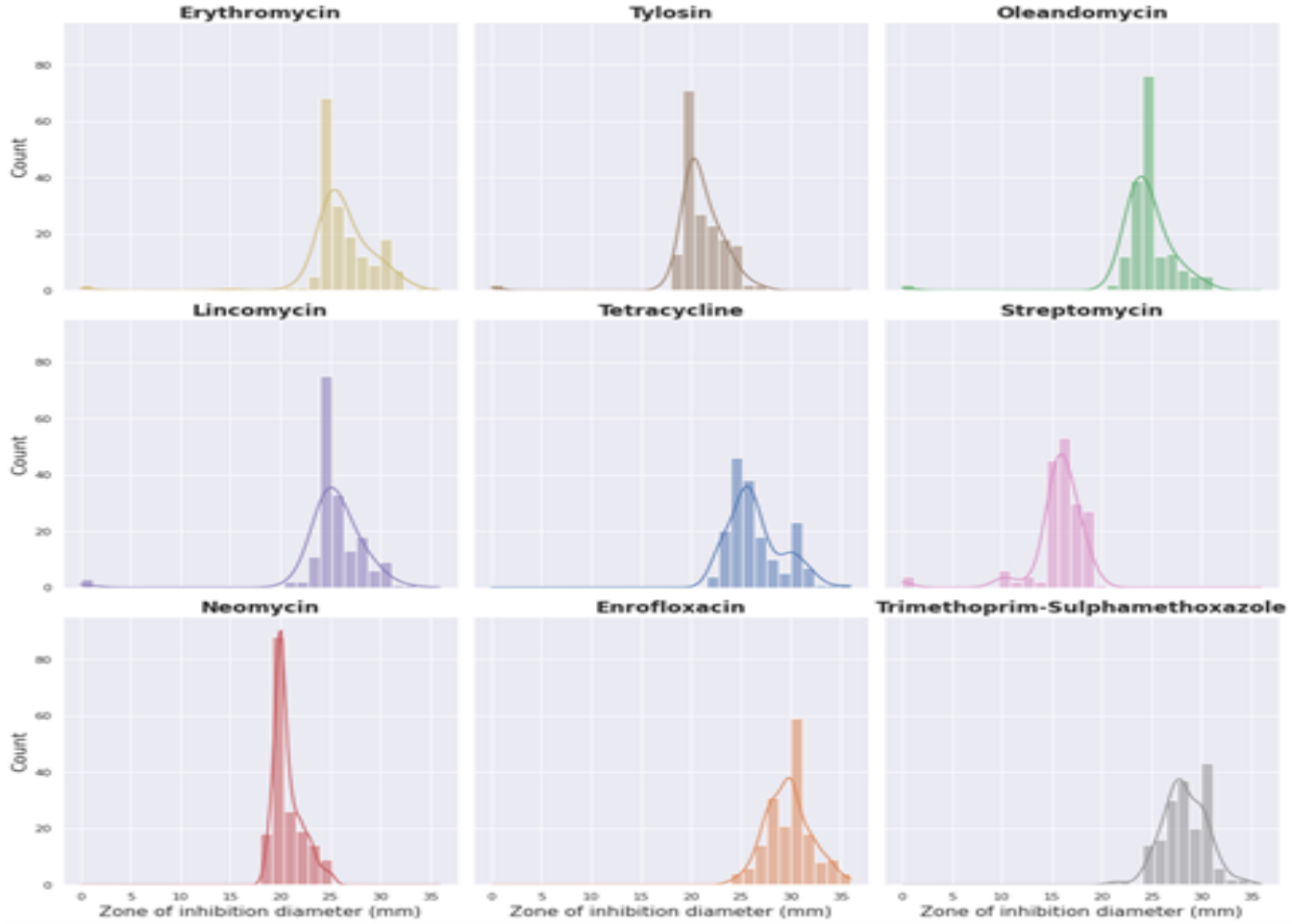

Figure 5.3: Frequency distribution histogram of the disk diffusion ZOI diameters (in mm) for the antimicrobials (erythromycin, tylosin, oleandomycin, lincomycin, tetracycline, streptomycin, neomycin, and trimethoprim-sulphamethoxazole). The kernel density estimation curves are laid over the histograms.

Table 5.4: Mean and range of ZOI diameters for non-β-lactam antimicrobials of susceptible S. aureus isolates from bovine mastitis divided by clonal complex.

Antimicrobials	Clonal complex	No of isolates	Range (mm)	Mean (95%CI) (mm)	Mean difference (95%CI) (mm)	P-Value*
Erythromycin	CC1	132	23 to 32	26.5 (26.1 to 26.9)	-0.49 (-1.52 to 0.55)	0.355
	CC97	22	24 to 34	26.9 (25.8 to 28.1)		
Tylosin	CC1	134	18 to 25	20.9 (20.7 to 21.2)	-1.26 (-2.01 to -0.50)	0.001
	CC97	22	20 to 27	22.2 (21.2 to 23.2)		
Oleandomycin	CC1	133[a]	21 to 31	24.6 (24.3 to 24.9)	-0.53 (-1.41 to 0.36)	0.242
	CC97	22	21 to 30	25.1 (24.0 to 26.2)		
Lincomycin	CC1	133	21 to 32	25.6 (25.2 to 25.9)	-0.99 (-1.93 to -0.05)	0.040
	CC97	22	21 to 33	26.6 (25.4 to 27.7)		
Tetracycline	CC1	132	23 to 32	26.4 (25.9 to 26.9)	-0.58 (-1.78 to 0.62)	0.339
	CC97	22	23 to 36	27 (25.7 to 28.3)		
Streptomycin	CC1	132	12 to 20	16.2 (15.9 to 16.4)	-0.25 (-1.01 to 0.49)	0.505
	CC97	14	14 to 19	16.4 (15.6 to 17.2)		
Neomycin	CC1	136	19 to 25	20.7 (20.5 to 20.9)	0.02 (-0.59 to 0.64)	0.939
	CC97	22	19 to 25	20.7 (20.0 to 21.3)		
Enrofloxacin	CC1	136	24 to 36	29.7 (29.3 to 30.1)	-0.99 (-2.00 to 0.004)	0.051
	CC97	22	28 to 36	30.7 (29.8 to 31.5)		
TMS	CC1	136	21 to 34	28.0 (27.6 to 28.4)	-1.24 (-2.20 to -0.27)	0.012
	CC97	21[a]	24 to 34	29.2 (28.2 to 30.3)		

[a] – One isolate was not recorded as the disk fell off the agar during incubation; * Calculated assuming equal variances; TMS-Trimethoprim-Sulphamethaxole

None of the other macrolides or lincosamides have standard disk diffusion veterinary interpretive criteria. For tylosin and oleandomycin, the two isolates containing the ermC gene had a ZOI of 0 mm; for lincomycin, three isolates had a ZOI of 0. This included the two isolates with the *ermC* gene. The kernel density estimation curves for tylosin, oleandomycin and lincomycin were all bell-shaped except for isolates with 0 mm ZOI. (see Figure 5.3). The mean and range of the ZOI (excluding those with a ZOI of 0 mm) for these three antimicrobials are presented in Table 5.4. In all three cases, the ZOI of susceptible CC97 isolates were larger than those of susceptible CC1 isolates. However, except for tylosin (P=0.001) and lincomycin (P=0.040), there was no statistical significance in the mean difference between ZOI of susceptible CC1 and CC97 isolates for the other macrolides (erythromycin: P=0.355 and oleandomycin: P=0.242) (see Table 5.4).

5.3.3 TETRACYCLINES

The CLSI standard defines *S. aureus* with ZOI $\geq$ 23 mm as susceptible, 18 to 22 mm as intermediate, and <22 mm as resistant to tetracycline. For this analysis, the intermediate category was considered resistant. Of the 174 *S. aureus* isolates, 4 (from different farms) were classified as phenotypically resistant (all had a ZOI of 22 mm) and the rest as phenotypically susceptible, with diameters ranging from 23 to 36 mm (mean of 26.6 mm). All 174 isolates were identified as containing the chromosomally encoded *tet38* resistance gene by all reference databases, except for Resfinder which did not identify this gene in any isolates. All the phenotypically resistant isolates belonged to CC1. There was no statistical significance in the mean difference between ZOI of susceptible CC1 and CC97 isolates (P=0.339) for tetracycline (Table 5.4).

5.3.4 AMINOGLYCOSIDES

Neither neomycin nor streptomycin have standard veterinary interpretive criteria. The kernel density estimation curve and the AST frequency histogram for streptomycin were bimodal, with a kernel density estimation cut-off point of 11 mm. Using this cut-off point to define susceptibility, 10/174 (5.74%) isolates were phenotypically resistant, while 164/174 were phenotypically susceptible. Of the resistant isolates

4/10 belonged to CC1 from three different farms and 6/10 to CC97 from one farm. CC97 isolates were 9.3 (95% CI 2.8 to 30.2) times more likely to be streptomycin-resistant than CC1 isolates.

All 174 isolates contained two known generic aminoglycoside resistance genes: *Aph* and *aac3*-Ik. These genes were identified across all the databases except for Resfinder, which did not find these two genes in any of the isolates.

The four resistant CC1 isolates all had ZOI of 0 mm. Three of these isolates contained the *str* gene (all databases), which codes for an enzyme that modifies streptomycin. Isolates with the *str* gene were 24.4 (95% CI 11.8 to 50.5) times more likely to be phenotypically resistant than str-negative isolates. Mean ZOI for isolates that were phenotypically susceptible to streptomycin were larger for CC97 than CC1, but there was no statistical significance in the mean difference between ZOI of susceptible CC1 and CC97 isolates (P=0.505) for streptomycin (Table 5.4).

Based on criteria established by the disk manufacturer (Oxoid 2013) for *Staphylococcus* spp., a ZOI $\geq$ 16 mm was classified as susceptible to neomycin. According to this criterion, all isolates in this study were phenotypically susceptible, with ZOI ranging from 19 to 25 mm (mean 20.7 mm). No neomycin-specific resistance genes were identified in the isolates. There was no statistical significance in the mean difference between ZOI of CC1 and CC97 isolates (P=0.939) for neomycin (see Table 5.4).

5.3.5 FLUOROQUINOLONES

The interpretive criteria for enrofloxacin according to the CLSI standard classifies *S. aureus* with a ZOI $\geq$ 23mm as susceptible. The ZOI of the isolates in the study ranged from 24 to 36 mm, categorising all as susceptible to enrofloxacin. All except four isolates had a chromosomally encoded *norA* gene, found in databases ARG-ANNOT and CARD. Three of the four isolates that lacked *norA* were from CC151, while one was from CC133 and all these isolates were from different farms belonging to the 2002-03 year of collection.

The ZOI of isolates that had *norA* ranged from 24 to 36 mm (with a mean of 29.7 mm) and that of the isolates that lacked *norA* ranged from 28 to 35 mm (with a mean

of 31.3 mm). The mean difference between isolates that had *norA* and that lacked *norA* was 1.5 mm (95% CI -0.8 to 3.8, P=0.192). There was no statistical significance in the mean difference between ZOI of CC1 and CC97 isolates (P=0.051) for enrofloxacin (see Table 5.4).

5.3.6 TRIMETHOPRIM-SULPHAMETHOXAZOLE

All 173 *S. aureus* isolates were susceptible to the trimethoprim-sulphamethoxazole combination according to the CLSI ((ZOI $\geq$ 16mm for a susceptible isolate) (one isolate was excluded as the disk fell off the agar). Their ZOI ranged from 21 to 35mm, with a mean of 28.3 mm. No gene mutations specific for trimethoprim and sulphamethoxazole resistance were found in any of the *S. aureus* isolates.

Mean ZOI for CC97 isolates was 1.24 mm larger than for CC1 isolates with statistical significance (P=0.012) (see Table 5.4).

5.4 DISCUSSION

The disk diffusion AST is still the most commonly used test for identifying resistant organisms in clinical veterinary practice in New Zealand and elsewhere. However, with the advent of new bioinformatic pipelines and the fall in sequencing costs, resistance gene searches via whole-genome sequencing could potentially become a viable alternative, as WGS has the potential to provide much more information to the end-user (Baker et al., 2018). The aim of this study was to assess the effectiveness of WGS-AMR by comparing genomic searches with the standard disk diffusion AST.

The most common phenotypic resistance identified in this study was penicillin resistance (59/173 isolates; 34.1%), with amoxicillin resistance being present at similar levels (55/174 isolates; 31.6%). Considering the variation due to the different sampling strategies, this level of resistance was comparable to the observations by Petrovski et al. (2011, 2015) (Table 5.5). Penicillin resistance was predicted by WGS-AMR and ResFinder with a specificity of 100% and a sensitivity of 71%. For amoxicillin, specificity was 100% and sensitivity 74%. The high specificity and

Table 5.5: The percentage of resistance observed for antimicrobials in this study along with those observed in Petrovski et al. (2011) and Petrovski et al. (2015) studies.

Antimicrobials (cont*)	Resistance (%)		
	This study	Petrovski et al. (2011)[a]	Petrovski et al. (2015) (cont[b])
Penicillin (6 μg)	34.1	26.9	20.6 (10 μg)
Amoxicllin (25 μg)	31.6	26.6	20.6 (25 μg)
Amox-clav (30 μg)	0	1.3	0 (30 μg)
Cloxacillin (5 μg)	0	2.1 (oxacillin)	0 (5 μg)
Erythromycin (15 μg)	2.9	25.3	0 (15 μg)
Lincomycin (15 μg)	1.7	33.9	99.1 (2 μg)
Streptomycin (10 μg)	5.7	28.3	0.9 (10 μg)
Neomycin (30 μg)	0	6.2	0 30 μg)
Tetracycline (30 μg)	2.3	1.8	0 (30 μg)
Enrofloxacin (5 μg)	0	ND	0 (5 μg)

Only antimicrobials that were used in more than one study are listed in this table. The resistance percentage from both Petrovski et al. (2011) and Petrovski et al. (2015) studies were calculated from the susceptibility percentage listed in the study. cont - content, Amox-clav - Amoxicillin and clavulanic acid. * - The concentration of antimicrobials used in this study. [a] - The concentration of antimicrobial used were not listed in this study. [b] - The disk content of antimicrobial used in this study is listed in parentheses

similarity between these two results is due to the fact that both drugs are inactivated by β-lactamases produced by the expression of the *blaZ* gene (Bush and Bradford, 2016). However, a study by Bradley et al. (2015b) found that WGS-AMR had a specificity of 88.3% and a sensitivity of 99.7% for penicillin, while the study by (Gordon et al., 2014) reported a specificity of 100% and a sensitivity of 99.1% for penicillin. Those studies had a prevalence of penicillin resistance among human *S. aureus* isolates of 80% and 88%, respectively. The lower sensitivity in the current study when compared to the studies by Bradley et al. (2015b) and Gordon et al. (2014) could be explained by a variation in the study design, specifically, the inclusion of short or low coverage contigs for *blaZ* gene as Gordon et al. (2014) found a high number of isolates with susceptible genotype with a resistant phenotype.

The sensitivity of WGS was much lower in this study than in Bradley et al. (2015b) and Gordon et al. (2014), because of the higher proportion of phenotypically resistant isolates which lacked the *blaZ* gene. Since the programs used in this study failed to detect any known genetic determinant of penicillin resistance in some phenotypically resistant isolates, it is unclear what determined the resistance in such isolates. Interestingly, the ZOI of isolates that were phenotypically resistant to penicillin and amoxicillin, but genotypically-negative were much larger than those of the genotypically-positive isolates (24.1 vs 16.8 mm, respectively), and their mean

difference was statistically significant. It would be interesting to see whether isolates that are only phenotypically resistant behave differently in the clinical setting from isolates that are also genotypically resistant.

None of the isolates were phenotypically resistant to amoxicillin-clavulanic acid based on the thresholds recommended by the disk manufacturer. This was consistent with previous studies done on *S. aureus* from bovine mastitis in New Zealand (McDougall et al., 2014; Petrovski et al., 2015). This is likely due to clavulanic acid being a specific β-lactamase inhibitor that prevents amoxicillin degradation by plasmid-mediated β-lactamases such as the enzyme coded by *blaZ* gene. Nevertheless, the presence of the *blaZ* gene did affect the ZOI of the *S. aureus* isolates (mean difference in ZOI between isolates with and without *blaZ* was 12.4 mm, and this difference was also statistically significant). This suggests some degradation of amoxicillin by the *S. aureus* β-lactamase does occur *in vitro*, even in the presence of clavulanic acid. However, this finding should be weighed with the fact that the threshold used for amoxicillin-clavulanic acid was not a standardised interpretative criteria.

Consistent with Petrovski et al. (2015), no isolates were found which were phenotypically resistant to cloxacillin (Table 5.5). As expected, the presence of the *blaZ* gene was not related to either phenotypic resistance or ZOI, as cloxacillin is an isoxazolyl penicillin specifically designed to resist β-lactamases encoded by the *blaZ* gene.

Another key finding of the WGS-AMR concerned cloxacillin. None of the isolates possessed the *mec* gene, which encodes the penicillin-binding protein PBP 2A, which has lower affinity for beta-lactam antimicrobials (especially isoxazolyl penicillins). This is the main explanation for why none of the isolates were phenotypically resistant to cloxacillin and also why no resistance was observed to cefoxitin, which is used as a drug for the *in vitro* detection of methicillin-resistant *S. aureus* (MRSA) (Fernandes et al., 2005).

Approximately 10% of human clinical *S. aureus* isolated in New Zealand are MRSA (Heffernan et al., 2016). However, even with the discovery of CC398 MRSA, a livestock-associated MRSA in 2011, the MRSA has been documented only in some countries in mastitis-causing *S. aureus* (Williamson et al., 2014a; Schnitt and Tenhagen, 2020). To date, only one MRSA from a bovine milk sample has been

identified in New Zealand (Grinberg et al., 2008), but its clinical significance was unclear as it was obtained from a mixed culture with an MSSA.

No phenotypic resistance was observed against ceftiofur, a third-generation cephalosporin. This result was expected, as only low resistance levels against ceftiofur for S. aureus have been recorded worldwide (Oliveira et al., 2012; Abdi et al., 2018) and ceftiofur is a β-lactamase resistant cephalosporin. This may also be related to the relatively low use of this antimicrobial in cattle in New Zealand, especially in young stock, and its classification as a 'red' antimicrobial that should be used on a case by case basis only when indicated by the New Zealand Veterinary Association (NZVA) (NZVA, 2019; MPI, 2019; Hillerton et al., 2021).

Phenotypic resistance to other classes of antimicrobials was minimal, ranging from 0 to 6% and thus, it was difficult to calculate the level of correlation between phenotypic resistance and the presence of resistance genes. However, in all cases except for one isolate tested against erythromycin and lincomycin and seven isolates tested against streptomycin, the observed phenotypic resistance was accompanied by the presence of a resistance gene against the corresponding drug. The resistance to aminoglycosides was similar to that reported by Petrovski et al. (2015) but lower than Petrovski et al. (2011) (Table 5.5). This pattern was similar for the resistance to neomycin (Table 5.5).

The resistance to erythromycin was comparable with Petrovski et al. (2015) while it was lower than that reported by Petrovski et al. (2011) (Table 5.5). Resistance to tylosin and oleandomycin were not tested in those studies. The reported prevalence of resistance to lincomycin was much higher in previous studies (Table 5.5) than that reported in the current study. The reason for this large discrepancy is likely to be related to the lack of specific interpretive veterinary criteria for lincomycin. In this study a ZOI of 0 mm was used to define resistance, but Petrovski et al. (2011) did not report how resistance was defined while Petrovski et al. (2015) mentioned that internal break points were used. Furthermore, part of the variation in the resistance levels may be attributable to the fact that a 2 μg lincomycin disk was used by Petrovski et al. (2015) study, while a 15 μg disk was used in the current study. (Petrovski et al., 2011) did not report the content of the lincomycin disks used in their study. This emphasises the urgent need for the standardisation of testing methods and interpretive criteria in veterinary medicine.

The current study's prevalence of resistance to enrofloxacin (0%) was consistent with the 0% resistance reported by Petrovski et al. (2015), although *norA*, an efflux pump gene was found in all but four isolates in all the databases except for Resfinder. This highlights the need for standardized definitions for maintaining antimicrobial resistance genes across all platforms and databases.

Four isolates were found to be intermediately susceptible/resistant to tetracycline, in line with the findings by Petrovski et al. (2011) and Petrovski et al. (2015) (Table 5.5). All the isolates including the four resistant isolates contained the *tet38* gene, a chromosomally encoded efflux pump. The expression of the *tet38* gene needs complex regulation by local and global regulators (Truong-Bolduc et al., 2005, 2015), so the lack of correlation between the phenotypic and genotypic tests in this study corroborates that the mere presence of this gene is unable to predict a phenotypic resistance to tetracycline.

With regards to the databases used, Resfinder was the most accurate in identifying phenotypic resistance, with 100% of the isolates with resistance genes defined by Resfinder being phenotypically resistant to the corresponding drug. Other databases identified false positives and also intrinsic factors such as *tet38* and *norA*. CARD database was the most comprehensive database, with 18 resistance factors both acquired and intrinsic reported, while Resfinder only reported 4 acquired resistance factors.

It would be interesting to analyse the data, using tools such as Pointfinder (Zankari et al., 2017), to identify mutations in any of the intrinsic resistance factors that could explain the discrepancies observed with tetracycline. This webtool was online during 2017, but then taken down from the website during the period of this study, and it was unable to be used. This software has been repackaged and published as a stand-alone software in 2021, after the completion of this analysis.

Genome-wide association studies (GWAS) would provide a better understanding of complex resistance in *S. aureus* in particular where multiple gene interactions are needed, such as, the regulation of efflux pumps and reducing permeability to antimicrobials (Blair et al., 2015). This could help in the understanding of the four intermediate resistant isolates to tetracycline in the current study, due to the complex nature of its expression. Machine learning (ML) approaches have been gaining interest in genome studies, and would be a better alternative for predicting

antimicrobial resistance, than the simple gene presence/absence analysis used in this study (Su et al., 2019). The approach uses training sets of genomes with known phenotypes, and is fed from an existing database and focussed on genetic features (such as SNP, substitutions, and indels). These genetic features are given scores or weights based on their contribution to antimicrobial resistance. This approach can also model interactions between resistance genes as well as between these genes and the genetic background of the strain (Su et al., 2019). The major disadvantage of the ML approach is the dependency on the quality of the data used to develop and test the model. If these data are not accurate, then the model will not be accurate either (Boolchandani et al., 2019). However, the ML method is useful in the GWAS aspect of the WGS-AST.

In conclusion, WGS-AMR based on gene presence/absence determinations has the potential to replace laboratory AST for bovine mastitis *S. aureus* as the costs and speed of sequencing improve, more resistance genes are identified, and the databases are updated. Our results suggest, the analysis of genes coding for β-lactam resistance may have high positive predictive values, but only modest negative predictive values. The same cannot be said for tetracycline resistance, which is highly regulated by complex gene regulation mechanisms. These results suggest machine learning approaches, once developed and validated, may be more accurate predictors of an isolate's antibiogram than mere gene presence/absence analysis.

6

IDENTIFICATION OF POTENTIAL PROTEIN VACCINE CANDIDATES THROUGH REVERSE VACCINOLOGY

6.1 INTRODUCTION

Staphylococcus aureus is one of the primary causative organisms of bovine mastitis in New Zealand and causes severe economic loss to the dairy industry (Petrovski, 2007; Hogeveen et al., 2011). The adaptation of *S. aureus* to the bovine mammary gland environment can make antimicrobial treatment ineffective even if the organisms are susceptible to those antimicrobials. Factors such as the ability of *S. aureus* to survive in the host epithelial cells and neutrophils and induce the formation of micro-abscesses are contributing factors to the failure of antimicrobial treatment (Barkema et al., 2006).

Bacterial cure rates in bovine mastitis after antimicrobial treatment vary markedly from 4 to 92% (Rainard et al., 2018). The considerable variability in the treatment efficacy is because these figures include data from cows treated during lactation and cows treated at drying off. The cure rates for cows treated at drying off are higher because of the longer duration of action of the antimicrobials used at this time (Barkema et al., 2006). Extended treatment with antimicrobials is usually associated with increased cure rates (Roy and Keefe, 2012). However, it can lead to an increase in withdrawal time, antimicrobial residues and reduction in milk production. Hence, preventative measures such as hygienic milking, culling of chronic cases, and vaccinations have been undertaken to reduce the spread of infections within the herd (Barkema et al., 2006).

Vaccination against *S. aureus* mastitis has been a focus of research for many years, but only a few products are commercially available. Currently, there are two vaccines against *S. aureus* mastitis in cattle which are commercially available in NZ: Lysigin (Boehringer Ingelheim Vetmedica Inc, USA) and StartVac (HIPRA, Spain). Lysigin contains a lysed culture of 5 different phage types (see section 1.1.4) of *S. aureus* while StartVac contains lysed culture of *S. aureus* SP140 along with the *E. coli* J5 antigen. Although these vaccines have been shown in controlled studies to decrease the clinical duration and severity of *S. aureus* intramammary infections(IMI) and increase serum and milk antibody concentrations against *S. aureus* (Pankey et al., 1985; Middleton et al., 2006; Luby et al., 2007; Middleton et al., 2009; Nickerson et al., 1999; Schukken et al., 2014; Bradley et al., 2015a; Bryan M, 2015; Guccione et al., 2017), their efficacy in preventing or clearing *S. aureus* IMI in cattle has been, at best, equivocal. The failure of these vaccines to effectively prevent new *S. aureus*

IMI may have been due to the failure of killed bacteria in the commercial vaccines to elicit a protective immune response that was sufficient to prevent IMI. This may have been due to the use of killed bacterial cells which generally produce a reduced immune-response compared with live vaccines (Bhardwaj, 2018). Other factors that may limit the efficacy of vaccines against *S. aureus* IMI include the biofilm production during the infection, which increases the resistance of *S. aureus* to immune factors, as biofilms are antiphagocytic, and the anti-inflammatory effect (Barrio et al., 2000) and facultative intracellular nature of *S. aureus,* which makes it inaccessible to antibodies (Hébert et al., 2000).

A DNA/protein vaccine targeted against the *S. aureus* fibronectin-binding protein and clumping factor A has shown promising results in stimulating both specific cell-mediated and humoral immune-responses (Shkreta et al., 2004). Inducing cell-mediated immune response might play a significant role in clearing IMI with *S. aureus* due to its localisation within epithelial and phagocytic cells.

It would be extremely beneficial to identify further candidate proteins/antigens which may increase the efficacy of *S. aureus* vaccines. This would seem to be a process ideally suited to reverse vaccinology. Reverse vaccinology (RV), first described by Rappuloi in the early 2000s (Rappuoli, 2000), has already helped in the prediction of potential vaccine candidates for bacterial and viral organisms against which conventional vaccinology methods have struggled to develop effective vaccines (Rodríguez-Ortega et al., 2006; Tettelin et al., 2006; Bambini and Rappuoli, 2009; Song et al., 2009; Carlos et al., 2015). The RV method utilises the entire genome or proteome of an organism, to predict potential vaccine candidates *in silico* using various bioinformatics tools. The advantages of the RV method are:

1) The identification of novel antigens which are missed out using conventional methods.

2) The identification of antigens that are not expressed *in vitro*.

3) A less time-consuming prediction of potential vaccine candidates when compared to the progressive, slow discovery and screening of candidates applied with conventional methods.

There are several bioinformatic packages specifically designed for RV. The input for

all the RV packages is the proteome or the protein sequences of the organism of interest. Based on their algorithmic approach, these packages can be classified as either decision-tree/filtering or machine learning/classifying.

Packages that use a decision-tree or filtering approach filter out protein sequences based on multiple criteria, including the proteins' features, their localisation, the probability of the protein being an adhesin, or the similarity to host proteins. The programs vary based on the filters used, as well as the default threshold for their filters. NERVE (Vivona et al., 2006), Vaxign (He et al., 2010) and VacSol (Rizwan et al., 2017) are examples of decision tree RV packages.

Machine learning or classifying RV tools assign the input protein sequences into either potential vaccine candidates (PVCs) or not-PVCs based on the training examples of PVCs and not-PVCs. The machine learning programs do not discard or filter out proteins but rank them based on the program-based scoring systems. Examples of machine-learning RV tools are Vaxign-ML (Ong et al., 2020) and VaxiJen (Doytchinova and Flower, 2007).

There are no studies on designing vaccine candidates using the RV method for *S. aureus* mastitis in cattle. Therefore, given that the population genetic structure of bovine *S. aureus* in New Zealand is stable in time, as determined in the study presented in Chapter 3, this study aimed to identify PVCs against *S. aureus* using the isolates used in the same Chapter 3.

6.2 MATERIALS AND METHODS

6.2.1 SELECTION OF ISOLATES

Three isolates (B003, B083 and B133) belonging to the dominant clonal complex (CC), CC1, which comprised ~70% of New Zealand isolates identified in Chapter 3 were selected at random for the initial *in silico* RV process. The three isolates were selected from each of the collection period described in Chapter 2. Three isolates were selected to avoid omitting important accessory genes and to check for concordance of the results through the pipeline. The proteome (the faa file of the

Figure 6.1: Flowchart of the reverse vaccinology pipeline used in this study which utilised the proteomes extrapolated from the whole genome sequences of *S. aureus* analysed in Chapter 3. The flowchart includes only the filtering steps.

Prokka output, see 2.4.5) of each of the selected isolates was used for the RV pipeline.

6.2.2 REVERSE VACCINOLOGY PIPELINE

6.2.2.1 PIPELINE'S GENERAL FEATURES

The RV pipeline used in this study is schematically described as a flowchart in Figure 6.1. Initially, the pipeline was implemented on the three pre-selected CC1 isolates.

The parameters of the various programs selected followed those recommended by Dalsass et al. (2019). The parameters were chosen based on the criteria of predicted exposure of the coded proteins to the host immune system, their ability to be cloned and purified for *in vivo* immunogenicity testing, and finally, their occurrence in all the

isolates of the major clonal complexes (CCs) identified in Chapter 3.

The criterion of exposure to the immune system was ascertained by algorithms that estimate their localisation (extracellular or cell-wall), the adhesin function probability, and the predicted physico-chemical antigenic properties of the proteins. Proteins' localisation and adhesin probability were predicted using the Vaxign2 (Ong et al., 2021) web tool, while their physico-chemical antigenic properties were analysed using VaxiJen.

The ability to be cloned and purified was ascertained through the analysis of the number of transmembrane helices (TMH) of the proteins. The number of TMH is essential for cloning and purifying the proteins using recombinant technology. Pizza et al. (2000) found that proteins with more than one TMH had a higher rate of expression failure. Hence, proteins with more than one TMH were filtered out.

An effective PVC should be present in all the isolates in the target population. Since there were not many bovine *S. aureus* mastitis genomes from NZ in genome databases, only the genomes sequences in the study described in Chapter 3 were used. The PVC's nucleotide sequences identified in the genomes of the first CC1 isolates were used to map them to the reads of the other isolates belonging to CC1, CC97, CC151 and CC8, which constituted about 96% of the isolates described in Chapter 3. The sections below report a detail description of the bioinformatic pipeline starting with the proteins from the three CC1 isolates, each processed separately.

6.2.2.2 PREDICTION OF LOCALISATION AND TRANSMEMBRANE HELICES BY VAXIGN2

Vaxign2 has two modalities (Ong et al., 2021): 1) Precompute query: it allows to select and upload genomes (more than 350 genomes of various bacteria and virus) which are uploaded into Vaxign2 and precomputed for PVC; 2) Dynamic analysis, which allows user-inputted proteins to be analysed. The dynamic analysis was used for this prediction, with the following menu options: Gram-positive bacterium; subcellular localisation; transmembrane helices (TMH); adhesin probability. The localisation is predicted by the program pSORT (Gardy et al., 2005) tool, which enables the selection of the proteins exposed to the host immune system, either present on the bacterial cell-wall or as extracellular secretions. The TMH are

predicted by the program HMMTOP (Tusnady and Simon, 2001a). The probability of the protein being an adhesin is calculated by the program SPAAN (Sachdeva et al., 2005). Adhesin proteins are essential for the attachment of bacteria to the host cells and are often exposed to the host immune system (Finlay and Falkow, 1997).

The Vaxign2 dynamic analysis produces a spreadsheet of all the proteins provided in the input, with their hypothetical localisation (cell-wall, extracellular, cytoplasmic membrane, cytoplasmic, and unknown) and the probability of the protein's localisation (ranging from 0 to 1), the adhesin function probability, the number of TMHs, and the protein length (as amino acid number).

In this study, the proteins that were identified as cell-wall and extracellular proteins with a probability over 0.8 were selected for further analysis (the localisation of the proteins with a probability of less than 0.75 is defined as 'unknown' by the Vaxign2 program). From this pool of proteins, all the proteins with TMH of less than 2 were selected. The adhesin probability was not used as a selection criterion, to avoid losing non-adhesin proteins that may still be immunogenic. The selection of these proteins was made from the spreadsheet output of Vaxign2. The sequences of these proteins were extracted from the proteome (faa file of Prokka output) using their locus tag through an in-house python script that utilised samtools faidx (Li et al., 2009). The in-house python script can be found in the github repository (`https://github.com/jabinnes/Thesis-files/blob/main/Scripts/Reverse_vaccinology.ipynb`) for the thesis.

6.2.2.3 PREDICTION OF ANTIGENICITY USING VAXIJEN

The protein sequences selected in section 6.2.2.2 were uploaded onto the web tool VaxiJen (Doytchinova and Flower, 2007). VaxiJen predicts antigenicity of the proteins using their physico-chemical properties, independent of their amino acid sequence alignment. The VaxiJen process was run with the threshold at the default value of 0.5, with 'bacterium' as the target organism.

The VaxiJen's output, for each isolate, is represented by the list of the input proteins with their identifier and a prediction probability score, and any protein with a prediction probability score larger than 0.5 is described as 'Probable ANTIGEN'. Proteins with a prediction probability score of less than 0.5 are termed 'Probable

NON-ANTIGEN'.

Using samtools and an in-house python script (https://github.com/jabinnes/Thesis-files/blob/main/Scripts/Reverse_vaccinology.ipynb) , the nucleotide and protein sequences of the proteins which were termed Probable ANTIGEN were extracted from the Prokka output of the three isolates.

The three isolates were run through these steps until this point. The protein sequences (extracted from B003 prokka output) that were present in all the three isolates were selected for further processing.

6.2.2.4 MAPPING THE NUCLEOTIDE SEQUENCES OF THE SELECTED PROTEINS TO THE READS OF THE MAJOR CLONAL COMPLEXES USING ARIBA

From this step, a single list of selected proteins present in all three isolates were used. This step was undertaken to verify that the selected proteins are present in all the major CCs (CC1, CC97, CC8, CC151) identified in this study (see Chapter 3). This was done to ensure that the vaccine candidates are universally present in bovine *S. aureus* in New Zealand. A local nucleotide database was created from the Prokka output, consisting of the proteins that were selected in the previous section. An in-house python script (https://github.com/jabinnes/Thesis-files/blob/main/Scripts/Reverse_vaccinology.ipynb) utilising samtools was used to extract the nucleotide sequences of the proteins, using their unique identifiers (i.e. the locus tag). ARIBA (Hunt et al., 2017) was used to map the sequence-reads of all the isolates belonging to the major CCs to the above mentioned nucleotide sequences. The summary report file produced by ARIBA lists the presence/absence of the nucleotide sequences. The gene sequence is considered present if there is a 95% identity to the sequences in the user-created local sequence database. Using this presence/absence file, protein sequences (by their locus tag) were then selected if they were present in all isolates belonging to the major clonal complexes (CC1, CC97, CC8, CC151).

6.2.2.5 PROTEIN SELECTION BASED ON VAXIGN-ML PROTEGENICITY SCORE AND SIMILARITY TO HUMAN PROTEINS

The shortlisted protein sequences were run through the dynamic query of the Vaxign2 web tool with the following options: Gram-positive bacterium; subcellular localisation, TMH, adhesin probability, Vaxign-ML and similarity to human proteins. The Vaxign-ML is a program that uses five machine learning algorithms to predict protective antigens based on their biological and physico-chemical properties (Ong et al., 2020). The Vaxign-ML produces a percentile rank score for each protein, termed the protegenicity score (Vaxign-ML score). The authors of Vaxign-ML found that the vaccines that contained recombinant proteins, either used in clinical trials or licensed for use, had proteins with a protegenicity score greater than 90 (Ong et al., 2020). Hence, proteins with a protegenicity score > 90 were selected for further analysis.

The similarity to human proteins was calculated in Vaxign2 using BLAST. This was done to minimise the presence of immunoglobulins against human proteins in cow milk intended for human consumption (Ulfman et al., 2018).

6.2.2.6 EPITOPE PREDICTION USING BCPREDS

The protein sequences shortlisted from the previous step were run through BCPreds: the B-cell epitope prediction server (`http://ailab-projects1.ist.psu.edu:8080/bcpred/predict.html`) using the BCPreds string kernel (El-Manzalawy et al., 2008) and Amino Acid Pair (AAP) antigenicity scale (Chen et al., 2007) algorithms. The BCPreds algorithm predicts multiple linear B-cell epitopes using machine learning that utilises sequences of varying lengths while the AAP algorithm predicts multiple linear B-cell epitopes based on the finding that particular amino acid pairs are favoured to be epitopes. The BCPreds and AAP algorithms use machine learning trained with sets of epitopes and non-epitope amino acid sequences. Based on the studies by Chen et al. (2007) and EL-Manzalawy et al. (2008), an epitope length of 20 amino acids and default specificity of 75% was used for both the BCPreds and AAP algorithm for all the potential vaccine candidate proteins selected from the previous step.

6.2.2.7 PROTEIN SEQUENCE VARIABILITY

The variability in the protein sequences can affect the efficacy of a vaccine if those variable regions occur in the epitope regions (Maiden, 2019). In order to assess the variable regions in the potential vaccine candidates compared to the proteome of the isolates of major CCs, blastp (Altschul et al., 1990) analysis was performed. The blastp analysis was done by comparing the protein sequences of the shortlisted PVCs (query) against the faa output file (subject) of the prokka annotation of the isolates of the major CCs (CC1, CC97, CC151 and CC8) with default parameters except for the output file format set to tabular format (outfmt-6). The protein sequences with the best hit (highest percent identity and coverage) were extracted from the faa output using samtools via an in-house python script (https://github.com/jabinnes/Thesis-files/blob/main/Scripts/Reverse_vaccinology.ipynb). The extracted sequences for each shortlisted protein were aligned separately using Geneious v.6.1 (Biomatters, Auckland, New Zealand). The alignment was visualized using Weblogo (Crooks et al., 2004) through their online platform (Weblogo v2.8.2; https://weblogo.berkeley.edu/logo.cgi). The Weblogo utilises sequence logos (Schneider and Stephens, 1990), a graphical representation consisting of amino acid symbols stacked for each position in the sequence. The overall height of the stack reveals the sequence conservation while the height of each symbol indicates the frequency of each amino acid at that position. This step is not a filtering step but helps identifying the variable regions of the proteins.

6.3 RESULTS

The proteomes used in the study belong to three isolates: B003, B083 and B133, and comprised 2565, 2560, 2492 open reading frames (ORFs), respectively.

Table 6.1: Distribution of proteins predicted by PSORTb through the Vaxign2 pipeline for the three isolates, according to their predicted localisations.

Isolate	Prediction of protein localisation	Number of proteins (%)
B003 (Total ORFs 2565)	Cytoplasmic	1292 (50.4)
	Cytoplasmic membrane	692 (27)
	Unknown	456 (17.8)
	Extracellular	85 (3.3)
	Cell-wall	40 (1.5)
B083 (Total ORFs 2560)	Cytoplasmic	1300 (50.8)
	Cytoplasmic membrane	690 (26.9)
	Unknown	447 (17.5)
	Extracellular	84 (3.3)
	Cell-wall	39 (1.5)
B133 (Total ORFs 2492)	Cytoplasmic	1261 (50.6)
	Cytoplasmic membrane	684 (27.4)
	Unknown	425 (17.1)
	Extracellular	83 (3.3)
	Cell-wall	39 (1.6)

6.3.1 PROTEIN SELECTION ACCORDING TO LOCALISATION

The distribution of the different protein localisations predicted by PSORTb through the Vaxign2 pipeline for the three isolates is presented in Table 6.1. The percentage of predicted extracellular proteins was 3.3% across the three isolates, while that of the predicted cell-wall proteins was 1.5%. At the end of this process 125, 123 and 122 proteins were selected from the genomes of B003, B083 and B133, respectively. An Upset plot showing the intersection of the selected proteins from the three isolates is shown in Figure 6.2. A total of 111 proteins were found in all three isolates.

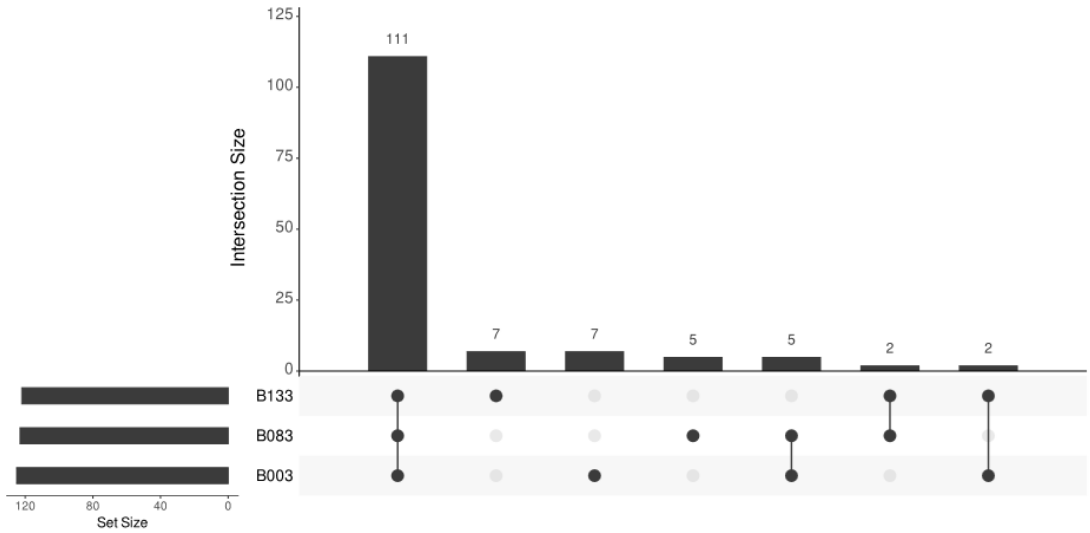

Figure 6.2: Upset plot of the intersection of the proteins selected based on the three isolates' localisation (extracellular and cell-wall). The number of proteins in each isolate; B003-125, B083- 123, and B133-122.

6.3.2 PROTEIN SELECTION ACCORDING TO THE NUMBER OF TRANSMEMBRANE HELICES

Proteins selected in the previous step were checked for the number of TMHs and proteins with <2 TMHs were selected for further processing. Five extracellular proteins from B003, B083 and B133 were discarded as they had more than two TMHs, leaving a total of 120, 118, and 117 proteins, respectively. The excluded extracellular proteins were the Thermonuclease, immunodominant staphylococcal antigen B, Serine protease Spl D, Serine Protease SplF, and a hypothetical protein. The remaining proteins are listed in Table 6.2.

Table 6.2: List of the proteins selected as potential vaccine candidates based on their localisation and number of transmembrane helices (TMH). The list indicates the presence of the proteins in one or more of the three isolates (B003-A, B083-B, and B133-C) used in this study along with the subcellular localisation, adhesin probability, number of TMH, length, and Vaxigen antigenicity probability scores. The rows that are italicised are the proteins that were eliminated from downstream processes based on a VaxiJen antigenicity probability score <0.5. Proteins with an adhesin probability >0.5 are predicted as more likely to be adhesin proteins.

Protein description	Presence in Isolates	Subcellular Localisation	Adhesin probability	Number of TMH	Length of proteins (amino acids)	VaxiJen score
14.7 kDa ribonuclease H-like protein	A, B, C	Cell-wall	0.26	0	133	0.6031
1-phosphatidylinositol phosphodi-esterase	*A, B, C*	*Extracellular*	*0.564*	*1*	*328*	*0.4602*
Acid shock protein	*A, B, C*	*Cell-wall*	*0.742*	*0*	*140*	*0.417*
Alkaline phosphatase 3	A, B, C	Extracellular	0.784	0	474	0.5414
Alpha-hemolysin	A, B, C	Extracellular	0.629	0	319	0.611
Antitoxin MazE	A, B, C	Extracellular	0.569	0	56	0.6105
Bifunctional autolysin	A, B, C	Extracellular	0.787	0	1256	0.5414
Bifunctional autolysin	*A, B, C*	*Extracellular*	*0.499*	*1*	*258*	*0.4191*
Bifunctional autolysin	*A, B, C*	*Extracellular*	*0.731*	*1*	*284*	*0.4352*
Cell division protein DivIB	A, B, C	Cell-wall	0.16	1	439	0.7737
Cell division protein DivIC	A, B, C	Cell-wall	0.141	1	130	0.687
Collagen adhesin	A, B, C	Cell-wall	0.857	1	1183	0.7419
ComG operon protein 3	A, B, C	Cell-wall	0.467	1	103	0.7813

Continued on next page

Protein description	Presence in Isolates	Subcellular Localisation	Adhesin probability	Number of TMH	Length of proteins (amino acids)	VaxiJen score
Disulfide bond formation protein D	*A, B, C*	*Cell wall*	*0.594*	*0*	*199*	*0.4703*
Enterotoxin type A	A, B, C	Extracellular	0.471	0	229	0.7545
Enterotoxin type H	A, B, C	Extracellular	0.634	0	241	0.6617
ESAT-6 secretion system extracellular protein A	A, B, C	Extracellular	0.593	0	97	0.5382
ESAT-6 secretion system extracellular protein B	*A, B, C*	*Extracellular*	*0.424*	*0*	*104*	*0.4197*
Extracellular matrix protein-binding protein emp	*A, B, C*	*Cell-wall*	*0.558*	*0*	*295*	*0.4222*
Extracellular matrix-binding protein ebh	A, B, C	Cell-wall	0.151	1	665	0.5437
Fibrinogen-binding protein	A, B, C	Extracellular	0.526	0	165	0.5307
Fibronectin-binding protein A	A, B, C	Cell-wall	0.2	1	69	0.5395
Fibronectin-binding protein A	B	Cell-wall	0.683	0	957	0.7719
Fibronectin-binding protein A	*C*	*Cell-wall*	*0.283*	*0*	*171*	*0.4465*
Gamma-hemolysin component A	A, B, C	Extracellular	0.575	0	321	0.6432
Gamma-hemolysin component B	A, B, C	Extracellular	0.687	0	325	0.649
Gamma-hemolysin component C	A, B, C	Extracellular	0.772	0	315	0.6942
Glutamyl endopeptidase	*A, B, C*	*Extracellular*	*0.845*	*0*	*333*	*0.3828*

Continued on next page

Protein description	Presence in Isolates	Subcellular Localisation	Adhesin probability	Number of TMH	Length of proteins (amino acids)	VaxiJen score
Glutathione hydrolase proenzyme	*A, B, C*	*Extracellular*	*0.192*	*0*	*668*	*0.4561*
Glutathione-binding protein GsiB	A, B, C	Cell-wall	0.644	0	491	0.5267
Glycyl-glycine endopeptidase LytM	A, B, C	Extracellular	0.899	0	316	0.8209
Holliday junction resolvase RecU	*A, B, C*	*Extracellular*	*0.434*	*0*	*208*	*0.4742*
Hyaluronate lyase	A, B, C	Extracellular	0.67	0	815	0.5511
hypothetical protein	A, C	Cell-wall	0.889	0	2017	1.1083
hypothetical protein	A, B, C	Cell-wall	0.868	0	1795	0.7935
hypothetical protein	A, B, C	Cell-wall	0.151	0	323	0.5034
hypothetical protein	A, B, C	Extracellular	0.658	1	457	0.5068
hypothetical protein	A, B, C	Cell-wall	0.593	0	772	0.6489
hypothetical protein	A, B, C	Extracellular	0.605	0	356	0.8309
hypothetical protein	A	Cell-wall	0.684	1	348	1.0494
hypothetical protein	A, B, C	Extracellular	0.557	0	272	0.8084
hypothetical protein	A, B, C	Extracellular	0.613	0	234	0.5875
hypothetical protein	A, B, C	Extracellular	0.515	0	232	0.5905
hypothetical protein	A, B, C	Cell-wall	0.825	0	686	0.9621
hypothetical protein	A, B, C	Extracellular	0.434	0	232	0.5688
hypothetical protein	A, B, C	Extracellular	0.475	0	232	0.7908

Continued on next page

Protein description	Presence in Isolates	Subcellular Localisation	Adhesin probability	Number of TMH	Length of proteins (amino acids)	VaxiJen score
hypothetical protein	A, B, C	Extracellular	0.723	0	231	0.6447
hypothetical protein	A, B, C	Extracellular	0.388	0	231	0.6155
hypothetical protein	A, B, C	Extracellular	0.667	0	230	0.7589
hypothetical protein	A, C	Extracellular	0.578	0	227	0.6262
hypothetical protein	A, B, C	Extracellular	0.626	0	226	0.7442
hypothetical protein	A, B, C	Cell-wall	0.559	1	198	0.6622
hypothetical protein	A, B, C	Extracellular	0.56	1	192	0.611
hypothetical protein	A, B, C	Extracellular	0.82	0	166	0.6002
hypothetical protein	A, B, C	Extracellular	0.752	1	152	1.2005
hypothetical protein	A, B, C	Extracellular	0.689	0	131	1.2026
hypothetical protein	A, B, C	Extracellular	0.546	0	105	0.5043
hypothetical protein	A, B, C	Extracellular	0.639	0	60	0.6157
hypothetical protein	A, B, C	Cell-wall	0.776	0	837	0.6816
hypothetical protein	A, B, C	Extracellular	0.264	0	50	0.7715
hypothetical protein	A, B, C	Extracellular	0	0	48	0.9847
hypothetical protein	C	Extracellular	0	0	43	0.8962
hypothetical protein	A, B, C	Extracellular	0	0	34	0.867
hypothetical protein	A, B, C	Extracellular	0	0	30	0.9814

Continued on next page

Protein description	Presence in Isolates	Subcellular Localisation	Adhesin probability	Number of TMH	Length of proteins (amino acids)	VaxiJen score
hypothetical protein	*A, B, C*	*Cell-wall*	*0.063*	*0*	*439*	*0.247*
hypothetical protein	*A, B*	*Extracellular*	*0.395*	*0*	*634*	*0.3501*
hypothetical protein	*B, C*	*Cell-wall*	*0.543*	*0*	*173*	*0.3628*
hypothetical protein	*B*	*Extracellular*	*0.162*	*0*	*58*	*0.3901*
hypothetical protein	*A, B, C*	*Extracellular*	*0.567*	*0*	*279*	*0.393*
hypothetical protein	*A, B, C*	*Extracellular*	*0.635*	*0*	*238*	*0.3989*
hypothetical protein	*A, B, C*	*Extracellular*	*0.611*	*0*	*143*	*0.4268*
hypothetical protein	*A, C*	*Extracellular*	*0.729*	*0*	*241*	*0.4417*
hypothetical protein	*B*	*Extracellular*	*0.738*	*0*	*241*	*0.4609*
hypothetical protein	*A, B, C*	*Extracellular*	*0.411*	*0*	*65*	*0.4705*
hypothetical protein	*A, B*	*Extracellular*	*0.687*	*0*	*481*	*0.4708*
hypothetical protein	*A, B, C*	*Cell-wall*	*0.439*	*1*	*635*	*0.4838*
hypothetical protein	*A, B, C*	*Extracellular*	*0.758*	*0*	*241*	*0.4927*
hypothetical protein	*A, B, C*	*Extracellular*	*0.569*	*0*	*300*	*0.4938*
Immunoglobulin G-binding protein A	A, B	Cell-wall	0.601	1	476	0.5976
Immunoglobulin G-binding protein A	C	Cell-wall	0.427	0	213	0.7585
Iron-regulated surface determinant protein A	A, B, C	Cell-wall	0.721	1	350	0.6946

Continued on next page

Protein description	Presence in Isolates	Subcellular Localisation	Adhesin probability	Number of TMH	Length of proteins (amino acids)	VaxiJen score
Iron-regulated surface determinant protein B	A, B, C	Cell-wall	0.709	1	645	0.7235
Iron-regulated surface determinant protein H	A, B, C	Cell-wall	0.592	0	895	0.659
Kinase-associated lipoprotein B	*A, B, C*	*Cell wall*	*0.367*	*0*	*127*	*0.477*
Lantimicrobial gallidermin	A, B, C	Extracellular	0	0	47	1.2684
Leucotoxin LukDv	A, B, C	Extracellular	0.734	0	327	0.6174
Leucotoxin LukDv	A, B, C	Extracellular	0.732	0	322	0.5464
Leucotoxin LukEv	A, B, C	Extracellular	0.749	0	311	0.6659
Leucotoxin LukEv	A, B, C	Extracellular	0.579	0	308	0.6506
Lipase 1	A, B, C	Extracellular	0.504	1	681	0.6694
Lipase 2	A, B, C	Extracellular	0.672	0	690	0.6623
N-acetylmuramoyl-L-alanine amidase domain-containing protein	A, B, C	Extracellular	0.526	0	619	0.6731
N-acetylmuramoyl-L-alanine amidase sle1	A, B, C	Cell-wall	0.905	0	334	0.6465
N-acetylmuramoyl-L-alanine amidase sle1	A, B, C	Cell-wall	0.849	0	265	0.6789

Continued on next page

Protein description	Presence in Isolates	Subcellular Localisation	Adhesin probability	Number of TMH	Length of proteins (amino acids)	VaxiJen score
Nickel-binding periplasmic protein	A, B, C	Cell-wall	0.756	0	532	0.6999
Phospholipase C	A, B, C	Extracellular	0.682	0	331	0.5623
Poly-beta-1,6-N-acetyl-D-glucosamine N-deacetylase	*A, B, C*	*Extracellular*	*0.532*	*1*	*290*	*0.2376*
putative deferrochelatase/peroxidase EfeN	*A, B, C*	*Extracellular*	*0.304*	*1*	*409*	*0.4546*
putative leukocidin-like protein 1	A, B, C	Extracellular	0.782	0	338	0.5946
putative leukocidin-like protein 2	A, B, C	Extracellular	0.432	1	350	0.643
putative malate:quinone oxidoreductase 1	*A, B, C*	*Cell-wall*	*0.238*	*0*	*492*	*0.4626*
putative malate:quinone oxidoreductase 2	*A, B, C*	*Cell-wall*	*0.264*	*0*	*498*	*0.3905*
putative protein	A, B, C	Cell-wall	0.552	1	241	0.6062
Putative surface protein	C	Cell-wall	0.346	0	280	0.5422
putative transglycosylase IsaA	A, B, C	Extracellular	0.928	0	233	0.8745
putative transglycosylase SceD	A, B, C	Extracellular	0.939	0	231	0.9649
Serine protease SplA	*A, B, C*	*Extracellular*	*0.596*	*0*	*235*	*0.39*
Serine protease SplB	*A, B, C*	*Extracellular*	*0.457*	*1*	*240*	*0.4367*

Continued on next page

Protein description	Presence in Isolates	Subcellular Localisation	Adhesin probability	Number of TMH	Length of proteins (amino acids)	VaxiJen score
Serine protease SplC	*A, B, C*	*Extracellular*	*0.479*	*0*	*239*	*0.4042*
Serine protease SplE	*A, B, C*	*Extracellular*	*0.631*	*1*	*238*	*0.3895*
Serine-aspartate repeat-containing protein C	B	Cell-wall	0.723	1	395	0.7601
Serine-aspartate repeat-containing protein C	*A, B, C*	*Cell-wall*	*0.329*	*1*	*51*	*0.3429*
Serine-aspartate repeat-containing protein E	*A, B, C*	*Cell-wall*	*0.432*	*1*	*51*	*0.4305*
Signal peptidase IB	A, B, C	Cell-wall	0.268	1	191	0.5887
Signal peptidase IB	*A, B, C*	*Cell-wall*	*0.376*	*1*	*174*	*0.3871*
Staphopain A	A, B, C	Extracellular	0.736	0	388	0.5219
Staphopain B	A, B, C	Extracellular	0.625	1	393	0.5797
Staphylocoagulase	A, B, C	Extracellular	0.46	0	508	0.5958
Staphylococcal secretory antigen SsaA	A, B, C	Extracellular	0.912	0	255	0.7687
Staphylococcal secretory antigen ssaA2	A, B, C	Extracellular	0.929	0	269	0.6952
Superoxide dismutase [Mn/Fe] 2	A, B, C	Extracellular	0.526	0	199	0.5284
Superoxide dismutase [Mn] 1	A, B, C	Extracellular	0.443	0	199	0.5903
Surface protein G	*A, B*	*Cell-wall*	*0.104*	*0*	*59*	*0.4362*

Continued on next page

Protein description	Presence in Isolates	Subcellular Localisation	Adhesin probability	Number of TMH	Length of proteins (amino acids)	VaxiJen score
Thermonuclease	*A, B, C*	*Extracellular*	*0.549*	*1*	*177*	*0.4952*
Trifunctional nucleotide phosphoesterase protein YfkN	*A, B, C*	*Cell-wall*	*0.223*	*0*	*511*	*0.4862*
Zinc metalloproteinase aureolysin	A, B, C	Extracellular	0.561	0	509	0.6013

6.3.3 PROTEIN SELECTION ACCORDING TO ANTIGENICITY

The selected proteins were processed using VaxiJen, which predicts antigenicity using their physico-chemical properties. Using the antigenicity probability score threshold of 0.5, 36, 38 and 35 proteins were filtered out, resulting in probable vaccine candidate proteins from B003, B083 and B133, respectively. The remaining proteins are the non-italicised proteins listed in Table 6.2. There were 47, 48, and 49 known proteins, 32, 27 and 29 hypothetical proteins, and 5, 5, and 6 putative proteins in the proteomes of B003, B083 and B133, respectively.

6.3.4 PROTEIN SELECTION ACCORDING TO THE MAPPING OF READS OF MAJOR CLONAL COMPLEX ISOLATES

The proteins remaining after the previous step were used to create a reference database and BLASTed against all the isolates belonging to major *S. aureus* clonal complexes identified in the study using ARIBA. The 28 identical proteins selected from each of the three isolates used in this study are listed in Table 6.3. These proteins were present across all the isolates from the major CCs (CC1, CC97, CC8, CC151) identified in Chapter 3. Of the identified potential vaccine candidates, 19 were known proteins, 6 were hypothetical proteins, and 3 were putative proteins. Out of the 28 proteins, 19 were highly likely to be adhesin proteins with the adhesin probability ranging from 0.526 to 0.939.

Table 6.3: List of proteins selected based on their subcellular localisation, number of transmembrane helices (TMH) and antigenicity score, and their presence in the *S. aureus* isolates belonging to the major clonal complexes identified in this study.

Protein description	Subcellular Localisation	Adhesin probability	Number of TMH	Length (amino acids)	VaxiJen score
Bifunctional autolysin	Extracellular	0.787	0	1256	0.5414
Iron-regulated surface determinant protein B	Cell-wall	0.709	1	645	0.7235
Fibrinogen-binding protein	Extracellular	0.526	0	165	0.5307
hypothetical protein	Extracellular	0.752	1	152	1.2005
Nickel-binding periplasmic protein	Cell-wall	0.756	0	532	0.6999
Cell division protein DivIC	Cell-wall	0.141	1	130	0.687
hypothetical protein	Extracellular	0.546	0	105	0.5043
ComG operon protein 3	Cell-wall	0.467	1	103	0.7813
ESAT-6 secretion system extracellular protein A	Extracellular	0.593	0	97	0.5382
Fibronectin-binding protein A	Cell-wall	0.2	1	69	0.5395
Antitoxin MazE	Extracellular	0.569	0	56	0.6105
hypothetical protein	Extracellular	0.264	0	50	0.7715
Staphylococcal secretory antigen ssaA2	Extracellular	0.929	0	269	0.6952
hypothetical protein	Extracellular	0	0	34	0.867
Glutathione-binding protein GsiB	Cell-wall	0.644	0	491	0.5267
Alkaline phosphatase 3	Extracellular	0.784	0	474	0.5414

Continued on next page

Protein description	Subcellular Localisation	Adhesin probability	Number of TMH	Length (amino acids)	VaxiJen score
Cell division protein DivIB	Cell-wall	0.16	1	439	0.7737
Staphopain B	Extracellular	0.625	1	393	0.5797
Iron-regulated surface determinant protein A	Cell-wall	0.721	1	350	0.6946
putative leukocidin-like protein 1	Extracellular	0.782	0	338	0.5946
hypothetical protein	Cell-wall	0.151	0	323	0.5034
N-acetylmuramoyl-L-alanine amidase sle1	Cell-wall	0.849	0	265	0.6789
putative transglycosylase IsaA	Extracellular	0.928	0	233	0.8745
putative transglycosylase SceD	Extracellular	0.939	0	231	0.9649
Superoxide dismutase [Mn/Fe] 2	Extracellular	0.526	0	199	0.5284
Superoxide dismutase [Mn] 1	Extracellular	0.443	0	199	0.5903
Signal peptidase IB	Cell-wall	0.268	1	191	0.5887
hypothetical protein	Extracellular	0.82	0	166	0.6002

6.3.5 PROTEIN SELECTION BASED ON VAXIGN-ML PROTEGENICITY SCORE AND SIMILARITY TO HUMAN PROTEINS

The Vaxign-ML protegenicity score of the 28 shortlisted proteins ranged from 60.9 to 99.7. The Vaxign-ML scores of the 28 proteins are reported in Table 6.4. Twenty-one had a Vaxign-ML protegenicity score above 90. The 7 proteins that scored lower than 90 were removed from further analysis.

Out of the 21 proteins with a Vaxign-ML score above 90, three were similar to human proteins. The three proteins were Alkaline phosphatase 3, Superoxide dismutase 1 and Superoxide dismutase 2. None of the remaining 18 proteins had the threshold similarity or coverage with the host proteins.

Out of those 18 proteins, 7 were cell-wall proteins and 11 were extracellular proteins. The predicted adhesin probability of two cell wall proteins, the cell division protein DivIB (0.16) and a hypothetical protein (0.15), were less than the threshold of 0.5 and thus less likely to be an adhesin. The rest of the 16 proteins had high adhesin probabilities ranging from 0.53 to 0.94. There were four hypothetical proteins, three putative proteins and 11 well-known proteins. The protein sequences of these PVCs are provided in Appendix A.1

Table 6.4: List of the proteins selected along with their Vaxign-ML score, localisation, adhesion probability, transmembrane helices (TMH) and their similarity to the human protein. The proteins discarded in the step with lower Vaxign-ML score and their similarity to human proteins are italicised.

Protein Name	Vaxign-ML Score	Localisation (Probability)	Adhesin Probability	TMH	Similar Human Protein
Bifunctional autolysin	99.7	Extracellular (1.00)	0.787	0	No
Glutathione-binding protein GsiB	99.6	Cell-wall (0.92)	0.644	0	No
Alkaline phosphatase 3	*99.5*	*Extracellular (0.97)*	*0.784*	*0*	*Yes*
Iron-regulated surface determinant protein B	99.5	Cell-wall (1.00)	0.709	1	No
Staphopain B	98.2	Extracellular (1.00)	0.625	1	No
putative leukocidin-like protein 1	97.8	Extracellular (0.97)	0.782	0	No
Nickel-binding periplasmic protein	97.8	Cell-wall (0.92)	0.756	0	No
Iron-regulated surface determinant protein A	95.1	Cell-wall (1.00)	0.721	1	No
ESAT-6 secretion system extracellular protein A	94.8	Extracellular (1.00)	0.593	0	No
putative transglycosylase SceD	93.1	Extracellular (1.00)	0.939	0	No

Continued on next page

Protein Name	Vaxign-ML Score	Localisation (Probability)	Adhesin Probability	TMH	Similar Human Protein
Staphylococcal secretory antigen ssaA2	92.8	Extracellular (1.00)	0.929	0	No
Fibrinogen-binding protein	92.8	Extracellular (1.00)	0.526	0	No
putative transglycosylase IsaA	91.4	Extracellular (1.00)	0.928	0	No
Cell division protein DivIB	90.9	Cell-wall (0.88)	0.16	1	No
hypothetical protein	90.9	Cell-wall (1.00)	0.151	0	No
N-acetylmuramoyl-L-alanine amidase sle1	90.9	Cell-wall (0.92)	0.849	0	No
Superoxide dismutase [Mn/Fe] 2	*90.9*	*Extracellular (0.97)*	*0.526*	*0*	*Yes*
Superoxide dismutase [Mn] 1	*90.9*	*Extracellular (0.97)*	*0.443*	*0*	*Yes*
hypothetical protein	90.9	Extracellular (1.00)	0.82	0	No
hypothetical protein	90.9	Extracellular (0.91)	0.752	1	No
hypothetical protein	90.9	Extracellular (0.97)	0.546	0	No

Continued on next page

Protein Name	Vaxign-ML Score	Localisation (Probability)	Adhesin Probability	TMH	Similar Human Protein
Antitoxin MazE	*89.3*	*Extracellular (0.89)*	*0.569*	*0*	*No*
hypothetical protein	*88*	*Extracellular (0.89)*	*0.264*	*0*	*No*
Fibronectin-bindign protein A	*80.5*	*Cell-wall (1.00)*	*0.2*	*1*	*No*
Cell division protein DivIC	*76.3*	*Cell-wall (0.88)*	*0.141*	*1*	*No*
ComG operon protein 3	*76.1*	*Cell-wall (0.82)*	*0.467*	*1*	*No*
hypothetical protein	*68.7*	*Extracellular (0.89)*	*0*	*0*	*No*
Signal peptidase IB	*60.9*	*Cell-wall (0.88)*	*0.268*	*1*	*No*

6.3.6 EPITOPE PREDICTION FROM SELECTED PROTEINS

The number of 20-mer B-cell epitopes and the ratio of the number of epitopes to protein length as predicted by BCPreds and AAP algorithm for the 18 PVC proteins are listed in Table 6.5. The number of epitopes ranged from 1 to 37 among the 18 potential vaccine candidate proteins when predicted using the BCPreds algorithm, while the AAP algorithm predicted between 0 and 40 epitopes. The epitopes of all these 18 PVCs are listed in Appendices A.3 and A.4.

The proteins were ranked in descending order based on their epitope density (ratio of the number of epitopes to the amino acids in the protein) as predicted by BCPreds, as high epitope density often increases the antigenicity and immunogenicity of the protein (Liu and Chen, 2005). The proteins were sorted first on BCPreds epitope density and then AAP epitope density.

Table 6.5: List of potential vaccine candidate proteins selected in this study ranked based on their epitope density calculated from BCPreds epitope prediction and AAP epitope prediction along with their localisation, adhesin probability, number of transmembrane helices (TMH), length in amino acids, VaixJen score of antigenicity, Vaxign-ML score and the number of epitopes and their ratio based on BCPREDS and AAP epitope prediction.

Rank	Protein description	Subcellular Localisation	Adhesin probability	TMH	Length (amino acids)	VaxiJen score	Vaxign-ML score	BCPreds		AAP	
								No. of epitopes	Ratio	No. of epitopes	Ratio
1	Staphylococcal secretory antigen ssaA2	Extracellular	0.929	0	269	0.6952	92.8	9	0.033	8	0.030
2	hypothetical protein C	Extracellular	0.752	1	152	1.2005	90.9	5	0.033	4	0.026
3	putative transglycosylase SceD	Extracellular	0.939	0	231	0.9649	93.1	7	0.030	8	0.035
4	putative transglycosylase IsaA	Extracellular	0.928	0	233	0.8745	91.4	7	0.030	7	0.030
5	N-acetylmuramoyl-L-alanine amidase sle1	Cell-wall	0.849	0	265	0.6789	90.9	8	0.030	7	0.026
6	Bifunctional autolysin	Extracellular	0.787	0	1256	0.5414	99.7	37	0.029	40	0.032
7	Iron-regulated surface determinant protein A	Cell-wall	0.721	1	350	0.6946	95.1	10	0.029	9	0.026
8	putative leukocidin-like protein 1	Extracellular	0.782	0	338	0.5946	97.8	9	0.027	9	0.027

Continued on next page

Rank	Protein description	Subcellular Localisation	Adhesin probability	TMH	Length (amino acids)	VaxiJen score	Vaxign-ML score	BCPreds		AAP	
								No. of epitopes	Ratio	No. of epitopes	Ratio
9	Iron-regulated surface determinant protein B	Cell-wall	0.709	1	645	0.7235	99.5	17	0.026	20	0.031
10	Cell division protein DivIB	Cell-wall	0.16	1	439	0.7737	90.9	11	0.025	13	0.030
11	Staphopain B	Extracellular	0.625	1	393	0.5797	98.2	10	0.025	8	0.020
12	hypothetical protein A	Extracellular	0.82	0	166	0.6002	90.9	4	0.024	4	0.024
13	Fibrinogen-binding protein	Extracellular	0.526	0	165	0.5307	92.8	4	0.024	2	0.012
14	ESAT-6 secretion system extracellular protein A	Extracellular	0.593	0	97	0.5382	94.8	2	0.021	2	0.021
15	Glutathione-binding protein GsiB	Cell-wall	0.644	0	491	0.5267	99.6	10	0.020	12	0.024
16	Nickel-binding periplasmic protein	Cell-wall	0.756	0	532	0.6999	97.8	10	0.019	11	0.021
17	hypothetical protein D	Extracellular	0.546	0	105	0.5043	90.9	2	0.019	2	0.019
18	hypothetical protein B	Cell-wall	0.151	0	323	0.5034	90.9	5	0.015	4	0.012

6.3.7 PROTEIN SEQUENCE VARIABILITY

Results of the blastp process to assess conservation of the protein sequences across proteome of the isolates belonging to major CCs are tabled in https://github.com/jabinnes/Thesis-files/blob/main/Reverse%20vaccinology/Protein%20conservation%20results.xlsx. The sequence logos visualisations of the protein sequence variability of all the PVCs are documented in https://github.com/jabinnes/Thesis-files/tree/main/Reverse%20vaccinology/Sequence%20logos%20images. The sequence variability was mostly CC-specific with the best hits of the blastp ranging from 88.5% to 100%. The proteins were highly conserved in the 140 CC1 isolates with no mismatches for 13/18 (72%) PVCs and even with the three PVCs with mismatches in CC1 isolates, the maximum number of mismatches was four. CC151 was the most variable compared to the PVC proteins (identified in CC1 isolate) with one of its isolate having a best hit of 88.5% to Staphylococcal secretory antigen ssaA2 and the most number of mismatches (20) to Iron-regulated surface determinant protein B.

6.3.8 DESCRIPTION OF THE SELECTED PROTEINS

The functions of proteins selected as PVCs are detailed below, emphasising their potential role in the virulence or antigenicity of *S. aureus*.

6.3.8.1 BIFUNCTIONAL AUTOLYSIN (Atl)

Bifunctional autolysin (Atl), also known as the major autolysin, is a well-known surface protein of *S.aureus* (Komatsuzawa et al., 1997). Atl, encoded by the *atl* gene, is a peptidoglycan hydrolase involved in peptidoglycan degradation. The process of peptidoglycan degradation of the cell-wall is essential for bacterial growth and division, cell-wall degradation and cell division (Yamada et al., 1996). The Atl protein has two domains, amidase (AM) and glucosaminidase (GM) (Komatsuzawa et al., 1997). The Atl protein is synthesised as a single precursor protein and is cleaved to produce the active AM and GM protein domains. The protein also plays an important role in biofilm development and binds to host extracellular proteins such as

fibronectin and vitronectin (Biswas et al., 2006). Increased immune response was found in mice vaccinated with recombinant autolysin. However, this result was not followed by bacterial challenge studies indicating the need to assess for a protective immune response of the vaccine (Haghighat et al., 2017). It was also found that vaccinating with autolysin improved antimicrobial treatment outcomes and also decreased the biofilm formation of *S. aureus* (Brady et al., 2011).

6.3.8.2 CELL DIVISION PROTEIN (DivIB)

The DivIB is involved in the cell division process of Gram-positive bacteria. Most of the studies regarding the function of DivIB were done on *Bacillus subtilis*. The DivIB protein is a peptidoglycan binding protein essential for completion of the cell septum during cell division (Bottomley et al., 2014). Binding of two other cell division proteins (DivIC and FtsL) to form a complex with DivIB is needed for the completion of cell septum formation (Noirclerc-Savoye et al., 2005).

6.3.8.3 ESAT-6 SECRETION SYSTEM EXTRACELLULAR PROTEIN A (EsxA)

The extracellular protein A (EsxA) is one of the proteins secreted by the ESAT-6 (also known as Type VII) secretion system (Burts et al., 2005). The EsxA protein, along with EsxB, modulates host cell apoptosis and the release of *S. aureus* from the host cells (Korea et al., 2014). EsxA and EsxB also control the human dendritic cell, a host immune regulator, by inducing apoptosis and thereby affecting the host immune response (Cruciani et al., 2017). *S. aureus* mutants lacking EsxA were reported to have a significantly reduced ability to establish tissue abscesses (Burts et al., 2005). Interestingly, the EsxA was found to be highly immunogenic in bovine mastitis (Misra et al., 2018). Vaccination with EsxA and four other proteins provided effective protection against *S. aureus* infection in a mouse model (Deng et al., 2019) and EsxA was found to induce T-cell based immune response (Zhang et al., 2015). The T helper cells have been reported to aid in clearing *S. aureus* in the mouse model (Zhao et al., 2015).

6.3.8.4 FIBRINOGEN BINDING PROTEIN (Efb)

The Fibrinogen binding protein (Efb) encoded by the *efb* gene is a secreted protein that forms a protective barrier over the organism by binding with host fibrinogen and plasma proteins to prevent phagocytosis of the bacteria by neutrophils (Ko et al., 2013). High levels of antibodies were reported against Efb in dairy cattle immunised with recombinant Efb protein (Boerhout et al., 2015). Vaccination with fibrinogen binding protein in a mouse mastitis model showed a decrease in the colonisation of the mammary gland by *S. aureus* (Mamo et al., 1994).

6.3.8.5 GLUTATHIONE BINDING PROTEIN (GsiB)

The glutathione binding protein (GsiB) is a transmembrane protein involved in glutathione transport into bacterial cells. The GsiB protein is part of the Glutathione importer (Gsi) containing GsiA (ATP binding protein), GsiC and GsiD (two inner membrane components) (Suzuki et al., 2005; Keseler et al., 2009; Wang et al., 2018a).

6.3.8.6 IRON-REGULATED SURFACE DETERMINANT PROTEINS (IsdA and IsdB)

Iron-regulated surface determinant proteins (Isd) are a cell-wall protein involved in the transfer of haem from haemoglobin (Skaar and Schneewind, 2004). One of the Isd determinants, the IsdA, encoded by the *isdA* gene, is an adhesin that binds with the host's haem as well as fibrinogen and fibronectin (Clarke et al., 2004). The IsdA protects *S. aureus* against host immune defences (Clarke et al., 2007; Clarke and Foster, 2008). The IsdA is expressed in bovine mastitis and is highly immunogenic (Misra et al., 2017). In the study by Misra et al. (2017) another Isd factor, IsdB was also identified as a potential vaccine candidate against bovine *S. aureus* mastitis. The IsdB protein is a predominant receptor for haemoglobin to enable the transfer of haem by the other Isd factors (Fonner et al., 2014). Along with Isd factors, IsdB was found to provide resistance against peroxidases and thus may be essential in *S. aureus* evasion of destruction within the host phagocytic cells (Palazzolo-Ballance et al., 2008). Immunisation using IsdA and IsdB conferred protection against *S. aureus* cleared the bacteria in infected animal models (Stranger-Jones et al., 2006), and

induced a cell-mediated immune response (Arlian and Tinker, 2011; Joshi et al., 2012). Vaccination with IsdA along with ClfA and cholera toxin elicited a significant immune response in bovine trials, but the protective response was not studied with challenge studies (Misra et al., 2018).

6.3.8.7 N-ACETYLMURAMOYL-L-ALANINE AMIDASE (Sle1)

The N-acetylmuramoyl-L-alanine amidase sle1 (Sle1) encoded by the *sle1* gene is involved in the cell separation by splitting the septum during cell division using its peptidoglycan hydrolase activity (Palazzolo-Ballance et al., 2008). The activity of Sle1 is similar to the AM fraction of the Atl protein.(See section 3.7.1) The Sle1 is also reported to be essential for β-lactam activity on *S. aureus* (Thalsø-Madsen et al., 2019).

6.3.8.8 NICKEL BINDING PERIPLASMIC PROTEIN (NikA)

The nickel binding periplasmic protein (NikA) encoded by the *nikA* gene is a membrane transporter competing for nickel in the host environment (Carlson et al., 2020). However, the importance of nickel for *S. aureus* pathogenesis in bovine mastitis has not been studied yet.

6.3.8.9 STAPHOPAIN B (SspB)

Staphopain B (SspB) encoded by the *sspB* gene is one of the cysteine proteases secreted by *S. aureus*. The SspB induces apoptosis in host neutrophils and monocytes thus protecting *S. aureus* from phagocytosis (Smagur et al., 2009a; Elmwall et al., 2017). SspB also induces phagocytosis of host neutrophils by the host macrophages (Smagur et al., 2009b).

6.3.8.10 STAPHYLOCOCCAL SECRETORY ANTIGEN (ssaA2)

The Staphylococcal secretory antigen (ssaA2) function is unknown, but the protein has been identified as an immunogenic protein (Lang et al., 2000).

6.3.8.11 HYPOTHETICAL PROTEINS

Four hypothetical proteins were identified as potential vaccine candidates in this study. One of the proteins (hypothetical protein A) was described as a surface protein (COG3942) by COG analysis, but the function was unknown. Hypothetical protein B was labelled as the domain of unknown function 1542 (DUF1542). The DUF1542 domain has been identified as part of surface proteins in *S. aureus* (Schroeder et al., 2009). The specific function of the DUF1542 has not been elucidated. The other two hypothetical proteins (hypothetical protein C and D) did not have a COG description.

6.3.8.12 PUTATIVE PROTEINS (Transglycosylase IsaA, SceD, leukocidin-like protein)

The two putative transglycosylases, IsaA and SceD are reported to have hydrolytic activity on the peptidoglycan. The inactivation of IsaA increased the expression of SceD, and SceD activity was increased in the presence of sodium chloride (Stapleton et al., 2007). This could explain the alteration of the cell size and permeability in the presence of NaCl (Vijaranakul et al., 1995). However, the function of these proteins in the pathogenesis of bovine mastitis has not been studied. Vaccination using the transglycosylase IsaA generated a strong immune response in mouse models, but no protective immunity was noticed against *S. aureus* challenge (van den Berg et al., 2015). However, passive immunisation using monoclonal antibodies of IsaA were found to be protective in challenge models (Lorenz et al., 2011). The exact function of the putative leukocidin-like protein has not been elucidated, but due to the similarity with the leukocidins, its function could be associated with the lysis of neutrophils (Spaan et al., 2017).

6.4 DISCUSSION

Developing an effective vaccine against *S. aureus* has been difficult as *S. aureus* is very proficient in evading host immune responses. The genetic structure and clonal diversity of bovine *S. aureus* in New Zealand has been stable over the last 15 years, as indicated by the study presented in Chapter 3. Hence, a recombinant vaccine using a combination of proteins identified in this study through RV could theoretically enable immunity against the majority of the strains circulating in New Zealand. This study has identified 18 potential vaccine candidate proteins. Eleven of these are proteins of known function, most of which have been implicated in *S. aureus* virulence *in vitro*. However, there have been no detailed studies of these proteins in the context of bovine mastitis. Some of the proteins identified in this study have been previously used as recombinant proteins for vaccination against either human *S. aureus* disease or bovine *S. aureus* mastitis.

The proteins identified as PVCs in this study were highly conserved and CC-specific. This CC-specificity demonstrates the need for a population genetic structure study of *S. aureus* and strain-based vaccine design. The immunological studies mentioned in section 6.3.8 corroborate that many of the proteins found in the current bioinformatic study are immunogenic proteins. Proteins with high epitope density have high antigenicity and immunogenicity (Liu and Chen, 2005), and thus, the proteins were ranked based on their epitope density to assist in selecting the proteins for future laboratory analysis for vaccine testing.

It has been postulated that vaccination with a combination of antigens provides better protection against *S. aureus* infections than vaccination with a single antigen (Stranger-Jones et al., 2006). Hence, some of the vaccine candidates identified in this study could be combined for vaccination against bovine *S. aureus* mastitis.

The current study has limitations. Firstly, there is no data indicating whether the selected PVCs are expressed during bovine mastitis. More data is needed on the role of these PVCs in bovine mastitis as currently there is only limited understanding of the role of these proteins in the pathogenesis of bovine mastitis. Although some of the vaccine candidates are known to be immunogenic, the protection against new *S. aureus* IMI via vaccination has yet to be established. The inherent limitation of reverse vaccinology to identify only protein candidates excludes polysaccharides and

other non-protein based vaccine candidates. The conservation of protein sequences among epitopes were not identified as this was beyond the scope of this study.

In summary, the population genetic study presented in Chapter 3 revealed a clonal *S. aureus* population causing bovine mastitis in New Zealand, which is stable in time and space, enabling the present study. The bioinformatic RV study presented here identified universally occurring genes coding for PVCs. The immunogenicity of some of these PVCs has been previously established in *in vitro* and *in vivo* models.

7

GENERAL DISCUSSION

Bovine mastitis is one of the most costly diseases affecting the dairy industry worldwide. In New Zealand, *S. aureus* is the second most common bacterial species isolated from bovine milk submitted to clinical microbiology laboratories for bacteriological analysis (Bates et al., 2020; McDougall, 2010). *S. aureus* is a highly clonal organism (Lindsay, 2010), and the clonal complex (CC), based on multilocus sequence typing (MLST), has been the most common operational taxonomic unit used to resolve the genetic variation among *S. aureus* strains in the last two decades (Enright et al., 2002; Kuhn et al., 2006). To date, no detailed studies have been published describing the population genetic structure of *S. aureus* causing bovine mastitis through the use of whole-genome sequencing (WGS) in New Zealand. This PhD project used this approach to analyse a relatively large sample of stored isolates obtained over 15 years.

The study presented in Chapter 3 represents the first step of the analysis, where WGS was used to explore the clonal composition of bovine *S. aureus*. Two decades ago, prior to the advent of Next Generation Sequencing and the development of bioinformatics pipelines for bacterial whole genome analysis, embarking on the seemingly simple MLST analysis of 200 isolates would have represented an enormous task requiring a total of 200 x 7 = 1,400 PCR reactions, followed by an equal number of Sanger sequencing reactions and manual editing of each sequencing output. Two hundred PCR-sequencing reactions would have been required to analyse the Spa types types. This is without mentioning the immense number of reactions that would have been required to determine virulence factors and antimicrobial resistance genes. Such a study would have not been feasible within the framework of a PhD project.

In many countries, bovine-adapted sequence types (STs), such as ST97, ST151, or ST133, which are not widely observed in humans, prevail in bovine mastitis. The mechanisms of host adaptation are not entirely clear, although *in vitro* studies of representative isolates demonstrated that some of these STs carry genes that code for ruminant-adaptive virulence factors, such as the ruminant-specific leukocidins. Data shown in Chapter 3 indicate these ruminant-specific STs are not dominant in New Zealand. Instead, CC1, a human-associated CC, and in particular ST1, is the dominant clonal lineage, accounting for almost 75% of the isolates. Moreover, the dominance of CC1/ST1 was observed on both the North and South Islands, and was stable over the 15 years of observation. This finding was interesting for a number of reasons. Firstly, CC1/ST1 is also the most common lineage infecting humans and pets

in New Zealand (Heffernan et al., 2015). Moreover, there was a common spa-type (t127) among cattle and human CC1 isolates. Secondly, CC1 is a known human-associated clonal complex, described less commonly in cattle worldwide (Thomas et al., 2021; Boss et al., 2016; Hoekstra et al., 2020). CC1 has been found in some areas of China and Australia, but not as a predominant strain (O'Dea et al., 2020; Wang et al., 2018b). This dominance of a single clonal lineage, CC1, in both bovine and humans in New Zealand, led to questioning whether there is a differentiation of host-adaptated lineages within this CC, similar to the host adaptation observed within CC5, to humans, or poultry (Murray et al., 2017).

The comparative genomics study presented in Chapter 4 was undertaken mainly to analyse the genomes of CC1 for the presence of genomic differentiation between the human and bovine samples. Such a study is crucial to understand cross transmission between the human and bovine hosts. Interestingly, the CC1 isolates clustered based on the host species at both core and accessory genome levels, and the lone small ruminant CC1 isolate clustered with the bovine isolates, demonstrating the existence of 'ruminant adaptation' within CC1, as previously demonstrated for ST97 and other 'typical' ruminant STs. Strikingly, analysis of individual genes also revealed the presence of host-specific/adaptive accessory genes in ruminant CC1 isolates, such as the lukM and lukF'-PV, which were absent in human isolates, and *vice-versa*, the absence of a number of human-associated genes (scn, sak, chp, lukF-PV, lukS-PV, sea) predominantly found in human CC1 isolates (Matuszewska et al., 2020; Sung et al., 2008). Analysis of the resistome in CC1 isolates also revealed host-specific signatures, with a high prevalence of *blaZ* (penicillin resistance) genes in human isolates (75%), when compared to cattle isolates (13%). In fact, penicillin resistance is almost universal in human *S. aureus* in New Zealand, whereas in cattle, the prevalence of penicillin resistance is about 20% (Heffernan et al., 2015; Petrovski et al., 2015).

The comparative study of antimicrobial resistome (AMR) and disk diffusion (DD) antimicrobial susceptibility testing presented in Chapter 5 was undertaken considering the potential use of resistome analysis to predict phenotypic resistance, particularly in surveys and monitoring of resistance. In the current study, all the isolates harbouring the *blaZ* gene were phenotypically resistant to penicillin, and WGS-AMR predicted *in vitro* penicillin resistance with a specificity of 100% and a sensitivity of 71%. Only a few phenotypic resistances were identified for other antimicrobials classes, with fewer resistance genes for them. The lower sensitivity

when compared to the studies by Bradley et al. (2015b) (a specificity of 88.3% and a sensitivity of 99.7%) and Gordon et al. (2014) (a specificity of 100% and a sensitivity of 99.1%) could be explained by variation in the study design, especially the inclusion of short or low coverage contigs for *blaZ* gene as Gordon et al. (2014) found a high number of isolates with susceptible genotype with a resistant phenotype.

Finally, the stable population structure characterised by the dominance of a single clonal lineage (CC1), followed by the presence of a limited number of ruminant-associated lineages such as ST97 and ST151, and the lack of evidence for clonal shifts over the 15 years of observation, enabled the development of the reverse vaccinology pipeline reported in Chapter 6. Eighteen potential vaccine candidates (PVCs) were shortlisted using established bioinformatic tools. These PVCs were either cell wall, or extracellular proteins present in all the isolates belonging to the major CCs infecting cattle (CC1, CC97, CC8 and CC151). The amino acid sequences of these PVCs were conserved within CCs. Some of these PVCs have been previously studied and found to be immunogenic in humans and cattle (Brady et al., 2011; Joshi et al., 2012; Deng et al., 2019). Further *in vitro* and *in vivo* studies will determine whether they can be used effectively in vaccines for cattle, particularly if their use can elicit the necessary immune response to protect against intramammary infections.

This research project has limitations, most of which have been discussed in the individual chapters. The main limitation was the limited metadata available on the isolates. The lack of clinical data did not allow associations between genetic markers and disease severity to be established. Establishing such associations would enable the identification of bacterial genetic markers of prognostic value. This could be an interesting research project for the future. The unavailability of farm addresses precluded ruling out the presence of multiple isolates from the same farm in the different collection periods, which could have biased some of the statistical analysis. This limitation is difficult to resolve when utilising isolates obtained from veterinary laboratories, due to the need to maintain confidentiality.

Despite the new findings, new research questions were also raised at the end of this study. Firstly, are there any geographically-specific adaptations of *S. aureus* in New Zealand that have enabled the dominance of CC1? Further studies are needed to understand the association of CC1 isolates from cattle in New Zealand with those identified worldwide to recognise any genomic adaptation. Secondly, how does the

disease severity caused by CC1 lineage strains in dairy cattle compare to other 'ruminant adapted' CCs? Investigations with clinical metadata would shed light on this question. Thirdly, if the isolates are from the same niche in different hosts, will there be any host-specific genes? Studies with isolates from cattle and humans affected by mastitis would help in answering that question. Nevertheless, the main prospective study would be to continue to analyse the 18 shortlisted PVCs using conventional *in vitro* studies such as the production of mapped epitopes through recombinant technology and efficacy studies through the use of antisera against the PVCs against *S. aureus in vitro*. Immunological studies are also needed to determine if natural antibodies are produced against these 18 PVCs in bovine milk. Finally, *in vivo* studies are needed to ascertain if these PVCs, either as a single or combined subunit vaccine, protect against new *S. aureus* intramammary infection or clear existing *S. aureus* infections.

In conclusion, using data from a period spanning over 15 years, this is the first comprehensive genomic epidemiology study of bovine *S. aureus* in New Zealand and one of a few such studies worldwide. The dominance in New Zealand of CC1 in both cattle and humans has not been previously reported in any other country. This has allowed us to identify host-specific/adaptive genes in CC1 in cattle and humans. This research has also demonstrated the value of WGS as a method of identifying PVCs using reverse vaccinology methodology.

Appendix A

APPENDIX

Table A.1: Selected QUAST metrics of the 188 bovine *S. aureus* genomes. The isolates designation (assigned names) is composed of five parts separated by an underscore. The first part denotes the isolate's serial number, the second part is a unique farm identifier, the third indicates the year of collection, the fourth is the sequence type (ST), and the last is clonal complex (CC). For example, isolate B064_F51_Y13_ST1_CC1 is serial number B64, it was isolated from farm 51 in 2013, and it belongs to ST1, CC1.

Unique Lab Identifier	**Assigned names**	**No of contigs**	**Largest contig**	**Total length**	**GC (%)**	**N50**	**L50**
BM109	B064_F51_Y13_ST1_CC1	62	314,882	2,754,728	32.77	97,610	9
BM110	B060_F48_Y13_ST1_CC1	26	491,105	2,763,202	32.76	287,602	4
BM111	B061_F48_Y13_ST1_CC1	25	567,876	2,763,079	32.76	207,181	4
BM112	B057_F45_Y13_ST1_CC1	36	385,067	2,765,689	32.75	175,544	6
BM113	B052_F41_Y13_ST1_CC1	25	436,806	2,721,767	32.72	253,836	4
BM114	B049_F38_Y13_ST78_CC78	27	691,658	2,726,968	32.68	417,220	3
BM116	B053_F06_Y13_ST1_CC1	24	490,289	2,750,761	32.69	205,220	5
BM117	B054_F42_Y13_ST1_CC1	37	349,125	2,765,731	32.76	145,856	6
BM118	B045_F34_Y13_ST1_CC1	46	368,837	2,788,066	32.78	136,345	6
BM119	B059_F47_Y13_ST1_CC1	26	623,913	2,766,595	32.76	196,362	5
BM123	B047_F36_Y13_ST97_CC97	41	395,255	2,790,039	32.75	150,245	6
BM124	B043_F33_Y13_ST1_CC1	61	234,638	2,818,447	32.74	145,543	8
BM126	B044_F33_Y13_ST1_CC1	32	286,587	2,728,918	32.72	148,077	7
BM128	B040_F30_Y13_ST4551_CC1	23	482,061	2,721,529	32.74	240,935	4
BM129	B036_F137_Y13_ST1_CC1	79	304,872	2,718,308	32.74	95,993	10

Continued on next page

Unique Lab Identifier	Assigned names*	No of contigs	Largest contig	Total length	GC (%)	N50	L50
BM130	B037_F28_Y13_ST1_CC1	33	495,935	2,858,814	32.74	241,078	5
BM131	B063_F50_Y13_ST1_CC1	64	210,211	2,761,161	32.77	88,799	11
BM132	B041_F31_Y13_ST1_CC1	17	710,952	2,704,123	32.74	349,232	3
BM133	B051_F40_Y13_ST1_CC1	44	314,912	2,806,070	32.8	138,786	7
BM134	B038_F29_Y13_ST1_CC1	52	385,094	2,761,487	32.75	113,171	7
BM136	B039_F29_Y13_ST1_CC1	40	314,823	2,761,077	32.75	146,005	7
BM137	B046_F35_Y13_ST97_CC97	33	338,476	2,785,197	32.66	193,918	6
BM33	B001_F01_Y13_ST6143_CC8	57	278,580	2,774,290	32.62	141,787	8
BM34	B002_F01_Y13_ST6143_CC8	51	582,983	2,773,006	32.62	141,300	5
BM35	B003_F02_Y13_ST1_CC1	43	318,535	2,761,627	32.74	157,063	7
BM36	B004_F03_Y13_ST1_CC1	35	330,824	2,767,359	32.76	147,978	7
BM37	B005_F04_Y13_ST97_CC97	50	312,043	2,756,477	32.68	128,369	8
BM38	B006_F05_Y13_ST1_CC1	48	314,890	2,815,245	32.79	215,364	6
BM39	B007_F02_Y13_ST1_CC1	52	206,390	2,763,789	32.75	116,560	9
BM40	B008_F05_Y13_ST1_CC1	54	222,284	2,763,454	32.77	103,940	9
BM41	B009_F07_Y13_ST1_CC1	60	210,432	2,755,987	32.74	78,426	12
BM42	B010_F07_Y13_ST1_CC1	41	298,562	2,764,882	32.75	144,573	8
BM43	B011_F03_Y13_ST1_CC1	21	528,736	2,766,711	32.76	241,813	4
BM45	B012_F08_Y13_ST1_CC1	33	314,829	2,721,127	32.72	204,401	6

Continued on next page

Unique Lab Identifier	Assigned names*	No of contigs	Largest contig	Total length	GC (%)	N50	L50
BM46	B013_F09_Y13_ST6143_CC8	61	276,498	2,774,006	32.62	109,220	8
BM49	B014_F10_Y13_ST1_CC1	29	560,453	2,767,294	32.75	245,960	4
BM51	B016_F12_Y13_ST1_CC1	61	224,773	2,717,288	32.7	107,283	10
BM52	B017_F134_Y13_ST1_CC1	28	717,266	2,765,499	32.78	240,249	4
BM53	B018_F13_Y13_ST1_CC1	43	221,073	2,759,815	32.74	139,360	9
BM54	B019_F14_Y13_ST97_CC97	56	307,535	2,764,446	32.69	128,691	8
BM55	B020_F15_Y13_ST1_CC1	25	562,568	2,762,747	32.77	240,960	4
BM57	B021_F16_Y13_ST1_CC1	22	363,676	2,749,994	32.77	220,105	5
BM58	B022_F17_Y13_ST1_CC1	26	427,790	2,824,123	32.69	167,631	6
BM59	B023_F18_Y13_ST199_CC15	48	476,169	2,792,015	32.78	232,228	5
BM60	B024_F19_Y13_ST1_CC1	42	526,635	2,844,613	32.74	241,134	4
BM61	B025_F13_Y13_ST97_CC97	27	458,894	2,791,526	32.73	197,756	5
BM63	B026_F20_Y13_ST97_CC97	28	515,455	2,759,741	32.67	201,557	5
BM64	B027_F21_Y13_ST1247_CC133	22	406,616	2,689,222	32.77	197,991	5
BM65	B028_F22_Y13_ST1_CC1	20	459,150	2,732,589	32.7	241,184	4
BM66	B029_F23_Y13_ST97_CC97	31	503,175	2,742,328	32.72	151,658	4
BM67	B030_F24_Y13_ST1_CC1	16	1,223,753	2,776,753	32.76	527,369	2
BM69	B031_F25_Y13_ST8_CC8	21	1,104,801	2,809,598	32.62	489,475	2
BM71	B032_F26_Y13_ST1_CC1	74	174,182	2,761,959	32.79	79,035	12

Continued on next page

Unique Lab Identifier	Assigned names*	No of contigs	Largest contig	Total length	GC (%)	N50	L50
BM72	B033_F08_Y13_ST1_CC1	22	834,895	2,842,477	32.75	314,798	3
BM73	B034_F19_Y13_ST97_CC97	44	263,682	2,786,290	32.75	122,359	9
BM75	B035_F27_Y13_ST1_CC1	37	389,919	2,763,976	32.77	139,757	6
BM76	B058_F46_Y13_ST1_CC1	34	792,851	2,796,976	32.79	240,982	4
BM77	B055_F43_Y13_ST1_CC1	18	621,355	2,721,688	32.72	323,792	3
BM81	B050_F39_Y13_ST1_CC1	30	557,587	2,749,482	32.73	217,648	4
BM85	B062_F49_Y13_ST97_CC97	41	326,699	2,776,363	32.74	154,963	6
BM87	B042_F32_Y13_ST1_CC1	20	1,015,894	2,766,050	32.77	272,938	3
BM89	B048_F37_Y13_ST1_CC1	29	413,014	2,766,078	32.76	200,771	5
BM90	B056_F44_Y13_ST1_CC1	21	1,072,182	2,766,844	32.77	363,589	2
BM92	B065_F52_Y13_ST1_CC1	78	155,173	2,790,988	32.75	88,684	14
H01	B066_F54_Y18_ST1_CC1	59	196,450	2,700,337	32.73	93,364	11
H02	B067_F55_Y18_ST97_CC97	58	216,481	2,740,234	32.62	87,061	11
H03	B068_F56_Y18_ST151_CC151	44	302,482	2,731,253	32.71	135,272	7
H04	B069_F55_Y18_ST97_CC97	75	252,566	2,741,029	32.64	67,643	14
H05	B070_F54_Y18_ST1_CC1	58	196,370	2,702,217	32.73	103,024	10
H06	B071_F57_Y18_ST151_CC151	56	334,767	2,730,158	32.71	106,992	8
H07	B072_F58_Y18_ST5367_CC5367	99	142,429	2,707,597	32.71	60,054	16
H08	B076_F59_Y18_ST1_CC1	65	231,523	2,720,330	32.73	108,461	9

Continued on next page

Unique Lab Identifier	Assigned names*	No of contigs	Largest contig	Total length	GC (%)	N50	L50
H09	B073_F55_Y18_ST97_CC97	85	161,822	2,735,676	32.64	66,876	14
H10	B077_F60_Y18_ST97_CC97	100	146,940	2,732,514	32.68	58,929	17
H11	B074_F55_Y18_ST97_CC97	100	123,729	2,735,115	32.67	49,842	18
H12	B075_F58_Y18_ST5367_CC5367	73	303,570	2,710,410	32.67	73,786	10
H13	B078_F61_Y18_ST1_CC1	45	289,097	2,761,465	32.74	139,833	7
S010	B083_F63_Y18_ST1_CC1	47	282,088	2,757,742	32.73	142,929	8
S105	B110_F62_Y18_ST6160_CC97	128	223,993	2,805,656	32.77	68,106	14
S107	B116_F62_Y18_ST6160_CC97	133	120,155	2,792,096	32.75	46,770	20
S108	B117_F62_Y18_ST6160_CC97	86	146,711	2,795,011	32.71	87,225	13
S109	B118_F62_Y18_ST6160_CC97	63	396,507	2,795,287	32.72	136,696	7
S110	B120_F62_Y18_ST6160_CC97	78	213,262	2,797,750	32.73	79,148	13
S117	B126_F73_Y18_ST1_CC1	80	188,651	2,754,069	32.8	69,190	14
S118	B125_F72_Y18_ST1_CC1	70	216,454	2,756,720	32.76	75,397	11
S138	B122_F71_Y18_ST1_CC1	83	132,352	2,757,684	32.77	64,616	16
S142	B124_F69_Y18_ST1_CC1	95	125,843	2,752,942	32.79	57,754	17
S144	B123_F69_Y18_ST1_CC1	40	315,319	2,754,454	32.75	140,615	7
S015	B084_F64_Y18_ST1_CC1	47	282,618	2,768,289	32.78	114,269	8
S153	B111_F70_Y18_ST1_CC1	65	263,405	2,717,322	32.72	96,399	10
S154	B112_F70_Y18_ST1_CC1	58	168,936	2,721,710	32.71	94,526	12

Continued on next page

Unique Lab Identifier	Assigned names*	No of contigs	Largest contig	Total length	GC (%)	N50	L50
S155	B113_F70_Y18_ST1_CC1	74	246,040	2,807,184	32.78	95,345	10
S156	B119_F62_Y18_ST1_CC1	59	269,294	2,787,392	32.74	110,020	9
S159	B114_F66_Y18_ST1_CC1	72	168,343	2,757,775	32.76	76,666	13
S166	B121_F62_Y18_ST1_CC1	59	278,895	2,744,299	32.77	84,324	10
S171	B115_F65_Y18_ST1_CC1	73	251,146	2,761,930	32.71	167,671	7
S181	B127_F63_Y18_ST1_CC1	55	206,648	2,710,217	32.71	99,037	10
S182	B128_F63_Y18_ST1_CC1	51	170,308	2,712,604	32.72	107,693	10
S183	B129_F63_Y18_ST1_CC1	34	313,233	2,714,208	32.71	171,092	6
S002	B079_F62_Y18_ST6160_CC97	74	214,222	2,745,651	32.66	72,992	12
S021	B085_F65_Y18_ST1_CC1	59	275,866	2,766,558	32.76	97,139	9
S024	B086_F66_Y18_ST1_CC1	49	296,821	2,783,842	32.74	108,347	8
S027	B087_F65_Y18_ST1_CC1	38	330,072	2,765,026	32.75	169,896	6
S028	B088_F65_Y18_ST1_CC1	32	531,722	2,766,256	32.75	162,255	5
S021	B089_F65_Y18_ST1_CC1	47	253,354	2,768,136	32.76	107,794	9
S024	B090_F67_Y18_ST6141_CC1	66	168,502	2,758,140	32.76	94,806	11
S025	B091_F67_Y18_ST6141_CC1	52	248,298	2,767,382	32.77	107,759	9
S019	B092_F65_Y18_ST1_CC1	123	148,300	2,760,072	32.79	44,467	21
S020	B093_F65_Y18_ST1_CC1	72	175,013	2,761,913	32.75	83,817	11
S021	B094_F65_Y18_ST1_CC1	62	345,724	2,789,136	32.78	145,685	7

Continued on next page

Unique Lab Identifier	Assigned names*	No of contigs	Largest contig	Total length	GC (%)	N50	L50
S023	B095_F68_Y18_ST1_CC1	41	340,981	2,761,447	32.77	148,638	6
S024	B096_F68_Y18_ST1_CC1	43	285,670	2,768,703	32.76	108,363	9
S028	B097_F63_Y18_ST1_CC1	36	517,241	2,764,311	32.76	235,814	5
S006	B080_F62_Y18_ST6160_CC97	63	251,703	2,744,632	32.68	85,674	11
S015	B098_F69_Y18_ST1_CC1	43	246,217	2,751,920	32.75	135,841	8
S026	B099_F69_Y18_ST1_CC1	41	322,685	2,758,399	32.77	189,069	6
S007	B081_F62_Y18_ST6160_CC97	43	357,090	2,748,850	32.67	105,275	8
S029	B100_F64_Y18_ST1_CC1	45	324,753	2,763,065	32.75	139,462	8
S020	B101_F64_Y18_ST1_CC1	35	331,930	2,762,849	32.77	167,378	6
S014	B102_F65_Y18_ST1_CC1	59	270,825	2,755,643	32.74	76,415	10
S026	B103_F65_Y18_ST6161_CC1	81	174,519	2,757,366	32.79	80,021	13
S027	B104_F65_Y18_ST1_CC1	76	200,625	2,759,532	32.74	77,456	12
S028	B105_F65_Y18_ST1_CC1	69	315,387	2,761,380	32.77	90,823	9
S029	B106_F65_Y18_ST1_CC1	44	362,907	2,766,292	32.74	132,820	8
S029	B082_F63_Y18_ST1_CC1	38	279,437	2,760,108	32.76	136,081	8
S026	B107_F70_Y18_ST6140_CC1	58	231,434	2,762,419	32.71	108,323	9
S027	B108_F62_Y18_ST1_CC1	47	322,278	2,765,789	32.74	114,388	7
S028	B109_F62_Y18_ST1_CC1	38	298,698	2,752,537	32.77	125,959	8
ST116	B136_F80_Y02_ST1_CC1	87	201,638	2,754,487	32.79	60,860	14

Continued on next page

Unique Lab Identifier	Assigned names*	No of contigs	Largest contig	Total length	GC (%)	N50	L50
ST117	B132_F76_Y02_ST5_CC5	125	146,694	2,691,030	32.81	41,794	18
ST119	B137_F81_Y02_ST508_CC45	98	183,544	2,743,438	32.73	60,600	15
ST120	B133_F77_Y02_ST1_CC1	53	198,896	2,716,576	32.71	115,986	9
ST130	B134_F78_Y02_ST705_CC151	96	224,476	2,649,310	32.66	79,471	11
ST132	B135_F79_Y02_ST8_CC8	74	231,686	2,780,488	32.72	74,953	12
ST135	B148_F92_Y02_ST1_CC1	80	205,173	2,756,264	32.76	62,405	14
ST136	B131_F75_Y02_ST97_CC97	109	118,943	2,783,539	32.79	54,700	16
ST137	B138_F82_Y02_ST8_CC8	92	191,519	2,799,264	32.72	65,413	14
ST143	B139_F83_Y02_ST1_CC1	151	114,042	2,841,611	32.73	43,354	22
ST144	B158_F102_Y02_ST6162_CC97	58	249,864	2,810,296	32.74	116,050	9
ST147	B142_F86_Y02_ST1_CC1	151	158,203	2,752,868	32.84	39,371	20
ST150	B149_F93_Y02_ST6163_CC1	38	314,848	2,740,947	32.69	148,246	6
ST153	B147_F91_Y02_ST1_CC1	61	187,779	2,718,511	32.71	95,300	11
ST154	B159_F103_Y02_ST1_CC1	60	277,376	2,718,434	32.72	97,110	10
ST155	B160_F104_Y02_ST1_CC1	65	283,010	2,777,703	32.75	84,383	10
ST156	B161_F105_Y02_ST1_CC1	40	314,910	2,721,320	32.71	148,258	7
ST157	B143_F87_Y02_ST1_CC1	50	245,479	2,763,021	32.74	139,719	8
ST159	B162_F106_Y02_ST1_CC1	39	279,565	2,721,155	32.72	131,057	8
ST160	B163_F107_Y02_ST1_CC1	67	170,542	2,754,525	32.74	101,552	11

Continued on next page

Unique Lab Identifier	Assigned names*	No of contigs	Largest contig	Total length	GC (%)	N50	L50
ST163	B164_F108_Y02_ST71_CC97	190	112,844	2,775,673	32.83	30,859	28
ST164	B150_F94_Y02_ST1_CC1	55	208,964	2,719,369	32.71	99,972	10
ST166	B151_F95_Y02_ST151_CC151	50	307,116	2,642,588	32.62	117,464	7
ST167	B165_F109_Y02_ST1_CC1	65	188,242	2,758,496	32.73	90,531	11
ST168	B154_F98_Y02_ST1_CC1	47	314,977	2,744,307	32.73	134,654	8
ST169	B166_F111_Y02_ST1_CC1	16	581,245	2,721,452	32.72	317,524	3
ST173	B144_F88_Y02_ST1_CC1	129	115,393	2,718,287	32.79	44,592	19
ST174	B167_F112_Y02_ST1_CC1	55	234,371	2,720,283	32.7	128,508	9
ST176	B168_F113_Y02_ST1_CC1	53	314,476	2,758,864	32.73	124,133	8
ST177	B155_F99_Y02_ST1_CC1	38	316,685	2,753,491	32.7	150,683	7
ST180	B153_F97_Y02_ST1_CC1	42	263,302	2,719,028	32.7	139,795	8
ST181	B169_F114_Y02_ST1_CC1	26	564,954	2,723,444	32.72	169,178	5
ST182	B170_F115_Y02_ST1_CC1	103	188,935	2,765,092	32.77	51,972	18
ST185	B152_F96_Y02_ST1_CC1	67	261,486	2,789,264	32.72	68,904	11
ST187	B171_F116_Y02_ST1_CC1	65	169,207	2,775,572	32.71	97,961	11
ST188	B172_F117_Y02_ST1_CC1	41	451,679	2,779,447	32.73	143,223	5
ST189	B173_F118_Y02_ST1_CC1	48	325,720	2,761,815	32.74	99,068	9
ST196	B174_F119_Y02_ST1_CC1	53	261,691	2,717,207	32.72	86,016	11
ST198	B177_F122_Y02_ST1_CC1	92	182,434	2,760,903	32.78	63,661	14

Continued on next page

Unique Lab Identifier	Assigned names*	No of contigs	Largest contig	Total length	GC (%)	N50	L50
ST199	B141_F85_Y02_ST97_CC97	51	272,649	2,783,731	32.73	87,356	10
ST200	B178_F123_Y02_ST1_CC1	90	210,838	2,758,163	32.79	62,734	14
ST202	B145_F89_Y02_ST151_CC151	43	307,123	2,646,583	32.62	140,484	7
ST203	B179_F124_Y02_ST151_CC151	25	834,735	2,650,246	32.63	288,577	3
ST204	B180_F125_Y02_ST6164_CC97	41	252,565	2,848,322	32.71	180,198	7
ST205	B157_F101_Y02_ST1_CC1	34	322,971	2,791,474	32.72	145,520	7
ST206	B140_F84_Y02_ST1_CC1	36	431,240	2,743,233	32.69	169,147	5
ST208	B181_F126_Y02_ST1_CC1	25	459,212	2,738,342	32.7	212,459	5
ST209	B182_F127_Y02_ST1_CC1	54	286,862	2,719,130	32.74	111,694	8
ST211	B175_F120_Y02_ST1_CC1	33	319,229	2,778,199	32.74	148,039	6
ST212	B176_F121_Y02_ST6165_CC133	88	199,808	2,692,451	32.71	72,541	12
ST213	B156_F100_Y02_ST1_CC1	42	314,891	2,719,482	32.72	148,326	6
ST215	B183_F128_Y02_ST1_CC1	35	448,527	2,721,734	32.71	148,001	6
ST216	B184_F129_Y02_ST151_CC151	36	337,435	2,645,135	32.62	165,229	5
ST217	B185_F130_Y02_ST1_CC1	42	314,830	2,769,230	32.7	143,625	7
ST218	B186_F131_Y02_ST1_CC1	52	276,529	2,738,987	32.67	139,737	8
ST219	B187_F132_Y02_ST1_CC1	59	243,694	2,715,693	32.7	98,976	10
ST220	B146_F90_Y02_ST6140_CC1	44	325,313	2,767,688	32.7	145,663	8
ST222	B188_F133_Y02_ST1_CC1	48	253,131	2,718,927	32.7	111,629	9

Continued on next page

Unique Lab Identifier	**Assigned names***	**No of contigs**	**Largest contig**	**Total length**	**GC (%)**	**N50**	**L50**
ST99	B130_F74_Y02_ST1_CC1	159	101,688	2,759,650	32.85	37,613	27

Table A.2: QUAST metrics of the 279 *S. aureus* isolate assemblies with their assigned identifier, the number of contigs for each isolate, the total length of the assembly (in base pairs (bp)), percentage of guanine and cytosine (GC%), Length of contig that using longer or equal length contigs produces 50% of the total length (N50; in base pairs (bp)) and the smallest number of contigs whose length sum makes up half of the total length (L50). The assigned identifiers contain the host species identifier (B-Bovine, S-Small ruminant, H-Human, C-Canine, F-Feline), Farm identifier (F; varies for host species), Year of isolation (Y; where available), the niche of isolation (Co- commensal, Cl- Clinical infection), sequence type (ST), clonal complex (CC). The two isolates that were discarded did no have any assigned identifiers.

Assigned identifier	No of contigs	Total length (bp)	GC (%)	N50	L50
B001_F01_Y13_ST6143_CC8	57	2,774,290.00	32.6	141,787	8
B002_F01_Y13_ST6143_CC8	51	2,773,006.00	32.6	141,300	5
B003_F02_Y13_ST1_CC1	43	2,761,627.00	32.7	157,063	7
B004_F03_Y13_ST1_CC1	35	2,767,359.00	32.8	147,978	7
B005_F04_Y13_ST97_CC97	50	2,756,477.00	32.7	128,369	8
B006_F05_Y13_ST1_CC1	48	2,815,245.00	32.8	215,364	6
B007_F02_Y13_ST1_CC1	52	2,763,789.00	32.8	116,560	9
B008_F05_Y13_ST1_CC1	54	2,763,454.00	32.8	103,940	9
B009_F07_Y13_ST1_CC1	60	2,755,987.00	32.7	78,426	12
B010_F07_Y13_ST1_CC1	41	2,764,882.00	32.8	144,573	8
B011_F03_Y13_ST1_CC1	21	2,766,711.00	32.8	241,813	4
B012_F08_Y13_ST1_CC1	33	2,721,127.00	32.7	204,401	6
B013_F09_Y13_ST6143_CC8	61	2,774,006.00	32.6	109,220	8
B014_F10_Y13_ST1_CC1	29	2,767,294.00	32.8	245,960	4
B016_F12_Y13_ST1_CC1	61	2,717,288.00	32.7	107,283	10
B017_F134_Y13_ST1_CC1	28	2,765,499.00	32.8	240,249	4
B018_F13_Y13_ST1_CC1	43	2,759,815.00	32.7	139,360	9
B019_F14_Y13_ST97_CC97	56	2,764,446.00	32.7	128,691	8
B020_F15_Y13_ST1_CC1	25	2,762,747.00	32.8	240,960	4
B021_F16_Y13_ST1_CC1	22	2,749,994.00	32.8	220,105	5
B022_F17_Y13_ST1_CC1	26	2,824,123.00	32.7	167,631	6
B023_F18_Y13_ST199_CC15	48	2,792,015.00	32.8	232,228	5
B024_F19_Y13_ST1_CC1	42	2,844,613.00	32.7	241,134	4
B025_F13_Y13_ST97_CC97	27	2,791,526.00	32.7	197,756	5
B026_F20_Y13_ST97_CC97	28	2,759,741.00	32.7	201,557	5

Continued on next page

Assigned identifier	No of contigs	Total length (bp)	GC (%)	N50	L50
B027_F21_Y13_ST1247_CC133	22	2,689,222.00	32.8	197,991	5
B028_F22_Y13_ST1_CC1	20	2,732,589.00	32.7	241,184	4
B029_F23_Y13_ST97_CC97	31	2,742,328.00	32.7	151,658	4
B030_F24_Y13_ST1_CC1	16	2,776,753.00	32.8	527,369	2
B031_F25_Y13_ST8_CC8	21	2,809,598.00	32.6	489,475	2
B032_F26_Y13_ST1_CC1	74	2,761,959.00	32.8	79,035	12
B033_F08_Y13_ST1_CC1	22	2,842,477.00	32.8	314,798	3
B034_F19_Y13_ST97_CC97	44	2,786,290.00	32.8	122,359	9
B035_F27_Y13_ST1_CC1	37	2,763,976.00	32.8	139,757	6
B036_F27_Y13_ST1_CC1	79	2,718,308.00	32.7	95,993	10
B037_F28_Y13_ST1_CC1	33	2,858,814.00	32.7	241,078	5
B038_F29_Y13_ST1_CC1	52	2,761,487.00	32.8	113,171	7
B039_F29_Y13_ST1_CC1	40	2,761,077.00	32.8	146,005	7
B040_F30_Y13_ST4551_CC1	23	2,721,529.00	32.7	240,935	4
B041_F31_Y13_ST1_CC1	17	2,704,123.00	32.7	349,232	3
B042_F32_Y13_ST1_CC1	20	2,766,050.00	32.8	272,938	3
B043_F33_Y13_ST1_CC1	61	2,818,447.00	32.7	145,543	8
B044_F33_Y13_ST1_CC1	32	2,728,918.00	32.7	148,077	7
B045_F34_Y13_ST1_CC1	46	2,788,066.00	32.8	136,345	6
B046_F35_Y13_ST97_CC97	33	2,785,197.00	32.7	193,918	6
B047_F36_Y13_ST97_CC97	41	2,790,039.00	32.8	150,245	6
B048_F37_Y13_ST1_CC1	29	2,766,078.00	32.8	200,771	5
B049_F38_Y13_ST78_CC78	27	2,726,968.00	32.7	417,220	3
B050_F39_Y13_ST1_CC1	30	2,749,482.00	32.7	217,648	4
B051_F40_Y13_ST1_CC1	44	2,806,070.00	32.8	138,786	7
B052_F41_Y13_ST1_CC1	25	2,721,767.00	32.7	253,836	4
B053_F06_Y13_ST1_CC1	24	2,750,761.00	32.7	205,220	5
B054_F42_Y13_ST1_CC1	37	2,765,731.00	32.8	145,856	6
B055_F43_Y13_ST1_CC1	18	2,721,688.00	32.7	323,792	3
B056_F44_Y13_ST1_CC1	21	2,766,844.00	32.8	363,589	2
B057_F45_Y13_ST1_CC1	36	2,765,689.00	32.8	175,544	6
B058_F46_Y13_ST1_CC1	34	2,796,976.00	32.8	240,982	4

Continued on next page

Assigned identifier	No of contigs	Total length (bp)	GC (%)	N50	L50
B059_F47_Y13_ST1_CC1	26	2,766,595.00	32.8	196,362	5
B060_F48_Y13_ST1_CC1	26	2,763,202.00	32.8	287,602	4
B061_F48_Y13_ST1_CC1	25	2,763,079.00	32.8	207,181	4
B062_F49_Y13_ST97_CC97	41	2,776,363.00	32.7	154,963	6
B063_F50_Y13_ST1_CC1	64	2,761,161.00	32.8	88,799	11
B064_F51_Y13_ST1_CC1	62	2,754,728.00	32.8	97,610	9
B065_F52_Y13_ST1_CC1	78	2,790,988.00	32.8	88,684	14
B066_F54_Y18_ST1_CC1	59	2,700,337.00	32.7	93,364	11
B067_F55_Y18_ST97_CC97	58	2,740,234.00	32.6	87,061	11
B068_F56_Y18_ST151_CC151	44	2,731,253.00	32.7	135,272	7
B069_F55_Y18_ST97_CC97	75	2,741,029.00	32.6	67,643	14
B070_F54_Y18_ST1_CC1	58	2,702,217.00	32.7	103,024	10
B071_F57_Y18_ST151_CC151	56	2,730,158.00	32.7	106,992	8
B072_F58_Y18_ST5367_CC5367	99	2,707,597.00	32.7	60,054	16
B073_F55_Y18_ST97_CC97	85	2,735,676.00	32.6	66,876	14
B074_F55_Y18_ST97_CC97	100	2,735,115.00	32.7	49,842	18
B075_F58_Y18_ST5367_CC5367	73	2,710,410.00	32.7	73,786	10
B076_F59_Y18_ST1_CC1	65	2,720,330.00	32.7	108,461	9
B077_F60_Y18_ST97_CC97	100	2,732,514.00	32.7	58,929	17
B078_F61_Y18_ST1_CC1	45	2,761,465.00	32.7	139,833	7
B079_F62_Y18_ST6160_CC97	74	2,745,651.00	32.7	72,992	12
B080_F62_Y18_ST6160_CC97	63	2,744,632.00	32.7	85,674	11
B081_F62_Y18_ST6160_CC97	43	2,748,850.00	32.7	105,275	8
B082_F63_Y18_ST1_CC1	38	2,760,108.00	32.8	136,081	8
B083_F63_Y18_ST1_CC1	47	2,757,742.00	32.7	142,929	8
B084_F64_Y18_ST1_CC1	47	2,768,289.00	32.8	114,269	8
B085_F65_Y18_ST1_CC1	59	2,766,558.00	32.8	97,139	9
B086_F66_Y18_ST1_CC1	49	2,783,842.00	32.7	108,347	8
B087_F65_Y18_ST1_CC1	38	2,765,026.00	32.8	169,896	6
B088_F65_Y18_ST1_CC1	32	2,766,256.00	32.8	162,255	5
B089_F65_Y18_ST1_CC1	47	2,768,136.00	32.8	107,794	9
B090_F67_Y18_ST6141_CC1	66	2,758,140.00	32.8	94,806	11

Continued on next page

Assigned identifier	No of contigs	Total length (bp)	GC (%)	N50	L50
B091_F67_Y18_ST6141_CC1	52	2,767,382.00	32.8	107,759	9
B092_F65_Y18_ST1_CC1	123	2,760,072.00	32.8	44,467	21
B093_F65_Y18_ST1_CC1	72	2,761,913.00	32.8	83,817	11
B094_F65_Y18_ST1_CC1	62	2,789,136.00	32.8	145,685	7
B095_F68_Y18_ST1_CC1	41	2,761,447.00	32.8	148,638	6
B096_F68_Y18_ST1_CC1	43	2,768,703.00	32.8	108,363	9
B097_F63_Y18_ST1_CC1	36	2,764,311.00	32.8	235,814	5
B098_F69_Y18_ST1_CC1	43	2,751,920.00	32.8	135,841	8
B099_F69_Y18_ST1_CC1	41	2,758,399.00	32.8	189,069	6
B100_F64_Y18_ST1_CC1	45	2,763,065.00	32.8	139,462	8
B101_F64_Y18_ST1_CC1	35	2,762,849.00	32.8	167,378	6
B102_F65_Y18_ST1_CC1	59	2,755,643.00	32.7	76,415	10
B103_F65_Y18_ST6161_CC1	81	2,757,366.00	32.8	80,021	13
B104_F65_Y18_ST1_CC1	76	2,759,532.00	32.7	77,456	12
B105_F65_Y18_ST1_CC1	69	2,761,380.00	32.8	90,823	9
B106_F65_Y18_ST1_CC1	44	2,766,292.00	32.7	132,820	8
B107_F70_Y18_ST6140_CC1	58	2,762,419.00	32.7	108,323	9
B108_F62_Y18_ST1_CC1	47	2,765,789.00	32.7	114,388	7
B109_F62_Y18_ST1_CC1	38	2,752,537.00	32.8	125,959	8
B110_F62_Y18_ST6160_CC97	128	2,805,656.00	32.8	68,106	14
B111_F70_Y18_ST1_CC1	65	2,717,322.00	32.7	96,399	10
B112_F70_Y18_ST1_CC1	58	2,721,710.00	32.7	94,526	12
B113_F70_Y18_ST1_CC1	74	2,807,184.00	32.8	95,345	10
B114_F66_Y18_ST1_CC1	72	2,757,775.00	32.8	76,666	13
B115_F65_Y18_ST1_CC1	73	2,761,930.00	32.7	167,671	7
B116_F62_Y18_ST6160_CC97	133	2,792,096.00	32.8	46,770	20
B117_F62_Y18_ST6160_CC97	86	2,795,011.00	32.7	87,225	13
B118_F62_Y18_ST6160_CC97	63	2,795,287.00	32.7	136,696	7
B119_F62_Y18_ST1_CC1	59	2,787,392.00	32.7	110,020	9
B120_F62_Y18_ST6160_CC97	78	2,797,750.00	32.7	79,148	13
B121_F62_Y18_ST1_CC1	59	2,744,299.00	32.8	84,324	10
B122_F71_Y18_ST1_CC1	83	2,757,684.00	32.8	64,616	16

Continued on next page

Assigned identifier	No of contigs	Total length (bp)	GC (%)	N50	L50
B123_F69_Y18_ST1_CC1	40	2,754,454.00	32.8	140,615	7
B124_F69_Y18_ST1_CC1	95	2,752,942.00	32.8	57,754	17
B125_F72_Y18_ST1_CC1	70	2,756,720.00	32.8	75,397	11
B126_F73_Y18_ST1_CC1	80	2,754,069.00	32.8	69,190	14
B127_F63_Y18_ST1_CC1	55	2,710,217.00	32.7	99,037	10
B128_F63_Y18_ST1_CC1	51	2,712,604.00	32.7	107,693	10
B129_F63_Y18_ST1_CC1	34	2,714,208.00	32.7	171,092	6
B130_F74_Y02_ST1_CC1	159	2,759,650.00	32.9	37,613	27
B131_F75_Y02_ST97_CC97	109	2,783,539.00	32.8	54,700	16
B132_F76_Y02_ST5_CC5	125	2,691,030.00	32.8	41,794	18
B133_F77_Y02_ST1_CC1	53	2,716,576.00	32.7	115,986	9
B134_F78_Y02_ST705_CC151	96	2,649,310.00	32.7	79,471	11
B135_F79_Y02_ST8_CC8	74	2,780,488.00	32.7	74,953	12
B136_F80_Y02_ST1_CC1	87	2,754,487.00	32.8	60,860	14
B137_F81_Y02_ST508_CC45	98	2,743,438.00	32.7	60,600	15
B138_F82_Y02_ST8_CC8	92	2,799,264.00	32.7	65,413	14
B139_F83_Y02_ST1_CC1	151	2,841,611.00	32.7	43,354	22
B140_F84_Y02_ST1_CC1	36	2,743,233.00	32.7	169,147	5
B141_F85_Y02_ST97_CC97	51	2,783,731.00	32.7	87,356	10
B142_F86_Y02_ST1_CC1	151	2,752,868.00	32.8	39,371	20
B143_F87_Y02_ST1_CC1	50	2,763,021.00	32.7	139,719	8
B144_F88_Y02_ST1_CC1	129	2,718,287.00	32.8	44,592	19
B145_F89_Y02_ST151_CC151	43	2,646,583.00	32.6	140,484	7
B146_F90_Y02_ST6140_CC1	44	2,767,688.00	32.7	145,663	8
B147_F91_Y02_ST1_CC1	61	2,718,511.00	32.7	95,300	11
B148_F92_Y02_ST1_CC1	80	2,756,264.00	32.8	62,405	14
B149_F93_Y02_ST6163_CC1	38	2,740,947.00	32.7	148,246	6
B150_F94_Y02_ST1_CC1	55	2,719,369.00	32.7	99,972	10
B151_F95_Y02_ST151_CC151	50	2,642,588.00	32.6	117,464	7
B152_F96_Y02_ST1_CC1	67	2,789,264.00	32.7	68,904	11
B153_F97_Y02_ST1_CC1	42	2,719,028.00	32.7	139,795	8
B154_F98_Y02_ST1_CC1	47	2,744,307.00	32.7	134,654	8

Continued on next page

Assigned identifier	No of contigs	Total length (bp)	GC (%)	N50	L50
B155_F99_Y02_ST1_CC1	38	2,753,491.00	32.7	150,683	7
B156_F100_Y02_ST1_CC1	42	2,719,482.00	32.7	148,326	6
B157_F101_Y02_ST1_CC1	34	2,791,474.00	32.7	145,520	7
B158_F102_Y02_ST6162_CC97	58	2,810,296.00	32.7	116,050	9
B159_F103_Y02_ST1_CC1	60	2,718,434.00	32.7	97,110	10
B160_F104_Y02_ST1_CC1	65	2,777,703.00	32.8	84,383	10
B161_F105_Y02_ST1_CC1	40	2,721,320.00	32.7	148,258	7
B162_F106_Y02_ST1_CC1	39	2,721,155.00	32.7	131,057	8
B163_F107_Y02_ST1_CC1	67	2,754,525.00	32.7	101,552	11
B164_F108_Y02_ST71_CC97	190	2,775,673.00	32.8	30,859	28
B165_F109_Y02_ST1_CC1	65	2,758,496.00	32.7	90,531	11
B166_F111_Y02_ST1_CC1	16	2,721,452.00	32.7	317,524	3
B167_F112_Y02_ST1_CC1	55	2,720,283.00	32.7	128,508	9
B168_F113_Y02_ST1_CC1	53	2,758,864.00	32.7	124,133	8
B169_F114_Y02_ST1_CC1	26	2,723,444.00	32.7	169,178	5
B170_F115_Y02_ST1_CC1	103	2,765,092.00	32.8	51,972	18
B171_F116_Y02_ST1_CC1	65	2,775,572.00	32.7	97,961	11
B172_F117_Y02_ST1_CC1	41	2,779,447.00	32.7	143,223	5
B173_F118_Y02_ST1_CC1	48	2,761,815.00	32.7	99,068	9
B174_F119_Y02_ST1_CC1	53	2,717,207.00	32.7	86,016	11
B175_F120_Y02_ST1_CC1	33	2,778,199.00	32.7	148,039	6
B176_F121_Y02_ST6165_CC133	88	2,692,451.00	32.7	72,541	12
B177_F122_Y02_ST1_CC1	92	2,760,903.00	32.8	63,661	14
B178_F123_Y02_ST1_CC1	90	2,758,163.00	32.8	62,734	14
B179_F124_Y02_ST151_CC151	25	2,650,246.00	32.6	288,577	3
B180_F125_Y02_ST6164_CC97	41	2,848,322.00	32.7	180,198	7
B181_F126_Y02_ST1_CC1	25	2,738,342.00	32.7	212,459	5
B182_F127_Y02_ST1_CC1	54	2,719,130.00	32.7	111,694	8
B183_F128_Y02_ST1_CC1	35	2,721,734.00	32.7	148,001	6
B184_F129_Y02_ST151_CC151	36	2,645,135.00	32.6	165,229	5
B185_F130_Y02_ST1_CC1	42	2,769,230.00	32.7	143,625	7
B186_F131_Y02_ST1_CC1	52	2,738,987.00	32.7	139,737	8

Continued on next page

Assigned identifier	No of contigs	Total length (bp)	GC (%)	N50	L50
B187_F132_Y02_ST1_CC1	59	2,715,693.00	32.7	98,976	10
B188_F133_Y02_ST1_CC1	48	2,718,927.00	32.7	111,629	9
B189_F37_Y13_ST1_CC1	17	2,723,352.00	32.7	283,546	3
C01_Cl_ST1_CC1	254	2,647,707.00	32.9	20,699	38
C02_Cl_ST5_CC5	155	2,689,721.00	32.8	41,849	22
C03_Cl_ST1_CC1	62	2,713,376.00	32.7	80,995	12
C04_Cl_ST582_CC15	68	2,705,708.00	32.7	91,664	9
C05_Cl_ST188_CC188	57	2,787,842.00	32.7	113,863	8
C06_Cl_ST188v_CC188	107	2,743,716.00	32.7	56,202	16
C07_Cl_ST30_CC30	211	2,721,135.00	32.9	26,451	33
C08_Cl_ST39_CC30	283	2,670,493.00	32.9	16,844	45
C09_Cl_ST5_CC5	36	2,723,148.00	32.7	136,816	6
C10_Co_ST1_CC1	90	2,835,585.00	32.7	53,096	17
C11_Co_ST1_CC1	153	2,812,768.00	32.8	43,345	22
C12_Co_ST1_CC1	114	2,762,225.00	32.8	48,382	18
C13_Co_ST5_CC5	83	2,708,890.00	32.8	74,004	11
C14_Co_ST5_CC5	56	2,721,153.00	32.8	88,107	9
C15_Co_ST5_CC5	46	2,751,786.00	32.7	115,622	8
C16_Co_ST188_CC188	111	2,756,229.00	32.8	47,375	17
C17_Co_ST30_CC30	77	2,787,647.00	32.7	99,421	9
C18_Co_ST34_CC30	86	2,833,664.00	32.8	81,298	11
C19_Co_ST1259_CC5	72	2,778,667.00	32.8	68,637	12
F01_Cl_ST1_CC1	91	2,663,680.00	32.8	62,559	16
F02_Cl_ST188_CC188	80	2,781,861.00	32.7	86,557	10
F03_Cl_ST188_CC188	349	2,656,755.00	33.0	12,831	57
F04_Cl_ST188_CC188	99	2,705,008.00	32.8	64,072	16
F05_Co_ST1_CC1	67	2,834,960.00	32.7	88,720	12
F06_Co_ST188v_CC188	164	2,733,644.00	32.8	36,172	29
H01_Cl_ST5_CC5	197	2,632,627.00	32.9	28,488	30
H02_Cl_ST835_CC5	76	2,825,927.00	32.7	90,051	11
H03_Cl_ST188_CC188	101	2,744,821.00	32.7	82,496	12
H04_Cl_ST188_CC188	118	2,739,204.00	32.8	46,427	20

Continued on next page

Assigned identifier	No of contigs	Total length (bp)	GC (%)	N50	L50
H05_Cl_ST188_CC188	32	2,829,562.00	32.7	165,426	6
H06_Cl_ST188_CC188	72	2,729,454.00	32.7	87,935	11
H07_Cl_ST1_CC1	111	2,720,779.00	32.8	51,270	17
H08_Cl_ST1_CC1	62	2,833,524.00	32.7	87,894	12
H09_Cl_ST1_CC1	31	2,782,692.00	32.7	177,309	5
H10_Cl_ST1_CC1	43	2,724,115.00	32.7	108,518	7
H11_Cl_ST1_CC1	34	2,836,853.00	32.7	134,271	6
H12_Cl_ST1_CC1	66	2,709,242.00	32.7	76,270	11
H13_Cl_ST15_CC15	64	2,755,001.00	32.8	84,062	11
H14_Cl_ST30_CC30	54	2,771,952.00	32.8	167,367	5
H15_Co_ST5_CC5	89	2,706,234.00	32.8	52,060	17
H16_Co_ST1_CC1	62	2,786,851.00	32.7	94,272	9
H17_Co_ST1_CC1	56	2,731,782.00	32.7	92,511	10
H18_Co_ST5_CC5	77	2,753,537.00	32.8	68,973	13
H19_Co_ST5_CC5	166	2,683,330.00	32.8	34,364	25
H20_Co_ST5_CC5	30	2,784,835.00	32.7	230,765	4
H21_Co_ST188_CC188	21	2,747,563.00	32.7	250,644	4
H22_Co_ST188_CC188	44	2,785,549.00	32.7	162,694	6
H23_Co_ST188_CC188	54	2,774,368.00	32.7	93,516	8
H24_Co_ST188_CC188	116	2,732,872.00	32.7	65,673	14
H25_Co_ST188_CC188	45	2,756,065.00	32.7	167,404	6
H26_Co_ST188_CC188	102	2,721,518.00	32.7	51,298	15
H27_Co_ST1_CC1	60	2,712,217.00	32.7	108,296	8
H28_Co_ST5_CC5	86	2,763,247.00	32.8	55,272	13
H29_Co_ST5_CC5	56	2,740,501.00	32.8	108,251	8
H30_Co_ST5_CC5	46	2,711,197.00	32.7	124,291	9
H31_Co_ST15_CC15	36	2,712,455.00	32.7	177,233	6
H32_Co_ST30_CC30	141	2,777,519.00	32.8	39,089	25
H33_Co_ST30_CC30	56	2,763,167.00	32.7	133,987	8
H34_Co_ST30v_CC30	96	2,754,099.00	32.7	67,237	12
Isolate discarded (small-ruminant)	33	2,401,438.00	34.0	168,713	4
Isolate discarded (small-ruminant)	342	3,338,706.00	34.3	161,727	5

Continued on next page

Assigned identifier	No of contigs	Total length (bp)	GC (%)	N50	L50
S001_F11_Y13_ST3615_CC1	26	2,721,880.00	32.7	210,757	5
S002_F53_Y18_ST6157_CC133	46	2,813,813.00	32.8	233,386	5
S01_F01_ST6137_CC133	31	2,721,885.00	32.7	227,007	3
S02_F02_ST133_CC133	61	2,829,571.00	32.8	190,359	5
S03_F02_ST5367_CC5367	73	2,740,490.00	32.7	94,046	8
S04_F02_ST5367_CC5367	63	2,740,468.00	32.7	101,317	8
S05_F03_ST6138_CC133	42	2,815,574.00	32.8	274,392	4
S06_F03_ST133_CC133	36	2,760,156.00	32.8	266,602	4
S08_F03_ST1640_CC1640	42	2,763,837.00	32.8	128,688	8
S09_F03_ST1640_CC1640	45	2,844,344.00	32.8	137,003	7
S10_F04_ST133_CC133	38	2,808,419.00	32.8	320,901	3
S11_F05_ST701_CC133	81	2,876,524.00	32.8	174,627	6
S12_F04_ST6139_CC133	45	2,821,024.00	32.8	266,909	4
S13_F06_ST8_CC8	44	2,791,226.00	32.8	305,097	4
S14_F02_ST692_CC692	24	2,719,443.00	32.7	345,482	4
S15_F07_ST1640_CC1640	33	2,778,616.00	32.8	175,458	5
S16_F08_ST5367_CC5367	62	2,734,260.00	32.7	93,543	8
S17_F08_ST8_CC8	26	2,768,411.00	32.7	275,769	3
S18_F08_ST8_CC8	27	2,768,248.00	32.7	507,516	2
S19_F09_ST133_CC133	47	2,778,526.00	32.8	189,986	6
S20_F09_ST8_CC8	34	2,758,892.00	32.7	275,863	2
S22_F08_ST8_CC8	25	2,767,953.00	32.7	363,570	3
S23_F09_ST8_CC8	31	2,788,018.00	32.7	304,915	3
S24_F08_ST6166_CC133	63	2,825,190.00	32.8	280,021	4
S25_F03_ST133_CC133	99	2,835,472.00	32.8	268,155	3
S26_F03_ST133_CC133	34	2,818,357.00	32.8	271,835	4
S27_F03_ST8_CC8	26	2,732,332.00	32.7	364,695	3
S28_F09_ST8_CC8	24	2,746,581.00	32.7	346,678	2
S29_F10_ST133_CC133	38	2,817,833.00	32.8	320,654	3
S30_F10_ST133_CC133	39	2,801,944.00	32.8	266,523	5

A.1 Aminoacid sequences (in fasta format) of the 18 potential vaccine candidate proteins shortlisted through the reverse vaccinology pipeline in Chapter 6.

>Staphylococcal secretory antigen (ssaA2)
MKKIATATIATAGFATIAIASGNQAHASEQDNYGYNPNDPTSYSYTYTIDAQGNYHYTWK
GNWHPSQLNQDNGYYSYYYYNGYNNYNYNNYNNGYSYNNYSRYNNYSNNNQSYNYNNYNS
YNTNSYRTGGLGASYSTSSNNVQVTTTMAPSSNGRSISSGYTSGRNLYTSGQCTYYVFDR
VGGKIGSTWGNASNWANAAARAGYTVNNTPKAGAIMQTTQGAYGHVAYVESVNSNGSVRV
SEMNYGYGPGVVTSRTISASQAAGYNFIH

>Hypothetical protein C
MKKLLITIVIIVLIGALGFAIYAAINHSTSSSNNQEQENKATHKKTETEENDQQDDSADK
ADKNQNNGNNVTNDNQPSSQSNVPNNHSTAPSNTQPAQPKDSNSNKNNTHDGGSQSSTNN
QNNQQQNTPPANKSNNTNTSSKQTPQAQSPKN

>Putative transglycosylase (SceD)
MKKTLLASSLAVGLGIVAGNAGHEAHASEADLNKASLAQMAQSNDQTLNQKPIEAGAYNY
TFDYEGFTYHFESDGTHFAWNYHATGTNGADMSAQAPATNNVAPSAVQANQVQSQEVEAP
QNAQTQQPQASTSNNSQVTATPTESKSSEGSSVNVNAHLKQIAQRESGGNIHAVNPTSGA
AGKYQFLQSTWDSVAPAKYKGVSPANAPESVQDAAAVKLYNTGGAGHWVTA

>Putative transglycosylase (IsaA)

MKKTIMASSLAVALGVTGYAAGTGHQAHAAEVNVDQAHLVDLAHNHQDQLNAAPIKDGAY
DIHFVKDGFQYNFTSNGTTWSWSYEAANGQTAGFSNVAGADYTTSYNQGSNVQSVSYNAQ
SSNSNVEAVSAPTYHNYSTSTTSSSVRLSNGNTAGATGSSAAQIMAQRTGVSASTWAAII
ARESNGQVNAYNPSGASGLFQTMPGWGPTNTVDQQINAAVKAYKAQGLGAWGF

>N-acetylmuramoyl-L-alanine amidase (sle1)
MKKLAFAITATSGAAAFLTHHDAQASTQHTVQSGESLWSIAQKYNTSVESIKQNNQLDNN
LVFPGQVISVGGSDAQNSSNTSPQAGSASSHTVQAGESLNIIASRYGVSVDQLMAANNLR
GYLIMPNQTLQIPNGGSGGTTPTATTGSNGNASSFNHQNLYTAGQCTWYVFDRRAQAGSP
ISTYWSDAKYWAGNAANDGYQVNNTPSVGSIMQSTPGPYGHVAYVERVNGDGSILISEMN
YTYGPYNMNYRTIPASEVSSYAFIH

>Bifunctional autolysin (Atl)
MAKKFNYKLPSMVALTLVGSAVTAHQVQAAETTQDQTTNKNVLDSNKVKATTEQAKAEVK
NPTQNISGTQVYQDPAIVQPKTANNKTGNAQVSQKVDTAQVNGDTRANQSATTNNTQPVA
KSTSTTAPKTNTNVTNAGYSLVDDEDDNSENQINPELIKSAAKPAALETQYKAAAPKAAT
TSAPKAKTEATPKVTTFSASAQPRSVAATPKTSLPKYKPQVNSSINDYIRKNNLKAPKIE
EDYTSYFPKYAYRNGVGRPEGIVVHDTANDRSTINGEISYMKNNYQNAFVHAFVDGDRII
ETAPTDYLSWGVGAVGNPRFINVEIVHTHDYASFARSMNNYADYAATQLQYYGLKPDSAE
YDGNGTVWTHYAVSKYLGGTDHADPHGYLRSHNYSYDQLYDLINEKYLIKMGKVAPWGTQ
STTTPTTPSKPTTPSKPSTGKLTVAANNGVAQIKPTNSGLYTTVYDKTGKATNEVQKTFA
VSKTATLGNQKFYLVQDYNSGNKFGWVKEGDVVYNTAKSPVNVNQSYSIKPGTKLYTVPW
GTSKQVAGSVSGSGNQTFKASKQQQIDKSIYLYGSVNGKSGWVSKAYLVDTAKPTPTPTP

KPSTPTTNNKLTVSSLNGVAQINAKNNGLFTTVYDKTGKPTKEVQKTFAVTKEASLGGNK
FYLVKDYNSPTLIGWVKQGDVIYNNAKSPVNVMQTYTVKPGTKLYSVPWGTYKQEAGAVS
GTGNQTFKATKQQQIDKSIYLFGTVNGKSGWVSKAYLAVPAAPKKAVAQPKTAVKAYTVT
KPQTTQTVSKIAQVKPNNTGIRASVYEKTAKNGAKYADRTFYVTKERAHGNETYVLLNNT
SHNIPLGWFNVKDLNVQNLGKEVKTTQKYTVNKSNNGLSMVPWGTKNQVILTGNNIAQGT
FNATKQVSVGKDVYLYGTINNRTGWVNAKDLTAPTAVKPTTSAAKDYNYTYVIKNGNGYY
YVTPNSDTAKYSLKAFNEQPFAVVKEQVINGQTWYYGKLSNGKLAWIKSTDLAKELIKYN
QTGMTLNQVAQIQAGLQYKPQVQRVPGKWTDANFNDVKHAMDTKRLAQDPALKYQFLRLD
QPQNISIDKINQFLKGKGVLENQGAAFNKAAQMYGINEVYLISHALLETGNGTSQLAKGA
DVVNNKVVTNSNTKYHNVFGIAAYDNDPLREGIKYAKQAGWDTVSKAIVGGAKFIGNSYV
KAGQNTLYKMRWNPAHPGTHQYATDIDWANINAKIIKGYYDKIGEVGKYFDIPQYK

>Iron-regulated surface determinant protein (IsdA)
MTKHYLNSKYQSEQRSSAMKKITMGTASIILGSLVYIGADSQQVNAATEATNATNNQSTQ
VSQATSQPINFQVQKDGSSEKSHMDDYMQHPGKVIKQNNKYYFQTVLNNPSFWKEYKFYN
ANNQELATTVVNDNKKADTRTINVAVEPGYKSLTTKVHIVVPQINYNHRYTTHLEFEKAI
PTLADAAKPNNVKPVQPKPAQPKTPTEQTKPVQPKVEKVKPTVTTTSKVEDNHSTKVVST
DTTKDQTKTQTAHTVKTAQTAQEQNKVQTPVKDVATAKSESNNQAVSDNKSQQTNKVTKH
NETPKQASKAKELPKTGLTSVDNFISTVAFATLALLGSLSLLLFKRKESK

>Putative leukocidin-like protein 1
MIKQLYKNITICTLALSTTFTVLPATSYAKINSEIKAVSEKNLDGDTKMYTRTATTSDSQ
KNITQSLQFNFLTEPNYDKETVFIKAKGTIGSGLRILDPNGYWNSTLRWPGSYSVSIQNV

DDNNNTNVTDFAPKNQDESREVKYTYGYKTGGDFSINRGGLTGNITKESNYSETISYQQP
SYRTLLDQSTSHKGVGWKVEAHLINNMGHDHTRQLTNDSDNRTKSEIFSLTRNGNLWAKD
NFTPKNKMPVTVSEGFNPEFLAVMSHDKKDEGKSKFVVHYKRSMDEFKIDWNRHGFWGYW
SGENHVDKKEEKLSALYEVDWKTHNVKFVKVLNDNEKK

>Iron-regulated surface determinant protein B (IsdB)
MNKQQKEFKSFYSIRKSSLGVASVAISTLLLLMSNGEAQAAAEETGGTNTEAQPKTEAVA
SPTTTSEKAPETKPVANAVSVSNKEVEAPTSETKEAKEVKEVKAPKETKEVKPAAKATNN
TYPILNQELREAIKNPAIKDKDHSAPNSRPIDFEMKKKDGTQQFYHYASSVKPARVIFTD
SKPEIELGLQSGQFWRKFEVYEGDKKLPIKLVSYDTVKDYAYIRFSVSNGTKAVKIVSST
HFNNKEEKYDYTLMEFAQPIYNSADKFKTEEDYKAEKLLAPYKKAKTLERQVYELNKIQD
KLPEKLKAEYKKKLEDTKKALDEQVKSAITEFQNVQPTNEKMTDLQDTKYVVYESVENNE
SMMDTFVKHPIKTGMLNGKKYMVMETTNDDYWKDFMVEGQRVRTISKDAKNNTRTIIFPY
VEGKTLYDAIVKVHVKTIDYDGQYHVRIVDKEAFTKANTDKSNKKEQQDNSAKKEATPAT
PSKPTPSPVEKESQKQDSQKDDNKQLPSVEKENDASSESGKDKTPATKPTKGEVESSSTT
PTKVVSTTQNVAKPTTASSKTTKDVVQTSAGSSEAKDSAPLQKANIKNTNDGHTQSQNNK
NTQENKAKSLPQTGEESNKDMTLPLMALLALSSIVAFVLPRKRKN

>Cell division protein (DivIB)
MDDKTKNDQQESNEDKDELELFTRNTSKKRRQRKRSKATHFSNQNKDDTSQQADFDEEIY
LINKDFKKEQSNDENNDSASSRANNNNIDDSTDSNIENEDYRYNQEIDDQNESNGIAFDN
EQPQSAPKEQNGDSNDEETVTKKERKSKVTQLKPLTLEEKRKLRRKRQKRIQYSVITILV
LLIAVILIYMFSPLSKIAHVNINGNNHVSTSKINKVLGVKNDSRMYTFSKKNAINDLEEN

PLIKSVEIHKQLPNTLNVDITENEIIALVKYKGKYLPLLENGKLLKGSNDVKINDAPVMD
GFKGTKEDDMIKALSEMTPEVRRYIAEVTYAPSKNKQSRIELFTTDGLQVIGDISTISKK
MKYYPQMSQSLSRDSSGKLKTRGYIDLSVGASFIPYRGNTSSQSESDKNVTKSSQEENQA
KEELQSVLNKINKQSSKNN

>Staphopain B (SspB)
MNSSYKSRVFNIISIIMVSMLILSLGAFANNNKAKADSHSKQLEINVKSDKVPQKVKDLA
QQQFAGYAKALDKQSNAKTGKYELGEAFKIYKFNGEEDNSYYYPVIKDGKIVYTLTLSPK
NKDDLNKSKEDMNYSVKISNFIAKDLDQIKDKNSNITVLTDEKGFYFEEDGKVRLVKATP
LPGNVKEKESAKTVSSKLKQELKNTVTPTKVEENEAIQEDQVQYENTLKNFKIREQQFDN
SWCAGFSMAALLNATKNTDTYNAHDIMRTLYPEVSEQDLPNCSTFPNQMIEYGKSQGRDI
HYQEGVPSYEQVDQLTKDNVGIMILAQSVSQNPNDPHLGHALAVVGNAKINDQEKLIYWN
PWDTELSIQDADSSLLHLSFNRDYNWYGSMIGY

>Hypothetical protein A
MKKLVTATTLTAGIGTALVGQAHHADAAENYTNYNNYNYNTTQTTTTTTTTTTSSISHS
GNLYTAGQCTWYVYDKVGGEIGSTWGNANNWAAAAQGAGFTVNHTPSKGAILQSSEGPFG
HVAYVESVNSDGSVTISEMNYSGGPFSVSSRTISASEAGNYNYIHI

>Fibrinogen-binding protein (Efb)
MKNKLIAKSLLTIAAIGITTTTIASTADASEGYGPREKKPVSINHNIVEYNDGTFKYQSR
PKFNSTPKYIKFKHDYNILEFNDGTFEYGARPQFNKPAAKTDATIKKEQKLIQAQNLVRE
FEKTHTVSAHRKAQKAVNLVSFEYKVKKMVLQERIDNVLKQGLVK

>ESAT-6 secretion system extracellular protein A (EsxA)
MAMIKMSPEEIRAKSQSYGQGSDQIRQILSDLTRAQGEIAANWEGQAFSRFEEQFQQLSP
KVEKFAQLLEEIKQQLNSTADAVQEQDQQLSNNFGLQ

>Glutathione-binding protein (GsiB)
MKFKRLATIFSAVLVLSGCGSMHSSGKDLNISLPLKTKSIAPYETDVPVKIGAAESLFKT
NDQGKIEKALVKSYHQPNDTTLDIELKDNIKFQNGQKLTAEKVKSSLENSMKKSDLVKYS
LPISSITAKGQKLTIKTNSAYPELVSELANPFMAIYDTDAKSDVNQTPVGTGPYQIKDY
QSRKISLSNFKDYWQGKPKLDHITVTYQEDGNNRVRNLESQKDDLITDVPVNKVQDIENN
QNLKVSKESGFRTSLLMYNHTNKKMTKSVREALDHIIDRQGIADHIYQGYAKPATSPFND
KIPYIKEPKLTKQNIEQAKTLLAKDGYTKEHPLKIKLITYDGRPELSKIAQVLQSDAKKA
NIEIDIKSVDDIEGYLKDRSAWDATMYSFGTIPRGDTGYFFNQAYKKDGAINKGDYNNSN
VDDLINQLNHTVDVKERHNISNDIIKLSSRDVPNSYIAYNDQIVAANAKVKNYKVTPEGI
YLIDYRTTIER

>Nickel-binding periplasmic protein (NikA)
MRKLTKMSAMLLASGLILTGCGGNKGLEEKKENKQLTYTTVKDIGDMNPHVYGGSMSAES
MIYEPLVRNTKDGIKPLLAKKWDVSEDGKTYTFHLRDDVKFHDGTPFDADAVKKNIDAVQ
QNKKLHSWLKISTLIDNVKVKDKYTVELNLKEAYQPALAELAMPRPYVFVSPKDFKNGTT
KDGVKKFDGTGPFKLGEHKKDESADFNKNDQYWGEKSKLNKVQAKVMPAGETAFLSMKKG
ETNFAFTDDRGTDSLDKDSLKQLKDTGDYQVKRSQPMNTKMLVVNSGKKDNAVSDKTVRQ
AIGHMVNRDKIAKEILDGQEKPATQLFAKNVTDINFDMPTRKYDLKKAESLLDEAGWKKG
KDSDVRQKDGKNLEMAMYYDKGSSSQKEQAEYLQAEFKKMGIKLNINGETSDKIAERRTS

GDYDLMFNQTWGLLYDPQSTIAAFKAKNGYESATSGIENKDKIYNSIDDAFKIQNGKERS
GAYKNILKQIDDEGIFIPISHGSMTVVAPKDLEKVSFTQSQYELPFNEMQYK

>Hypothetical protein D
MKKNFIGKSILSIAAISLTVSTFAGESHAQTKVEKYNEYQTNFKKQVNKKVVDAQKAVNL
FKRTRTVATHRKAQRAVNLIHFQHSYEKKKLQRQIDLVLKYNTLK

>Hypothetical protein B
MTNADVAYLLHDEKNEIREIEPVISRKASAREQLTTLFNDKKQAIEANIQATVEERNSIL
AQLQNIYDTAIGQIDQDRSNAQVDKTASLNLQTIHDLDVHPIKKPDAEKTINDDLARVTA
LVQNYRKVSDRNKADALKAITALKLQMDEELKTARTNADVDAVLKRFNVALSDIEAVITE
KENSLLRIDNIAQQTYAKFKAIATPEQLAKVKVLIDQYVADGNRMIDEDATLNDIKQHTQ
FIVDEILAIKLPAEAMKVSPKVIQPAPKVCTPIKKEETHESRKVEKELPNTGSEEMDLPL
KEFALITGAALLARRRTKNEKES

Table A.3: List of epitopes along with their position in the protein sequence and epitope score of the shortlisted potential vaccine candidates identified using BCPreds in Chapter 6.

Glutathione-binding protein (GsiB)		
Position	Epitope	Score
390	GTIPRGDTGYFFNQAYKKDG	0.996
291	AKPATSPFNDKIPYIKEPKL	0.988
224	DLITDVPVNKVQDIENNQNL	0.984
161	KSDVNQTPVGTGPYQIKDYK	0.975
194	WQGKPKLDHITVTYQEDGNN	0.96
465	AANAKVKNYKVTPEGIYLID	0.953
34	PLKTKSIAPYETDVPVKIGA	0.901
70	LVKSYHQPNDTTLDIELKDN	0.827
326	GYTKEHPLKIKLITYDGRPE	0.823
368	SVDDIEGYLKDRSAWDATMY	0.784

Bifunctional autolysin		
Position	Epitope	Score
591	TAKPTPTPTPKPSTPTTNNK	1
421	STTTPTTPSKPTTPSKPSTG	1
173	AAAPKAATTSAPKAKTEATP	1
118	PVAKSTSTTAPKTNTNVTNA	1
932	TAPTAVKPTTSAAKDYNYTY	0.999
1142	VVNNKVVTNSNTKYHNVFGI	0.998
75	PAIVQPKTANNKTGNAQVSQ	0.994
796	PNNTGIRASVYEKTAKNGAK	0.993
758	AVPAAPKKAVAQPKTAVKAY	0.992
506	WVKEGDVVYNTAKSPVNVNQ	0.986
458	SGLYTTVYDKTGKATNEVQK	0.985
953	IKNGNGYYYVTPNSDTAKYS	0.983
244	TSYFPKYAYRNGVGRPEGIV	0.982
300	IETAPTDYLSWGVGAVGNPR	0.981
630	FTTVYDKTGKPTKEVQKTFA	0.979
683	YNNAKSPVNVMQTYTVKPGT	0.978
706	SVPWGTYKQEAGAVSGTGNQ	0.978
1227	DWANINAKIIKGYYDKIGEV	0.977
1038	YKPQVQRVPGKWTDANFNDV	0.975

202	QPRSVAATPKTSLPKYKPQV	0.967
551	SGSGNQTFKASKQQQIDKSI	0.94
350	QYYGLKPDSAEYDGNGTVWT	0.938
223	SSINDYIRKNNLKAPKIEED	0.937
1205	NTLYKMRWNPAHPGTHQYAT	0.925
527	YSIKPGTKLYTVPWGTSKQV	0.912
869	YTVNKSNNGLSMVPWGTKNQ	0.871
910	GKDVYLYGTINNRTGWVNAK	0.867
265	HDTANDRSTINGEISYMKNN	0.867
661	FYLVKDYNSPTLIGWVKQGD	0.86
54	QAKAEVKNPTQNISGTQVYQ	0.851
1014	KELIKYNQTGMTLNQVAQIQ	0.841
145	EDDNSENQINPELIKSAAKP	0.834
97	DTAQVNGDTRANQSATTNNT	0.805
731	KQQQIDKSIYLFGTVNGKSG	0.8
989	INGQTWYYGKLSNGKLAWIK	0.785
1079	LDQPQNISIDKINQFLKGKG	0.771
484	TATLGNQKFYLVQDYNSGNK	0.76

Cell division protein (DivIB)		
Position	Epitope	Score
406	SDKNVTKSSQEENQAKEELQ	1
23	TRNTSKKRRQRKRSKATHFS	0.999
68	KEQSNDENNDSASSRANNNN	0.999
158	EEKRKLRRKRQKRIQYSVIT	0.996
2	DDKTKNDQQESNEDKDELEL	0.995
120	NEQPQSAPKEQNGDSNDEET	0.987
89	DDSTDSNIENEDYRYNQEID	0.97
385	IDLSVGASFIPYRGNTSSQS	0.951
196	KIAHVNINGNNHVSTSKINK	0.949
293	INDAPVMDGFKGTKEDDMIK	0.911
352	GDISTISKKMKYYPQMSQSL	0.875

Staphopain B		
Position	Epitope	Score
200	QELKNTVTPTKVEENEAIQE	0.997
174	RLVKATPLPGNVKEKESAKT	0.985

97	EDNSYYYPVIKDGKIVYTLT	0.984
290	IEYGKSQGRDIHYQEGVPSY	0.976
68	AKALDKQSNAKTGKYELGEA	0.919
26	GAFANNNKAKADSHSKQLEI	0.884
350	INDQEKLIYWNPWDTELSIQ	0.872
229	KNFKIREQQFDNSWCAGFSM	0.763
145	DLDQIKDKNSNITVLTDEKG	0.751
256	KNTDTYNAHDIMRTLYPEVS	0.744

Iron-regulated surface determinant protein A (IsdA)		
Position	Epitope	Score
187	AKPNNVKPVQPKPAQPKTPT	1
44	VNAATEATNATNNQSTQVSQ	0.998
209	TKPVQPKVEKVKPTVTTTSK	0.997
137	ADTRTINVAVEPGYKSLTTK	0.99
276	TAKSESNNQAVSDNKSQQTN	0.987
231	DNHSTKVVSTDTTKDQTKTQ	0.972
298	TKHNETPKQASKAKELPKTG	0.916
158	HIVVPQINYNHRYTTHLEFE	0.865
108	NNPSFWKEYKFYNANNQELA	0.852
80	EKSHMDDYMQHPGKVIKQNN	0.774

Putative leukocidin-like protein 1		
Position	Epitope	Score
116	SIQNVDDNNNTNVTDFAPKN	1
240	DNFTPKNKMPVTVSEGFNPE	0.981
294	HGFWGYWSGENHVDKKEEKL	0.962
155	SINRGGLTGNITKESNYSET	0.944
209	HDHTRQLTNDSDNRTKSEIF	0.911
95	RILDPNGYWNSTLRWPGSYS	0.894
261	LAVMSHDKKDEGKSKFVVHY	0.875
182	YRTLLDQSTSHKGVGWKVEA	0.758

Hypothetical protein B		
Position	Epitope	Score
262	VIQPAPKVCTPIKKEETHES	0.997
11	HDEKNEIREIEPVISRKASA	0.997
71	IGQIDQDRSNAQVDKTASLN	0.953

95	HDLDVHPIKKPDAEKTINDD	0.859
144	KLQMDEELKTARTNADVDAV	0.855

Staphylococcal secretory antigen (ssaA2)		
Position	Epitope	Score
122	NTNSYRTGGLGASYSTSSNN	0.999
181	VGGKIGSTWGNASNWANAAA	0.992
205	TVNNTPKAGAIMQTTQGAYG	0.992
238	VRVSEMNYGYGPGVVTSRTI	0.985
101	SRYNNYSNNNQSYNYNNYNS	0.982
51	AQGNYHYTWKGNWHPSQLNQ	0.956
158	SSGYTSGRNLYTSGQCTYYV	0.944
26	HASEQDNYGYNPNDPTSYSY	0.892
80	YNGYNNYNYNNYNNGYSYNN	0.87

N-acetylmuramoyl-L-alanine amidase (sle1)		
Position	Epitope	Score
135	GGSGGTTPTATTGSNGNASS	1
76	QNSSNTSPQAGSASSHTVQA	0.999
213	QSTPGPYGHVAYVERVNGDG	0.996
192	AGNAANDGYQVNNTPSVGSI	0.963
171	FDRRAQAGSPISTYWSDAKY	0.958
241	YTYGPYNMNYRTIPASEVSS	0.946
43	KYNTSVESIKQNNQLDNNLV	0.924
20	HHDAQASTQHTVQSGESLWS	0.897

Putative transglycosylase (IsaA)		
Position	Epitope	Score
126	VEAVSAPTYHNYSTSTTSSS	1
101	DYTTSYNQGSNVQSVSYNAQ	0.987
200	FQTMPGWGPTNTVDQQINAA	0.98
66	KDGFQYNFTSNGTTWSWSYE	0.977
179	IIARESNGQVNAYNPSGASG	0.976
17	TGYAAGTGHQAHAAEVNVDQ	0.898
147	RLSNGNTAGATGSSAAQIMA	0.848

Iron-regulated surface determinant protein B (IsdB)		
Position	Epitope	Score

471	SAKKEATPATPSKPTPSPVE	1
540	TPTKVVSTTQNVAKPTTASS	1
513	NDASSESGKDKTPATKPTKG	1
48	TNTEAQPKTEAVASPTTTSE	0.999
72	TKPVANAVSVSNKEVEAPTS	0.998
404	TISKDAKNNTRTIIFPYVEG	0.997
586	IKNTNDGHTQSQNNKNTQEN	0.991
95	EAKEVKEVKAPKETKEVKPA	0.99
154	EMKKKDGTQQFYHYASSVKP	0.985
347	DTKYVVYESVENNESMMDTF	0.972
133	IKNPAIKDKDHSAPNSRPID	0.948
382	MVMETTNDDYWKDFMVEGQR	0.924
236	IVSSTHFNNKEEKYDYTLME	0.912
450	DKEAFTKANTDKSNKKEQQD	0.881
280	APYKKAKTLERQVYELNKIQ	0.863
257	AQPIYNSADKFKTEEDYKAE	0.764
607	AKSLPQTGEESNKDMTLPLM	0.756

Putative transglycosylase (SceD)		
Position	Epitope	Score
139	TATPTESKSSEGSSVNVNAH	0.999
112	VQSQEVEAPQNAQTQQPQAS	0.999
91	DMSAQAPATNNVAPSAVQAN	0.994
195	APAKYKGVSPANAPESVQDA	0.989
164	QRESGGNIHAVNPTSGAAGK	0.914
70	HFESDGTHFAWNYHATGTNG	0.822
42	QSNDQTLNQKPIEAGAYNYT	0.817

Hypothetical protein A		
Position	Epitope	Score
82	GSTWGNANNWAAAAQGAGFT	0.999
38	NYNTTQTTTTTTTTTTSSI	0.993
113	QSSEGPFGHVAYVESVNSDG	0.958
6	TATTLTAGIGTALVGQAHHA	0.849

Fibrinogen-binding protein		
Position	Epitope	Score

82	NDGTFEYGARPQFNKPAAKT	0.995
17	GITTTTIASTADASEGYGPR	0.976
38	KKPVSINHNIVEYNDGTFKY	0.89
59	SRPKFNSTPKYIKFKHDYNI	0.808

Hypothetical protein C		
Position	Epitope	Score
124	QQQNTPPANKSNNTNTSSKQ	1
24	AINHSTSSSNNQEQENKATH	1
81	SNVPNNHSTAPSNTQPAQPK	1
59	DKADKNQNNGNNVTNDNQPS	0.998
103	NSNKNNTHDGGSQSSTNNQN	0.995

Hypothetical protein D		
Position	Epitope	Score
77	VNLIHFQHSYEKKKLQRQID	0.818
28	HAQTKVEKYNEYQTNFKKQV	0.806

Nickel-binding periplasmic protein		
Position	Epitope	Score
355	AGWKKGKDSDVRQKDGKNLE	0.994
377	MYYDKGSSSQKEQAEYLQAE	0.982
173	KDFKNGTTKDGVKKFDGTGP	0.978
281	MLVVNSGKKDNAVSDKTVRQ	0.94
416	ERRTSGDYDLMFNQTWGLLY	0.937
36	LTYTTVKDIGDMNPHVYGGS	0.936
443	AFKAKNGYESATSGIENKDK	0.922
79	AKKWDVSEDGKTYTFHLRDD	0.873
317	DGQEKPATQLFAKNVTDINF	0.862
236	SMKKGETNFAFTDDRGTDSL	0.803

ESAT-6 secretion system extracellular protein A		
Position	Epitope	Score
8	PEEIRAKSQSYGQGSDQIRQ	0.972
77	NSTADAVQEQDQQLSNNFGL	0.81

Table A.4: List of epitopes along with their position in the protein sequence and epitope score of the shortlisted potential vaccine candidates identified using AAPPred in Chapter 6.

Glutathione-binding protein GsiB		
Position	Epitope	Score
155	IYDTDAKSDVNQTPVGTGPY	1
32	SLPLKTKSIAPYETDVPVKI	1
406	KKDGAINKGDYNNSNVDDLI	1
222	KDDLITDVPVNKVQDIENNQ	1
290	YAKPATSPFNDKIPYIKEPK	1
195	QGKPKLDHITVTYQEDGNNR	1
377	KDRSAWDATMYSFGTIPRGD	1
355	SDAKKANIEIDIKSVDDIEG	1
63	QGKIEKALVKSYHQPNDTTL	1
263	KKMTKSVREALDHIIDRQGI	1
128	AKGQKLTIKTNSAYPELVSE	0.134
8	TIFSAVLVLSGCGSMHSSGK	0.078

Bifunctional autolysin		
Position	Epitope	Score
518	KSPVNVNQSYSIKPGTKLYT	1
105	TRANQSATTNNTQPVAKSTS	1
622	INAKNNGLFTTVYDKTGKPT	1
161	AAKPAALETQYKAAAPKAAT	1
74	DPAIVQPKTANNKTGNAQVS	1
926	VNAKDLTAPTAVKPTTSAAK	1
1040	PQVQRVPGKWTDANFNDVKH	1
446	ANNGVAQIKPTNSGLYTTVY	1
252	YRNGVGRPEGIVVHDTANDR	1
588	LVDTAKPTPTPTPKPSTPTT	1
799	TGIRASVYEKTAKNGAKYAD	1
203	PRSVAATPKTSLPKYKPQVN	1
759	VPAAPKKAVAQPKTAVKAYT	1
346	ATQLQYYGLKPDSAEYDGNG	1
1217	PGTHQYATDIDWANINAKII	1
714	QEAGAVSGTGNQTFKATKQQ	1
468	TGKATNEVQKTFAVSKTATL	1

673	IGWVKQGDVIYNNAKSPVNV	1
415	APWGTQSTTTPTTPSKPTTP	1
49	KATTEQAKAEVKNPTQNISG	1
1170	REGIKYAKQAGWDTVSKAIV	1
868	KYTVNKSNNGLSMVPWGTKN	1
1142	VVNNKVVTNSNTKYHNVFGI	1
126	TAPKTNTNVTNAGYSLVDDE	1
1008	KSTDLAKELIKYNQTGMTLN	1
549	SVSGSGNQTFKASKQQQIDK	1
182	SAPKAKTEATPKVTTFSASA	1
1101	ENQGAAFNKAAQMYGINEVY	1
1196	GNSYVKAGQNTLYKMRWNPA	1
956	GNGYYYVTPNSDTAKYSLKA	1
982	AVVKEQVINGQTWYYGKLSN	1
495	VQDYNSGNKFGWVKEGDVVY	1
229	IRKNNLKAPKIEEDYTSYFP	0.999
295	DGDRIIETAPTDYLSWGVGA	0.998
15	LTLVGSAVTAHQVQAAETTQ	0.97
736	DKSIYLFGTVNGKSGWVSKA	0.909
820	TFYVTKERAHGNETYVLLNN	0.903
321	INVEIVHTHDYASFARSMNN	0.742
368	WTHYAVSKYLGGTDHADPHG	0.709
905	KQVSVGKDVYLYGTINNRTG	0.064

Cell division protein DivIB		
Position	Epitope	Score
123	PQSAPKEQNGDSNDEETVTK	1
20	ELFTRNTSKKRRQRKRSKAT	1
407	DKNVTKSSQEENQAKEELQS	1
386	DLSVGASFIPYRGNTSSQSE	1
362	KYYPQMSQSLSRDSSGKLKT	1
303	KGTKEDDMIKALSEMTPEVR	1
279	LENGKLLKGSNDVKINDAPV	1
74	ENNDSASSRANNNNIDDSTD	1
41	FSNQNKDDTSQQADFDEEIY	1
245	SVEIHKQLPNTLNVDITENE	0.999

102	RYNQEIDDQNESNGIAFDNE	0.999
149	VTQLKPLTLEEKRKLRRKRQ	0.421
195	SKIAHVNINGNNHVSTSKIN	0.378

Staphopain B		
Position	Epitope	Score
177	KATPLPGNVKEKESAKTVSS	1
102	YYPVIKDGKIVYTLTLSPKN	1
290	IEYGKSQGRDIHYQEGVPSY	1
200	QELKNTVTPTKVEENEAIQE	1
222	VQYENTLKNFKIREQQFDNS	1
268	RTLYPEVSEQDLPNCSTFPN	0.993
64	FAGYAKALDKQSNAKTGKYE	0.963
344	VVGNAKINDQEKLIYWNPWD	0.443

Iron-regulated surface determinant protein A (IsdA)		
Position	Epitope	Score
297	VTKHNETPKQASKAKELPKT	1
266	KVQTPVKDVATAKSESNNQA	1
190	NNVKPVQPKPAQPKTPTEQT	1
52	NATNNQSTQVSQATSQPINF	1
118	FYNANNQELATTVVNDNKKA	1
211	PVQPKVEKVKPTVTTTSKVE	1
146	VEPGYKSLTTKVHIVVPQIN	1
9	KYQSEQRSSAMKKITMGTAS	0.925
83	HMDDYMQHPGKVIKQNNKYY	0.727

Putative leukocidin-like protein 1		
Position	Epitope	Score
114	SVSIQNVDDNNNTNVTDFAP	1
239	KDNFTPKNKMPVTVSEGFNP	1
307	DKKEEKLSALYEVDWKTHNV	1
23	LPATSYAKINSEIKAVSEKN	1
135	NQDESREVKYTYGYKTGGDF	1
207	MGHDHTRQLTNDSDNRTKSE	1
166	TKESNYSETISYQQPSYRTL	1
267	DKKDEGKSKFVVHYKRSMDE	0.997

45	GDTKMYTRTATTSDSQKNIT	0.938

Hypothetical protein B		
Position	Epitope	Score
9	LLHDEKNEIREIEPVISRKA	1
262	VIQPAPKVCTPIKKEETHES	1
221	DGNRMIDEDATLNDIKQHTQ	1
93	TIHDLDVHPIKKPDAEKTIN	0.951

Staphylococcal secretory antigen ssaA2		
Position	Epitope	Score
123	TNSYRTGGLGASYSTSSNNV	1
192	ASNWANAAARAGYTVNNTPK	1
248	GPGVVTSRTISASQAAGYNF	1
145	TTTMAPSSNGRSISSGYTSG	1
169	TSGQCTYYVFDRVGGKIGST	1
48	TIDAQGNYHYTWKGNWHPSQ	0.999
217	QTTQGAYGHVAYVESVNSNG	0.999
27	ASEQDNYGYNPNDPTSYSYT	0.169

N-acetylmuramoyl-L-alanine amidase sle1		
Position	Epitope	Score
128	QTLQIPNGGSGGTTPTATTG	1
214	STPGPYGHVAYVERVNGDGS	1
64	PGQVISVGGSDAQNSSNTSP	1
243	YGPYNMNYRTIPASEVSSYA	1
187	DAKYWAGNAANDGYQVNNTP	0.999
19	THHDAQASTQHTVQSGESLW	0.988
160	LYTAGQCTWYVFDRRAQAGS	0.891

Putative transglycosylase IsaA		
Position	Epitope	Score
143	SSSVRLSNGNTAGATGSSAA	1
181	ARESNGQVNAYNPSGASGLF	1
214	QQINAAVKAYKAQGLGAWGF	1
98	AGADYTTSYNQGSNVQSVSY	1
74	TSNGTTWSWSYEAANGQTAG	1
121	SSNSNVEAVSAPTYHNYSTS	1

44	HNHQDQLNAAPIKDGAYDIH	0.898

Iron-regulated surface determinant protein B (IsdB)		
Position	Epitope	Score
42	AEETGGTNTEAQPKTEAVAS	1
74	PVANAVSVSNKEVEAPTSET	1
508	SVEKENDASSESGKDKTPAT	1
114	AAKATNNTYPILNQELREAI	1
479	ATPSKPTPSPVEKESQKQDS	1
279	LAPYKKAKTLERQVYELNKI	1
597	QNNKNTQENKAKSLPQTGEE	1
411	NNTRTIIFPYVEGKTLYDAI	1
552	AKPTTASSKTTKDVVQTSAG	1
455	TKANTDKSNKKEQQDNSAKK	1
228	SNGTKAVKIVSSTHFNNKEE	1
529	PTKGEVESSSTTPTKVVSTT	1
167	YASSVKPARVIFTDSKPEIE	1
135	NPAIKDKDHSAPNSRPIDFE	1
198	FEVYEGDKKLPIKLVSYDTV	1
324	QVKSAITEFQNVQPTNEKMT	1
364	DTFVKHPIKTGMLNGKKYMV	1
386	TTNDDYWKDFMVEGQRVRTI	0.999
434	HVKTIDYDGQYHVRIVDKEA	0.966
576	KDSAPLQKANIKNTNDGHTQ	0.609

Putative transglycosylase SceD		
Position	Epitope	Score
102	VAPSAVQANQVQSQEVEAPQ	1
130	ASTSNNSQVTATPTESKSSE	1
187	LQSTWDSVAPAKYKGVSPAN	1
164	QRESGGNIHAVNPTSGAAGK	1
80	WNYHATGTNGADMSAQAPAT	1
47	TLNQKPIEAGAYNYTFDYEG	1
18	AGNAGHEAHASEADLNKASL	0.995
208	PESVQDAAAVKLYNTGGAGH	0.989

Hypothetical protein A

Position	Epitope	Score
142	SGGPFSVSSRTISASEAGNY	1
102	VNHTPSKGAILQSSEGPFGH	1
26	DAAENYTNYNNYNYNTTQTT	0.365
59	HSGNLYTAGQCTWYVYDKVG	0.105
Fibrinogen-binding protein		
Position	Epitope	Score
29	ASEGYGPREKKPVSINHNIV	1
90	ARPQFNKPAAKTDATIKKEQ	1
Hypothetical protein C		
Position	Epitope	Score
126	QNTPPANKSNNTNTSSKQTP	1
25	INHSTSSSNNQEQENKATHK	1
94	TQPAQPKDSNSNKNNTHDGG	1
69	NNVTNDNQPSSQSNVPNNHS	1
Hypothetical protein D		
Position	Epitope	Score
84	HSYEKKKLQRQIDLVLKYNT	0.294
34	EKYNEYQTNFKKQVNKKVVD	0.074
Nickel-binding periplasmic protein		
Position	Epitope	Score
342	KYDLKKAESLLDEAGWKKGK	1
317	DGQEKPATQLFAKNVTDINF	1
282	LVVNSGKKDNAVSDKTVRQA	1
445	KAKNGYESATSGIENKDKIY	1
475	NGKERSGAYKNILKQIDDEG	1
406	INGETSDKIAERRTSGDYDL	1
13	ASGLILTGCGGNKGLEEKKE	1
59	ESMIYEPLVRNTKDGIKPLL	1
503	SMTVVAPKDLEKVSFTQSQY	0.993
119	VQQNKKLHSWLKISTLIDNV	0.948
261	KQLKDTGDYQVKRSQPMNTK	0.108
ESAT-6 secretion system extracellular protein A		

Position	Epitope	Score
4	IKMSPEEIRAKSQSYGQGSD	1
76	LNSTADAVQEQDQQLSNNFG	0.999

Bibliography

D. M. Aanensen, E. J. Feil, M. T. G. Holden, J. Dordel, C. A. Yeats, A. Fedosejev, R. Goater, S. Castillo-Ramírez, J. Corander, C. Colijn, M. A. Chlebowicz, L. Schouls, M. Heck, G. Pluister, R. Ruimy, G. Kahlmeter, J. Åhman, E. Matuschek, A. W. Friedrich, J. Parkhill, S. D. Bentley, B. G. Spratt, and H. Grundmann. Whole-genome sequencing for routine pathogen surveillance in public health: a population snapshot of invasive *Staphylococcus aureus* in europe. *mBio*, 7(3):e00444–16, 2016. doi: 10.1128/mBio.00444-16. URL http://mbio.asm.org/content/7/3/e00444-16.abstract.

R. D. Abdi, B. E. Gillespie, J. Vaughn, C. Merrill, S. I. Headrick, D. B. Ensermu, D. H. D'Souza, G. E. Agga, R. A. Almeida, S. P. Oliver, and O. Kerro Dego. Antimicrobial resistance of *Staphylococcus aureus* isolates from dairy cows and genetic diversity of resistant isolates. *Foodborne Pathogens and Disease*, 15(7): 449–458, 2018. ISSN 1535-3141. doi: 10.1089/fpd.2017.2362. URL https://doi.org/10.1089/fpd.2017.2362.

E. P. Abraham, A. D. Gardner, E. Chain, N. G. Heatley, C. M. Fletcher, and M. A. Jennings. Further observations on penicillin. *Lancet*, 2:177–189, 1941. ISSN 0140-6736. URL <GotoISI>://WOS:000188728900120.

P. R. F. Adkins, S. Dufour, J. N. Spain, M. J. Calcutt, T. J. Reilly, G. C. Stewart, and J. R. Middleton. Cross-sectional study to identify staphylococcal species isolated from teat and inguinal skin of different-aged dairy heifers. *Journal of Dairy Science*, 101(4):3213–3225, 2018. ISSN 1525-3198 (Electronic) 0022-0302 (Linking). doi: 10.3168/jds.2017-13974. URL https://www.ncbi.nlm.nih.gov/pubmed/29397170.

M. Ahmadi, S. M. R. Rohani, and N. Ayremlou. Detection of *Staphylococcus aureus* in milk by pcr. *Comparative Clinical Pathology*, 19(1):91–94, 2010. ISSN

1618-5641. doi: 10.1007/s00580-009-0901-0. URL https://dx.doi.org/10.1007/s00580-009-0901-0.

M. Akkou, C. Bouchiat, K. Antri, M. Bes, A. Tristan, O. Dauwalder, P. Martins-Simoes, J. P. Rasigade, J. Etienne, F. Vandenesch, N. Ramdani-Bouguessa, and F. Laurent. New host shift from human to cows within *Staphylococcus aureus* involved in bovine mastitis and nasal carriage of animal's caretakers. *Vet Microbiology*, 223:173–180, 2018. ISSN 1873-2542 (Electronic) 0378-1135 (Linking). doi: 10.1016/j.vetmic.2018.08.003. URL https://www.ncbi.nlm.nih.gov/pubmed/30173744.

E. Aklilu and C. Hui Ying. First *mecC* and *mecA* positive livestock-associated methicillin resistant *Staphylococcus aureus* (*mecC* mrsa/la-mrsa) from dairy cattle in malaysia. *Microorganisms*, 8(2):147, 2020. ISSN 2076-2607. doi: 10.3390/microorganisms8020147. URL https://pubmed.ncbi.nlm.nih.gov/31973159https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7074714/.

M. T. Alam, r. Petit, Robert A., E. K. Crispell, T. A. Thornton, K. N. Conneely, Y. Jiang, S. W. Satola, and T. D. Read. Dissecting vancomycin-intermediate resistance in *Staphylococcus aureus* using genome-wide association. *Genome biology and evolution*, 6(5):1174–1185, 2014. ISSN 1759-6653. doi: 10.1093/gbe/evu092. URL https://pubmed.ncbi.nlm.nih.gov/24787619https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4040999/.

B. P. Alcock, A. R. Raphenya, T. T. Y. Lau, K. K. Tsang, M. Bouchard, A. Edalatmand, W. Huynh, A.-L. V. Nguyen, A. A. Cheng, S. Liu, S. Y. Min, A. Miroshnichenko, H.-K. Tran, R. E. Werfalli, J. A. Nasir, M. Oloni, D. J. Speicher, A. Florescu, B. Singh, M. Faltyn, A. Hernandez-Koutoucheva, A. N. Sharma, E. Bordeleau, A. C. Pawlowski, H. L. Zubyk, D. Dooley, E. Griffiths, F. Maguire, G. L. Winsor, R. G. Beiko, F. S. L. Brinkman, W. W. L. Hsiao, G. V. Domselaar, and A. G. McArthur. Card 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Research*, 48(D1): D517–D525, 2019. ISSN 0305-1048. doi: 10.1093/nar/gkz935. URL https://doi.org/10.1093/nar/gkz935.

F. Alonzo and V. J. Torres. The bicomponent pore-forming leucocidins of

Staphylococcus aureus. *Microbiology and Molecular Biology Reviews*, 78(2):199, 2014. doi: 10.1128/MMBR.00055-13. URL http://mmbr.asm.org/content/78/2/199.abstract.

r. Alonzo, F., L. Kozhaya, S. A. Rawlings, T. Reyes-Robles, A. L. DuMont, D. G. Myszka, N. R. Landau, D. Unutmaz, and V. J. Torres. Ccr5 is a receptor for *Staphylococcus aureus* leukotoxin ed. *Nature*, 493(7430):51–5, 2013. ISSN 0028-0836 (Print) 0028-0836. doi: 10.1038/nature11724.

S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. Basic local alignment search tool. *Journal of Molecular Biology*, 215(3):403–10, 1990. ISSN 0022-2836 (Print) 0022-2836. doi: 10.1016/s0022-2836(05)80360-2.

R. I. Aminov. A brief history of the antibiotic era: lessons learned and challenges for the future. *Frontiers in microbiology*, 1:134–134, 2010. ISSN 1664-302X. doi: 10.3389/fmicb.2010.00134. URL https://pubmed.ncbi.nlm.nih.gov/21687759https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3109405/.

J. C. Anderson. Mechanisms of staphylococcal virulence in relation to bovine mastitis. *British Veterinary Journal*, 132(3):229–245, 1976. ISSN 0007-1935. doi: https://doi.org/10.1016/S0007-1935(17)34682-1. URL https://www.sciencedirect.com/science/article/pii/S0007193517346821.

S. Andrews, 2010. URL http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

M. Antonio, N. McFerran, and M. J. Pallen. Mutations affecting the rossman fold of isoleucyl-trna synthetase are correlated with low-level mupirocin resistance in *Staphylococcus aureus*. *Antimicrobial Agents and Chemotherapy*, 46(2):438–442, 2002. doi: 10.1128/aac.46.2.438-442.2002. URL https://aac.asm.org/content/aac/46/2/438.full.pdf.

C. R. Arciola, D. Campoccia, S. Ravaioli, and L. Montanaro. Polysaccharide intercellular adhesin in biofilm: structural and regulatory aspects. *Frontiers in Cellular and Infection Microbiology*, 5(7), 2015. ISSN 2235-2988. doi: 10.3389/fcimb.2015.00007. URL https://www.frontiersin.org/article/10.3389/fcimb.2015.00007.

M. Á. Argudín, M. C. Mendoza, and M. R. Rodicio. Food poisoning and staphylococcus aureus enterotoxins. *Toxins*, 2(7):1751–1773, 2010.

B. M. Arlian and J. K. Tinker. Mucosal immunization with a *Staphylococcus aureus* isda-cholera toxin a2/b chimera induces antigen-specific th2-type responses in mice. *Clinical and vaccine immunology : CVI*, 18(9):1543–1551, 2011. ISSN 1556-679X 1556-6811. doi: 10.1128/CVI.05146-11. URL https://pubmed.ncbi.nlm.nih.gov/21734065https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3165230/.

L. Armand-Lefevre, R. Ruimy, and A. Andremont. Clonal comparison of *Staphylococcus aureus* isolates from healthy pig farmers, human controls, and pigs. *Emerging infectious diseases*, 11(5):711–714, 2005. ISSN 1080-6040 1080-6059. doi: 10.3201/eid1105.040866. URL https://pubmed.ncbi.nlm.nih.gov/15890125https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3320358/.

S. Ashraf, J. Cheng, and X. Zhao. Clumping factor a of *Staphylococcus aureus* interacts with annexina2 on mammary epithelial cells. *Scientific Reports*, 7(1): 40608, 2017. ISSN 2045-2322. doi: 10.1038/srep40608. URL https://dx.doi.org/10.1038/srep40608.

H. Atalla, C. Gyles, C. L. Jacob, H. Moisan, F. Malouin, and B. Mallard. Characterization of a staphylococcus aureus small colony variant (scv) associated with persistent bovine mastitis. *Foodborne pathogens and disease*, 5(6):785–799, 2008.

J. Bachmann, C. Helmschrodt, A. Richter, W. Heuwieser, and S. Bertulat. Residue concentration of cefquinome after intramammary dry cow therapy and short dry periods. *Journal of Dairy Science*, 101(8):7540–7550, 2018. ISSN 0022-0302. doi: https://doi.org/10.3168/jds.2017-13826. URL https://www.sciencedirect.com/science/article/pii/S0022030218305277.

F. Bahrami, D. L. Morris, and M. H. Pourgholami. Tetracyclines: Drugs with huge therapeutic potential. *Mini-Reviews in Medicinal Chemistry*, 12(1):44–52, 2012. ISSN 1389-5575/1875-5607. doi: http://dx.doi.org/10.2174/138955712798868977. URL http://www.eurekaselect.com/node/89329/article.

S. Baker, N. Thomson, F.-X. Weill, and K. E. Holt. Genomic insights into the emergence and spread of antimicrobial-resistant bacterial pathogens. *Science (New York, N.Y.)*, 360(6390):733–738, 2018. ISSN 1095-9203 0036-8075. doi: 10.1126/science.aar3777. URL https://pubmed.ncbi.nlm.nih.gov/29773743https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6510332/.

A. Baldwin, M. Loughlin, J. Caubilla-Barron, E. Kucerova, G. Manning, C. Dowson, and S. Forsythe. Multilocus sequence typing of cronobacter sakazakii and cronobacter malonaticus reveals stable clonal structures with clinical significance which do not correlate with biotypes. *BMC Microbiology*, 9(1):223, 2009. ISSN 1471-2180. doi: 10.1186/1471-2180-9-223. URL https://dx.doi.org/10.1186/1471-2180-9-223.

S. Bambini and R. Rappuoli. The use of genomics in microbial vaccine development. *Drug discovery today*, 14(5-6):252–260, 2009. ISSN 1359-6446.

A. Bankevich, S. Nurk, D. Antipov, A. A. Gurevich, M. Dvorkin, A. S. Kulikov, V. M. Lesin, S. I. Nikolenko, S. Pham, A. D. Prjibelski, A. V. Pyshkin, A. V. Sirotkin, N. Vyahhi, G. Tesler, M. A. Alekseyev, and P. A. Pevzner. Spades: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of computational biology : a journal of computational molecular cell biology*, 19(5): 455–477, 2012. ISSN 1557-8666 1066-5277. doi: 10.1089/cmb.2012.0021. URL https://pubmed.ncbi.nlm.nih.gov/22506599https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3342519/.

G. K. Bar-Gal, S. E. Blum, L. Hadas, R. Ehricht, S. Monecke, and G. Leitner. Host-specificity of *Staphylococcus aureus* causing intramammary infections in dairy animals assessed by genotyping and virulence genes. *Veterinary Microbiology*, 176(1):143–154, 2015. ISSN 0378-1135. doi: https://doi.org/10.1016/j.vetmic.2015.01.007. URL https://www.sciencedirect.com/science/article/pii/S0378113515000243.

M. Barber. Methicillin-resistant staphylococci. *Journal of Clinical Pathology*, 14(4): 385–393, 1961. ISSN 0021-9746. doi: 10.1136/jcp.14.4.385. URL https://dx.doi.org/10.1136/jcp.14.4.385.

M. Barber and M. Rozwadowska-Dowzenko. Infection by penicillin-resistant staphylococci. *Lancet*, pages 641–4, 1948.

H. W. Barkema, Y. H. Schukken, and R. N. Zadoks. Invited review: The role of cow, pathogen, and treatment regimen in the therapeutic success of bovine *Staphylococcus aureus* mastitis. *Journal of Dairy Science*, 89(6):1877–1895, 2006. ISSN 0022-0302. doi: 10.3168/jds.s0022-0302(06)72256-1. URL https://dx.doi.org/10.3168/jds.s0022-0302(06)72256-1.

J. Barlow. Mastitis therapy and antimicrobial susceptibility: a multispecies review with a focus on antibiotic treatment of mastitis in dairy cattle. *Journal of Mammary Gland Biology and Neoplasia*, 16(4):383–407, 2011. ISSN 1573-7039. doi: 10.1007/s10911-011-9235-z. URL https://doi.org/10.1007/s10911-011-9235-z.

J. R. Barreiro, J. L. Gonçalves, R. Grenfell, R. F. Leite, L. Juliano, and M. V. Santos. Direct identification of bovine mastitis pathogens by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry in pre-incubated milk. *Brazilian journal of microbiology : [publication of the Brazilian Society for Microbiology]*, 49(4):801–807, 2018. ISSN 1678-4405 1517-8382. doi: 10.1016/j.bjm.2018.04.012. URL https://www.ncbi.nlm.nih.gov/pubmed/30177270https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6175725/.

F. F. Barrett, R. F. McGehee, and M. Finland. Methicillin-resistant *Staphylococcus aureus* at boston city hospital. *New England Journal of Medicine*, 279(9):441–448, 1968. doi: 10.1056/nejm196808292790901. URL https://www.nejm.org/doi/full/10.1056/NEJM196808292790901.

B. Barrio, F. Vangroenweghe, H. Dosogne, and C. Burvenich. Decreased neutrophil bactericidal activity during phagocytosis of a slime-producing *Staphylococcus aureus* strain. *Veterinary Research*, 31(6):603–9, 2000. ISSN 0928-4249 (Print) 0928-4249. doi: 10.1051/vetres:2000143.

M. D. Bartels, A. Petersen, P. Worning, J. B. Nielsen, H. Larner-Svensson, H. K. Johansen, L. P. Andersen, J. O. Jarløv, K. Boye, A. R. Larsen, and H. Westh. Comparing whole-genome sequencing with sanger sequencing for spa typing of methicillin-resistant *Staphylococcus aureus*. *Journal of clinical microbiology*, 52 (12):4305–4308, 2014. ISSN 1098-660X 0095-1137. doi: 10.1128/JCM.01979-14. URL https://pubmed.ncbi.nlm.nih.gov/25297335https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4313303/.

R. Baselga, I. Albizu, and B. Amorena. *Staphylococcus aureus* capsule and slime as virulence factors in ruminant mastitis. a review. *Veterinary Microbiology*, 39(3): 195–204, 1994. ISSN 0378-1135. doi: https://doi.org/10.1016/0378-1135(94)90157-0. URL https://www.sciencedirect.com/science/article/pii/0378113594901570.

A. Bates, R. Laven, O. Bork, M. Hay, J. McDowell, and B. Saldias. Selective and deferred treatment of clinical mastitis in seven new zealand dairy herds. *Preventive Veterinary Medicine*, 176:104915, 2020. ISSN 0167-5877. doi: https://doi.org/10.1016/j.prevetmed.2020.104915. URL http://www.sciencedirect.com/science/article/pii/S0167587719305392.

D. Baxby. *Jenner's smallpox vaccine: the riddle of vaccinia virus and its origin*, volume 14. Heinemann Educational Books London, 1981.

J. D. Baxter, G. W. Rogers, S. B. Spencer, and R. J. Eberhart. The effect of milking machine liner slip on new intramammary infections. *Journal of Dairy Science*, 75 (4):1015–8, 1992. ISSN 0022-0302 (Print) 0022-0302. doi: 10.3168/jds.S0022-0302(92)77844-8.

S. Ben Ayed, I. Boutiba-Ben Boubaker, E. Samir, and S. Ben Redjeb. Prevalence of agr specificity groups among methicilin resistant *Staphylococcus aureus* circulating at charles nicolle hospital of tunis. *Pathol Biol (Paris)*, 54(8-9):435–8, 2006. ISSN 0369-8114 (Print) 0369-8114. doi: 10.1016/j.patbio.2006.07.010.

J. B. Benoit, D. N. Frank, and M. T. Bessesen. Genomic evolution of *Staphylococcus aureus* isolates colonizing the nares and progressing to bacteremia. *PloS one*, 13 (5):e0195860, 2018.

S. P. Bergin, T. L. Holland, V. G. Fowler, and S. Y. C. Tong. *Bacteremia, Sepsis, and Infective Endocarditis Associated with Staphylococcus aureus*, pages 263–296. Springer International Publishing, 2015. ISBN 0070-217X. doi: 10.1007/82_2015_5001. URL https://dx.doi.org/10.1007/82_2015_5001.

E. A. Berry and J. E. Hillerton. The effect of selective dry cow treatment on new intramammary infections. *Journal of Dairy Science*, 85(1):112–121, 2002. ISSN 0022-0302. doi: https://doi.org/10.3168/jds.S0022-0302(02)74059-9. URL https://www.sciencedirect.com/science/article/pii/S0022030202740599.

S. Besier, A. Ludwig, V. Brade, and T. A. Wichelhaus. Molecular analysis of fusidic acid resistance in *Staphylococcus aureus*. *Molecular Microbiology*, 47(2):463–469, 2003. ISSN 0950-382X. doi: 10.1046/j.1365-2958.2003.03307.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1046/j.1365-2958.2003.03307.x.

S. Bhardwaj. *Chapter 21 - Vaccines*, pages 341–353. Academic Press, Boston, 2018. ISBN 978-0-12-802103-3. doi: https://doi.org/10.1016/B978-0-12-802103-3.00022-5. URL https://www.sciencedirect.com/science/article/pii/B9780128021033000225.

R. Biswas, L. Voggu, U. K. Simon, P. Hentschel, G. Thumm, and F. Götz. Activity of the major staphylococcal autolysin atl. *FEMS Microbiol Lett*, 259(2):260–8, 2006. ISSN 0378-1097 (Print) 0378-1097. doi: 10.1111/j.1574-6968.2006.00281.x.

J. E. Blair and R. E. Williams. Phage typing of staphylococci. *Bulletin of the World Health Organization*, 24(6):771–784, 1961. ISSN 0042-9686. URL https://www.ncbi.nlm.nih.gov/pubmed/20604092https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2555522/.

J. M. A. Blair, M. A. Webber, A. J. Baylay, D. O. Ogbolu, and L. J. V. Piddock. Molecular mechanisms of antibiotic resistance. *Nature Reviews Microbiology*, 13 (1):42–51, 2015. ISSN 1740-1534. doi: 10.1038/nrmicro3380. URL https://doi.org/10.1038/nrmicro3380.

D. S. Blanc, P. Francioli, and P. M. Hauser. Poor value of pulsed-field gel electrophoresis to investigate long-term scale epidemiology of methicillin-resistant *Staphylococcus aureus*. *Infection, Genetics and Evolution*, 2(2):145–148, 2002. ISSN 1567-1348. doi: 10.1016/s1567-1348(02)00093-x. URL https://dx.doi.org/10.1016/s1567-1348(02)00093-x.

J. S. Blevins, A. F. Gillaspy, T. M. Rechtin, B. K. Hurlburt, and M. S. Smeltzer. The staphylococcal accessory regulator (sar) represses transcription of the *Staphylococcus aureus* collagen adhesin gene (cna) in an agr-independent manner. *Mol Microbiol*, 33(2):317–26, 1999. ISSN 0950-382X (Print) 0950-382x. doi: 10.1046/j.1365-2958.1999.01475.x.

P. Blowey, Roger; Edmondson. *Teat and udder defences against mastitis*, pages 20–32. CABI Publishing, UK, 2010. doi: 10.1079/9781845935504.0000.

E. Boerhout, M. Vrieling, L. Benedictus, I. Daemen, L. Ravesloot, V. Rutten, P. Nuijten, J. van Strijp, A. Koets, and S. Eisenberg. Immunization routes in cattle impact the levels and neutralizing capacity of antibodies induced against *S. aureus* immune evasion proteins. *Veterinary Research*, 46(1):115, 2015. ISSN 1297-9716. doi: 10.1186/s13567-015-0243-7. URL https://doi.org/10.1186/s13567-015-0243-7.

A. M. Bolger, M. Lohse, and B. Usadel. Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics*, 30(15):2114–2120, 2014. ISSN 1460-2059. doi: 10.1093/bioinformatics/btu170. URL https://dx.doi.org/10.1093/bioinformatics/btu170.

M. Boolchandani, A. W. D'Souza, and G. Dantas. Sequencing-based methods and resources to study antimicrobial resistance. *Nature Reviews Genetics*, 20(6): 356–370, 2019.

V. Bortolaia, R. S. Kaas, E. Ruppe, M. C. Roberts, S. Schwarz, V. Cattoir, A. Philippon, R. L. Allesoe, A. R. Rebelo, A. F. Florensa, L. Fagelhauer, T. Chakraborty, B. Neumann, G. Werner, J. K. Bender, K. Stingl, M. Nguyen, J. Coppens, B. B. Xavier, S. Malhotra-Kumar, H. Westh, M. Pinholt, M. F. Anjum, N. A. Duggett, I. Kempf, S. Nykäsenoja, S. Olkkola, K. Wieczorek, A. Amaro, L. Clemente, J. Mossong, S. Losch, C. Ragimbeau, O. Lund, and F. M. Aarestrup. Resfinder 4.0 for predictions of phenotypes from genotypes. *J Antimicrob Chemother*, 2020. ISSN 0305-7453. doi: 10.1093/jac/dkaa345.

R. Boss, A. Cosandey, M. Luini, K. Artursson, M. Bardiau, F. Breitenwieser, E. Hehenberger, T. Lam, M. Mansfeld, A. Michel, G. Mösslacher, J. Naskova, S. Nelson, O. Podpečan, A. Raemy, E. Ryan, O. Salat, P. Zangerl, A. Steiner, and H. U. Graber. Bovine *Staphylococcus aureus*: Subtyping, evolution, and zoonotic transfer. *Journal of Dairy Science*, 99(1):515–528, 2016. ISSN 0022-0302. doi: 10.3168/jds.2015-9589. URL https://dx.doi.org/10.3168/jds.2015-9589.

A. L. Bottomley, A. F. Kabli, A. F. Hurd, R. D. Turner, J. Garcia-Lara, and S. J. Foster. *Staphylococcus aureus* divib is a peptidoglycan-binding protein that is required for a morphological checkpoint in cell division. *Molecular Microbiology*, 94(5): 1041–1064, 2014. ISSN 0950-382X. doi: https://doi.org/10.1111/mmi.12813. URL https://doi.org/10.1111/mmi.12813.

D. Bouchard, V. Peton, S. Almeida, C. Le Maréchal, A. Miyoshi, V. Azevedo, N. Berkova, L. Rault, P. François, J. Schrenzel, S. Even, D. Hernandez, and Y. Le Loir. Genome sequence of *Staphylococcus aureus* newbould 305, a strain associated with mild bovine mastitis. *Journal of bacteriology*, 194(22):6292–6293, 2012. ISSN 1098-5530 0021-9193. doi: 10.1128/JB.01188-12. URL https://pubmed.ncbi.nlm.nih.gov/23105046https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3486374/.

A. J. Bradley, J. E. Breen, B. Payne, V. White, and M. J. Green. An investigation of the efficacy of a polyvalent mastitis vaccine using different vaccination regimens under field conditions in the united kingdom. *Journal of Dairy Science*, 98(3): 1706–1720, 2015a. ISSN 0022-0302. doi: https://doi.org/10.3168/jds.2014-8332. URL http://www.sciencedirect.com/science/article/pii/S0022030214008741.

P. Bradley, N. C. Gordon, T. M. Walker, L. Dunn, S. Heys, B. Huang, S. Earle, L. J. Pankhurst, L. Anson, M. de Cesare, P. Piazza, A. A. Votintseva, T. Golubchik, D. J. Wilson, D. H. Wyllie, R. Diel, S. Niemann, S. Feuerriegel, T. A. Kohl, N. Ismail, S. V. Omar, E. G. Smith, D. Buck, G. McVean, A. S. Walker, T. E. A. Peto, D. W. Crook, and Z. Iqbal. Rapid antibiotic-resistance predictions from genome sequence data for *Staphylococcus aureus* and mycobacterium tuberculosis. *Nature Communications*, 6(1):10063, 2015b. ISSN 2041-1723. doi: 10.1038/ncomms10063. URL https://doi.org/10.1038/ncomms10063.

R. A. Brady, G. A. May, J. G. Leid, M. L. Prior, J. W. Costerton, and M. E. Shirtliff. Resolution of*Staphylococcus aureus* biofilm infection using vaccination and antibiotic treatment. *Infection and Immunity*, 79(4):1797, 2011. doi: 10.1128/IAI.00451-10. URL http://iai.asm.org/content/79/4/1797.abstract.

K. Breyne, R. W. Honaker, Z. Hobbs, M. Richter, M. Żaczek, T. Spangler, J. Steenbrugge, R. Lu, A. Kinkhabwala, B. Marchon, E. Meyer, and L. Mokres. Efficacy and safety of a bovine-associated *Staphylococcus aureus* phage cocktail in a murine model of mastitis. *Frontiers in Microbiology*, 8(2348), 2017. ISSN 1664-302X. doi: 10.3389/fmicb.2017.02348. URL https://www.frontiersin.org/article/10.3389/fmicb.2017.02348.

P. Briaud, J. Baude, S. Bastien, L. Camus, F. Vandenesch, and K. Moreau. Nork, a

novel norfloxacin efflux pump in *Staphylococcus aureus*. *bioRxiv*, page 850768, 2019. doi: 10.1101/850768. URL https://www.biorxiv.org/content/biorxiv/early/2019/11/24/850768.full.pdf.

M. J. Brittnacher, C. Fong, H. Hayden, M. Jacobs, M. Radey, and L. Rohmer. Pgat: a multistrain analysis resource for microbial genomes. *Bioinformatics*, 27(17): 2429–2430, 2011.

B. M. Broker, D. Mrochen, and V. Péton. The t cell response to *Staphylococcus aureus*. *Pathogens*, 5(1), 2016. ISSN 2076-0817. doi: 10.3390/pathogens5010031.

E. Brouillette, B. G. Talbot, and F. Malouin. The fibronectin-binding proteins of *Staphylococcus aureus* may promote mammary gland colonization in a lactating mouse model of mastitis. *Infection and Immunity*, 71(4):2292–2295, 2003. ISSN 0019-9567. doi: 10.1128/iai.71.4.2292-2295.2003. URL https://dx.doi.org/10.1128/iai.71.4.2292-2295.2003.

E. Brouillette, A. Martinez, B. J. Boyll, N. E. Allen, and F. Malouin. Persistence of a staphylococcus aureus small-colony variant under antibiotic pressure in vivo. *FEMS Immunology & Medical Microbiology*, 41(1):35–41, 2004.

C. Bru, E. Courcelle, S. Carrère, Y. Beausse, S. Dalmar, and D. Kahn. The prodom database of protein domain families: more emphasis on 3d. *Nucleic acids research*, 33(suppl_1):D212–D215, 2005.

M. Bryan and S. Hea. A survey of antimicrobial use in dairy cows from farms in four regions of new zealand. *New Zealand Veterinary Journal*, 65(2):93–98, 2017. ISSN 0048-0169. doi: 10.1080/00480169.2016.1256794. URL https://dx.doi.org/10.1080/00480169.2016.1256794.

H. S. Y. Bryan M. *Evaluation of* Staphylococcus aureus *vaccine in dairy cattle in New Zealand*, volume Pan Pacific (NZVA and AVA) Veterinary Conference 2015 of *Proceedings of the Australian Veterinary Association*. Australian Veterinary Association, 2015.

K. E. Budd, F. McCoy, S. Monecke, P. Cormican, J. Mitchell, and O. M. Keane. Extensive genomic diversity among bovine-adapted *Staphylococcus aureus*: Evidence for a genomic rearrangement within cc97. *Plos One*, 10(8):e0134592, 2015. ISSN 1932-6203. doi: 10.1371/journal.pone.0134592. URL https://dx.doi.org/10.1371/journal.pone.0134592.

S. Burgess and N. French. Antimicrobial resistant bacteria in dairy cattle: A review. *New Zealand Food Safety and Science Research Centre*, 2017. URL https://www.nzfssrc.org.nz/node/79.

J. L. Burton and R. J. Erskine. Immunity and mastitis some new ideas for an old disease. *Veterinary Clinics: Food Animal Practice*, 19(1):1–45, 2003. ISSN 0749-0720.

M. L. Burts, W. A. Williams, K. DeBord, and D. M. Missiakas. Esxa and esxb are secreted by an esat-6-like system that is required for the pathogenesis of *Staphylococcus aureus* infections. *Proc Natl Acad Sci U S A*, 102(4):1169–74, 2005. ISSN 0027-8424 (Print) 0027-8424. doi: 10.1073/pnas.0405620102.

K. Bush. Past and present perspectives on β-lactamases. *Antimicrobial Agents and Chemotherapy*, 62(10):e01076–18, 2018. doi: 10.1128/aac.01076-18. URL https://aac.asm.org/content/aac/62/10/e01076-18.full.pdf.

K. Bush and P. A. Bradford. β-lactams and β-lactamase inhibitors: an overview. *Cold Spring Harbor perspectives in medicine*, 6(8):a025247, 2016.

C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, and T. L. Madden. Blast+: architecture and applications. *BMC Bioinformatics*, 10(1):421, 2009. ISSN 1471-2105. doi: 10.1186/1471-2105-10-421. URL https://dx.doi.org/10.1186/1471-2105-10-421.

R. Cantón, D. M. Livermore, M. I. Morosini, J. Díaz-Regañón, G. M. Rossolini, P. S. Group, and P. S. Group. Etest® versus broth microdilution for ceftaroline mic determination with *Staphylococcus aureus*: results from premium, a european multicentre study. *Journal of Antimicrobial Chemotherapy*, page dkw442, 2016.

P. Carlos, V. Roupie, S. Holbert, F. Ascencio, K. Huygen, G. Gomez-Anduro, M. Branger, M. Reyes-Becerril, and C. Angulo. In silico epitope analysis of unique and membrane associated proteins from mycobacterium avium subsp. paratuberculosis for immunogenicity and vaccine evaluation. *Journal of theoretical biology*, 384:1–9, 2015. ISSN 0022-5193.

S. K. Carlson, D. L. Erickson, and E. Wilson. *Staphylococcus aureus* metal acquisition in the mastitic mammary gland. *Microbial Pathogenesis*, 144:104179, 2020. ISSN 0882-4010. doi: https://doi.org/10.1016/j.micpath.2020.104179. URL

https:
//www.sciencedirect.com/science/article/pii/S0882401020304666.

G. P. Carter, M. B. Schultz, S. L. Baines, A. Gonçalves da Silva, H. Heffernan, A. Tiong, P. H. Pham, I. R. Monk, T. P. Stinear, B. P. Howden, et al. Topical antibiotic use coselects for the carriage of mobile genetic elements conferring resistance to unrelated antimicrobials in staphylococcus aureus. *Antimicrobial agents and chemotherapy*, 62(2):e02000–17, 2018.

M. Castanheira, H. S. Sader, D. J. Farrell, R. E. Mendes, and R. N. Jones. Activity of ceftaroline-avibactam tested against gram-negative organism populations, including strains expressing one or more β-lactamases and methicillin-resistant *Staphylococcus aureus* carrying various staphylococcal cassette chromosome mec types. *Antimicrobial agents and chemotherapy*, 56(9):4779–4785, 2012. ISSN 1098-6596 0066-4804. doi: 10.1128/AAC.00817-12. URL https://pubmed.ncbi.nlm.nih.gov/22733066https:
//www.ncbi.nlm.nih.gov/pmc/articles/PMC3421892/.

I. G. Castilho, S. T. A. Dantas, H. Langoni, J. P. Araújo, A. Fernandes, F. C. L. Alvarenga, L. Maia, D. Q. Cagnini, and V. L. M. Rall. Host-pathogen interactions in bovine mammary epithelial cells and hela cells by *Staphylococcus aureus* isolated from subclinical bovine mastitis. *Journal of Dairy Science*, 100(8): 6414–6421, 2017. ISSN 0022-0302. doi: 10.3168/jds.2017-12700. URL https://dx.doi.org/10.3168/jds.2017-12700.

J. J. Champoux. Dna topoisomerases: Structure, function, and mechanism. *Annual Review of Biochemistry*, 70(1):369–413, 2001. ISSN 0066-4154. doi: 10.1146/annurev.biochem.70.1.369. URL https://dx.doi.org/10.1146/annurev.biochem.70.1.369.

G. H. Chapman. The significance of sodium chloride in studies of staphylococci. *Journal of bacteriology*, 50(2):201–203, 1945. ISSN 0021-9193 1098-5530. URL https://pubmed.ncbi.nlm.nih.gov/16560988https:
//www.ncbi.nlm.nih.gov/pmc/articles/PMC374126/.

J. Chen, H. Liu, J. Yang, and K. C. Chou. Prediction of linear b-cell epitopes using amino acid pair antigenicity scale. *Amino Acids*, 33(3):423–8, 2007. ISSN 0939-4451. doi: 10.1007/s00726-006-0485-9.

L. Chen, J. Yang, J. Yu, Z. Yao, L. Sun, Y. Shen, and Q. Jin. Vfdb: a reference database for bacterial virulence factors. *Nucleic Acids Res*, 33(Database issue): D325–8, 2005. ISSN 0305-1048 (Print) 0305-1048. doi: 10.1093/nar/gki008.

Q. Chen, S. Xie, X. Lou, S. Cheng, X. Liu, W. Zheng, Z. Zheng, and H. Wang. Biofilm formation and prevalence of adhesion genes among *Staphylococcus aureus* isolates from different food sources. *MicrobiologyOpen*, 9(1):e00946, 2020. ISSN 2045-8827. doi: https://doi.org/10.1002/mbo3.946. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/mbo3.946.

W. N. Cheng and S. G. Han. Bovine mastitis: risk factors, therapeutic strategies, and alternative treatments - a review. *Asian-Australasian journal of animal sciences*, 33 (11):1699–1713, 2020. ISSN 1011-2367 1976-5517. doi: 10.5713/ajas.20.0156. URL https://pubmed.ncbi.nlm.nih.gov/32777908https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7649072/.

O. Chesneau, K. Tsvetkova, and P. Courvalin. Resistance phenotypes conferred by macrolide phosphotransferases. *FEMS Microbiology Letters*, 269(2):317–322, 2007. ISSN 0378-1097. doi: 10.1111/j.1574-6968.2007.00643.x. URL https://doi.org/10.1111/j.1574-6968.2007.00643.x.

A. L. Cheung, K. J. Eberhardt, E. Chung, M. R. Yeaman, P. M. Sullam, M. Ramos, A. S. Bayer, et al. Diminished virulence of a sar-/agr-mutant of *Staphylococcus aureus* in the rabbit model of endocarditis. *The Journal of clinical investigation*, 94 (5):1815–1822, 1994.

Y. Chien and A. L. Cheung. Molecular interactions between two global regulators, sar and agr, in *Staphylococcus aureus*. *J Biol Chem*, 273(5):2645–52, 1998. ISSN 0021-9258 (Print) 0021-9258. doi: 10.1074/jbc.273.5.2645.

I. Chopra and M. Roberts. Tetracycline antibiotics: mode of action, applications, molecular biology, and epidemiology of bacterial resistance. *Microbiology and molecular biology reviews : MMBR*, 65(2):232–260, 2001. ISSN 1092-2172 1098-5557. doi: 10.1128/MMBR.65.2.232-260.2001. URL https://pubmed.ncbi.nlm.nih.gov/11381101https://www.ncbi.nlm.nih.gov/pmc/articles/PMC99026/.

K. Y. L. Chua, T. P. Stinear, and B. P. Howden. Functional genomics of *Staphylococcus aureus*. *Briefings in Functional Genomics*, 12(4):305–315, 2013.

ISSN 2041-2649. doi: 10.1093/bfgp/elt006. URL https://dx.doi.org/10.1093/bfgp/elt006.

C. U. Chukwudi. rrna binding sites and the molecular mechanism of action of the tetracyclines. *Antimicrobial agents and chemotherapy*, 60(8):4433–4441, 2016. ISSN 1098-6596 0066-4804. doi: 10.1128/AAC.00594-16. URL https://www.ncbi.nlm.nih.gov/pubmed/27246781https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4958212/.

J. L. B. Cineros and O. Lund. Kmerfinderjs: A client-server method for fast species typing of bacteria over slow internet connections. *bioRxiv*, page 145284, 2017. doi: 10.1101/145284. URL http://biorxiv.org/content/early/2017/06/02/145284.abstract.

S. R. Clarke and S. J. Foster. Isda protects *Staphylococcus aureus* against the bactericidal protease activity of apolactoferrin. *Infect Immun*, 76(4):1518–26, 2008. ISSN 0019-9567 (Print) 0019-9567. doi: 10.1128/iai.01530-07.

S. R. Clarke, M. D. Wiltshire, and S. J. Foster. Isda of *Staphylococcus aureus* is a broad spectrum, iron-regulated adhesin. *Mol Microbiol*, 51(5):1509–19, 2004. ISSN 0950-382X (Print) 0950-382x. doi: 10.1111/j.1365-2958.2003.03938.x.

S. R. Clarke, R. Mohamed, L. Bian, A. F. Routh, J. F. Kokai-Kun, J. J. Mond, A. Tarkowski, and S. J. Foster. The *Staphylococcus aureus* surface protein isda mediates resistance to innate defenses of human skin. *Cell Host Microbe*, 1(3): 199–212, 2007. ISSN 1931-3128. doi: 10.1016/j.chom.2007.04.005.

CLSI. Performance standards for antimicrobial disk and dilution susceptibility tests for bacteria isolated from animals. Report, Clinical and Laboratory Standards Institute, 2018.

CLSI. Performance standards for antimicrobial disk and dilution susceptibility tests for bacteria isolated from animals. Report, Clinical and Laboratory Standards Institute, 2018a.

F. Coll, R. McNerney, M. D. Preston, J. A. Guerra-Assunção, A. Warry, G. Hill-Cawthorne, K. Mallard, M. Nair, A. Miranda, A. Alves, et al. Rapid determination of anti-tuberculosis drug resistance from whole-genome sequences. *Genome medicine*, 7(1):1–10, 2015.

P. D. Constable and D. E. Morin. Treatment of clinical mastitis: Using antimicrobial susceptibility profiles for treatment decisions. *Veterinary Clinics of North America: Food Animal Practice*, 19(1):139–155, 2003. ISSN 0749-0720. doi: https://doi.org/10.1016/S0749-0720(02)00068-3. URL https://www.sciencedirect.com/science/article/pii/S0749072002000683.

G. A. Contreras and J. M. Rodríguez. Mastitis: comparative etiology and epidemiology. *Journal of mammary gland biology and neoplasia*, 16(4):339–356, 2011. ISSN 1083-3021.

J. E. Coombe, S. M. Tymms, and M. Humphris. Antimicrobial stewardship in the dairy industry: responding to the threat of antimicrobial resistance. *Australian Veterinary Journal*, 97(7):231–232, 2019. ISSN 0005-0423. doi: https://doi.org/10.1111/avj.12807. URL https://doi.org/10.1111/avj.12807.

S. S. Costa, M. Viveiros, L. Amaral, and I. Couto. Multidrug efflux pumps in *Staphylococcus aureus*: an update. *The open microbiology journal*, 7:59–71, 2013. ISSN 1874-2858. doi: 10.2174/1874285801307010059. URL https://www.ncbi.nlm.nih.gov/pubmed/23569469https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3617543/.

J. Cote-Gravel and F. Malouin. Symposium review: Features of *Staphylococcus aureus* mastitis pathogenesis that guide vaccine development strategies. *Journal of Dairy Science*, 102(5):4727–4740, 2019. ISSN 1525-3198 (Electronic) 0022-0302 (Linking). doi: 10.3168/jds.2018-15272. URL https://www.ncbi.nlm.nih.gov/pubmed/30580940.

M. P. Cox, D. A. Peterson, and P. J. Biggs. Solexaqa: At-a-glance quality assessment of illumina second-generation sequencing data. *BMC Bioinformatics*, 11(1):485, 2010. ISSN 1471-2105. doi: 10.1186/1471-2105-11-485. URL https://doi.org/10.1186/1471-2105-11-485.

G. E. Crooks, G. Hon, J.-M. Chandonia, and S. E. Brenner. Weblogo: a sequence logo generator. *Genome research*, 14(6):1188–1190, 2004.

M. P. Crotty, T. Krekel, C.-A. D. Burnham, and D. J. Ritchie. New gram-positive agents: the next generation of oxazolidinones and lipoglycopeptides. *Journal of*

Clinical Microbiology, 54(9):2225–2232, 2016. doi: 10.1128/jcm.03395-15. URL https://jcm.asm.org/content/jcm/54/9/2225.full.pdf.

M. Cruciani, M. P. Etna, R. Camilli, E. Giacomini, Z. A. Percario, M. Severa, S. Sandini, F. Rizzo, V. Brandi, G. Balsamo, F. Polticelli, E. Affabris, A. Pantosti, F. Bagnoli, and E. M. Coccia. *Staphylococcus aureus* esx factors control human dendritic cell functions conditioning th1/th17 response. *Front Cell Infect Microbiol*, 7:330, 2017. ISSN 2235-2988. doi: 10.3389/fcimb.2017.00330.

C. Cucarella, M. A. Tormo, C. Úbeda, M. P. Trotonda, M. Monzón, C. Peris, B. Amorena, I. n. Lasa, and J. R. Penadés. Role of biofilm-associated protein bap in the pathogenesis of bovine *Staphylococcus aureus*. *Infection and Immunity*, 72 (4):2177–2185, 2004. doi: 10.1128/iai.72.4.2177-2185.2004. URL https://iai.asm.org/content/iai/72/4/2177.full.pdf.

D. Cue, M. Lei, and C. Lee. Genetic regulation of the intercellular adhesion locus in staphylococci. *Frontiers in Cellular and Infection Microbiology*, 2(38), 2012. ISSN 2235-2988. doi: 10.3389/fcimb.2012.00038. URL https://www.frontiersin.org/article/10.3389/fcimb.2012.00038.

L. Cui, X. Ma, K. Sato, K. Okuma, F. C. Tenover, E. M. Mamizuka, C. G. Gemmell, M.-N. Kim, M.-C. Ploy, N. El-Solh, V. Ferraz, and K. Hiramatsu. Cell wall thickening is a common feature of vancomycin resistance in *Staphylococcus aureus*. *Journal of clinical microbiology*, 41(1):5–14, 2003. ISSN 0095-1137 1098-660X. doi: 10.1128/jcm.41.1.5-14.2003. URL https://pubmed.ncbi.nlm.nih.gov/12517819https://www.ncbi.nlm.nih.gov/pmc/articles/PMC149586/.

L. Cui, A. Iwamoto, J.-Q. Lian, H.-m. Neoh, T. Maruyama, Y. Horikawa, and K. Hiramatsu. Novel mechanism of antibiotic resistance originating in vancomycin-intermediate *Staphylococcus aureus*. *Antimicrobial agents and chemotherapy*, 50(2):428–438, 2006. ISSN 0066-4804 1098-6596. doi: 10.1128/AAC.50.2.428-438.2006. URL https://pubmed.ncbi.nlm.nih.gov/16436693https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1366884/.

C. Cuny, L. H. Wieler, and W. Witte. Livestock-associated mrsa: The impact on humans. *Antibiotics (Basel, Switzerland)*, 4(4):521–543, 2015. ISSN 2079-6382.

doi: 10.3390/antibiotics4040521. URL https://pubmed.ncbi.nlm.nih.gov/27025639https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4790311/.

D. M. Daigle, D. W. Hughes, and G. D. Wright. Prodigious substrate specificity of aac(6′)-aph(2″), an aminoglycoside antibiotic resistance determinant in enterococci and staphylococci. *Chemistry and biology*, 6(2):99–110, 1999. ISSN 1074-5521. doi: 10.1016/s1074-5521(99)80006-4. URL https://dx.doi.org/10.1016/s1074-5521(99)80006-4.

G. E. Dale, C. Broger, A. D' Arcy, P. G. Hartman, R. DeHoogt, S. Jolidon, I. Kompis, A. M. Labhardt, H. Langen, H. Locher, M. G. P. Page, D. Stüber, R. L. Then, B. Wipf, and C. Oefner. A single amino acid substitution in *Staphylococcus aureus* dihydrofolate reductase determines trimethoprim resistance 1 1 edited by t.richmond. *Journal of Molecular Biology*, 266(1):23–30, 1997. ISSN 0022-2836. doi: https://doi.org/10.1006/jmbi.1996.0770. URL http://www.sciencedirect.com/science/article/pii/S0022283696907706.

M. Dalsass, A. Brozzi, D. Medini, and R. Rappuoli. Comparison of open-source reverse vaccinology programs for bacterial vaccine antigen discovery. *Frontiers in Immunology*, 10(113), 2019. ISSN 1664-3224. doi: 10.3389/fimmu.2019.00113. URL https://www.frontiersin.org/article/10.3389/fimmu.2019.00113.

S. E. Damodaran and S. Madhan. Telavancin: A novel lipoglycopeptide antibiotic. *Journal of pharmacology & pharmacotherapeutics*, 2(2):135–137, 2011. ISSN 0976-5018 0976-500X. doi: 10.4103/0976-500X.81918. URL https://pubmed.ncbi.nlm.nih.gov/21772784https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3127350/.

S. F. Darwish and H. A. E. Asfour. Investigation of biofilm forming ability in staphylococci causing bovine mastitis using phenotypic and genotypic assays. *The Scientific World Journal*, 2013:378492, 2013. ISSN 2356-6140. doi: 10.1155/2013/378492. URL https://doi.org/10.1155/2013/378492.

B. D. Davis, L. L. Chen, and P. C. Tai. Misread protein creates membrane channels: an essential step in the bactericidal action of aminoglycosides. *Proceedings of the National Academy of Sciences of the United States of America*, 83(16):6164–6168, 1986. ISSN 0027-8424 1091-6490. doi: 10.1073/pnas.83.16.6164. URL https://pubmed.ncbi.nlm.nih.gov/2426712.

J. A. Davis, S. R. Farrah, and A. C. Wilkie. Selective growth of *Staphylococcus aureus* from flushed dairy manure wastewater using acriflavine-supplemented mannitol salt agar. *Letters in Applied Microbiology*, 0(0):060425081229010, 2006. ISSN 0266-8254. doi: 10.1111/j.1472-765x.2006.01915.x. URL https://dx.doi.org/10.1111/j.1472-765X.2006.01915.x.

J. J. Davis, S. Boisvert, T. Brettin, R. W. Kenyon, C. Mao, R. Olson, R. Overbeek, J. Santerre, M. Shukla, A. R. Wattam, R. Will, F. Xia, and R. Stevens. Antimicrobial resistance prediction in patric and rast. *Scientific reports*, 6: 27930–27930, 2016. ISSN 2045-2322. doi: 10.1038/srep27930. URL https://pubmed.ncbi.nlm.nih.gov/27297683https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4906388/.

P. A. d'Azevedo, I. Siquiera, J. Gugel, A. L. S. Antunes, C. Secchi, J. Pasternak, and M. D. V. Martino. Evaluation of the automated system vitek2 for identification and antimicrobial susceptibility testing of brazilian gram-positive cocci strains. *Brazilian Journal of Infectious Diseases*, 13:107–110, 2009.

N. W. M. De Jong, M. Vrieling, B. L. Garcia, G. Koop, M. Brettmann, P. C. Aerts, M. Ruyken, J. A. G. Van Strijp, M. Holmes, E. M. Harrison, B. V. Geisbrecht, and S. H. M. Rooijakkers. Identification of a staphylococcal complement inhibitor with broad host specificity in equid*Staphylococcus aureus*strains. *Journal of Biological Chemistry*, 293(12):4468–4477, 2018. ISSN 0021-9258. doi: 10.1074/jbc.ra117.000599. URL https://dx.doi.org/10.1074/jbc.RA117.000599.

O. K. Dego, J. E. van Dijk, and H. Nederbragt. Factors involved in the early pathogenesis of bovine *Staphylococcus aureus* mastitis with emphasis on bacterial adhesion and invasion. a review. *Veterinary Quarterly*, 24(4):181–198, 2002. ISSN 0165-2176. doi: 10.1080/01652176.2002.9695135. URL https://doi.org/10.1080/01652176.2002.9695135.

J. R. E. del Castillo. Tetracyclines. *Antimicrobial Therapy in Veterinary Medicine*, pages 257–268, 2013. ISSN 9781118675014. doi: doi:10.1002/9781118675014.ch1510.1002/9781118675014.ch15. URL https://doi.org/10.1002/9781118675014.ch15.

C. Del Fabbro, S. Scalabrin, M. Morgante, and F. M. Giorgi. An extensive evaluation of read trimming effects on illumina ngs data analysis. *PLOS ONE*, 8(12):e85024,

2013. doi: 10.1371/journal.pone.0085024. URL https://doi.org/10.1371/journal.pone.0085024.

S. Delgado, P. García, L. Fernández, E. Jiménez, M. Rodríguez-Baños, R. del Campo, and J. M. Rodríguez. Characterization of *Staphylococcus aureus* strains involved in human and bovine mastitis. *FEMS Immunology & Medical Microbiology*, 62(2): 225–235, 2011. ISSN 0928-8244. doi: 10.1111/j.1574-695X.2011.00806.x. URL https://doi.org/10.1111/j.1574-695X.2011.00806.x.

O. A. Dellagostin, A. A. Grassmann, C. Rizzi, R. A. Schuch, S. Jorge, T. L. Oliveira, A. J. McBride, and D. D. Hartwig. Reverse vaccinology: an approach for identifying leptospiral vaccine candidates. *International Journal of Molecular Sciences*, 18(1):158, 2017.

H. A. Deluyker, S. N. Van Oye, and J. F. Boucher. Factors affecting cure and somatic cell count after pirlimycin treatment of subclinical mastitis in lactating cows. *Journal of Dairy Science*, 88(2):604–14, 2005. ISSN 0022-0302 (Print) 0022-0302. doi: 10.3168/jds.S0022-0302(05)72724-7.

H. Demirci, F. t. Murphy, E. Murphy, S. T. Gregory, A. E. Dahlberg, and G. Jogl. A structural basis for streptomycin-induced misreading of the genetic code. *Nature communications*, 4:1355–1355, 2013. ISSN 2041-1723. doi: 10.1038/ncomms2346. URL https://pubmed.ncbi.nlm.nih.gov/23322043https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3552334/.

J. Deng, X. Wang, B.-Z. Zhang, P. Gao, Q. Lin, R. Y.-T. Kao, K. Gustafsson, K.-Y. Yuen, and J.-D. Huang. Broad and effective protection against *Staphylococcus aureus* is elicited by a multivalent vaccine formulated with novel antigens. *mSphere*, 4(5):e00362–19, 2019. doi: 10.1128/mSphere.00362-19. URL http://msphere.asm.org/content/4/5/e00362-19.abstract.

R. Deora, T. Tseng, and T. K. Misra. Alternative transcription factor sigmasb of *Staphylococcus aureus*: characterization and role in transcription of the global regulatory locus sar. *J Bacteriol*, 179(20):6355–9, 1997. ISSN 0021-9193 (Print) 0021-9193. doi: 10.1128/jb.179.20.6355-6359.1997.

A. Derbise, G. De Cespedes, and N. E. Solh. Nucleotide sequence of the*Staphylococcus aureus* transposon, tn5405, carrying aminoglycosides resistance

genes. *Journal of Basic Microbiology*, 37(5):379–384, 1997. ISSN 0233-111X. doi: 10.1002/jobm.3620370511. URL https://dx.doi.org/10.1002/jobm.3620370511.

S. Deresinski. Vancomycin in combination with other antibiotics for the treatment of serious methicillin-resistant *Staphylococcus aureus* infections. *Clinical Infectious Diseases*, 49(7):1072–1079, 2009. ISSN 1058-4838. doi: 10.1086/605572. URL https://doi.org/10.1086/605572.

L. Devriese, L. Van Damme, and L. Fameree. Methicillin (cloxacillin)-resistant staphylococcus aureus strains isolated from bovine mastitis cases. *Zentralblatt für Veterinärmedizin Reihe B*, 19(7):598–605, 1972.

L. A. Devriese, D. Nzuambe, and C. Godard. Identification and characteristics of staphylococci isolated from lesions and normal skin of horses. *Veterinary Microbiology*, 10(3):269–277, 1985. ISSN 0378-1135. doi: 10.1016/0378-1135(85)90052-5. URL https://dx.doi.org/10.1016/0378-1135(85)90052-5.

N. B. Dhanawade, D. R. Kalorey, R. Srinivasan, S. B. Barbuddhe, and N. V. Kurkure. Detection of intercellular adhesion genes and biofilm production in *Staphylococcus aureus* isolated from bovine subclinical mastitis. *Veterinary Research Communications*, 34(1):81–89, 2010. ISSN 1573-7446. doi: 10.1007/s11259-009-9326-0. URL https://doi.org/10.1007/s11259-009-9326-0.

R. Di Caprio, S. Lembo, L. Di Costanzo, A. Balato, and G. Monfrecola. Anti-inflammatory properties of low and high doxycycline doses: An *in vitro* study. *Mediators of Inflammation*, 2015:1–10, 2015. ISSN 0962-9351. doi: 10.1155/2015/329418. URL https://dx.doi.org/10.1155/2015/329418.

Y. Ding, Y. Onodera, J. C. Lee, and D. C. Hooper. Norb, an efflux pump in *Staphylococcus aureus* strain mw2, contributes to bacterial fitness in abscesses. *Journal of Bacteriology*, 190(21):7123–7129, 2008. doi: 10.1128/jb.00655-08. URL https://jb.asm.org/content/jb/190/21/7123.full.pdf.

Y. Ding, Y. Fu, J. C. Lee, and D. C. Hooper. *Staphylococcus aureus* nord, a putative efflux pump coregulated with the opp1 oligopeptide permease, contributes selectively to fitness *in vivo*. *Journal of bacteriology*, 194(23):6586–6593, 2012.

ISSN 1098-5530 0021-9193. doi: 10.1128/JB.01414-12. URL https://pubmed.ncbi.nlm.nih.gov/23042988https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3497520/.

K. E. Dingle, F. M. Colles, D. R. A. Wareing, R. Ure, A. J. Fox, F. E. Bolton, H. J. Bootsma, R. J. L. Willems, R. Urwin, and M. C. J. Maiden. Multilocus sequence typing system for campylobacter jejuni. *Journal of Clinical Microbiology*, 39(1): 14–23, 2001. doi: 10.1128/jcm.39.1.14-23.2001. URL https://jcm.asm.org/content/jcm/39/1/14.full.pdf.

R. T. Dingwell, K. E. Leslie, P. Sabour, D. Lepp, and J. Pacan. Influence of the genotype of *Staphylococcus aureus*, determined by pulsed-field gel electrophoresis, on dry-period elimination of subclinical mastitis in canadian dairy herds. *Can J Vet Res*, 70(2):115–20, 2006. ISSN 0830-9000 (Print) 0830-9000.

P. Dönnes and O. Kohlbacher. Svmhc: a server for prediction of mhc-binding peptides. *Nucleic acids research*, 34(suppl_2):W194–W197, 2006.

J. Dordel, C. Kim, M. Chung, M. Pardos de la Gándara, M. T. J. Holden, J. Parkhill, H. de Lencastre, S. D. Bentley, and A. Tomasz. Novel determinants of antibiotic resistance: identification of mutated loci in highly methicillin-resistant subpopulations of methicillin-resistant *Staphylococcus aureus*. *mBio*, 5(2): e01000–e01000, 2014. ISSN 2150-7511. doi: 10.1128/mBio.01000-13. URL https://pubmed.ncbi.nlm.nih.gov/24713324https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3993859/.

W. R. Dowdle, E. De Gourville, O. M. Kew, M. A. Pallansch, and D. J. Wood. Polio eradication: the opv paradox. *Rev Med Virol*, 13(5):277–91, 2003. ISSN 1052-9276 (Print) 1052-9276. doi: 10.1002/rmv.401.

I. A. Doytchinova and D. R. Flower. Vaxijen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics*, 8(1):4, 2007. ISSN 1471-2105. doi: 10.1186/1471-2105-8-4. URL https://doi.org/10.1186/1471-2105-8-4.

C. M. Duarte, P. P. Freitas, and R. Bexiga. Technological advances in bovine mastitis diagnosis: an overview. *Journal of Veterinary Diagnostic Investigation*, 27(6): 665–672, 2015. ISSN 1040-6387. doi: 10.1177/1040638715603087. URL https://doi.org/10.1177/1040638715603087.

P. Duckert, S. Brunak, and N. Blom. Prediction of proprotein convertase cleavage sites. *Protein Engineering Design and Selection*, 17(1):107–112, 2004.

S. Dufour, J. Labrie, and M. Jacques. The mastitis pathogens culture collection. *Microbiology resource announcements*, 8(15):e00133–19, 2019. ISSN 2576-098X. doi: 10.1128/MRA.00133-19. URL https://pubmed.ncbi.nlm.nih.gov/30975807https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6460030/.

K. Dziewanowska, J. M. Patti, C. F. Deobald, K. W. Bayles, W. R. Trumble, and G. A. Bohach. Fibronectin binding protein and host cell tyrosine kinase are required for internalization of *Staphylococcus aureus* by epithelial cells. *Infect Immun*, 67(9): 4673–8, 1999. ISSN 0019-9567 (Print) 0019-9567. doi: 10.1128/iai.67.9.4673-4678.1999.

R. Ekblom and J. B. Wolf. A field guide to whole-genome sequencing, assembly and annotation. *Evolutionary applications*, 7(9):1026–1042, 2014.

S. El-Gebali, J. Mistry, A. Bateman, S. R. Eddy, A. Luciani, S. C. Potter, M. Qureshi, L. J. Richardson, G. A. Salazar, A. Smart, et al. The pfam protein families database in 2019. *Nucleic acids research*, 47(D1):D427–D432, 2019.

Y. El-Manzalawy, D. Dobbs, and V. Honavar. Predicting linear b-cell epitopes using string kernels. *Journal of molecular recognition : JMR*, 21(4):243–255, 2008. ISSN 0952-3499 1099-1352. doi: 10.1002/jmr.893. URL https://pubmed.ncbi.nlm.nih.gov/18496882https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2683948/.

Y. EL-Manzalawy, D. Dobbs, and V. Honavar. Predicting linear b-cell epitopes using string kernels. *Journal of Molecular Recognition: An Interdisciplinary Journal*, 21 (4):243–255, 2008.

J. Elmwall, J. Kwiecinski, M. Na, A. A. Ali, V. Osla, L. N. Shaw, W. Wang, K. Sävman, E. Josefsson, J. Bylund, T. Jin, A. Welin, and A. Karlsson. Galectin-3 is a target for proteases involved in the virulence of *Staphylococcus aureus*. *Infect Immun*, 85(7), 2017. ISSN 0019-9567 (Print) 0019-9567. doi: 10.1128/iai.00177-17.

O. Emanuelsson, S. Brunak, G. Von Heijne, and H. Nielsen. Locating proteins in the cell using targetp, signalp and related tools. *Nature protocols*, 2(4):953–971, 2007.

M. C. Enright and B. G. Spratt. A multilocus sequence typing scheme for streptococcus pneumoniae: identification of clones associated with serious invasive disease. *Microbiology*, 144(11):3049–3060, 1998. ISSN 1350-0872. doi: 10.1099/00221287-144-11-3049. URL https://dx.doi.org/10.1099/00221287-144-11-3049.

M. C. Enright and B. G. Spratt. Multilocus sequence typing. *Trends in Microbiology*, 7(12):482–487, 1999. ISSN 0966-842X. doi: 10.1016/s0966-842x(99)01609-1. URL https://dx.doi.org/10.1016/S0966-842X(99)01609-1.

M. C. Enright, N. P. Day, C. E. Davies, S. J. Peacock, and B. G. Spratt. Multilocus sequence typing for characterization of methicillin-resistant and methicillin-susceptible clones of *Staphylococcus aureus*. *Journal of clinical microbiology*, 38(3):1008–1015, 2000. ISSN 0095-1137 1098-660X. URL https://www.ncbi.nlm.nih.gov/pubmed/10698988https://www.ncbi.nlm.nih.gov/pmc/articles/PMC86325/.

M. C. Enright, B. G. Spratt, A. Kalia, J. H. Cross, and D. E. Bessen. Multilocus sequence typing of streptococcus pyogenes and the relationships between emm type and clone. *Infection and Immunity*, 69(4):2416–2427, 2001. ISSN 0019-9567. doi: 10.1128/iai.69.4.2416-2427.2001. URL https://dx.doi.org/10.1128/IAI.69.4.2416-2427.2001.

M. C. Enright, D. A. Robinson, G. Randle, E. J. Feil, H. Grundmann, and B. G. Spratt. The evolutionary history of methicillin-resistant *Staphylococcus aureus* (mrsa). *Proceedings of the National Academy of Sciences*, 99(11):7687–7692, 2002. ISSN 0027-8424. doi: 10.1073/pnas.122108599. URL https://dx.doi.org/10.1073/pnas.122108599.

R. Erskine, 2020. URL https://www.msdvetmanual.com/reproductive-system/mastitis-in-large-animals/mastitis-in-cattle.

S. Y. Essack. The development of β-lactam antibiotics in response to the evolution of β-lactamases. *Pharmaceutical Research*, 18(10):1391–1399, 2001. ISSN 1573-904X. doi: 10.1023/A:1012272403776. URL https://doi.org/10.1023/A:1012272403776.

E. A. Ezekwe, C. Weng, and J. A. Duncan. Adam10 cell surface expression but not activity is critical for *Staphylococcus aureus* α-hemolysin-mediated activation of the nlrp3 inflammasome in human monocytes. *Toxins*, 8(4):95, 2016. ISSN 2072-6651. URL https://www.mdpi.com/2072-6651/8/4/95.

D. Falush and R. Bowden. Genome-wide association mapping in bacteria? *Trends in Microbiology*, 14(8):353–355, 2006. ISSN 0966-842X. doi: https://doi.org/10.1016/j.tim.2006.06.003. URL https://www.sciencedirect.com/science/article/pii/S0966842X0600148X.

R. Fang, J. Cui, T. Cui, H. Guo, H. K. Ono, C.-H. Park, M. Okamura, A. Nakane, and D.-L. Hu. Staphylococcal enterotoxin c is an important virulence factor for mastitis. *Toxins*, 11(3):141, 2019. ISSN 2072-6651. URL https://www.mdpi.com/2072-6651/11/3/141.

E. J. Feil, J. E. Cooper, H. Grundmann, D. A. Robinson, M. C. Enright, T. Berendt, S. J. Peacock, J. M. Smith, M. Murphy, B. G. Spratt, C. E. Moore, and N. P. J. Day. How clonal is *Staphylococcus aureus*? *Journal of Bacteriology*, 185(11): 3307–3316, 2003. ISSN 0021-9193. doi: 10.1128/jb.185.11.3307-3316.2003. URL https://dx.doi.org/10.1128/JB.185.11.3307-3316.2003.

E. J. Feil, B. C. Li, D. M. Aanensen, W. P. Hanage, and B. G. Spratt. eburst: Inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. *Journal of Bacteriology*, 186(5):1518–1530, 2004. ISSN 0021-9193. doi: 10.1128/jb.186.5.1518-1530.2004. URL https://dx.doi.org/10.1128/JB.186.5.1518-1530.2004.

F. Fenner, R. Wittek, and K. Dumbell. The global spread, control, and eradication of smallpox. *The Orthopoxviruses*, pages 317–352, 1989.

C. J. Fernandes, L. A. Fernandes, P. Collignon, and R. on behalf of the Australian Group on Antimicrobial. Cefoxitin resistance as a surrogate marker for the detection of methicillin-resistant *Staphylococcus aureus*. *Journal of Antimicrobial Chemotherapy*, 55(4):506–510, 2005. ISSN 0305-7453. doi: 10.1093/jac/dki052. URL https://doi.org/10.1093/jac/dki052.

R. Fernandez-Munoz, R. E. Monro, D. Vazquez, and R. Torres-Pinedo. Substrate-and antibiotic-binding sites at the peptidyl-transferase centre of *Escherichia coli* ribosomes. studies on the chloramphenicol, lincomycin and erythromycin sites.

European Journal of Biochemistry, 23(1):185–193, 1971. ISSN 0014-2956. doi: 10.1111/j.1432-1033.1971.tb01607.x. URL https://dx.doi.org/10.1111/j.1432-1033.1971.tb01607.xhttps://febs.onlinelibrary.wiley.com/doi/full/10.1111/j.1432-1033.1971.tb01607.x?sid=nlm%3Apubmed.

J. C. Ferreira, M. S. Gomes, E. C. R. Bonsaglia, I. F. Canisso, E. F. Garrett, J. L. Stewart, Z. Zhou, and F. S. Lima. Comparative analysis of four commercial on-farm culture methods to identify bacteria associated with clinical mastitis in dairy cattle. *PLOS ONE*, 13(3):e0194211, 2018. ISSN 1932-6203. doi: 10.1371/journal.pone.0194211. URL https://dx.doi.org/10.1371/journal.pone.0194211.

G. Filioussis, M. Kachrimanidou, G. Christodoulopoulos, M. Kyritsi, C. Hadjichristodoulou, M. Adamopoulou, A. Tzivara, S. K. Kritas, and A. Grinberg. Bovine mastitis caused by a multidrug-resistant, mcr-1-positive (colistin-resistant), extended-spectrum β-lactamase–producing escherichia coli clone on a greek dairy farm. *Journal of dairy science*, 103(1):852–857, 2020.

J. E. Finan, A. E. Rosato, T. M. Dickinson, D. Ko, and G. L. Archer. Conversion of oxacillin-resistant staphylococci from heterotypic to homotypic resistance expression. *Antimicrobial agents and chemotherapy*, 46(1):24–30, 2002. ISSN 0066-4804 1098-6596. doi: 10.1128/aac.46.1.24-30.2002. URL https://pubmed.ncbi.nlm.nih.gov/11751106https://www.ncbi.nlm.nih.gov/pmc/articles/PMC126971/.

B. B. Finlay and S. Falkow. Common themes in microbial pathogenicity revisited. *Microbiology and molecular biology reviews : MMBR*, 61(2):136–169, 1997. ISSN 1092-2172 1098-5557. URL https://pubmed.ncbi.nlm.nih.gov/9184008https://www.ncbi.nlm.nih.gov/pmc/articles/PMC232605/.

R. D. Finn, J. Clements, and S. R. Eddy. Hmmer web server: interactive sequence similarity searching. *Nucleic Acids Research*, 39(suppl):W29–W37, 2011. ISSN 0305-1048. doi: 10.1093/nar/gkr367. URL https://dx.doi.org/10.1093/nar/gkr367.

J. R. Fitzgerald. Livestock-associated *Staphylococcus aureus*: origin, evolution and public health threat. *Trends in Microbiology*, 20(4):192–198, 2012. ISSN

0966-842X. doi: https://doi.org/10.1016/j.tim.2012.01.006. URL http://www.sciencedirect.com/science/article/pii/S0966842X12000170.

J. R. Fitzgerald, S. R. Monday, T. J. Foster, G. A. Bohach, P. J. Hartigan, W. J. Meaney, and C. J. Smyth. Characterization of a putative pathogenicity island from bovine *Staphylococcus aureus* encoding multiple superantigens. *Journal of Bacteriology*, 183(1):63–70, 2001. doi: 10.1128/jb.183.1.63-70.2001. URL https://jb.asm.org/content/jb/183/1/63.full.pdf.

J.-P. Flandrois, G. Lina, and O. Dumitrescu. Mubii-tb-db: a database of mutations associated with antibiotic resistance in mycobacterium tuberculosis. *BMC bioinformatics*, 15(1):1–9, 2014.

R. Fleischmann, M. Adams, O. White, R. Clayton, E. Kirkness, A. Kerlavage, C. Bult, J. Tomb, B. Dougherty, J. Merrick, and E. Al. Whole-genome random sequencing and assembly of haemophilus influenzae rd. *Science*, 269(5223): 496–512, 1995. ISSN 0036-8075. doi: 10.1126/science.7542800. URL https://dx.doi.org/10.1126/science.7542800.

A. C. Fluit, M. R. Visser, and F.-J. Schmitz. Molecular detection of antimicrobial resistance. *Clinical microbiology reviews*, 14(4):836–871, 2001.

B. A. Fonner, B. P. Tripet, B. J. Eilers, J. Stanisich, R. K. Sullivan-Springhetti, R. Moore, M. Liu, B. Lei, and V. Copié. Solution structure and molecular determinants of hemoglobin binding of the first neat domain of isdb in *Staphylococcus aureus*. *Biochemistry*, 53(24):3922–33, 2014. ISSN 0006-2960 (Print) 0006-2960. doi: 10.1021/bi5005188.

T. J. Foster. Antibiotic resistance in *Staphylococcus aureus*. current status and future prospects. *FEMS Microbiology Reviews*, 41(3):430–449, 2017. ISSN 0168-6445. doi: 10.1093/femsre/fux007. URL https://doi.org/10.1093/femsre/fux007.

T. J. Foster, J. A. Geoghegan, V. K. Ganesh, and M. Höök. Adhesion, invasion and evasion: the many functions of the surface proteins of *Staphylococcus aureus*. *Nature Reviews Microbiology*, 12(1):49–62, 2014. ISSN 1740-1526. doi: 10.1038/nrmicro3161. URL https://dx.doi.org/10.1038/nrmicro3161.

B. Fournier, X. Zhao, T. Lu, K. Drlica, and D. C. Hooper. Selective targeting of topoisomerase iv and dna gyrase in *Staphylococcus aureus*: Different patterns of

quinolone- induced inhibition of dna synthesis. *Antimicrobial Agents and Chemotherapy*, 44(8):2160–2165, 2000. doi: 10.1128/aac.44.8.2160-2165.2000. URL https://aac.asm.org/content/aac/44/8/2160.full.pdf.

C. Fournier, P. Kuhnert, J. Frey, R. Miserez, M. Kirchhofer, T. Kaufmann, A. Steiner, and H. U. Graber. Bovine *Staphylococcus aureus*: Association of virulence genes, genotypes and clinical outcome. *Research in Veterinary Science*, 85(3):439–448, 2008. ISSN 0034-5288. doi: https://doi.org/10.1016/j.rvsc.2008.01.010. URL http://www.sciencedirect.com/science/article/pii/S0034528808000106.

L. K. Fox and D. S. Adams. The ability of the enzyme-linked immunosorbent assay to detect antibody against *Staphylococcus aureus* in milk following experimental intramammary infection. *Journal of Veterinary Medicine, Series B*, 47(7):517–526, 2000. ISSN 0931-1793. doi: https://doi.org/10.1046/j.1439-0450.2000.00379.x. URL https://doi.org/10.1046/j.1439-0450.2000.00379.x.

L. K. Fox, R. N. Zadoks, and C. T. Gaskins. Biofilm production by *Staphylococcus aureus* associated with intramammary infection. *Vet Microbiology*, 107(3-4):295–9, 2005. ISSN 0378-1135 (Print) 0378-1135. doi: 10.1016/j.vetmic.2005.02.005.

A. P. Francisco, M. Bugalho, M. Ramirez, and J. A. Carriço. Global optimal eburst analysis of multilocus typing data using a graphic matroid approach. *BMC bioinformatics*, 10(1):1–15, 2009.

K. T. Franck, H. Gumpert, B. Olesen, A. R. Larsen, A. Petersen, J. Bangsborg, P. Albertsen, H. Westh, and M. D. Bartels. Staphylococcal aureus enterotoxin c and enterotoxin-like l associated with post-partum mastitis. *Frontiers in Microbiology*, 8(173), 2017. ISSN 1664-302X. doi: 10.3389/fmicb.2017.00173. URL https://www.frontiersin.org/article/10.3389/fmicb.2017.00173.

H. Frase, M. Toth, and S. B. Vakulenko. Revisiting the nucleotide and aminoglycoside substrate specificity of the bifunctional aminoglycoside acetyltransferase(6′)-ie/aminoglycoside phosphotransferase(2″)-ia enzyme. *Journal of Biological Chemistry*, 287(52):43262–43269, 2012. ISSN 0021-9258. doi: 10.1074/jbc.m112.416453. URL https://dx.doi.org/10.1074/jbc.M112.416453.

M. Freick, Y. Frank, K. Steinert, A. Hamedy, O. Passarge, and A. Sobiraj. Mastitis vaccination using a commercial polyvalent vaccine or a herd-specific *Staphylococcus aureus* vaccine. *Tierärztliche Praxis Ausgabe G: Großtiere/Nutztiere*, 44(04):219–229, 2016.

H. M. Frénay, A. E. Bunschoten, L. M. Schouls, W. J. van Leeuwen, C. M. Vandenbroucke-Grauls, J. Verhoef, and F. R. Mooi. Molecular typing of methicillin-resistant *Staphylococcus aureus* on the basis of protein a gene polymorphism. *Eur J Clin Microbiol Infect Dis*, 15(1):60–4, 1996. ISSN 0934-9723 (Print) 0934-9723. doi: 10.1007/bf01586186.

S. Fuchs, H. Mehlan, J. Bernhardt, A. Hennig, S. Michalik, K. Surmann, J. Pané-Farré, A. Giese, S. Weiss, L. Backert, A. Herbig, K. Nieselt, M. Hecker, U. Völker, and U. Mäder. Aureowiki - the repository of the *Staphylococcus aureus* research and annotation community. *International Journal of Medical Microbiology*, 308(6):558–568, 2018. ISSN 1438-4221. doi: https://doi.org/10.1016/j.ijmm.2017.11.011. URL https://www.sciencedirect.com/science/article/pii/S1438422117304629. Infections by Gram-positive pathobionts *Staphylococcus aureus* and Streptococcus pneumoniae – from colonization to invasive infections.

J. M. Fueyo, M. C. Mendoza, M. R. Rodicio, J. Muñiz, M. A. Alvarez, and M. C. Martín. Cytotoxin and pyrogenic toxin superantigen gene profiles of *Staphylococcus aureus* associated with subclinical mastitis in dairy cows and relationships with macrorestriction genomic profiles. *Journal of clinical microbiology*, 43(3):1278–1284, 2005. ISSN 0095-1137 1098-660X. doi: 10.1128/JCM.43.3.1278-1284.2005. URL https://pubmed.ncbi.nlm.nih.gov/15750096https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1081256/.

K. Fursova, A. Sorokin, S. Sokolov, T. Dzhelyadin, I. Shulcheva, M. Shchannikova, D. Nikanova, O. Artem'eva, N. Zinovieva, and F. Brovko. Virulence factors and phylogeny of *Staphylococcus aureus* associated with bovine mastitis in russia based on genome sequences. *Frontiers in Veterinary Science*, 7(135), 2020. ISSN 2297-1769. doi: 10.3389/fvets.2020.00135. URL https://www.frontiersin.org/article/10.3389/fvets.2020.00135.

G. F. Gad, M. A. El-Feky, M. S. El-Rehewy, M. A. Hassan, H. Abolella, and R. M.

El-Baky. Detection of icaa, icad genes and biofilm production by *Staphylococcus aureus* and staphylococcus epidermidis isolated from urinary tract catheterized patients. *J Infect Dev Ctries*, 3(5):342–51, 2009. ISSN 1972-2680. doi: 10.3855/jidc.241.

E. K. Ganda, R. S. Bisinotto, D. H. Decter, and R. C. Bicalho. Evaluation of an on-farm culture system (accumast) for fast identification of milk pathogens associated with clinical mastitis in dairy cows. *PLoS One*, 11(5):e0155314, 2016. ISSN 1932-6203. doi: 10.1371/journal.pone.0155314.

L. Garcia-Alvarez, M. T. G. Holden, H. Lindsay, C. R. Webb, D. F. J. Brown, M. D. Curran, E. Walpole, K. Brooks, D. J. Pickard, C. Teale, J. Parkhill, S. D. Bentley, G. F. Edwards, E. K. Girvan, A. M. Kearns, B. Pichon, R. L. R. Hill, A. R. Larsen, R. L. Skov, S. J. Peacock, D. J. Maskell, and M. A. Holmes. Meticillin-resistant *Staphylococcus aureus* with a novel meca homologue in human and bovine populations in the uk and denmark: a descriptive study. *The Lancet. Infectious diseases*, 11(8):595–603, 2011. ISSN 1474-4457 1473-3099. doi: 10.1016/S1473-3099(11)70126-8. URL https://pubmed.ncbi.nlm.nih.gov/21641281https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3829197/.

J. L. Gardy, M. R. Laird, F. Chen, S. Rey, C. Walsh, M. Ester, and F. S. Brinkman. Psortb v. 2.0: expanded prediction of bacterial protein subcellular localization and insights gained from comparative proteome analysis. *Bioinformatics*, 21(5): 617–623, 2005. ISSN 1460-2059.

E. Gasteiger, C. Hoogland, A. Gattiker, M. R. Wilkins, R. D. Appel, A. Bairoch, et al. Protein identification and analysis tools on the expasy server. *The proteomics protocols handbook*, pages 571–607, 2005.

R. Gaynes. The discovery of penicillin—new insights after more than 75 years of clinical use. *Emerging Infectious Diseases*, 23(5):849–853, 2017. ISSN 1080-6040 1080-6059. doi: 10.3201/eid2305.161556. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5403050/.

N. H. Georgopapadakou and F. Y. Liu. Penicillin-binding proteins in bacteria. *Antimicrobial Agents and Chemotherapy*, 18(1):148–157, 1980. doi: 10.1128/aac.18.1.148. URL https://aac.asm.org/content/aac/18/1/148.full.pdf.

N. H. Georgopapadakou, B. A. Dix, and Y. R. Mauriz. Possible physiological functions of penicillin-binding proteins in *Staphylococcus aureus*. *Antimicrobial Agents and Chemotherapy*, 29(2):333–336, 1986. doi: 10.1128/aac.29.2.333. URL https://aac.asm.org/content/aac/29/2/333.full.pdf.

V. Gerke and S. E. Moss. Annexins: From structure to function. *Physiological Reviews*, 82(2):331–371, 2002. ISSN 0031-9333. doi: 10.1152/physrev.00030.2001. URL https://doi.org/10.1152/physrev.00030.2001.

M. M. Giuliani, J. Adu-Bobie, M. Comanducci, B. Aricò, S. Savino, L. Santini, B. Brunelli, S. Bambini, A. Biolchi, B. Capecchi, et al. A universal vaccine for serogroup b meningococcus. *Proceedings of the National Academy of Sciences*, 103(29):10834–10839, 2006.

R. Gleckman, N. Blagg, and D. W. Joubert. Trimethoprim: mechanisms of action, antimicrobial activity, bacterial resistance, pharmacokinetics, adverse reactions, and therapeutic indications. *Pharmacotherapy*, 1(1):14–20, 1981. ISSN 0277-0008 (Print) 0277-0008. doi: 10.1002/j.1875-9114.1981.tb03548.x.

A. Glenny and B. E. Hopkins. Diphtheria toxoid as an immunising agent. *British journal of experimental pathology*, 4(5):283, 1923.

S. Gnerre, I. Maccallum, D. Przybylski, F. J. Ribeiro, J. N. Burton, B. J. Walker, T. Sharpe, G. Hall, T. P. Shea, S. Sykes, A. M. Berlin, D. Aird, M. Costello, R. Daza, L. Williams, R. Nicol, A. Gnirke, C. Nusbaum, E. S. Lander, and D. B. Jaffe. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proceedings of the National Academy of Sciences*, 108(4): 1513–1518, 2011. ISSN 0027-8424. doi: 10.1073/pnas.1017351108. URL https://dx.doi.org/10.1073/pnas.1017351108.

T. Goerge, M. B. Lorenz, S. van Alen, N. O. Hübner, K. Becker, and R. Köck. Mrsa colonization and infection among persons with occupational livestock exposure in europe: Prevalence, preventive options and evidence. *Vet Microbiology*, 200:6–12, 2017. ISSN 0378-1135. doi: 10.1016/j.vetmic.2015.10.027.

P. R. Gonzales, M. W. Pesesky, R. Bouley, A. Ballard, B. A. Biddy, M. A. Suckow, W. R. Wolter, V. A. Schroeder, C.-A. D. Burnham, S. Mobashery, M. Chang, and G. Dantas. Synergistic, collaterally sensitive β-lactam combinations suppress

resistance in mrsa. *Nature Chemical Biology*, 11(11):855–861, 2015. ISSN 1552-4450. doi: 10.1038/nchembio.1911. URL https://dx.doi.org/10.1038/nchembio.1911.

B. Gonzalez-Zorn and P. Courvalin. vana-mediated high level glycopeptide resistance in mrsa. *The Lancet Infectious Diseases*, 3(2):67–68, 2003. ISSN 1473-3099. doi: 10.1016/s1473-3099(03)00510-3. URL https://dx.doi.org/10.1016/s1473-3099(03)00510-3.

S. J. Goodswen, P. J. Kennedy, and J. T. Ellis. Vacceed: a high-throughput in silico vaccine candidate discovery pipeline for eukaryotic pathogens based on reverse vaccinology. *Bioinformatics*, 30(16):2381–2383, 2014. ISSN 1367-4803. doi: 10.1093/bioinformatics/btu300. URL https://doi.org/10.1093/bioinformatics/btu300.

N. C. Gordon, J. R. Price, K. Cole, R. Everitt, M. Morgan, J. Finney, A. M. Kearns, B. Pichon, B. Young, D. J. Wilson, M. J. Llewelyn, J. Paul, T. E. A. Peto, D. W. Crook, A. S. Walker, and T. Golubchik. Prediction of *Staphylococcus aureus* antimicrobial resistance by whole-genome sequencing. *Journal of clinical microbiology*, 52(4):1182–1191, 2014. ISSN 1098-660X 0095-1137. doi: 10.1128/JCM.03117-13. URL https://pubmed.ncbi.nlm.nih.gov/24501024https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3993491/.

E. C. Gotschlich, T. Y. Liu, and M. S. Artenstein. Human immunity to the meningococcus : Iii. preparation and immunochemical properties of the group a, group b, and group c meningococcal polysaccharides. *Journal of Experimental Medicine*, 129(6):1349–1365, 1969. ISSN 0022-1007. doi: 10.1084/jem.129.6.1349. URL https://doi.org/10.1084/jem.129.6.1349.

H. Graveland, J. A. Wagenaar, H. Heesterbeek, D. Mevius, E. van Duijkeren, and D. Heederik. Methicillin resistant *Staphylococcus aureus* st398 in veal calf farming: Human mrsa carriage related with animal antimicrobial usage and farm hygiene. *PLOS ONE*, 5(6):e10990, 2010. doi: 10.1371/journal.pone.0010990. URL https://doi.org/10.1371/journal.pone.0010990.

M. J. Green, L. E. Green, G. F. Medley, Y. H. Schukken, and A. J. Bradley. Influence of dry period bacterial intramammary infection on clinical mastitis in dairy cows.

Journal of Dairy Science, 85(10):2589–99, 2002. ISSN 0022-0302 (Print) 0022-0302. doi: 10.3168/jds.S0022-0302(02)74343-9.

H. D. Gresham, J. H. Lowrance, T. E. Caver, B. S. Wilson, A. L. Cheung, and F. P. Lindberg. Survival of*Staphylococcus aureus* inside neutrophils contributes to infection. *The Journal of Immunology*, 164(7):3713, 2000. doi: 10.4049/jimmunol.164.7.3713. URL http://www.jimmunol.org/content/164/7/3713.abstract.

A. Grinberg, N. Lopez-Villalobos, K. Lawrence, and M. Nulsen. Prediction of penicillin resistance in *Staphylococcus aureus* isolates from dairy cows with mastitis, based on prior test results. *New Zealand Veterinary Journal*, 53(5): 332–335, 2005. ISSN 0048-0169. doi: 10.1080/00480169.2005.36569. URL https://doi.org/10.1080/00480169.2005.36569.

A. Grinberg, D. D. Kingsbury, I. R. Gibson, B. M. Kirby, H. J. Mack, and D. Morrison. Clinically overt infections with methicillin-resistant *Staphylococcus aureus* in animals in new zealand: A pilot study. *New Zealand Veterinary Journal*, 56(5):237–242, 2008. ISSN 0048-0169. doi: 10.1080/00480169.2008.36840. URL https://doi.org/10.1080/00480169.2008.36840.

A. Grinberg, P. J. Biggs, J. Zhang, S. Ritchie, Z. Oneroa, C. O'Neill, A. Karkaba, N. S. Velathanthiri, and G. W. Coombs. Genomic epidemiology of methicillin-susceptible *Staphylococcus aureus* across colonisation and skin and soft tissue infection. *J Infect*, 75(4):326–335, 2017. ISSN 1532-2742 (Electronic) 0163-4453 (Linking). doi: 10.1016/j.jinf.2017.07.010. URL https://www.ncbi.nlm.nih.gov/pubmed/28782565.

P. A. Gross and F. A. Ennis. Influenza vaccine: split-product versus whole-virus types—how do they differ?, 1977.

T. Grunert, B. Stessl, F. Wolf, D. O. Sordelli, F. R. Buzzola, and M. Ehling-Schulz. Distinct phenotypic traits of *Staphylococcus aureus* are associated with persistent, contagious bovine intramammary infections. *Scientific reports*, 8(1):15968–15968, 2018. ISSN 2045-2322. doi: 10.1038/s41598-018-34371-1. URL https://pubmed.ncbi.nlm.nih.gov/30374136https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6206001/.

G. G. Guay and D. M. Rothstein. Expression of the tetk gene from *Staphylococcus aureus* in *Escherichia coli*: comparison of substrate specificities of teta(b), teta(c), and tetk efflux proteins. *Antimicrobial agents and chemotherapy*, 37(2):191–198, 1993. ISSN 0066-4804 1098-6596. doi: 10.1128/aac.37.2.191. URL https://pubmed.ncbi.nlm.nih.gov/8452348https://www.ncbi.nlm.nih.gov/pmc/articles/PMC187637/.

J. Guccione, A. Pesce, M. Pascale, C. Salzano, G. Tedeschi, L. D'Andrea, A. De Rosa, and P. Ciaramella. Efficacy of a polyvalent mastitis vaccine against *Staphylococcus aureus* on a dairy mediterranean buffalo farm: results of two clinical field trials. *BMC veterinary research*, 13(1):29–29, 2017. ISSN 1746-6148. doi: 10.1186/s12917-017-0944-4. URL https://pubmed.ncbi.nlm.nih.gov/28103866https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5248485/.

A. Guidry, A. Fattom, A. Patel, C. O'Brien, S. Shepherd, and J. Lohuis. Serotyping scheme for *Staphylococcus aureus* isolated from cows with mastitis. *American journal of veterinary research*, 59(12):1537–1539, 1998. ISSN 0002-9645. URL http://europepmc.org/abstract/MED/9858402.

C. M. Guinane, N. L. Ben Zakour, M. A. Tormo-Mas, L. A. Weinert, B. V. Lowder, R. A. Cartwright, D. S. Smyth, C. J. Smyth, J. A. Lindsay, K. A. Gould, A. Witney, J. Hinds, J. P. Bollback, A. Rambaut, J. R. Penadés, and J. R. Fitzgerald. Evolutionary genomics of *Staphylococcus aureus* reveals insights into the origin and molecular basis of ruminant host adaptation. *Genome biology and evolution*, 2: 454–466, 2010. ISSN 1759-6653. doi: 10.1093/gbe/evq031. URL https://pubmed.ncbi.nlm.nih.gov/20624747https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2997551/.

C. Guldimann, K. J. Boor, M. Wiedmann, and V. Guariglia-Oropeza. Resilience in the face of uncertainty: Sigma factor b fine-tunes gene expression to support homeostasis in gram-positive bacteria. *Applied and Environmental Microbiology*, 82(15):4456–4469, 2016. doi: 10.1128/aem.00714-16. URL https://aem.asm.org/content/aem/82/15/4456.full.pdf.

R. Gupta, E. Jung, A. A. Gooley, K. L. Williams, S. Brunak, and J. Hansen. Scanning the available dictyostelium discoideum proteome for o-linked glcnac glycosylation sites using neural networks. *Glycobiology*, 9(10):1009–1022, 1999.

R. Gupta, E. Jung, and S. Brunak. Netnglyc 1.0 server. *Center for biological sequence analysis, technical university of Denmark available from: http://www. cbs. dtu dk/services/NetNGlyc*, 2004.

S. K. Gupta, B. R. Padmanabhan, S. M. Diene, R. Lopez-Rojas, M. Kempf, L. Landraud, and J.-M. Rolain. Arg-annot, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. *Antimicrobial agents and chemotherapy*, 58(1):212–220, 2014.

A. Gurevich, V. Saveliev, N. Vyahhi, and G. Tesler. Quast: quality assessment tool for genome assemblies. *Bioinformatics*, 29(8):1072–1075, 2013. ISSN 1460-2059. doi: 10.1093/bioinformatics/btt086. URL https://dx.doi.org/10.1093/bioinformatics/btt086.

A. F. Haag, J. R. Fitzgerald, and J. R. Penadés. *Staphylococcus aureus* in animals. *Microbiol Spectr*, 7(3), 2019. ISSN 2165-0497. doi: 10.1128/microbiolspec.GPP3-0060-2019.

C. J. Hackbarth and H. F. Chambers. blai and blar1 regulate beta-lactamase and pbp 2a production in methicillin-resistant *Staphylococcus aureus*. *Antimicrobial Agents and Chemotherapy*, 37(5):1144–1149, 1993. ISSN 0066-4804. doi: 10.1128/aac.37.5.1144. URL https://dx.doi.org/10.1128/aac.37.5.1144.

J. Hacker and J. B. Kaper. Pathogenicity islands and the evolution of microbes. *Annual Review of Microbiology*, 54(1):641–679, 2000. ISSN 0066-4227. doi: 10.1146/annurev.micro.54.1.641. URL https://dx.doi.org/10.1146/annurev.micro.54.1.641.

S. Haghighat, S. D. Siadat, S. M. Sorkhabadi, A. A. Sepahi, and M. Mahdavi. Cloning, expression and purification of autolysin from methicillin-resistant *Staphylococcus aureus*: potency and challenge study in balb/c mice. *Mol Immunol*, 82:10–18, 2017. ISSN 0161-5890. doi: 10.1016/j.molimm.2016.12.013.

M. Haghkhah and Z. Lotfi. Study on the frequency of spa gene in *Staphylococcus aureus* isolates from human infections and its relationship with meca gene. *International Journal of Infectious Diseases*, 45:137, 2016. ISSN 1201-9712. doi: 10.1016/j.ijid.2016.02.336. URL https://dx.doi.org/10.1016/j.ijid.2016.02.336.

L. Hall-Stoodley, J. W. Costerton, and P. Stoodley. Bacterial biofilms: from the natural environment to infectious diseases. *Nat Rev Microbiol*, 2(2):95–108, 2004. ISSN 1740-1526 (Print) 1740-1526. doi: 10.1038/nrmicro821.

O. Hammer, D. A. Harper, and P. D. Ryan. Past: Paleontological statistics software package for education and data analysis. *Palaeontologia electronica*, 4(1):9, 2001.

I. C. Hampele, A. D'Arcy, G. E. Dale, D. Kostrewa, J. Nielsen, C. Oefner, M. G. P. Page, H.-J. Schönfeld, D. Stüber, and R. L. Then. Structure and function of the dihydropteroate synthase from *Staphylococcus aureus*11edited by r. huber. *Journal of Molecular Biology*, 268(1):21–30, 1997. ISSN 0022-2836. doi: https://doi.org/10.1006/jmbi.1997.0944. URL http://www.sciencedirect.com/science/article/pii/S002228369790944X.

W. P. Hanage, C. Fraser, and B. G. Spratt. The impact of homologous recombination on the generation of diversity in bacteria. *Journal of Theoretical Biology*, 239(2): 210–219, 2006. ISSN 0022-5193. doi: 10.1016/j.jtbi.2005.08.035. URL https://dx.doi.org/10.1016/j.jtbi.2005.08.035.

H. Hanaki, K. Kuwahara-Arai, S. Boyle-Vavra, R. S. Daum, H. Labischinski, and K. Hiramatsu. Activated cell-wall synthesis is associated with vancomycin resistance in methicillin-resistant *Staphylococcus aureus* clinical strains mu3 and mu50. *J Antimicrob Chemother*, 42(2):199–209, 1998a. ISSN 0305-7453 (Print) 0305-7453. doi: 10.1093/jac/42.2.199.

H. Hanaki, H. Labischinski, Y. Inaba, N. Kondo, H. Murakami, and K. Hiramatsu. Increase in glutamine-non-amidated muropeptides in the peptidoglycan of vancomycin-resistant *Staphylococcus aureus* strain mu50. *J Antimicrob Chemother*, 42(3):315–20, 1998b. ISSN 0305-7453 (Print) 0305-7453. doi: 10.1093/jac/42.3.315.

B. J. Hartman and A. Tomasz. Low-affinity penicillin-binding protein associated with beta-lactam resistance in *Staphylococcus aureus*. *Journal of bacteriology*, 158(2): 513–516, 1984. ISSN 0021-9193 1098-5530. URL https://www.ncbi.nlm.nih.gov/pubmed/6563036https://www.ncbi.nlm.nih.gov/pmc/articles/PMC215458/.

H. Hasman, A. Moodley, L. Guardabassi, M. Stegger, R. Skov, and F. M. Aarestrup. Spa type distribution in *Staphylococcus aureus* originating from pigs, cattle and

poultry. *Veterinary microbiology*, 141:326–31, 2009. doi: 10.1016/j.vetmic.2009.09.025.

S. Hathroubi, M. A. Mekni, P. Domenico, D. Nguyen, and M. Jacques. Biofilms: microbial shelters against antibiotics. *Microbial Drug Resistance*, 23(2):147–156, 2017. ISSN 1076-6294.

T. Hauschild, A. T. Feßler, K. Kadlec, C. Billerbeck, and S. Schwarz. Detection of the novel vga(e) gene in methicillin-resistant *Staphylococcus aureus* cc398 isolates from cattle and poultry. *Journal of Antimicrobial Chemotherapy*, 67(2):503–504, 2011. ISSN 0305-7453. doi: 10.1093/jac/dkr446. URL https://doi.org/10.1093/jac/dkr446.

M. Haveri, A. Roslöf, L. Rantala, and S. Pyörälä. Virulence genes of bovine *Staphylococcus aureus* from persistent and nonpersistent intramammary infections with different clinical characteristics. *J Appl Microbiol*, 103(4):993–1000, 2007. ISSN 1364-5072 (Print) 1364-5072. doi: 10.1111/j.1365-2672.2007.03356.x.

Y. He, Z. Xiang, and H. L. Mobley. Vaxign: the first web-based vaccine design program for reverse vaccinology and applications for vaccine development. *J Biomed Biotechnol*, 2010:297505, 2010. ISSN 1110-7243 (Print) 1110-7243. doi: 10.1155/2010/297505.

A. Hébert, K. Sayasith, S. Sénéchal, P. Dubreuil, and J. Lagacé. Demonstration of intracellular *Staphylococcus aureus* in bovine mastitis alveolar cells and macrophages isolated from naturally infected cow milk. *FEMS Microbiology Letters*, 193(1):57–62, 2000. ISSN 0378-1097. doi: 10.1111/j.1574-6968.2000.tb09402.x. URL https://doi.org/10.1111/j.1574-6968.2000.tb09402.x.

H. Heffernan and S. Bakker. 2017 survey of methicillin-resistant *Staphylococcus aureus* (mrsa). *Porirua: Nosocomial Infections Laboratory, Institute of Environmental Science and Research Ltd*, 2018.

H. Heffernan, S. Bakker, R. Woodhouse, K. Dyet, and D. Williamson. Demographics, antimicrobial susceptibility and molecular epidemiology of *Staphylococcus aureus* in new zealand, 2014. *Institute of Environmental Science and Research Limited, Wellington, New Zealand*, 2015.

H. Heffernan, S. Bakker, K. Dyet, and D. Williamson. Annual survey of methicillin-resistant *Staphylococcus aureus* (mrsa), 2015. *Porirua: Nosocomial Infections Laboratory, Institute of Environmental Science and Research Ltd*, 2016.

R. Heidemann Olsen, H. Christensen, S. Kabell, and M. Bisgaard. Characterization of prevalent bacterial pathogens associated with pododermatitis in table egg layers. *Avian Pathology*, 47(3):281–285, 2018. ISSN 0307-9457. doi: 10.1080/03079457.2018.1440066. URL https://dx.doi.org/10.1080/03079457.2018.1440066.

S. M. Hensen, M. J. Pavicić, J. A. Lohuis, and B. Poutrel. Use of bovine primary mammary epithelial cells for the comparison of adherence and invasion ability of *Staphylococcus aureus* strains. *Journal of Dairy Science*, 83(3):418–29, 2000. ISSN 0022-0302 (Print) 0022-0302. doi: 10.3168/jds.S0022-0302(00)74898-3.

D. Hernandez, P. Francois, L. Farinelli, M. Osteras, and J. Schrenzel. De novo bacterial genome sequencing: Millions of very short reads assembled on a desktop computer. *Genome Research*, 18(5):802–809, 2008. ISSN 1088-9051. doi: 10.1101/gr.072033.107. URL https://dx.doi.org/10.1101/gr.072033.107.

D. Hernandez, R. Tewhey, J. B. Veyrieras, L. Farinelli, M. Osteras, P. Francois, and J. Schrenzel. De novo finished 2.8 mbp *Staphylococcus aureus* genome assembly from 100 bp short and long range paired-end reads. *Bioinformatics*, 30(1):40–49, 2014. ISSN 1367-4803. doi: 10.1093/bioinformatics/btt590. URL https://dx.doi.org/10.1093/bioinformatics/btt590.

M. Hilleman, W. McAleer, E. Buynak, and A. McLean. The preparation and safety of hepatitis b vaccine. *Journal of Infection*, 7:3–8, 1983.

J. Hillerton, M. Bryan, B. Beattie, D. Scott, A. Millar, and N. French. Use of antimicrobials for food animals in new zealand: updated estimates to identify a baseline to measure targeted reductions. *New Zealand Veterinary Journal*, 69(3): 180–185, 2021.

K. Hiramatsu. Methicillin-resistant *Staphylococcus aureus* clinical strain with reduced vancomycin susceptibility. *Journal of Antimicrobial Chemotherapy*, 40(1): 135–136, 1997. ISSN 1460-2091. doi: 10.1093/jac/40.1.135. URL https://dx.doi.org/10.1093/jac/40.1.135.

K. Hiramatsu. Vancomycin-resistant *Staphylococcus aureus* : a new model of antibiotic resistance. *The Lancet Infectious Diseases*, 1(3):147–155, 2001. ISSN 1473-3099. doi: 10.1016/s1473-3099(01)00091-3. URL https://dx.doi.org/10.1016/s1473-3099(01)00091-3.

K. Hiramatsu, K. Asada, E. Suzuki, K. Okonogi, and T. Yokota. Molecular cloning and nucleotide sequence determination of the regulator region of meca gene in methicillin-resistant *Staphylococcus aureus* (mrsa). *FEBS Letters*, 298(2-3): 133–136, 1992. ISSN 0014-5793. doi: 10.1016/0014-5793(92)80039-j. URL https://febs.onlinelibrary.wiley.com/doi/abs/10.1016/0014-5793%2892%2980039-J.

G. H. Hitchings and J. J. Burchall. Inhibition of folate biosynthesis and function as a basis for chemotherapy. *Adv Enzymol Relat Areas Mol Biol*, 27:417–68, 1965. ISSN 0065-258X (Print) 0065-258x. doi: 10.1002/9780470122723.ch9.

J. Hoekstra, A. L. Zomer, V. P. M. G. Rutten, L. Benedictus, A. Stegeman, M. P. Spaninks, T. W. Bennedsgaard, A. Biggs, S. De Vliegher, D. H. Mateo, R. Huber-Schlenstedt, J. Katholm, P. Kovács, V. Krömker, G. Lequeux, P. Moroni, L. Pinho, S. Smulski, K. Supré, J. M. Swinkels, M. A. Holmes, T. J. G. M. Lam, and G. Koop. Genomic analysis of european bovine *Staphylococcus aureus* from clinical versus subclinical mastitis. *Scientific Reports*, 10(1):18172, 2020. ISSN 2045-2322. doi: 10.1038/s41598-020-75179-2. URL https://doi.org/10.1038/s41598-020-75179-2.

H. Hogeveen, K. Huijps, and T. Lam. Economic aspects of mastitis: New developments. *New Zealand Veterinary Journal*, 59(1):16–23, 2011. ISSN 0048-0169. doi: 10.1080/00480169.2011.547165. URL https://dx.doi.org/10.1080/00480169.2011.547165.

H. Hogeveen, W. Steeneveld, and C. A. Wolf. Production diseases reduce the efficiency of dairy production: A review of the results, methods, and approaches regarding the economics of mastitis. *Annual Review of Resource Economics*, 11(1): 289–312, 2019. ISSN 1941-1340. doi: 10.1146/annurev-resource-100518-093954. URL https://doi.org/10.1146/annurev-resource-100518-093954.

M. Hossain, S. Paul, M. Hossain, M. Islam, and M. Alam. Bovine mastitis and its therapeutic strategy doing antibiotic sensitivity test. *Austin J Vet Sci Anim Husb*, 4 (1):1030, 2017.

C. Hot, N. Berthet, and O. Chesneau. Characterization of a novel gene responsible for lincosamide and streptogramin a resistance in *Staphylococcus sciuri*. *Antimicrobial Agents and Chemotherapy*, 58(6):3335–3341, 2014. doi: 10.1128/aac.02797-13. URL https://aac.asm.org/content/aac/58/6/3335.full.pdf.

F. Huber, B. Ehrensperger, C. Hatz, F. Chappuis, S. Bühler, and G. Eperon. Safety of live vaccines on immunosuppressive or immunomodulatory therapy—a retrospective study in three swiss travel clinics. *Journal of Travel Medicine*, 25(1), 2018. ISSN 1708-8305. doi: 10.1093/jtm/tax082. URL https://doi.org/10.1093/jtm/tax082.

M. Hunt, A. E. Mather, L. Sánchez-Busó, A. J. Page, J. Parkhill, J. A. Keane, and S. R. Harris. Ariba: rapid antimicrobial resistance genotyping directly from sequencing reads. *Microbial genomics*, 3(10):e000131–e000131, 2017. ISSN 2057-5858. doi: 10.1099/mgen.0.000131. URL https://pubmed.ncbi.nlm.nih.gov/29177089https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5695208/.

J. N. Huxley, M. J. Greent, L. E. Green, and A. J. Bradley. Evaluation of the efficacy of an internal teat sealer during the dry period. *Journal of Dairy Science*, 85(3): 551–61, 2002. ISSN 0022-0302 (Print) 0022-0302. doi: 10.3168/jds.s0022-0302(02)74108-8.

D. Hyatt, G.-L. Chen, P. F. Locascio, M. L. Land, F. W. Larimer, and L. J. Hauser. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*, 11(1):119, 2010. ISSN 1471-2105. doi: 10.1186/1471-2105-11-119. URL https://dx.doi.org/10.1186/1471-2105-11-119.

D. IaI and A. Tonevitskiĭ. Linear b-cell epitope prediction. *Molekuliarnaia biologiia*, 43(1):166–174, 2009.

E. A. Idelevich, C. Kreis, B. Löffler, and G. Peters. Staphylococcus aureus-*Associated Musculoskeletal Infections*, pages 229–261. Springer International Publishing, 2016. ISBN 0070-217X. doi: 10.1007/82_2016_19. URL https://dx.doi.org/10.1007/82_2016_19.

M. Inouye, H. Dashnow, L.-A. Raven, M. B. Schultz, B. J. Pope, T. Tomita, J. Zobel,

and K. E. Holt. Srst2: rapid genomic surveillance for public health and hospital microbiology labs. *Genome medicine*, 6(11):1–16, 2014.

T. Ito and K. Hiramatsu. Acquisition of methicillin resistance and progression of multiantibiotic resistance in methicillin-resistant *Staphylococcus aureus*. *Yonsei Medical Journal*, 39(6):526, 1998. ISSN 0513-5796. doi: 10.3349/ymj.1998.39.6.526. URL https://dx.doi.org/10.3349/ymj.1998.39.6.526.

A. Iwasaki and S. B. Omer. Why and how vaccines work. *Cell*, 183(2):290–295, 2020.

S. D. Jackman, B. P. Vandervalk, H. Mohamadi, J. Chu, S. Yeo, S. A. Hammond, G. Jahesh, H. Khan, L. Coombe, R. L. Warren, and I. Birol. Abyss 2.0: resource-efficient assembly of large genomes using a bloom filter. *Genome Research*, 27(5):768–777, 2017. ISSN 1088-9051. doi: 10.1101/gr.214346.116. URL https://dx.doi.org/10.1101/gr.214346.116.

S. Jaeger, F. Virchow, P. R. Torgerson, M. Bischoff, B. Biner, S. Hartnack, and S. R. Rüegg. Test characteristics of milk amyloid a elisa, somatic cell count, and bacteriological culture for detection of intramammary pathogens that cause subclinical mastitis. *Journal of Dairy Science*, 100(9):7419–7426, 2017. ISSN 0022-0302. doi: https://doi.org/10.3168/jds.2016-12446. URL https://www.sciencedirect.com/science/article/pii/S0022030217306070.

V. Jaiswal, S. K. Chanumolu, A. Gupta, R. S. Chauhan, and C. Rout. Jenner-predict server: prediction of protein vaccine candidates (pvcs) in bacteria based on host-pathogen interactions. *BMC bioinformatics*, 14(1):1–11, 2013.

H. Jamali, B. Radmehr, and S. Ismail. Short communication: Prevalence and antibiotic resistance of *Staphylococcus aureus* isolated from bovine clinical mastitis. *Journal of Dairy Science*, 97(4):2226–2230, 2014. ISSN 0022-0302. doi: https://doi.org/10.3168/jds.2013-7509. URL https://www.sciencedirect.com/science/article/pii/S0022030214001246.

S. Jarraud, G. J. Lyon, A. M. Figueiredo, G. Lina, F. Vandenesch, J. Etienne, T. W. Muir, and R. P. Novick. Exfoliatin-producing strains define a fourth agr specificity group in *Staphylococcus aureus*. *J Bacteriol*, 182(22):6517–22, 2000. ISSN 0021-9193 (Print) 0021-9193. doi: 10.1128/jb.182.22.6517-6522.2000.

S. Jarraud, M. A. Peyrat, A. Lim, A. Tristan, M. Bes, C. Mougel, J. Etienne, F. Vandenesch, M. Bonneville, and G. Lina. *egc*, a highly prevalent operon of enterotoxin gene, forms a putative nursery of superantigens in *Staphylococcus aureus*. *The Journal of Immunology*, 166(1):669–677, 2001. doi: 10.4049/jimmunol.166.1.669. URL https://www.jimmunol.org/content/jimmunol/166/1/669.full.pdf.

S. O. Jensen and B. R. Lyon. Genetics of antimicrobial resistance in *Staphylococcus aureus*. *Future Microbiol*, 4(5):565–82, 2009. ISSN 1746-0913. doi: 10.2217/fmb.09.30.

M. C. Jespersen, B. Peters, M. Nielsen, and P. Marcatili. Bepipred-2.0: improving sequence-based b-cell epitope prediction using conformational epitopes. *Nucleic acids research*, 45(W1):W24–W29, 2017.

Y. Ji. *Methicillin-Resistant* Staphylococcus aureus *(MRSA) Protocols: Cutting-Edge Technologies and Advancements*. Springer, 2020.

Y. Jian and M. Li. A narrative review of single-nucleotide polymorphism detection methods and their application in studies of *Staphylococcus aureus*. *Journal of Bio-X Research*, 4(1):1–9, 2021. ISSN 2096-5672. doi: 10.1097/jbr.0000000000000071. URL https://journals.lww.com/jbioxresearch/Fulltext/2021/03000/A_narrative_review_of_single_nucleotide.1.aspx.

H. Jiang, R. Lei, S.-W. Ding, and S. Zhu. Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC Bioinformatics*, 15(1):182, 2014. ISSN 1471-2105. doi: 10.1186/1471-2105-15-182. URL https://dx.doi.org/10.1186/1471-2105-15-182.

H. Jick. Adverse reactions to trimethoprim-sulfamethoxazole in hospitalized patients. *Reviews of Infectious Diseases*, 4(2):426–428, 1982. ISSN 1058-4838. doi: 10.1093/clinids/4.2.426. URL https://dx.doi.org/10.1093/clinids/4.2.426.

K. G. Joensen, F. Scheutz, O. Lund, H. Hasman, R. S. Kaas, E. M. Nielsen, and F. M. Aarestrup. Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic *Escherichia coli*. *J Clin Microbiol*, 52(5): 1501–10, 2014. ISSN 0095-1137 (Print) 0095-1137. doi: 10.1128/jcm.03617-13.

M. B. Johansen, L. Kiemer, and S. Brunak. Analysis and prediction of mammalian protein glycation. *Glycobiology*, 16(9):844–853, 2006.

K. A. Jolley, J. E. Bray, and M. C. J. Maiden. Open-access bacterial population genomics: Bigsdb software, the pubmlst.org website and their applications. *Wellcome open research*, 3:124–124, 2018. ISSN 2398-502X. doi: 10.12688/wellcomeopenres.14826.1. URL https://www.ncbi.nlm.nih.gov/pubmed/30345391https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6192448/.

A. Joshi, G. Pancari, L. Cope, E. P. Bowman, D. Cua, R. A. Proctor, and T. McNeely. Immunization with *Staphylococcus aureus* iron regulated surface determinant b (isdb) confers protection via th17/il17 pathway in a murine sepsis model. *Human vaccines & immunotherapeutics*, 8(3):336–346, 2012. ISSN 2164-554X 2164-5515. doi: 10.4161/hv.18946. URL https://pubmed.ncbi.nlm.nih.gov/22327491https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3426080/.

T. H. Jukes and C. R. Cantor. *CHAPTER 24 - Evolution of Protein Molecules*, pages 21–132. Academic Press, 1969. ISBN 978-1-4832-3211-9. doi: https://doi.org/10.1016/B978-1-4832-3211-9.50009-7. URL http://www.sciencedirect.com/science/article/pii/B9781483232119500097.

S. Jünemann, F. J. Sedlazeck, K. Prior, A. Albersmeier, U. John, J. Kalinowski, A. Mellmann, A. Goesmann, A. von Haeseler, J. Stoye, and D. Harmsen. Updating benchtop sequencing performance comparison. *Nat Biotechnol*, 31(4):294–6, 2013. ISSN 1087-0156. doi: 10.1038/nbt.2522.

G. W. Kaatz, R. V. Thyagarajan, and S. M. Seo. Effect of promoter region mutations and mgra overexpression on transcription of nora, which encodes a *Staphylococcus aureus* multidrug efflux transporter. *Antimicrobial agents and chemotherapy*, 49 (1):161–169, 2005. ISSN 0066-4804 1098-6596. doi: 10.1128/AAC.49.1.161-169.2005. URL https://pubmed.ncbi.nlm.nih.gov/15616291https://www.ncbi.nlm.nih.gov/pmc/articles/PMC538897/.

K. Kadlec, M. Entorf, and T. Peters. Occurrence and characteristics of livestock-associated methicillin-resistant *Staphylococcus aureus* in quarter milk

samples from dairy cows in germany. *Frontiers in Microbiology*, 10(1295), 2019. ISSN 1664-302X. doi: 10.3389/fmicb.2019.01295. URL https://www.frontiersin.org/article/10.3389/fmicb.2019.01295.

B. Kahl, M. Herrmann, A. S. Everding, H. G. Koch, K. Becker, E. Harms, R. A. Proctor, and G. Peters. Persistent infection with small colony variant strains of *Staphylococcus aureus* in patients with cystic fibrosis. *J Infect Dis*, 177(4):1023–9, 1998. ISSN 0022-1899 (Print) 0022-1899. doi: 10.1086/515238.

A. H. Kampen, T. Tollersrud, and A. Lund. *Staphylococcus aureus* capsular polysaccharide types 5 and 8 reduce killing by bovine neutrophils *in vitro*. *Infection and Immunity*, 73(3):1578, 2005. doi: 10.1128/IAI.73.3.1578-1583.2005. URL http://iai.asm.org/content/73/3/1578.abstract.

I. Kanevsky-Mullarky, M. Lehtimäki, W. Wark, and W. Mwangi. The role of th17+ t cells and neutrophils in mucosal immunity against*Staphylococcus aureus* mastitis. (p4256). *The Journal of Immunology*, 190(1 Supplement):54.2, 2013. URL http://www.jimmunol.org/content/190/1_Supplement/54.2.abstract.

M. Karahan, M. N. Açık, and B. Çetinkaya. Investigation of toxin genes by polymerase chain reaction in *Staphylococcus aureus* strains isolated from bovine mastitis in turkey. *Foodborne Pathogens and Disease*, 6(8):1029–1035, 2009. ISSN 1535-3141. doi: 10.1089/fpd.2009.0304. URL https://doi.org/10.1089/fpd.2009.0304.

W. W. Karakawa, J. M. Fournier, W. F. Vann, R. Arbeit, R. S. Schneerson, and J. B. Robbins. Method for the serological typing of the capsular polysaccharides of *Staphylococcus aureus*. *Journal of clinical microbiology*, 22(3):445–447, 1985. ISSN 0095-1137 1098-660X. URL https://www.ncbi.nlm.nih.gov/pubmed/3930565https://www.ncbi.nlm.nih.gov/pmc/articles/PMC268430/.

K. Kawai, Y. Kondo, Y. Shinozuka, R. Kawata, S. Kaneko, H. Iwano, M. Enokidani, A. Watanabe, F. Yuliza-Purba, N. Isobe, and T. Kurumisawa. Immune response during the onset of coliform mastitis in dairy cows vaccinated with startvac®. *Animal Science Journal*, 92(1):e13502, 2021. ISSN 1344-3941. doi: https://doi.org/10.1111/asj.13502. URL https://doi.org/10.1111/asj.13502.

G. Keefe. Update on control of *Staphylococcus aureus* and streptococcus agalactiae for management of mastitis. *Veterinary Clinics of North America: Food Animal Practice*, 28(2):203–216, 2012. ISSN 0749-0720. doi: 10.1016/j.cvfa.2012.03.010. URL https://dx.doi.org/10.1016/j.cvfa.2012.03.010.

K. Kengsakul, K. Sathirapongsasuti, and S. Punyagupta. Fatal myeloencephalitis following yellow fever vaccination in a case with hiv infection. *J Med Assoc Thai*, 85(1):131–4, 2002. ISSN 0125-2208 (Print) 0125-2208.

J. I. Kerslake, P. R. Amer, P. L. O'Neill, S. L. Wong, J. R. Roche, and C. V. C. Phyn. Economic costs of recorded reasons for cow mortality and culling in a pasture-based dairy industry. *Journal of Dairy Science*, 101(2):1795–1803, 2018. ISSN 0022-0302. doi: 10.3168/jds.2017-13124. URL https://dx.doi.org/10.3168/jds.2017-13124.

I. M. Keseler, C. Bonavides-Martínez, J. Collado-Vides, S. Gama-Castro, R. P. Gunsalus, D. A. Johnson, M. Krummenacker, L. M. Nolan, S. Paley, I. T. Paulsen, M. Peralta-Gil, A. Santos-Zavaleta, A. G. Shearer, and P. D. Karp. Ecocyc: a comprehensive view of *Escherichia coli* biology. *Nucleic Acids Res*, 37(Database issue):D464–70, 2009. ISSN 0305-1048 (Print) 0305-1048. doi: 10.1093/nar/gkn751.

G. G. Khachatourians. Agricultural use of antibiotics and the evolution and transfer of antibiotic-resistant bacteria. *CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne*, 159(9):1129–1136, 1998. ISSN 0820-3946 1488-2329. URL https://pubmed.ncbi.nlm.nih.gov/9835883https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1229782/.

M. A. Khan, C. H. Kim, I. Kakoma, E. Morin, R. D. Hansen, W. L. Hurley, D. N. Tripathy, and B. K. Baek. Detection of *Staphylococcus aureus* in milk by use of polymerase chain reaction analysis. *Am J Vet Res*, 59(7):807–13, 1998. ISSN 0002-9645 (Print) 0002-9645.

A. B. Khodursky and N. R. Cozzarelli. The mechanism of inhibition of topoisomerase iv by quinolone antibacterials. *Journal of Biological Chemistry*, 273 (42):27668–27677, 1998. ISSN 0021-9258. doi: 10.1074/jbc.273.42.27668. URL https://dx.doi.org/10.1074/jbc.273.42.27668.

A. B. Khodursky, E. L. Zechiedrich, and N. R. Cozzarelli. Topoisomerase iv is a target of quinolones in *Escherichia coli*. *Proceedings of the National Academy of Sciences*, 92(25):11801–11805, 1995. doi: 10.1073/pnas.92.25.11801. URL https://www.pnas.org/content/pnas/92/25/11801.full.pdf.

L. Kiemer, J. D. Bendtsen, and N. Blom. Netacet: prediction of n-terminal acetylation sites. *Bioinformatics*, 21(7):1269–1270, 2005.

C. H. Kim, M. Khan, D. E. Morin, W. L. Hurley, D. N. Tripathy, M. Kehrli, A. O. Oluoch, and I. Kakoma. Optimization of the pcr for detection of *Staphylococcus aureus* nuc gene in bovine milk. *Journal of Dairy Science*, 84(1):74–83, 2001. ISSN 0022-0302. doi: 10.3168/jds.s0022-0302(01)74454-2. URL https://dx.doi.org/10.3168/jds.s0022-0302(01)74454-2.

A. Kimura, D. Toneatto, A. Kleinschmidt, H. Wang, and P. Dull. Immunogenicity and safety of a multicomponent meningococcal serogroup b vaccine and a quadrivalent meningococcal crm197 conjugate vaccine against serogroups a, c, w-135, and y in adults who are at increased risk for occupational exposure to meningococcal isolates. *Clinical and Vaccine Immunology*, 18(3):483–486, 2011.

J. Kluytmans, A. van Belkum, and H. Verbrugh. Nasal carriage of *Staphylococcus aureus*: epidemiology, underlying mechanisms, and associated risks. *Clinical microbiology reviews*, 10(3):505–520, 1997. ISSN 0893-8512 1098-6618. URL https://www.ncbi.nlm.nih.gov/pubmed/9227864https://www.ncbi.nlm.nih.gov/pmc/articles/PMC172932/.

Y.-P. Ko, A. Kuipers, C. M. Freitag, I. Jongerius, E. Medina, W. J. van Rooijen, A. N. Spaan, K. P. M. van Kessel, M. Höök, and S. H. M. Rooijakkers. Phagocytosis escape by a *Staphylococcus aureus* protein that connects complement and coagulation proteins at the bacterial surface. *PLOS Pathogens*, 9(12):e1003816, 2013. doi: 10.1371/journal.ppat.1003816. URL https://doi.org/10.1371/journal.ppat.1003816.

M. Kokot, M. Długosz, and S. Deorowicz. Kmc 3: counting and manipulating k-mer statistics. *Bioinformatics*, 33(17):2759–2761, 2017. ISSN 1367-4803. doi: 10.1093/bioinformatics/btx304. URL https://doi.org/10.1093/bioinformatics/btx304.

H. Komatsuzawa, M. Sugai, S. Nakashima, S. Yamada, A. Matsumoto, T. Oshida, and H. Suginaka. Subcellular localization of the major autolysin, atl and its processed proteins in *Staphylococcus aureus*. *Microbiol Immunol*, 41(6):469–79, 1997. ISSN 0385-5600 (Print) 0385-5600. doi: 10.1111/j.1348-0421.1997.tb01880.x.

K.-F. Kong, L. Schneper, and K. Mathee. Beta-lactam antibiotics: from antibiosis to resistance and bacteriology. *APMIS*, 118(1):1–36, 2010. ISSN 0903-4641. doi: 10.1111/j.1600-0463.2009.02563.x. URL https://dx.doi.org/10.1111/j.1600-0463.2009.02563.x.

C. G. Korea, G. Balsamo, A. Pezzicoli, C. Merakou, S. Tavarini, F. Bagnoli, D. Serruto, and M. Unnikrishnan. Staphylococcal esx proteins modulate apoptosis and release of intracellular *Staphylococcus aureus* during infection in epithelial cells. *Infect Immun*, 82(10):4144–53, 2014. ISSN 0019-9567 (Print) 0019-9567. doi: 10.1128/iai.01576-14.

L. Koreen, S. V. Ramaswamy, E. A. Graviss, S. Naidich, J. M. Musser, and B. N. Kreiswirth. spa typing method for discriminating among*Staphylococcus aureus* isolates: Implications for use of a single marker to detect genetic micro- and macrovariation. *Journal of Clinical Microbiology*, 42(2):792, 2004. doi: 10.1128/JCM.42.2.792-799.2004. URL http://jcm.asm.org/content/42/2/792.abstract.

A. Krogh, B. Larsson, G. von Heijne, and E. L. Sonnhammer. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. *Journal of Molecular Biology*, 305(3):567–80, 2001. ISSN 0022-2836 (Print) 0022-2836. doi: 10.1006/jmbi.2000.4315.

G. Kuhn, P. Francioli, and D. S. Blanc. Evidence for clonal evolution among highly polymorphic genes in methicillin-resistant *Staphylococcus aureus*. *Journal of Bacteriology*, 188(1):169–178, 2006. ISSN 0021-9193. doi: 10.1128/jb.188.1.169-178.2006. URL https://dx.doi.org/10.1128/JB.188.1.169-178.2006.

I. Kullik, P. Giachino, and T. Fuchs. Deletion of the alternative sigma factor sigmab in *Staphylococcus aureus* reveals its function as a global regulator of virulence genes. *Journal of bacteriology*, 180(18):4814–4820, 1998. ISSN 0021-9193 1098-5530. doi: 10.1128/JB.180.18.4814-4820.1998. URL

https://pubmed.ncbi.nlm.nih.gov/9733682https://www.ncbi.nlm.nih.gov/pmc/articles/PMC107504/.

I. I. Kullik and P. Giachino. The alternative sigma factor sigmab in *Staphylococcus aureus*: regulation of the sigb operon in response to growth phase and heat shock. *Arch Microbiol*, 167(2-3):151–9, 1997. ISSN 0302-8933. doi: 10.1007/s002030050428.

M. Kuroda, T. Ohta, I. Uchiyama, T. Baba, H. Yuzawa, I. Kobayashi, L. Cui, A. Oguchi, K.-I. Aoki, Y. Nagai, J. Lian, T. Ito, M. Kanamori, H. Matsumaru, A. Maruyama, H. Murakami, A. Hosoyama, Y. Mizutani-Ui, N. K. Takahashi, T. Sawano, R.-I. Inoue, C. Kaito, K. Sekimizu, H. Hirakawa, S. Kuhara, S. Goto, J. Yabuzaki, M. Kanehisa, A. Yamashita, K. Oshima, K. Furuya, C. Yoshino, T. Shiba, M. Hattori, N. Ogasawara, H. Hayashi, and K. Hiramatsu. Whole genome sequencing of meticillin-resistant *Staphylococcus aureus*. *The Lancet*, 357(9264): 1225–1240, 2001. ISSN 0140-6736. doi: 10.1016/s0140-6736(00)04403-2. URL https://dx.doi.org/10.1016/s0140-6736(00)04403-2.

K. Kuzma, E. Malinowski, H. Lassa, and A. Klossowska. Detection of genes for enterotoxins and toxic shock syndrome toxin-1 in *Staphylococcus aureus* isolated from bovine mastitis. *BULLETIN-VETERINARY INSTITUTE IN PULAWY*, 47(2): 419–426, 2003. ISSN 0042-4870.

S. M. Lakin, C. Dean, N. R. Noyes, A. Dettenwanger, A. S. Ross, E. Doster, P. Rovira, Z. Abdo, K. L. Jones, J. Ruiz, et al. Megares: an antimicrobial resistance database for high throughput sequencing. *Nucleic acids research*, 45(D1): D574–D580, 2017.

A. Lammers, P. J. M. Nuijten, and H. E. Smith. The fibronectin binding proteins of *Staphylococcus aureus* are required for adhesion to and invasion of bovine mammary gland cells. *FEMS Microbiology Letters*, 180(1):103–109, 1999. ISSN 0378-1097. doi: 10.1111/j.1574-6968.1999.tb08783.x. URL https://doi.org/10.1111/j.1574-6968.1999.tb08783.x.

H. Landin, M. J. Mörk, M. Larsson, and K. P. Waller. Vaccination against *Staphylococcus aureus* mastitis in two swedish dairy herds. *Acta Veterinaria Scandinavica*, 57(1):81, 2015. ISSN 1751-0147. doi: 10.1186/s13028-015-0171-6. URL https://doi.org/10.1186/s13028-015-0171-6.

S. Lang, M. A. Livesley, P. A. Lambert, W. A. Littler, and T. S. Elliott. Identification of a novel antigen from staphylococcus epidermidis. *FEMS Immunol Med Microbiol*, 29(3):213–20, 2000. ISSN 0928-8244 (Print) 0928-8244. doi: 10.1111/j.1574-695X.2000.tb01525.x.

B. Langmead and S. L. Salzberg. Fast gapped-read alignment with bowtie 2. *Nature Methods*, 9(4):357–359, 2012. ISSN 1548-7105. doi: 10.1038/nmeth.1923. URL https://doi.org/10.1038/nmeth.1923.

B. Langmead, C. Trapnell, M. Pop, and S. L. Salzberg. Ultrafast and memory-efficient alignment of short dna sequences to the human genome. *Genome Biology*, 10(3):R25, 2009. ISSN 1465-6906. doi: 10.1186/gb-2009-10-3-r25. URL https://dx.doi.org/10.1186/gb-2009-10-3-r25.

J. Lannergård, T. Norström, and D. Hughes. Genetic determinants of resistance to fusidic acid among clinical bacteremia isolates of *Staphylococcus aureus*. *Antimicrobial Agents and Chemotherapy*, 53(5):2059–2065, 2009. doi: 10.1128/aac.00871-08. URL https://aac.asm.org/content/aac/53/5/2059.full.pdf.

D. Laslett. Aragorn, a program to detect trna genes and tmrna genes in nucleotide sequences. *Nucleic Acids Research*, 32(1):11–16, 2004. ISSN 1362-4962. doi: 10.1093/nar/gkh152. URL https://dx.doi.org/10.1093/nar/gkh152.

B. S. Laursen, H. P. Sørensen, K. K. Mortensen, and H. U. Sperling-Petersen. Initiation of protein synthesis in bacteria. *Microbiology and molecular biology reviews : MMBR*, 69(1):101–123, 2005. ISSN 1092-2172 1098-5557. doi: 10.1128/MMBR.69.1.101-123.2005. URL https://pubmed.ncbi.nlm.nih.gov/15755955https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1082788/.

R. Laxminarayan, P. Matsoso, S. Pant, C. Brower, J.-A. Røttingen, K. Klugman, and S. Davies. Access to effective antimicrobials: a worldwide challenge. *The Lancet*, 387(10014):168–175, 2016. ISSN 01406736. doi: 10.1016/s0140-6736(15)00474-2.

K. Y. Le and M. Otto. Quorum-sensing regulation in staphylococci—an overview. *Frontiers in Microbiology*, 6, 2015. ISSN 1664-302X. doi:

10.3389/fmicb.2015.01174. URL https://dx.doi.org/10.3389/fmicb.2015.01174.

R. Leclercq. Mechanisms of resistance to macrolides and lincosamides: Nature of the resistance elements and their clinical implications. *Clinical Infectious Diseases*, 34 (4):482–492, 2002. ISSN 1058-4838. doi: 10.1086/324626. URL https://doi.org/10.1086/324626.

J. H. Lee. Methicillin (oxacillin)-resistant *Staphylococcus aureus* strains isolated from major food animals and their potential transmission to humans. *Applied and environmental microbiology*, 69(11):6489–6494, 2003.

P. K. Lee, B. N. Kreiswirth, J. R. Deringer, S. J. Projan, W. Eisner, B. L. Smith, E. Carlson, R. P. Novick, and P. M. Schlievert. Nucleotide sequences and biologic properties of toxic shock syndrome toxin 1 from ovine- and bovine-associated *Staphylococcus aureus*. *The Journal of Infectious Diseases*, 165(6):1056–1063, 1992. ISSN 0022-1899. doi: 10.1093/infdis/165.6.1056. URL https://doi.org/10.1093/infdis/165.6.1056.

S. H. I. Lee, B. L. C. Mangolin, J. L. Gonçalves, D. V. Neeff, M. P. Silva, A. G. Cruz, and C. A. F. Oliveira. Biofilm-producing ability of *Staphylococcus aureus* isolates from brazilian dairy farms. *Journal of Dairy Science*, 97(3):1812–1816, 2014. ISSN 0022-0302. doi: https://doi.org/10.3168/jds.2013-7387. URL https://www.sciencedirect.com/science/article/pii/S002203021400040X.

M. Leijon, E. Atkins, K. P. Waller, and K. Artursson. Longitudinal study of staphylococcus aureus genotypes isolated from bovine clinical mastitis. *Journal of Dairy Science*, 2021.

R. Leinonen, H. Sugawara, and M. Shumway. The sequence read archive. *Nucleic Acids Research*, 39(Database):D19–D21, 2011. ISSN 0305-1048. doi: 10.1093/nar/gkq1019. URL https://dx.doi.org/10.1093/nar/gkq1019.

M. Levine, D. Herrington, G. Losonsky, B. Tall, J. Kaper, J. Ketley, C. Tacket, and S. Cryz. Safety, immunogenicity, and efficacy of recombinant live oral cholera vaccines, cvd 103 and cvd 103-hgr. *The Lancet*, 332(8609):467–470, 1988.

M. M. Levine, C. Ferreccio, R. E. Black, C. O. Tacket, R. Germanier, and C. T. Committee. Progress in vaccines against typhoid fever. *Clinical Infectious Diseases*, 11(Supplement_3):S552–S567, 1989.

D. Li, C.-M. Liu, R. Luo, K. Sadakane, and T.-W. Lam. Megahit: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de bruijn graph. *Bioinformatics*, 31(10):1674–1676, 2015a. ISSN 1460-2059. doi: 10.1093/bioinformatics/btv033. URL https://dx.doi.org/10.1093/bioinformatics/btv033.

D. Li, R. Luo, C. M. Liu, C. M. Leung, H. F. Ting, K. Sadakane, H. Yamashita, and T. W. Lam. Megahit v1.0: A fast and scalable metagenome assembler driven by advanced methodologies and community practices. *Methods*, 102:3–11, 2016. ISSN 1046-2023. doi: 10.1016/j.ymeth.2016.02.020.

H. Li. Aligning sequence reads, clone sequences and assembly contigs with bwa-mem. *arXiv preprint arXiv:1303.3997*, 2013.

H. Li. Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. *Bioinformatics*, 32(14):2103–2110, 2016. ISSN 1367-4803. doi: 10.1093/bioinformatics/btw152. URL https://dx.doi.org/10.1093/bioinformatics/btw152.

H. Li, 2018. URL https://github.com/lh3/seqtk.

H. Li and R. Durbin. Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics*, 25(14):1754–1760, 2009. ISSN 1367-4803. doi: 10.1093/bioinformatics/btp324. URL https://dx.doi.org/10.1093/bioinformatics/btp324.

H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin. The sequence alignment/map format and samtools. *Bioinformatics*, 25(16):2078–2079, 2009. ISSN 1367-4803. doi: 10.1093/bioinformatics/btp352. URL https://dx.doi.org/10.1093/bioinformatics/btp352.

L. Li, L. Zhou, L. Wang, H. Xue, and X. Zhao. Characterization of methicillin-resistant and -susceptible staphylococcal isolates from bovine milk in northwestern china. *PloS one*, 10(3):e0116699–e0116699, 2015b. ISSN 1932-6203. doi: 10.1371/journal.pone.0116699. URL https://pubmed.ncbi.nlm.nih.gov/25756992https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4355487/.

L. Li, L. Wang, Y. Gao, J. Wang, and X. Zhao. Effective antimicrobial activity of plectasin-derived antimicrobial peptides against *Staphylococcus aureus* infection in mammary glands. *Frontiers in microbiology*, 8:2386–2386, 2017. ISSN 1664-302X. doi: 10.3389/fmicb.2017.02386. URL https://www.ncbi.nlm.nih.gov/pubmed/29255451https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5723147/.

S. Li, R. L. Skov, X. Han, A. R. Larsen, J. Larsen, M. Sørum, M. Wulf, A. Voss, K. Hiramatsu, and T. Ito. Novel types of staphylococcal cassette chromosome mec elements identified in clonal complex 398 methicillin-resistant *Staphylococcus aureus* strains. *Antimicrobial agents and chemotherapy*, 55(6):3046–3050, 2011. ISSN 1098-6596 0066-4804. doi: 10.1128/AAC.01475-10. URL https://pubmed.ncbi.nlm.nih.gov/21422209https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3101438/.

W. Li, G. C. Atkinson, N. S. Thakor, U. Allas, C.-c. Lu, K.-Y. Chan, T. Tenson, K. Schulten, K. S. Wilson, V. Hauryliuk, and J. Frank. Mechanism of tetracycline resistance by ribosomal protection protein tet(o). *Nature Communications*, 4(1): 1477, 2013. ISSN 2041-1723. doi: 10.1038/ncomms2470. URL https://doi.org/10.1038/ncomms2470.

D. Lim and N. C. J. Strynadka. Structural basis for the β-lactam resistance of pbp2a from methicillin-resistant *Staphylococcus aureus*. *Nature Structural Biology*, 9 (11):870–876, 2002. ISSN 1545-9985. doi: 10.1038/nsb858. URL https://doi.org/10.1038/nsb858.

J. A. Lindsay. Genomic variation and evolution of *Staphylococcus aureus*. *Int J Med Microbiol*, 300(2-3):98–103, 2010. ISSN 1618-0607 (Electronic) 1438-4221 (Linking). doi: 10.1016/j.ijmm.2009.08.013. URL https://www.ncbi.nlm.nih.gov/pubmed/19811948.

J. A. Lindsay and M. T. G. Holden. *Staphylococcus aureus* : superbug, super genome? *Trends in Microbiology*, 12(8):378–385, 2004. ISSN 0966-842X. doi: 10.1016/j.tim.2004.06.004. URL https://dx.doi.org/10.1016/j.tim.2004.06.004.

J. A. Lindsay, C. E. Moore, N. P. Day, S. J. Peacock, A. A. Witney, R. A. Stabler, S. E. Husain, P. D. Butcher, and J. Hinds. Microarrays reveal that each of the ten

dominant lineages of *Staphylococcus aureus* has a unique combination of surface-associated and regulatory genes. *Journal of Bacteriology*, 188(2):669–676, 2006. doi: 10.1128/JB.188.2.669-676.2006. URL https://journals.asm.org/doi/abs/10.1128/JB.188.2.669-676.2006.

L. Lingzhi, G. Haojie, G. Dan, M. Hongmei, L. Yang, J. Mengdie, Z. Chengkun, and Z. Xiaohui. The role of two-component regulatory system in β-lactam antibiotics resistance. *Microbiological Research*, 215:126–129, 2018. ISSN 0944-5013. doi: 10.1016/j.micres.2018.07.005. URL https://dx.doi.org/10.1016/j.micres.2018.07.005.

J. L. Lister and A. R. Horswill. Staphylococcus aureus biofilms: recent developments in biofilm dispersal. *Frontiers in cellular and infection microbiology*, 4:178, 2014.

B. Liu and M. Pop. Ardb—antibiotic resistance genes database. *Nucleic acids research*, 37(suppl_1):D443–D447, 2009.

H. Liu, N. K. Archer, C. A. Dillen, Y. Wang, A. G. Ashbaugh, R. V. Ortines, T. Kao, S. K. Lee, S. S. Cai, R. J. Miller, M. C. Marchitto, E. Zhang, D. P. Riggins, R. D. Plaut, S. Stibitz, R. S. Geha, and L. S. Miller. *Staphylococcus aureus* epicutaneous exposure drives skin inflammation via il-36-mediated t cell responses. *Cell Host Microbe*, 22(5):653–666 e5, 2017. ISSN 1934-6069 (Electronic) 1931-3128 (Linking). doi: 10.1016/j.chom.2017.10.006. URL https://www.ncbi.nlm.nih.gov/pubmed/29120743.

J. Liu, D. Chen, B. M. Peters, L. Li, B. Li, Z. Xu, and M. E. Shirliff. Staphylococcal chromosomal cassettes mec (sccmec): A mobile genetic element in methicillin-resistant *Staphylococcus aureus*. *Microb Pathog*, 101:56–67, 2016. ISSN 0882-4010. doi: 10.1016/j.micpath.2016.10.028.

W. Liu and Y. H. Chen. High epitope density in a single protein molecule significantly enhances antigenicity as well as immunogenicity: a novel strategy for modern vaccine development and a preliminary investigation about b cell discrimination of monomeric proteins. *Eur J Immunol*, 35(2):505–14, 2005. ISSN 0014-2980 (Print) 0014-2980. doi: 10.1002/eji.200425749.

Y. Liu, A. C. Manna, C. H. Pan, I. A. Kriksunov, D. J. Thiel, A. L. Cheung, and G. Zhang. Structural and function analyses of the global regulatory protein sara

from *Staphylococcus aureus*. *Proc Natl Acad Sci U S A*, 103(7):2392–7, 2006. ISSN 0027-8424 (Print) 0027-8424. doi: 10.1073/pnas.0510439103.

M. Lombard, P.-P. Pastoret, and A. Moulin. A brief history of vaccines and vaccination. *Revue Scientifique et Technique-Office International des Epizooties*, 26(1):29–48, 2007.

U. Lorenz, B. Lorenz, T. Schmitter, K. Streker, C. Erck, J. Wehland, J. Nickel, B. Zimmermann, and K. Ohlsen. Functional antibodies targeting isaa of*Staphylococcus aureus* augment host immune response and open new perspectives for antibacterial therapy. *Antimicrobial Agents and Chemotherapy*, 55 (1):165, 2011. doi: 10.1128/AAC.01144-10. URL http://aac.asm.org/content/55/1/165.abstract.

F. D. Lowy. Antimicrobial resistance: the example of *Staphylococcus aureus*. *The Journal of clinical investigation*, 111(9):1265–1273, 2003. ISSN 0021-9738. doi: 10.1172/JCI18535. URL https://www.ncbi.nlm.nih.gov/pubmed/12727914https://www.ncbi.nlm.nih.gov/pmc/articles/PMC154455/.

C. Lozano, C. Aspiroz, A. Rezusta, E. Gómez-Sanz, C. Simon, P. Gómez, C. Ortega, M. J. Revillo, M. Zarazaga, and C. Torres. Identification of novel vga(a)-carrying plasmids and a tn5406-like transposon in meticillin-resistant *Staphylococcus aureus* and staphylococcus epidermidis of human and animal origin. *International Journal of Antimicrobial Agents*, 40(4):306–312, 2012. ISSN 0924-8579. doi: https://doi.org/10.1016/j.ijantimicag.2012.06.009. URL http://www.sciencedirect.com/science/article/pii/S0924857912002646.

C. D. Luby, J. R. Middleton, J. Ma, C. L. Rinehart, S. Bucklin, C. Kohler, and J. W. Tyler. Characterization of the antibody isotype response in serum and milk of heifers vaccinated with a *Staphylococcus aureus* bacterin (lysigintm). *Journal of Dairy Research*, 74(2):239–246, 2007. ISSN 0022-0299. doi: 10.1017/S0022029907002476. URL https://www.cambridge.org/core/article/characterization-of-the-antibody-isotype-response-in-serum-and-milk-of-heife E8744D00F96BB5EC184E2210449E99D6.

R. Luo, B. Liu, Y. Xie, Z. Li, W. Huang, J. Yuan, G. He, Y. Chen, Q. Pan, Y. Liu, J. Tang, G. Wu, H. Zhang, Y. Shi, Y. Liu, C. Yu, B. Wang, Y. Lu, C. Han, D. W.

Cheung, S.-M. Yiu, S. Peng, Z. Xiaoqian, G. Liu, X. Liao, Y. Li, H. Yang, J. Wang, T.-W. Lam, and J. Wang. Soapdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience*, 1(1):18, 2012. ISSN 2047-217X. doi: 10.1186/2047-217x-1-18. URL https://dx.doi.org/10.1186/2047-217x-1-18.

B. Löffler, M. Hussain, M. Grundmeier, M. Brück, D. Holzinger, G. Varga, J. Roth, B. C. Kahl, R. A. Proctor, and G. Peters. *Staphylococcus aureus* panton-valentine leukocidin is a very potent cytotoxic factor for human neutrophils. *PLoS Pathog*, 6 (1):e1000715, 2010. ISSN 1553-7366 (Print) 1553-7366. doi: 10.1371/journal.ppat.1000715.

J. Ma, J. Cocchiaro, and J. C. Lee. Evaluation of serotypes of *Staphylococcus aureus* strains used in the production of a bovine mastitis bacterin. *Journal of Dairy Science*, 87(1):178–182, 2004. ISSN 0022-0302. doi: https://doi.org/10.3168/jds.S0022-0302(04)73156-2. URL https://www.sciencedirect.com/science/article/pii/S0022030204731562.

C. M. Macleod. The inhibition of the bacteriostatic action of sulfonamide drugs by substances of animal and bacterial origin. *Journal of Experimental Medicine*, 72 (3):217–232, 1940. ISSN 0022-1007. doi: 10.1084/jem.72.3.217. URL https://dx.doi.org/10.1084/jem.72.3.217.

T. Madsen. Vaccination against whooping cough. *Journal of the American Medical Association*, 101(3):187–188, 1933.

T. Magoč and S. L. Salzberg. Flash: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*, 27(21):2957–63, 2011. ISSN 1367-4803 (Print) 1367-4803. doi: 10.1093/bioinformatics/btr507.

G. Magro, S. Biffani, G. Minozzi, R. Ehricht, S. Monecke, M. Luini, and R. Piccinini. Virulence genes of *S. aureus* from dairy cow mastitis and contagiousness risk. *Toxins*, 9(6):195, 2017. ISSN 2072-6651. doi: 10.3390/toxins9060195. URL https://pubmed.ncbi.nlm.nih.gov/28635647https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5488045/.

M. C. J. Maiden. The impact of nucleotide sequence analysis on meningococcal vaccine development and assessment. *Frontiers in immunology*, 9:3151, 2019.

M. C. J. Maiden, J. A. Bygraves, E. Feil, G. Morelli, J. E. Russell, R. Urwin, Q. Zhang, J. Zhou, K. Zurth, D. A. Caugant, I. M. Feavers, M. Achtman, and B. G. Spratt. Multilocus sequence typing: A portable approach to the identification of clones within populations of pathogenic microorganisms. *Proceedings of the National Academy of Sciences*, 95(6):3140–3145, 1998. ISSN 0027-8424. doi: 10.1073/pnas.95.6.3140. URL https://dx.doi.org/10.1073/pnas.95.6.3140.

D. Maione, I. Margarit, C. D. Rinaudo, V. Masignani, M. Mora, M. Scarselli, H. Tettelin, C. Brettoni, E. T. Iacobini, R. Rosini, et al. Identification of a universal group b streptococcus vaccine by multiple genome screen. *Science*, 309(5731): 148–150, 2005.

S. Mairpady Shambat, N. Siemens, I. R. Monk, D. B. Mohan, S. Mukundan, K. C. Krishnan, S. Prabhakara, J. Snäll, A. Kearns, F. Vandenesch, M. Svensson, M. Kotb, B. Gopal, G. Arakere, and A. Norrby-Teglund. A point mutation in agrc determines cytotoxic or colonizing properties associated with phenotypic variants of st22 mrsa strains. *Scientific Reports*, 6(1):31360, 2016. ISSN 2045-2322. doi: 10.1038/srep31360. URL https://doi.org/10.1038/srep31360.

N. Malachowa and F. R. DeLeo. Mobile genetic elements of *Staphylococcus aureus*. *Cell Mol Life Sci*, 67(18):3057–71, 2010. ISSN 1420-682x. doi: 10.1007/s00018-010-0389-4.

F. B. Malcata, P. T. Pepler, E. L. O'Reilly, N. Brady, P. D. Eckersall, R. N. Zadoks, and L. Viora. Point-of-care tests for bovine clinical mastitis: what do we have and what do we need? *Journal of Dairy Research*, 87(S1):60–66, 2020. ISSN 0022-0299.

W. Mamo, G. Fröman, and T. Wadström. Interaction of sub-epithelial connective tissue components with *Staphylococcus aureus* and coagulase-negative staphylococci from bovine mastitis. *Vet Microbiology*, 18(2):163–76, 1988. ISSN 0378-1135 (Print) 0378-1135. doi: 10.1016/0378-1135(88)90062-4.

W. Mamo, M. Bodén, and J.-I. Flock. Vaccination with *Staphylococcus aureus* fibrinogen binding proteins (fgbps) reduces colonisation of *S. aureus* in a mouse mastitis model. *FEMS Immunology & Medical Microbiology*, 10(1):47–53, 1994. ISSN 0928-8244. doi: 10.1111/j.1574-695X.1994.tb00010.x. URL https://doi.org/10.1111/j.1574-695X.1994.tb00010.x.

E. M. Mansion-de Vries, N. Knorr, J.-H. Paduch, C. Zinke, M. Hoedemaker, and V. Krömker. A field study evaluation of petrifilm™ plates as a 24-h rapid diagnostic test for clinical mastitis on a dairy farm. *Preventive Veterinary Medicine*, 113(4):620–624, 2014. ISSN 0167-5877. doi: https://doi.org/10.1016/j.prevetmed.2013.11.019. URL http://www.sciencedirect.com/science/article/pii/S016758771300370X.

J. M. Manson, J. M. Smith, and G. M. Cook. Persistence of vancomycin-resistant enterococci in new zealand broilers after discontinuation of avoparcin use. *Applied and environmental microbiology*, 70(10):5764–5768, 2004.

S. M. Marcy and J. O. Klein. The isoxazolyl penicillins: oxacillin, cloxacillin, and dicloxacillin. *Med Clin North Am*, 54(5):1127–43, 1970. ISSN 0025-7125 (Print) 0025-7125.

M. Margulies, M. Egholm, W. E. Altman, S. Attiya, J. S. Bader, L. A. Bemben, J. Berka, M. S. Braverman, Y.-J. Chen, Z. Chen, S. B. Dewell, L. Du, J. M. Fierro, X. V. Gomes, B. C. Godwin, W. He, S. Helgesen, C. H. Ho, G. P. Irzyk, S. C. Jando, M. L. I. Alenquer, T. P. Jarvie, K. B. Jirage, J.-B. Kim, J. R. Knight, J. R. Lanza, J. H. Leamon, S. M. Lefkowitz, M. Lei, J. Li, K. L. Lohman, H. Lu, V. B. Makhijani, K. E. McDade, M. P. McKenna, E. W. Myers, E. Nickerson, J. R. Nobile, R. Plant, B. P. Puc, M. T. Ronan, G. T. Roth, G. J. Sarkis, J. F. Simons, J. W. Simpson, M. Srinivasan, K. R. Tartaro, A. Tomasz, K. A. Vogt, G. A. Volkmer, S. H. Wang, Y. Wang, M. P. Weiner, P. Yu, R. F. Begley, and J. M. Rothberg. Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, 437(7057):376–380, 2005. ISSN 0028-0836. doi: 10.1038/nature03959. URL https://dx.doi.org/10.1038/nature03959.

V. F. Marques, H. A. Santos, T. H. Santos, D. A. Melo, S. M. Coelho, I. S. Coelho, and M. Souza. Expression of icaa and icad genes in biofilm formation in *Staphylococcus aureus* isolates from bovine subclinical mastitis. *Pesquisa Veterinária Brasileira*, 41, 2021. ISSN 0100-736X.

M. Martin. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, 17(1):10, 2011. ISSN 2226-6089. doi: 10.14806/ej.17.1.200. URL https://dx.doi.org/10.14806/ej.17.1.200.

M. J. Martin, S. E. Thottathil, and T. B. Newman. Antibiotics overuse in animal agriculture: A call to action for health care providers. *Am J Public Health*, 105(12):

2409–10, 2015. ISSN 1541-0048 (Electronic) 0090-0036 (Linking). doi: 10.2105/AJPH.2015.302870. URL https://www.ncbi.nlm.nih.gov/pubmed/26469675.

C. Martín-Bourgon. Phage-typing of staphylococcus aureus from cases of bacteraemia. *Journal of Hospital Infection*, 6(4):429–433, 1985.

K. B. Martins, P. Y. Faccioli, M. F. Bonesso, S. Fernandes, A. A. Oliveira, A. Dantas, L. F. Zafalon, and M. Cunha. Characteristics of resistance and virulence factors in different species of coagulase-negative staphylococci isolated from milk of healthy sheep and animals with subclinical mastitis. *Journal of Dairy Science*, 100(3): 2184–2195, 2017. ISSN 1525-3198 (Electronic) 0022-0302 (Linking). doi: 10.3168/jds.2016-11583. URL https://www.ncbi.nlm.nih.gov/pubmed/28109594.

A. Mason, D. Foster, P. Bradley, T. Golubchik, M. Doumith, N. C. Gordon, B. Pichon, Z. Iqbal, P. Staves, D. Crook, A. S. Walker, A. Kearns, and T. Peto. Accuracy of different bioinformatics methods in detecting antibiotic resistance and virulence factors from *Staphylococcus aureus* whole-genome sequences. *Journal of Clinical Microbiology*, 56(9):e01815–17, 2018. doi: 10.1128/JCM.01815-17. URL http://jcm.asm.org/content/56/9/e01815-17.abstract.

W. J. Mason, J. S. Blevins, K. Beenken, N. Wibowo, N. Ojha, and M. S. Smeltzer. Multiplex pcr protocol for the diagnosis of staphylococcal infection. *Journal of Clinical Microbiology*, 39(9):3332–3338, 2001. ISSN 0095-1137. doi: 10.1128/jcm.39.9.3332-3338.2001. URL https://dx.doi.org/10.1128/JCM.39.9.3332-3338.2001.

T. Matsunaga, S.-i. Kamata, N. Kakiichi, and K. Uchida. Characteristics of staphylococcus aureus isolated from peracute, acute and chronic bovine mastitis. *Journal of Veterinary Medical Science*, 55(2):297–300, 1993.

E. Matuschek, D. F. Brown, and G. Kahlmeter. Development of the eucast disk diffusion antimicrobial susceptibility testing method and its implementation in routine microbiology laboratories. *Clinical Microbiology and Infection*, 20(4): O255–O266, 2014.

M. Matuszewska, G. G. R. Murray, E. M. Harrison, M. A. Holmes, and L. A. Weinert. The evolutionary genomics of host specificity in *Staphylococcus aureus*. *Trends in*

Microbiology, 28(6):465–477, 2020. ISSN 0966-842X. doi: https://doi.org/10.1016/j.tim.2019.12.007. URL https://www.sciencedirect.com/science/article/pii/S0966842X19303221.

D. Matzov, Z. Eyal, R. I. Benhamou, M. Shalev-Benami, Y. Halfon, M. Krupkin, E. Zimmerman, H. Rozenberg, A. Bashan, M. Fridman, and A. Yonath. Structural insights of lincosamides targeting the ribosome of *Staphylococcus aureus*. *Nucleic acids research*, 45(17):10284–10292, 2017. ISSN 1362-4962 0305-1048. doi: 10.1093/nar/gkx658. URL https://pubmed.ncbi.nlm.nih.gov/28973455https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5622323/.

T. Mazzei, E. Mini, A. Novelli, and P. Periti. Chemistry and mode of action of macrolides. *J Antimicrob Chemother*, 31 Suppl C:1–9, 1993. ISSN 0305-7453 (Print) 0305-7453. doi: 10.1093/jac/31.suppl_c.1.

P. R. McAdam, A. Holmes, K. E. Templeton, and J. R. Fitzgerald. Adaptive evolution of *Staphylococcus aureus* during chronic endobronchial infection of a cystic fibrosis patient. *PloS one*, 6(9):e24301, 2011. ISSN 1932-6203.

L. K. McDougal, C. D. Steward, G. E. Killgore, J. M. Chaitram, S. K. McAllister, and F. C. Tenover. Pulsed-field gel electrophoresis typing of oxacillin-resistant *Staphylococcus aureus* isolates from the united states: Establishing a national database. *Journal of Clinical Microbiology*, 41(11):5113–5120, 2003. ISSN 0095-1137. doi: 10.1128/jcm.41.11.5113-5120.2003. URL https://dx.doi.org/10.1128/jcm.41.11.5113-5120.2003.

S. McDougall. A randomised, non-inferiority trial of a new cephalonium dry-cow therapy. *New Zealand Veterinary Journal*, 58(1):45–58, 2010. ISSN 0048-0169. doi: 10.1080/00480169.2010.65060. URL https://doi.org/10.1080/00480169.2010.65060.

S. McDougall, D. Arthur, M. Bryan, J. Vermunt, and A. Weir. Clinical and bacteriological response to treatment of clinical mastitis with one of three intramammary antibiotics. *New Zealand Veterinary Journal*, 55(4):161–170, 2007. ISSN 0048-0169. doi: 10.1080/00480169.2007.36762. URL https://dx.doi.org/10.1080/00480169.2007.36762.

S. McDougall, H. Hussein, and K. Petrovski. Antimicrobial resistance in *Staphylococcus aureus*, *Streptococcus uberis* and *Streptococcus dysgalactiae* from dairy cows with mastitis. *New Zealand Veterinary Journal*, 62(2):68–76, 2014. ISSN 0048-0169. doi: 10.1080/00480169.2013.843135. URL https://doi.org/10.1080/00480169.2013.843135.

S. McDougall, L. Clausen, J. Hintukainen, and J. Hunnam. Randomized, controlled, superiority study of extended duration of therapy with an intramammary antibiotic for treatment of clinical mastitis. *Journal of Dairy Science*, 102(5):4376–4386, 2019. ISSN 0022-0302. doi: 10.3168/jds.2018-15141. URL https://dx.doi.org/10.3168/jds.2018-15141.

S. McDougall, J. Penry, and D. Dymock. Antimicrobial susceptibilities in dairy herds that differ in dry cow therapy usage. *Journal of Dairy Science*, 2021. ISSN 0022-0302. doi: https://doi.org/10.3168/jds.2020-19925. URL https://www.sciencedirect.com/science/article/pii/S0022030221005981.

C. R. E. McEvoy, B. Tsuji, W. Gao, T. Seemann, J. L. Porter, K. Doig, D. Ngo, B. P. Howden, and T. P. Stinear. Decreased vancomycin susceptibility in *Staphylococcus aureus* caused by tempering of walkr expression. *Antimicrobial Agents and Chemotherapy*, 57(7):3240–3249, 2013. doi: 10.1128/aac.00279-13. URL https://aac.asm.org/content/aac/57/7/3240.full.pdf.

P. T. McNamee, J. A. Smyth, and J. A. Smyth. Bacterial chondronecrosis with osteomyelitis ('femoral head necrosis') of broiler chickens: A review. *Avian Pathology*, 29(4):253–270, 2000. ISSN 0307-9457. doi: 10.1080/03079450050118386. URL https://dx.doi.org/10.1080/03079450050118386.

D. Medini, C. Donati, H. Tettelin, V. Masignani, and R. Rappuoli. The microbial pan-genome. *Current Opinion in Genetics & Development*, 15(6):589–594, 2005. ISSN 0959-437X. doi: https://doi.org/10.1016/j.gde.2005.09.006. URL https://www.sciencedirect.com/science/article/pii/S0959437X05001759.

E. N. Meeusen, J. Walker, A. Peters, P. P. Pastoret, and G. Jungersen. Current status of veterinary vaccines. *Clin Microbiol Rev*, 20(3):489–510, table of contents, 2007. ISSN 0893-8512 (Print) 0893-8512 (Linking). doi: 10.1128/CMR.00005-07. URL https://www.ncbi.nlm.nih.gov/pubmed/17630337.

P. L. Mehndiratta and P. Bhalla. Use of antibiotics in animal agriculture & emergence of methicillin-resistant *Staphylococcus aureus* (mrsa) clones: need to assess the impact on public health. *The Indian journal of medical research*, 140(3):339–344, 2014. ISSN 0971-5916 0975-9174. URL https://pubmed.ncbi.nlm.nih.gov/25366200https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4248379/.

M. B. Melchior, H. Vaarkamp, and J. Fink-Gremmels. Biofilms: A role in recurrent mastitis infections? *The Veterinary Journal*, 171(3):398–407, 2006. ISSN 1090-0233. doi: https://doi.org/10.1016/j.tvjl.2005.01.006. URL https://www.sciencedirect.com/science/article/pii/S1090023305000316.

A. Mellmann, T. Weniger, C. Berssenbrügge, J. Rothgänger, M. Sammeth, J. Stoye, and D. Harmsen. Based upon repeat pattern (burp): an algorithm to characterize the long-term evolution of *Staphylococcus aureus* populations based on spa polymorphisms. *BMC Microbiology*, 7(1):98, 2007. ISSN 1471-2180. doi: 10.1186/1471-2180-7-98. URL https://dx.doi.org/10.1186/1471-2180-7-98.

O. Melter and B. Radojevič. Small colony variants of *Staphylococcus aureus*–review. *Folia Microbiol (Praha)*, 55(6):548–58, 2010. ISSN 0015-5632. doi: 10.1007/s12223-010-0089-3.

G. Méric, M. Miragaia, M. De Been, K. Yahara, B. Pascoe, L. Mageiros, J. Mikhail, L. G. Harris, T. S. Wilkinson, J. Rolo, S. Lamble, J. E. Bray, K. A. Jolley, W. P. Hanage, R. Bowden, M. C. J. Maiden, D. Mack, H. De Lencastre, E. J. Feil, J. Corander, and S. K. Sheppard. Ecological overlap and horizontal gene transfer in *Staphylococcus aureus* and staphylococcus epidermidis. *Genome Biology and Evolution*, 7(5):1313–1328, 2015. ISSN 1759-6653. doi: 10.1093/gbe/evv066. URL https://dx.doi.org/10.1093/gbe/evv066.

J. R. Middleton. *Staphylococcus aureus* antigens and challenges in vaccine development. *Expert Review of Vaccines*, 7(6):805–815, 2008. ISSN 1476-0584. doi: 10.1586/14760584.7.6.805. URL https://doi.org/10.1586/14760584.7.6.805.

J. R. Middleton and L. K. Fox. Influence of *Staphylococcus aureus* strain on mammary quarter milk production. *Vet Rec*, 150(13):411–3, 2002. ISSN 0042-4900 (Print) 0042-4900. doi: 10.1136/vr.150.13.411.

J. R. Middleton, J. Ma, C. L. Rinehart, V. N. Taylor, C. D. Luby, and B. J. Steevens. Efficacy of different lysigin formulations in the prevention of *Staphylococcus aureus* intramammary infection in dairy heifers. *J Dairy Res*, 73(1):10–9, 2006. ISSN 0022-0299 (Print) 0022-0299. doi: 10.1017/s0022029905001354.

J. R. Middleton, C. D. Luby, and D. S. Adams. Efficacy of vaccination against staphylococcal mastitis: a review and new data. *Vet Microbiology*, 134(1-2):192–8, 2009. ISSN 0378-1135 (Print) 0378-1135 (Linking). doi: 10.1016/j.vetmic.2008.09.053. URL https://www.ncbi.nlm.nih.gov/pubmed/19010613.

J. R. Middleton, L. K. Fox, G. Pighetti, and C. Petersson-Wolfe. Laboratory handbook on bovine mastitis, 3rd ed. *National Mastitis Council Inc*, 2017.

A. Mikheenko, A. Prjibelski, V. Saveliev, D. Antipov, and A. Gurevich. Versatile genome assembly evaluation with quast-lg. *Bioinformatics*, 34(13):i142–i150, 2018. ISSN 1367-4803. doi: 10.1093/bioinformatics/bty266. URL https://dx.doi.org/10.1093/bioinformatics/bty266.

M. L. Miller, B. Soufi, C. Jers, N. Blom, B. Macek, and I. Mijakovic. Netphosbac–a predictor for ser/thr phosphorylation sites in bacterial proteins. *Proteomics*, 9(1): 116–125, 2009.

N. Misra, T. F. Wines, C. L. Knopp, M. A. McGuire, and J. K. Tinker. Expression, immunogenicity and variation of iron-regulated surface protein a from bovine isolates of *Staphylococcus aureus*. *FEMS microbiology letters*, 364(9):fnx082, 2017. ISSN 1574-6968 0378-1097. doi: 10.1093/femsle/fnx082. URL https://pubmed.ncbi.nlm.nih.gov/28430959https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5430615/.

N. Misra, X. Pu, D. N. Holt, M. A. McGuire, and J. K. Tinker. Immunoproteomics to identify *Staphylococcus aureus* antigens expressed in bovine milk during mastitis. *Journal of dairy science*, 101(7):6296–6309, 2018. ISSN 1525-3198 0022-0302. doi: 10.3168/jds.2017-14040. URL https://pubmed.ncbi.nlm.nih.gov/29729920https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6610883/.

S. Monecke, P. Kuhnert, H. Hotzel, P. Slickers, and R. Ehricht. Microarray based study on virulence-associated genes and resistance determinants of *Staphylococcus*

aureus isolates from cattle. *Veterinary Microbiology*, 125(1):128–140, 2007. ISSN 0378-1135. doi: https://doi.org/10.1016/j.vetmic.2007.05.016. URL https://www.sciencedirect.com/science/article/pii/S0378113507002623.

S. Monecke, P. Slickers, and R. Ehricht. Assignment of *Staphylococcus aureus* isolates to clonal complexes based on microarray analysis and pattern recognition. *FEMS Immunology & Medical Microbiology*, 53(2):237–251, 2008. ISSN 0928-8244. doi: https://doi.org/10.1111/j.1574-695X.2008.00426.x. URL https://doi.org/10.1111/j.1574-695X.2008.00426.x.

Y. Mora-Hernández, E. Vera Murguía, J. Stinenbosch, P. Hernández Jauregui, J. M. van Dijl, and G. Buist. Molecular typing and antimicrobial resistance profiling of 33 mastitis-related *Staphylococcus aureus* isolates from cows in the comarca lagunera region of mexico. *Scientific Reports*, 11(1):6912, 2021. ISSN 2045-2322. doi: 10.1038/s41598-021-86453-2. URL https://doi.org/10.1038/s41598-021-86453-2.

E. Morfeldt, D. Taylor, A. von Gabain, and S. Arvidson. Activation of alpha-toxin translation in *Staphylococcus aureus* by the trans-encoded antisense rna, rnaiii. *The EMBO journal*, 14(18):4569–4577, 1995. ISSN 0261-4189 1460-2075. URL https://www.ncbi.nlm.nih.gov/pubmed/7556100https://www.ncbi.nlm.nih.gov/pmc/articles/PMC394549/.

T. Mork, H. J. Jorgensen, M. Sunde, B. Kvitle, S. Sviland, S. Waage, and T. Tollersrud. Persistence of staphylococcal species and genotypes in the bovine udder. *Vet Microbiology*, 159(1-2):171–80, 2012. ISSN 1873-2542 (Electronic) 0378-1135 (Linking). doi: 10.1016/j.vetmic.2012.03.034. URL https://www.ncbi.nlm.nih.gov/pubmed/22503603.

MPI. 2011-2014 antibiotic sales analysis. Report, Ministry for Primary Industries (MPI), 2016. URL https://www.mpi.govt.nz/dmsdocument/14497-2011-2014-antibiotic-sales-analysis.

MPI. Analysis of risk pathways for the introduction of *Mycoplasma bovis* into new zealand, 2017. URL https://www.mpi.govt.nz/dmsdocument/28050-Pathways-Report-Redacted.pdf.

MPI. Antibiotic sales analysis 2017. Report, Ministry for Primary Industries (MPI), 2019. URL https://www.mpi.govt.nz/dmsdocument/37886/direct.

D. M. Mrochen, L. M. Fernandes de Oliveira, D. Raafat, and S. Holtfreter. *Staphylococcus aureus* host tropism and its implications for murine infection models. *International journal of molecular sciences*, 21(19):7061, 2020. ISSN 1422-0067. doi: 10.3390/ijms21197061. URL https://pubmed.ncbi.nlm.nih.gov/32992784https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7582387/.

M. P. Murphy, D. A. Niedziela, F. C. Leonard, and O. M. Keane. The *in vitro* host cell immune response to bovine-adapted *Staphylococcus aureus* varies according to bacterial lineage. *Scientific Reports*, 9(1):6134, 2019. ISSN 2045-2322. doi: 10.1038/s41598-019-42424-2. URL https://doi.org/10.1038/s41598-019-42424-2.

D. L. Murray, G. S. Prasad, C. A. Earhart, B. A. Leonard, B. N. Kreiswirth, R. P. Novick, D. H. Ohlendorf, and P. M. Schlievert. Immunobiologic and biochemical properties of mutants of toxic shock syndrome toxin-1. *The Journal of Immunology*, 152(1):87–95, 1994. URL https://www.jimmunol.org/content/jimmunol/152/1/87.full.pdf.

S. Murray, B. Pascoe, G. Meric, L. Mageiros, K. Yahara, M. D. Hitchings, Y. Friedmann, T. S. Wilkinson, F. J. Gormley, D. Mack, et al. Recombination-mediated host adaptation by avian staphylococcus aureus. *Genome biology and evolution*, 9(4):830–842, 2017.

T. Nakamura, K. D. Yamada, K. Tomii, and K. Katoh. Parallelization of mafft for large-scale multiple sequence alignments. *Bioinformatics*, 34(14):2490–2492, 2018. ISSN 1367-4803. doi: 10.1093/bioinformatics/bty121. URL https://doi.org/10.1093/bioinformatics/bty121.

A. A. Neyfakh, C. M. Borsch, and G. W. Kaatz. Fluoroquinolone resistance protein nora of *Staphylococcus aureus* is a multidrug efflux transporter. *Antimicrob Agents Chemother*, 37(1):128–9, 1993. ISSN 0066-4804 (Print) 0066-4804. doi: 10.1128/aac.37.1.128.

S. Nickerson, W. Owens, G. Tomita, and P. Widel. Vaccinating dairy heifers with a *Staphylococcus aureus* bacterin reduces mastitis at calving. *Large Animal Practice*, 1999. ISSN 1092-7603.

H. Nielsen. Predicting secretory proteins with signalp. In *Protein function prediction*, pages 59–73. Springer, 2017.

J. B. Nielsen, M. P. Caulfield, and J. O. Lampen. Lipoprotein nature of bacillus licheniformis membrane penicillinase. *Proc Natl Acad Sci U S A*, 78(6):3511–5, 1981. ISSN 0027-8424 (Print) 0027-8424. doi: 10.1073/pnas.78.6.3511.

W. Noble. Co-transfer of vancomycin and other resistance genes from enterococcus faecalis nctc 12201 to *Staphylococcus aureus*. *FEMS Microbiology Letters*, 93(2): 195–198, 1992. ISSN 0378-1097. doi: 10.1016/0378-1097(92)90528-v. URL https://dx.doi.org/10.1016/0378-1097(92)90528-v.

M. Noirclerc-Savoye, A. Le Gouëllec, C. Morlot, O. Dideberg, T. Vernet, and A. Zapun. *in vitro* reconstitution of a trimeric complex of divib, divic and ftsl, and their transient co-localization at the division site in streptococcus pneumoniae. *Molecular Microbiology*, 55(2):413–424, 2005. ISSN 0950-382X. doi: https://doi.org/10.1111/j.1365-2958.2004.04408.x. URL https://doi.org/10.1111/j.1365-2958.2004.04408.x.

S. Notcovich, G. DeNicolo, S. H. Flint, N. B. Williamson, K. Gedye, A. Grinberg, and N. Lopez-Villalobos. Biofilm-forming potential of *Staphylococcus aureus* isolated from clinical mastitis cases in new zealand. *Veterinary sciences*, 5(1):8, 2018. ISSN 2306-7381. doi: 10.3390/vetsci5010008. URL https://pubmed.ncbi.nlm.nih.gov/29351199https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5876566/.

R. P. Novick. Autoinduction and signal transduction in the regulation of staphylococcal virulence. *Molecular Microbiology*, 48(6):1429–1449, 2003. ISSN 0950-382X. doi: 10.1046/j.1365-2958.2003.03526.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1046/j.1365-2958.2003.03526.x.

R. P. Novick and E. Geisinger. Quorum sensing in staphylococci. *Annual Review of Genetics*, 42(1):541–564, 2008. ISSN 0066-4197. doi: 10.1146/annurev.genet.42.110807.091640. URL https://dx.doi.org/10.1146/annurev.genet.42.110807.091640.

D. Nurjadi, A. O. Olalekan, F. Layer, A. O. Shittu, A. Alabi, B. Ghebremedhin, F. Schaumburg, J. Hofmann-Eifler, P. J. J. Van Genderen, E. Caumes, R. Fleck, F. P. Mockenhaupt, M. Herrmann, W. V. Kern, S. Abdulla, M. P. Grobusch, P. G.

Kremsner, C. Wolz, and P. Zanger. Emergence of trimethoprim resistance gene dfrg in *Staphylococcus aureus* causing human infection and colonization in sub-saharan africa and its import to europe. *Journal of Antimicrobial Chemotherapy*, 69(9):2361–2368, 2014. ISSN 0305-7453. doi: 10.1093/jac/dku174. URL https://doi.org/10.1093/jac/dku174.

S. Nurk, A. Bankevich, D. Antipov, A. Gurevich, A. Korobeynikov, A. Lapidus, A. Prjibelsky, A. Pyshkin, A. Sirotkin, Y. Sirotkin, R. Stepanauskas, J. McLean, R. Lasken, S. R. Clingenpeel, T. Woyke, G. Tesler, M. A. Alekseyev, and P. A. Pevzner. Assembling genomes and mini-metagenomes from highly chimeric reads. *Journal of computational biology*, pages 158–170, 2013.

A. K. Nyman, K. Persson Waller, U. Emanuelson, and J. Frössling. Sensitivity and specificity of pcr analysis and bacteriological culture of milk samples for identification of intramammary infections in dairy cows using latent class analysis. *Prev Vet Med*, 135:123–131, 2016. ISSN 0167-5877. doi: 10.1016/j.prevetmed.2016.11.009.

NZHerald. Dairy industry ups the war against mastitis. *NZ Herald*, Jun 2012. URL https://www.nzherald.co.nz/business/dairy-industry-ups-the-war-against-mastitis/AUPUKYYTYZE34YH74OYR6G5X24/.

NZVA. Antimicrobial resistance a significant global challenge for animal and human health, Nov 2019. URL https://www.nzva.org.nz/news/amr-challenge/.

M. M. Obaidat, A. E. Bani Salman, and A. A. Roess. High prevalence and antimicrobial resistance of meca *Staphylococcus aureus* in dairy cattle, sheep, and goat bulk tank milk in jordan. *Tropical Animal Health and Production*, 50(2): 405–412, 2018. ISSN 1573-7438. doi: 10.1007/s11250-017-1449-7. URL https://doi.org/10.1007/s11250-017-1449-7.

F. G. O'Brien, C. Price, W. B. Grubb, and J. E. Gustafson. Genetic characterization of the fusidic acid and cadmium resistance determinants of *Staphylococcus aureus* plasmid pub101. *Journal of Antimicrobial Chemotherapy*, 50(3):313–321, 2002. ISSN 0305-7453. doi: 10.1093/jac/dkf153. URL https://doi.org/10.1093/jac/dkf153.

F. P. O'Hara, J. A. Suaya, G. T. Ray, R. Baxter, M. L. Brown, R. M. Mera, N. M. Close, E. Thomas, and H. Amrine-Madsen. spa typing and multilocus sequence typing show comparable performance in a macroepidemiologic study of *Staphylococcus aureus* in the united states. *Microbial Drug Resistance*, 22(1): 88–96, 2016. ISSN 1076-6294. doi: 10.1089/mdr.2014.0238. URL https://dx.doi.org/10.1089/mdr.2014.0238.

R. Olaniyi, C. Pozzi, L. Grimaldi, and F. Bagnoli. Staphylococcus aureus-*Associated Skin and Soft Tissue Infections: Anatomical Localization, Epidemiology, Therapy and Potential Prophylaxis*, pages 199–227. Springer International Publishing, 2016. ISBN 0070-217X. doi: 10.1007/82_2016_32. URL https://dx.doi.org/10.1007/82_2016_32.

N. A. O'Leary, M. W. Wright, J. R. Brister, S. Ciufo, D. Haddad, R. McVeigh, B. Rajput, B. Robbertse, B. Smith-White, D. Ako-Adjei, et al. Reference sequence (refseq) database at ncbi: current status, taxonomic expansion, and functional annotation. *Nucleic acids research*, 44(D1):D733–D745, 2016.

D. Oliveira, A. Borges, and M. Simões. *Staphylococcus aureus* toxins and their molecular activity in infectious diseases. *Toxins*, 10(6):252, 2018. ISSN 2072-6651. doi: 10.3390/toxins10060252. URL https://pubmed.ncbi.nlm.nih.gov/29921792https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6024779/.

L. Oliveira, H. Langoni, C. Hulland, and P. L. Ruegg. Minimum inhibitory concentrations of *Staphylococcus aureus* recovered from clinical and subclinical cases of bovine mastitis. *Journal of Dairy Science*, 95(4):1913–1920, 2012. ISSN 0022-0302. doi: https://doi.org/10.3168/jds.2011-4938. URL https://www.sciencedirect.com/science/article/pii/S0022030212001555.

S. P. Oliver and C. National Mastitis. *Microbiological procedures for the diagnosis of bovine udder infection and determination of milk quality*. Verona, WI : NMC, 2004.

K. Omoe, D.-L. Hu, H. Takahashi-Omoe, A. Nakane, and K. Shinagawa. Comprehensive analysis of classical and newly described staphylococcal superantigenic toxin genes in *Staphylococcus aureus* isolates. *FEMS microbiology letters*, 246(2):191–198, 2005. ISSN 1574-6968.

K. Omoe, D.-L. Hu, H. K. Ono, S. Shimizu, H. Takahashi-Omoe, A. Nakane, T. Uchiyama, K. Shinagawa, and K. Imanishi. Emetic potentials of newly identified staphylococcal enterotoxin-like toxins. *Infection and immunity*, 81(10): 3627–3631, 2013. ISSN 1098-5522 0019-9567. doi: 10.1128/IAI.00550-13. URL https://pubmed.ncbi.nlm.nih.gov/23876808https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3811759/.

A. J. O'Neill and I. Chopra. Molecular basis of fusb-mediated resistance to fusidic acid in *Staphylococcus aureus*. *Molecular Microbiology*, 59(2):664–676, 2006. ISSN 0950-382X. doi: 10.1111/j.1365-2958.2005.04971.x. URL https://doi.org/10.1111/j.1365-2958.2005.04971.x.

A. J. O'Neill, F. McLaws, G. Kahlmeter, A. S. Henriksen, and I. Chopra. Genetic basis of resistance to fusidic acid in staphylococci. *Antimicrobial Agents and Chemotherapy*, 51(5):1737–1740, 2007. doi: 10.1128/aac.01542-06. URL https://aac.asm.org/content/aac/51/5/1737.full.pdf.

E. Ong, H. Wang, M. U. Wong, M. Seetharaman, N. Valdez, and Y. He. Vaxign-ml: supervised machine learning reverse vaccinology model for improved prediction of bacterial protective antigens. *Bioinformatics*, 36(10):3185–3191, 2020. ISSN 1367-4803. doi: 10.1093/bioinformatics/btaa119. URL https://doi.org/10.1093/bioinformatics/btaa119.

E. Ong, M. F. Cooke, A. Huffman, Z. Xiang, M. U. Wong, H. Wang, M. Seetharaman, N. Valdez, and Y. He. Vaxign2: the second generation of the first web-based vaccine design program using reverse vaccinology and machine learning. *Nucleic Acids Research*, 2021. ISSN 0305-1048. doi: 10.1093/nar/gkab279. URL https://doi.org/10.1093/nar/gkab279.

J. O'Shea, E. O'Callaghan, and W. J. Meaney. Effect of machine milking on new mastitis infections. *Irish Journal of Agricultural Research*, 23(2/3):155–171, 1984. ISSN 05787483. URL www.jstor.org/stable/25556088.

L. H. Otero, A. Rojas-Altuve, L. I. Llarrull, C. Carrasco-Lopez, M. Kumarasiri, E. Lastochkin, J. Fishovitz, M. Dawley, D. Hesek, M. Lee, J. W. Johnson, J. F. Fisher, M. Chang, S. Mobashery, and J. A. Hermoso. How allosteric control of *Staphylococcus aureus* penicillin binding protein 2a enables methicillin resistance and physiological function. *Proceedings of the National Academy of Sciences*, 110

(42):16808–16813, 2013. ISSN 0027-8424. doi: 10.1073/pnas.1300118110. URL https://dx.doi.org/10.1073/pnas.1300118110.

M. Otto. Staphylococcal infections: mechanisms of biofilm maturation and detachment as critical determinants of pathogenicity. *Annu Rev Med*, 64:175–88, 2013. ISSN 1545-326X (Electronic) 0066-4219 (Linking). doi: 10.1146/annurev-med-042711-140023. URL https://www.ncbi.nlm.nih.gov/pubmed/22906361.

W. E. Owens. Isolation of *Staphylococcus aureus* l forms from experimentally induced bovine mastitis. *Journal of clinical microbiology*, 25(10):1956–1961, 1987. ISSN 0095-1137 1098-660X. doi: 10.1128/JCM.25.10.1956-1961.1987. URL https://pubmed.ncbi.nlm.nih.gov/3667916https://www.ncbi.nlm.nih.gov/pmc/articles/PMC269375/.

W. E. Owens, J. L. Watts, R. L. Boddie, and S. C. Nickerson. Antibiotic treatment of mastitis: comparison of intramammary and intramammary plus intramuscular therapies. *Journal of Dairy Science*, 71(11):3143–7, 1988. ISSN 0022-0302 (Print) 0022-0302. doi: 10.3168/jds.S0022-0302(88)79915-4.

Oxoid. 28912 oxoid fda cartridge tables:1, 2011. URL http://www.oxoid.com/pdf/uk/2011-CLSI&FDA-table-update.pdf.

M. O'Dea, R. J. Abraham, S. Sahibzada, T. Lee, D. Jordan, T. Laird, S. Pang, N. Buller, M. Stegger, G. W. Coombs, D. J. Trott, and S. Abraham. Antimicrobial resistance and genomic insights into bovine mastitis-associated *Staphylococcus aureus* in australia. *Veterinary Microbiology*, 250:108850, 2020. ISSN 0378-1135. doi: https://doi.org/10.1016/j.vetmic.2020.108850. URL https://www.sciencedirect.com/science/article/pii/S0378113520309883.

A. J. Page, C. A. Cummins, M. Hunt, V. K. Wong, S. Reuter, M. T. G. Holden, M. Fookes, D. Falush, J. A. Keane, and J. Parkhill. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics*, 31(22):3691–3693, 2015. ISSN 1367-4803. doi: 10.1093/bioinformatics/btv421. URL https://dx.doi.org/10.1093/bioinformatics/btv421.

A. M. Palazzolo-Ballance, M. L. Reniere, K. R. Braughton, D. E. Sturdevant, M. Otto, B. N. Kreiswirth, E. P. Skaar, and F. R. DeLeo. Neutrophil microbicides induce a pathogen survival response in community-associated

methicillin-resistant*Staphylococcus aureus*. *The Journal of Immunology*, 180(1): 500, 2008. doi: 10.4049/jimmunol.180.1.500. URL http://www.jimmunol.org/content/180/1/500.abstract.

J. W. Pankey, N. T. Boddie, J. L. Watts, and S. C. Nickerson. Evaluation of protein a and a commercial bacterin1 as vaccines against *Staphylococcus aureus* mastitis by experimental challenge. *Journal of Dairy Science*, 68(3):726–731, 1985. ISSN 0022-0302. doi: https://doi.org/10.3168/jds.S0022-0302(85)80879-1. URL http://www.sciencedirect.com/science/article/pii/S0022030285808791.

D. Papadopoulou and N. V. Sipsas. Comparison of national clinical practice guidelines and recommendations on vaccination of adult patients with autoimmune rheumatic diseases. *Rheumatol Int*, 34(2):151–63, 2014. ISSN 0172-8172. doi: 10.1007/s00296-013-2907-9.

E. P. Papagiannitsis and Constantinos. Resistance of staphylococci to macrolides-lincosamides- streptogramins b (mlsb): Epidemiology and mechanisms of resistance. *IntechOpen*, 2018. doi: DOI:10.5772/intechopen.75192.

M. G. Papich, editor. *Florfenicol*, pages 327–329. W.B. Saunders, St. Louis, fourth edition edition, 2016. ISBN 978-0-323-24485-5. doi: https://doi.org/10.1016/B978-0-323-24485-5.00264-3. URL https://www.sciencedirect.com/science/article/pii/B9780323244855002643.

N. Pardi, M. J. Hogan, F. W. Porter, and D. Weissman. mrna vaccines — a new era in vaccinology. *Nature Reviews Drug Discovery*, 17(4):261–279, 2018. ISSN 1474-1784. doi: 10.1038/nrd.2017.243. URL https://doi.org/10.1038/nrd.2017.243.

M. A. Parenti, S. M. Hatfield, and J. J. Leyden. Mupirocin: a topical antibiotic with a unique structure and mechanism of action. *Clin Pharm*, 6(10):761–70, 1987. ISSN 0278-2677 (Print) 0278-2677.

A. K. Park, H. Kim, and H. J. Jin. Phylogenetic analysis of rrna methyltransferases, erm and ksga, as related to antibiotic resistance. *FEMS Microbiology Letters*, pages no–no, 2010. ISSN 0378-1097. doi: 10.1111/j.1574-6968.2010.02031.x. URL https://dx.doi.org/10.1111/j.1574-6968.2010.02031.x.

L. Pasteur. The attenuation of the causal agent of fowl cholera. *CR Acad Sci*, 91: 673–680, 1880.

J. M. Patti, B. L. Allen, M. J. McGavin, and M. Höök. Mscramm-mediated adherence of microorganisms to host tissues. *Annu Rev Microbiol*, 48:585–617, 1994. ISSN 0066-4227 (Print) 0066-4227. doi: 10.1146/annurev.mi.48.100194.003101.

R. Peakall and P. E. Smouse. Genalex 6.5: genetic analysis in excel. population genetic software for teaching and research—an update. *Bioinformatics*, 28(19): 2537–2539, 2012. ISSN 1367-4803. doi: 10.1093/bioinformatics/bts460. URL https://doi.org/10.1093/bioinformatics/bts460.

O. Pérez, A. Batista-Duharte, E. González, C. Zayas, J. Balboa, M. Cuello, O. Cabrera, M. Lastre, and V. E. Schijns. Human prophylactic vaccine adjuvants and their determinant role in new vaccine formulations. *Brazilian Journal of Medical and Biological Research*, 45(8):681–692, 2012.

V. K. Pérez, D. A. Custódio, E. M. Silva, J. de Oliveira, A. S. Guimarães, M. A. Brito, A. F. Souza-Filho, M. B. Heinemann, A. P. Lage, and E. M. Dorneles. Virulence factors and antimicrobial resistance in staphylococcus aureus isolated from bovine mastitis in brazil. *Brazilian Journal of Microbiology*, 51(4):2111–2122, 2020.

V. K. C. Perez, G. M. d. Costa, A. S. Guimarães, M. B. Heinemann, A. P. Lage, and E. M. S. Dorneles. Relationship between virulence factors and antimicrobial resistance in *Staphylococcus aureus* from bovine mastitis. *Journal of Global Antimicrobial Resistance*, 22:792–802, 2020. ISSN 2213-7165. doi: https://doi.org/10.1016/j.jgar.2020.06.010. URL http://www.sciencedirect.com/science/article/pii/S2213716520301557.

C. S. Petersson-Wolfe, I. K. Mullarky, and G. M. Jones. Staphylococcus aureus *mastitis: cause, detection, and control*. Virginia Cooperative Extension, 2010. URL https://vtechworks.lib.vt.edu/bitstream/handle/10919/48390/404-229_pdf.pdf?sequence=1&isAllowed=y.

V. Peton and Y. Le Loir. *Staphylococcus aureus* in veterinary medicine. *Infect Genet Evol*, 21:602–15, 2014. ISSN 1567-7257 (Electronic) 1567-1348 (Linking). doi: 10.1016/j.meegid.2013.08.011. URL https://www.ncbi.nlm.nih.gov/pubmed/23974078.

K. R. Petrovski. *Bovine mastitis in New Zealand*. Thesis, IVABS, 2007. URL http://ezproxy.massey.ac.nz/login?url=http://search.ebscohost.

com/login.aspx?direct=true$&$db=cat00245a$&$AN=massey.b1976658$&$site=eds-live$&$scope=sitehttp://hdl.handle.net/10179/7103.

K. R. Petrovski, C. Heuer, T. J. Parkinson, and N. B. Williamson. The incidence and aetiology of clinical bovine mastitis on 14 farms in northland, new zealand. *N Z Vet J*, 57(2):109–15, 2009. ISSN 0048-0169 (Print) 0048-0169 (Linking). doi: 10.1080/00480169.2009.36887. URL https://www.ncbi.nlm.nih.gov/pubmed/19471330.

K. R. Petrovski, R. A. Laven, and N. Lopez-Villalobos. A descriptive analysis of the antimicrobial susceptibility of mastitis-causing bacteria isolated from samples submitted to commercial diagnostic laboratories in new zealand (2003–2006). *New Zealand Veterinary Journal*, 59(2):59–66, 2011. ISSN 0048-0169. doi: 10.1080/00480169.2011.552853. URL https://doi.org/10.1080/00480169.2011.552853.

K. R. Petrovski, A. Grinberg, N. B. Williamson, M. E. Abdalla, N. Lopez-Villalobos, T. J. Parkinson, I. G. Tucker, and P. Rapnicki. Susceptibility to antimicrobials of mastitis-causing *Staphylococcus aureus*, *Streptococcus uberis* and str. dysgalactiae from new zealand and the usa as assessed by the disk diffusion test. *Aust Vet J*, 93 (7):227–33, 2015. ISSN 0005-0423. doi: 10.1111/avj.12340.

N. Petrovsky. Comparative safety of vaccine adjuvants: A summary of current evidence and future needs. *Drug Safety*, 38(11):1059–1074, 2015. ISSN 1179-1942. doi: 10.1007/s40264-015-0350-4. URL https://doi.org/10.1007/s40264-015-0350-4.

S. Pichette-Jolette, G. Millette, E. Demontier, D. Bran-Barrera, M. Cyrenne, C. Ster, D. Haine, G. Keefe, F. Malouin, and J. P. Roy. Partial prediction of the duration and the clinical status of *Staphylococcus aureus* bovine intramammary infections based on the phenotypic and genotypic analysis of isolates. *Veterinary Microbiology*, 228:188–195, 2019. ISSN 0378-1135. doi: https://doi.org/10.1016/j.vetmic.2018.11.024. URL https://www.sciencedirect.com/science/article/pii/S0378113518311532.

S. Pietrokovski. Searching databases of conserved sequence regions by aligning protein multiple-alignments. *Nucleic acids research*, 24(19):3836–3845, 1996.

M. G. Pinho, H. de Lencastre, and A. Tomasz. An acquired and a native penicillin-binding protein cooperate in building the cell wall of drug-resistant staphylococci. *Proceedings of the National Academy of Sciences*, 98(19): 10886–10891, 2001a. doi: 10.1073/pnas.191260798. URL https://www.pnas.org/content/pnas/98/19/10886.full.pdf.

M. G. Pinho, S. R. Filipe, H. de Lencastre, and A. Tomasz. Complementation of the essential peptidoglycan transpeptidase function of penicillin-binding protein 2 (pbp2) by the drug resistance protein pbp2a in *Staphylococcus aureus*. *Journal of Bacteriology*, 183(22):6525–6531, 2001b. doi: 10.1128/jb.183.22.6525-6531.2001. URL https://jb.asm.org/content/jb/183/22/6525.full.pdf.

M. Pizza, V. Scarlato, V. Masignani, M. M. Giuliani, B. Aricò, M. Comanducci, G. T. Jennings, L. Baldi, E. Bartolini, B. Capecchi, C. L. Galeotti, E. Luzzi, R. Manetti, E. Marchetti, M. Mora, S. Nuti, G. Ratti, L. Santini, S. Savino, M. Scarselli, E. Storni, P. Zuo, M. Broeker, E. Hundt, B. Knapp, E. Blair, T. Mason, H. Tettelin, D. W. Hood, A. C. Jeffries, N. J. Saunders, D. M. Granoff, J. C. Venter, E. R. Moxon, G. Grandi, and R. Rappuoli. Identification of vaccine candidates against serogroup b meningococcus by whole-genome sequencing. *Science*, 287(5459): 1816–1820, 2000. doi: 10.1126/science.287.5459.1816. URL https://science.sciencemag.org/content/sci/287/5459/1816.full.pdf.

P. J. Planet, A. Narechania, L. Chen, B. Mathema, S. Boundy, G. Archer, and B. Kreiswirth. Architecture of a species: Phylogenomics of *Staphylococcus aureus*. *Trends Microbiol*, 25(2):153–166, 2017. ISSN 1878-4380 (Electronic) 0966-842X (Linking). doi: 10.1016/j.tim.2016.09.009. URL https://www.ncbi.nlm.nih.gov/pubmed/27751626.

A. Pournajaf, A. Ardebili, L. Goudarzi, M. Khodabandeh, T. Narimani, and H. Abbaszadeh. Pcr-based identification of methicillin–resistant *Staphylococcus aureus* strains and their antibiotic resistance profiles. *Asian pacific journal of tropical biomedicine*, 4:S293–S297, 2014.

G. Prévost, B. Pottecher, M. Dahlet, M. Bientz, J. M. Mantz, and Y. Piémont. Pulsed field gel electrophoresis as a new epidemiological tool for monitoring methicillin-resistant *Staphylococcus aureus* in an intensive care unit. *Journal of Hospital Infection*, 17(4):255–269, 1991. ISSN 0195-6701. doi:

10.1016/0195-6701(91)90270-i. URL https://dx.doi.org/10.1016/0195-6701(91)90270-I.

J. Price, N. Claire Gordon, D. Crook, M. Llewelyn, and J. Paul. The usefulness of whole genome sequencing in the management of *Staphylococcus aureus* infections. *Clinical Microbiology and Infection*, 19(9):784–789, 2013. ISSN 1198-743X. doi: https://doi.org/10.1111/1469-0691.12109. URL https://www.sciencedirect.com/science/article/pii/S1198743X14631943.

R. A. Proctor, P. van Langevelde, M. Kristjansson, J. N. Maslow, and R. D. Arbeit. Persistent and relapsing infections associated with small-colony variants of *Staphylococcus aureus*. *Clin Infect Dis*, 20(1):95–102, 1995. ISSN 1058-4838 (Print) 1058-4838. doi: 10.1093/clinids/20.1.95.

R. A. Proctor, C. Von Eiff, B. C. Kahl, K. Becker, P. McNamara, M. Herrmann, and G. Peters. Small colony variants: a pathogenic form of bacteria that facilitates persistent and recurrent infections. *Nature Reviews Microbiology*, 4(4):295–305, 2006.

S. J. Projan, S. Moghazeh, and R. P. Novick. Nucleotide sequence of ps194, a streptomycin-resistance plasmid from *Staphylococcus aureus*. *Nucleic Acids Res*, 16(5):2179–87, 1988. ISSN 0305-1048 (Print) 0305-1048. doi: 10.1093/nar/16.5.2179.

D. Raafat, M. Otto, K. Reppschläger, J. Iqbal, and S. Holtfreter. Fighting *Staphylococcus aureus* biofilms with monoclonal antibodies. *Trends in microbiology*, 27(4):303–322, 2019. ISSN 1878-4380 0966-842X. doi: 10.1016/j.tim.2018.12.009. URL https://pubmed.ncbi.nlm.nih.gov/30665698https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6420399/.

H. Rahimi, H. Dastmalchi Saei, and M. Ahmadi. Nasal carriage of *Staphylococcus aureus*: Frequency and antibiotic resistance in healthy ruminants. *Jundishapur Journal of Microbiology*, 8(10), 2015. ISSN 2008-3645. doi: 10.5812/jjm.22413. URL https://dx.doi.org/10.5812/jjm.22413.

O. Rahman, S. P. Cummings, D. J. Harrington, and I. C. Sutcliffe. Methods for the bioinformatic identification of bacterial lipoproteins encoded in the genomes of

gram-positive bacteria. *World Journal of Microbiology and Biotechnology*, 24(11): 2377–2382, 2008.

P. Rainard, G. Foucras, J. R. Fitzgerald, J. L. Watts, G. Koop, and J. R. Middleton. Knowledge gaps and research priorities in *Staphylococcus aureus* mastitis control. *Transboundary and Emerging Diseases*, 65(S1):149–165, 2018. ISSN 1865-1674. doi: https://doi.org/10.1111/tbed.12698. URL https://doi.org/10.1111/tbed.12698.

V. L. M. Rall, E. S. Miranda, I. G. Castilho, C. H. Camargo, H. Langoni, F. F. Guimarães, J. P. Araújo Júnior, and A. Fernandes Júnior. Diversity of staphylococcus species and prevalence of enterotoxin genes isolated from milk of healthy cows and cows with subclinical mastitis. *Journal of Dairy Science*, 97(2): 829–837, 2014. ISSN 0022-0302. doi: https://doi.org/10.3168/jds.2013-7226. URL https://www.sciencedirect.com/science/article/pii/S0022030213008576.

R. Rappuoli. Reverse vaccinology. *Curr Opin Microbiol*, 3(5):445–50, 2000. ISSN 1369-5274 (Print) 1369-5274. doi: 10.1016/s1369-5274(00)00119-3.

R. J. Reece and A. Maxwell. Dna gyrase: Structure and function. *Critical Reviews in Biochemistry and Molecular Biology*, 26(3-4):335–375, 1991. ISSN 1040-9238. doi: 10.3109/10409239109114072. URL https://doi.org/10.3109/10409239109114072.

L. B. Reller, M. Weinstein, J. H. Jorgensen, and M. J. Ferraro. Antimicrobial susceptibility testing: A review of general principles and contemporary practices. *Clinical Infectious Diseases*, 49(11):1749–1755, 2009. ISSN 1058-4838. doi: 10.1086/647952. URL https://doi.org/10.1086/647952.

Q. Ren, G. Liao, Z. Wu, J. Lv, and W. Chen. Prevalence and characterization of *Staphylococcus aureus* isolates from subclinical bovine mastitis in southern xinjiang, china. *Journal of Dairy Science*, 103(4):3368–3380, 2020. ISSN 0022-0302. doi: https://doi.org/10.3168/jds.2019-17420. URL https://www.sciencedirect.com/science/article/pii/S0022030220300709.

T. Reyes-Robles, r. Alonzo, F., L. Kozhaya, D. B. Lacy, D. Unutmaz, and V. J. Torres. *Staphylococcus aureus* leukotoxin ed targets the chemokine receptors cxcr1 and

cxcr2 to kill leukocytes and promote infection. *Cell Host Microbe*, 14(4):453–9, 2013. ISSN 1931-3128 (Print) 1931-3128. doi: 10.1016/j.chom.2013.09.005.

T. Reyes-Robles, A. Lubkin, r. Alonzo, F., D. B. Lacy, and V. J. Torres. Exploiting dominant-negative toxins to combat *Staphylococcus aureus* pathogenesis. *EMBO Rep*, 17(3):428–40, 2016. ISSN 1469-221X (Print) 1469-221x. doi: 10.15252/embr.201540994.

E. J. Richardson, R. Bacigalupe, E. M. Harrison, L. A. Weinert, S. Lycett, M. Vrieling, K. Robb, P. A. Hoskisson, M. T. G. Holden, E. J. Feil, G. K. Paterson, S. Y. C. Tong, A. Shittu, W. van Wamel, D. M. Aanensen, J. Parkhill, S. J. Peacock, J. Corander, M. Holmes, and J. R. Fitzgerald. Gene exchange drives the ecological success of a multi-host bacterial pathogen. *Nature Ecology & Evolution*, 2(9): 1468–1478, 2018. ISSN 2397-334X. doi: 10.1038/s41559-018-0617-0. URL https://doi.org/10.1038/s41559-018-0617-0.

A. L. Ridler, G. Rout-Brown, K. J. Flay, N. Velathanthiri, and A. Grinberg. Defects and bacterial pathogens in udders of non-dairy breed ewes from new zealand. *New Zealand Journal of Agricultural Research*, pages 1–9, 2021. ISSN 0028-8233. doi: 10.1080/00288233.2021.1905005. URL https://doi.org/10.1080/00288233.2021.1905005.

R. B. Rindsig, R. G. Rodewald, A. R. Smith, and S. L. Spahr. Complete versus selective dry cow therapy for mastitis control1. *Journal of Dairy Science*, 61(10): 1483–1497, 1978. ISSN 0022-0302. doi: https://doi.org/10.3168/jds.S0022-0302(78)83753-9. URL https://www.sciencedirect.com/science/article/pii/S0022030278837539.

M. Rizwan, A. Naz, J. Ahmad, K. Naz, A. Obaid, T. Parveen, M. Ahsan, and A. Ali. Vacsol: a high throughput in silico pipeline to predict potential therapeutic targets in prokaryotic pathogens using subtractive reverse vaccinology. *BMC bioinformatics*, 18(1):1–7, 2017. ISSN 1471-2105.

M. C. Roberts. Tetracycline resistance determinants: mechanisms of action, regulation of expression, genetic mobility, and distribution. *FEMS Microbiology Reviews*, 19(1):1–24, 1996. ISSN 0168-6445. doi: 10.1111/j.1574-6976.1996.tb00251.x. URL https://doi.org/10.1111/j.1574-6976.1996.tb00251.x.

M. C. Roberts, J. Sutcliffe, P. Courvalin, L. B. Jensen, J. Rood, and H. Seppala. Nomenclature for macrolide and macrolide-lincosamide-streptogramin b resistance determinants. *Antimicrobial Agents and Chemotherapy*, 43(12):2823–2830, 1999. doi: 10.1128/aac.43.12.2823. URL https://aac.asm.org/content/aac/43/12/2823.full.pdf.

D. A. Robinson and M. C. Enright. Evolutionary models of the emergence of methicillin-resistant *Staphylococcus aureus*. *Antimicrobial Agents and Chemotherapy*, 47(12):3926–3934, 2003. ISSN 0066-4804. doi: 10.1128/aac.47.12.3926-3934.2003. URL https://dx.doi.org/10.1128/AAC.47.12.3926-3934.2003.

D. A. Robinson, A. B. Monk, J. E. Cooper, E. J. Feil, and M. C. Enright. Evolutionary genetics of the accessory gene regulator (agr) locus in *Staphylococcus aureus*. *Journal of bacteriology*, 187(24):8312–8321, 2005. ISSN 0021-9193 1098-5530. doi: 10.1128/JB.187.24.8312-8321.2005. URL https://pubmed.ncbi.nlm.nih.gov/16321935https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1317016/.

A. Rodloff, T. Bauer, S. Ewig, P. Kujath, and E. Müller. Susceptible, intermediate, and resistant–the intensity of antibiotic action. *Deutsches Ärzteblatt International*, 105(39):657, 2008.

M. J. Rodríguez-Ortega, N. Norais, G. Bensi, S. Liberatori, S. Capo, M. Mora, M. Scarselli, F. Doro, G. Ferrari, and I. Garaguso. Characterization and identification of vaccine candidate proteins through analysis of the group a streptococcus surface proteome. *Nature biotechnology*, 24(2):191–197, 2006. ISSN 1546-1696.

S. H. M. Rooijakkers, M. Ruyken, A. Roos, M. R. Daha, J. S. Presanis, R. B. Sim, W. J. B. Van Wamel, K. P. M. Van Kessel, and J. A. G. Van Strijp. Immune evasion by a staphylococcal complement inhibitor that acts on c3 convertases. *Nature Immunology*, 6(9):920–927, 2005. ISSN 1529-2908. doi: 10.1038/ni1235. URL https://dx.doi.org/10.1038/ni1235.

J. I. Ross, E. A. Eady, J. H. Cove, W. J. Cunliffe, S. Baumberg, and J. C. Wootton. Inducible erythromycin resistance in staphlyococci is encoded by a member of the atp-binding transport super-gene family. *Molecular Microbiology*, 4(7):1207–1214,

1990. ISSN 0950-382X. doi: 10.1111/j.1365-2958.1990.tb00696.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2958.1990.tb00696.x.

S. J. Rowland and K. G. Dyke. Characterization of the staphylococcal beta-lactamase transposon tn552. *Embo j*, 8(9):2761–73, 1989. ISSN 0261-4189 (Print) 0261-4189.

S. J. Rowland and K. G. Dyke. Tn552, a novel transposable element from *Staphylococcus aureus*. *Mol Microbiol*, 4(6):961–75, 1990. ISSN 0950-382X (Print) 0950-382x. doi: 10.1111/j.1365-2958.1990.tb00669.x.

J.-P. Roy and G. Keefe. Systematic review: What is the best antibiotic treatment for *Staphylococcus aureus* intramammary infection of lactating cows in north america? *Veterinary Clinics of North America: Food Animal Practice*, 28(1):39–50, 2012. ISSN 0749-0720. doi: https://doi.org/10.1016/j.cvfa.2011.12.004. URL http://www.sciencedirect.com/science/article/pii/S0749072011000788.

S. Roy, C. Coldren, A. Karunamurthy, N. S. Kip, E. W. Klee, S. E. Lincoln, A. Leon, M. Pullambhatla, R. L. Temple-Smolkin, K. V. Voelkerding, C. Wang, and A. B. Carter. Standards and guidelines for validating next-generation sequencing bioinformatics pipelines: A joint recommendation of the association for molecular pathology and the college of american pathologists. *The Journal of Molecular Diagnostics*, 20(1):4–27, 2018. ISSN 1525-1578. doi: https://doi.org/10.1016/j.jmoldx.2017.11.003. URL https://www.sciencedirect.com/science/article/pii/S1525157817303732.

E. Royster, S. Godden, D. Goulart, A. Dahlke, P. Rapnicki, and J. Timmerman. Evaluation of the minnesota easy culture system ii bi-plate and tri-plate for identification of common mastitis pathogens in milk. *Journal of Dairy Science*, 97 (6):3648–3659, 2014. ISSN 0022-0302. doi: 10.3168/jds.2013-7748. URL https://dx.doi.org/10.3168/jds.2013-7748.

G. Sachdeva, K. Kumar, P. Jain, and S. Ramachandran. Spaan: a software program for prediction of adhesins and adhesin-like proteins using neural networks. *Bioinformatics*, 21(4):483–491, 2005. ISSN 1460-2059.

D. E. Salmon and T. Smith. On a new method of producing immunity from contagious diseases. *Am Vet Rev*, 10:63–69, 1886.

A. J. Salter. Overview. trimethoprim-sulfamethoxazole: An assessment of more than 12 years of use. *Reviews of Infectious Diseases*, 4(2):196–236, 1982. ISSN 1058-4838. doi: 10.1093/clinids/4.2.196. URL https://dx.doi.org/10.1093/clinids/4.2.196.

M. L. Sanchez and R. N. Jones. E test, an antimicrobial susceptibility testing method with broad clinical and epidemiologic application. *Antimicrobic Newsletter*, 8(1): 1–7, 1992.

A. Sandgren, M. Strong, P. Muthukrishnan, B. K. Weiner, G. M. Church, and M. B. Murray. Tuberculosis drug resistance mutation database. *PLoS medicine*, 6(2): e1000002, 2009.

F. Sanger, S. Nicklen, and A. R. Coulson. Dna sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*, 74(12):5463–5467, 1977. ISSN 0027-8424. doi: 10.1073/pnas.74.12.5463. URL https://dx.doi.org/10.1073/pnas.74.12.5463.

N. Saraswathy and P. Ramalingam. *8 - Genome sequence assembly and annotation*, pages 109–121. Woodhead Publishing, 2011. ISBN 978-1-907568-10-7. doi: https://doi.org/10.1533/9781908818058.109. URL http://www.sciencedirect.com/science/article/pii/B9781907568107500080.

C. Sartori, R. Boss, I. Ivanovic, and H. U. Graber. Development of a new real-time quantitative pcr assay for the detection of *Staphylococcus aureus* genotype b in cow milk, targeting the new gene adlb. *Journal of Dairy Science*, 100(10):7834–7845, 2017. ISSN 0022-0302. doi: 10.3168/jds.2017-12820. URL https://dx.doi.org/10.3168/jds.2017-12820.

T. Sasaki, S. Tsubakishita, Y. Tanaka, A. Sakusabe, M. Ohtsuka, S. Hirotaki, T. Kawakami, T. Fukata, and K. Hiramatsu. Multiplex-pcr method for species identification of coagulase-positive staphylococci. *Journal of Clinical Microbiology*, 48(3):765–769, 2010. ISSN 0095-1137. doi: 10.1128/jcm.01232-09. URL https://dx.doi.org/10.1128/JCM.01232-09.

N. A. Saunders and A. Holmes. *Multilocus Sequence Typing (MLST) of* Staphylococcus aureus, pages 71–85. Humana Press, Totowa, NJ, 2007. ISBN 978-1-59745-468-1. doi: 10.1007/978-1-59745-468-1_6. URL https://doi.org/10.1007/978-1-59745-468-1_6.

D.-J. Scheffers and M. G. Pinho. Bacterial cell wall synthesis: New insights from localization studies. *Microbiology and Molecular Biology Reviews*, 69(4):585–607, 2005. doi: 10.1128/mmbr.69.4.585-607.2005. URL https://mmbr.asm.org/content/mmbr/69/4/585.full.pdf.

C. G. M. Scherpenzeel, H. Hogeveen, L. Maas, and T. Lam. Economic optimization of selective dry cow treatment. *Journal of Dairy Science*, 101(2):1530–1539, 2018. ISSN 0022-0302. doi: 10.3168/jds.2017-13076.

J. M. Schifano, R. Edifor, J. D. Sharp, M. Ouyang, A. Konkimalla, R. N. Husson, and N. A. Woychik. Mycobacterial toxin mazf-mt6 inhibits translation through cleavage of 23s rrna at the ribosomal a site. *Proceedings of the National Academy of Sciences*, 110(21):8501–8506, 2013. ISSN 0027-8424. doi: 10.1073/pnas.1222031110. URL https://dx.doi.org/10.1073/pnas.1222031110.

K. Schlotter, R. Ehricht, H. Hotzel, S. Monecke, M. Pfeffer, and K. Donat. Leukocidin genes *lukF-P83* and *lukM* are associated with *Staphylococcus aureus* clonal complexes 151, 479 and 133 isolated from bovine udder infections in thuringia, germany. *Veterinary Research*, 43(1):42, 2012. ISSN 1297-9716. doi: 10.1186/1297-9716-43-42. URL https://doi.org/10.1186/1297-9716-43-42.

R. Schneerson, O. Barrera, A. Sutton, and J. B. Robbins. Preparation, characterization, and immunogenicity of haemophilus influenzae type b polysaccharide-protein conjugates. *J Exp Med*, 152(2):361–76, 1980. ISSN 0022-1007 (Print) 0022-1007. doi: 10.1084/jem.152.2.361.

T. D. Schneider and R. M. Stephens. Sequence logos: a new way to display consensus sequences. *Nucleic acids research*, 18(20):6097–6100, 1990.

A. Schnitt and B.-A. Tenhagen. Risk factors for the occurrence of methicillin-resistant *Staphylococcus aureus* in dairy herds: An update. *Foodborne pathogens and disease*, 17(10):585–596, 2020. ISSN 1556-7125 1535-3141. doi: 10.1089/fpd.2019.2638. URL https://pubmed.ncbi.nlm.nih.gov/31433237https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7549011/.

K. Schroeder, M. Jularic, S. M. Horsburgh, N. Hirschhausen, C. Neumann, A. Bertling, A. Schulte, S. Foster, B. E. Kehrel, G. Peters, and C. Heilmann. Molecular characterization of a novel *Staphylococcus aureus* surface protein (sasc) involved in cell aggregation and biofilm accumulation. *PloS one*, 4(10): e7567–e7567, 2009. ISSN 1932-6203. doi: 10.1371/journal.pone.0007567. URL https://pubmed.ncbi.nlm.nih.gov/19851500https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2761602/.

Y. H. Schukken, V. Bronzo, C. Locatelli, C. Pollera, N. Rota, A. Casula, F. Testa, L. Scaccabarozzi, R. March, D. Zalduendo, R. Guix, and P. Moroni. Efficacy of vaccination on *Staphylococcus aureus* and coagulase-negative staphylococci intramammary infection dynamics in 2 dairy herds. *Journal of Dairy Science*, 97 (8):5250–5264, 2014. ISSN 0022-0302. doi: https://doi.org/10.3168/jds.2014-8008. URL http://www.sciencedirect.com/science/article/pii/S0022030214003907.

D. C. Schwartz and C. R. Cantor. Separation of yeast chromosome-sized dnas by pulsed field gradient gel electrophoresis. *Cell*, 37(1):67–75, 1984. ISSN 0092-8674. doi: https://doi.org/10.1016/0092-8674(84)90301-5. URL http://www.sciencedirect.com/science/article/pii/0092867484903015.

S. Schwarz, A. T. Feßler, T. Hauschild, C. Kehrenberg, and K. Kadlec. Plasmid-mediated resistance to protein biosynthesis inhibitors in staphylococci. *Annals of the New York Academy of Sciences*, 1241(1):82–103, 2011. ISSN 0077-8923.

S. Schwarz, J. Shen, S. Wendlandt, A. T. Feßler, Y. Wang, K. Kadlec, and C.-M. Wu. Plasmid-mediated antimicrobial resistance in staphylococci and other firmicutes. *Microbiology spectrum*, 2(6):2.6. 13, 2014. ISSN 2165-0497.

C. Seah, D. C. Alexander, L. Louie, A. Simor, D. E. Low, J. Longtin, and R. G. Melano. Mupb, a new high-level mupirocin resistance mechanism in *Staphylococcus aureus*. *Antimicrobial agents and chemotherapy*, 56(4):1916–1920, 2012. ISSN 1098-6596 0066-4804. doi: 10.1128/AAC.05325-11. URL https://pubmed.ncbi.nlm.nih.gov/22252810https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3318397/.

T. Seemann. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*, 30(14):

2068–2069, 2014. ISSN 1367-4803. doi: 10.1093/bioinformatics/btu153. URL https://dx.doi.org/10.1093/bioinformatics/btu153.

T. Seemann. barnap. unpublished, 2018a. URL https://github.com/tseemann/barrnap.

T. Seemann. shovill. unpublished, 2018b. URL https://github.com/tseemann/shovill.

T. Seemann. snippy. unpublished, 2018c. URL https://github.com/tseemann/snippy.

T. Seemann. mlst. unpublished, 2020. URL https://github.com/tseemann/mlst.

J.-i. Sekiguchi, P. Tharavichitkul, T. Miyoshi-Akiyama, V. Chupia, T. Fujino, M. Araake, A. Irie, K. Morita, T. Kuratsuji, and T. Kirikae. Cloning and characterization of a novel trimethoprim-resistant dihydrofolate reductase from a nosocomial isolate of *Staphylococcus aureus* cm.s2 (imcj1454). *Antimicrobial Agents and Chemotherapy*, 49(9):3948–3951, 2005. doi: 10.1128/aac.49.9.3948-3951.2005. URL https://aac.asm.org/content/aac/49/9/3948.full.pdf.

D. Serruto, M. J. Bottomley, S. Ram, M. M. Giuliani, and R. Rappuoli. The new multicomponent vaccine against meningococcal serogroup b, 4cmenb: immunological, functional and structural characterization of the antigens. *Vaccine*, 30:B87–B97, 2012.

A. Seykora and B. McDaniel. Udder and teat morphology related to mastitis resistance: a review. *Journal of Dairy Science*, 68(8):2087–2093, 1985. ISSN 0022-0302.

C. G. Shapiro-Shapin. Pearl kendrick, grace eldering, and the pertussis vaccine. *Emerging infectious diseases*, 16(8):1273, 2010.

L. Shaw, E. Golonka, J. Potempa, and S. J. Foster. The role and regulation of the extracellular proteases of *Staphylococcus aureus*. *Microbiology (Reading)*, 150(Pt 1):217–228, 2004. ISSN 1350-0872 (Print) 1350-0872. doi: 10.1099/mic.0.26634-0.

M. A. Shepheard, V. M. Fleming, T. R. Connor, J. Corander, E. J. Feil, C. Fraser, and W. P. Hanage. Historical zoonoses and other changes in host tropism of *Staphylococcus aureus*, identified by phylogenetic analysis of a population dataset. *PLoS One*, 8(5):e62369, 2013. ISSN 1932-6203 (Electronic) 1932-6203 (Linking). doi: 10.1371/journal.pone.0062369. URL https://www.ncbi.nlm.nih.gov/pubmed/23667472.

A. Shimizu, J. Kawano, J. Ozaki, N. Sasaki, S. Kimura, M. Kamada, S. Anzai, H. Saito, and H. Sato. Characteristics of *Staphylococcus aureus* isolated from lesions of horses. *The Journal of veterinary medical science*, 53(4):601–606, 1991. ISSN 0916-7250. doi: 10.1292/jvms.53.601. URL https://dx.doi.org/10.1292/jvms.53.601.

H. R. Shinefield and S. Black. Prevention of *Staphylococcus aureus* infections: advances in vaccine development. *Expert Review of Vaccines*, 4(5):669–676, 2005. ISSN 1476-0584. doi: 10.1586/14760584.4.5.669. URL https://doi.org/10.1586/14760584.4.5.669.

A. O. Shittu, E. E. Udo, and J. Lin. Insights on virulence and antibiotic resistance: a review of the accessory genome of *Staphylococcus aureus*. *Wounds*, 19(9):237, 2007.

L. Shkreta, B. G. Talbot, M. S. Diarra, and P. Lacasse. Immune responses to a dna/protein vaccination strategy against *Staphylococcus aureus* induced mastitis in dairy cows. *Vaccine*, 23(1):114–126, 2004. ISSN 0264-410X. doi: https://doi.org/10.1016/j.vaccine.2004.05.002. URL http://www.sciencedirect.com/science/article/pii/S0264410X04003688.

B. Shopsin, M. Gomez, S. O. Montgomery, D. H. Smith, M. Waddington, D. E. Dodge, D. A. Bost, M. Riehman, S. Naidich, and B. N. Kreiswirth. Evaluation of protein a gene polymorphic region dna sequencing for typing of *Staphylococcus aureus* strains. *Journal of clinical microbiology*, 37(11):3556–3563, 1999. ISSN 0095-1137 1098-660X. URL https://www.ncbi.nlm.nih.gov/pubmed/10523551https://www.ncbi.nlm.nih.gov/pmc/articles/PMC85690/.

B. Shopsin, M. Gomez, M. Waddington, M. Riehman, and B. N. Kreiswirth. Use of coagulase gene (*coa*) repeat region nucleotide sequences for typing of

methicillin-resistant *Staphylococcus aureus* strains. *Journal of Clinical Microbiology*, 38(9):3453–3456, 2000. URL https://jcm.asm.org/content/jcm/38/9/3453.full.pdf.

S. A. Showsh, E. H. De Boever, and D. B. Clewell. Vancomycin resistance plasmid in enterococcus faecalis that encodes sensitivity to a sex pheromone also produced by *Staphylococcus aureus*. *Antimicrobial agents and chemotherapy*, 45(7):2177–2178, 2001. ISSN 0066-4804 1098-6596. doi: 10.1128/aac.45.7.2177-2178.2001. URL https://www.ncbi.nlm.nih.gov/pubmed/11441824https://www.ncbi.nlm.nih.gov/pmc/articles/PMC90626/.

J. T. Simpson and R. Durbin. Efficient *de novo* assembly of large genomes using compressed data structures. *Genome Research*, 22(3):549–556, 2012. ISSN 1088-9051. doi: 10.1101/gr.126953.111. URL https://dx.doi.org/10.1101/gr.126953.111.

J. T. Simpson, K. Wong, S. D. Jackman, J. E. Schein, S. J. M. Jones, and I. Birol. Abyss: A parallel assembler for short read sequence data. *Genome Research*, 19 (6):1117–1123, 2009. ISSN 1088-9051. doi: 10.1101/gr.089532.108. URL https://dx.doi.org/10.1101/gr.089532.108.

B. Sinha, P. P. François, O. Nüße, M. Foti, O. M. Hartford, P. Vaudaux, T. J. Foster, D. P. Lew, M. Herrmann, and K.-H. Krause. Fibronectin-binding protein acts as *Staphylococcus aureus* invasin via fibronectin bridging to integrin $\alpha\,\beta 1$. *Cellular Microbiology*, 1(2):101–117, 1999. ISSN 1462-5814. doi: https://doi.org/10.1046/j.1462-5822.1999.00011.x. URL https://doi.org/10.1046/j.1462-5822.1999.00011.x.

J. A. Siqueira, C. Speeg-Schatz, F. I. Freitas, J. Sahel, H. Monteil, and G. Prévost. Channel-forming leucotoxins from *Staphylococcus aureus* cause severe inflammatory reactions in a rabbit eye model. *J Med Microbiol*, 46(6):486–94, 1997. ISSN 0022-2615 (Print) 0022-2615. doi: 10.1099/00222615-46-6-486.

E. P. Skaar and O. Schneewind. Iron-regulated surface determinants (isd) of *Staphylococcus aureus*: stealing iron from heme. *Microbes Infect*, 6(4):390–7, 2004. ISSN 1286-4579 (Print) 1286-4579. doi: 10.1016/j.micinf.2003.12.008.

O. Skold. Sulfonamide resistance: mechanisms and trends. *Drug Resist Updat*, 3(3): 155–160, 2000. ISSN 1368-7646. doi: 10.1054/drup.2000.0146.

R. Skov, R. Smyth, A. Larsen, A. Bolmstrom, A. Karlsson, K. Mills, N. Frimodt-Moller, and G. Kahlmeter. Phenotypic detection of methicillin resistance in *Staphylococcus aureus* by disk diffusion testing and etest on mueller-hinton agar. *Journal of clinical microbiology*, 44(12):4395–4399, 2006.

J. Smagur, K. Guzik, M. Bzowska, M. Kuzak, M. Zarebski, T. Kantyka, M. Walski, B. Gajkowska, and J. Potempa. Staphylococcal cysteine protease staphopain b (sspb) induces rapid engulfment of human neutrophils and monocytes by macrophages. *Biol Chem*, 390(4):361–71, 2009a. ISSN 1431-6730 (Print) 1431-6730. doi: 10.1515/bc.2009.042.

J. Smagur, K. Guzik, L. Magiera, M. Bzowska, M. Gruca, I. B. Thøgersen, J. J. Enghild, and J. Potempa. A new pathway of staphylococcal pathogenesis: apoptosis-like death induced by staphopain b in human neutrophils and monocytes. *J Innate Immun*, 1(2):98–108, 2009b. ISSN 1662-811X (Print) 1662-811x. doi: 10.1159/000181014.

E. M. Smith, L. E. Green, G. F. Medley, H. E. Bird, L. K. Fox, Y. H. Schukken, J. V. Kruze, A. J. Bradley, R. N. Zadoks, and C. G. Dowson. Multilocus sequence typing of intercontinental bovine *Staphylococcus aureus* isolates. *Journal of Clinical Microbiology*, 43(9):4737–4743, 2005. doi: 10.1128/jcm.43.9.4737-4743.2005. URL https://jcm.asm.org/content/jcm/43/9/4737.full.pdf.

T. H. Smith, L. K. Fox, and J. R. Middleton. Outbreak of mastitis caused by one strain of *Staphylococcus aureus* in a closed dairy herd. *J Am Vet Med Assoc*, 212 (4):553–6, 1998. ISSN 0003-1488 (Print) 0003-1488.

R. W. Smyth and G. Kahlmeter. Mannitol salt agar-cefoxitin combination as a screening medium for methicillin-resistant *Staphylococcus aureus*. *Journal of Clinical Microbiology*, 43(8):3797–3799, 2005. ISSN 0095-1137. doi: 10.1128/jcm.43.8.3797-3799.2005. URL https://dx.doi.org/10.1128/JCM.43.8.3797-3799.2005.

J.-i. Sohn and J.-W. Nam. The present and future of de novo whole-genome assembly. *Briefings in bioinformatics*, 19(1):23–40, 2018.

J. Sol, O. C. Sampimon, J. J. Snoep, and Y. H. Schukken. Factors associated with bacteriological cure after dry cow treatment of subclinical staphylococcal mastitis

with antibiotics. *Journal of Dairy Science*, 77(1):75–9, 1994. ISSN 0022-0302 (Print) 0022-0302. doi: 10.3168/jds.S0022-0302(94)76930-7.

J. Sol, O. C. Sampimon, J. J. Snoep, and Y. H. Schukken. Factors associated with bacteriological cure during lactation after therapy for subclinical mastitis caused by *Staphylococcus aureus*. *Journal of Dairy Science*, 80(11):2803–2808, 1997. ISSN 0022-0302. doi: https://doi.org/10.3168/jds.S0022-0302(97)76243-X. URL https://www.sciencedirect.com/science/article/pii/S002203029776243X.

L. Song, L. Florea, and B. Langmead. Lighter: fast and memory-efficient sequencing error correction without counting. *Genome Biology*, 15(11):509, 2014. ISSN 1474-760X. doi: 10.1186/s13059-014-0509-9. URL https://doi.org/10.1186/s13059-014-0509-9.

Y. Song, T. La, N. D. Phillips, M. I. Bellgard, and D. J. Hampson. A reverse vaccinology approach to swine dysentery vaccine development. *Veterinary microbiology*, 137(1-2):111–119, 2009. ISSN 0378-1135.

A. Souvorov, R. Agarwala, and D. J. Lipman. Skesa: strategic k-mer extension for scrupulous assemblies. *Genome Biology*, 19(1):153, 2018. ISSN 1474-760X. doi: 10.1186/s13059-018-1540-z. URL https://doi.org/10.1186/s13059-018-1540-z.

A. N. Spaan, M. Vrieling, P. Wallet, C. Badiou, T. Reyes-Robles, E. A. Ohneck, Y. Benito, C. J. de Haas, C. J. Day, M. P. Jennings, G. Lina, F. Vandenesch, K. P. van Kessel, V. J. Torres, J. A. van Strijp, and T. Henry. The staphylococcal toxins γ-haemolysin ab and cb differentially target phagocytes by employing specific chemokine receptors. *Nat Commun*, 5:5438, 2014. ISSN 2041-1723. doi: 10.1038/ncomms6438.

A. N. Spaan, J. A. G. van Strijp, and V. J. Torres. Leukocidins: staphylococcal bi-component pore-forming toxins find their receptors. *Nature reviews. Microbiology*, 15(7):435–447, 2017. ISSN 1740-1534 1740-1526. doi: 10.1038/nrmicro.2017.27. URL https://pubmed.ncbi.nlm.nih.gov/28420883https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5621924/.

C. M. Spahn, G. Blaha, R. K. Agrawal, P. Penczek, R. A. Grassucci, C. A. Trieber, S. R. Connell, D. E. Taylor, K. H. Nierhaus, and J. Frank. Localization of the ribosomal protection protein tet(o) on the ribosome and the mechanism of tetracycline resistance. *Mol Cell*, 7(5):1037–45, 2001. ISSN 1097-2765 (Print) 1097-2765. doi: 10.1016/s1097-2765(01)00238-6.

B. G. Spratt. Multilocus sequence typing: molecular typing of bacterial pathogens in an era of rapid dna sequencing and the internet. *Current opinion in microbiology*, 2 (3):312–316, 1999.

T. Stachyra, P. Levasseur, M. C. Pechereau, A. M. Girard, M. Claudon, C. Miossec, and M. T. Black. *in vitro* activity of the beta-lactamase inhibitor nxl104 against kpc-2 carbapenemase and enterobacteriaceae expressing kpc carbapenemases. *J Antimicrob Chemother*, 64(2):326–9, 2009. ISSN 0305-7453. doi: 10.1093/jac/dkp197.

J. E. Stajich. The bioperl toolkit: Perl modules for the life sciences. *Genome Research*, 12(10):1611–1618, 2002. ISSN 1088-9051. doi: 10.1101/gr.361602. URL https://dx.doi.org/10.1101/gr.361602.

M. R. Stapleton, M. J. Horsburgh, E. J. Hayhurst, L. Wright, I.-M. Jonsson, A. Tarkowski, J. F. Kokai-Kun, J. J. Mond, and S. J. Foster. Characterization of isaa and sced, two putative lytic transglycosylases of *Staphylococcus aureus*. *Journal of bacteriology*, 189(20):7316–7325, 2007. ISSN 0021-9193 1098-5530. doi: 10.1128/JB.00734-07. URL https://pubmed.ncbi.nlm.nih.gov/17675373https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2168438/.

N. M. Steele. *Mastitis pathogen identification using polymerase chain reaction in New Zealand milk samples*. Thesis, IVABS, 2015. URL http://ezproxy.massey.ac.nz/login?url=http://search.ebscohost.com/login.aspx?direct=true$&$db=ir00033a$&$AN=massnz.10179.6892$&$site=eds-live$&$scope=site$&$authtype=sso$&$custid=s3027306http://hdl.handle.net/10179/6892.

C. Steentoft, S. Y. Vakhrushev, H. J. Joshi, Y. Kong, M. B. Vester-Christensen, K. T.-B. Schjoldager, K. Lavrsen, S. Dabelsteen, N. B. Pedersen, L. Marcos-Silva, et al. Precision mapping of the human o-galnac glycoproteome through simplecell technology. *The EMBO journal*, 32(10):1478–1488, 2013.

P. S. Stewart and J. William Costerton. Antibiotic resistance of bacteria in biofilms. *The Lancet*, 358(9276):135–138, 2001. ISSN 0140-6736. doi: https://doi.org/10.1016/S0140-6736(01)05321-1. URL https://www.sciencedirect.com/science/article/pii/S0140673601053211.

A. W. Stott, G. M. Jones, G. J. Gunn, M. Chase-Topping, R. W. Humphry, H. Richardson, and D. N. Logue. Optimum replacement policies for the control of subclinical mastitis due to s.aureus in dairy cows. *Journal of Agricultural Economics*, 53(3):627–644, 2002. ISSN 0021-857X. doi: https://doi.org/10.1111/j.1477-9552.2002.tb00041.x. URL https://doi.org/10.1111/j.1477-9552.2002.tb00041.x.

Y. K. Stranger-Jones, T. Bae, and O. Schneewind. Vaccine assembly from surface proteins of *Staphylococcus aureus*. *Proceedings of the National Academy of Sciences*, 103(45):16942, 2006. doi: 10.1073/pnas.0606863103. URL http://www.pnas.org/content/103/45/16942.abstract.

B. Strommenger, C. Kettlitz, G. Werner, and W. Witte. Multiplex pcr assay for simultaneous detection of nine clinically relevant antibiotic resistance genes in *Staphylococcus aureus*. *Journal of clinical microbiology*, 41(9):4089–4094, 2003.

B. Strommenger, C. Kettlitz, T. Weniger, D. Harmsen, A. W. Friedrich, and W. Witte. Assignment of staphylococcus isolates to groups by spa typing, smai macrorestriction analysis, and multilocus sequence typing. *Journal of Clinical Microbiology*, 44(7):2533–2540, 2006. ISSN 0095-1137. doi: 10.1128/jcm.00420-06. URL https://dx.doi.org/10.1128/JCM.00420-06.

B. Strommenger, C. Braulke, D. Heuck, C. Schmidt, B. Pasemann, U. Nübel, and W. Witte. spa typing of *Staphylococcus aureus* as a frontline tool in epidemiological typing. *Journal of Clinical Microbiology*, 46(2):574–581, 2008. doi: 10.1128/jcm.01599-07. URL https://jcm.asm.org/content/jcm/46/2/574.full.pdf.

M. Su, S. W. Satola, and T. D. Read. Genome-based prediction of bacterial antibiotic resistance. *Journal of clinical microbiology*, 57(3):e01405–18, 2019.

J. M. L. Sung and J. A. Lindsay. *Staphylococcus aureus* strains that are hypersusceptible to resistance gene transfer from enterococci. *Antimicrobial*

Agents and Chemotherapy, 51(6):2189–2191, 2007. ISSN 0066-4804. doi: 10.1128/aac.01442-06. URL https://dx.doi.org/10.1128/AAC.01442-06.

J. M. L. Sung, D. H. Lloyd, and J. A. Lindsay. *Staphylococcus aureus* host specificity: comparative genomics of human versus animal isolates by multi-strain microarray. *Microbiology*, 154(7):1949–1959, 2008. ISSN 1350-0872. doi: 10.1099/mic.0.2007/015289-0. URL https://dx.doi.org/10.1099/mic.0.2007/015289-0.

L. Sutra and B. Poutrel. Virulence factors involved in the pathogenesis of bovine intramammary infections due to *Staphylococcus aureus*. *Journal of Medical Microbiology*, 40(2):79–89, 1994. ISSN 1473-5644. doi: https://doi.org/10.1099/00222615-40-2-79. URL https://www.microbiologyresearch.org/content/journal/jmm/10.1099/00222615-40-2-79.

H. Suzuki, T. Koyanagi, S. Izuka, A. Onishi, and H. Kumagai. The *yliA*, -b, -c, and -d genes of *Escherichia coli* k-12 encode a novel glutathione importer with an atp-binding cassette. *J Bacteriol*, 187(17):5861–7, 2005. ISSN 0021-9193 (Print) 0021-9193. doi: 10.1128/jb.187.17.5861-5867.2005.

P. Szweda, M. SCHIELMAnn, S. Milewski, A. FRAnKOWSKA, and A. JAKuBCZAK. Biofilm production and presence of *ica* and *bap* genes in staphylococcus aureus strains isolated from cows with mastitis in the eastern poland. *Pol J Microbiol*, 61(1):65–9, 2012.

S. Takuno, T. Kado, R. P. Sugino, L. Nakhleh, and H. Innan. Population genomics in bacteria: A case study of *Staphylococcus aureus*. *Molecular Biology and Evolution*, 29(2):797–809, 2012. ISSN 0737-4038. doi: 10.1093/molbev/msr249. URL https://doi.org/10.1093/molbev/msr249.

M. Tanaka, T. Wang, Y. Onodera, Y. Uchida, and K. Sato. Mechanism of quinolone resistance in *Staphylococcus aureus*. *Journal of Infection and Chemotherapy*, 6(3):131–139, 2000. ISSN 1341-321X. doi: https://doi.org/10.1007/s101560070010. URL http://www.sciencedirect.com/science/article/pii/S1341321X0071302X.

Y. Tanizawa, T. Fujisawa, and Y. Nakamura. Dfast: a flexible prokaryotic genome

annotation pipeline for faster genome publication. *Bioinformatics*, 34(6): 1037–1039, 2018.

T. Tatusova, M. DiCuccio, A. Badretdin, V. Chetvernin, E. P. Nawrocki, L. Zaslavsky, A. Lomsadze, K. D. Pruitt, M. Borodovsky, and J. Ostell. Ncbi prokaryotic genome annotation pipeline. *Nucleic acids research*, 44(14):6614–6624, 2016.

F. C. Tenover, R. D. Arbeit, R. V. Goering, P. A. Mickelsen, B. E. Murray, D. H. Persing, and B. Swaminathan. Interpreting chromosomal dna restriction patterns produced by pulsed-field gel electrophoresis: criteria for bacterial strain typing. *J Clin Microbiol*, 33(9):2233–9, 1995. ISSN 0095-1137 (Print) 0095-1137.

T. Tenson, M. Lovmar, and M. Ehrenberg. The mechanism of action of macrolides, lincosamides and streptogramin b reveals the nascent peptide exit path in the ribosome. *Journal of Molecular Biology*, 330(5):1005–1014, 2003. ISSN 0022-2836. doi: https://doi.org/10.1016/S0022-2836(03)00662-4. URL http://www.sciencedirect.com/science/article/pii/S0022283603006624.

H. Tettelin, N. J. Saunders, J. Heidelberg, A. C. Jeffries, K. E. Nelson, J. A. Eisen, K. A. Ketchum, D. W. Hood, J. F. Peden, R. J. Dodson, et al. Complete genome sequence of neisseria meningitidis serogroup b strain mc58. *Science*, 287(5459): 1809–1815, 2000.

H. Tettelin, D. Medini, C. Donati, and V. Masignani. Towards a universal group b streptococcus vaccine using multistrain genome analysis. *Expert review of vaccines*, 5(5):687–694, 2006. ISSN 1476-0584.

I. Thalsø-Madsen, F. R. Torrubia, L. Xu, A. Petersen, C. Jensen, and D. Frees. The sle1 cell wall amidase is essential for β-lactam resistance in community-acquired methicillin-resistant *Staphylococcus aureus* usa300. *Antimicrobial Agents and Chemotherapy*, 64(1):e01931–19, 2019. doi: 10.1128/aac.01931-19. URL https://aac.asm.org/content/aac/64/1/e01931-19.full.pdf.

A. Thomas, S. Chothe, M. Byukusenge, T. Mathews, T. Pierre, S. Kariyawasam, E. Luley, S. Kuchipudi, and B. Jayarao. Prevalence and distribution of multilocus sequence types of *Staphylococcus aureus* isolated from bulk tank milk and cows with mastitis in pennsylvania. *PLOS ONE*, 16(3):e0248528, 2021. doi: 10.1371/journal.pone.0248528. URL https://doi.org/10.1371/journal.pone.0248528.

C. Thorpe, L. Edwards, R. Snelgrove, O. Finco, A. Rae, G. Grandi, R. Guilio, and T. Hussell. Discovery of a vaccine antigen that protects mice from chlamydia pneumoniae infection. *Vaccine*, 25(12):2252–2260, 2007.

J. Tomlin, M. J. Pead, D. H. Lloyd, S. Howell, F. Hartmann, H. A. Jackson, and P. Muir. Methicillin-resistant *Staphylococcus aureus* infections in 11 dogs. *Veterinary Record*, 144(3):60–64, 1999. ISSN 0042-4900. doi: 10.1136/vr.144.3.60. URL https://dx.doi.org/10.1136/vr.144.3.60.

P. L. Toutain, A. Bousquet-Mélou, P. Damborg, A. A. Ferran, D. Mevius, L. Pelligand, K. T. Veldman, and P. Lees. En route towards european clinical breakpoints for veterinary antimicrobial susceptibility testing: A position paper explaining the vetcast approach. *Front Microbiol*, 8:2344, 2017. ISSN 1664-302X (Print) 1664-302x. doi: 10.3389/fmicb.2017.02344.

Q. C. Truong-Bolduc, P. M. Dunman, J. Strahilevitz, S. J. Projan, and D. C. Hooper. Mgra is a multiple regulator of two new efflux pumps in *Staphylococcus aureus*. *J Bacteriol*, 187(7):2395–405, 2005. ISSN 0021-9193 (Print) 0021-9193. doi: 10.1128/jb.187.7.2395-2405.2005.

Q. C. Truong-Bolduc, J. Strahilevitz, and D. C. Hooper. Norc, a new efflux pump regulated by mgra of *Staphylococcus aureus*. *Antimicrobial agents and chemotherapy*, 50(3):1104–1107, 2006. ISSN 0066-4804 1098-6596. doi: 10.1128/AAC.50.3.1104-1107.2006. URL https://pubmed.ncbi.nlm.nih.gov/16495280https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1426456/.

Q. C. Truong-Bolduc, G. R. Bolduc, H. Medeiros, J. M. Vyas, Y. Wang, and D. C. Hooper. Role of the tet38 efflux pump in *Staphylococcus aureus* internalization and survival in epithelial cells. *Infect Immun*, 83(11):4362–72, 2015. ISSN 0019-9567 (Print) 0019-9567. doi: 10.1128/iai.00723-15.

G. E. Tusnady and I. Simon. The hmmtop transmembrane topology prediction server. *Bioinformatics*, 17(9):849–850, 2001a. ISSN 1460-2059.

G. E. Tusnady and I. Simon. The hmmtop transmembrane topology prediction server. *Bioinformatics*, 17(9):849–850, 2001b.

E. Udo, L. Jacob, and B. Mathew. Genetic analysis of methicillin-resistant *Staphylococcus aureus* expressing high- and low-level mupirocin resistance. *Journal of Medical Microbiology*, 50(10):909–915, 2001. ISSN 0022-2615. doi: https://doi.org/10.1099/0022-1317-50-10-909. URL https://www.microbiologyresearch.org/content/journal/jmm/10.1099/0022-1317-50-10-909.

M. Uhlen, B. Guss, B. Nilsson, S. Gatenbeck, L. Philipson, and M. Lindberg. Complete sequence of the staphylococcal gene encoding protein a. a gene evolved through multiple duplications. *J Biol Chem*, 259(3):1695–702, 1984. ISSN 0021-9258 (Print) 0021-9258.

L. H. Ulfman, J. H. W. Leusen, H. F. J. Savelkoul, J. O. Warner, and R. J. J. van Neerven. Effects of bovine immunoglobulins on immune function, allergy, and infection. *Frontiers in nutrition*, 5:52–52, 2018. ISSN 2296-861X. doi: 10.3389/fnut.2018.00052. URL https://pubmed.ncbi.nlm.nih.gov/29988421https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6024018/.

A. Ullmann. Pasteur-koch: Distinctive ways of thinking about infectious diseases. *Microbe-American Society for Microbiology*, 2(8):383, 2007.

P. Valenzuela, A. Medina, W. J. Rutter, G. Ammerer, and B. D. Hall. Synthesis and assembly of hepatitis b virus surface antigen particles in yeast. *Nature*, 298(5872): 347–350, 1982.

J. Valle, A. Toledo-Arana, C. Berasain, J. M. Ghigo, B. Amorena, J. R. Penadés, and I. Lasa. Sara and not sigmab is essential for biofilm development by *Staphylococcus aureus*. *Mol Microbiol*, 48(4):1075–87, 2003. ISSN 0950-382X (Print) 0950-382x. doi: 10.1046/j.1365-2958.2003.03493.x.

F. van Bambeke, M.-P. Mingeot-Leclercq, Y. Glupczynski, and P. M. Tulkens. *Infectious Diseases (Fourth Edition)*, pages 1162–1180.e1. Elsevier, 2017. ISBN 978-0-7020-6285-8. doi: https://doi.org/10.1016/B978-0-7020-6285-8.00137-4. URL http://www.sciencedirect.com/science/article/pii/B9780702062858001374.

A. van Belkum, D. C. Melles, S. V. Snijders, W. B. van Leeuwen, H. F. L. Wertheim, J. L. Nouwen, H. A. Verbrugh, and J. Etienne. 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(4):1555–1557, 2006. ISSN 0095-1137 1098-660X. doi: 10.1128/JCM.44.4.1555-1557.2006. URL https://pubmed.ncbi.nlm.nih.gov/16597892https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1448628/.

S. van den Berg, D. G. A. M. Koedijk, J. W. Back, J. Neef, A. Dreisbach, J. M. van Dijl, I. A. J. M. Bakker-Woudenberg, and G. Buist. Active immunization with an octa-valent *Staphylococcus aureus* antigen mixture in models of *S. aureus* bacteremia and skin infection in mice. *PLOS ONE*, 10(2):e0116847, 2015. doi: 10.1371/journal.pone.0116847. URL https://doi.org/10.1371/journal.pone.0116847.

B. H. P. Van Den Borne, M. Nielen, G. Van Schaik, M. B. Melchior, T. J. G. M. Lam, and R. N. Zadoks. Host adaptation of bovine *Staphylococcus aureus* seems associated with bacteriological cure after lactational antimicrobial treatment. *Journal of Dairy Science*, 93(6):2550–2558, 2010. ISSN 0022-0302. doi: 10.3168/jds.2009-2971. URL https://dx.doi.org/10.3168/jds.2009-2971.

W. J. B. van Wamel, S. H. M. Rooijakkers, M. Ruyken, K. P. M. van Kessel, and J. A. G. van Strijp. 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(4): 1310–1315, 2006. doi: 10.1128/jb.188.4.1310-1315.2006. URL https://jb.asm.org/content/jb/188/4/1310.full.pdf.

F. Vandenesch, G. Lina, and T. Henry. *Staphylococcus aureus* hemolysins, bi-component leukocidins, and cytolytic peptides: A redundant arsenal of membrane-damaging virulence factors? *Frontiers in Cellular and Infection Microbiology*, 2(12), 2012. ISSN 2235-2988. doi: 10.3389/fcimb.2012.00012. URL https://www.frontiersin.org/article/10.3389/fcimb.2012.00012.

P. Vannuffel and C. Cocito. Mechanism of action of streptogramins and macrolides. *Drugs*, 51(1):20–30, 1996. ISSN 1179-1950. doi: 10.2165/00003495-199600511-00006. URL https://doi.org/10.2165/00003495-199600511-00006.

N. P. Varela, R. Friendship, C. Dewey, and A. Valdivieso. Comparison of agar dilution and e-test for antimicrobial susceptibility testing of campylobacter coli isolates recovered from 80 ontario swine farms. *Canadian journal of veterinary research*, 72(2):168, 2008.

A. Varley, J. Sule, and A. Absalom. Principles of antibiotic therapy. *Continuing Education in Anaesthesia Critical Care & Pain*, 9(6):184–188, 2009. ISSN 1743-1816. doi: 10.1093/bjaceaccp/mkp035. URL https://doi.org/10.1093/bjaceaccp/mkp035.

K. A. Veh, R. C. Klein, C. Ster, G. Keefe, P. Lacasse, D. Scholl, J. P. Roy, D. Haine, S. Dufour, B. G. Talbot, A. O. B. Ribon, and F. Malouin. Genotypic and phenotypic characterization of *Staphylococcus aureus* causing persistent and nonpersistent subclinical bovine intramammary infections during lactation or the dry period. *Journal of Dairy Science*, 98(1):155–168, 2015. ISSN 0022-0302. doi: https://doi.org/10.3168/jds.2014-8044. URL https://www.sciencedirect.com/science/article/pii/S0022030214007280.

T. Vesikari, A. Forstén, M. G. Desole, G. Ferrera, M. Caubet, N. Mesaros, and D. Boutriau. A combined haemophilus influenzae type b neisseria meningitidis serogroup c tetanus toxoid conjugate vaccine is immunogenic and well-tolerated when coadministered with diphtheria, tetanus, acellular pertussis hepatitis b–inactivated poliovirus at 3, 5 and 11 months of age: Results of an open, randomized, controlled study. *The Pediatric infectious disease journal*, 32(5): 521–529, 2013.

D. Viana, L. Selva, P. Segura, J. R. Penadés, and J. M. Corpa. Genotypic characterization of *Staphylococcus aureus* strains isolated from rabbit lesions. *Veterinary Microbiology*, 121(3-4):288–298, 2007. ISSN 0378-1135. doi: 10.1016/j.vetmic.2006.12.003. URL https://dx.doi.org/10.1016/j.vetmic.2006.12.003.

U. Vijaranakul, M. J. Nadakavukaren, B. L. de Jonge, B. J. Wilkinson, and R. K. Jayaswal. Increased cell size and shortened peptidoglycan interpeptide bridge of nacl-stressed *Staphylococcus aureus* and their reversal by glycine betaine. *J Bacteriol*, 177(17):5116–21, 1995. ISSN 0021-9193 (Print) 0021-9193. doi: 10.1128/jb.177.17.5116-5121.1995.

S. Vivona, F. Bernante, and F. Filippini. Nerve: new enhanced reverse vaccinology environment. *BMC biotechnology*, 6(1):1–8, 2006. ISSN 1472-6750.

M. Vos and X. Didelot. A comparison of homologous recombination rates in bacteria and archaea. *The ISME Journal*, 3(2):199–208, 2009. ISSN 1751-7362. doi: 10.1038/ismej.2008.93. URL https://dx.doi.org/10.1038/ismej.2008.93.

P. Vos, G. Garrity, D. Jones, N. R. Krieg, W. Ludwig, F. A. Rainey, K.-H. Schleifer, and W. B. Whitman. *Bergey's manual of systematic bacteriology: Volume 3: The Firmicutes*, volume 3. Springer Science & Business Media, 2011.

A. Voss, F. Loeffen, J. Bakker, C. Klaassen, and M. Wulf. Methicillin-resistant*Staphylococcus aureus*in pig farming. *Emerging Infectious Diseases*, 11(12):1965–1966, 2005. ISSN 1080-6040. doi: 10.3201/eid1112.050428. URL https://dx.doi.org/10.3201/eid1112.050428.

M. Vrieling, K. Koymans, D. Heesterbeek, P. Aerts, V. Rutten, C. De Haas, K. Van Kessel, A. Koets, R. Nijland, and J. van Strijp. Bovine *Staphylococcus aureus* secretes the leukocidin lukmf' to kill migrating neutrophils through ccr1. *MBio*, 6(3):e00335–15, 2015a.

M. Vrieling, K. J. Koymans, D. A. Heesterbeek, P. C. Aerts, V. P. Rutten, C. J. de Haas, K. P. van Kessel, A. P. Koets, R. Nijland, and J. A. van Strijp. Bovine *Staphylococcus aureus* secretes the leukocidin lukmf' to kill migrating neutrophils through ccr1. *mBio*, 6(3):e00335, 2015b. doi: 10.1128/mBio.00335-15.

M. Vrieling, E. M. Boerhout, G. F. van Wigcheren, K. J. Koymans, T. G. Mols-Vorstermans, C. J. C. de Haas, P. C. Aerts, I. J. J. M. Daemen, K. P. M. van Kessel, A. P. Koets, V. P. M. G. Rutten, P. J. M. Nuijten, J. A. G. van Strijp, and L. Benedictus. Lukmf' is the major secreted leukocidin of bovine *Staphylococcus aureus* and is produced *in vivo* during bovine mastitis. *Scientific Reports*, 6(1): 37759, 2016. ISSN 2045-2322. doi: 10.1038/srep37759. URL https://doi.org/10.1038/srep37759.

D. E. Waldron and J. A. Lindsay. Sau1: a novel lineage-specific type i restriction-modification system that blocks horizontal gene transfer into *Staphylococcus aureus* and between *S. aureus* isolates of different lineages.

Journal of Bacteriology, 188(15):5578–5585, 2006. ISSN 0021-9193. doi: 10.1128/jb.00418-06. URL https://dx.doi.org/10.1128/JB.00418-06.

B. J. Walker, T. Abeel, T. Shea, M. Priest, A. Abouelliel, S. Sakthikumar, C. A. Cuomo, Q. Zeng, J. Wortman, S. K. Young, and A. M. Earl. Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLOS ONE*, 9(11):e112963, 2014. doi: 10.1371/journal.pone.0112963. URL https://doi.org/10.1371/journal.pone.0112963.

M. Wang, Y. Wei, W. Yu, L. Wang, L. Zhai, X. Li, X. Wang, H. Zhang, Z. Feng, L. Yu, Y. Yu, J. Ma, and Y. Cui. Identification of a conserved linear b-cell epitope in the *Staphylococcus aureus* gapc protein. *Microb Pathog*, 118:39–47, 2018a. ISSN 1096-1208 (Electronic) 0882-4010 (Linking). doi: 10.1016/j.micpath.2018.03.007. URL https://www.ncbi.nlm.nih.gov/pubmed/29522802.

S.-C. Wang, C.-M. Wu, S.-C. Xia, Q. Yong-Hua, L.-N. Xia, and J.-Z. Shen. Distribution of superantigenic toxin genes in *Staphylococcus aureus* isolates from milk samples of bovine subclinical mastitis cases in two major diary production regions of china. *Veterinary microbiology*, 137(3-4):276–281, 2009. ISSN 0378-1135.

W. Wang, X. Lin, T. Jiang, Z. Peng, J. Xu, L. Yi, F. Li, S. Fanning, and Z. Baloch. Prevalence and characterization of *Staphylococcus aureus* cultured from raw milk taken from dairy cows with mastitis in beijing, china. *Frontiers in Microbiology*, 9 (1123), 2018b. ISSN 1664-302X. doi: 10.3389/fmicb.2018.01123. URL https://www.frontiersin.org/article/10.3389/fmicb.2018.01123.

M. Waskom, O. Botvinnik, D. O'Kane, P. Hobson, S. Lukauskas, D. C. Gemperline, T. Augspurger, Y. Halchenko, J. B. Cole, J. Warmenhoven, J. D. Ruiter, C. Pye, S. Hoyer, J. Vanderplas, S. Villalba, G. Kunter, E. Quintero, P. Bachant, M. Martin, K. Meyer, A. Miles, Y. Ram, T. Yarkoni, M. L. Williams, C. Evans, C. Fitzgerald, , Brian, C. Fonnesbeck, A. Lee, and A. Qalieh. Mwaskom/seaborn: V0.8.1 (september 2017), 2017. URL https://zenodo.org/record/883859.

A. R. Wattam, D. Abraham, O. Dalay, T. L. Disz, T. Driscoll, J. L. Gabbard, J. J. Gillespie, R. Gough, D. Hix, R. Kenyon, D. Machi, C. Mao, E. K. Nordberg, R. Olson, R. Overbeek, G. D. Pusch, M. Shukla, J. Schulman, R. L. Stevens, D. E.

Sullivan, V. Vonstein, A. Warren, R. Will, M. J. Wilson, H. S. Yoo, C. Zhang, Y. Zhang, and B. W. Sobral. Patric, the bacterial bioinformatics database and analysis resource. *Nucleic Acids Res*, 42(Database issue):D581–91, 2014. ISSN 0305-1048 (Print) 0305-1048. doi: 10.1093/nar/gkt1099.

A. R. Wattam, J. J. Davis, R. Assaf, S. Boisvert, T. Brettin, C. Bun, N. Conrad, E. M. Dietrich, T. Disz, J. L. Gabbard, et al. Improvements to patric, the all-bacterial bioinformatics database and analysis resource center. *Nucleic acids research*, 45 (D1):D535–D542, 2017.

J. L. Watts, M. T. Sweeney, and B. V. Lubbers. Antimicrobial susceptibility testing of bacteria of veterinary origin. *Antimicrobial Resistance in Bacteria from Livestock and Companion Animals*, pages 17–32, 2018. ISSN 9781683670520. doi: https://doi.org/10.1128/9781555819804.ch2. URL https://doi.org/10.1128/9781555819804.ch2.

B. Weisblum. Erythromycin resistance by ribosome modification. *Antimicrobial Agents and Chemotherapy*, 39(3):577–585, 1995. ISSN 0066-4804. doi: 10.1128/aac.39.3.577. URL https://dx.doi.org/10.1128/aac.39.3.577.

O. Wellnitz and R. M. Bruckmaier. The innate immune response of the bovine mammary gland to bacterial infection. *The veterinary journal*, 192(2):148–152, 2012.

S. Wendlandt, C. Lozano, K. Kadlec, E. Gómez-Sanz, M. Zarazaga, C. Torres, and S. Schwarz. The enterococcal abc transporter gene *lsa*(e) confers combined resistance to lincosamides, pleuromutilins and streptogramin a antibiotics in methicillin-susceptible and methicillin-resistant *Staphylococcus aureus*. *Journal of Antimicrobial Chemotherapy*, 68(2):473–475, 2012. ISSN 0305-7453. doi: 10.1093/jac/dks398. URL https://doi.org/10.1093/jac/dks398.

H. F. Wertheim, D. C. Melles, M. C. Vos, W. van Leeuwen, A. van Belkum, H. A. Verbrugh, and J. L. Nouwen. The role of nasal carriage in *Staphylococcus aureus* infections. *The Lancet infectious diseases*, 5(12):751–762, 2005. ISSN 1473-3099.

C. A. Wesson, L. E. Liou, K. M. Todd, G. A. Bohach, W. R. Trumble, and K. W. Bayles. *Staphylococcus aureus* agr and sar global regulators influence internalization and induction of apoptosis. *Infection and immunity*, 66(11): 5238–5243, 1998. ISSN 0019-9567 1098-5522. doi:

10.1128/IAI.66.11.5238-5243.1998. URL https://pubmed.ncbi.nlm.nih.gov/9784528https://www.ncbi.nlm.nih.gov/pmc/articles/PMC108654/.

I. Wiegand, K. Hilpert, and R. E. Hancock. Agar and broth dilution methods to determine the minimal inhibitory concentration (mic) of antimicrobial substances. *Nature protocols*, 3(2):163–175, 2008.

A. Wieser, L. Schneider, J. Jung, and S. Schubert. Maldi-tof ms in microbiological diagnostics—identification of microorganisms and beyond (mini review). *Applied Microbiology and Biotechnology*, 93(3):965–974, 2012. ISSN 0175-7598. doi: 10.1007/s00253-011-3783-4. URL https://dx.doi.org/10.1007/s00253-011-3783-4.

J. Williams, S. GR, S. GL, and G. DL. A clinical evaluation of *Staphylococcus aureus* bacterin in the control of staphylococcal mastitis in cows. *Veterinary medicine, small animal clinician*, 1975.

D. A. Williamson, S. Bakker, G. W. Coombs, H. Tan, S. Monecke, and H. Heffernan. Emergence and molecular characterization of clonal complex 398 (cc398) methicillin-resistant *Staphylococcus aureus* (mrsa) in new zealand. *J Antimicrob Chemother*, 69(5):1428–30, 2014a. ISSN 0305-7453. doi: 10.1093/jac/dkt499.

D. A. Williamson, S. Monecke, H. Heffernan, S. R. Ritchie, S. A. Roberts, A. Upton, M. G. Thomas, and J. D. Fraser. High usage of topical fusidic acid and rapid clonal expansion of fusidic acid–resistant *Staphylococcus aureus*: A cautionary tale. *Clinical Infectious Diseases*, 59(10):1451–1454, 2014b. ISSN 1058-4838. doi: 10.1093/cid/ciu658. URL https://doi.org/10.1093/cid/ciu658.

D. J. Wilson, R. N. Gonzalez, and P. M. Sears. Segregation or use of separate milking units for cows infected with *Staphylococcus aureus*: effects on prevalence of infection and bulk tank somatic cell count. *Journal of Dairy Science*, 78(9): 2083–5, 1995. ISSN 0022-0302 (Print) 0022-0302. doi: 10.3168/jds.S0022-0302(95)76834-5.

D. N. Wilson. Ribosome-targeting antibiotics and mechanisms of bacterial resistance. *Nature Reviews Microbiology*, 12(1):35–48, 2014. ISSN 1740-1526. doi: 10.1038/nrmicro3155. URL https://dx.doi.org/10.1038/nrmicro3155.

S. Wingett and S. Andrews. Fastq screen: A tool for multi-genome mapping and quality control [version 2; peer review: 4 approved]. *F1000Research*, 7(1338), 2018. doi: 10.12688/f1000research.15931.2. URL http://openr.es/dqv.

L. Wondrack, M. Massa, B. V. Yang, and J. Sutcliffe. Clinical strain of *Staphylococcus aureus* inactivates and causes efflux of macrolides. *Antimicrobial Agents and Chemotherapy*, 40(4):992–998, 1996. doi: 10.1128/aac.40.4.992. URL https://aac.asm.org/content/aac/40/4/992.full.pdf.

A. M. Woodin. Purification of the two components of leucocidin from *Staphylococcus aureus*. *The Biochemical journal*, 75(1):158–165, 1960. ISSN 0264-6021 1470-8728. doi: 10.1042/bj0750158. URL https://pubmed.ncbi.nlm.nih.gov/13845860https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1204343/.

A. Woods. Science, disease and dairy production in britain, c. 1927 to 1980. *Agricultural History Review*, 62(2):294–314, 2014. URL https://www.ingentaconnect.com/content/bahs/agrev/2014/00000062/00000002/art00008.

WorldHealthOrganization. Types of vaccine, 2018. URL https://vaccine-safety-training.org/types-of-vaccine-overview.html.

WorldHealthOrganization. Who and unicef warn of a decline in vaccinations during covid-19, 2020. URL https://www.who.int/news/item/15-07-2020-who-and-unicef-warn-of-a-decline-in-vaccinations-during-covid-19.

S. Yamada, M. Sugai, H. Komatsuzawa, S. Nakashima, T. Oshida, A. Matsumoto, and H. Suginaka. An autolysin ring associated with cell separation of *Staphylococcus aureus*. *J Bacteriol*, 178(6):1565–71, 1996. ISSN 0021-9193 (Print) 0021-9193. doi: 10.1128/jb.178.6.1565-1571.1996.

T. Yamada, N. Tochimaru, S. Nakasuji, E. Hata, H. Kobayashi, M. Eguchi, J. Kaneko, Y. Kamio, T. Kaidoh, and S. Takeuchi. Leukotoxin family genes in *Staphylococcus aureus* isolated from domestic animals and prevalence of *lukM–lukF-PV* genes by bacteriophages in bovine isolates. *Veterinary Microbiology*, 110(1):97–103, 2005. ISSN 0378-1135. doi: https://doi.org/10.1016/j.vetmic.2005.07.006. URL https://www.sciencedirect.com/science/article/pii/S0378113505002531.

S. P. Yazdankhah, A. L. Hellemann, K. Rønningen, and E. Olsen. Rapid and sensitive detection of staphylococcus species in milk by elisa based on monodisperse magnetic particles. *Vet Microbiology*, 62(1):17–26, 1998. ISSN 0378-1135 (Print) 0378-1135. doi: 10.1016/s0378-1135(98)00193-x.

J. Ye, S. McGinnis, and T. L. Madden. Blast: improvements for better sequence analysis. *Nucleic acids research*, 34(suppl_2):W6–W9, 2006.

N. Y. Yu, J. R. Wagner, M. R. Laird, G. Melli, S. Rey, R. Lo, P. Dao, S. C. Sahinalp, M. Ester, L. J. Foster, et al. Psortb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics*, 26(13):1608–1615, 2010.

R. N. Zadoks, J. R. Middleton, S. McDougall, J. Katholm, and Y. H. Schukken. Molecular epidemiology of mastitis pathogens of dairy cattle and comparative relevance to humans. *Journal of Mammary Gland Biology and Neoplasia*, 16(4): 357–372, 2011. ISSN 1083-3021. doi: 10.1007/s10911-011-9236-y. URL https://dx.doi.org/10.1007/s10911-011-9236-y.

P. Zangerl and H. Asperger. *Chapter 6 Media used in the detection and enumeration of* Staphylococcus aureus, volume 37, pages 91–110. Elsevier, 2003. ISBN 0079-6352. doi: https://doi.org/10.1016/S0079-6352(03)80009-7. URL http://www.sciencedirect.com/science/article/pii/S0079635203800097.

E. Zankari, H. Hasman, S. Cosentino, M. Vestergaard, S. Rasmussen, O. Lund, F. M. Aarestrup, and M. V. Larsen. Identification of acquired antimicrobial resistance genes. *Journal of antimicrobial chemotherapy*, 67(11):2640–2644, 2012.

E. Zankari, R. Allesøe, K. G. Joensen, L. M. Cavaco, O. Lund, and F. M. Aarestrup. Pointfinder: a novel web tool for wgs-based detection of antimicrobial resistance associated with chromosomal point mutations in bacterial pathogens. *Journal of Antimicrobial Chemotherapy*, 72(10):2764–2768, 2017. ISSN 0305-7453. doi: 10.1093/jac/dkx217. URL https://doi.org/10.1093/jac/dkx217.

D. R. Zerbino. Using the velvet de novo assembler for short-read sequencing technologies. *Current protocols in bioinformatics*, Chapter 11:Unit–11.5, 2010. ISSN 1934-340X 1934-3396. doi: 10.1002/0471250953.bi1105s31. URL https://www.ncbi.nlm.nih.gov/pubmed/20836074https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2952100/.

B. Z. Zhang, Y. H. Hua, B. Yu, C. C. Y. Lau, J. P. Cai, S. Y. Zheng, W. C. Yam, R. Y. T. Kao, K. H. Sze, B. J. Zheng, K. Y. Yuen, and J. D. Huang. Recombinant esat-6-like proteins provoke protective immune responses against invasive *Staphylococcus aureus* disease in a murine model. *Infection and Immunity*, 83(1): 339, 2015. doi: 10.1128/IAI.02498-14. URL http://iai.asm.org/content/83/1/339.abstract.

H. Zhang, Y. Zheng, H. Gao, P. Xu, M. Wang, A. Li, M. Miao, X. Xie, Y. Deng, H. Zhou, and H. Du. Identification and characterization of *Staphylococcus aureus* strains with an incomplete hemolytic phenotype. *Frontiers in cellular and infection microbiology*, 6:146–146, 2016. ISSN 2235-2988. doi: 10.3389/fcimb.2016.00146. URL https://pubmed.ncbi.nlm.nih.gov/27917374https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5114236/.

Y. Zhao, M. Zhou, Y. Gao, H. Liu, W. Yang, J. Yue, and D. Chen. Shifted t helper cell polarization in a murine *Staphylococcus aureus* mastitis model. *PLOS ONE*, 10(7): e0134797, 2015. doi: 10.1371/journal.pone.0134797. URL https://doi.org/10.1371/journal.pone.0134797.

ZionMarketResearch. Global veterinary vaccine market expected to reach usd 12.10 billion by 2021: Zion market research, Nov 2016. URL https://www.globenewswire.com/en/news-release/2016/11/16/890528/0/en/Global-Veterinary-Vaccine-Market-expected-to-Reach-USD-12-10-Billion-by-2021html.

G. Ziv and M. Storper. Intramuscular treatment of subclinical staphylococcal mastitis in lactating cows with penicillin g, methicillin and their esters. *J Vet Pharmacol Ther*, 8(3):276–83, 1985. ISSN 0140-7783 (Print) 0140-7783. doi: 10.1111/j.1365-2885.1985.tb00957.x.