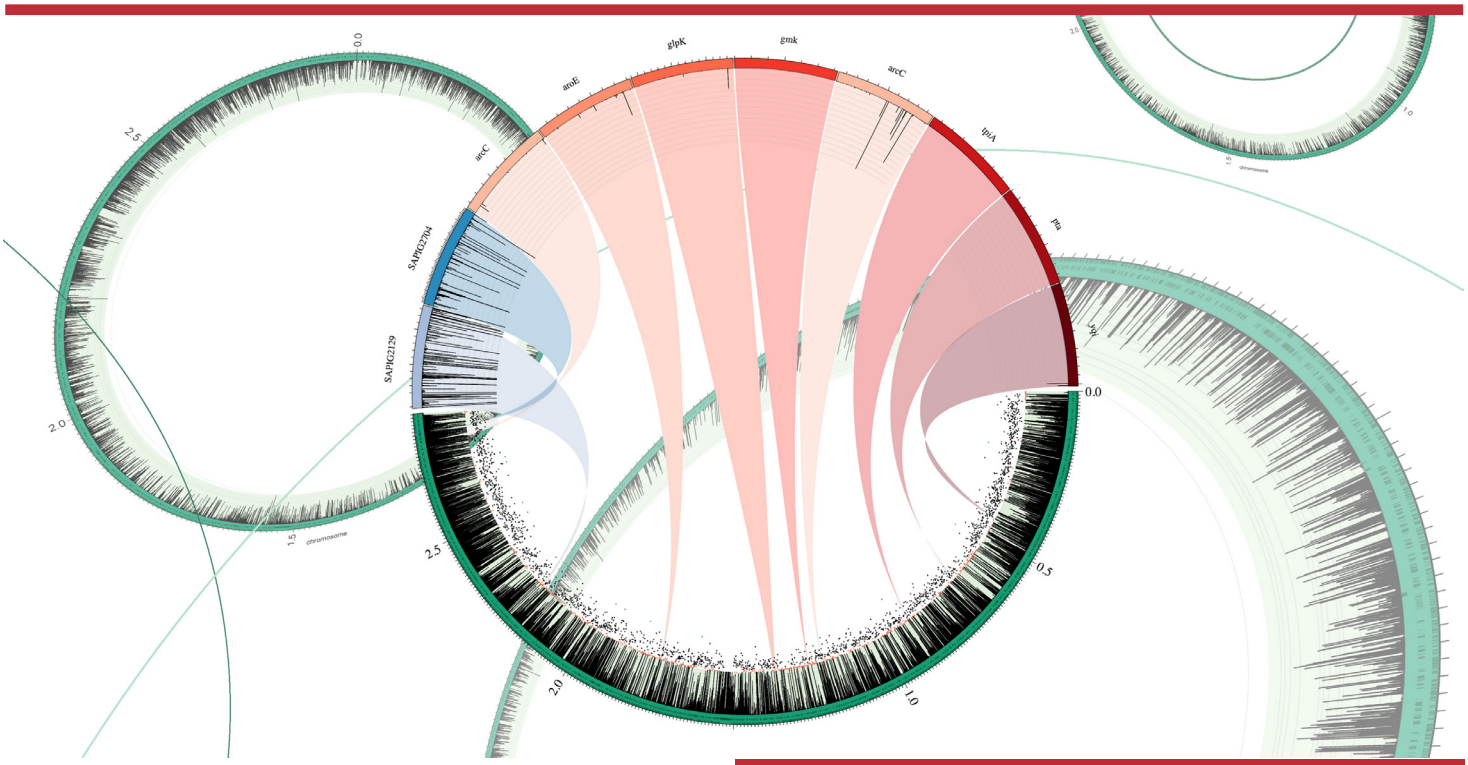
Characterization of methicillin-resistant *Staphylococcus aureus* Sequence Type 398



Mette Theilgaard Christiansen
PhD Thesis
2014

DTU Food
National Food Institute

Characterization of methicillin-resistant S*taphylococcus aureus* Sequence Type 398

PhD thesis by

Mette Theilgaard Christiansen, cand.scient.techn.

Revised version

Submitted February 2014

National Food Institute

Technical University of Denmark

Division for Epidemiology and Microbial Genomics

Kemitorvet building 204, 2800 Kgs. Lyngby, DK

Preface

The thesis is a result of a three-year PhD project carried out at the national Food Institute at the Technical University of Denmark in the Division of Epidemiology and Microbial Genomics. The project was supervised by Professor Frank Møller Aarestrup as main supervisor and Senior Researchers Henrik Hasman and Yvonne Agersø as co-supervisors. The project was conducted from June 2010 to August 2013 and a six month external research stay was included from October 2011 to March 2012 at Department of Veterinary Medicine, University of Cambridge, UK under the supervision of Senior Lecturer Dr. Mark Holmes. The work was funded by the Danish Ministry of Food, Agriculture and Fisheries (Grant no.: 3304-FVFP-09-F-002-1) and The Technical University of Denmark.

The thesis consists of an introduction and three manuscripts all presented in three chapters. Chapter 1 includes a short introduction to *Staphylococcus aureus* and to some of the methods used for bacterial characterization. Studies characterizing *S. aureus* and especially *S. aureus* multilocus sequence type 398 (ST398) are presented. In chapter 2 the results obtained in the three manuscripts are summarized and discussed. Chapter 3 contains the three manuscripts (manuscript I-III). Manuscript I describes the generation, verification, and evaluation of a high-throughput approach for bacterial characterization. In manuscript II a genomic screen was performed to identify genes important for *S. aureus* ST398 survival in a porcine reservoir. Manuscript III represents the development of a bioinformatic tool that can be used for virulence profiling of *S. aureus* using whole genome sequence data.

Supervisors:

Professor Frank M. Aarestrup (main supervisor), Research Group of Microbial Genomics and Antimicrobial Resistance, Division of Epidemiology and Microbial Genomics, National food ‘institute, Technical University of Denmark

Senior Scientist Henrik Hasman (co-supervisor), Research Group of Microbial Genomics and Antimicrobial Resistance, Division of Epidemiology and Microbial Genomics, National food ‘institute, Technical University of Denmark

Senior Scientist Yvonne Agersø (co-supervisor), Research Group of Microbial Genomics and Antimicrobial Resistance, Division of Epidemiology and Microbial Genomics, National food ‘institute, Technical University of Denmark

Assessment committee:

Senior Scientist Lina Cavaco, Research Group of Microbial Genomics and Antimicrobial Resistance, Division of Epidemiology and Microbial Genomics, National food ‘institute, Technical University of Denmark

Professor Jody A. Lindsay, Research Centre for Infection and Immunity, Division of Clinical Sciences, St. George’s University of London, UK

Professor Henrik Westh, Department of Clinical Microbiology, Hvidovre Hospital, DK

Acknowledgements

There are a lot of people I want to thank, who in one way or another have contributed to the completion of this work.

First of all I would like to thank my supervisor Frank Aarestrup for the last three years. It has been hard, fun, frustrating, challenging but most of all educational. Also thank you Henrik and Yvonne for your co-supervision and for, at times, putting Frank's ideas into a more realistic context.

Another very important person I would like to thank is Mark Holmes, my external supervisor at Cambridge University. You opened up your lab for me without hesitation. You and your colleagues helped me through a though period where no experiments would work and you have opened a door of opportunity in the UK for me – thank you!

Acknowledgement and many thanks to all my colleagues at DTU Food! You have definitely made my stay at DTU a pleasant one and without your assistance I would have been lost.

Thanks to Maria and Rita for being the best officemates one could ever hope for. You have been great in every aspect. A special thanks goes to Rolf! You have helped me tremendously with my data analysis and you have contributed to my bioinformatic skills in a way that I'm now not afraid to take a job using bioinformatics on a daily basis.

I would also like to thank all my wonderful friends for putting up with me and my talk about MRSA the past years. It must have been tiring at times, but also very exciting, right? Anne, thank you so much for proof reading everything, it has really been a great help.

Thank you to my family, you're great and I love you. Mom and dad, thank you for always being supportive and proud of me no matter what!

Above all, thanks to my amazing husband. Kenneth, you have been more than any, the greatest support, suffering with me at times and celebrating at others. Your high expectations continue to push me and you make me feel like I can do anything. Thank you!

English summary

Staphylococcus aureus is an opportunistic pathogen that colonizes the nares and skin surfaces of several animal species, including man. *S. aureus* can cause a wide variety of infections ranging from superficial soft tissue and skin infections to severe and deadly systemic infections. Traditionally *S. aureus* and methicillin-resistant *Staphylococcus aureus* (MRSA) have been associated with hospitals, but during the past decades MRSA has emerged in the community and now a new branch of MRSA has been found in association with livestock (LA-MRSA). A specific lineage (multilocus sequence type 398 (ST398)) has been particularly successful in colonization of pigs and ST398 has become the most frequently reported MRSA strain found in associated with livestock.

Currently the understanding of the successful colonization and transmission of LA-MRSA ST398 in pigs are limited and mainly based on observational field surveys. The aim of this work was to develop a high-throughput approach for genotypic and phenotypic characterization of LA-MRSA ST398 in the porcine reservoir.

The thesis represents three studies (manuscript I-III). In manuscript I a genome-saturated transposon mutant library was generated and Transposon Directed Inserted site Sequencing (TraDIS) was for the first time assessed in an LA-MRSA ST398 strain. Using this high-throughput approach, genes essential for LA-MRSA ST398 survival under laboratory conditions and in whole porcine blood *in vitro* were identified. In manuscript II, genes important for LA-MRSA ST398 survival on porcine skin and nasal epithelium *ex vivo* were identified. These genes could represent targets for de-colonization, which could help prevent further spread and adaption of LA-MRSA ST398. Manuscript III describes the construction of the *S. aureus* VirulenceFinder database. The database can be applied for identification of virulence genes in *S. aureus* using whole genome

sequence data. The *S. aureus* VirulenceFinder will be part of the tool package generated for the Centre for Genomic Epidemiology (CGE) (www.genomicepidemiology.org).

Dansk resumé

Staphylococcus aureus er en opportunistisk patogen, der koloniserer næsen og huden hos forskellige dyrearter, inklusive mennesker. *S. aureus* kan forårsage en lang række forskellige infektioner, der bl.a. omfatter overfladiske hudinfektioner og mere alvorlige systemiske infektioner. Traditionelt har *S. aureus* og methicillin-resistente *Staphylococcus aureus* (MRSA) været associeret med hospitalserhvervet infektioner, men i de seneste årtier har MRSA spredt sig til resten af samfundet, og en ny gren af MRSA med association til produktionsdyr er blevet identificeret (LA-MRSA). En specifik slægt har vist sig at være yderst succesfuld til at kolonisere grise, og denne slægt (ST398) er nu den hyppigst rapporterede MRSA stamme fundet i association med produktionsdyr.

Den nuværende viden, om hvorfor LA-MRSA ST398 er succesfuld i kolonisation og spredning hos grise, er begrænset og primært baseret på overvågnings-studier. Formålet med dette studie var at udvikle en metode til at udføre en omfattende genotypisk og fænotypisk karaktering af LA-MRSA ST398 i et grise-reservoir.

Denne afhandling repræsenterer tre studier (manuskript I-III). I manuskript I blev der genereret et genom-mættet transposon mutant bibliotek, og 'Transposon Directed Inserted site Sequencing (TraDIS)' blev for første gang anvendt på en LA-MRSA ST398 stamme. Ved brugen af disse metoder blev gener essentielle for LA-MRSA ST398s overlevelse under laboratorie forhold og i svine blod identificeret. I manuskript II blev gener vigtige for LA-MRSA ST398 overlevelse på svine hud og svine-næseepitel bestemt. Disse gener kan potentielt udgøre nye mål for afkolonisering og dermed forebygge videre spredning og tilpasning af LA-MRSA ST398. Manuskript III beskriver konstruktionen af en *S. aureus* VirulenceFinder database. Denne kan anvendes til identifikation af virulens gener i *S. aureus* hel-genom sekvens data. *S. aureus*

VirulenceFinder er en del af den redskabspakke, der bliver genereret for Center for Genomisk Epidemiologi (CGE) (www.genomicepidemiology.org).

Table of Contents

List of abbreviations

S. aureus	*Staphylococcus aureus*
MSSA	Methicillin-sensitive *Staphylococcus aureus*
MRSA	Methicillin-resistant *Staphylococcus aureus*
HA-MRSA	Healthcare-associated Methicillin-resistant *Staphylococcus aureus*
CA-MRSA	Community- associated Methicillin-resistant *Staphylococcus aureus*
LA-MRSA	Livestock-associated Methicillin-resistant *Staphylococcus aureus*
TMDH	Transposon Mediated Differential Hybridization
TraDIS	Transposon Directed Insertion site Sequencing
WGS	Whole genome sequencing
DNA	Deoxyribonucleic acid
°C	Degrees (Celsius)
g	Gram
rt	Room temperature
MIC	Minimum inhibitory concentration
min	Minutes
CC	Clonal complex
ST	Sequence Type
PCR	Polymerase Chain Reaction
Q-PCR	Real-time quantitative PCR
BLAST	Basic Local Alignment Search Tool
bp	Base pair
NCBI	National Center for Biotechnology Information
Mb	Megabases
NaCl	Sodium chloride
WT	Wild-Type
BHI	Brain Heart Infusion
TSB	Tryptic Soy Broth
TSA	Tryptic Soy Agar
LB	Luria Broth
mg	milligram
µg	Microgram

CFU	Colony forming units
app.	Approximate
AAC6'-APH2'	6´-acetyltransferase-2´´-phosphotransferase (gentamicin resistance)
OD	Optical density
CaCl	Calcium chloride
M	Molar
mM	Millimolar
ml	Millilitre
µl	Microlitre
rmp	Rounds per minute
EDTA	Ethylenediaminetetraacetic acid
gen	Gentamicin
ery	Erythromycin
tet	Tetracycline
chl	Chloramphenicol
ng	Nanogram
Tn	Transposon
Q	Quality
COG	Cluster of Orthologous groups
SCC*mec*	Staphylococcal Cassette Chromosome mec
RNA	Ribonucleic acid
mRNA	messenger RNA
PBS	Phosphate buffered saline
KCl	Potassium chloride
$MgSO_4$	Magnesium Sulphate
Hz	Hertz
DMEM	Dulbecco's Modified Eagle Medium
RPMI	Roswell Park Memorial Institute medium
U	Unit
CO_2	Carbon Dioxide
SMIT	Size Marker Identification Technology

Background

Methicillin-resistant *Staphylococcus aureus* (MRSA) is one of the most frequent cases of hospital- and community-associated infections and constitute a major burden on society world-wide (DeLeo et al., 2010; Otto, 2012). Resistance to beta-lactam antibiotics, which are the most widely used group of antibiotics, makes MRSA infections very difficult to treat and only very few alternative drugs are currently available for treatment. Vancomycin is the first-line treatment of severe MRSA infections (DeLeo et al., 2010), it is however less efficient, requires intravenous administration, and resistance has already been reported in the form of vancomycin-resistant *Staphylococcus aureus* (VRSA) (Robinson and Enright, 2003).

MRSA has primarily been considered as a hospital-associated (HA) pathogen but has emerged in the community in the recent decades (DeLeo et al., 2010; Verkade and Kluytmans, 2013). Community-associated (CA) MRSA differ from the HA-MRSA (Diep and Otto, 2008) as they show a more virulent phenotypic profile. They frequently produce the Panton-Valentine leukocidin, which is a toxin often associated with severe skin infections (Vandenesch et al., 2003).

More recently a third group of MRSA has emerged. This group has been associated with livestock and especially pigs (Smith and Pearson, 2011; Voss et al., 2005). These strains termed livestock-associated (LA) MRSA make up yet a different profile compared to HA- and CA-MRSA. LA-MRSA has been identified to be less aggressive, not encoding many of the toxins often associated with *S. aureus* (Hallin et al., 2011). LA-MRSA is often found in related to colonization, but can in some cases cause illness of different severity in both animals and humans, and LA-MRSA is in fact the most frequent cause of porcine skin infections (Cuny et al., 2010; van Duijkeren et al., 2004). LA-MRSA differ in their resistance pattern compared to the human-associated MRSA, by expressing resistance to tetracycline (Price et al., 2012). Tetracycline is repeatedly used in the swine production industry and is most likely the driving force for tetracycline resistance in LA-MRSA

(Schijffelen et al., 2010). Many of the LA-MRSA strains belong to lineage multilocus sequence type 398 (ST398), and they show a broader host range compared to most other *S. aureus* lineages (Verkade and Kluytmans, 2013). A whole genome sequenced LA-MRSA ST398 isolate has shown genotypic traits that could imply increased ability to take up foreign DNA as it contains multiple integrative conjugative elements combined with the absence of a type I restriction and modification system (Schijffelen et al., 2010). It has been postulated, that LA-MRSA originated as methicillin-sensitive *S. aureus* in humans and were transferred to pigs where they acquired methicillin and tetracycline resistance via the uptake of mobile genetic elements, and then transferred back to humans (Price et al., 2012).

Even though LA-MRSA has been the subject of several studies most of them are based on survey data (E M Broens et al., 2011b; Els M Broens et al., 2011) and only few have investigated bacterial ecology in the porcine reservoir (Moodley et al., 2012; Tulinski et al., 2013). The potential of LA-MRSA transmission and adaption is still unknown and further investigations into why ST398 has successfully colonized so many different animal species are needed to help understand how we might prevent similar problems in the future.

Objectives and research approach

The overall objective of this study was to identify genes of importance for the emergence and spread of LA-MRSA in food animals. The focus was on lineage multilocus sequence type 398 (ST398) as it has become the most commonly reported MRSA strain associated with livestock in the recent years (Smith and Pearson, 2011). Identification of such genes could assist in a better understanding of the ecology of ST398 in the porcine reservoir and facilitate the identification of targets in intervention strategies. The following objectives were defined:

1. Develop high-throughput approach for bacterial characterization.
2. Perform a comprehensive phenotypic and genotypic characterization of LA-MRSA ST398.
3. Investigate genes important for LA-MRSA ST398 colonization in the porcine reservoir.

Most of the information on ST398 colonization known to date has been based on survey data. In this project a different approach has been taken. High-throughput methods were applied to simultaneously link all genes within a genome to phenotypes. This approach may facilitate the discovery of new gene function and can highlight which genes are essential for bacterial survival in specific environments.

Manuscript presentation

The thesis includes three manuscripts, each of which represents studies performed to answer the objectives. In manuscript I and II a high-throughput method was developed and used for investigation of genes important for LA-MRSA ST398 survival in the porcine reservoir, as this reservoir contributes to the on-going spread and adaptation of LA-MRSA. In manuscript III a bioinformatic tool was constructed to define virulence profiles of *S. aureus* using whole genome sequence data. This database will be freely available online and can contribute to the identification of virulence genes in LA-MRSA isolated from different host reservoirs.

Manuscript I entitled, *Genome-wide high-throughput screening to investigate essential genes involved in methicillin-resistant Staphylococcus aureus Sequence Type 398 survival* represents a method paper. The **aim** of the work was to generate a high complexity transposon mutant library and assess the application of Transposon directed inserted site sequencing (TraDIS) in LA-MRSA ST398. The manuscript describes the generation and verification of a genome-saturated transposon

mutant library. The data obtained in the manuscript verifies that the high-throughput genotypic approach TraDIS can be assed for transposon insertion site identification in *S. aureus.* Genes essential for LA-MRSA ST398 survival under laboratory conditions were identified. In addition the manuscript describes an evaluation of the approach, where the mutant library was screened in whole porcine blood *in vitro* and mutant composition pre- and post- selection was compared. ***Manuscript accepted for publication in PloS One (publication date Feb. 12th 2014).***

Manuscript II entitled, *Genes important for survival of livestock-associated methicillin-resistant Staphylococcus aureus Sequence Type 398 in the porcine reservoir* was carried out as part of an externship conducted at Department of Veterinary Medicine, University of Cambridge, UK. The **aim** of the study was to identify genes important for LA-MRSA ST398 survival on porcine skin and nasal epithelium, as these locations are believed to be relevant habitats for LA-MRSA ST398. Two *ex vivo* models, using porcine nasal epithelium and porcine skin tissue, were developed. By screening the transposon mutant library (generated in manuscript I) within the models, a number of genes important for bacterial survival on porcine explants, were identified. These genes could represent targets for de-colonization, which could help prevent further spread and adaptation of LA-MRSA ST398. ***Manuscript in preparation.***

Manuscript III entitled, *Identification of virulence genes in whole genome sequenced Staphylococcus aureus* describes the construction of the *S. aureus* VirulenceFinder database. The **aim** of the study was to develop a database which can define virulence profiles in *S. aureus* using sequence data. The generated database can be applied for identification of previous described virulence genes. *S. aureus* VirulenceFinder is part of the tool package generated for the Centre for Genomic Epidemiology (CGE) (www.genomicepidemiology.org). CGE aims at generating

bioinformatic tools for handling whole genome sequence information, useful for outbreak investigation, epidemiological surveillance, source tracking and diagnostics. The service is publically available through web servers. ***Manuscript in preparation.***

Chapter 1

Introduction

1. *Staphylococcus aureus*

Staphylococcus aureus is a facultative anaerobic Gram-positive coccus that normally is arranged in grape-like clusters. They are non-motile and often golden-yellow pigmented cells. The primarily colonization sites are the anterior part of the nares and skin surfaces. The organism is non spore forming but is resistant to dry conditions and high salt concentrations, which is essential when colonising the skin surface. There are more than 50 species and sub-species of Staphylococci of which *S. aureus* is often associated with pathogenicity in humans. *S. aureus* is distinguished from the other species by its ability to clot blood plasma by the action of the enzyme coagulase (Foster, 2009; Harris et al., 2002).

S. aureus is a commensal commonly found to colonize several animal species, including humans. Around 20-40% of the human population are carriers and some humans are intermediate carriers whereas others are persistent carries (Foster, 2009; Williams, 1963). The difference in colonization implies that host factors are important elements for successful bacterial colonization.
If given the opportunity *S. aureus* can cause infection, most commonly at sites of lowered host resistance such as damaged skin or mucosal membranes. The bacteria possesses a large number of cell-associated and extracellular virulence factors, some of which contribute to the ability of the organism to overcome the host immune defence and to invade and colonize the tissue (Foster, 2009). *S. aureus* can cause a wide variety of infections ranging from superficial soft tissue and skin infections like pimples, boils and abscesses to severe systemic infections like bacteraemia, endocarditic, pneumonia and toxic shock syndrome (Otto, 2012).

S. aureus has a clonal population structure and it does not undergo extensive genomic recombination but rather random nucleotide mutations and horizontal gene transfer are the main factor for diversity (Holmes and Zadoks, 2011).

Due to the clinical relevance a large number of typing methods with different discriminative power, are being used to study the population genetics of *S. aureus*. Pulse-field gel electrophoresis (PFGE) is a method that can detect rapidly accumulating genetic variation by looking at a genetic fingerprint. The methods can be used to study outbreaks or the phylogeny of small populations (Tenover and Arbeit, 1995). s*pa*-typing and multilocus sequence typing (MLST) are other methods looking at more slowly accumulating genetic variations and are used for the investigation of global epidemiology and population genetics. s*pa*-typing is based on DNA sequencing of the polymorphic 24 base pair tandem repeat of the 3-prime end of the *S. aureus*-specific staphylococcal protein A (*spa* gene). In MLST typing, a bacterial isolate is assigned an allelic profile relating to nucleotide sequences of seven housekeeping genes and based on the allelic profile the isolate will be given a sequence type (ST). If isolates differs in only one allele they will be given different sequence types but are said to be in the same clonal complex (CC) (Urwin and Maiden, 2003; Fitzgerald et al. 2001). MLST is less discriminative than PFGE and s*pa*-typing and a multilocus sequence type (ST) linage can contain several (often related) s*pa*-types.

In the recent years next generation sequencing has become increasingly available. By using a whole genome sequencing approach, one will obtain all the genetic information, and this can potentially be translated into information about population structure, genetic variation, outbreak investigation, global epidemiology, typing, and diagnostics (Hall, 2007; Price et al., 2012; Zankari et al., 2012).

1.1 Methicillin-resistant *Staphylococcus aureus* (MRSA)

S. aureus is known for its potential to adapt to a selective pressure from antibiotics. In 1948 the first report of a penicillin-resistant *S. aureus* was published, which was only shortly after the introduction of penicillin in the clinical practice (Barber and Rozwadowska-dowzenko, 1948). In the late 1950s methicillin was introduced as treatment of infections with penicillin resistant strains and soon after, in 1961, methicillin-resistant *S. aureus* (MRSA) was identified in the UK, and was after that recognized as a hospital-associated pathogen worldwide (DeLeo et al., 2010; Jevons et al., 1963; Otto, 2012).

Methicillin-resistance is in staphylococci conferred by the carriage of the Staphylococcal Cassette Chromosome *mec* (SCC*mec*). The SCC*mec* cassette is a mobile genetic element that includes the *mecA* gene encoding the penicillin binding protein (PBP) 2a, which shows low affinity for beta-lactam antibiotics such as penicillin and methicillin, and allows, even in the presence of such antibiotics, the cell to produce a functional cell wall and therefore survive beta-lactam treatment (Hartman and Tomasz, 1981). The structure of the SCC*mec* cassette is diverse and is classified based on the combination of the cassette chromosome recombinase (*ccr*) the *mec*-class. The *ccr* genes are responsible for mobility of the element, whereas the *mec*-class relates to beta-lactam resistance and its regulation. In addition to the *ccr* and *mec* genes some non-essential junkyard regions are included in SCC*mec* typing (Kondo et al., 2007). Some SCC*mec* cassettes include other resistance genes besides the *mecA* gene, which encode resistance to other antibiotics and/or heavy metals. Such genes are part of integrated copies of plasmids or transposons (Grundmann et al., 2006).

Recently a divergent *mecA* homologue, termed *mecC* (formerly $mecA_{LGA251}$), was identified in *S. aureus* isolates from dairy cattle in United Kingdom (García-Álvarez et al., 2011). *mecC* shows 70 % nucleotide identity to *mecA* and is located in a novel SCC*mec* element designated *SCCmec* type

XI. Routine culture and susceptibility testing will identify *S. aureus* isolates containing *mecC*, however, molecular confirmatory methods will not identify them as MRSA (García-Álvarez et al., 2011). *S. aureus* containing *mecC* has been found in a range of multilocus sequence type lineages isolated from humans and other animal species (Cuny et al., 2011; García-Álvarez et al., 2011; Harrison et al., 2013; Laurent et al., 2012).

Colonization with MRSA is normally asymptomatic in healthy individuals but elderly, immune-compromised people, and post-operation patients have a significant higher risk for developing symptomatic infections. It has been estimated that patients with MRSA have a pronounced higher risk of mortality compared to other patients (Grundmann et al., 2006; Memorial, 2005).

1.1.1 Hospital- and Community-associated MRSA

MRSA infections were traditionally hospital-associated (HA-MRSA) and have been a major public health issue for the last 50 years causing severe nososcomial infections worldwide. However in the past approximately 10-15 years, an increased number of infections due to community-associated MRSA (CA-MRSA) has been reported in Europe and the US (DeLeo et al., 2010; Verkade and Kluytmans, 2013). CA-MRSA infections are mainly seen in healthy individuals with no recent contact to the healthcare system.

There are about ten human multilocus lineages distributed globally (CC1, CC5, CC8, CC12, CC15 CC22, CC25, CC30, CC45, CC51) of which CC1, CC5, CC8, CC22, CC30 and CC45 contain the most common MRSA strains (Feil and Cooper, 2003; Holmes and Zadoks, 2011).
Typically CA-MRSA is genetically distinct from HA-MRSA (Diep and Otto 2008). They frequently carry the SCC*mec* type IV or V and they often harbour the Planton-Valentine leukocidin

(PVL) that is associated with skin and soft tissue infections (DeLeo et al., 2010; Vandenesch et al., 2003). The SCC*mec* types IV and V are smaller in comparison to the SCC*mec* types I, II and III often carried by HA-MRSA, which may facilitate a lower metabolic cost and a more efficiently transferred element between CA strains.

The genetic differences seen between HA-MRSA and CA-MRSA correlate with the host environment (David and Daum, 2010; Diep and Otto, 2008). HA-MRSA often infects immune-compromised individuals in hospital settings where a high antibiotic selective pressure are present, whereas CA-MRSA, infecting healthy individuals, requires fewer resistance genes, additional virulence factors, and a growth advantage in the more competitive environment outside the hospital setting.

1.1.2 Livestock-associated MRSA

A third emerging branch of *S. aureus* has been identified in association with livestock animals. The isolation of MRSA from animals was first reported in 1972. This was following the detection of MRSA in milk from mastitic cows (Devriese et al., 1972) but was at that time most likely associated with human to animal transmission of an MRSA strain acquired by the farmer during hospitalisation. Since then livestock-associated MRSA (LA-MRSA) has been identified in various animals in several European countries, the US and Asia (Smith and Pearson, 2011). Population genetic studies have identified certain genotypes to be associated with specific host species. Multilocus sequence type ST71, ST97, ST126, ST133 and ST151 are often found among ruminants, and are the major course of bovine mastitis, whereas ST5 is often associated with poultry and ST9, ST433, ST398 are often found in pigs (Armand-Lefevre, 2005; Holmes and Zadoks, 2011; Lowder et al., 2009; Moodley et al., 2012; Nickerson, 2009).

1.1.2.1 Clonal Complex 398

The first LA-MRSA isolated from pigs was reported in France in 2005 (Armand-Lefevre, 2005) and the same clonal complex (CC398), was discovered as being widespread in pigs in the Netherlands (Voss et al., 2005). Multilocus sequence type 398 (ST398), belonging to CC398, has shown a broader host-spectrum compared to most other MRSA, and have until now been found in pigs, cattle, veal calves, horses, poultry, turkeys, companion animals as well as humans (Verkade and Kluytmans, 2013).

The main reservoir for LA-MRSA CC398, as well as for MSSA ST398 (Hasman et al., 2010), seems to be pigs and many of the early studies on swine LA-MRSA CC398 were carried out in the Netherlands. Here the prevalence of HA-MRSA is generally low but LA-MRSA CC398 is found to be widespread (Smith and Pearson, 2011; Voss et al., 2005). Dutch prevalence studies report that the number of CC398 positive farms varies from 23 % to 81 %, whereas the prevalence in individual pigs varies from 11 % to 39 % (E M Broens et al., 2011a; de Neeling et al., 2007; van Duijkeren et al., 2008). Only around five years after the first LA-MRSA isolate was found in pigs, a prevalence study examining swine breeding farms in Europe found swine MRSA to be present in pig facilities in Austria, Belgium, Cyprus, the Czech Republic, the Netherlands, Denmark, Finland, France, Germany, Hungary, Italy, Luxembourg, Poland, Portugal, Slovakia and Spain (Food and Authority, 2010). The prevalence of MRSA CC398 in pigs at slaughter in Denmark in 2012 was estimated to 77 % (Agersø et al., 2012). LA-MRSA CC398 has also been found in high prevalence in veal calves, with 28 % of the calves and 88 % of the farms tested being positive (Graveland et al., 2010).

The main risk factors for human colonization with CC398 are direct exposure to pigs and veal calves or sharing a household with people who are in direct contact with these animals (Graveland

et al., 2011; van den Broek et al., 2009). In addition living in regions with high densities of livestock, has been proposed as a risk factor for human colonization with CC398, despite not having direct contact with pigs (Feingold et al., 2012). By the end of 2008, 42 % of all newly identified MRSA strains in humans in the Netherlands belonged to CC398 (Verkade and Kluytmans, 2013). MRSA was found in 232 human cases in Denmark in 2012, which was an increase from the previous years (42 in 2009, 111 in 2010, and 164 in 2011) despite no targeted screening for CC398 in 2012. The majority of the human cases identified in Denmark were from persons with documented close contact to pigs or household members to pig handlers (Agersø et al., 2012).

ST398 do not typically cause illness in pigs but nonetheless, it is the most frequent cause of porcine skin infections (Cuny et al., 2010; van Duijkeren et al., 2004). In addition *S. aureus* is frequently isolated from lesions observed during post-mortem inspection, especially from abscesses in lungs and udder (O'Mahony et al., 2005; Strommenger et al., 2006).
LA-MRSA ST398 infections in human are rare compared to infections caused by HA- or CA-MRSA, but has been seen of various severities. Human to human transmission is not as pronounced as for other MRSA types, but recently several studies have identified ST398 infections in humans with no previous contact to animals. These infections are primarily caused by methicillin-sensitive *S. aureus* (MSSA) ST398 (Agersø et al., 2012; Bhat et al., 2009).

The porcine reservoir constitutes an important niche for adaptation and transmission of LA-MRSA ST398. Therapeutic treatment of pigs with oxytetracycline and treatment of complete flocks with prophylactic oxytetracycline are both common practices in farming (van Duijkeren et al., 2004). The vast majority of *S. aureus* ST398 isolated from pigs show tetracycline resistance, and tetracycline resistance is most likely one of the responsible agents for the selection of ST398

isolates (Hasman et al., 2010; Schijffelen et al., 2010). A recent study has demonstrated that ST398 originated as MSSA in humans, was then transmitted to the pig reservoir where it acquired methicillin and tetracycline resistance and are now being transferred back to humans (Price et al., 2012). LA-MRSA CC398 harbors most often the smaller SCC*mec* cassette types IV or V. Specifically the SCC*mec* subtype Vc (2C5&5) encoding the cadmium-zinc resistance gene *czrC* are often found among LA-MRSA (Cavaco et al., 2011; Price et al., 2012). This emphasizes that the industrialization of pig production with high livestock densities, frequent animal transmission between farms and the use of antibiotics and heavy metals generate an ideal environment for this highly adaptable opportunistic pathogen.

During the past decade, ST398 has been rapidly emerging and has now become the most commonly reported MRSA strain found in association with livestock (Smith and Pearson, 2011). This highlights that better intervention strategies to control the spread are needed. However, the transmission between animals and between farms is most likely multi factorial, which complicates efforts to control spread of LA-MRSA ST398.

Even though LA-MRSA ST398 has been subject of epidemiologic research on farms and hospitals, various central questions remain unanswered. Profession and geographic regional location have been recognized as risk factors for human colonization, but specific genetic factors facilitating zoonotic transmission remains unidentified. The potential of LA-MRSA transmission and adaptation are still unknown and further investigations into why ST398 have successfully colonized so many different animal species are needed to help understand how we might prevent similar problems in the future.

The porcine reservoir seems to constitute a very important environment for ST398 adaption and transmission, however, little is known about which genes in the ST398 genome are important for

persistent porcine carriage. Identification of essential genes for porcine colonization could constitute targets for decolonization in an attempt to control the spread of ST398. A comprehensive phenotypic and genotypic characterization of LA-MRSA ST398 may help to better understand how ST398 has become the most dominant MRSA strain within livestock. Essential genes for porcine colonization could constitute markers in future epidemiology, surveying pathogens associated with pig farming. In addition, as ST398 displays resistance to various antibiotics, identification of essential genes for bacterial infection could highlight new potential targets for therapeutic agents.

2. Methods for bacterial characterization

Comprehensive insight into bacterial behaviour is crucial to overcome and prevent bacterial infections. To gain a better understanding of bacterial ecology in specific environments various methods can be applied. Some characterize only bacterial phenotypes and some only the genotypes, but to gain the most comprehensive bacterial characterization, a combination must be applied.

2.1 Model systems

Model systems mimicking natural environments are essential to understand bacterial behaviour. Various *in vitro*, *in vivo* or *ex vivo* models to investigate bacterial colonization and infection have been described. Such models are used in different ways to study adhesion, colonization, virulence and differences between strains.

As mentioned above *S. aureus* colonizes the nares and skin surfaces of several animal species. *In vitro* adhesion and colonization studies have been performed using desquamated nasal epithelial cells, skin corneocytes, epidermal keratinocytes and keratin (Corrigan et al., 2009; Moodley and Espinosa-Gongora, 2012; O'Brien et al., 2002). Cells originating from different hosts has been used for investigating host specificity of different lineages (Corrigan et al., 2009; Moodley and Espinosa-

Gongora, 2012). Whole blood and immune cells isolated from blood have been used to evaluate gene expression, bacterial survival and invasion, and host immune evasion (den Reijer et al., 2013; Malachowa et al., 2011) (manuscript 1). *In vitro* models are useful however such models lack several host components present in a natural *S. aureus* environment.

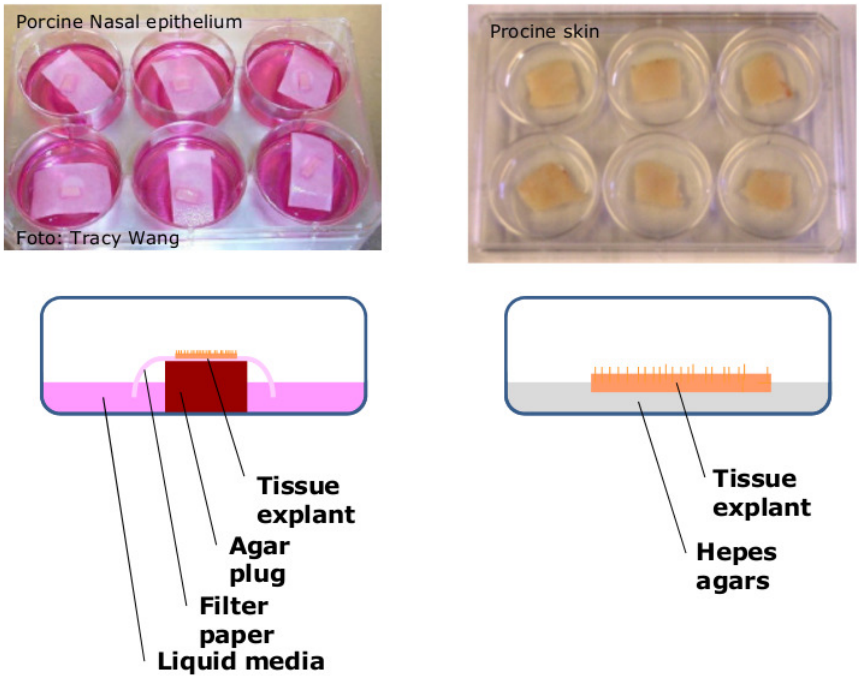
Therefore animal models are often used and they are useful models to study colonization and infection. Murine and rat models have been developed to mimic *S. aureus* colonization in humans (Kiser et al., 1999; Kokai-Kun, 2008), but as the murine nasal cavity is not a natural habitat for *S. aureus*, this model system is not optimal to study *S. aureus* colonization (González-Zorn and Senna, 2005). Pigs are, however, natural hosts and piglets have also been applied in colonization and persistent carriage studies (Els M Broens et al., 2011; Crombé et al., 2012; Moodley and Espinosa-Gongora, 2012).

The nematode *Caenorhabditis elegans* constitutes an alternative model to mammalian animal models. A variety of bacteria, including *S. aureus* can kill *C. elegans* and there seems to be a high degree of correlation between virulence factors required for nematode killing and virulence in vertebrates (Sifri et al., 2005). It has been used as a simple surrogate model to study infection (Bae et al., 2004; Begun et al., 2005). A hallmark feature of invasive *S. aureus* disease is its ability to cause bacteraemia which can lead to severe systemic infections. Murine and rabbit models have been used for studying bacteraemia caused by *S. aureus* (Benton et al., 2004; Coulter et al., 1998; Diep et al., 2008a; Mei et al., 1997).

An alternative to the animal models is freshly isolated tissues from various animals, in which *ex vivo* adhesion, colonization and infection studies and be performed (Tulinski et al., 2013) (manuscript II). The advantages of using tissue explants are that it resembles the natural host environment to a higher degree than *in vitro* systems, it is easier to set up, and more cost effective compared to *in vivo* animal models. Tissue from larger animals than rodents can be applied and as

pigs constitute a large natural reservoir for *S. aureus* this approach can help to highlight bacterial features important for successful porcine colonization. Examples of porcine *ex vivo* models are illustrated in Figure 1.

Figure 1: Porcine e*x vivo* models.



The figure illustrates two ex vivo porcine models. On the left, porcine nasal epithelium tissue explants are placed on filter paper on agar plugs. The filter paper feeds the tissue with media to sustain tissue viability. The exterior surface of the nasal tissue is infected with S. aureus. On the right porcine skin tissue explants are embedded in HEPES agar, leaving the skin surface exposed. S. aureus is inoculated onto the skin surface. Both models are applied in manuscript II.

Wild-type strains can be tested in model systems to investigate and compare phenotypes. To associate a bacterial genotype with a phenotype, mutants, lacking specific gene function can be tested in these model systems. This approach is applied in the investigation of genes important for bacterial colonization and infection. For this purpose single gene knockouts or complex mutant

libraries can be generated. Single gene knockouts are primarily used for investigation of specific genes with an already known or hypothesized function, whereas genome saturated mutant libraries facilitate a high-throughput screening for investigation of all genes within a genome simultaneously. Both are strong tools to associate genes with phenotypes.

2.1.1 Single gene knockout

Gene knockout is a genetic technique that, via different approaches, makes a gene in an organism non-functional. Once the gene of interest has been "knocked out" the so-called knockout organism can be tested in various functional assays to gain knowledge about gene function. Conclusions are drawn from the difference between the knockout organism and the wild-type strain.
The gene knockout approach is often used for investigation of pathogenic bacteria to show that a gene found in such strains encodes a product that contribute to the disease caused by the pathogen. The principal behind the use of genetic manipulation to investigate genes encoding virulence factors was first formulated by Standley Falkow in 1988 and was based on Koch's postulates (Falkow, 1988). The basic premise is that by using genetic manipulation a gene encoding a putative virulence factor can be inactivated and the mutant can be tested for loss of virulence capacity in a virulence model. In addition, Falkow described that it is necessary to demonstrate that by complementation the virulence capacity can be restored to wild-type level.

Gene manipulation in Gram-negative bacteria is generally easier compared to gene manipulation in Gram-positives, as the thick peptidoglycan cell wall harboured by Gram-positive bacteria hampers the manipulation. In addition, the majority of *S. aureus* strains possess a strong restriction modification barrier that hinders the uptake of foreign DNA (Monk et al., 2012; Monk and Foster, 2012). The first step in generation of knockout mutant is to introduce a vector, which includes a

sequence identical to the gene of interest or its flanking regions, into the strain of interest. For vector DNA to be transferred into *S. aureus* and not undergo destruction by the restriction and modification systems, the DNA must pass through a modified *S. aureus* laboratory strain (Bae et al., 2008). *S. aureus* RN4220 is a laboratory strain that has been mutated in one of the restriction systems, which allows it to take up foreign DNA, which has been cloned in for example *E. coli*. Once the foreign DNA has been replicated by RN4220 the methylation pattern will be of *S. aureus* origin and thereby acceptable for uptake by most wild-type *S. aureus* strains.

Introduction of DNA can be performed via for example transformation or transduction. The most widely used approach is transformation, which is the transfer of free DNA. Transformation can be difficult in *S. aureus* and have mostly been done in laboratory strains like in *S. aureus* RN4220 described above. However, new methods are being developed for *S. aureus* gene manipulation (Monk and Foster, 2012). Transduction is based on the usage of a virus that infects bacteria, a so-called bacteriophage. When a bacteriophage infects a bacterial cell, it utilises the cell machinery for viral DNA replication. Once viral DNA replication is complete, the virus transmits to other bacterial cells through a lytic cycle, killing the host cell. In this process bacterial DNA can, by accident, be packed into the viral capsid and once the bacteriophage infects a new cell, bacterial DNA can be transferred from one bacterium to another. Transduction is only an intra-species approach as bacteriophages are species specific and enter the bacterial cells by recognition specific surface bound receptors (Novick, 1991). Generally DNA transfers with low frequencies in *S. aureus*, complicating gene manipulation in this species.

Once the vector DNA has been introduced, homologous recombination can occur and the gene of interest can be replaced with a selection marker and in that way inactivated. The mutant is tested in parallel to the wild-type in for example an infection model. If the mutant displays reduced virulence compared to the wild-type, it can be postulated that the gene, which has been “knocked out”,

encodes a virulence factor. In addition to testing the mutant in various assays, complementation experiments must be performed for a definitive conclusion. Such complementation can be done by reintroducing gene function in the mutant or by testing the gene function in a surrogate host lacking pathogenicity.

2.1.2 Transposon mutant libraries

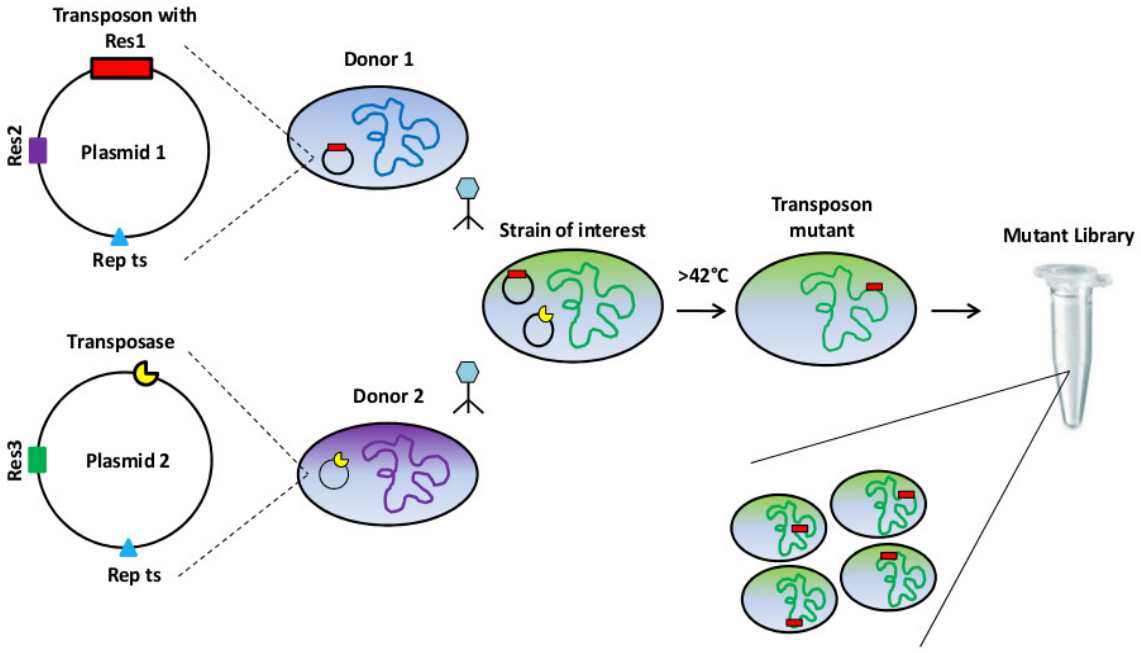
Screening of a transposon mutant library is similar to the gene knockout approach. However, it is a high-throughput method that facilitates investigation of all genes within a genome simultaneously. A transposon mutant library is composed of numerous mutants, where each mutant has intergrated a transposon at a random position within the bacterial genome. The approach is based on a negative selection strategy, where transposon inserts into functional genes will result in mutants with attenuated fitness, or a complete inability to survive, and recovery of only those mutants with inserts in non-essential genes.

Transposon mutagenesis can be used for building a library of random mutants (Figure 2). Typically a two-plasmid-system is used for generation of transposon mutant libraries in *S. aureus* (Bae et al., 2004; Chaudhuri et al., 2009; Fey et al., 2013) (manuscript I). One of the plasmids carries a mariner transposon, which is required for insertion in *S. aureus* genomes, and another plasmid carries a transposase facilitating the transposition event. Both plasmids contain a temperature sensitive origin of replication, which allows for removal of the plasmids by increasing growth temperature (Bae et al., 2008).

The transposon inserts randomly into the bacterial genome and when a gene is inserted with a transposon the gene function is potentially disrupted. If a gene essential for cell survival is disrupted the mutant will die, and as a result no essential genes will be present in the mutant library.

The transposition event happens once per cell as only one transposon is present within each cell. Every cell in the library will by chance contain a transposon at a unique genomic position. The transposon will contain a selection marker which will enable selection of the transposon mutants. A transposon mutant library can be generated via liquid transposition (Figure 2) resulting in a pool of thousands of mutants or transposition can be conducted on solid phase. Solid phase transposition enables separation of the individual mutants but is more laborious than liquid transposition and often results in mutant libraries containing fewer mutants compared to mutant libraries generated by liquid transposition.

Figure 2: Generation of a transposon mutant library.



The figure illustrates a schematic presentation of how a transposon mutant library can be generated in S. aureus using a two-plasmid-system. Plasmid 1 carries the transposon which includes a resistance marker (Res1), a temperature sensitive origin of replication (Rep ts), and a second resistance marker (Res2) carried on the plasmid backbone. Plasmid 2 carries a transposase essential for the transposition event, a temperature sensitive origin of replication (Rep ts), and a resistance marker (Res3) carried on the plasmid backbone. The plasmids are carried by two different donor cells (Donor 1 and Donor 2) and are via two separate rounds of transduction introduced into the strain of interest.

After the transposition event (illustrated as liquid transposition) the culture is grown at high temperatures to facilitate plasmid loss (no plasmid replication >30°C). The result is a transposon mutant library consisting of mutants with one transposon inserted randomly once per bacterial genome.

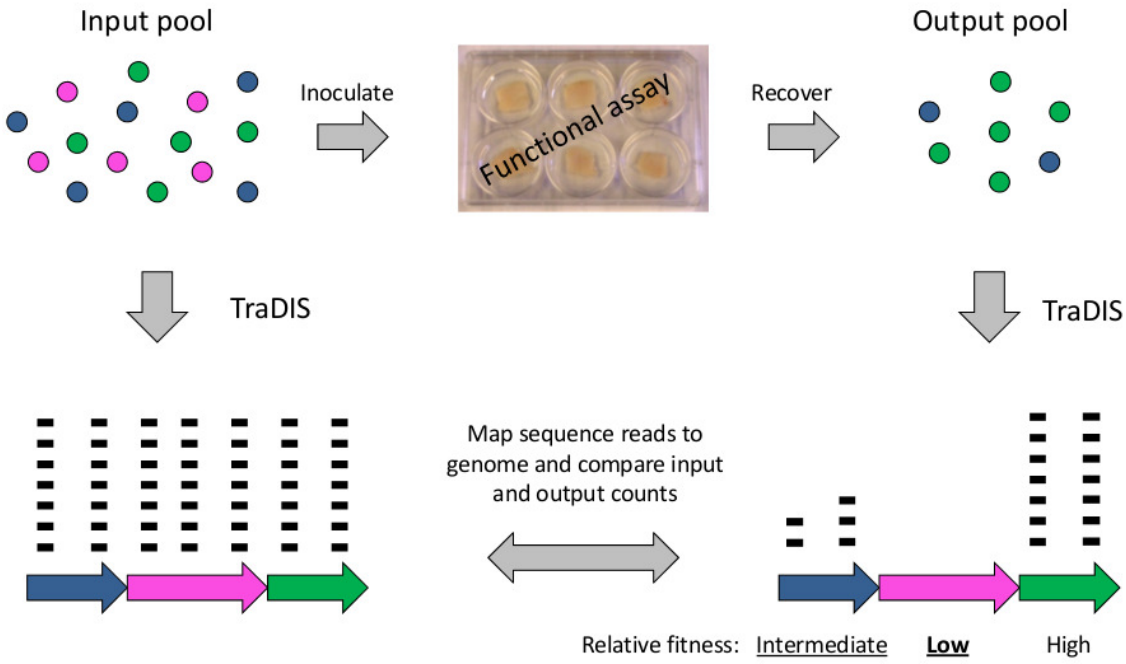
Once the library is generated and validated the transposon insertion sites must be identified to characterize the mutant pool. Various genetic techniques can be applied to identify the flanking regions of the transposon insertion sites.

For an optimal output it is desirable that the genotypic approach is high-throughput to match the large number of mutants generated using transposon mutagenesis. A high-throughput genotypic approach has been developed by Chaudhuri *et al.* (Chaudhuri et al., 2009). It is a DNA microarray and PCR-based method called Transposon Mediated Differential Hybridization (TMDH). Once the mutant library has been generated, genomic DNA is digested with a restriction enzyme and labelled RNA run-offs are produced from outward facing promoters integrated into the flanking ends of the transposon. The labelled RNA is hybridized to a tiling oligonucleotide microarray. Probes that are downstream of the transposon give a positive "on" signal while other probes give an "off" signal. Small genes (<300 bp) have fewer transposon insert possibilities compared to larger genes. Such small genes are only covered by a low number of probes resulting in a poor signal and they can be problematic to detect as "on" signals. For these genes a laborious PCR step using a transposon specific primer is necessary. In addition, in some microarray features the distinction between positive and negative signals can be difficult. The TMDH approach was applied in the first comprehensive study identifying essential genes in *S. aureus* (Chaudhuri et al., 2009).

Another genotypic strategy, superior to the microarray approach, is based on high-throughput sequencing. Langridge *et al.* developed a system named Transposon Directed Insertion site Sequencing (TraDIS), which uses a transposon specific primer, enabling sequencing of the genomic target region flanking the transposon insertion sites (Langridge et al., 2009). The primer is designed

in such a manner that the first 10 bp in each sequence read is of transposon origin. As the method is of a "digital" nature, any sequence read that have the 10 bp transposon tag sequence with adjacent genomic sequence is almost certainly an indication of the transposon insertion site. Importantly, this sequencing procedure not only identifies essential genes under different environmental conditions, but also provides an estimate of the relative importance of gene function (Langridge et al., 2009). Statistical analysis can be performed with tools like R for a quantitative comparisons between samples (Anders and Huber, 2012).

Figure 3: TraDIS approach.



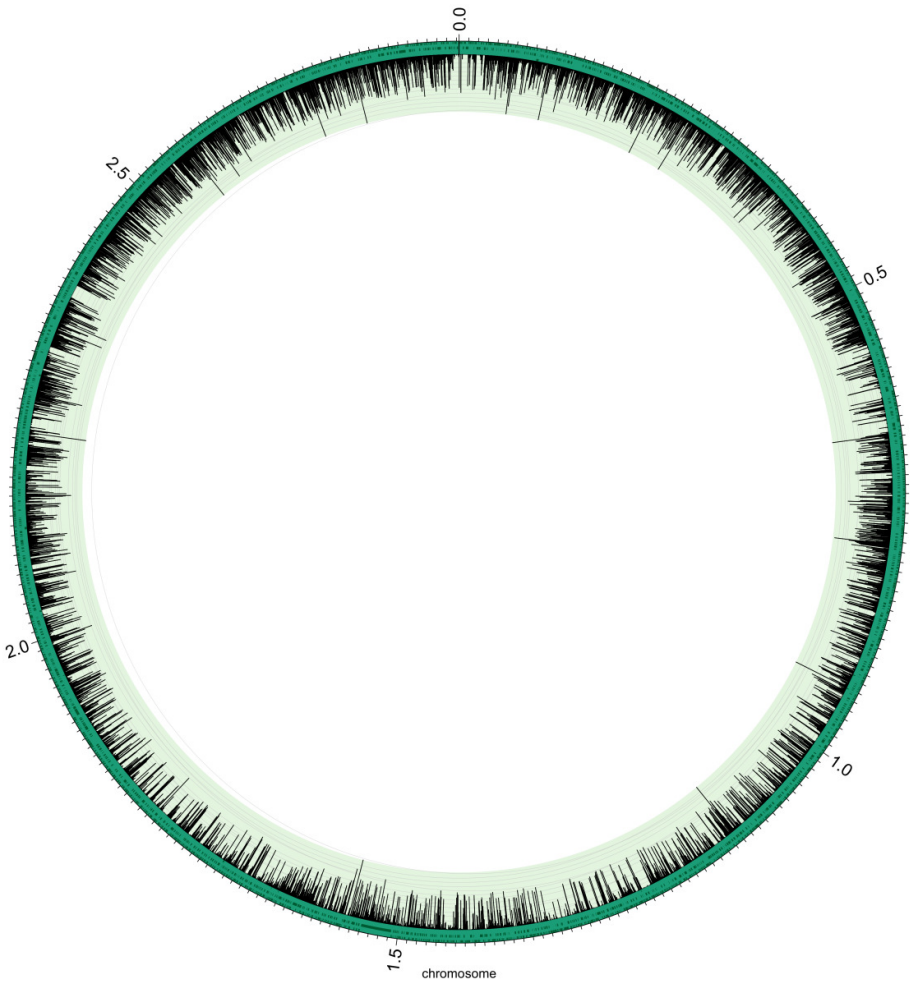
The figure illustrates how the TraDIS approach was applied in manuscript 1 and 2. An input pool of the transposon mutant library was screened in a functional assay and an output mutant pool was recovered. DNA from both input and output, representing mutants pre- and post- selection, were purified and sequenced on the Illumina platform. Sequence reads from the input and output samples were mapped to a reference genome and compared. Mutants present in input but absent in output (pink) represent a gene that is essential for survival in the specific environment defined in the functional assay. Mutants recovered in decreased numbers in output compared to input (blue) represent a gene that is

to some degree important for survival in the defined environment. Whereas mutants present in comparable numbers in input and output (green) represent a gene that is non-essential in the functional assay.

The sequencing approach has been used by Langridge *et al.*, Khatiwara *et al.*, Pickard *et al.* and Chaudhuri *et al.* to study essential and conditional essential genes in *Salmonella* Typhi and *Salmonella* Typhimurium (Chaudhuri et al., 2013; Khatiwara et al., 2012; Langridge et al., 2009; Pickard et al., 2013). Manuscript I describes, for the first time, the use of TraDIS in an *S. aureus* isolates and Figure 3 illustrate how TraDIS was applied in manuscript I and II.

The optimal mutant library will compose a genome-saturated library. Such a library will consist of mutants, each containing one transposon insertion site at a unique position, increasing the likelihood of every functional gene being disrupted. To verify that the transposon has been inserted throughout the genome within the mutant pool, each insertion site can be identified and mapped against a reference genome, revealing potential "hot spots" or confirming that all the genes, within the genome have been inserted with a transposon. This can be visualized by a genome atlas as shown in Figure 4.

Figure 4: Visualization of genomic insertion sites – genome atlas.



The figure shows a visualisation of the genome-saturated transposon mutant library generated in manuscript 1. The LA-MRSA ST398 S0385 reference genome (GenBank accession no. AM990992) is illustrated by the green outer circle and the black spikes illustrate staked sequence reads aligned to the reference genome. Each sequence read represent a transposon insertion site. The figure illustrate that no "hot spots" for transposon insertion were identified and almost every region within the genome has been inserted with a transposon.

Both TMDH and TraDIS require a cut-off strategy separating essential and non-essential genes, as a transposon can insert into non-functional parts of the gene without disrupting gene function. In the TMDH approach, PCR foot-printing is used to evaluate the microarray screen to verify the true essential and true non-essential genes, if an exhaustive essential gene list is required (Chaudhuri et

al., 2009). Using TraDIS, a normalized insertion index can be calculated of each gene and plotted against insertion frequency. Such a plot will have a bimodal distribution with two peaks and the local minimum separating those can define a cut-off, separating essential from non-essential genes (Langridge et al., 2009) (illustrated in Figure 2 in manuscript I).

When interpreting the data it is important to recognize that environmental and experimental factors have unintended consequences for the output data. When comparing samples pre- and post-selection, only non-essential genes can be studied, as mutants with inserts in essential genes will not be present in the mutant library. In addition, transposon insertions may affect the expression of downstream genes or operons, causing polar mutations that lead to incorrect identification of essential genes in a defined environment. For definitive identification of gene function it is necessary to generate single gene knockouts and test those in the same functional assays used for the screenings. However, since a large number of genes are listed as having no known function and there is inconsiderable value in generating evidence for the phenotypes resulting from the possession of these genes, high-throughput methods can help to narrow the pool of genes to be investigated further.

2.2 Expression studies

Gene function is an important bacterial characteristic. However as many genes are only expressed under certain conditions, expression studies and information about the bacterial transcriptome are equally important. Knowledge about when a gene is expressed provides a deeper insight into gene function and this can be crucial when investigating new targets for antimicrobial agents.

Microarray was first described in 1995 and since then, the method has been used extensively in various studies (Schena et al., 1995). Microarray is a high-throughput automated approach consisting of multiple probes deposited or directly synthesized on a surface in an ordered fashion. The probes can be made of nucleic acid, proteins, carbohydrates or antibodies. On a DNA microarray nucleic acid probes are deposited on a planar glass surface, which is coated with a chemical reactive group to ensure efficient binding of the probes to the surface. To identify target genes, DNA samples are labelled chemically or enzymatic. The labelled samples are hybridized onto the array and washed. The remaining signal from the bound nucleic acids, specifically interacting with the probes deposited on the array, is measured using a confocal microarray scanner. Only probes hybridized with target DNA will give a signal thus identifying the gene with the related DNA motif in the sample (Huyghe et al., 2009).

DNA microarray can be used for identification of genes in multiple regions within a bacterial genome, or it can be applied to samples consisting of different genomes. However, only the genes with a target probe will be identified. A variety of genes, e.g. genes encoding virulence factors, phylogenetic markers or antibiotic resistance, have been employed on microbial characterization microarrays (Hallin et al., 2011; Sung et al., 2008). Microarray have been applied widely in expression studies, where mRNA is isolated from a bacterial culture in a defined environment and translated back into cDNA before quantified on the microarray (den Reijer et al., 2013; Malachowa et al., 2011). The transcriptomic approach can emphasize if specific genes are up- or down-regulated under specific physiological conditions. A significant change in transcript will highlight the importance of gene function under the environmental conditions. In addition, microarray can be used for other applications like comparative genome hybridization, microbial community characterization and single nucleotide polymorphism (SNP) analysis (Huyghe et al., 2009).

2.3 Whole genome sequencing and comparative genomics

Several bioinformatic tools have been developed to characterize bacteria. Such tools are primarily based on a homology strategy, where nucleotide identity to already defined genes are used to describe new gene function (Hall, 2007). In recent years, whole genome sequencing (WGS) has become increasingly available. There have been huge improvements in sequencing technologies and the cost has gone down significantly. This gives rise to a new approach within diagnostics and surveillance, where WGS can be utilized for species identification, evolutionary clustering (Price et al., 2012), identification of resistance (Zankari et al., 2012) and virulence markers, just to mention a few of the many applications. WGS has enormous potential as it contains all the information, however, the biggest challenge with the appliance of WGS is to interpret the large amount of data retrieved with this technology. To translate large amounts of DNA sequences into functional information requires bioinformatic tools that are standardized and simple to use.

The improvements within WGS have boosted the approach of comparative genomics, where bacteria, as well as other organisms, can be compared on a genome level (Price et al., 2012). A complete bacterial genotype obtained by WGS can stand alone (Schijffelen et al., 2010), but by comparing the complete genotype from different strains living in different environments, genetic traits can potentially be identified, which can explain the successful colonization of a given environment.

3. Characterization of *Staphylococcus aureus*

The methods introduced have been used for characterization of *S. aureus* and various results from previous studies will be presented in this section. The focus will mainly be on *S. aureus* CC398.

3.1 Essential genes

Transposon mutagenesis has been used to define essential genes in *S. aureus* (Bae et al., 2004; Chaudhuri et al., 2009; Fey et al., 2013) (manuscript I). A schematic overview of the results obtained in the different studies can be found in manuscript I supporting figures Table S2.

Bae *et al*. generated an unsaturated transposon mutant library consisting of 10,325 transposon mutants. The insertion sites were amplified by PCR and sequenced and 450-550 genes were identified as essential for *S. aureus* strain Newman under laboratory conditions (Bae et al., 2004).
Chaurhuri *et al.* generated the first genome-saturated transposon mutant library generated in *S. aureus*. The library was generated in *S. aureus* strain SH1000 and 351 genes were proposed essential for growth under laboratory condition. The TMDH approach was used for identification of transposon insertion sites (Chaudhuri et al., 2009).
Fey *et al.* identified 579 open reading frames which were not disrupted by a transposon in *S. aureus* strain JE2 (derived from a USA300 isolate) and these genes were proposed as essential for growth under laboratory conditions. They used a high-throughput sequencing method to identify transposon insertion sites (Fey et al., 2013).

The proposed essential genes are classified into functional categories and represents genes involved in DNA and RNA metabolism, protein synthesis, cell envelope, carbon metabolism, respiratory pathways, nucleotide biosynthesis and metabolism and cofactors. A fairly large portion of the essential genes are of unknown function underlining the need for further investigations.
Some differences are seen between the lists of proposed essential genes in *S. aureus*. Any attempt to define the minimum set of essential genes will inevitably be influenced by the conditions under

which the experiment is performed. A gene may be scored as essential in a particular experiment because it is required for survival following exposure to a particular stress inherent in the methods or because it is involved in uptake or metabolism of the particular nutrients provided in the growth media. The differences found in the studies presented could be a result of differences in methodology, experimental conditions or true differences between strains.

3.2 Adhesion and colonization

3.2.1 The nares.

Adhesion is the first step in colonization and infection. *S. aureus* colonizes both the nares and skin surfaces of several animal species, including humans, but the most frequent site of carriage is the moist squamous nasal epithelium of the anterior nares. It has been postulated that the ability of *S. aureus* to adhere to the nares is widely determined by its ability to adhere to desquamated cells on the epithelial surface of the nasal vestibules (Corrigan et al., 2009; Foster, 2009). As "only" 20-40% of the human population are intermediate or persistent carriers of *S. aureus*, it is clear that host factors play an important role in colonization (Foster, 2009).

Clumping factor B (ClfB encoded by the *clfB* gene) and iron regulated surface determinant protein A (IdsA encoded by the *isdA* gene) are surface expressed proteins utilized by *S. aureus* for adhesion to desquamated epithelial cells *in vitro*. In addition, they have been shown to be important for colonization of the nares of rodents *in vivo* and in humans in the case of ClfB (Clarke et al., 2004; Schaffer et al., 2006; Wertheim et al., 2008). ClfB binds human type 1 cytokeratin 10 found on the surface of human nasal cells (O'Brien et al., 2002). IsdA is only expressed under iron-limited conditions, which the bacterium often encounters in a host environment. It binds a number of different substrates including fibronectin, fibrinogen and several proteins associated with the cell

envelope of desquamated nasal epithelial cells such as cytokeratin 10. Both proteins constitute attractive candidates as antigens for a colonization-blocking vaccine (Clarke et al., 2006, 2004). By testing gene knockouts *in vitro* Corrigan *et al.*demonstrated that *S. aureus* strain Newman adherence to human desquamated nasal epithelial cells is multifactorial and involves the serine-aspartic acid repeat surface proteins SdrC and SdrD as well as ClfB and IsdA (Corrigan et al., 2009).

In addition to proteinaceous adhesins, cell wall teichoic acids (WTA) and capsular polysaccharides have been shown to be involved in nasal adhesion. Further investigations have shown by expression analysis that various adhesion factors are expressed at different stages in nasal colonization. Genes involved in WTA biosynthesis are primarily expressed in the initial stage of colonization whereas *clfB* and *idsA* are up-regulated at a later stage (Burian et al., 2010; Kiser et al., 1999).

S. aureus surface protein (Sas) G and X have also been demonstrated to bind nasal epithelial cells (Li et al., 2012; Roche et al., 2003). SasX is encoded on a mobile genetic element (MGE) occurring predominantly in ST239 MRSA strains, which are the most frequent source of MRSA infections in Asia. It has been shown to contribute to colonization, biofilm formation, immune evasion and virulence in animal infection models (Li et al., 2012; Otto, 2012).

Pigs constitute an important reservoir for the spread and adaption of *S. aureus* ST398. Current knowledge on colonization and transmission of LA-MRSA in pigs is limited and mainly based on observational field surveys, but recently, *in vivo* pig colonization models have been applied (E M Broens et al., 2011b; Els M Broens et al., 2011; Crombé et al., 2012; Moodley and Espinosa-Gongora, 2012). Transmission quantification studies indicated that LA-MRSA ST398 easily spread among pigs and once a pig is colonized, there is a high probability of persistence, even without antimicrobial use (Els M Broens et al., 2011; Crombé et al., 2012).

Incubation in pigs yielded however variable results, which is possibly due to unstable colonization. To obtain stable colonization, porcine *ex vivo* model systems constitute an excellent alternative to animal experiments. *Ex vivo* models mimics the natural host environment but under more controlled conditions. In addition, several explants from one animal can be derived allowing for replicates within the same genetic background. Such models have been developed by Tulinski *et al.* as well as in manuscript II (Tulinski et al., 2013) (manuscript II).

Tulinski *et al.* studied MRSA ST398 colonization of porcine nasal epithelial explants. Three different isolates were tested. One of the MRSA isolates was isolated from a carrier pig (S0462), one from a human case of endocarditis (S0385) and a beta hemolysin (Hlb) laboratory mutant (Δhlb = S0385-2) derived from the S0385 isolate. Different porcine colonization properties were observed suggesting differences in interaction of the different isolates and the tissue. All isolates showed an initial decline in attached cells, which could indicate bacterial adaptation to the environment. After prolonged incubation, the isolate from the carrier pig showed an increase in cell number. However, the bacteria number was unaltered for the isolate from the humane case of endocarditis and decreased for the corresponding *hlb* mutant (Tulinski et al., 2013).

Hlb is an exotoxin produced by *S. aureus* for complete lyses of red blood cells. Hlb production has also been demonstrated to damage keratinocytes and subsequent lead to colonization of skin (Katayama et al., 2013). MRSA Mu50, a human derived MRSA isolate, showed a similar pattern in porcine nasal colonization as the MRSA ST398 S0385 isolate. The similarity between the isolate Mu50 and S0385 suggests that the S0385 isolate potentially has lost some porcine specificity despite being of porcine origin. This underlines the adaptive potential of ST398.

3.2.2 The Skin

The skin surface constitutes a harsh environment and *S. aureus* must overcome surroundings that are constantly changing. Sweating and drying of the skin mean considerable changes in osmolarity, salt concentration and pH, in addition to mechanical stress. Host defence plays a significant role for bacterial survival and resistance to antimicrobial peptides produced by the host is likely to be of major importance for the ability of *S. aureus* to survive on skin surfaces (Foster, 2009; Otto, 2012). Fatty acids present in sebum are part of the anti-bacterial defence of the skin and in order to defend itself *S. aureus* produces IsdA which makes the cell surface more resistance to these molecules. It has been shown that IsdA mutants are more sensitive to killing *in vitro* by bactericidal lipids and the mutants survive poorly on human skin compared to wild-type (Clarke et al., 2007).

In some CA-MRSA strains the presence of the arginine catabolic mobile element (ACME), which is linked to the SCC*mec* element, has been proposed to be important for pH haemostasis in the acid environment of the skin (Diep et al., 2008b; Foster, 2009). ACME has been identified in *S. epidermidis*, a commensal of the skin in humans, but was not identified in all the CA-MRSA isolates investigated by Diep *et al.* (Diep et al., 2006). In addition Hallin *et al.* did not find ACME in 16 LA-MRSA ST398 isolates using a microarray approach (Hallin et al., 2011). This indicates that other factors, than the ACME which was identified as important for human skin colonization, are important for skin colonization of animals.

3.3 Infection

S. aureus is an opportunistic pathogen that is capable of causing a variety of infections ranging from minor soft tissue and skin infections to life-threatening systemic infections (Ekkelenkamp et al., 2006; Hasman et al., 2010; Huijsdens et al., 2006). Successful infection in a specific host is multifactorial and depends on virulence factors produced by *S. aureus*. Both secreted and cell surface-associated proteins can promote adhesion to host extracellular matrices, damage host cells,

and facilitate host immune evasion (Fluit, 2012; Foster, 2005). Manuscript III, supporting material Table S2 shows an overview of the virulence factors that have been described in *S. aureus*. The genes are categorized based on functionality like adherence, exoenzymes, host immune evasion, secretion system and toxins.

The LA-MRSA ST398 lineage is mainly associated with porcine colonization and porcine skin infections (Cuny et al., 2010; van Duijkeren et al., 2008) and encodes generally not as many of the traditionally human described virulence genes compared to HA-MRSA and CA-MRSA lineages. Table 1 in manuscript III illustrates a virulence profile of the whole genome sequenced LA-MRSA ST398 S0385 genome using the *S. aureus* VirulenceFinder generated and described in manuscript III (manuscript III Table 1).

3.3.1 Toxins

S. aureus encodes toxin like hemolysins, enterotoxins, exotoxins, exfoloative toxins, toxic shock syndrome toxin (tsst) and leukotoxins as the Panton-Valentine leukocidin (PVL) (Foster, 2009). Different *S. aureus* strains encode different toxins. Exfoliative toxins, tsst and PVL are only present in some clones, because they are encoded on MGEs, whereas alpha and gamma hemolysin are encoded in the core-genome and produced by most strains. Differential expression of core-genome encoded genes will nevertheless result in differences in pathogenesis (Novick et al., 1993).

Generally, the ST398 lineage is not associated with any of the human-associated enterotoxins (Golding et al., 2012; Hallin et al., 2011). A reduced toxicity could to some extent hide the bacteria from the immune system and may therefore facilitate a more stable and successful colonization of the host. This could explain why this particular lineage shows a broader host capacity compared to

most other *S. aureus* lineages. However, there are a number of un-described genes in the LA-MRSA stains which could encode virulence factors associated with infections in animals.

The world-wide emergence of CA-MRSA has been linked to the carriage of the PVL genes (DeLeo et al., 2010). The specific role of PVL in pathogenesis has been much debated. Based on a gene knockout screen, the PVL genes have been identified as contributing transiently to CA-MRSA pathogenesis in a rabbit bacteraemia model (Diep et al., 2008a). In addition there is a strong association between PVL and severe skin infections in humans (Lina et al., 1999). By comparative genomics the PLV genes have been identified in some human-associated MSSA ST398 isolates (Price et al., 2012) (manuscript III), which could contribute to increased virulence in these strains.

3.3.2 Host immune evasion

S. aureus produces several surface-associated components that increase bacterial resistance to phagocytosis (Foster, 2005). This is primarily obtained via anti-opsonic means disguising the bacterium from immune cell recognition. If *S. aureus* is recognized and potentially engulfed by phagocytic cells it is equipped with mechanisms promoting intracellular survival in addition to killing of host cells and manipulation of the adaptive immune response. Some of the *S. aureus* host immune evasion factors are presented in the following section.

Resistance to phagocytosis is an important bacterial feature to avoid being killed by the host immune system. *S. aureus* expressed various factors that contribute to a reduction in phagocytosis. The surface protein A, which is encoded by the *spa* gene, has demonstrated anti-phagocytic effects *in vitro* (Foster, 2009). It contains several domains that each binds to the Fc region of IgG (Forsgren and Sjöquist, 1966). This interaction coats the bacterium with IgG molecules in the incorrect

direction, which prevents any recognition by the neutrophil Fc receptor and activation of the complement system. Neutrophils and the complement system are both important candidates in an innate immune response. In addition, protein A has been identified as a virulence factor *in vivo* (Palmqvist et al., 2002).

The surface-associated clumping factor A (ClfA) binds fibrinogen and enhances virulence *in vivo* (Josefsson et al., 2001). This is most likely due to impaired recognition of opsonins resulting in increased resistance to phagocytosis (Higgins et al., 2006). Most *S. aureus* strains express a microcapsule that is composed of capsular polysaccharides (O'Riordan and Lee, 2004; Roghmann et al., 2005). Capsular serotype 5 and 8 is associated with increased virulence in animal infection models and the presence of a capsule has been shown to reduce bacterial uptake by human neutrophils *in vitro* (Luong and Lee, 2002; Nilsson et al., 1997; Thakker et al., 1998). In a microarray study both clumping factor (*clfA*) and capsular type 5 (*cap5A*) were identified in ST398 isolates (Hallin et al., 2011).

S. aureus displays several mechanisms to evade the host immune system. Beside anti-phagocytotic capacities the pathogen encodes virulence factors that can kill host immune cells, modulate the immune response and facilitate bacterial survival within phagocytotic cells.

Leukotoxins are cytotoxins that target leukocytes. *S. aureus* expresses different leukotoxins, of which only gamma hemolysin can lyse the membrane of both humane erythrocytes and humane leukocytes. The staphylococcal *g*amma hemolysins are bi-component and two active toxin (AB or CB) can be formed by combining the class-S components (HlgA or HlgC) with the class-F-component HlgB (Dalla Serra et al., 2005). An expression study showed that the gamma hemolysin components in *S. aureus* were up-regulated during short-term incubation in human blood *in vitro*. *hlg*ABCgene knockouts did, however, show similar virulence as the wild-type in a murine skin

infection model, and bacterial survival and neutrophil lysis after phagocytosis were similar between mutants and wild-type (Malachowa et al., 2011). The gamma-hemolysin components were not identified within the ST398 isolates investigated by Hallin *et al.* (Hallin et al., 2011).

The bacterial encoded MHC class II-analogue protein Map (also called Eap) can bind the T-cell receptor on T cells resulting in alteration in T cell function and causing a reduction in T cell proliferation. The protein can also manipulate the adaptive immune response by shifting a Th1 response to a Th2 response. This manipulation could explain why *map/eap* mutants are rapidly cleared compared to wild-type *in vivo* (Haggar et al., 2005). In addition, a high concentration of Map protein can have similar effect as a superantigen, stimulating apoptosis of B and T cells (Foster, 2009). A previously published study has identified the *map/eap* gene in MRSA ST398 isolates (Hallin et al., 2011).

When engulfed by phagocytes the bacterium encounter negatively charged antimicrobial defensins, which are secreted into the phagosome. *S. aureus* secretes proteins that can neutralise cationic defensins. The staphylokinase (sak), which is a prothrombin activator, can dissolve fibrin clots and cleave IgG and complement factor C3, both of which have potent defensin-binding effects (Foster, 2009). The *sak* gene has been associated with *S. aureus* host specificity (Sung et al., 2008).

S. aureus is an opportunistic pathogen shifting from being a colonising agent to cause infection in response to changes to host environments. It has the capacity to switch on selective sets of genes to enhance its chance for survival. This includes the regulation of virulence genes, which needs to be differentially expressed at different stages of infection. The expression is controlled by global regulatory systems, such as Agr, SarA, SaeRS and the alternative transcription factor sigmaB (σ^B)

(Otto, 2012; Pané-Farré et al., 2006). Inactivation of the *sarA* and *agr* loci has been shown to result in reduced virulence in several staphylococcal *in vivo* infection models (Abdelnour et al., 1993; Cheung et al., 2004; Nilsson et al., 1997).

3.3.3 High-throughput screening of S. aureus virulence genes

Various studies have used *S. aureus* transposon mutant libraries to screen for genes involved in infection. A collection of 6,300 *S. aureus* mutants were screened *in vivo* in a murine systemic infection model (Benton et al., 2004). 24 attenuated mutants were identified. The transposon inserts were identified by DNA size marker identification technology (SMIT). The mutants grouped into four functional classes, small molecule biosynthetic enzymes, cell surface binding and transport proteins, signal transduction systems, and anaerobic energy generation, as well as several conserved hypothetical proteins of unknown function. Mutations in genes encoding secreted virulence factors, such as hemolysins were not isolated (Benton et al., 2004).

Another study screening 1,248 *S. aureus* transposon mutants *in vivo*, in a murine bacteraemia model, did not identify previous described virulence genes as important for bacterial survival (Mei et al., 1997). Fifty attenuated mutants were identified and approximately half represented genes with unknown function. They recovered several mutants with insertion in the *fem*AB operon. FemA and FemB are involved in methicillin resistance and *fem*A mutants have shown a reduced cell wall turnover in growing cells, reduced whole-cell autolysis under non-growing conditions and increased methicillin sensitivity (Maidhof et al., 1991). In addition they recovered mutants with mutations in other cell surface components, like components of cell wall peptidoglycan, surface adhesion integrity, membrane transport, lipoprotein modification and genes affecting the capsule. Mutations affecting the tricarboxylic acid (TCA) cycle were also identified as important for survival in the

bacteraemia model. Such mutations were correlated to capsule production by Mei *et al*., as respiratory activity is essential for capsule production during some stages of *S. aureus* growth *in vitro* and capsule size can have a significant effect on virulence (Dassy and Fournier, 1996; Mei et al., 1997).

Begun *et al*. screened 2,950 transposon *S. aureus* mutants in a *C. elegans*-killing model and identified 10 unique mutants with mutations in TCA cycle components, nucleic acid metabolism/DNA replication, transporter, and miscellaneous proteins (Begun et al., 2005).

These findings underlines that the TCA cycle has a critical role in *S. aureus* pathogenesis and that the maintenance of the bacterial cell wall and cell wall transport are essential for successful colonization and infection.

3.4 Host specificity

It is generally believed that clones display a high degree of host specialization. Phylogeny has indicated that clones isolated from one host species tend to be uncommon in other species. There has however been increasing evidence that some lineages have a broader host-spectrum (McCarthy et al., 2012).

To study host specificity Moodley *et al*. used an *in vitro* skin corneocytes model to quantify adhesion in five *S. aureus* lineages (Moodley and Espinosa-Gongora, 2012). The human specific lineage ST36 showed preferred adhesion to human derived corneocytes whereas the pig-associated lineage ST433 showed preferred adhesion to porcine derived corneocytes. This was confirmed by *in*

vivo colonization of piglets, where ST433 was better at colonising than ST36. No differences were found in adhesion properties between a human and a pig derived ST398 isolate.

Uhlemann *et al.* compared adhesion properties of MRSA ST398 isolates derived from pigs and MSSA ST398 isolates, isolated from humans with no previous pig contact. The MSSA ST398 adhered significantly better to human derived skin keratinocytes *in vitro* compared to the MRSA ST398 isolates. However, no significant difference was observed in adhesion of the MRSA ST398 isolates to human- or pig-derived skin keratinocytes. The genomes differed in the content of mobile genetic elements (MGEs) and in surface-associated adhesion genes (Uhlemann et al., 2012)..

To elucidate the molecular mechanism underlying *S. aureus* host specificity several studies comparing human and animal derived strains have been performed. A micro-array based study revealed that six livestock-associated *S. aureus* ST398 isolates were distinct from more than 2,000 *S. aureus* isolates from humans (Belkum et al., 2008). Another study used a microarray-based comparative genomic approach to study genes associated with host specificity (Sung et al., 2008). Sung *et al.* found fibronectin binding protein A (*fnbA*), coagulase (*coa*) and cell wall-associated fibronectin binding protein (*ebh*) to be of most significance in relation to host specificity. They emphasized that it is likely that minor variation in other surface proteins are important as well, but these are too small to be detected by microarray. They also found a low incidence of the *scn*, *chp* and *sak* genes in the animal isolates. These genes are typically found as part of a prophage and implicate immune evasion in the human host. The staphylococcal complement inhibitor (*scn*) reduces phagocytosis by neutrophils and has been found to be specific to humans (Rooijakkers et al., 2005). *chp* encodes a chemotaxis inhibitory protein that modulates the chemokine response preventing neutrophil chemotaxis and activation, whereas *sak* encodes an anti-opsonin and inhibitor of defensins (Wamel, 2006).

Price *et al.* compared 89 MRSA and MSSA ST398 strains isolated from different hosts using WGS (Price et al., 2012). They identified a prophage encoding innate immune modulators specific for the human-derived isolates. *scn* was identified in all, *chp* in the majority, and *sak* in a third of the human originating isolates. In addition, only one of the 70 pig originating isolates contained the prophage. The tetracycline resistance gene *tetM*, encoded on an MGE, was identified in all the pig originating isolates and none of the human isolates. Price *et al.* suggests that ST398 originated in humans as MSSA and acquired methicillin (primarily SCC*mec* type Vc 5C2&5) and tetracycline resistance after the introduction to livestock (Price et al., 2012). This indicates the use of antibiotics (like tetracycline and beta-lactams) and heavy metals (like mercury) in food animal production is likely selecting for MRSA ST398 in pigs.

The *S. aureus* ST398 isolate S0385 isolated from a human case of endocarditis has been genome sequenced and annotated and can be found in GenBank under the accession no. AM990992 (Schijffelen et al., 2010). The ϕSA3 prophage found to be associated to *S. aureus* of human origin was not identified in this isolate and it has therefore been defined as a LA-MRSA isolate (Price et al., 2012; Schijffelen et al., 2010). The S03985 isolate harbour a *scn* homolog encoded on a pathogenecity island, which could be specific for the porcine host. Schijffelen *et al.* stress that this could mean that genes targeting animal immune systems could be found in *S. aureus* strains isolated from animals. However, there may only be relatively few conserved differences between human and animal isolates and that genes determining host specificity are difficult to identify even though most *S. aureus* lineages seem to be host specific (Sung et al., 2008).

The sequence analysis of the S0385 genome showed that the isolate was relatively different to other non-ST398 *S. aureus* genome sequences. These differences were identified in unique MGEs and

most of the elements harboured determinants for virulence and antimicrobial resistance. In addition, this isolate lacked one of the restriction and modification systems, which could make it more prone to up-take of foreign DNA (Schijffelen et al., 2010). These features may allow ST398 to adapt to new niches and could explain, at least in part, the broad host range this lineage is able to colonize.

Chapter 2

4. Summary and discussion of the results from the manuscripts

4.1 Manuscript I

The aim of this study was to generate a high complexity transposon mutant library and assess the application of TraDIS in *S. aureus* Sequence Type 398 (ST398), belonging to CC398. The generated transposon mutant library was screened in BHI and porcine blood in order to identify genes essential for ST398 to survive under these conditions.

The strain LA-MRSA ST398 S0385 was selected for this study because this isolate was the first CC398 isolate to be whole genome sequenced and annotated. The TraDIS method is based on next generation sequencing for comparing mutant composition pre and post selection. This approach requires an annotated reference strain to identify transposon insertions into open reading frames. The S0385 isolate was isolated from a human case of endocarditis but is considered a livestock-associated strain as it contains the Tn916 transposon encoding tetracycline resistance and the strain do not contain any of the phage associated genes often found in *S. aureus* isolated from humans.

The transposon mutant library was generated using a two plasmid system. One of the plasmids carried a Tn5-derived transposon with an erythromycin resistance marker (*erm*) and mariner mosaic ends. The mariner mosaic ends constitute inverted repeats, which is required for transposon insertion into the *S. aureus* genome. In addition the plasmid backbone contains a chloramphenicol resistance marker (*cat*). The other plasmid contains a transposase that is responsible for the insertion and excision of the mariner transposon. The transposase-carrying plasmid was modified in this study to contain gentamicin resistance (AAC6'-APH2') as the selection marker because the original selection marker was based on tetracycline resistance (*tet*). Thus, the new tool is also

available for future genetic manipulation in other tetracycline resistant strains. Both plasmids contain temperature-sensitive origins of replication only allowing plasmid replication at 30°C or below. The plasmids have previously been used to generate a high complexity transposon mutant library in an *S. aureus* laboratory strain (Chaudhuri et al., 2009). Both plasmids were transduced into the strain of interest using the *S. aureus* specific bacteriophage ϕ11, which has been described for genetic manipulation previously (Novick, 1991). After transduction the cells will contain one set of plasmids and within these cells the transposition event were conducted. During transposition the mariner transposon was inserted at a random TA di-nucleotide position in the *S. aureus* genome. Once the transposon insertion mutants were generated the mutant pool was grown at 43°C to facilitate plasmid loss. To eliminate the plasmids the mutant pool was passaged up to four times at 43°C and serial dilution and plating were used to determine the mutant library size and plasmid loss. At this temperature the plasmids will not replicate and thus their presence in the mutant population will be diluted. After each passage, mutant library aliquots were diluted and cultured on BHI agar plates containing erythromycin, chloramphenicol or gentamicin. Approximately 10^6 CFU/ml showed erythromycin resistance but chloramphenicol sensitivity. This showed chromosomal integration of the transposon and 100 % plasmid loss of the transposon-carrying plasmid. After growth at 43°C for two generations, 70 % of the erythromycin resistant mutants showed sensitivity to gentamicin, indicating that approximately 30 % of the mutants still contained the transposase-carrying plasmid. To increase the plasmid loss two additional passage at 43°C were conducted. This resulted in approximately 93 % plasmid loss after both the third and fourth growth passages. Growth at high temperatures will induce selection on the mutant library and will influence the specificity of further downstream screenings and therefore, the passage at high temperatures was terminated after three passages. However, due to the incomplete plasmid loss, around 7 % of the mutants will contain a transposase. The transposase could facilitate excision and re-insertion of the

transposon and thereby influence the stability of the mutant library. To avoid relocation of the transposon nutrient-rich broth was supplemented with erythromycin at each growth step and thus, the mutants were in that way continuously exposed to erythromycin. This will aid stability of the transposon insertions as the erythromycin resistance gene will not be transcribed in an excised transposon which will make the mutants sensitive to erythromycin. After generating the mutant library, various screenings assays where performed, where the mutant composition in an input pool was compared to a mutant composition in an output pool. If any genomic transposon relocation had taken place despite the presence of erythromycin, the mutant composition would have altered unintentionally. Such potential alterations could however only take place in a minor proportion of the mutants and would only be included in the analysis if the change in mutant composition happens at the output pool level, as only mutants present in the input pool will be considered in the final evaluation. Additionally this will only influence the results if the unintended transposon relocation happens with a similar frequency in all biological replicates.

The transposon mutant library was validated using Linker-PCR and sequencing. Linker-PCR is a method to validate if the transposon had inserted randomly throughout the genome. A transposon-specific forward primer facing outwards and a linker-specific reverse primer were used. The reverse primer will not recognize its target before after the first round of amplification with the transposon-specific primer and therefore, only transposon insertion sites will be exponentially amplified. Random mutants from the mutant library were selected and the transposon insertion site was amplified using linker-PCR, sequenced, and mapped against the reference genome to identify the genomic insertion site (see manuscript I supporting Figures S2 and S3).

Once the genome-saturated mutant library was generated and validated, the library was screened for genes important for bacterial survival under specific condition. Genomic DNA from mutant pools were sequenced using Transposon Directed Insertion site Sequencing (TraDIS). The sequencing was performed using a custom sequencing primer, sequencing from the 5' end of the transposon and into the genomic DNA flanking the transposon insert. In a sequencing run, one lane from an Illumina flow cell generated a minimum of 40 million reads of 43 bp plus index reads. The first 10 bp of each read constitute the Tn sequence, which were stripped from the reads. The remaining reads were between 10-23 bp in length. The sequence reads were mapped to the reference strain (Accession no. AM990992). Reads of down to 10 pb in length were allowed in this analysis, as all the sample genomes were identical to the reference genome. It is very likely that when comparing two identical genomes even small reads of 10 bp will map correctly. However, it might be more likely that a 10 bp read will map to more than one position within the reference strain. The aligner tool Bowtie 2.0 was used for mapping the sequence reads to the reference genome. By default, all reads mapping more than once to the reference genome will randomly be mapped to only one position and such reads will be given a low mapping quality score. When defining the number of unique insertion sites only reads with a high quality score will be taken into account and this number will therefore be based on reads mapping only once to the reference genome.
It is known that the *S. aureus* genome contains duplicate regions, which is important to recognize when evaluating essential genes with zero transposon insertion sites. For this evaluation all reads was considered despite mapping quality score and only the genes with zero reads mapping was proposed as essential in this study.

The high-throughput approach is based on a negative selection strategy. If an essential gene has been disrupted by a transposon insertion the mutant will not be viable and thereby not present in the

mutant pool. This strategy was used in the study to identify essential genes under laboratory conditions. A total of 152 genes had zero transposon inserts and were proposed as essential for LA-MRSA ST398 survival under laboratory conditions (manuscript I supporting figures Table S1). As gene function can be maintained with few inserts in non-functional parts of a gene, genes with a low number of inserts was also considered as essential or advantageous. In manuscript I, 526 genes with only few transposon inserts were identified and therefore evaluated as beneficial for growth under the laboratory conditions (manuscript I supporting figures Table S2).

A comparison between previous studies all identifying *S. aureus* essential genes under laboratory conditions using high-throughput approaches was performed (Bae et al., 2004; Chaudhuri et al., 2009; Fey et al., 2013). Some differences between the lists of proposed essential genes in *S. aureus* were identified and can be found in manuscript I supporting figures Table S3.

Any attempt to define the minimum set of essential genes will inevitably be influenced by the conditions under which the experiment is performed. A gene may be scored as essential in a particular experiment because it is required for survival following exposure to a particular stress inherent in the methods or because it is involved in uptake or metabolism of the particular nutrients provided in the growth media. An example of this is the requirement for extended incubation of *S. aureus* at high temperatures (>43°C) to facilitate loss of the temperature-sensitive plasmids. Consequently, genes required for high temperature survival will be scored as putative essential. Thus the differences found in the studies presented in Table S3 could either be a result of differences in methodology and experimental conditions or true differences between strains. In this study an insertion index was calculated and a cut-off was defined to identify genes with a low number of transposon inserts as beneficial for growth under laboratory conditions. The application of the insertion index was introduced by the authors of the original TraDIS paper (Langridge et al., 2009) but was not applied in the previous studies identifying *S. aureus* essential genes (Bae et al.,

2004; Chaudhuri et al., 2009; Fey et al., 2013). The selection of the cut-off separating essential/beneficial from non-essential genes is an important consideration. The cut-off defined in this study may not be optimal and could in part be the reason for the difference seen in the comparison with previous studies (manuscript I supporting figures Table S3).

A complex transposon mutant library is a very sensitive system and even small differences in library generation and experimental conditions may influence the output. To generate a true comparison of *S. aureus* essential genes in various strains the transposon mutant library should be generated under the same conditions using the same approach for identification of the transposon insertion sites and for the sake of clarity, it might be better to use less complex libraries containing a lower number of mutants.

4.1.1 Evaluation of a high-throughput screening in whole porcine blood

To assess the appliance of the high-throughput screening approach, the transposon mutant library was screened in whole porcine blood *in vitro*. Two 50 ml falcon tubes were filled with approximately 10 ml heparinised whole porcine blood and each tube was inoculated with 0.5 ml of the mutant pool (8.8 x 10^7 cells). DNA was extracted from the input mutant pool (~10^9 cells) representing the mutant composition before screening the library in whole porcine blood. The blood samples were incubated for 24 hours at 37°C with aeration. The following day the blood cultures were tested for viable counts (1.4 x 10^7 CFU/ml) and 500 µl (~10^7 cells) from each blood-culture were inoculated into 2x 10 ml BHI supplemented with 5 mg/l erythromycin, to increase the bacterial/blood cell ratio prior to DNA extraction, and incubated over night at 37°C with aeration. This resulted in two rounds of growth selection: one selection round in whole porcine blood followed by a selection round in BHI. After the second round of selection, DNA was extracted from ~10^9 of the mutants from each blood culture and stored as output pools representing two biological

replicates. No specific cell viability tests were performed on the blood cells, but it has been shown previously that whole-blood units stored at room temperature maintain cellular counts and coagulation activity for up to 72 hours (Hughes et al., 2007). In addition, in previous experiments an initial decrease in bacterial cell counts was observed when incubating the transposon mutant library in whole porcine blood, which could reflect neutrophil killing (see manuscript I Figure S4).

To identify genes representing mutants with altered fitness after screening in whole porcine blood *in vitro*, gDNA from the input pool and the output pools were extracted and sequenced. The number of reads corresponding to each transposon insertion site in the input pool was compared to the number of reads mapping to the equivalent position in the output pools using the DESeq package in R
The raw read counts were expected to follow an approximately normal distribution. However, based on a frequency distribution plot, read counts below 2^4 showed inconsistency with this assumption and were therefore considered as noise and not used in the analysis (< 0.05 % of the reads were discarded). The reason for this noise is not understood but was seen repeatedly in all the samples. The sequence reads could potentially be chimeric reads that contains a Tn sequence and a part of a genomic position, but do not correspond to a true insertion site. A potential way of avoiding such chimeric reads could be to use paired end sequencing where both ends of the transposon are used for defining a transposon insertion site.
The read counts, corresponding to transposon insertion sites, were normalized with a size factor to account for variation in the total number of reads obtained from each samples. The ratio of input:output reads counts were determined and referred to as a $\log_2$ fold change. A negative $\log_2$ fold change reflects an attenuated mutant and was determined when the number of read counts from input pool to output pool decreased and thereby illustrated a decrease in mutant clones after selection. For strongly attenuated mutants, zero clones were present in the output pools and the $\log_2$

fold change was defined as minus infinity for such mutants. For each individual mutant, the hypothesis that the fitness score was equal to zero and thereby that the mutant was present at equivalent levels in the input and output pools was tested for, using a negative binomial distribution as implemented in DESeq (Chaudhuri et al., 2013). DESeq models variance under the assumption that mutants with comparable levels of sequence coverage exhibit similar levels of dispersion. The model was fitted only from those mutants from which replicate data was available and the resultant model was then applied to data derived from all mutants to estimate *P* values.

Twenty-three mutants were identified with a specific significant reduction in fitness after selection in whole blood (manuscript I Table 3). Some mutations related to carbon metabolism via regulation of the TCA cycle, enzymes involved in gluconeogenesis and galactose metabolism. Several mutants could be linked to the cell wall and pH shock, in the form of amino acid metabolism, transport, pigmentation and cell wall repair. Mutations involved in regulation, which could affect the transcription of virulence genes, were also identified. Eight of the 23 genes representing attenuated mutants were of unknown function. Mutation in one gene of unknown function resulted in a hypercompetitive mutant. For an overview of the genes see manuscript I Table 4.

The transposon mutant library was incubated in whole porcine blood *in vitro* for 24 hours. This could partly reflect why many metabolic genes were identified as important for whole porcine blood survival in this study. However, an incubation period of 24 hours was specifically selected based on initial growth experiments performed in whole porcine blood *in vitro* (manuscript I Figure S4). These experiments showed an initial decrease in bacterial population size, which could be explained by phagocytosis and potential bacterial killing by host immune cells. The mutant population size returned to an equivalent size of the inoculated population after 24 hours, and at this point the

mutants had potentially seen all the selective elements within whole blood. Genes important for immune evasion will have undergone selection in a similar manner as the metabolic genes but as *S. aureus* encodes a larger variety of immune evasion genes it is justifiable to conclude that none of these are singlehandedly responsible for survivin the immune response. This may explain why no immune evasion genes were identified as important for whole blood survival.

The results indicate that key genes for survival in porcine blood cultures may not be genes involved for iron uptake, such as hemolysins and sideophors, and immune evasion but may be genes associated with the ability to utilize the available carbon hydrates in blood, which is supported to by previous studies (den Reijer et al., 2013; Malachowa et al., 2011; Mei et al., 1997). In two of these studies it was observed, that up- or down-regulated genes were mainly involved in cellular metabolism or had an unknown function (den Reijer et al., 2013; Malachowa et al., 2011). A previous study screening 1248 transposon *S. aureus* mutants in an *in vivo* murine bacteraemia model identified 50 genes as being important for whole blood survival, half of which had unknown function and the rest with an involvement in nutrient biosynthesis and surface metabolism (Mei et al., 1997). Furthermore, they identified genes important for the tricarboxylic acid cycle (TCA cycle) and in this study we identified the *icd* gene, a TCA cycle regulator, as important for *in vitro* survival in porcine blood. This indicates that the TCA cycle and carbon metabolism have important functions for bacterial survival in blood from different hosts *in vivo* and *in vitro*. The *fem*A and *fem*B genes were previously identified as important for whole blood survival *in vivo* (Mei et al., 1997). However, we found *fem*A and *fem*B mutants to have a growth disadvantage under laboratory conditions which is consistent with other studies identifying *S. aureus* essential genes (Bae et al., 2004; Chaudhuri et al., 2009; Fey et al., 2013).

In this study, a high complexity transposon mutant library was successfully generated in an LA-MRSA ST398 WT isolate and evaluated using the TraDIS system. *S. aureus* ST398 essential genes were identified and comparable with previous studies. Twenty-four genes were evaluated as being important for specific *in vitro* whole porcine blood survival, of which carbon metabolism, pH shock and regulation were related. For further evaluation of the genes identified as important for whole porcine blood survival it is necessary to generate single knock-out mutants and test these in the same assay as used in the high-throughput screening. In addition, it could be valuable to evaluate the single mutants in blood from different donor and under *in vivo* conditions.

4.2 Manuscript II

In this study, a transposon mutant library consisting of approximately one million LA-MRSA ST398 mutants was screened to identify genes important for survival in the porcine environment. The mutant library was generated and validated as described in manuscript I. The mutant library was screened in an *ex vivo* porcine skin model and an *ex vivo* porcine nasal epithelial model. For this purpose, two 6-month-old pigs, a male (Pig_1) and a female (Pig_2) were collected from the same farm with two weeks in between. They were euthanized by intravenous overdose of pentobarbitone and the tissue was collected immediately postmortem after obtaining the farm owner's permission for the use of their pigs in this study.

4.2.1 Porcine e*x vivo* skin model

The *ex vivo* porcine skin model was prepared as described previously (Maisch et al., 2007). Briefly, the skin areas behind the ears were washed with chlorhexidine soap and disinfected with 70 % ethanol before epilation with a sterile razor. A squared skin piece of around 8 x 8 cm was removed from the pigs and the adipose tissue beneath the dermis was removed with a scalpel. The skin was

dissected under sterile conditions into 2 cm^2 pieces, placed in 6-well plates and embedded in Hepes agar leaving the skin surface uncovered (manuscript II Figure 1). The skin pieces were disinfected with 70 % ethanol followed by washing with PBS three times. Swabs were taken from the washed skin surface to test for surface contamination and no such was found on any of the prepared skin explants.

Aliquots, of the transposon mutant library generated as described in manuscript I were grown over night in BHI broth supplemented with erythromycin. gDNA was extracted from the mutant culture (~10^9 cells) representing the input mutant pool.

In the porcine skin survival assay, 10 µl of up-concentrated stationary mutant culture (~10^{11} cells) were inoculated onto the porcine skin surface and incubated under atmospheric conditions at 32°C for ~24 or ~48 hours and duplicates were generated for each incubation period from both Pig_1 and Pig_2. After incubation the skin explants were homogenized and 9 x 10^7 - 2.5 x 10^8 CFU/ml was recovered after ~24 hours and 2.1 x 10^8 – 4.1 x 10^8 CFU/ml was recovered after ~48 hours incubation on the skin explants. A decrease in mutant cell count was observed suggesting an initial selection on the mutant pool. A slight increase in cell counts were observed between ~24 and ~48 hours incubation from an average of ~2 x 10^8 to ~3 x 10^8 CFU/ml, which propose that the mutants that are present on the skin explants are viable.

S. aureus expresses different surface proteins depending on growth phase (Foster, 2009) and therefore both exponentially and stationary grown cell were used in the skin adhesion assay. gDNA was extracted from an exponentially (OD_{600} 0.5-0.8) and stationary grown transposon mutant culture (~10^9 cells from each growth phase), representing mutant input pools. 10 µl of up-concentrated exponentially grown and stationary grown cells (~10^{11} from each growth phase) were

inoculated onto the porcine skin surface and incubated under atmospheric conditions at 32°C for ~20 hours. Four replicates for each growth phase were performed on tissue explants from Pig_1.
A decrease in cell counts was observed between the cells recovered in the adhesion assay (an average of ~5.4 x 10^7 CFU/ml) compared to the cells recovered in the survival assay after 24 hours incubation (an average of ~2 x 10^8 CFU/ml). This indicates that some mutants were lost in the washing step preformed in the adhesion assay. A lower number of mutants were recovered after porcine skin adhesion with exponential cells compared to stationary cells (an average of ~1.3 x 10^7 and ~5.4 x 10^7 CFU/ml respectively), which could point to that the stationary grown mutants adhere better to the porcine skin explants.
The cell suspensions, recovered from all the skin explants from both the survival and the adhesion assays were re-inoculated into 10 ml fresh BHI supplemented with 5 mg/l erythromycin to select for transposon mutant and reduce growth of the natural porcine skin microbiota. gDNA was extracted and sequenced from all the output replicates.

The mutant composition in input and output was evaluated using the DESeq package in R. The read counts corresponding to transposon insertion sites were normalized to account for variation in the total number of reads obtained from each sample. The ratio of input:output read counts were determined and referred to as a $\log_2$ fold change, which will be referred to as a fitness score. A negative fitness score reflects an attenuated mutant. An attenuated mutant was determined when the number of read counts from input pool to output pool decreased and thereby illustrated a decrease in mutant clones after selection. For strongly attenuated mutants, zero clones will be present in the output pools and the $\log_2$ fold change was defined as minus infinity and a fitness-score of -12 was assigned to such mutants. For each individual mutant, the hypothesis that the fitness score was equal to zero, e.i. the mutant was present at equivalent levels in the input and output pools, was

tested for using a negative binomial distribution as implemented in DESeq, as has been done previously (Chaudhuri et al., 2013). DESeq models variance under the assumption that the mutants with comparable levels of sequence coverage exhibit similar levels of dispersion. The model was fitted only from those mutants from which replicate data was available, which was in this case primarily sequence read counts from output pools, as no biological replicates were available from input pools. The resultant model was then applied to data derived from all mutants to estimate *P* values.

In the porcine skin survival study 27 genes were identified to be associated with alteration in fitness and therefore defined as important for LA-MRSA ST398 isolate S0385 survival on porcine skin. The genes selected represent mutants that had a significant change in fitness (*P* level ≤0.01) when screened on skin explants isolated from both pigs (two replicates from Pig_1 and from Pig_2). The genes are listed in manuscript II Table S1. Twenty-two genes illustrated attenuated mutants and fourteen of these genes have been identified as essential/beneficial for growth under laboratory conditions previously (manuscript I Table S1 and S2). Eight genes represent mutants with a reduced fitness specifically in the porcine skin survival assay and they are described as hypothetical proteins, regulators and transporters mainly. The S0385 strain contains 3 circular plasmids (Schijffelen et al., 2010) and after two days incubation on the porcine skin explants, mutants with transposon insert into the replication protein Rep located in plasmid 3 (PSAPIG030001) showed a drop in fitness. The plasmid is annotated to encode two different genes, the replication protein and a transcriptional regulator (SAPIG030002), one of which might be important for porcine skin survival. In addition, there were five genes representing hypercompetitive mutants in the porcine skin survival assay, of which two a reductase, one encodes a phage integrase and two encode repressors. These functions might not be important in porcine skin survival *ex vivo* but could be

essential in other more natural environments were competition and selection, are important factors for bacterial survival.

The transposon mutant library was also screened in a porcine skin adhesion assay to identify genes important for skin surface attachment. Only the genes specifically important for skin attachment were of interest and therefore genes identified as essential/beneficial for growth under laboratory conditions (manuscript I Table S1 and S2) were removed from the gene lists.

Sixty-eight genes were identified as representing mutants with significant reduced fitness (P level $\leq$ 0.05) when the transposon mutant library was selected in the adhesion assay as stationary grown cells (manuscript II Table S2). Twenty-nine genes representing mutants with attenuated fitness were identified when screening the transposon mutant library as exponentially grown cells in the adhesion assay (manuscript II Table S3).

In general various genes encoding enzymes, secreted proteins and surface-proteins represented the mutants with the most profound loss in fitness in the skin adhesion assays (manuscript II Table S2 and S3). Clumping factor B (clfB) and another fibrinogen-binding protein (SAPIG1154) were evaluated as important for skin adhesion. ClfB has previously been evaluated to be involved in human nasal adhesion and carriage (Corrigan et al., 2009). Immunoglobulin G binding protein A and staphylococcal secretory antigen ssA1 and ssA2 were identified as important for skin adhesion in addition to cap5A and cap5D, which are involved in capsular polysaccharide biosynthesis. Protein A and capsular polysaccharide inhibit phagocytosis (Foster, 2009) and the staphylococcal secretory antigens have predicted immunogenic function. This indicates that immune evasion and modulation are important features for the initial *S. aureus* ST398 colonization of porcine skin.

When comparing the list of genes obtained in the porcine skin survival and adhesion assay using stationary grown cells, eight genes were evaluated as important for both adhesion and survival in the skin model. Of these, six genes represent attenuated mutants and two genes hypercompetitive mutants (manuscript II Table 1). Eight genes were identified as important for porcine skin survival and adhesion using exponential grown cells, two of which showed increased fitness and six attenuated mutants with reduced fitness (manuscript II Table 2). Three genes showed inconsistency between the skin survival assay after 1 day of incubation and the skin adhesion assay using exponentially grown mutants.

Three genes, encoding a DNA-binding response regulator (*BecR*), an ABC transporter (*BecB*) and an export ATP-binding protein (*BecA*) were proposed as important for ST398 porcine skin survival (manuscript II Table 3). The genes, which are part of the BecAB transporter system, each showed a significant reduction in fitness when inserted with a transposon. The BecAB transporter system is similar to a *Bacillus subtilis* ABC transporter, which was previously defined as responsible for bacitracin efflux in Bacillus (Ohki et al., 2003). However, it could be that this ABC transporter system has other functions than bacitracin resistance as a *becS* (bacitracin sensing) mutant was not identified with reduced fitness in the skin survival model.

Another gene that was evaluated as important for porcine skin survival was *esaB* (manuscript II Table 3), which is a negative regulator of *esaC*. EsaC production and secretion is increased when Staphylococci replicate in serum or infected hosts (Burts et al., 2008). EsaB and EsaC are defined as being involved in *S. aureus* virulence and are required for persistent infection, *esaB* mutants fail to repress *esaC* and bacteria lacking *esaB* function will overproduce EsaC. Even though overexpression of EsaC is the natural response when *S. aureus* is replicating in host tissue, animals and humans mount an immune response to EsaC during infection (Burts et al., 2008), which could

explain why a constitutive overexpression of EsaC, in the e*saB* mutants, might not be in favour of the pathogen. In addition, a constitutive expression can have a metabolic cost on the mutant. Enzymes involved in membrane lipid metabolism and galactose metabolism were also identified as important for porcine skin survival in manuscript 2 (manuscript II Table 3).

4.2.2 Porcine e*x vivo* nasal epithelial model

The genome-saturated ST398 transposon mutant library was screened in a porcine *ex vivo* nasal epithelial survival model. The library was tested on explants from two different pigs. For isolation of nasal epithelial tissue, the pig head was removed from the carcass and immediately used for isolation of the nasal septum, leaving the lining nasal epithelial tissue intact. The tissue was washed in Dulbecco's Modified Eagle Medium (DMEM) supplemented enrofloxacin, streptomycin, and Fungizone to remove that natural microbiota. The antibiotic wash was followed by antibiotic-free washes in DMEM (for details, see method section in manuscript II). The nasal epithelium was dissected from the underlying cartilage of the nasal septum and divided into pieces of approximate 0.5 x 0.5 cm^2, in a sterile environment. Antibiotic residual test was performed on a bacterial lawn of LA-MRSA ST398 S0385 and here no growth clear zone was observed. The tissue pieces were placed on filter-paper overlying agar-plugs with the external side facing up-wards. The agar-plugs were arranged in 6-Well plates with a DMEM reservoir, moistening the filter paper and in that way nourishing the tissue (manuscript II Figure 1).

Aliquots of the transposon mutant library generated in manuscript I, were grown over night in BHI broth supplemented with erythromycin and gDNA was extracted from the mutant culture (~10^9 cells), representing the input mutant pool. From the mutant input culture, 5-10 µl of up-concentrated cells (~10^{11} cells) were inoculated onto the prepared nasal epithelium and incubated at 37°C plus 5

% CO_2 for ~24 hours (duplicates from Pig_1 and Pig_2). After incubation, the epithelial tissue was homogenized and 2.7 x 10^8 – 4.2 x 10^{10} CFU/ml was recovered after ~24 hours incubation on the nasal explants. A decrease in mutant cell count was observed suggesting selection on the mutant pool. The cell suspensions were re-inoculated into 10 ml fresh BHI supplemented erythromycin and incubated over night and gDNA was extracted (~10^9 cells), representing mutant output pools. gDNA from input and output pools were sequenced and mutant composition in input and output was evaluated like described for the porcine skin assay.

Four genes with specific importance for nasal epithelial survival were found in this study, two of which showed decrease in fitness and two with increased fitness. Manuscript II Table 3 shows the genes that were identified with a significant change in fitness score (P level ≤ 0.05) on both pigs.

Sixteen genes were identified as important for nasal epithelial survival, encoding proteins involved in regulation, metabolic enzymes, cell wall components and hypothetical proteins.

An aminoacyltransferase gene (*femA*) was identified as specifically important for ST398 nasal survival in both pigs (manuscript II Table 4). The isolate S0385 used in the screen contains four different genes of various lengths all described as encoding aminoacyltransferase FemA (SAPIG1375, SAPIG1248, SAPIG1250, SAPIG2462). In manuscript I, one of the four FemA encoding genes (SAPIG1375) was identified as advantageous for *S. aureus* survival under laboratory conditions (manuscript I Table S1). However, a different FemA-encoding gene (SAPIG1248) was identified as important for survival in the porcine *ex vivo* nasal epithelium survival model. FemA is involved in methicillin resistance and *fem*A mutants have shown a reduced cell wall turnover in growing cells, reduced whole-cell autolysis under non-growing conditions and increased methicillin sensitivity (Maidhof et al., 1991). The alteration in the cell wall in *femA* mutants could reduce bacterial resistance to the host immune response.

A previous study using porcine nasal explant identified beta-toxin gene *hlb* as a *S. aureus* S0385 nasal colonization factor, based on CFU quantifications of S0385 wild type and S0385 beta hemolysin mutant (Tulinski et al., 2013). SAPIG2471 encoding beta hemolysin was not among the genes identified as most significant for nasal epithelium survival in this study. When inspecting the raw count data, a decrease in read count from input to output for the beta hemolysin genes was indentified in three of the four replicates (data not shown), but this reduction was not defined as significant using the DESeq package in R. In the genome-wide screening approach, all mutants are compared relatively to each other and only the mutants with the most significant change will be identified. This could explain at least in part the inconsistency between the findings when using a high-throughput approach compared to a single mutant knockout strategy. Both methodologies are very useful for combining phenotypes to genotypes.

Some consistency was found between the genes identified as important for nasal epithelial survival when comparing the results obtained from the two pigs. It is known that many host factors are involved in *S. aureus* colonization (Foster, 2009) and the differences seen between the two pigs could be related to several factors, like genetic host variation, immune status, gender or simple differences between pig-replicates obtained when using this model system. Unfortunately, it was not possible to repeat the experiments on explants isolated from other pigs. However, as the genes presented here only illustrate genes which were identified as important for survival on several replicate explants isolated from both pigs, they should be considered as genes relevant for survival in the porcine reservoir. They constitute good gene candidates for generation of single knockout mutants, which should be tested within the same assays for a complete definition of the genes.

Pigs are important for the spread of ST398 and the identification of genes important for bacterial survival in the porcine reservoir could contribute to a better understanding of LA-MRSA ST398 ecology. Genes encoding transporters and metabolic enzymes were identified as relevant for porcine skin survival and genes encoding regulatory proteins, metabolic enzymes and cell wall components were proposed important for porcine nasal epithelium survival in this study. The genes could constitute targets for MRSA decolonization in pigs and thereby prevent further spread and adaption within the ST398 lineage. However, further investigations are needed to gain a more specific understanding of their role in bacterial survival.

4.3 Manuscript III

In this study the construction of the *S. aureus* VirulenceFinder is presented. The database is a web server that utilises whole genome sequence data from *S. aureus* genomes to extract a virulence profile and will be freely available through the Centre of Genomic Epidemiology (CGE) web services.

The database was built from sequences obtained from the NCBI nucleotide database. All known *S. aureus* virulence genes were listed (manuscript III Table 1) and sequences from these genes were used for building the database. The sequences were selected from 31 different *S. aureus* strains, which have been whole genome sequenced and annotated and can be found in GenBank (manuscript III Table S1). The sequences representing the virulence genes included in the *S. aureus* VirulenceFinder database were selected based on the annotations and gene descriptions found in the NCBI gene database and it can therefore not be excluded that virulence genes with a complex annotation or description were not included in the database.

For the first evaluation of the database the genome of the assembled LA-MRSA ST398 S0385 isolate (accession no. AM990992) was screened for virulence genes. The fasta file of the completed S0385 genome was submitted to an *S. aureus* VirulenceFinder alpha version and the threshold for nucleotide identity was set to ID=98 %. The virulence profile of S0385 was evaluated. Sixty-three different genes defined as virulence genes in the database were identified. The whole genome sequence dataset of S0385 isolate was used in building the database and it was therefore expected that all the identified virulence genes would show 100 % identity to a sequence within database. However, 16 of the identified virulence genes showed <100 % identity. This indicates that the annotation of these 16 genes was not recognized as virulence genes and these sequences were therefore not included in the database. Despite the incomplete collection of sequences from this genome the virulence genes were identified by lowering the default identity threshold to 98 % nucleotide identity. The ID threshold is by default set to 100 % nucleotide identity but by lowering the threshold, gene variants not included in to the database can be identified.
Correlating with previous findings in ST398 isolates, the VirulenceFinder identified *fnbA*, *clfA*, *cna*, *cap5A* and *eap/map* in the ST398 S0385 genome (Fluit, 2012; Hallin et al., 2011).

For further evaluation of the database previously whole genome sequenced *S. aureus* ST398 isolates, originating from various host origins, were screened for virulence profiles (Price et al., 2012). All together 89 *S. aureus* ST398 isolates were screened using the database but only 14 representatives were included in manuscript III (Table 2). The 14 isolates represent different clusters of the 89 isolates and both human- and porcine-originating isolates were analyzed. The 14 *S. aureus* ST398 isolates constitute useful candidates for initial screenings using the VirulenceFinder as they have been analyzed in a previous study (Price et al., 2012). A threshold of 95 % nucleotide identity (ID=95 %) was selected for this evaluation. The ID threshold can be set by

the user and a less stringent threshold is recommended as the alpha version of the database only contains variants of each virulence gene originating from 31 different *S. aureus* genomes used in building the database. A stringent threshold may result in some variation missed when using the database. However, a less stringent threshold will result in a considerably larger output that requires more analysis and potentially includes false positives.

Overall, the profiles of the 14 ST398 isolates were similar except for one isolate (13349_6), which also has been found to be an outlier previously (Price et al., 2012). The four isolates originating from a human host were positive for the *scn* gene, whereas none of the isolates originating from pigs contain the staphylococcal complement inhibitor. The *scn* and *sak* genes are both markers for strains of human origin (Price et al., 2012; Sung et al., 2008). Even though the *scn* was not identified in the isolates originating from pigs, a staphylococcal complement inhibitor variant has been identified on a pathogenicity island in ST398 S0385 (Schijffelen et al., 2010). Two of the human originating isolates contained both the *sak* gene and the two Panton-Valentine leukocidin encoding genes *lukF-PV* and *lukS-PV*. This is in agreement with the finding in Price *et al.* (2012) showing that some LA *S. aureus* strains are highly virulent, as is common for many of the CA *S. aureus* strains (Price et al., 2012).

Four of the 14 ST398 isolates originated from a human host. They all contained the *SdrC* and *SdrD* gene, whereas only *SdrC* was identified in all the ten isolates from porcine origin, when using the defined threshold (manuscript III Table 2). This might indicate that only *SdrC* is essential for adhesion to the porcine nares. However, phenotypic studies are needed for conformation.

Even though previous studies have emphasized that *S. aureus* ST398 do not contain any enterotoxins, two enterotoxin-encoding genes (*sep* and *sea*) were identified in 13 ST398 isolates using the VirulenceFinder (manuscript III Table 2). s*ep* was identified in all 13 ST398 isolates with

100% nucleotide identity to an annotated *sep* gene from the whole genome sequenced ST398 S0385 isolate, whereas *sea* was identified with only ~96 % nucleotide identify to an annotated *sea* gene from a ST80 CA-MRSA isolate. The enterotoxin P (*sep*) was originally defined after the full genome sequencing of *S. aureus* N315 (Omoe et al., 2005), however, the *sep* gene identified within the 13 ST398 isolates using the *S. aureus* Virulence Finder (GenBank gene SAPIG1666) showed similarity to two different genes within the N315 genome (GenBank gene SA1429 and SA1430). These genes encode an enterotoxin homolog and a protein similar to enterotoxin A precursor. This indicates that the *sep* gene identified in the 13 ST398 isolates is not the same as the original *sep* gene defined in *S. aureus* N315 (GenBank gene SA1761). However, ST398 might contain enterotoxin-like proteins that can be identified when using whole genome sequence data. Results obtained with the S. *aureus* Virulence Finder will reflect sequence and annotation quality found in the NCBI nucleotide database and might sometimes require further investigation.

Clumping factor A (*clfA*), protein A (*spa*) and capsular serotype 5 (*cap5A*) were found in all 14 ST398 isolates (manuscript III Table 2). The gamma-hemolysin components were not identified within the ST398 isolates investigated previously (Hallin et al., 2011). However, using the VirulenceFinder the gamma-components were identified in the 14 ST398 isolates. *hlgA* and *hlgC* were identified with >99 % identity to the sequences included in the database, whereas *hlgB* was identified with >95 % identity to a sequence from the database.

The *map*/*eap* gene was also identified in all the ST398 MRSA and MSSA isolates investigated. The *sak* gene was identified in two of the 14 ST398 isolates. These two isolates were MSSA ST398 of human origin. Two other MSSA ST398 isolates of human origin did not test positive for the *sak* gene when using the *S. aureus* VirulenceFinder.

Resistance and virulence profiles can help elucidate the approach for optimal treatment and define the virulence capacity of the infectious agent. Such information is crucial at hospitals in diagnostics and such profiles can as well be applied in local and global surveillance studies.

The *S. aureus* VirulenceFinder database generated in this study comprises an informative tool for whole genome sequence data to identify virulence genes in *S. aureus* genomes. The current version of the *S. aureus* VirulenceFinder is an alpha version and some adjustments of the included sequences are needed. The *S. aureus* Virulence Finder database will be part of the tool package found on the CGE webpage (www.genomicepidemiology.org). Here, tools like MLST and ResFinder are already available and additional tools for phylogenetic studies are under development.

5. Concluding remarks and future perspectives

S. aureus has multiple ways of thwarting the host immune system. The bacterium is able to colonize various hosts silently and under certain conditions cause infections of different severity. Various methods have been used to characterize *S. aureus* both as colonizer and as infectious agent. *S. aureus* and especially MRSA have been of world-wide importance for many years. The latest branch of MRSA is the LA-MRSA, which have been emerging the past decade. This group shows a broader host-spectrum compared to most other MRSA and a different virulence profile with fewer toxin-encoding genes. Little is known about the ecology of ST398 on farms, however, it is assumed that the use of antibiotics in the production animal industry has been the key force, driving emerge and spread of MRSA ST398. Even though some studies have shown that ST398 transmits less frequent among humans than human *S. aureus* strains, the transmissibility of ST398 still needs further investigation. ST398 has been the most commonly reported MRSA strain associated with livestock in recent years (Smith and Pearson, 2011), but knowledge on colonization and

transmission of LA-MRSA in pigs is limited and mainly based on observational field surveys (E M Broens et al., 2011b; Els M Broens et al., 2011). We are only in the beginning of understanding the role of these strains in the epidemiology of human *S. aureus* ST398 colonization and disease.

The three manuscripts that should be regarded as the main body of this thesis supplement the current knowledge about LA-MRSA ST398 characterization. It was demonstrated how high-throughput approaches can be utilized to perform a comprehensive phenotypic and genotypic characterization of a ST398 isolate.

Overall, the results showed that essential genes in ST398 seem to be similar to other *S. aureus* lineages, with few exceptions. Genes involved in the TCA cycle, membrane transport and pH shock are potentially important for ST398 survival in whole porcine blood *in vitro*. These findings correlate with previous studies investigating genes important for *S. aureus* survival *in vitro* in human blood and *in vivo* in a bacteraemia murine model. Membrane transport was identified as one of the main factors for ST398 survival on porcine skin and a FemA encoding gene was identified as essential for nasal epithelial survival *ex vivo*. The *ex vivo* nasal survival studies showed variations between explants isolated from two different porcine hosts, which underlines that host factors are important for nasal survival. Even though ST398 is not generally associated with enterotoxins, ST398 appears to contain some enterotoxin-like encoding genes. The staphylococcal complement inhibitor (*scn*) was, as shown previously, identified as a marker for host specificity of ST398, separating isolates of porcine and human origin. Additional serine-aspartic acid repeat surface proteins SdrC and SdrD may be of different importance for nasal colonization in pigs and humans.

The high-throughput approach applied in this work should be considered as a screen identifying genes that are potentially essential/beneficial for bacterial survival in a defined environment. For

definitive identification of gene function, it is necessary to generate single knockout mutants and test those in the same assays as used in the high-throughput screening. However, since a large number of genes are listed as having unknown function and there is a lack of correlation between phenotype and genotype, high-throughput methods, like the once developed and used in this work, will help to narrow the pool of genes to be investigated further.

The genes identified here as important for porcine survival could potentially constitute targets for MRSA decolonization within the porcine reservoir. By blocking transcription of these genes, LA-MRSA isolates will be attenuated in fitness, which could result in a reduction of LA-MRSA spread between pigs.

References

Abdelnour, a, Arvidson, S., Bremell, T., Rydén, C., Tarkowski, a, 1993. The accessory gene regulator (agr) controls Staphylococcus aureus virulence in a murine arthritis model. Infect. Immun. 61, 3879–85.

Agersø, Y., Hald, T., Helwigh, B., 2012. DANMAP 2012-use of antimicrobial agents and occurrence of antimicrobial resistance in bacteria from food animals, food and humans in Denmark. DANMAP 2012 101–102.

Anders, S., Huber, W., 2012. Differential expression of RNA-Seq data at the gene level – the DESeq package (No. DESeq version 1.13.0), European Molecular Biology Laboratory (EMBL), Heidelberg, Germany. European Molecular Biology Laboratory (EMBL), Heidelberg, Germany.

Armand-Lefevre, L., 2005. Clonal comparison of Staphylococcus aureus isolates from healthy pig farmers, human controls, and pigs. Emerg. Infect. Dis. 11, 711–714.

Bae, T., Banger, A.K., Wallace, A., Glass, E.M., Aslund, F., Schneewind, O., Missiakas, D.M., 2004. Staphylococcus aureus virulence genes identified by bursa aurealis mutagenesis and nematode killing. Proc. Natl. Acad. Sci. U. S. A. 101, 12312–7.

Bae, T., Glass, E.M., Schneewind, O., Missiakas, D., 2008. Generating a collection of insertion mutations in the Staphylococcus aureus genome using bursa aurealis. Methods Mol. Biol. 416, 103–16.

Barber, M., Rozwadowska-dowzenko, M., 1948. Infection by penicicllin-resisatnt staphylocooci. Lancet 641–644.

Begun, J., Sifri, C.D., Goldman, S., Stephen, B., Ausubel, F.M., Calderwood, S.B., 2005. Staphylococcus aureus Virulence Factors Identified by Using a High-Throughput Caenorhabditis elegans-Killing Model. Infect. Immun. 73, 872–877.

Belkum, A. Van, Melles, D.C., Peeters, J.K., Leeuwen, W.B. Van, Spalburg, E., Neeling, A.J. De, 2008. Methicillin-Resistant and -Susceptible Staphylococcus aureus Sequence Type 398 in Pigs and Humans. Emerg. Infect. Dis. 14, 479–483.

Benton, B.M., Zhang, J.P., Bond, S., Pope, C., Lee, L., Winterberg, K.M., Molly, B., Buysse, J.M., Christian, T., Schmid, M.B., 2004. Large-Scale Identification of Genes Required for Full Virulence of Staphylococcus aureus. Infect. Immun. 186, 8478–8489.

Bhat, M., Dumortier, C., Taylor, B.S., Miller, M., Vasquez, G., Yunen, J., Brudney, K., Rodriguez-Taveras, C., Rojas, R., Leon, P., Lowy, F.D., 2009. Staphylococcus aureus ST398, New York City and Dominican Republic. Emerg. Infect. Dis. 15, 285–287.

Broens, E M, Graat, E. a M., Van der Wolf, P.J., Van de Giessen, a W., De Jong, M.C.M., 2011a. Prevalence and risk factor analysis of livestock associated MRSA-positive pig herds in The Netherlands. Prev. Vet. Med. 102, 41–9.

Broens, E M, Graat, E. a M., van der Wolf, P.J., van de Giessen, a W., van Duijkeren, E., Wagenaar, J. a, van Nes, a, Mevius, D.J., de Jong, M.C.M., 2011b. MRSA CC398 in the pig production chain. Prev. Vet. Med. 98, 182–9.

Broens, Els M, Graat, E. a M., Van der Wolf, P.J., Van de Giessen, A.W., De Jong, M.C.M., 2011. Transmission of methicillin resistant Staphylococcus aureus among pigs during transportation from farm to abattoir. Vet. J. 189, 302–5.

Burian, M., Rautenberg, M., Kohler, T., Fritz, M., Krismer, B., Unger, C., Hoffmann, W.H., Peschel, A., Wolz, C., Goerke, C., 2010. Temporal expression of adhesion factors and activity of global regulators during establishment of Staphylococcus aureus nasal colonization. J. Infect. Dis. 201, 1414–21.

Burts, M., DeDent, A., Missiakas, D., 2008. EsaC substrate for the ESAT-6 secretion pathway and its role in persistent infections of Staphylococcus aureus. Mol. Microbiol. 69, 736–46.

Cavaco, L.M., Hasman, H., Aarestrup, F.M., 2011. Zinc resistance of Staphylococcus aureus of animal origin is strongly associated with methicillin resistance. Vet. Microbiol. 150, 344–8.

Chaudhuri, R.R., Allen, A.G., Owen, P.J., Shalom, G., Stone, K., Harrison, M., Burgis, T. a, Lockyer, M., Garcia-Lara, J., Foster, S.J., Pleasance, S.J., Peters, S.E., Maskell, D.J., Charles, I.G., 2009. Comprehensive identification of essential Staphylococcus aureus genes using Transposon-Mediated Differential Hybridisation (TMDH). BMC Genomics 10, 291.

Chaudhuri, R.R., Morgan, E., Peters, S.E., Pleasance, S.J., Hudson, D.L., Davies, H.M., Wang, J., van Diemen, P.M., Buckley, A.M., Bowen, A.J., Pullinger, G.D., Turner, D.J., Langridge, G.C., Turner, a. K., Parkhill, J., Charles, I.G., Maskell, D.J., Stevens, M.P., 2013. Comprehensive Assignment of Roles for Salmonella Typhimurium Genes in Intestinal Colonization of Food-Producing Animals. PLoS Genet. 9, e1003456.

Cheung, A.L., Bayer, A.S., Zhang, G., Gresham, H., Xiong, Y.-Q., 2004. Regulation of virulence determinants in vitro and in vivo in Staphylococcus aureus. FEMS Immunol. Med. Microbiol. 40, 1–9.

Clarke, S.R., Brummell, K.J., Horsburgh, M.J., McDowell, P.W., Mohamad, S.A.S., Stapleton, M.R., Acevedo, J., Read, R.C., Day, N.P.J., Peacock, S.J., Mond, J.J., Kokai-Kun, J.F., Foster, S.J., 2006. Identification of in vivo-expressed antigens of Staphylococcus aureus and their use in vaccinations for protection against nasal carriage. J. Infect. Dis. 193, 1098–108.

Clarke, S.R., Mohamed, R., Bian, L., Routh, A.F., Kokai-Kun, J.F., Mond, J.J., Tarkowski, A., Foster, S.J., 2007. The Staphylococcus aureus surface protein IsdA mediates resistance to innate defenses of human skin. Cell Host Microbe 1, 199–212.

Clarke, S.R., Wiltshire, M.D., Foster, S.J., 2004. IsdA of Staphylococcus aureus is a broad spectrum, iron-regulated adhesin. Mol. Microbiol. 51, 1509–19.

Corrigan, R.M., Miajlovic, H., Foster, T.J., 2009. Surface proteins that promote adherence of Staphylococcus aureus to human desquamated nasal epithelial cells. BMC Microbiol. 9, 22.

Coulter, S.N., Schwan, W.R., Ng, E.Y., Langhorne, M.H., Ritchie, H.D., Westbrock-Wadman, S., Hufnagle, W.O., Folger, K.R., Bayer, a S., Stover, C.K., 1998. Staphylococcus aureus genetic loci impacting growth and survival in multiple infection environments. Mol. Microbiol. 30, 393–404.

Crombé, F., Vanderhaeghen, W., Dewulf, J., Hermans, K., Haesebrouck, F., Butaye, P., 2012. Colonization and transmission of methicillin-resistant Staphylococcus aureus ST398 in nursery piglets. Appl. Environ. Microbiol. 78, 1631–4.

Cuny, C., Friedrich, A., Kozytska, S., Layer, F., Nübel, U., Ohlsen, K., Strommenger, B., Walther, B., Wieler, L., Witte, W., 2010. Emergence of methicillin-resistant Staphylococcus aureus (MRSA) in different animal species. Int. J. Med. Microbiol. 300, 109–17.

Cuny, C., Layer, F., Strommenger, B., Witte, W., 2011. Rare occurrence of methicillin-resistant Staphylococcus aureus CC130 with a novel mecA homologue in humans in Germany. PLoS One 6, e24360.

Dalla Serra, M., Coraiola, M., Viero, G., Comai, M., Potrich, C., Ferreras, M., Baba-Moussa, L., Colin, D. a, Menestrina, G., Bhakdi, S., Prévost, G., 2005. Staphylococcus aureus bicomponent gamma-hemolysins, HlgA, HlgB, and HlgC, can form mixed pores containing all components. J. Chem. Inf. Model. 45, 1539–45.

Dassy, B., Fournier, J.M., 1996. Respiratory activity is essential for post-exponential-phase production of type 5 capsular polysaccharide by Staphylococcus aureus. Infect. Immun. 64, 2408–14.

David, M.Z., Daum, R.S., 2010. Community-associated methicillin-resistant Staphylococcus aureus: epidemiology and clinical consequences of an emerging epidemic. Clin. Microbiol. Rev. 23, 616–87.

De Neeling, a J., van den Broek, M.J.M., Spalburg, E.C., van Santen-Verheuvel, M.G., Dam-Deisz, W.D.C., Boshuizen, H.C., van de Giessen, a W., van Duijkeren, E., Huijsdens, X.W., 2007. High prevalence of methicillin resistant Staphylococcus aureus in pigs. Vet. Microbiol. 122, 366–72.

DeLeo, F.R., Otto, M., Kreiswirth, B.N., Chambers, H.F., 2010. Community-associated meticillin-resistant Staphylococcus aureus. Lancet 375, 1557–68.

Den Reijer, P.M., Lemmens-den Toom, N., Kant, S., Snijders, S. V, Boelens, H., Tavakol, M., Verkaik, N.J., van Belkum, A., Verbrugh, H. a, van Wamel, W.J.B., 2013. Characterization of the humoral immune response during Staphylococcus aureus bacteremia and global gene expression by Staphylococcus aureus in human blood. PLoS One 8, e53391.

Devriese, L., Van Damme, R., Fameree, L., 1972. Methicillin (cloxacillin)-resistant Staphylocoocus aureus strains isolated from bovine mastitis cases. Vet. Med. 19, 598–605.

Diep, B., Gill, S., Chang, R., Phan, T., 2006. Complete genome sequence of USA300, an epidemic clone of community-acquired meticillin-resistant Staphylococcus aureus. Lancet 367, 731–739.

Diep, B., Otto, M., 2008. The role of virulence determinants in community-associated MRSA pathogenesis. Trends Microbiol. 16, 361–369.

Diep, B.A., Palazzolo-Ballance, A.M., Tattevin, P., Basuino, L., Braughton, K.R., Whitney, A.R., Chen, L., Kreiswirth, B.N., Otto, M., DeLeo, F.R., Chambers, H.F., 2008a. Contribution of Panton-Valentine leukocidin in community-associated methicillin-resistant Staphylococcus aureus pathogenesis. PLoS One 3, e3198.

Diep, B.A., Stone, G.G., Basuino, L., Graber, C.J., Miller, A., des Etages, S.-A., Jones, A., Palazzolo-Ballance, A.M., Perdreau-Remington, F., Sensabaugh, G.F., DeLeo, F.R., Chambers, H.F., 2008b. The arginine catabolic mobile element and staphylococcal chromosomal cassette mec linkage: convergence of virulence and resistance in the USA300 clone of methicillin-resistant Staphylococcus aureus. J. Infect. Dis. 197, 1523–30.

Ekkelenkamp, M.B., Sekkat, M., Carpaij, N., Troelstra, A., Bonten, M.J.M., 2006. [Endocarditis due to meticillin-resistant Staphylococcus aureus originating from pigs]. Ned. Tijdschr. Geneeskd. 150, 2442–7.

Falkow, S., 1988. Molecular Koch's postulates applied to microbial pathogenicity. Rev. Infect. Dis. 10 Suppl 2, S274–6.

Feil, E., Cooper, J., 2003. How clonal is Staphylococcus aureus? J. Bacteriol. 185, 3307–3316.

Feingold, B.J., Silbergeld, E.K., Curriero, F.C., Cleef, B.A.G.L. Van, Heck, M.E.O.C., Kluytmans, J.A.J.W., 2012. Livestock Density as Risk Factor for Resistant Staphylococcus aureus , the Netherlands. Emerg. Infect. Dis. 18, 1841–1849.

Fey, P.D., Endres, J.L., Yajjala, V.K., Yajjala, K., Widhelm, T.J., Boissy, R.J., Bose, J.L., Bayles, W., 2013. A Genetic Resource for Rapid and Comprehensive Phenotype. MBio 4, e00537–12.

Fitzgerald, J.R., Sturdevant, D.E., Mackie, S.M., Gill, S.R., Musser, J.M., 2001. Evolutionary genomics of Staphylococcus aureus: insights into the origin of methicillin-resistant strains and the toxic shock syndrome epidemic. Proc. Natl. Acad. Sci. U. S. A. 98, 8821–6.

Fluit, a C., 2012. Livestock-associated Staphylococcus aureus. Clin. Microbiol. Infect. 18, 735–44.

Food, E., Authority, S., 2010. Analysis of the baseline survey on the prevalence of methicillin-resistant Staphylococcus aureus (MRSA) in holdings with breeding pigs , in the EU , 2008 1 Part B□: factors associated with MRSA contamination of holdings 8, 1–67.

Forsgren, a, Sjöquist, J., 1966. "Protein A" from S. aureus. I. Pseudo-immune reaction with human gamma-globulin. J. Immunol. 97, 822–7.

Foster, T.J., 2005. Immune evasion by staphylococci. Nat. Rev. Microbiol. 3, 948–58.

Foster, T.J., 2009. Colonization and infection of the human host by staphylococci: adhesion, survival and immune evasion. Vet. Dermatol. 20, 456–70.

García-Álvarez, L., Holden, M.T.G., Lindsay, H., Webb, C.R., Brown, D.F.J., Curran, M.D., Walpole, E., Brooks, K., Pickard, D.J., Teale, C., Parkhill, J., Bentley, S.D., Edwards, G.F., Girvan, E.K., Kearns, A.M., Pichon, B., Hill, R.L.R., Larsen, A.R., Skov, R.L., Peacock, S.J., Maskell, D.J., Holmes, M. a, 2011. Meticillin-resistant Staphylococcus aureus with a novel mecA homologue in human and bovine populations in the UK and Denmark: a descriptive study. Lancet Infect. Dis. 11, 595–603.

Golding, G.R., Bryden, L., Levett, P.N., McDonald, R.R., Wong, A., Graham, M.R., Tyler, S., Van Domselaar, G., Mabon, P., Kent, H., Butaye, P., Smith, T.C., Kadlec, K., Schwarz, S., Weese, S.J., Mulvey, M.R., 2012. whole-genome sequence of livestock-associated st398 methicillin-resistant staphylococcus aureus Isolated from Humans in Canada. J. Bacteriol. 194, 6627–8.

González-Zorn, B., Senna, J., 2005. Bacterial and host factors implicated in nasal carriage of methicillin-resistant Staphylococcus aureus in mice. Infect. Immun. 73, 1847–1851.

Graveland, H., Wagenaar, J. a, Bergs, K., Heesterbeek, H., Heederik, D., 2011. Persistence of livestock associated MRSA CC398 in humans is dependent on intensity of animal contact. PLoS One 6, e16830.

Graveland, H., Wagenaar, J. a, Heesterbeek, H., Mevius, D., van Duijkeren, E., Heederik, D., 2010. Methicillin resistant Staphylococcus aureus ST398 in veal calf farming: human MRSA carriage related with animal antimicrobial usage and farm hygiene. PLoS One 5, e10990.

Grundmann, H., Aires-de-Sousa, M., Boyce, J., Tiemersma, E., 2006. Emergence and resurgence of meticillin-resistant Staphylococcus aureus as a public-health threat. Lancet 368, 874–85.

Haggar, A., Shannon, O., Norrby-Teglund, A., Flock, J.-I., 2005. Dual effects of extracellular adherence protein from Staphylococcus aureus on peripheral blood mononuclear cells. J. Infect. Dis. 192, 210–7.

Hall, N., 2007. Advanced sequencing technologies and their wider impact in microbiology. J. Exp. Biol. 210, 1518–25.

Hallin, M., De Mendonça, R., Denis, O., Lefort, A., El Garch, F., Butaye, P., Hermans, K., Struelens, M.J., 2011. Diversity of accessory genome of human and livestock-associated ST398 methicillin resistant Staphylococcus aureus strains. Infect. Genet. Evol. 11, 290–9.

Harris, L.G., Foster, S.J., Richards, R.G., 2002. An introduction to Staphylococcus aureus, and techniques for identifying and quantifying S. aureus adhesins in relation to adhesion to biomaterials: review. Eur. Cell. Mater. 4, 39–60.

Harrison, E.M., Paterson, G.K., Holden, M.T.G., Larsen, J., Stegger, M., Larsen, A.R., Petersen, A., Skov, R.L., Christensen, J.M., Bak Zeuthen, A., Heltberg, O., Harris, S.R., Zadoks, R.N., Parkhill, J., Peacock, S.J., Holmes, M. a, 2013. Whole genome sequencing identifies zoonotic transmission of MRSA isolates with the novel mecA homologue mecC. EMBO Mol. Med. 5, 509–15.

Hartman, B., Tomasz, A., 1981. Altered penicillin-binding proteins in methicillin-resistant strains of Altered Penicillin-Binding Proteins in Methicillin-Resistant Strains of Staphylococcus aureus. Antimicrob. Agents Chemother. 19, 726–735.

Hasman, H., Moodley, a, Guardabassi, L., Stegger, M., Skov, R.L., Aarestrup, F.M., 2010. Spa type distribution in Staphylococcus aureus originating from pigs, cattle and poultry. Vet. Microbiol. 141, 326–31.

Higgins, J., Loughman, A., van Kessel, K.P.M., van Strijp, J. a G., Foster, T.J., 2006. Clumping factor A of Staphylococcus aureus inhibits phagocytosis by human polymorphonuclear leucocytes. FEMS Microbiol. Lett. 258, 290–6.

Holmes, M. a, Zadoks, R.N., 2011. Methicillin resistant S. aureus in human and bovine mastitis. J. Mammary Gland Biol. Neoplasia 16, 373–82.

Hughes, J.D., Macdonald, V.W., Hess, J.R., 2007. Warm storage of whole blood for 72 hours. Transfusion 47, 2050–6.

Huijsdens, X.W., Dijke, B.J. Van, Spalburg, E., Van, M.G., Heck, M.E.O.C., Pluister, G.N., Voss, A., Wannet, W.J.B., Neeling, A.J. De, 2006. Community-acquired MRSA and pig-farming. Ann. Clin. Microbiol. Antimicrob. 5, 1–4.

Huyghe, A., Francois, P., Schrenzel, J., 2009. Characterization of microbial pathogens by DNA microarrays. Infect. Genet. Evol. 9, 987–95.

Jevons, M., Coe, A., Parker, M., 1963. Methicillin resistance in staphylococci. Lancet 904–907.

Josefsson, E., Hartford, O., O'Brien, L., Patti, J.M., Foster, T., 2001. Protection against experimental Staphylococcus aureus arthritis by vaccination with clumping factor A, a novel virulence determinant. J. Infect. Dis. 184, 1572–80.

Katayama, Y., Baba, T., Sekine, M., Fukuda, M., Hiramatsu, K., 2013. Beta-hemolysin promotes skin colonization by Staphylococcus aureus. J. Bacteriol. 195, 1194–203.

Khatiwara, A., Jiang, T., Sung, S.-S., Dawoud, T., Kim, J.N., Bhattacharya, D., Kim, H.-B., Ricke, S.C., Kwon, Y.M., 2012. Genome scanning for conditionally essential genes in Salmonella enterica Serotype Typhimurium. Appl. Environ. Microbiol. 78, 3098–107.

Kiser, K.B., Cantey-Kiser, J.M., Lee, J.C., 1999. Development and characterization of a Staphylococcus aureus nasal colonization model in mice. Infect. Immun. 67, 5001–6.

Kokai-Kun, J.F., 2008. The cotton rat as a model for Staphylococcus aureus nasal colonization in humans. Methods Mol. Biol. 431, 241–242.

Kondo, Y., Ito, T., Ma, X.X., Watanabe, S., Kreiswirth, B.N., Etienne, J., Hiramatsu, K., 2007. Combination of multiplex PCRs for staphylococcal cassette chromosome mec type assignment: rapid identification system for mec, ccr, and major differences in junkyard regions. Antimicrob. Agents Chemother. 51, 264–74.

Langridge, G.C., Phan, M., Turner, D.J., Perkins, T.T., Parts, L., Haase, J., Charles, I., Maskell, D.J., Peters, S.E., Dougan, G., Wain, J., Parkhill, J., Turner, A.K., 2009. Simultaneous assay of every Salmonella Typhi gene using one million transposon mutants. Genome Res. 2308–2316.

Laurent, F., Chardon, H., Haenni, M., Bes, M., Reverdy, M., Madec, J., Lagier, E., Vandenesch, F., Tristan, A., 2012. MRSA harboring mec A Variant Gene. Emerg. Infect. Dis. 18, 1465–1467.

Li, M., Du, X., Villaruz, A.E., Diep, B.A., Wang, D., Song, Y., Tian, Y., Hu, J., Yu, F., Lu, Y., Otto, M., 2012. MRSA epidemic linked to a quickly spreading colonization and virulence determinant. Nat. Med. 18, 816–9.

Lina, G., Piémont, Y., Godail-Gamot, F., Bes, M., Peter, M.O., Gauduchon, V., Vandenesch, F., Etienne, J., 1999. Involvement of Panton-Valentine leukocidin-producing Staphylococcus aureus in primary skin infections and pneumonia. Clin. Infect. Dis. 29, 1128–32.

Lowder, B. V, Guinane, C.M., Ben Zakour, N.L., Weinert, L. a, Conway-Morris, A., Cartwright, R. a, Simpson, a J., Rambaut, A., Nübel, U., Fitzgerald, J.R., 2009. Recent human-to-poultry host jump, adaptation, and pandemic spread of Staphylococcus aureus. Proc. Natl. Acad. Sci. U. S. A. 106, 19545–50.

Luong, T., Lee, C., 2002. Overproduction of type 8 capsular polysaccharide augments Staphylococcus aureus virulence. Infect. Immun. 70, 3389–3395.

Maidhof, H., Reinicke, B., Blümel, P., Berger-Bächi, B., Labischinski, H., 1991. femA, which encodes a factor essential for expression of methicillin resistance, affects glycine content of peptidoglycan in methicillin-resistant and methicillin-susceptible Staphylococcus aureus strains. J. Bacteriol. 173, 3507–13.

Maisch, T., Bosl, C., Szeimies, R.-M., Love, B., Abels, C., 2007. Determination of the antibacterial efficacy of a new porphyrin-based photosensitizer against MRSA ex vivo. Photochem. Photobiol. Sci. 6, 545–51.

Malachowa, N., Whitney, A.R., Kobayashi, S.D., Sturdevant, D.E., Kennedy, A.D., Braughton, K.R., Shabb, D.W., Diep, B.A., Chambers, H.F., Otto, M., DeLeo, F.R., 2011. Global changes in Staphylococcus aureus gene expression in human blood. PLoS One 6, e18617.

McCarthy, A.J., Lindsay, J. a, Loeffler, A., 2012. Are all meticillin-resistant Staphylococcus aureus (MRSA) equal in all hosts? Epidemiological and genetic comparison between animal and human MRSA. Vet. Dermatol. 23, 267–75, e53–4.

Mei, J.M., Nourbakhsh, F., Ford, C.W., Holden, D.W., 1997. Identification of Staphylococcus aureus virulence genes in a murine model of bacteraemia using signature-tagged mutagenesis. Mol. Microbiol. 26, 399–407.

Memorial, N., 2005. The Burden of Staphylococcus aureus Infections on Hospitals in the United States. Arch Intern Med 165, 1756–1761.

Monk, I., Shah, I., Xu, M., Tan, M., Foster, T., 2012. Transforming the untransformable: application of direct transformation to manipulate genetically Staphylococcus aureus and Staphylococcus epidermidis. MBio 3, 1–11.

Monk, I.R., Foster, T.J., 2012. Genetic manipulation of Staphylococci-breaking through the barrier. Front. Cell. Infect. Microbiol. 2, 49.

Moodley, A., Espinosa-Gongora, C., 2012. Comparative host specificity of human-and pig-associated Staphylococcus aureus clonal lineages. PLoS One 7, 1–7.

Moodley, A., Espinosa-Gongora, C., Nielsen, S.S., McCarthy, A.J., Lindsay, J. a, Guardabassi, L., 2012. Comparative host specificity of human- and pig- associated Staphylococcus aureus clonal lineages. PLoS One 7, e49344.

Nickerson, S.C., 2009. Control of heifer mastitis: antimicrobial treatment-an overview. Vet. Microbiol. 134, 128–35.

Nilsson, I., Lee, J., Bremell, T., 1997. The role of staphylococcal polysaccharide microcapsule expression in septicemia and septic arthritis. Infect. Immun. 65, 4216–4221.

Novick, R., 1991. Genetic systems in Staphylococci. Methods Enzymol. 204.

Novick, R.P., Ross, H.F., Projan, S.J., Kornblum, J., Kreiswirth, B., Moghazeh, S., 1993. Synthesis of staphylococcal virulence factors is controlled by a regulatory RNA molecule. EMBO J. 12, 3967–75.

O'Brien, L.M., Walsh, E.J., Massey, R.C., Peacock, S.J., Foster, T.J., 2002. Staphylococcus aureus clumping factor B (ClfB) promotes adherence to human type I cytokeratin 10: implications for nasal colonization. Cell. Microbiol. 4, 759–70.

O'Mahony, R., Abbott, Y., Leonard, F.C., Markey, B.K., Quinn, P.J., Pollock, P.J., Fanning, S., Rossney, a S., 2005. Methicillin-resistant Staphylococcus aureus (MRSA) isolated from animals and veterinary personnel in Ireland. Vet. Microbiol. 109, 285–96.

O'Riordan, K., Lee, J., 2004. Staphylococcus aureus capsular polysaccharides. Clin. Microbiol. Rev. 17.

Ohki, R., Tateno, K., Masuyama, W., Moriya, S., Kobayashi, K., Ogasawara, N., 2003. The BceRS two-component regulatory system induces expression of the bacitracin transporter, BceAB, in Bacillus subtilis. Mol. Microbiol. 49, 1135–1144.

Omoe, K., Imanishi, K., Hu, D., Kato, H., Fugane, Y., Abe, Y., Hamaoka, S., Watanabe, Y., Nakane, A., Uchiyama, T., Shinagawa, K., 2005. Characterization of Novel Staphylococcal Enterotoxin-Like Toxin Type P. Infect. Immun. 73, 5540–5546.

Otto, M., 2012. MRSA virulence and spread. Cell. Microbiol. 14, 1513–21.

Palmqvist, N., Foster, T., Tarkowski, A., Josefsson, E., 2002. Protein A is a virulence factor in Staphylococcus aureus arthritis and septic death. Microb. Pathog. 33, 239–249.

Pané-Farré, J., Jonas, B., Förstner, K., Engelmann, S., Hecker, M., 2006. The sigmaB regulon in Staphylococcus aureus and its regulation. Int. J. Med. Microbiol. 296, 237–58.

Pickard, D., Kingsley, R. a, Hale, C., Turner, K., Sivaraman, K., Wetter, M., Langridge, G., Dougan, G., 2013. A Genomewide Mutagenesis Screen Identifies Multiple Genes Contributing to Vi Capsular Expression in Salmonella enterica Serovar Typhi. J. Bacteriol. 195, 1320–6.

Price, L., Stegger, M., Hasman, H., Aziz, M., 2012. Staphylococcus aureus CC398: host adaptation and emergence of methicillin resistance in livestock. MBio 3, e00305–11.

Robinson, D., Enright, M., 2003. Evolutionary models of the emergence of methicillin-resistant Staphylococcus aureus. Antimicrob. Agents Chemother. 47, 3926–3934.

Roche, F.M., Meehan, M., Foster, T.J., 2003. The Staphylococcus aureus surface protein SasG and its homologues promote bacterial adherence to human desquamated nasal epithelial cells. Microbiology 149, 2759–67.

Roghmann, M., Taylor, K.L., Gupte, a, Zhan, M., Johnson, J. a, Cross, a, Edelman, R., Fattom, a I., 2005. Epidemiology of capsular and surface polysaccharide in Staphylococcus aureus infections complicated by bacteraemia. J. Hosp. Infect. 59, 27–32.

Rooijakkers, S.H.M., Ruyken, M., Roos, A., Daha, M.R., Presanis, J.S., Sim, R.B., van Wamel, W.J.B., van Kessel, K.P.M., van Strijp, J. a G., 2005. Immune evasion by a staphylococcal complement inhibitor that acts on C3 convertases. Nat. Immunol. 6, 920–7.

Schaffer, A.C., Solinga, R.M., Cocchiaro, J., Portoles, M., Kiser, K.B., Risley, A., Randall, S.M., Valtulina, V., Speziale, P., Walsh, E., Foster, T., Lee, J.C., 2006. Immunization with Staphylococcus aureus clumping factor B, a major determinant in nasal carriage, reduces nasal colonization in a murine model. Infect. Immun. 74, 2145–53.

Schena, M., Shalon, D., Davis, R., Brown, P., 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science (80-.). 270, 467–470.

Schijffelen, M.J., Boel, C.H.E., van Strijp, J.A.G., Fluit, A.C., 2010. Whole genome analysis of a livestock-associated methicillin-resistant Staphylococcus aureus ST398 isolate from a case of human endocarditis. BMC Genomics 11, 376.

Sifri, C.D., Begun, J., Ausubel, F.M., 2005. The worm has turned--microbial virulence modeled in Caenorhabditis elegans. Trends Microbiol. 13, 119–27.

Smith, T.C., Pearson, N., 2011. The emergence of Staphylococcus aureus ST398. Vector Borne Zoonotic Dis. 11, 327–39.

Strommenger, B., Kehrenberg, C., Kettlitz, C., Cuny, C., Verspohl, J., Witte, W., Schwarz, S., 2006. Molecular characterization of methicillin-resistant Staphylococcus aureus strains from pet animals and their relationship to human isolates. J. Antimicrob. Chemother. 57, 461–5.

Sung, J.M.-L., Lloyd, D.H., Lindsay, J. a, 2008. Staphylococcus aureus host specificity: comparative genomics of human versus animal isolates by multi-strain microarray. Microbiology 154, 1949–59.

Tenover, F., Arbeit, R., 1995. Interpreting chromosomal DNA restriction patterns produced by pulsed-field gel electrophoresis: criteria for bacterial strain typing. J. Clin. Microbiol. 33, 2233–2239.

Thakker, M., Park, J., Carey, V., Lee, J., 1998. Staphylococcus aureus serotype 5 capsular polysaccharide is antiphagocytic and enhances bacterial virulence in a murine bacteremia model. Infect. Immun. 66, 5183–5189.

Tulinski, P., Fluit, A.C., van Putten, J.P.M., de Bruin, A., Glorieux, S., Wagenaar, J. a, Duim, B., 2013. An ex vivo porcine nasal mucosa explants model to study MRSA colonization. PLoS One 8, e53783.

Uhlemann, A., Porcella, S., Trivedi, S., 2012. Identification of a highly transmissible animal-independent Staphylococcus aureus ST398 clone with distinct genomic and cell adhesion properties. MBio 3, e00027–12.

Urwin, R., Maiden, M.C.J., 2003. Multi-locus sequence typing: a tool for global epidemiology. Trends Microbiol. 11, 479–487.

Van den Broek, I.V.F., van Cleef, B. a G.L., Haenen, A., Broens, E.M., van der Wolf, P.J., van den Broek, M.J.M., Huijsdens, X.W., Kluytmans, J. a J.W., van de Giessen, a W., Tiemersma, E.W., 2009. Methicillin-resistant Staphylococcus aureus in people living and working in pig farms. Epidemiol. Infect. 137, 700–8.

Van Duijkeren, E., Box, a T. a, Heck, M.E.O.C., Wannet, W.J.B., Fluit, a C., 2004. Methicillin-resistant staphylococci isolated from animals. Vet. Microbiol. 103, 91–7.

Van Duijkeren, E., Ikawaty, R., Broekhuizen-Stins, M.J., Jansen, M.D., Spalburg, E.C., de Neeling, a J., Allaart, J.G., van Nes, a, Wagenaar, J. a, Fluit, a C., 2008. Transmission of methicillin-resistant Staphylococcus aureus strains between different kinds of pig farms. Vet. Microbiol. 126, 383–9.

Vandenesch, F., Naimi, T., Enright, M.C., Lina, G., Nimmo, G.R., Heffernan, H., Liassine, N., Bes, M., Greenland, T., Reverdy, M.-E., Etienne, J., 2003. Community-acquired methicillin-resistant Staphylococcus aureus carrying Panton-Valentine leukocidin genes: worldwide emergence. Emerg. Infect. Dis. 9, 978–84.

Verkade, E., Kluytmans, J., 2013. Livestock-associated Staphylococcus aureus CC398: Animal reservoirs and human infections. Infect. Genet. Evol.

Voss, A., Loeffen, F., Bakker, J., 2005. Methicillin-resistant Staphylococcus aureus in pig farming. Emerg. Infect. Dis. 11, 2004–2005.

Wamel, W. van, 2006. The innate immune modulators staphylococcal complement inhibitor and chemotaxis inhibitory protein of Staphylococcus aureus are located on β-hemolysin-Convertin Bactiophages. J. Bacteriol. 188, 1310–1315.

Wertheim, H.F.L., Walsh, E., Choudhurry, R., Melles, D.C., Boelens, H.A.M., Miajlovic, H., Verbrugh, H.A., Foster, T., van Belkum, A., 2008. Key role for clumping factor B in Staphylococcus aureus nasal colonization of humans. PLoS Med. 5, e17.

Williams, R., 1963. Healthy carriage of Staphylococcus aureus: its prevalence and importance. Bacteriol. Rev. 27, 56–71.

Zankari, E., Hasman, H., Cosentino, S., Vestergaard, M., Rasmussen, S., Lund, O., Aarestrup, F.M., Larsen, M.V., 2012. Identification of acquired antimicrobial resistance genes. J. Antimicrob. Chemother. 67, 2640–4.

Chapter 3

OPEN ACCESS Freely available online

PLOS ONE

Genome-Wide High-Throughput Screening to Investigate Essential Genes Involved in Methicillin-Resistant *Staphylococcus aureus* Sequence Type 398 Survival

Mette T. Christiansen[1]*¤a, Rolf S. Kaas[1], Roy R. Chaudhuri[2¤b], Mark A. Holmes[2], Henrik Hasman[1], Frank M. Aarestrup[1]*

1 National Food Institute, Technical University of Denmark, Kongens Lyngby, Denmark, **2** Department of Veterinary Medicine, University of Cambridge, Cambridge, England, United Kingdom

Abstract

Livestock-associated methicillin-resistant *Staphylococcus aureus* (LA-MRSA) Sequence Type 398 (ST398) is an opportunistic pathogen that is able to colonize and cause disease in several animal species including humans. To better understand the adaptation, evolution, transmission and pathogenic capacity, further investigations into the importance of the different genes harboured by LA-MRSA ST398 are required. In this study we generated a genome-wide transposon mutant library in an LA-MRSA ST398 isolate to evaluate genes important for bacterial survival in laboratory and host-specific environments. The transposon mutant library consisted of approximately 1 million mutants with around 140,000 unique insertion sites and an average number of unique inserts per gene of 44.8. We identified LA-MRSA ST398 essential genes comparable to other high-throughput *S. aureus* essential gene studies. As ST398 is the most common MRSA isolated from pigs, the transposon mutant library was screened in whole porcine blood. Twenty-four genes were specifically identified as important for bacterial survival in porcine blood. Mutations in 23 of these genes resulted in attenuated bacterial fitness. Seven of the 23 genes were of unknown function, whereas 16 genes were annotated with functions predominantly related to carbon metabolism, pH shock and a variety of regulations and only indirectly to virulence factors. Mutations in one gene of unknown function resulted in a hypercompetitive mutant. Further evaluation of these genes is required to determine their specific relevance in blood survival.

Citation: Christiansen MT, Kaas RS, Chaudhuri RR, Holmes MA, Hasman H, et al. (2014) Genome-Wide High-Throughput Screening to Investigate Essential Genes Involved in Methicillin-Resistant *Staphylococcus aureus* Sequence Type 398 Survival. PLoS ONE 9(2): e89018. doi:10.1371/journal.pone.0089018

Editor: Stefan Bereswill, Charité-University Medicine Berlin, Germany

Received October 30, 2013; **Accepted** January 14, 2014; **Published** February 12, 2014

Funding: This work was funded in part by the Danish Ministry of Food, Agriculture and Fisheries (Grant no. 3304-FVFP-09-F-002-1) and The Technical University of Denmark. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: m.christiansen@ucl.ac.uk (MTC); fmaa@food.dtu.dk (FMA)

¤a Current address: Immunity and Infection, University College London, London, United Kingdom
¤b Current address: Department of Molecular Biology and Biotechnology, University of Sheffield, Sheffield, United Kingdom

Introduction

Bacterial genomes contain between 470 to more than 9,000 different genes [1,2], many of which have unknown function. Detailed information on the importance and function of all genes within the genome is essential to understand bacterial survival and adaptation, especially for bacteria that may change between ecological stages as colonizers and pathogens and for those that may infect multiple hosts. Homology studies and other bioinformatic analyses of bacterial genomes have enabled prediction of gene function for many genes. However, there is still a shortage of data associating gene function with uncharacterized genes and characterized genes with phenotypes [3], as well as data on the relative importance of different genes for bacterial isolates living in different niches.

Transposon mutagenesis is a high-throughput method for functional phenotypic studies that can be utilised to associate genes to phenotypes. The method has been used to generate genome-saturated mutant libraries in several bacterial genomes [4–12]. The approach is based on a negative selection strategy, where transposon inserts into functional genes will result in mutants with attenuated fitness, or a complete inability to survive, and subsequent recovery of only those mutants with inserts in non-essential genes. The flanking regions of the transposon inserts can be identified and the composition of mutant libraries can be compared, pre- and post selection, resulting in identification of essential genes in a defined environment.

One genotypic approach for identifying transposon insertion sites, developed by Chaudhuri *et al.* (2009), is a DNA microarray and PCR-based method called Transposon Mediated Differential Hybridization (TMDH) [8]. This approach was applied in the first comprehensive study identifying essential genes in *Staphylococcus aureus*. Another genotypic strategy is based on high-throughput sequencing. Langridge *et al.* (2009) developed a system named Transposon Directed Insertion site Sequencing (TraDIS) which uses a transposon specific primer, enabling sequencing of the

genomic target region flanking the transposon insertion sites [9]. The sequencing approach has been used by Langridge *et al.* (2009), Khatiwara *et al.* (2012), Pickard *et al.* (2013) and Chaudhuri *et al.* (2013) to study essential and conditionally essential genes in *Salmonella* Typhi and *Salmonella* Typhimurium [9–11,13], but has not been applied previously to study *S. aureus* or other Gram positive bacteria. Importantly, this procedure not only identifies essential genes under different environmental conditions, but also provides an estimate of the relative importance of the presence or absence of genes.

S. aureus is an opportunistic pathogen that normally colonizes the host asymptomatically but given the opportunity, may cause a variety of pathogenic infections [14]. Some *S. aureus* clones are more successful human pathogens than others, and some show a high degree of host specificity for different animal species [15,16]. Recently, a specific linage belonging to clonal complex 398 (CC398), most likely of human origin, has spread among livestock globally, acquired methicillin resistance and is now transferring back to humans leading to both colonization and disease [17]. Pigs constitute a large reservoir for livestock-associated methicillin-resistant *S. aureus* (LA-MRSA) CC398 and contribute to an ongoing spread and genetic adaptation. Comparative genomic studies have identified a few phage associated genes that appear to be correlated with virulence in humans, but no genes of importance for successful colonization or infection in livestock or other animals have been identified [18]. A greater understanding of the pathogenicity and transmission of CC398 requires further investigations into the survival mechanisms utilized by this lineage.

The aim of this study was to generate a high complexity transposon mutant library and assess the application of TraDIS in *S. aureus* Sequence Type 398 (ST398), belonging to CC398. The generated transposon mutant library was screened in laboratory and host specific environments in order to identify genes essential for ST398 to survive under the given conditions. Even though ST398 is mainly associated with pig colonization and skin infections [19,20], *S. aureus* has potential to cause bacteraemia in pigs as well as in humans [15]. In this study whole porcine blood was applied for evaluation of the method.

Materials and Methods

Bacterial strains and culture conditions

The whole genome sequenced wild type (WT) livestock-associated methicillin-resistant *S. aureus* ST398 (Genbank accession AM990992) [21] and *S. aureus* RN4220 were grown in Brain Heart Infusion (BHI) (Oxoid, Difco) broth at 37°C with aeration. *S. aureus* SH1000 pMARGH2b, *S. aureus* SH1000 pFA545 and *S. aureus* RN4220 pFA545gen were grown in BHI or Tryptic Soy Broth (TSB) (Oxoid) with 5 mg/l erythromycin (Sigma), 5 mg/l tetracycline (Sigma) and 16 mg/l gentamicin (Sigma) respectively, at 30°C with aeration. For solid growth BHI agar, sheep blood agar plates (Oxoid) or Tryptic Soy Agar (TSA) (Oxoid) were applied and supplemented with the appropriate antibiotic if needed. *Escherichia coli* DH10 was cultured in Luria Broth (LB) at 37°C with aeration or on LB agar plates (Sigma).

Plasmids

The plasmids pMARGK2b and pFA545 previously described by Chaudhuri *et al.* (2009) were used for generating a transposon mutant library in the whole genome sequenced LA-MRSA ST398 S0385 isolate. The pMARGK2b plasmid contains a mariner transposon which includes an erythromycin resistance selection marker. The plasmid backbone holds a chloramphenicol resistance selection marker and a temperature-sensitive origin replication (replication at ≤30°C). The pFA545 encodes a transposase, a temperature-sensitive origin of replication (replication at ≤30°C) and a tetracycline resistance selection marker [8]. As the LA-MRSA ST398 S0385 isolate displays natural tetracycline resistance the pFA545 plasmid was purified (Qiagen tip100) and modified. Forward primer *KpnI* and reverse primer *SpeI* (see Table 1) were used for amplification of the AAC6′-APH2′ gene encoding gentamicin resistance from MRSA MU50 DNA, The PCR product and the original pFA545 were digested with *SpeI* and *KpnI* (New England Biolabs). The digested products were ligated using T4 DNA ligase (Fermentas). The modified pFA545 including the AAC6′-APH2′ gene (pFA545gen) was transformed into *E. coli* DH10 competent cells (Invitrogen), amplified (selected on ampicillin 100 mg/l or gentamicin 4–8 mg/l) and purified using the QIAprep spin column (Qiagen). An *EcoRV* (Fermentas) digest was performed on the purified original pFA545 (predicted digest products 7729 bp, 2038 bp, 312 bp → giving a total size of 10,079 bp) and the modified pFA545gen (predicted digest products 10,432 bp, 312 bp → 10,744 bp in total) and band patterns were compared on a 0.8% agarose gel (data not shown). pFA545gen was transformed into *S. aureus* RN4220 by electroporation.

Construction of transposon mutant library

pMARGK2b and pFA545gen were transduced into *S. aureus* ST398 S0385 in two separate rounds of transduction using the *S. aureus* bacteriophage ϕ11. Donor cells (SH1000 pMARGH2b or RN4220 pFA545gen) grown to mid-exponential phase OD_{600} 0.5–0.8 were mixed in a 1:1 ratio with two fold dilutions of phage in a 0.9% NaCl solution enriched with 10 mM $CaCl_2$. Following 5 min absorption at room temperature (rt.), the cells were plated in a TSB-top-agar solution (TSB, 0.5 mM $CaCl_2$, 0.5% agar) onto TSA plates supplemented with the appropriate antibiotics and incubated at 30°C over night. Top agar from plates with high phage titre were isolated, centrifuged (7,000 rpm, 10 min.) and sterile filtered using a 0.45 µm Millipore filter. Recipient cells (*S. aureus* ST398 S0385) were grown to OD_{600} 1–1.2, cells harvested by centrifugation (11,000 rpm, 10 min.) and re-suspended in TSB with 0.5 mM $CaCl_2$. Prepared recipient cells and phage lysate were mixed in different ratios (100:1–100:15), incubated at rt. for 5 min, followed by the addition of 0.5 mM $CaCl_2$ and incubated additionally 20 min. at rt. 0.02 M ice cold sodium citrate was added and mixed by vortexing. Cells were isolated by centrifugation (4000 rpm, 20 min, 4°C), re-suspended in 0.02 M sodium citrate, plated and incubated on BHI agar enriched with 0.2 mM sodium citrate and the appropriate antibiotic at 30°C over night. Transductants were sub-cultivated on selective plates containing the appropriate antibiotics and tested in an *ermB* and AAC6′-APH2′ PCR. Transductant, MRSA ST398 S0385 pMARGK2b pFA545gen was cultured at 30°C (plasmid replication at ≤30°C) with aeration in BHI supplemented with 5 mg/l erythromycin and 16 mg/l gentamicin and stored at −80°C in 0.5 ml aliquots ($>10^6$ cells) with 50% glycerol.

The transposon mutant library was generated as described by Chaudhuri et al. (2009) with some modifications [8]. A 0.5 ml aliquot was inoculated into 100 ml BHI containing 5 mg/l erythromycin and chloramphenicol (Sigma) and 16 mg/l gentamicin and incubated at 30°C with aeration until the culture reaches OD_{600} 0.4. Cells were recovered from 30 ml culture by centrifugation (4000 rpm, 10 min) and re-suspended in 600 ml BHI containing 5 mg/l erythromycin pre-warmed to 43°C. The culture was grown at 43°C with aeration until the culture reached an OD_{600} 0.4. 30 ml culture was recovered by centrifugation (4000 rpm for 10 min) and re-suspended in 600 ml BHI

Table 1. Primers.

Name	Sequence (orientation 5′ - 3′)	Source
Forward primer *KpnI*	GTGGGTACCTTAAFCCTAGAGCTTGCCATGTATATG	This study
Reverse primer *SpeI*	CTCACTAGTGTCTGGACTTGACTCACTTCC	This study
254 oligo	CGACTGGACCTGGA	J. H. Wang
256 oligo	GATAAGCAGGGATCGGAACCTCCAGGTCCAGTCG	J. H. Wang
ForwardTnL	CTTAAGTTTGCTTCGATGACTGG	This study
Reverse primer 258	GATAAGCAGGGATCGGAACC	J. H. Wang
ErmB forward 26	GGAACATCTGTGGTATGGCG	This study
ErmB reverse 27	CATTTAACGACGAAACTGGC	This study
Transposon-specific primer	AATGATACGGCGACCACCGAGATCTACACCTGAATTACCCTGTTATCCCTATTTAGGTGAC	Langridge *et al.* (2009)
P5	AATGATACGGCGACCACCGA	Illumina
P7	CAAGCAGAAGACGGCATACGA	Illumina
Sequencing primer	GACACTATAGAAGAGACCGGGGACTTATCAGC	This study

The table lists the primers used in the experimental approach. It includes primer name, nucleotide sequence and orientation, and source.
doi:10.1371/journal.pone.0089018.t001

containing 5 mg/l erythromycin pre-warmed to 43°C and the culture was grown at 43°C with aeration over night. The following day 30 ml culture was recovered and re-suspended in 600 ml BHI containing 5 mg/l erythromycin pre-warmed to 43°C and grown at 43°C with aeration over night and the same procedure was repeated one more day resulting in a 3rd generation transposon mutant library. Each day cells were plated on BHI plates containing 5 mg/l erythromycin, 5 mg/l chloramphenicol or 16 mg/l gentamicin and grown at 37°C over night. The growth pattern demonstrated a 100% cure of pMARGK2b, ~93% cure of pFA545gen and successful transposition of the transposon. Transposon mutants were stored in 0.5 ml ($>10^6$ cells) 50% glycerol aliquots at −80°C until further use.

Mutant library verification

Linker PCR was used to verify the complexity of the generated transposon mutant library. DNA was extracted (Gram positive DNA extraction Epicentre – lysing the cells with Ready-Lyse Lysozyme over night) from the transposon mutant pool in addition to DNA from 15 randomly isolated colonies (BHI plates containing 5 mg/l erythromycin) representing 15 random transposon mutants from the library. The DNA was digested with *RsaI* (Promega) and purified using a Minielute PCR purification kit (Qiagen). Adaptor molecules were made by mixing a 1:1 ratio (100 µM) of oligo 254 and 256 (see Table 1), denatured at 95°C for 3 min. in annealing buffer (10× annealing buffer = 100 mM Tris pH8, 500 mM NaCl, 10 mM EDTA) and annealed at room temperature for 1 hour (store at −20°C). Adaptors and digested DNA were ligated using a Quick DNA Ligase (New England Biolabs) followed by purification using a PCR purification kit (Qiagen). A PCR with primers ForwardTnL and reverse primer 258 (see Table 1) and Hotstar taq polymerase (Qiagen) was conducted with the following conditions: Hot-start 15 min at 95°C, 30 cycles of denaturation for 45 sec at 94°C, annealing 1 min at 55°C and elongation for 2 min at 72°C and a final elongation for 5 min at 72°C. The PCR products were visualised on a 2% NuSieve GTG Agarose gel (Lonza) (3 hours, 100 volts).

Passage of transposon mutant library in broth

A 0.5 ml mutant library aliquot ($>10^6$ cells) was inoculated in 10 ml BHI supplemented with 5 mg/l erythromycin and incubated over night at 37°C with aeration. 500 µl of the culture was re-inoculated into fresh BHI supplemented with 5 mg/l erythromycin and incubated over night at 37°C with aeration. The passage of the transposon mutant library was repeated three times. After each passage the library was tested for viable counts (results not shown) and DNA (from ~10^9 cells) was extracted using Easy-DNA kit (Invitrogen) which was stored at −20°C.

Ethical statement

The study protocol was submitted to the ethical review committee at the University of Cambridge, Department of Veterinary Medicine, who reported that post mortem collection of blood following the slaughter of male pigs, surplus to a breeding program, is not a regulated procedure and provided ethical approval. The UK Animals (Scientific Procedures) Act 1986 allows for the use of animal tissues and blood in research that comes from animals not regulated by the Act. These animals were slaughtered by a method of killing identified in Schedule 1 of the Act. In this case, a 6-month-old male pig was euthanized by intravenous overdose of pentobarbitone and the blood was collected immediately postmortem into heparinised containers after obtaining the farm owner's permission for the use of their pigs in this study.

Whole porcine blood survival

Two 50 ml falcon tubes were filled with approximately 10 ml heparinised whole porcine blood and each tube was inoculated with 0.5 ml mutant library aliquot (8.8×10^7 cells). DNA was extracted from pooled mutant library aliquots (~10^9 cells) using MasterPure Gram Positive Purification Kit (Epicentre) and stored as input pools (replicates) at −20°C. The blood samples were incubated for 24 hours at 37°C with aeration. The following day the blood cultures were tested for viable counts (1.4×10^7 CFU/ml) and 500 µl (~10^7 cells) from each blood-culture were inoculated into 2× 10 ml BHI supplemented with 5 mg/l erythromycin, to increase the bacterial/blood cell ratio prior to DNA extraction, and incubated over night at 37°C with aeration. This resulted in two rounds of growth selection, one selection round in whole porcine blood followed by a selection round in BHI. After the second round of selection DNA was extracted from ~10^9 of the mutants and stored at −20°C as output pools (replicates).

Library preparation for Illumina sequencing

For the TraDIS approach the library preps were prepared as described by Langridge *et al.* (2009) with modifications [9]. 3–5 μg of DNA from input and output pools were fragmented to an average size of approximately 200 bp by Covaris E210. The size profile was evaluated with Agilent 2100 Bioanalyzer on a DNA1000 chip. The fragmented DNA was prepared for sequencing on an Illumina platform using the SureSelect XT Library Prep Kit-ILM (Agilent). The ligated fragments were amplified using a transposon-specific primer (see Table 1) and the multiplexing PCR primer index 1–8 supplied in the SureSelect Library Prep Kit. The PCR was run for 22 cycles with 200–400 ng template-DNA per reaction to amplify the transposon insert and junction sites. The PCR products were cleaned using 0.8× Agencourt AMPure XP beads (Ramcon) to remove DNA fragments below 200 bp. The quality of the amplified products was assessed using an Agilent 2100 bioanalyzer on a high Sensitivity DNA chip and quantified by Q-PCR with primers P5 and P7 (see Table 1). The libraries were pooled in a 1:1 molar ratio and sequenced on an Illumina Hiseq2000 platform for 43 cycles plus index read using a custom sequencing primer (see Table 1) resulting in reads with the initial 10 bp being transposon insert specific followed by the junction region.

Sequencing analysis and statistics

Sequence reads from the Illumina FASTQ files were sorted by index and evaluated for the 10 bp transposon (Tn) sequence CAACCTGTTA allowing 1 mismatch, using the program Sabre (https://github.com/najoshi/sabre). The Tn and adapter sequences, as well as short reads (<10 nucleotides) and nucleotides with poor base call quality (<Q15), were stripped using Cutadapt [22] and the junction regions were extracted and mapped to the reference genome (AM990992) using Bowtie 2.0 [23]. An in-house script was used to identify the precise transposon insertion sites and quantify the number of reads mapping to the open reading frames within the reference genome. The program Circos [24] was applied for a genome wide visualization of the transposon mutant library.

The number of unique transposon insertion sites for any given gene was calculated and divided by the average gene length using an in-house script (insertion index calculation). Genes with zero or few transposon insertions sites were categorised based on function using the COG (Cluster of Orthologous groups) database [27,28], as described in Khatiwara *et al.* (2012) [10]. They were plotted as a percentage of all the COG categorised genes encoded by the reference genome.

The transposon mutant library was screened in whole porcine blood *in vitro* and mutants from input and output pools were compared using the DESeq package in R [25] enabling identification of significant differences in mutant composition pre- and post- selection. The approach was as described in Anders and Huber (2012) [26] and the settings are defined in Figure S1. The read counts, corresponding to transposon insertion sites were normalized to account for variation in the total number of reads obtained from each samples. The ratio of input:output read counts were determined and referred to as a $\log_2$ fold change. A negative $\log_2$ fold change reflects an attenuated mutant whereas a positive $\log_2$ fold change mirror a hypercompetitive mutant. For each individual mutant, the hypothesis that the fitness score was equal to zero and thereby that the mutant was present at equivalent levels in the input and output pools was tested for, using a negative binomial distribution as implemented in DESeq. The model was fitted only from those mutants from which replicate data was available and the resultant model was then applied to data derived from all mutants to estimate *P* values. An attenuated mutant was determined when the number of read counts from input pool to output pool significantly decreased and a hypercompetitive mutant was determined when the number of read counts from input to output pool significantly increased.

The raw sequence data will be available in the NCBI Sequence Read Archive (SRA) upon publication (Accession: SRR1056406 - SRR1056422).

Results

A construct for manipulation of LA-MRSA ST398

The transposon mutant library was generated in the whole genome sequenced wild type LA-MRSA ST398 S0385 isolate using a two plasmid system. One of the plasmids carried a Tn5 derived transposon with an erythromycin resistance marker and mariner mosaic ends, which was required for use in *S. aureus*. As most LA-MRSA ST398 harbour natural resistance to tetracycline, the tetracycline resistance marker in the transposase-bearing plasmid was substituted with a gentamicin resistance cassette, as S0385 was found, by susceptibility testing to be susceptible to gentamicin (Minimal Inhibitory Concentration, MIC = 0.5 mg/l). The tetracycline resistance gene was removed from the plasmid and the AAC6′-APH2′ gene originating from MRSA MU50 encoding gentamicin resistance was inserted into the plasmid at a position that facilitate the usage of the tetracycline resistance gene promoter. The plasmids were successfully transduced into the S0385 isolate.

Transposon mutant library

A high complexity mariner transposon mutant library was generated in the whole genome sequenced wild type LA-MRSA ST398 S0385 isolate. Serial dilution and plating on BHI agar plates containing the appropriate antibiotic determined a mutant library size of $\sim 10^6$ mutants, a 100% plasmid loss of the transposon carrying plasmid and approximately 93% plasmid loss of the transposase-carrying plasmid. Due to the incomplete loss of the transposase bearing plasmid, nutrient-rich broth was supplemented with erythromycin at each growth step to ensure that the genomic insertion of the transposon was maintained. Linker PCR and DNA sequencing was used to verify transposon insert throughout the bacterial genome (Figures S2 and S3).

Validation of the mutant library

DNA was isolated from the raw transposon mutant library and prepared for Illumina sequencing and sequenced on the HiSeq2000 platform. The sequencing was performed using a custom sequencing primer, sequencing from the 5′ end of the transposon and into the genomic DNA flanking the transposon insert.

In a sequencing run, one lane from an Illumina flow cell generated a minimum of 40 million reads of 43 bp plus index reads. The first 10 bp of each read constitutes the Tn sequence. Each lane was multiplexed with seven or eight samples, resulting in a minimum of 165 million nucleotides that represent the actual target DNA per sample. *S. aureus* S0385 has a total of 2777 annotated genes with an average length of 874 bp resulting in an average of 67x gene-coverage.

One mismatch was allowed when matching the Tn sequence. When using the HiSeq platform a lower quality of the Tn sequence was obtained in comparison to the quality of the target regions, as the Tn sequence is identical in all the reads. The sample used for validation had a total output of ~7.1 million reads and of these the Tn sequence was identified in ~6 million reads.

Tn sequence and adapter sequence were stripped and the reads (10–23 bp in length) were mapped to the reference genome. ~4.5 million reads were mapped exactly one time and 140,330 unique insertion sites were identified. The average distance between unique insertion sites was 20.5 bp and by utilising an average gene length of 911 bp (average gene length for genes containing an insert), the average number of unique inserts per gene was 44.8. The top row of Table 2 shows an overview of the transposon inserts recovered from the raw mutant library.

The distribution of the reads aligned to the reference chromosome is illustrated in Figure 1 by the right semicircle of the genome atlas. Reads are demonstrated as black spikes that are aligned to the reference genome, which is illustrated by the outermost green circle. The distribution of the aligned reads shows a high complexity transposon mutant library with inserts throughout the chromosome and no specific hotspots for transposon insertion.

Transposon insertion into a non-functional part of a gene may not disrupt gene function so it is necessary to define a threshold to separate essential/beneficial genes from non-essential genes. An insertion index was calculated by dividing the number of unique insertion sites for any given gene by the average gene length for genes containing an insert. Figure 2 illustrates a density plot based on the calculated insertion index for each gene. This plot separates genes with a low number of transposon inserts from genes with a high number of inserts (see Figure 2). The left most peak shows genes with a low number of inserts representing mutants with a decrease in fitness, which could lead to total loss of cellular survival or an arrested cell cycle, whereas the right most peak illustrates genes with a high number of inserts, representing viable mutants. The local minimum separating the peaks suggests that a cut-off value of around 0.02 would be suitable to distinguish essential/beneficial genes from non-essential genes.

The seven housekeeping genes *aroE*, *glpK*, *gmk*, *pta*, *tpiA*, *yqiL* and *arcC* used for Multi Locus Sequence Typing (MLST), shown in red in the left semicircle of Figure 1, represent potential candidates of essential genes within the *S. aureus* genome. One of the MLST genes (*tpiA* (SAPIG0853)) mapped zero reads, four genes mapped few reads (*pta* (SAPIG0662), *gmk* (SAPIG1207), *yqiL* (SAPIG0434) and *glpK* (SAPIG1302)) resulting in insertion indices below the cut-off (<0.02), identifying five of the MLST genes as essential/beneficial using this system. *aroE* (SAPIG1661) and *arcC* had insertion indices above the cut-off defining them as non-essential.

SAPIG2704 and SAPIG2129 (see Figure 1), shown in the left semicircle of Figure 1, encode serine-rich adhesin for platelets and cardiolipin synthetase, respectively, and are examples of two non-essential genes from the S0385 genome. A high number of reads mapped to these open reading frames, indicating that there was no significant loss of fitness when these genes were disrupted by transposon insertions.

LA-MRSA ST398 genes important for growth

The mutant library was grown for three passages in nutrient-rich broth at 37°C to identify genes essential for growth in this substrate. Table 2 shows an overview of the sequence analysis from passage 0 to passage 3. 71–75% of the reads containing the Tn tag sequence were found to map the reference genome. The number of unique insertion sites was between 97,000 and 162,000 with 31–51 unique insertion sites per gene. The number of unique insertion sites showed an initial decrease between passage 0 and passage 1. The decrease could illustrate that the transposon mutant library contains slow growing mutants, which will not be identified in the first growth passage. The passages were performed 3 times to increase selection sensitivity and to reduce the presence of arrested and dead cells.

A total of 152 genes mapping zero reads were identified from the mutant pool after three passages under laboratory conditions – of these, 100 were protein-coding genes, 4 encoded ribosomal RNAs (rRNA) and 48 transfer-RNAs (tRNAs). These genes are proposed to be essential for bacterial survival under laboratory conditions. In addition, 526 genes had only a few mapped reads and had an insertion index below the calculated cut-off of 0.02, indicating that these may also be important for growth (Tables S1 and S2). Genes with few transposon insertion sites may have maintained gene but cannot be identified as true essential genes and are therefore referred to as genes beneficial for bacterial survival under laboratory conditions. The protein coding genes were categorised based on functionality using the COG database and plotted as percentage of all the COG categorised genes in the WT (see Figure 3). Some genes were categorised as belonging to several COGs. The proposed essential gene list includes representatives of all the major functional COGs except group B (chromatin structure and dynamics) and N (cell motility). Representatives in V (defence mechanisms) were only identified when including the genes with few inserts (insertion index <0.02). Protein-coding genes involved in translation (COG group J), cell division (COG group D), coenzyme transport and metabolism (COG group H), and intracellular trafficking, secretion and vascular transport (COG group U) had the largest number of representatives in the proposed essential and beneficial gene sets. Approximately 9% of the proposed essential and beneficial protein-coding genes were of unknown function or not related to any COG group.

Table 2. Overview of the raw Transposon mutant Library and the passages in BHI - Illumina sequence data.

	Total no. of reads	Read with Tn tag (≤1 mismatch)	Reads mapped exactly 1 time	No. of unique insertion sites	Average no. of unique insertion sites per gene
Raw library	7,129,995	6,070,601	4,503,675 (75.88%)	140,330	44.8
Passage 0	7,564,547	5,931,390	4,284,574 (73.97%)	136,440	42.4
Passage 1	10,503,621	8,586,527	6,003,415 (71.14%)	97,236	31.2
Passage 2	10,316,723	8,481,909	6,017,839 (72.35%)	115,921	37
Passage 3	13,618,447	11,261,919	7,899,885 (71.54%)	162,228	51

The table shows the output from the raw transposon mutant library and the three passages in BHI. The number of reads recovered after trimming and alignment were identified and the number of unique insertion sites per gene was calculated. The sequence data of the raw mutant library was obtained from one lane of a flow cell which was multiplexed with eight samples. The sequence data from the three passages were obtained from one lane of a flow cell that was multiplexed with seven samples. The sequencing was performed on a Hiseq2000 platform.

doi:10.1371/journal.pone.0089018.t002

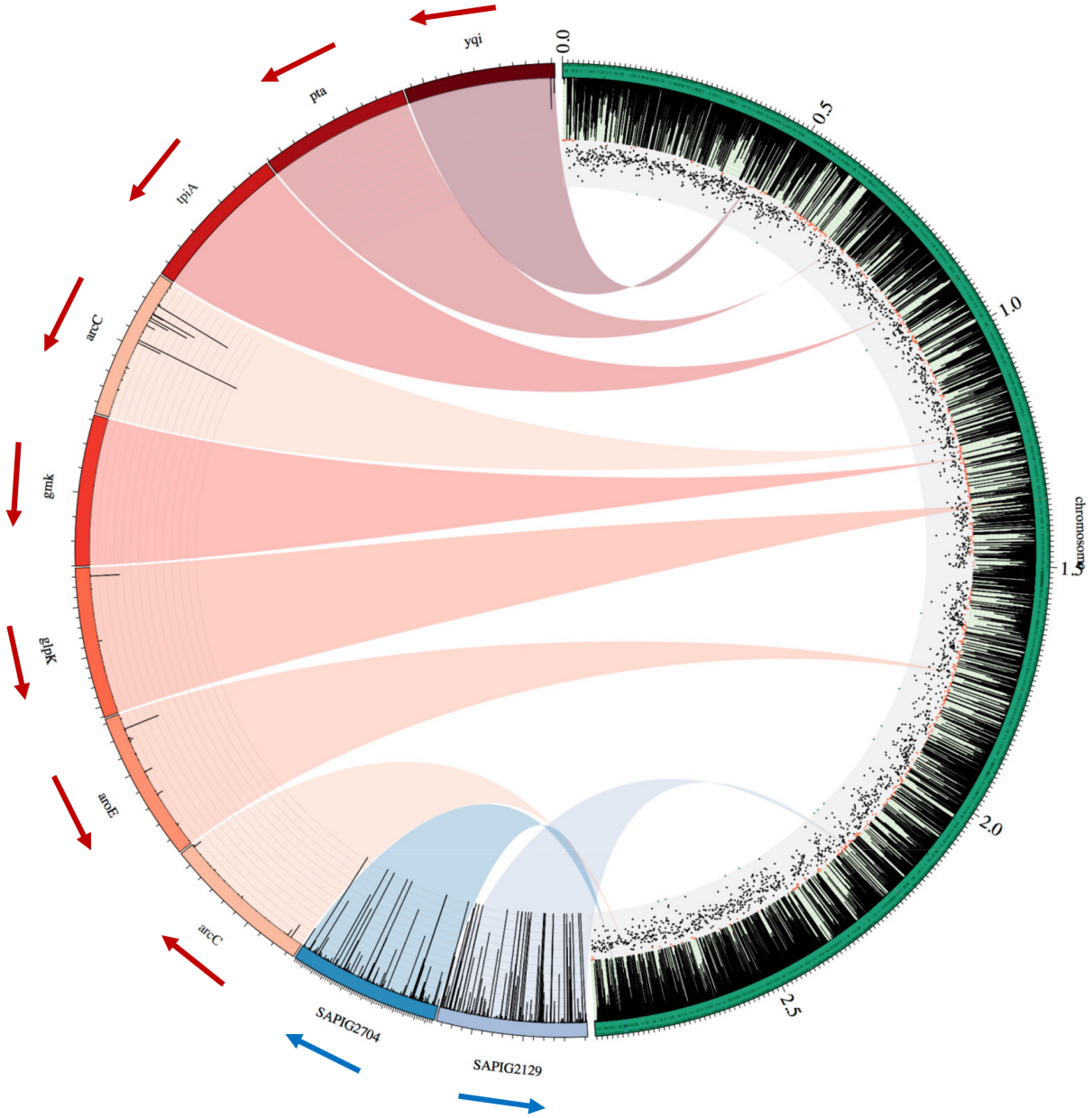


Figure 1. Genome atlas. Right semicircle: The green band in the outermost part of the semicircle illustrates the reference chromosome (AM990992) with the size of 2,872,582 bp. The three circular plasmids harboured by the reference are not included. The black spikes connected to the green semicircle shows the distribution of the reads from the raw transposon mutant library aligned to the reference strain. The black and red dots indicate positions within the reference with large number of reads (insertion index >0.02) and low number of reads (insertion index <0.02) respectively. Left semicircle: The red colours show of zoom of the seven MLST genes (*arcC* represented twice due to two copies of this particular gene) and the black spikes illustrated in some of the genes show reads mapping within the open reading frame. The arrows indicate transcription direction. The zoom of SAPIG2704 and SAPIG2129, visualised in blue colours, show examples of two genes with a large number of read mapping throughout the open reading frames.
doi:10.1371/journal.pone.0089018.g001

Survival in whole porcine blood

The mutant library was grown in porcine blood and DNA from mutants pre- and post- selection was prepared as input and output pools, respectively. The blood samples were inoculated with the transposon mutant library and incubated for 24 hours. Previous growth experiments in whole porcine blood showed an initial decrease in cell counts but after 24 hours of incubation the number of mutants returned to a population size equivalent to the inoculums (see Figure S4 for more details). The total number of reads corresponding to transposon insertion sites in the input pool was compared to the total number of reads mapping to the equivalent position in the output data. The read counts are expected to follow an approximately normal distribution but the data showed some noise in the lower end and read counts below 2^4 were considered as noise based on a frequency distribution plot (data not shown). Using the DESeq package in R the effective size of each sequence library was estimated based on the read counts and the estimated size factors were used for normalization of the

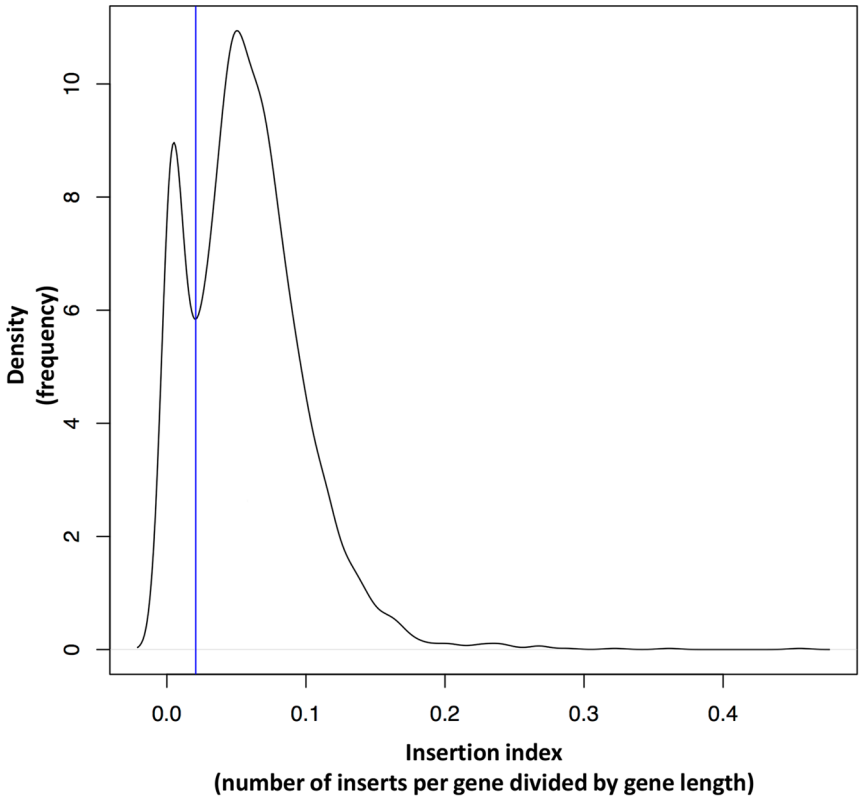


Figure 2. Density plot - Insertion index distribution. The figure shows a density plot illustrating the distribution of insertion indices (number of transposon inserts per gene divided by an average gene length). The plot indicates the density according to which the insertion indices are distributed and it shows that the insertion indices have a bimodal frequency distribution. The leftmost peak represents the genes with zero or very few insertions, whereas the rightmost peak represents the genes with a large number of insertions. The vertical line piercing the local minimum and separating the two peaks, defines the cut-off sorting genes as either, essential/beneficial or non- essential/neutral for bacterial fitness in a given environment.
doi:10.1371/journal.pone.0089018.g002

data. To contrast the two conditions and highlight a possible differential composition in mutants, recovered pre- and post-selection, the variance of reads mapping each gene was estimated and subsequently tested using a negative binomial test. Ratios of normalized read counts in the input and output samples were determined and expressed as a $\log_2$ fold change. A negative $\log_2$ fold change corresponds to a decrease in read counts from input to output and indicates attenuated mutants, whereas a positive $\log_2$ fold change reflects an increase in read counts from input to output.

Only the mutants that were uniquely attenuated under the selective conditions were of interest. The mutant composition pre-selection in whole porcine blood was compared to the mutant composition post-selection. The genes representing the mutants with the most significant change in clone number were identified. To eliminate general selection due to growth in BHI the mutant library was selected for an equivalent number of growth rounds in BHI and genes representing mutants with the most significant change in clone number were identified. The two gene lists were compared and the genes specific for survival in whole porcine blood were identified (see Figure S5).

Transposon inserts in 23 genes induced a significant decrease in fitness (negative $\log_2$ fold change) and transposon inserts in one gene induced a significant hypercompetitive mutant (positive $\log_2$ fold change), all as a consequence of being selected in porcine blood (see Table 3). Six of the mutants, illustrated with a minus infinity (-inf) $\log_2$ fold change in Table 3, were represented in the input pool but totally absent, with zero read counts, in the output pools. Seven of the 23 genes are defined as encoding hypothetical proteins with unknown function. Additionally two genes were of unknown function, whereas fifteen could be assigned a potential function (see Table 3 and Table 4).

Discussion

The purpose of the work was to generate a high complexity transposon mutant library and assess the application of TraDIS in *S. aureus* ST398. LA-MRSA ST398 was selected for this study as it shows different host infection/colonization patterns compared to most other MRSA strains. The isolation of MRSA from animals was first reported in 1972 [29], but was at that time most likely associated with human to animal transmission of an MRSA strain acquired by the farmer during hospitalisation. More recently, a specific lineage belonging to CC398, most likely of human origin, has spread among livestock globally, acquired methicillin resistance and is now transferring back to humans leading to both colonisation and disease [17,30]. ST398 is able to adapt to various host environments and continues to emerge worldwide both in livestock and also to some extent in hospital settings [31].

When interpreting the data it is important to recognize that the environment and other factors resulting from the experimental design can have unintended consequences on the output data. Nutrient-broth was supplemented with erythromycin to maintain the genomic insertion of the transposon and high temperatures

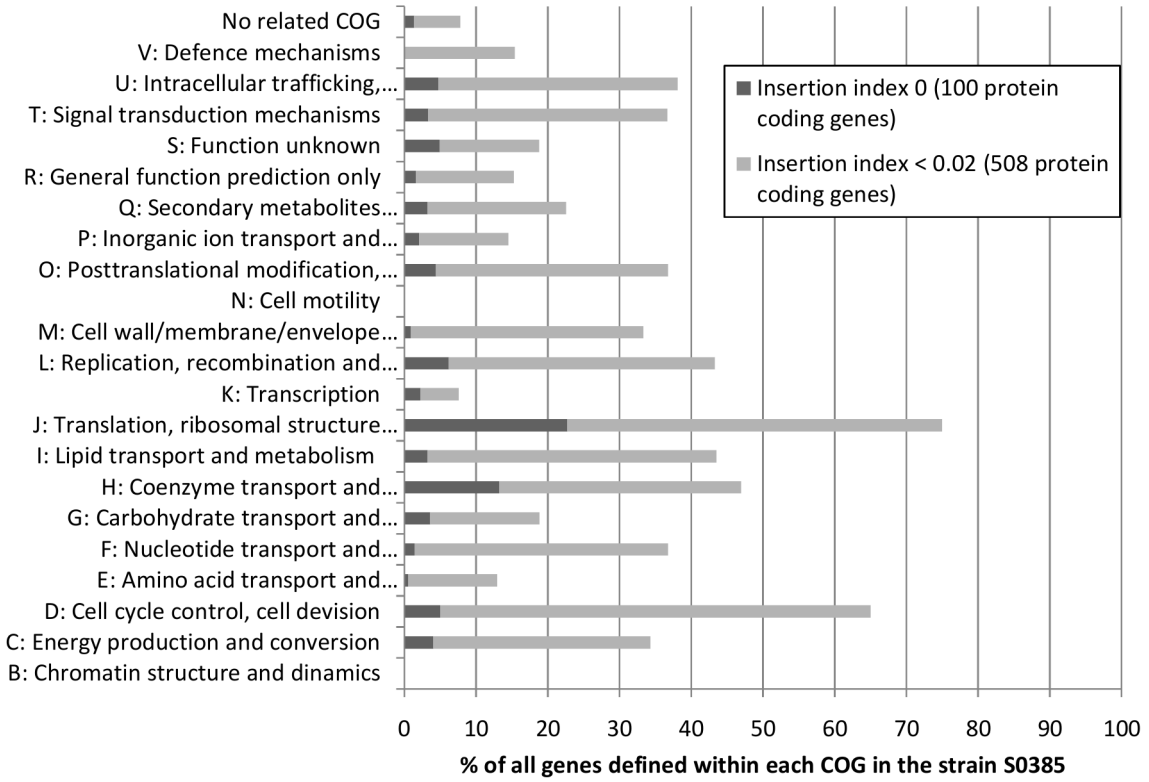


Figure 3. Proposed essential genes classified by functionality. The proposed essential genes for growth under laboratory conditions were classified by functionality and plotted as a percentage of all genes within each functional group encoded by the reference strain. The genes were assigned a functionality based on the COG database and these groups are illustrated on the vertical axis. The dark grey columns represent the proposed essential protein-coding genes with zero inserts, whereas the light grey columns add the protein-coding genes with few inserts (insertion index <0.02), which were proposed beneficial for growth under laboratory conditions.
doi:10.1371/journal.pone.0089018.g003

were applied to promote plasmid loss, both of which may influence the output when screening for essential genes. For example it has been reported that incubation at high temperatures in the presence of erythromycin enriches for mutants of the *sae* system, which is a two-component system involved in regulation of some virulence genes [32]. Enrichment of a regulatory system could have an unattended effect on the transposon mutant composition. In addition, transposon insertions may affect the expression of downstream genes or operons, causing polar mutation that leads to incorrect identification of essential genes in a defined environment. For definitive identification of gene function it is necessary to generate single knockout mutants and test those in the same functional assays used in the screenings. However, since a large number of genes are listed as having no known function and there is inconsiderable value in generating evidence for the phenotypes resulting from the possession of these genes, high-throughput methods can help to narrow the pool of genes to be investigated further.

In our study we generated a transposon mutant library consisting of $\sim 10^6$ mutants and we identified around 140,000 unique insertion sites. The transposon mutant library generated in *Salmonella* Typhi by Langridge *et al.* (2009) [9] yielded 370,000 unique insertion sites, which may be explained by the fact that the *Salmonella* genome is more than 2 Mb larger than the *S. aureus* genome and so provides the potential for a higher number of unique insertion sites. Langridge *et al.* showed an average of one insertion site for every 15–20 bp, which was similar in this study showing an average insertion site for every 20.5 bp. The sequence data (Table 2), linker PCR data (Figures S2 and S3) and the coverage atlas (Figure 1) showed a successful generation of a high complexity mutant library with transposon inserts throughout the bacterial genome, comparable to the mutant library generation in *Salmonella* Typhi [9].

The MLST genes are housekeeping genes and are expected to be essential for cell viability [33]. However not all seven MLST genes were defined as essential in this study. The *glpK*, *gmk*, *pta*, *tpiA* and *yqiL* MLST genes were identified as essential or beneficial with zero or few transposon inserts, whereas *aroE* and *arcC* were defined as non-essential. *tpiA*, *pta*, *gmk* and *yqiL* have all been identified as essential previously (see Table S3) [6,8,12]. The *arcC* gene encoding carbamate kinase has a paralogous gene at a different locus within the S0385 genome, which also encodes carbamate kinase. When one of the *arcC* homologues is disrupted by the transposon insert the transcript of the other may take over and this could explain an insertion index above the cut-off for both *arcC* genes (SAPIG1164 and SAPIG2682).

Gene SAPIG2704 and SAPIG2129, which encode serine-rich adhesin for platelets and cardiolipin synthetase respectively, constitute two examples of genes defined as non-essential for S0385 survival under laboratory conditions in this study. Figure 1 illustrates that a high number of reads mapped within these open reading frames. Serine-rich adhesins are postulated to be important for bacterial binding to platelets as part of the pathogenesis in infective endocarditis in humans [34]. The S0385 isolate was isolated from a human case of endocarditis [35], where serine-rich adhesins may be essential, but when

Table 3. Genes representing 23 attenuated mutants and 1 hypercompetitive mutant when selected in whole porcine blood.

ID (gene)	Read Count Input	Read Count Output		
Mean	**Mean**	**Log2**	**Fold Change**	***P*-value**
SAPIG2099	428.16	0.00	-inf	0.0237
SAPIG1465	317.79	0.00	-inf	0.0156
SAPIG2108	203.34	0.00	-inf	0.0288
SAPIG0429	196.84	0.00	-inf	0.0298
SAPIG1848	164.21	0.00	-inf	0.0354
SAPIG0633	143.38	0.00	-inf	0.0398
SAPIG0142	726.16	1.10	−9.3631	0.0024
SAPIG1650	492.68	1.10	−8.8088	0.0128
SAPIG1041	799.62	2.20	−8.5048	0.0025
SAPIG1748	352.17	1.10	−8.3244	0.0249
SAPIG1921	306.01	1.10	−8.1163	0.0430
SAPIG0315	605.24	2.21	−8.1003	0.0090
SAPIG2670	256.41	1.10	−7.8613	0.0406
SAPIG2057	469.89	2.21	−7.7351	0.0203
SAPIG1726	853.28	4.40	−7.5998	0.0038
SAPIG1977	526.66	3.30	−7.3165	0.0272
SAPIG0258	446.43	3.31	−7.0762	0.0426
SAPIG1054	1430.32	12.11	−6.8836	0.0019
SAPIG1096	694.20	6.60	−6.7168	0.0137
SAPIG2156	529.71	5.51	−6.5882	0.0296
SAPIG0647	1739.26	31.92	−5.7678	0.0497
SAPIG2568	751.50	15.43	−5.6056	0.0259
SAPIG2639	1449.27	77.16	−4.2314	0.0360
SAPIG0185	120.11	2868.20	4.5777	0.0328

The table lists the mutants that significantly changed in clone number from input to output (pre- and post-selection in whole porcine blood). The top 23 genes represent the mutants that were significantly reduced in number of clones after selection in whole porcine blood. The lowermost gene represents the mutant that significantly increased in clone number after selection in whole porcine blood. Mean read count input and Mean read count output represent the mean number of reads mapping within the defined gene. The differences between the mean values are illustrated by a log fold change from input to output and a negative $\log_2$ fold change indicating changes in fitness. A negative $\log_2$ fold change defines attenuation in fitness whereas a positive $\log_2$ fold change defines increase in fitness. The p-value shows the level of significance.

doi:10.1371/journal.pone.0089018.t003

transferring the isolate to a laboratory environment, these adhesins might lose their importance for bacterial survival. Cardiolipin synthetase are involved in conversion of bacterial membrane phosphatidylglycerol (PG) to cardiolipin (CL) when the bacteria progress from exponential growth phase to stationary and when phagocytosed by human neutrophils [36]. The *S. aureus* S0385 genome contains two open reading frames (Cls1: SAPIG1324 and Cls2: SAPIG2129) encoding cardiolipin synthetases. Cls2 is primarily responsible for CL accumulation under stationary phase [36], but when SAPIG2129 encoding Cls2 is disrupted by transposon insert, the homologous Cls1 may take over. The examples above illustrate the sensitivity of this methodology for identifying essential/beneficial or non-essential genes.

A total of 152 *S. aureus* S0385 genes had zero transposon inserts and were therefore proposed as essential genes, while 526 genes, with a low number of transposon inserts, were proposed as beneficial for growth under laboratory conditions. Table S1 shows the lists of proposed essential genes and Table S2 the list of proposed beneficial genes. Table S3 shows a comparison with previously described *S. aureus* essential genes using high complexity transposon mutant libraries [6,8,12].

Of the 526 genes (insertion index <0.02) proposed here as beneficial, 268 genes have been described as essential in *S. aureus* previously (see Table S2). The 258 proposed beneficial genes that have not been described as essential previously encode proteins involved in DNA repair, replication and recombination, which indicate that the high temperatures applied to promote plasmid loss under the mutant library construction induced as expected bacterial stress conditions. These genes are therefore evaluated as beneficial for ST398 survival in this study due to the specific conditions applied in the experimental setup. When ranking the genes with insertion indices <0.02, it is clear that, as the insertion index increases and approaches the cut-off (0.02), there is an increase in number of genes that have not been described as essential in *S. aureus* previously (see Table S2). The ranking and knowledge from previous studies could indicate an insertion index cut-off of approximate 0.007 instead of 0.02. This shows that the selection of the cut-off separating essential/beneficial from non-essential genes is an important consideration.

The differences found between this study and previous studies defining essential genes could be due to differences in methodology, sensitivity of the methods, environmental conditions or true differences between bacterial strains. However, the results need to

Table 4. Description of the genes identified as important for *S. aureus* ST398 survival in whole porcine blood.

ID (gene)	Description	Process	Whole blood survival
SAPIG2099 (*leuD*)	3-isopropylmalate dehydratase, small subunit	Leucine biosynthesis (amino acid biosynthesis)	Oxidative stress and pH shock. Stringent response (cellular adaptation to nutrient limiting conditions).
SAPIG1465 *(aroB)*	3-dehydroquinate synthase	Nucleotide and amino acid metabolism (aromatic amino acid metabolism)	Oxidative stress and pH shock.
SAPIG0429	Hypothetical protein	Unknown	?
SAPIG2108	Phosphoserine phosphatase, RsbU	Up-regulation of σ^B (alternative sigma factor)	σ^B influences expression of a variety of genes including virulence genes under stress and specific environmental conditions.
SAPIG1848	Hypothetical protein	Unknown	?
SAPIG0633	tRNA-specific adenosine deaminase	Unknown	?
SAPIG0142	NAD dependent epimerase/dehydratase family protein	Galactose metabolism	Glucose depletion. Galactose metabolism (galactose molecules compose important components of the surface bound antigens located on red blood cells).
SAPIG1650 *(lepA)*	GTP-binding protein	Specific function unknown	LepA protein homologous to translation factors that binds ribosomes.
SAPIG1041 *(menD)*	2-succinyl-6-hydroxy-2, 4-cyclohexadiene-1-carboxylic acid synthase/2-oxoglutarate decarboxylase	Menaquinone biosynthetic pathway	Respiration. Involved in protection against haem toxicity
SAPIG1748 *(icd)*	Isocitrate dehydrogenase (IDH), NADP-dependent (icd gene)	Regulation of tricarboxylic acid (TCA) cycle	Icd up-regulation under acidic conditions. Regulation of the TCA cycle.
SAPIG1921	RNA methyltransferase, TrmH family, group 2	RNA metabolism	Regulation – balance between transcript and degradation of mRNA.
SAPIG0315	Hypothetical protein	Unknown	?
SAPIG2670	Hypothetical protein	Unknown	?
SAPIG2057	Aspartate transaminase	Aminoacid metabolism.	Decrease in pH.
SAPIG1726	HemA concentration negative effector hemX	Transport	ABC-type transport system. C ytochrome c biogenesis.
SAPIG1977	Response regulator protein VraR	Regulator of cell wall damage stress response	Response to cell wall damage.
SAPIG0258	PTS system galactitol-specific enzyme II B component	Galactose metabolism	Glucose depletion. Galactose metabolism (galactose molecules compose important components of the surface bound antigens located on red blood cells).
SAPIG1054	Beta-lactamase		
SAPIG1096	Spermidine/putrescine ABC transporter ATP-binding subunit	ABC transporter involved in ion homeostasis	pH shock/changes.
SAPIG2156	Hypothetical protein	Unknown	?
SAPIG0647	Indigoidine systhesis protein	Secondary metabolite composing a blue pigment.	Oxidative stress – ph shock.
SAPIG2568 *(fbp)*	Fructose-1,6-bisphosphatase	Gluconeogenesis	Response to depletion of glucose.
SAPIG2639 *(pyrD)*	Dihydroorotate oxidase	Pyrimidine biosynthesis	Nucleic acids biosynthesis.
SAPIG0185	pANL51	Unknown function	?

doi:10.1371/journal.pone.0089018.t004

be verified by additional studies to provide further evidence of the essential nature of these genes.

None of the proposed essential genes were defined within the group of Defence mechanisms (COG group V), but four of the proposed beneficial genes were categorised as belonging to COG group V. These four genes, SAPIG1054, SAPIG1375, SAPIG1376 and SAPIG2314, encode beta-lactamase *ampC* and aminoacyltransferase *femA*, *femB* and *femX* respectively. *femA*, *femB* and *femX* have been identified as essential genes in previous studies [6,8] and it has been shown that *femA* and *femB* mutants have a reduced peptidoglycan (PG) glycin content compared with $femA^+$ and $femB^+$ strains [37,38]. The staphylococcal cell wall plays an important role in infection and pathogenicity, but based on our data these cell-wall impairments may also have wider influence on cell growth and survival in general. However, it has also been demonstrated that *femAB* null mutants harbouring an erythromycin resistance marker lead to a low level of erythromycin resistance, which may be due to a higher uncontrolled influx of erythromycin through the impaired cell-wall [39]. The presence of erythromycin in the nutrient-broth used in this study could explain the decreased fitness identified for the *femA*, *femB* and *femX* mutants.

Overall 24 genes were identified with a significant change in fitness after whole porcine blood incubation. Twenty-three of these genes were identified as giving a significant reduction in bacterial fitness when inserted with a transposon and selected *in vitro* in

porcine blood. Mutation in one gene resulted in a hypercompetitive mutant post-selection in whole porcine blood.

No specific cell viability tests were performed on the blood cells, but it has been shown previously that whole-blood units stored at room temperature maintained cellular counts and coagulation activity for up to 72 hours [40]. In addition, in previous experiments an initial decrease in bacterial cell counts was observed when incubating the transposon mutant library in whole porcine blood, which could reflect neutrophil killing. It is therefore reasonable in this case to believe that the genes identified are important for survival in whole porcine blood under *in vitro* conditions.

The 23 genes identified in the attenuated mutants represent mutants showing the greatest reduction in cell count when comparing input and output pools. However, they are unlikely to be the only genes important for survival in porcine blood. For example, mutants with transposon inserts in essential genes are absent in the input pools and a potential difference between input and output pools for those essential genes will not be detected and they can therefore not be considered as important for whole porcine blood survival in this experiment.

Seven of the 24 genes are defined as hypothetical genes of unknown function and two other proteins were annotated with unknown function. Fifteen genes were annotated to be predominantly involved in carbon metabolism, pH shock, regulation and transport (see Table 4) [41–47]. This indicates that key genes for survival in porcine blood cultures may not be genes involved in iron uptake such as hemolysins and sideophors, but may be genes associated with the ability to utilize the available carbon hydrates in blood, regulation at different levels as well as survival under extreme pH conditions. This is supported by previous studies analysing global gene expression of *S. aureus* under *in vitro* conditions of short-term culture in human blood [48,49]. In these studies, it was observed that up- or down regulated genes were mainly involved in cellular metabolism or had an unknown function. A previous study screening 1248 transposon *S. aureus* mutants in an *in vivo* murine bacteraemia model identified 50 genes as being important for whole blood survival, half of which had unknown function and the rest with an involvement in nutrient biosynthesis and surface metabolism [50]. Furthermore they identified genes important for the tricarboxylic acid cycle (TCA cycle) and in this study we identified the *icd* gene, a TCA cycle regulator, as important for *in vitro* survival in porcine blood. This indicates that the TCA cycle and carbon metabolism, have important functions for bacterial survival in blood *in vivo* and *in vitro* and in blood from different hosts. The *fem*A and *fem*B genes were previously identified as important for whole blood survival *in vivo* [50]. However, we found *fem*A and *fem*B mutants to have a growth disadvantage under laboratory conditions which is in correlation with other studies identifying *S. aureus* essential genes [6,12,13].

The transposon mutant library was incubated in whole porcine blood *in vitro* for 24 hours. This could partly reflect why many metabolic genes were identified as important for whole porcine blood survival in this study. However, an incubation period of 24 hours was specifically selected based on initial growth experiments in whole porcine blood *in vitro* (Figure S4). These experiments showed an initial decrease in bacterial population size, which could be explained by phagocytosis and potential bacterial killing by host immune cells. The mutant population size returned to an equivalent size of the inoculated population after 24 hours, and at this point the mutants had potentially seen all the selective elements within whole blood. Genes important for immune evasion will have undergone selection in a similar manner as the metabolic genes. *S. aureus* encodes however various immune evasion genes and it is justifiable to conclude that none of these are singlehandedly responsible for survival of the immune response, which could explain why none of these genes were identified as important for whole blood survival. Even though no specific virulence genes were identified as being important for blood survival in this study they might have important functions in more specific infection models.

In this study, we successfully generated a high complexity transposon mutant library in an LA-MRSA ST398 WT isolate and evaluated it using the TraDIS system. We identified *S. aureus* ST398 essential genes comparable with previous studies. Twenty-four genes were evaluated as being important for specific *in vitro* whole porcine blood survival, of which carbon metabolism, pH shock and regulation were related. For further evaluation of these genes, we aim to generate single knockout mutants and test these for survival in porcine blood, as well as in blood from other relevant donors. In addition, the generated transposon mutant library will be used in a screen for survival and colonization in other host relevant environments such as on porcine skin and nasal epithelium.

Supporting Information

Figure S1 Commands and settings used in R for the statistical analysis.
(TIF)

Figure S2 Whole mutant library and single colony verification. The gels show the result of the linker PCR used for library validation. The left gel shows squared in red a low complexity mutant library with a laddering of the smears. The blue squared lanes illustrate the same high complexity transposon mutant library from passage 0 (lane 2) to passage 3 (lane 5). The third generation transposon mutant library shows a smear with no specific bands. The right gel represents 15 randomly picked single mutant colonies isolated from the third generation transposon mutant library, each giving a band of different size indicating that the transposon has inserted at different locations with the genome.
(TIF)

Figure S3 Genome atlas identifying transposon inserts of 11 random isolated mutants. The genome atlas illustrates by black marks in the outer most circle 11 different transposon insertion sites within the reference genome. The insertion sites were identified based on sequencing 11 of the 15 randomly picked mutant colonies described in figure S1. The fragments from the 11 mutants were sequenced and aligning to the reference genome. The blue and red parts of the atlas indicate forward and reverse transcriptional direction of the open reading frames within the reference genome.
(TIF)

Figure S4 Growth profile of transposon mutant library in whole porcine blood *in vitro*. The figure shows the growth profile of the transposon mutant population in whole porcine blood *in vitro*. Mutant population size was determined at specific time-points to identify functionality of the blood immune cells. After 24 hours incubation *in vitro* the mutant population size was equivalent to the inoculated population size (indicated by the red circle).
(TIF)

Figure S5 Experimental setup for identification of genes important for bacterial growth in whole porcine blood. The mutant composition in input pool pre-selection in whole porcine blood (Input pool - library aliquot) were compared with mutant composition in output pool post-selection in porcine blood

(output pool BHI – second generation library). The mutants identified with a significant change in number of clones represent genes important for whole porcine blood survival in addition to growth BHI. The mutant composition in output pool post-selection in porcine blood (output pool BHI – second generation library) was compared to mutant composition after growth in BHI (BHI – second generation library). The mutants identified with a significant change in number of clones in both of the comparisons were evaluated as specific for survival in whole porcine blood *in vitro*.
(TIF)

Table S1 Proposed essential genes.
(XLSX)

Table S2 Proposed beneficial genes.
(XLSX)

Table S3 Comparison of essential gene lists of *S. aureus*.
(XLSX)

Acknowledgments

We thank Alex Hurd from University of Sheffield, UK for the plasmids used in the cloning construct. We acknowledge Morten Rasmussen at Copenhagen University, DK for indispensable help in the library preparation setup and the staff at the Sequencing Centre at Copenhagen University, DK for generating the sequences. In addition we thank J. H. Wang and Dan Tucker from the Department of Veterinary Medicine, University of Cambridge, UK for help with the transposon mutant library verification and collection of porcine blood.

Author Contributions

Conceived and designed the experiments: MTC HH FMA. Performed the experiments: MTC. Analyzed the data: MTC RSK RRC FMA. Contributed reagents/materials/analysis tools: MTC RSK RRC MAH HH FMA. Wrote the paper: MTC RSK RRC MAH HH FMA.

References

1. Fraser CM, Gocayne JD, White O, Adams MD, Rebecca A, et al. (2013) The Minimal Gene Complement *Mycoplasma genitalium* of. Science 270: 397–403.
2. Schneiker S, Perlova O, Kaiser O, Gerth K, Alici A, et al. (2007) Complete genome sequence of the myxobacterium *Sorangium cellulosum*. Nature biotechnology 25: 1281–1289.
3. Hall N (2007) Advanced sequencing technologies and their wider impact in microbiology. The Journal of experimental biology 210: 1518–1525.
4. Salama N, Shepherd B, Falkow S (2004) Global transposon mutagenesis and essential gene analysis of *Helicobacter pylori*. Journal of bacteriology 186: 7926–7935.
5. Liberati NT, Urbach JM, Miyata S, Lee DG, Drenkard E, et al. (2006) An ordered, nonredundant library of *Pseudomonas aeruginosa* strain PA14 transposon insertion mutants. Proceedings of the National Academy of Sciences of the United States of America 103: 2833–2838.
6. Bae T, Banger AK, Wallace A, Glass EM, Aslund F, et al. (2004) *Staphylococcus aureus* virulence genes identified by bursa aurealis mutagenesis and nematode killing. Proceedings of the National Academy of Sciences of the United States of America 101: 12312–12317.
7. Sassetti CM, Boyd DH, Rubin EJ (2003) Genes required for mycobacterial growth defined by high density mutagenesis. Molecular microbiology 48: 77–84.
8. Chaudhuri RR, Allen AG, Owen PJ, Shalom G, Stone K, et al. (2009) Comprehensive identification of essential *Staphylococcus aureus* genes using Transposon-Mediated Differential Hybridisation (TMDH). BMC genomics 10: 291.
9. Langridge GC, Phan M, Turner DJ, Perkins TT, Parts L, et al. (2009) Simultaneous assay of every *Salmonella* Typhi gene using one million transposon mutants. Genome research 19: 2308–2316. doi:10.1101/gr.097097.109
10. Khatiwara A, Jiang T, Sung S-S, Dawoud T, Kim JN, et al. (2012) Genome scanning for conditionally essential genes in *Salmonella enterica* Serotype Typhimurium. Applied and environmental microbiology 78: 3098–3107.
11. Pickard D, Kingsley RA, Hale C, Turner K, Sivaraman K, et al. (2013) A Genomewide Mutagenesis Screen Identifies Multiple Genes Contributing to Vi Capsular Expression in *Salmonella enterica* Serovar Typhi. Journal of bacteriology 195: 1320–1326.
12. Fey PD, Endres JL, Yajjala VK, Yajjala K, Widhelm TJ, et al. (2013) A Genetic Resource for Rapid and Comprehensive Phenotype. MBio 4: e00537–12. doi:10.1128/mBio.00537-12.Editor
13. Chaudhuri RR, Morgan E, Peters SE, Pleasance SJ, Hudson DL, et al. (2013) Comprehensive Assignment of Roles for *Salmonella* Typhimurium Genes in Intestinal Colonization of Food-Producing Animals. PLoS Genetics 9: e1003456.
14. Plata K, Rosato AE, Wegrzyn G (2009) *Staphylococcus aureus* as an infectious agent: overview of biochemistry and molecular genetics of its pathogenicity. Acta biochimica Polonica 56: 597–612.
15. Hasman H, Moodley A, Guardabassi L, Stegger M, Skov RL, et al. (2010) Spa type distribution in *Staphylococcus aureus* originating from pigs, cattle and poultry. Veterinary microbiology 141: 326–331.
16. Sung JM-L, Lloyd DH, Lindsay JA (2008) *Staphylococcus aureus* host specificity: comparative genomics of human versus animal isolates by multi-strain microarray. Microbiology (Reading, England) 154: 1949–1959.
17. Price L, Stegger M, Hasman H, Aziz M (2012) *Staphylococcus aureus* CC398: host adaptation and emergence of methicillin resistance in livestock. MBio 3: e00305–11.
18. Uhlemann A, Porcella S, Trivedi S (2012) Identification of a highly transmissible animal-independent *Staphylococcus aureus* ST398 clone with distinct genomic and cell adhesion properties. MBio 3: e00027–12.
19. Van Duijkeren E, Ikawaty R, Broekhuizen-Stins MJ, Jansen MD, Spalburg EC, et al. (2008) Transmission of methicillin-resistant *Staphylococcus aureus* strains between different kinds of pig farms. Veterinary microbiology 126: 383–389.
20. Cuny C, Friedrich A, Kozytska S, Layer F, Nübel U, et al. (2010) Emergence of methicillin-resistant *Staphylococcus aureus* (MRSA) in different animal species. International journal of medical microbiology: IJMM 300: 109–117.
21. Schijffelen MJ, Boel CHE, van Strijp JAG, Fluit AC (2010) Whole genome analysis of a livestock-associated methicillin-resistant *Staphylococcus aureus* ST398 isolate from a case of human endocarditis. BMC genomics 11: 376.
22. Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet journal 17: 10–12.
23. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nature methods 9: 357–359.
24. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, et al. (2009) Circos: an information aesthetic for comparative genomics. Genome research 19: 1639–1645.
25. Anders S, Huber W (2010) Differential expression analysis for sequence count data. Genome biology 11: R106.
26. Anders S, Huber W (2012) Differential expression of RNA-Seq data at the gene level – the DESeq package. Heidelberg, Germany: European Molecular Biology Laboratory (EMBL).
27. Tatusov RL (1997) A Genomic Perspective on Protein Families. Science 278: 631–637.
28. Tatusov RL, Galperin MY, Natale DA, Koonin EV (2000) The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic acids research 28: 33–36.
29. Devriese L, Van Damme R, Fameree L (1972) Methicillin (cloxacillin)-resistant *Staphylocoocus aureus* strains isolated from bovine mastitis cases. Vet Med 19: 598–605.
30. Verkade E, Kluytmans J (2013) Livestock-associated *Staphylococcus aureus* CC398: Animal reservoirs and human infections. Infection, genetics and evolution.
31. Smith TC, Pearson N (2011) The emergence of *Staphylococcus aureus* ST398. Vector borne and zoonotic diseases (Larchmont, NY) 11: 327–339.
32. Sun F, Cho H, Jeong D-W, Li C, He C, et al. (2010) Aureusimines in *Staphylococcus aureus* are not involved in virulence. PloS one 5: e15703.
33. Saunders NA, Holmes A (2007) Multilocus Sequence Typing (MLST) of *Staphylococcus aureus*. Methods in molecular biology (Clifton, NJ) 391: 71–85.
34. Siboo I, Chambers H, Sullam P (2005) Role of SraP, a serine-rich surface protein of *Staphylococcus aureus*, in binding to human platelets. Infection and immunity 73: 2273–2280.
35. Fluit AC (2012) Livestock-associated Staphylococcus aureus. Clinical Microbiology and Infection 18: 735–744.
36. Koprivnjak T, Zhang D, Ernst CM, Peschel A, Nauseef WM, et al. (2011) Characterization of *Staphylococcus aureus* cardiolipin synthases 1 and 2 and their contribution to accumulation of cardiolipin in stationary phase and within phagocytes. Journal of bacteriology 193: 4134–4142.
37. Henze U, Sidow T, Wecke J, Labischinski H, Berger-Bächi B (1993) Influence of *femB* on methicillin resistance and peptidoglycan metabolism in *Staphylococcus aureus*. Journal of bacteriology 175: 1612–1620.
38. Maidhof H, Reinicke B, Blümel P, Berger-Bächi B, Labischinski H (1991) *femA*, which encodes a factor essential for expression of methicillin resistance, affects glycine content of peptidoglycan in methicillin-resistant and methicillin-susceptible *Staphylococcus aureus* strains. Journal of bacteriology 173: 3507–3513.
39. Rohrer S, Ehlert K, Tschierske M, Labischinski H, Berger-Bächi B (1999) The essential *Staphylococcus aureus* gene *fmhB* is involved in the first step of

peptidoglycan pentaglycine interpeptide formation. Proceedings of the National Academy of Sciences of the United States of America 96: 9351–9356.
40. Hughes JD, Macdonald VW, Hess JR (2007) Warm storage of whole blood for 72 hours. Transfusion 47: 2050–2056.
41. Somerville G, Morgan C (2002) *Staphylococcus aureus* aconitase inactivation unexpectedly inhibits post-exponential-phase growth and enhances stationary-phase survival. Infection and immunity 70: 6373–6382.
42. Prasad UV, Vasu D, Kumar YN, Kumar PS, Yeswanth S, et al. (2013) Cloning, Expression and Characterization of NADP-Dependent Isocitrate Dehydrogenase from *Staphylococcus aureus*. Applied biochemistry and biotechnology 169: 862–869.
43. Rigby KM, DeLeo FR (2012) Neutrophils in innate host defense against *Staphylococcus aureus* infections. Seminars in immunopathology 34: 237–259.
44. Anderson K, Roux C (2010) Characterizing the effects of inorganic acid and alkaline shock on the *Staphylococcus aureus* transcriptome and messenger RNA turnover. FEMS Immunol Med Microbiol 60: 208–250.
45. Pané-Farré J, Jonas B, Förstner K, Engelmann S, Hecker M (2006) The sigmaB regulon in *Staphylococcus aureus* and its regulation. International journal of medical microbiology: IJMM 296: 237–258.
46. Somerville GA, Cockayne A, Dürr M, Peschel A, Otto M, et al. (2003) Synthesis and deformylation of *Staphylococcus aureus* delta-toxin are linked to tricarboxylic acid cycle activity. Journal of bacteriology 185: 6686–6694.
47. Wakeman CA, Hammer ND, Stauff DL, Attia AS, Anzaldi LL, et al. (2012) Menaquinone biosynthesis potentiates haem toxicity in *Staphylococcus aureus*. Molecular microbiology 86: 1376–1392.
48. Den Reijer PM, Lemmens-den Toom N, Kant S, Snijders SV, Boelens H, et al. (2013) Characterization of the humoral immune response during *Staphylococcus aureus* bacteremia and global gene expression by *Staphylococcus aureus* in human blood. PloS one 8: e53391.
49. Malachowa N, Whitney AR, Kobayashi SD, Sturdevant DE, Kennedy AD, et al. (2011) Global changes in *Staphylococcus aureus* gene expression in human blood. PloS one 6: e18617.
50. Mei JM, Nourbakhsh F, Ford CW, Holden DW (1997) Identification of *Staphylococcus aureus* virulence genes in a murine model of bacteraemia using signature-tagged mutagenesis. Molecular microbiology 26: 399–407.

Supporting figures

Figure S1. Commands and settings used in R for the statistical analysis.

```
Description of the sample
> BloodDesign <- data.frame(
 row.names = colnames( CountTable7.4 ),
 condition = c( "untreated", "untreated", "treated", "treated" ),
 libType = c( "single-end", "single-end", "single-end","single-end"
) )

Create a condition factor
> conds <- factor( c( "untreated", "untreated", "treated", "treated"
))

Examplify a CountDataSet (cds) which is the central structure in the
DESeq package

> library( DESeq )
> cds <- newCountDataSet( CountTable7.4, conds )

Acess counts
> head( counts(cds) )

Histogram to check for noise in data
> hist(log2(CountTable7.4$sample_46),100)

Romoval of noise
> cds <- newCountDataSet( CountTable7.5[CountTable7.5$sample_46>16,
], conds )

Access the count data
> head( counts(cds) )

Estimate the size factor (coverage)

> cds <- estimateSizeFactors( cds )
> sizeFactors ( cds )

Normalize count data according to size factor
> head( counts( cds, normalized=TRUE ) )

Estimate dispersion
> cds <- estimateDispersions( cds, fitType="local" )

Inspect the intermediate steps for the dispersion estimation
> str( fitInfo(cds) )
```

To visualize these steps plot the per-gene estimates against the normalized mean expressions per gene and then overlay the fitted curve

```
> plotDispEsts <- function( cds )
 {
 plot(
 rowMeans( counts( cds, normalized=TRUE ) ),
 fitInfo(cds)$perGeneDispEsts,
 pch = '.', log="xy" )
 xg <- 10^seq( -.5, 5, length.out=300 )
 lines( xg, fitInfo(cds)$dispFun( xg ), col="red" )
 }
```

Calling the function preduces the plot

```
> plotDispEsts( cds )
```

Dispersion values used by the subsequent testing are stored in the feature data slot of cds

```
> head( fData(cds) )
```

Fit to model based on the negative binomial distribution

```
> res <- nbinomTest( cds, "untreated", "treated" )
> head ( res )
```

Plot the log2 fold change against the base means, colouring in red those genes that are significant at 5% level

```
> plotDE <- function( res )
 plot(
 res$baseMean,
 res$log2FoldChange,
 log="x", pch=20, cex=.3,
 col = ifelse( res$pval < .05, "red", "black" ) )
> plotDE( res )
```

Filter for significant genes according to some chosen threshold

```
> resSig <- res[ res$pval < 0.05, ]
```

List the most significantly differentially expressed genes

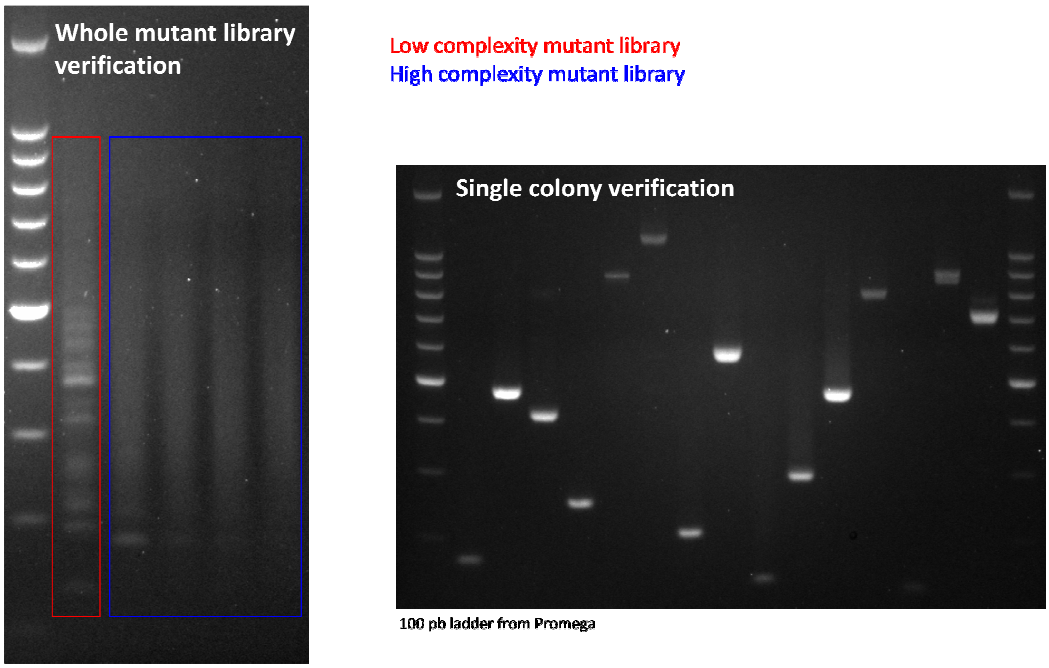
```
> head( resSig[ order(resSig$pval), ] )
```

To save the output file use the R functions write.table and write.csv

Export to excel

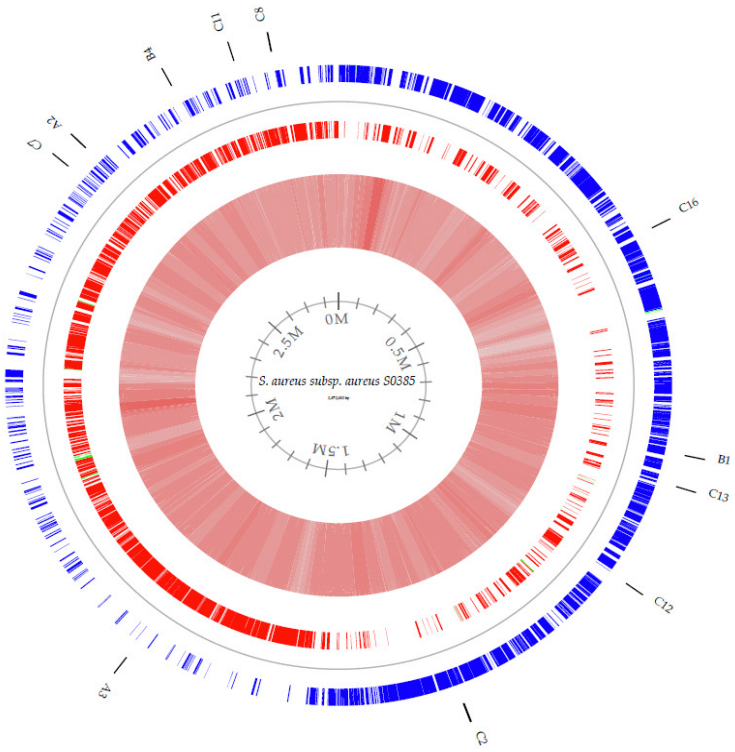
```
> write.csv( resSig, "datafile7.4_0.05.csv" )
```

Figure S2. Whole mutant library and single colony verification.



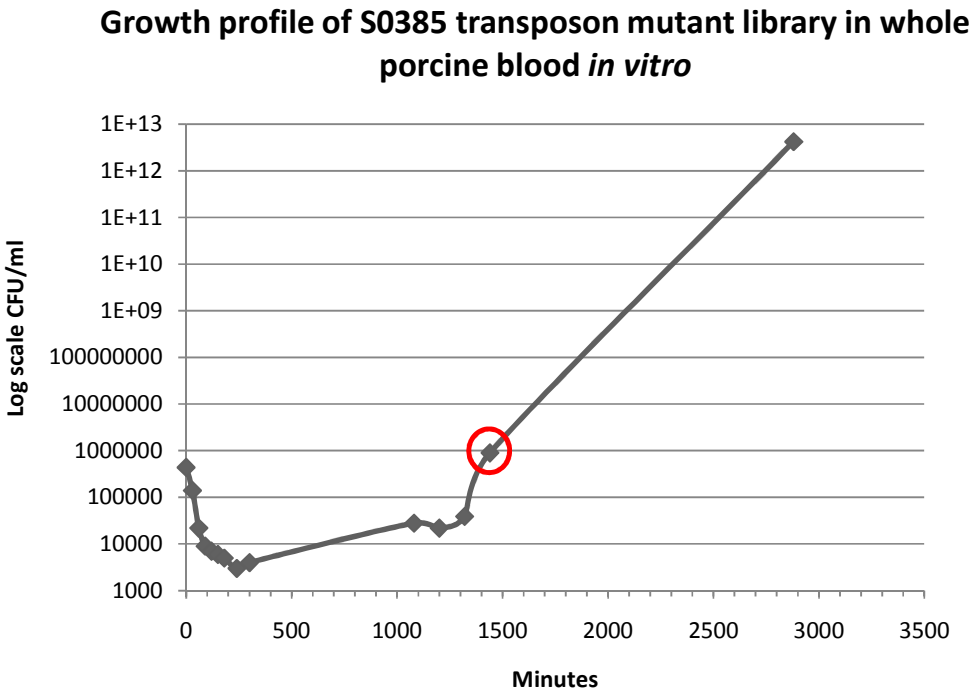
The gels show the result of the linker PCR used for library validation. The left gel shows squared in red a low complexity mutant library with a laddering of the smears. The blue squared lanes illustrate the same high complexity transposon mutant library from passage 0 (lane 2) to passage 3 (lane 5). The third generation transposon mutant library shows a smear with no specific bands. The right gel represents 15 randomly picked single mutant colonies isolated from the third generation transposon mutant library, each giving a band of different size indicating that the transposon has inserted at different locations with the genome.

Figure S3. Genome atlas identifying transposon inserts of 11 random isolated mutants.



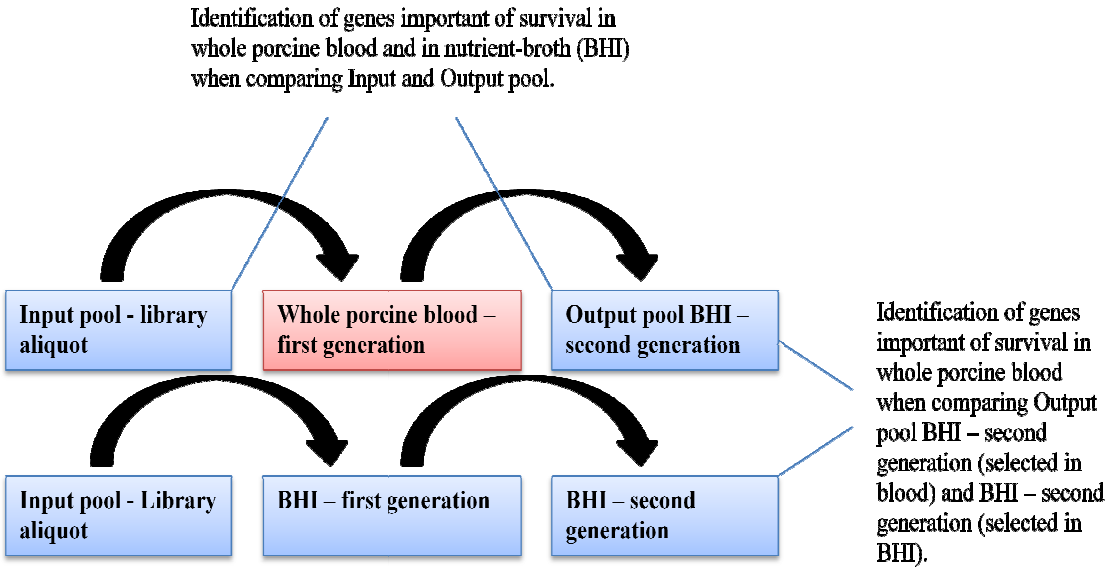
The genome atlas illustrates by black marks in the outer most circle 11 different transposon insertion sites within the reference genome. The insertion sites were identified based on sequencing 11 of the 15 randomly picked mutant colonies described in figure S1. The fragments from the 11 mutants were sequenced and aligning to the reference genome. The blue and red parts of the atlas indicate forward and reverse transcriptional direction of the open reading frames within the reference genome.

Figure S4. Growth profile of transposon mutant library in whole porcine blood *in vitro*.



The figure shows the growth profile of the transposon mutant population in whole porcine blood *in vitro*. Mutant population size was determined at specific time-points to identify functionality of the blood immune cells. After 24 hours incubation *in vitro* the mutant population size was equivalent to the inoculated population size (indicated by the red circle).

Figure S5. Experimental setup for identification of genes important for bacterial growth in whole porcine blood.



The mutant composition in input pool pre-selection in whole porcine blood (Input pool - library aliquot) were compared with mutant composition in output pool post-selection in porcine blood (output pool BHI – second generation library). The mutants identified with a significant change in number of clones represent genes important for whole porcine blood survival in addition to growth BHI. The mutant composition in output pool post-selection in porcine blood (output pool BHI – second generation library) was compared to mutant composition after growth in BHI (BHI – second generation library). The mutants identified with a significant change in number of clones in both of the comparisons were evaluated as specific for survival in whole porcine blood *in vitro*.

Table S1. Proposed essential genes.

Table S2. Proposed beneficial genes.

Table S3. Comparison of essential gene lists in *S. aureus*.

Table S1. Proposed essential genes

Gene ID	Gene Type	Insertion index	Name	Essential in previous studies	Gene description	Function	COG
ID=SAPIG0003	CDS	0	yaaA	y			s
ID=SAPIG0020	CDS	0	SAPIG0020	y			K, T
ID=SAPIG0339	CDS	0	SAPIG0339	n	repressor	Translational regulator	no related COG
ID=SAPIG0442	CDS	0	SAPIG0442	y			S
ID=SAPIG0445	CDS	0	rpsF	y			J
ID=SAPIG0447	CDS	0	rpsR	y			J
ID=SAPIG0545	CDS	0	SAPIG0545	y			S
ID=SAPIG0568	CDS	0	SAPIG0568	y			J
ID=SAPIG0580	CDS	0	SAPIG0580	y			H
ID=SAPIG0581	CDS	0	SAPIG0581	y			H
ID=SAPIG0610	CDS	0	secE	y			U
ID=SAPIG0612	CDS	0	rplK	y			J
ID=SAPIG0613	CDS	0	rplA	y			J
ID=SAPIG0615	CDS	0	rplL	y			J
ID=SAPIG0619	CDS	0	SAPIG0619	n	hypothetical protein	Unknown	J
ID=SAPIG0623	CDS	0	tuf	y			J
ID=SAPIG0641	CDS	0	SAPIG0641	y			S
ID=SAPIG0664	CDS	0	SAPIG0664	y			H
ID=SAPIG0687	CDS	0	SAPIG0687	n	domain of unknown function	Unknown	S
ID=SAPIG0697	CDS	0	SAPIG0697	y			K
ID=SAPIG0807	CDS	0	nrdI	y			S
ID=SAPIG0853	CDS	0	SAPIG0853	y			G

ID=SAPIG0860	CDS	0	smpB	y			O
ID=SAPIG0917	CDS	0	dltC	n	D-alanyl carrier protein	Cell wall	I, Q
ID=SAPIG0934	CDS	0	SAPIG0934	n	Na(+)/H(+) antiporter subunit	Transport	P
ID=SAPIG0938	CDS	0	SAPIG0938	y			P
ID=SAPIG1012	CDS	0	SAPIG1012	n	hypothetical protein	Unknown	no related COG
ID=SAPIG1026	CDS	0	SAPIG1026	n	hypothetical protein	Unknown	S
ID=SAPIG1039	CDS	0	SAPIG1039	y			H
ID=SAPIG1055	CDS	0	SAPIG1055	n	cytochrome aa3 quinol oxidase	Respiration	C
ID=SAPIG1059	CDS	0	SAPIG1059	n	hypothetical protein	Unknown	no related COG
ID=SAPIG1061	CDS	0	SAPIG1061	y			H
ID=SAPIG1080	CDS	0	SAPIG1080	y			G
ID=SAPIG1081	CDS	0	SAPIG1081	n	phosphoenolpyruvate-protein p	Carbon metabolism	G
ID=SAPIG1085	CDS	0	SAPIG1085	n	potassium uptake protein TrkA	Transport	G
ID=SAPIG1121	CDS	0	SAPIG1121	y			H
ID=SAPIG1124	CDS	0	rpmF	y			J
ID=SAPIG1179	CDS	0	SAPIG1179	y			M
ID=SAPIG1183	CDS	0	ftsZ	y			D
ID=SAPIG1209	CDS	0	SAPIG1209	y			H
ID=SAPIG1238	CDS	0	rpsP	y			J
ID=SAPIG1241	CDS	0	rplS	y			J
ID=SAPIG1254	CDS	0	xerC	n	tyrosine recombinase XerC	DNA metabolism	L
ID=SAPIG1269	CDS	0	SAPIG1269	y			K
ID=SAPIG1270	CDS	0	SAPIG1270	y			J
ID=SAPIG1274	CDS	0	SAPIG1274	y			H
ID=SAPIG1275	CDS	0	rpsO	y			J
ID=SAPIG1277	CDS	0	SAPIG1277	y			R
ID=SAPIG1285	CDS	0	SAPIG1285	y			I
ID=SAPIG1355	CDS	0	SAPIG1355	y			S
ID=SAPIG1356	CDS	0	SAPIG1356	y			L
ID=SAPIG1427	CDS	0	SAPIG1427	y			H
ID=SAPIG1444	CDS	0	SAPIG1444	n	hypothetical protein	Unknown	no related COG
ID=SAPIG1449	CDS	0	recU	y			R
ID=SAPIG1471	CDS	0	SAPIG1471	y			L

ID=SAPIG1483	CDS	0	SAPIG1483	y			C
ID=SAPIG1564	CDS	0	SAPIG1564	y			P
ID=SAPIG1576	CDS	0	SAPIG1576	y			G
ID=SAPIG1582	CDS	0	SAPIG1582	n	2-oxoisovalerate dehydrogenase subunit beta (branched-c		C
ID=SAPIG1635	CDS	0	SAPIG1635	y			R
ID=SAPIG1641	CDS	0	rpsU	y			J
ID=SAPIG1659	CDS	0	SAPIG1659	y			H
ID=SAPIG1680	CDS	0	SAPIG1680	y			L
ID=SAPIG1686	CDS	0	SAPIG1686	y			E
ID=SAPIG1696	CDS	0	SAPIG1696	y			J
ID=SAPIG1704	CDS	0	SAPIG1704	y			L
ID=SAPIG1708	CDS	0	rpmA	y			J
ID=SAPIG1710	CDS	0	rplU	y			J
ID=SAPIG1725	CDS	0	hemC	n	porphobilinogen deaminase	Heme biosynthesis	H
ID=SAPIG1738	CDS	0	SAPIG1738	y			L
ID=SAPIG1772	CDS	0	rpsD	y			J
ID=SAPIG1789	CDS	0	SAPIG1789	n	hypothetical protein	Unknown	no related COG
ID=SAPIG1833	CDS	0	SAPIG1833	n	hypothetical protein	Unknown	no related COG
ID=SAPIG1995	CDS	0	SAPIG1995	y			J
ID=SAPIG2067	CDS	0	groS	y			O
ID=SAPIG2088	CDS	0	SAPIG2088	y			O
ID=SAPIG2091	CDS	0	SAPIG2091	n	hypothetical protein	Unknown	R
ID=SAPIG2150	CDS	0	atpE	n	ATP synthase F0, C subunit	Metabolic pathways	C
ID=SAPIG2162	CDS	0	rpmE	y			J
ID=SAPIG2218	CDS	0	SAPIG2218	y			S
ID=SAPIG2270	CDS	0	rpsI	y			J
ID=SAPIG2271	CDS	0	rplM	y			no related COG
ID=SAPIG2278	CDS	0	rpsK	y			J
ID=SAPIG2279	CDS	0	rpsM	y			no related COG
ID=SAPIG2280	CDS	0	rpmJ	y			J
ID=SAPIG2282	CDS	0	SAPIG2282	y			F
ID=SAPIG2284	CDS	0	rplO	y			J
ID=SAPIG2285	CDS	0	rpmD	y			no related COG

ID=SAPIG2286	CDS	0	rpsE	y		J	
ID=SAPIG2287	CDS	0	rplR	y		no related COG	
ID=SAPIG2289	CDS	0	rpsH	y		no related COG	
ID=SAPIG2290	CDS	0	rpsN	y		J	
ID=SAPIG2294	CDS	0	rpsQ	y		J	
ID=SAPIG2295	CDS	0	rpmC	y		no related COG	
ID=SAPIG2298	CDS	0	rplV	y		J	
ID=SAPIG2299	CDS	0	rpsS	y		no related COG	
ID=SAPIG2300	CDS	0	rplB	y		J	
ID=SAPIG2554	CDS	0	SAPIG2554	y		no related COG	
ID=SAPIG2758	CDS	0	SAPIG2758	y		J	
ID=SAPIG2759	CDS	0	rpmH	y		no related COG	100
ID=SAPIG2762	rRNA	0	SAPIG2762	y			
ID=SAPIG2772	rRNA	0	SAPIG2772	y			
ID=SAPIG2774	rRNA	0	SAPIG2774	y			
ID=SAPIG2775	rRNA	0	SAPIG2775	y	4		
ID=SAPIG0018	tRNA	0	SAPIG0018	y			
ID=SAPIG0019	tRNA	0	SAPIG0019	y			
ID=SAPIG0584	tRNA	0	SAPIG0584	y			
ID=SAPIG0586	tRNA	0	SAPIG0586	y			
ID=SAPIG0587	tRNA	0	SAPIG0587	y			
ID=SAPIG0588	tRNA	0	SAPIG0588	y			
ID=SAPIG0589	tRNA	0	SAPIG0589	y			
ID=SAPIG0590	tRNA	0	SAPIG0590	y			
ID=SAPIG0591	tRNA	0	SAPIG0591	y			
ID=SAPIG0592	tRNA	0	SAPIG0592	y			
ID=SAPIG1021	tRNA	0	SAPIG1021	y			
ID=SAPIG1169	tRNA	0	SAPIG1169	y			
ID=SAPIG1889	tRNA	0	SAPIG1889	y			
ID=SAPIG1890	tRNA	0	SAPIG1890	y			
ID=SAPIG1891	tRNA	0	SAPIG1891	y			
ID=SAPIG1892	tRNA	0	SAPIG1892	y			
ID=SAPIG1894	tRNA	0	SAPIG1894	y			

ID=SAPIG1895	tRNA	0	SAPIG1895	y
ID=SAPIG1926	tRNA	0	SAPIG1926	y
ID=SAPIG1927	other	0	SAPIG1927	?
ID=SAPIG1928	tRNA	0	SAPIG1928	y
ID=SAPIG1929	tRNA	0	SAPIG1929	y
ID=SAPIG1931	tRNA	0	SAPIG1931	y
ID=SAPIG1932	tRNA	0	SAPIG1932	y
ID=SAPIG1933	tRNA	0	SAPIG1933	y
ID=SAPIG1934	tRNA	0	SAPIG1934	y
ID=SAPIG1935	tRNA	0	SAPIG1935	y
ID=SAPIG1936	tRNA	0	SAPIG1936	y
ID=SAPIG1937	tRNA	0	SAPIG1937	y
ID=SAPIG1938	tRNA	0	SAPIG1938	y
ID=SAPIG1940	tRNA	0	SAPIG1940	y
ID=SAPIG1941	tRNA	0	SAPIG1941	y
ID=SAPIG1942	tRNA	0	SAPIG1942	y
ID=SAPIG1943	tRNA	0	SAPIG1943	y
ID=SAPIG1944	tRNA	0	SAPIG1944	y
ID=SAPIG1945	tRNA	0	SAPIG1945	y
ID=SAPIG1946	tRNA	0	SAPIG1946	y
ID=SAPIG1947	tRNA	0	SAPIG1947	y
ID=SAPIG1948	tRNA	0	SAPIG1948	y
ID=SAPIG1949	tRNA	0	SAPIG1949	y
ID=SAPIG1952	tRNA	0	SAPIG1952	y
ID=SAPIG1953	tRNA	0	SAPIG1953	y
ID=SAPIG2102	tRNA	0	SAPIG2102	y
ID=SAPIG2220	tRNA	0	SAPIG2220	y
ID=SAPIG2221	tRNA	0	SAPIG2221	y
ID=SAPIG2222	tRNA	0	SAPIG2222	y
ID=SAPIG2224	tRNA	0	SAPIG2224	y
ID=SAPIG2225	tRNA	0	SAPIG2225	y

y=yes	Essential in previous studies	
n=no	Essential in previous studies	hypothetical proteins

Table S2. Proposed beneficial genes

Gene ID	Insertion index	Name	Essential in previous studies	COG	Gene description (protein coding)	Function
ID=SAPIG1210	0,00042	SAPIG1210	y	L		**cut-off 0.007**
ID=SAPIG0622	0,00048	fusA	y	J		
ID=SAPIG1226	0,00049	recG	n	K, L	ATP-dependent DNA helicase RecG	DNA repair
ID=SAPIG1288	0,00064		y	R		
ID=SAPIG0468	0,00065	SAPIG0468	y	F		
ID=SAPIG1571	0,00067	zwf	y	G		
ID=SAPIG1994	0,00069	SAPIG1994	y	J		
ID=SAPIG1993	0,00070		y	J		
ID=SAPIG2145	0,00071	atpD	n	C	ATP synthase F1, beta subunit	Metabolic pathways
ID=SAPIG1182	0,00071	ftsA	y	D		
ID=SAPIG1739	0,00071		y	J		
ID=SAPIG1727	0,00074	hemA	n	H	glutamyl-tRNA reductase	Metabolic pathways
ID=SAPIG0547	0,00075	SAPIG0547	n	E	Orn/Lys/Arg decarboxylase	Amino acid transport and metabolism
ID=SAPIG0945	0,00075	pgi	y	G		
ID=SAPIG2163	0,00076	rho	n	K	transcription termination factor Rho	Transcription
ID=SAPIG1718	0,00076	valS	y	J		
ID=SAPIG0575	0,00077		y	D		
ID=SAPIG1707	0,00077		n	R	Spo0B-associated GTP-binding protei	unknown
ID=SAPIG2595	0,00078	SAPIG2595	y	no related COG		
ID=SAPIG1581	0,00078		n	C	lipoamide acyltransferase component	Metabolic pathways
ID=SAPIG2314	0,00079	SAPIG2314	y	V		
ID=SAPIG1694	0,00079	hisS	y	J		
ID=SAPIG1235	0,00080	SAPIG1235	y	U		

ID=SAPIG1811	0,00083	leuS	y	J		
ID=SAPIG1457	0,00083	SAPIG1457	y	J		
ID=SAPIG1764	0,00083	ackA	n	C	acetate kinase	Metabolic pathways
ID=SAPIG1013	0,00085		n	M	UDP-glucose diacylglycerol glucosyltr	glycerolipid metabolism
ID=SAPIG0267	0,00085	SAPIG0267	n	M	cdp-glycerol:poly(glycerophosphate)	Cell envelope biogenesis
ID=SAPIG2596	0,00086	SAPIG2596	y	I		
ID=SAPIG0949	0,00086	addB	n	L	ATP-dependent nuclease subunit B	Nucleic acid metabolism
ID=SAPIG1662	0,00091		y	R		
ID=SAPIG1419	0,00093	murG	y	M		
ID=SAPIG0823	0,00095	SAPIG0823	n	M	probable undecaprenyl-phosphate N-	
ID=SAPIG0577	0,00096	SAPIG0577	n	O	putative Cell division protease FtsH h	Cell division
ID=SAPIG0851	0,00099	gap	y	G		
ID=SAPIG1998	0,00100	ligA	y	L		
ID=SAPIG1057	0,00101	qoxB	n	C	cytochrome aa3 quinol oxidase, subu	Metabolic pathways
ID=SAPIG0662	0,00101	pta	y	C		
ID=SAPIG1228	0,00101	plsX	y	I		
ID=SAPIG1752	0,00103	pfkA	y	G		
ID=SAPIG1468	0,00104		y	H		
ID=SAPIG0462	0,00104	SAPIG0462	n	no related COG	hypothetical protein	Unknown
ID=SAPIG1428	0,00104	thyA	y	F		
ID=SAPIG1753	0,00106	accA	y	I		
ID=SAPIG0840	0,00107	trxB	y	O		
ID=SAPIG0814	0,00108	murB	y	M		
ID=SAPIG1646	0,00109	dnaK	y	O		
ID=SAPIG2209	0,00111	glmS	y	M		
ID=SAPIG1627	0,00111	SAPIG1627	y	L		
ID=SAPIG1632	0,00111	era	y	R		
ID=SAPIG1220	0,00114	rsgA	y	R		
ID=SAPIG2167	0,00116	fba	y	G		
ID=SAPIG0544	0,00118		y	L		
ID=SAPIG2754	0,00119	SAPIG2754	y	K		
ID=SAPIG2159	0,00120	hemK	y	J		
ID=SAPIG2172	0,00124	coaA	y	H		

ID=SAPIG1146	0,00125	murI	y	M		
ID=SAPIG1756	0,00125	SAPIG1756	y	L		
ID=SAPIG1290	0,00125	SAPIG1290	n	S	hypothetical protein	Unknown
ID=SAPIG2147	0,00133	atpA	y	C		
ID=SAPIG1736	0,00134	SAPIG1736	y	E		
ID=SAPIG1014	0,00135	SAPIG1014	y	M		
ID=SAPIG1240	0,00136	trmD	y	J		
ID=SAPIG0915	0,00137	dltA	y	E		
ID=SAPIG1558	0,00138	SAPIG1558	n	K, T	transcriptional regulatory protein Res	Transcription regulator
ID=SAPIG0618	0,00138	rpoC	y	K		
ID=SAPIG1260	0,00138	pyrH	y	F		
ID=SAPIG1234	0,00140	smc	y	D		
ID=SAPIG0808	0,00142	SAPIG0808	y	F		
ID=SAPIG1460	0,00143	SAPIG1460	y	R		
ID=SAPIG1664	0,00146	mtnN	y	R		
ID=SAPIG1592	0,00147	accC	y	I		
ID=SAPIG1311	0,00149	glnA	y	E		
ID=SAPIG1476	0,00152	cmk	y	F		
ID=SAPIG1682	0,00152	alaS	y	J		
ID=SAPIG1985	0,00152	SAPIG1985	y	M		
ID=SAPIG0014	0,00152	SAPIG0014	y	T		
ID=SAPIG1473	0,00153	engA	y	R		
ID=SAPIG2297	0,00153	rpsC	y	no related COG		
ID=SAPIG2283	0,00155	SAPIG2283	y	no related COG		
ID=SAPIG1737	0,00155	thrS	y	J		
ID=SAPIG1221	0,00155	rpe	y	G		
ID=SAPIG1722	0,00155	hemL	n	H	glutamate-1-semialdehyde-2,1-amin	Metabolic pathways
ID=SAPIG0604	0,00156	cysE	y	E		
ID=SAPIG1222	0,00156	SAPIG1222	y	H		
ID=SAPIG1717	0,00157	SAPIG1717	y	H		
ID=SAPIG1647	0,00159	grpE	y	O		
ID=SAPIG0980	0,00161	fabF	y	I, Q		
ID=SAPIG0899	0,00161	SAPIG0899	y	E		

ID=SAPIG1732	0,00164	SAPIG1732	y	R		
ID=SAPIG1056	0,00165	quxC	n	C	cytochrome aa3 quinol oxidase, subu	Metabolic pathways
ID=SAPIG1838	0,00168	metK	y	H		
ID=SAPIG1278	0,00169	SAPIG1278	n	D	dna translocase ftsk (dna translocase	Cell division
ID=SAPIG0617	0,00169	rpoB	y	K		
ID=SAPIG1728	0,00169		y	R		
ID=SAPIG1645	0,00175	dnaJ	y	O		
ID=SAPIG1685	0,00179	trmU	y	J		
ID=SAPIG1178	0,00179	SAPIG1178	y	M		
ID=SAPIG1123	0,00179	SAPIG1123	y	R		
ID=SAPIG1594	0,00179	efp	y	J		
ID=SAPIG1187	0,00180	SAPIG1187	y	S		
ID=SAPIG1261	0,00180	frr	y	J		
ID=SAPIG1807	0,00181	SAPIG1807	y	R		
ID=SAPIG1626	0,00181	SAPIG1626	y	K		
ID=SAPIG1088	0,00181	def	y	J		
ID=SAPIG1058	0,00182	SAPIG1058	n	C	quinol oxidase subunit 2 (Quinol oxid	Metabolic pathways
ID=SAPIG2547	0,00183	SAPIG2547	n	no related COG	phosphomannomutase	Carbohydrate metabolism
ID=SAPIG1560	0,00184	scpB	y	K		
ID=SAPIG2291	0,00185		y	no related COG		
ID=SAPIG0666	0,00186	SAPIG0666	y	I		
ID=SAPIG2160	0,00186	prfA	y	J		
ID=SAPIG2169	0,00186	pyrG	y	F		
ID=SAPIG2288	0,00186	SAPIG2288	y	J		
ID=SAPIG1271	0,00189	infB	y	J		
ID=SAPIG2398	0,00190	fni	y	C		
ID=SAPIG1276	0,00191	SAPIG1276	n	J	polyribonucleotide nucleotidyltransfe	RNA degradation
ID=SAPIG0950	0,00192	addA	n	L	recombination helicase AddA	Recombination
ID=SAPIG2149	0,00192	atpF	n	C	ATP synthase F0, B subunit	Metabolic pathways
ID=SAPIG1361	0,00198	SAPIG1361	y	S		
ID=SAPIG0446	0,00198		n	L	single-stranded DNA-binding protein	DNA replication, recombination, repair
ID=SAPIG0614	0,00200	SAPIG0614	y	J		
ID=SAPIG1583	0,00201		n	C	2-oxoisovalerate dehydrogenase sub	Metabolic pathways

ID=SAPIG0583	0,00202	lysS	y	J		
ID=SAPIG1788	0,00202	ccpA	n	K	catabolite control protein A	Transcription regulator
ID=SAPIG1091	0,00204		y	C		
ID=SAPIG1652	0,00205	holA	y	L		
ID=SAPIG1723	0,00205	hemB	n	H	delta-aminolevulinic acid dehydratase	Metabolic pathways
ID=SAPIG0809	0,00206	SAPIG0809	y	F		
ID=SAPIG0005	0,00207	gyrB	y	L		
ID=SAPIG0566	0,00207	prsA	y	E, F		
ID=SAPIG1266	0,00208	polC	n	L	DNA polymerase	DNA replication
ID=SAPIG1136	0,00209	rnhC	n	L	ribonuclease hiii	DNA replication
ID=SAPIG2277	0,00212	rpoA	n	no related COG	DNA-directed RNA polymerase, alpha	RNA polymerase
ID=SAPIG0979	0,00212		y	I		
ID=SAPIG2756	0,00213	gidA	y	D		
ID=SAPIG1215	0,00214	fmt	y	J		
ID=SAPIG1593	0,00215	accB	y	I		
ID=SAPIG1630	0,00216	glyS	y	J		
ID=SAPIG1229	0,00216	fabD	y	I		
ID=SAPIG1570	0,00217	rnz	y	R		
ID=SAPIG1180	0,00222	murD	y	M		
ID=SAPIG0015	0,00224	rplI	y	J		
ID=SAPIG1291	0,00227	SAPIG1291	n	C	pyruvate ferredoxin oxidoreductase a	Metabolic pathways
ID=SAPIG1116	0,00230	SAPIG1116	n	S	regulatory protein YlbF	Replication, recombination and repair
ID=SAPIG2296	0,00230	rplP	y	J		
ID=SAPIG1092	0,00232		n	C	dihydrolipoyllysine-residue acetyltran	carbon metabolism
ID=SAPIG1011	0,00232	SAPIG1011	n	G, E, P, R	MFS-type transporter	Transport
ID=SAPIG1957	0,00233	hemL	n	H	glutamate-1-semialdehyde-2,1-amin	Metabolic pathways
ID=SAPIG0009	0,00233	serS	y	J		
ID=SAPIG1754	0,00233	accD	y	I		
ID=SAPIG2142	0,00237	murA	y	M		
ID=SAPIG1375	0,00238	SAPIG1375	y	V		
ID=SAPIG0663	0,00239	SAPIG0663	y	H		
ID=SAPIG1617	0,00241	SAPIG1617	n	M	penicillin-binding Protein dimerisation	Peptidoglycan biosynthesis
ID=SAPIG2007	0,00243	nadE	y	H		

ID=SAPIG1110	0,00244	SAPIG1110	y	D		
ID=SAPIG2066	0,00247	groL	y	O		
ID=SAPIG1357	0,00250	parC	y	L		
ID=SAPIG1997	0,00250	SAPIG1997	n	R	sex pheromone staph-cAM373	
ID=SAPIG0716	0,00252	SAPIG0716	y	M, G		
ID=SAPIG1802	0,00253	SAPIG1802	y	M		
ID=SAPIG0470	0,00254	SAPIG0470	n	no related COG	bovine pathogenicity island protein O	Transcription regulator
ID=SAPIG0573	0,00254	SAPIG0573	y	D		
ID=SAPIG1284	0,00254	SAPIG1284	n	S	hypothetical protein	Unknown
ID=SAPIG0918	0,00255	SAPIG0918	y	M		
ID=SAPIG1268	0,00255	nusA	y	K		
ID=SAPIG1263	0,00255	cdsA	y	I		
ID=SAPIG1262	0,00259	uppS	y	I		
ID=SAPIG0897	0,00262	sufC	y	O		
ID=SAPIG1981	0,00264	map	y	J		
ID=SAPIG0927	0,00267	SAPIG0927	y	Q		
ID=SAPIG0932	0,00267	SAPIG0932	y	C, P		
ID=SAPIG0004	0,00270	SAPIG0004	n	L	DNA replication and repair protein Re	DNA repair
ID=SAPIG1310	0,00271	SAPIG1310	y	K		
ID=SAPIG2293	0,00271	rplN	y	no related COG		
ID=SAPIG2008	0,00272	pncB	y	H		
ID=SAPIG0467	0,00273	guaB	y	F		
ID=SAPIG1984	0,00273		y	R		
ID=SAPIG2463	0,00273		n	no related COG	L-cystine import ATP-binding protein	transport
ID=SAPIG1999	0,00274	pcrA	y	L		
ID=SAPIG1697	0,00274	SAPIG1697	y	K, T		
ID=SAPIG2151	0,00274	atpB	n	C	ATP synthase F0, A subunit	Metabolic pathways
ID=SAPIG2112	0,00278	acpS	y	J, K, L		
ID=SAPIG0268	0,00279		y	I		
ID=SAPIG1693	0,00283	aspS	y	J		
ID=SAPIG1134	0,00283	pheS	y	J		
ID=SAPIG1282	0,00284		n	I, Q, R	3-ketoacyl-	fatty acid biosynthesis
ID=SAPIG0016	0,00286	dnaB	y	L		

ID=SAPIG1897	0,00286	hemG	n	H	protoporphyrinogen oxidase	Metabolic pathways
ID=SAPIG1899	0,00289	hemE	n	H	uroporphyrinogen decarboxylase	Metabolic pathways
ID=SAPIG1445	0,00290		y	D		
ID=SAPIG1191	0,00290	ileS	y	J		
ID=SAPIG1453	0,00291	SAPIG1453	y	L		
ID=SAPIG2388	0,00291	rpiA	y	G		
ID=SAPIG1237	0,00292	ffh	y	U		
ID=SAPIG0933	0,00292	SAPIG0933	y	P		
ID=SAPIG1265	0,00293	proS	y	J		
ID=SAPIG1770	0,00295	ezrA	y	D		
ID=SAPIG2216	0,00295	glmM	y	G		
ID=SAPIG0565	0,00296	glmU	y	M		
ID=SAPIG2090	0,00302	SAPIG2090	y	o		
ID=SAPIG2303	0,00302	rplC	y	no related COG		
ID=SAPIG0830	0,00302	prfB	y	J		
ID=SAPIG0992	0,00303	trpS	y	J		
ID=SAPIG1181	0,00303	SAPIG1181	n	M	cell division protein	Cell division
ID=SAPIG1793	0,00304	murC	y	M		
ID=SAPIG0665	0,00305	mvaD	y	I		
ID=SAPIG0855	0,00307	eno	y	G		
ID=SAPIG1456	0,00309	SAPIG1456	y	H		
ID=SAPIG0796	0,00309	SAPIG0796	y	M		
ID=SAPIG0713	0,00310	SAPIG0713	n	K	iron dependent repressor	Transcription regulator
ID=SAPIG1709	0,00312		y	J		
ID=SAPIG1376	0,00317	SAPIG1376	y	V		
ID=SAPIG1141	0,00317	trx	y	O, C		
ID=SAPIG2302	0,00321	rplD	y	J		
ID=SAPIG1207	0,00321	gmk	y	F		
ID=SAPIG2304	0,00324	rpsJ	n	J	ribosomal protein S10	
ID=SAPIG0550	0,00324		y	L		
ID=SAPIG0600	0,00326	SAPIG0600	n	O	egative regulator of genetic compete	Posttranslational modification
ID=SAPIG0854	0,00329	gpmI	y	G		
ID=SAPIG0564	0,00330	SAPIG0564	n	M	SpoVG superfamily	Regulatory protein

ID=SAPIG0852	0,00336	pgk	y	G		
ID=SAPIG1563	0,00338	xerD	n	L	tyrosine recombinase XerD	DNA replication, recombination, and repair
ID=SAPIG1243	0,00339	SAPIG1243	y	R		
ID=SAPIG0846	0,00340	clpP	y	U, O		
ID=SAPIG0930	0,00340		y	P		
ID=SAPIG2133	0,00344	SAPIG2133	y	U		
ID=SAPIG1292	0,00346	SAPIG1292	n	C	ferrodoxin oxidoreductase beta subu	TCA cycle
ID=SAPIG1086	0,00353	SAPIG1086	y	R		
ID=SAPIG0829	0,00355	secA	y	U		
ID=SAPIG1735	0,00357	infC	y	J		
ID=SAPIG0688	0,00361	argS	y	J		
ID=SAPIG1816	0,00362	SAPIG1816	n	I	hydrolase, alpha/beta fold family, pu	Lipid metabolism
ID=SAPIG2301	0,00362	rplW	y	no related COG		
ID=SAPIG0611	0,00364	nusG	n	K	transcription termination/antitermina	Transcription regulator
ID=SAPIG0001	0,00367	dnaA	y	L		
ID=SAPIG1960	0,00368		n	J	protein LMOf2365_1711	unknown
ID=SAPIG0717	0,00369	SAPIG0717	y	M, G		
ID=SAPIG0576	0,00370	hpt	y	F		
ID=SAPIG2148	0,00370	atpH	n	C	ATP synthase F1, delta subunit	m
ID=SAPIG2170	0,00377	rpoE	n	K	DNA-directed RNA polymerase, delta	RNA polymerase
ID=SAPIG1663	0,00379	yqeG	y	R		
ID=SAPIG2158	0,00382	SAPIG2158	y	S		
ID=SAPIG0898	0,00382	sufD	y	O		
ID=SAPIG1287	0,00383	recA	n	L	protein RecA	DNA repair
ID=SAPIG1730	0,00384	tig	n	O	trigger factor	
ID=SAPIG1763	0,00385	SAPIG1763	n	T	universal stress protein family	Cellular stress
ID=SAPIG1252	0,00385	topA	y	L		
ID=SAPIG1239	0,00397	rimM	y	J		
ID=SAPIG1344	0,00402	tkt	y	G		
ID=SAPIG1363	0,00407	SAPIG1363	y	K		
ID=SAPIG1230	0,00408	fabG	y	I, Q, R		
ID=SAPIG1552	0,00410	SAPIG1552	n	no related COG	phage repressor	Transcription regulator, DNA reapir
ID=SAPIG1345	0,00412		n	S	hypothetical protien	unknown

ID=SAPIG2125	0,00416	SAPIG2125	y	D		
ID=SAPIG2464	0,00417	SAPIG2464	n	E, H	L-cystine transport system permease	Transport
ID=SAPIG2546	0,00419	SAPIG2546	n	no related COG	modification methylase EcoRII (Cytos	Transcription regulator
ID=SAPIG1584	0,00422	lpdA	y	C		
ID=SAPIG0844	0,00423	SAPIG0844	n	S	hypothetical protein	Unknown
ID=SAPIG1757	0,00425	SAPIG1757	n	R	DHH family protein	
ID=SAPIG1908	0,00425	SAPIG1908	n	R	3'-5' exoribonuclease YhaM	
ID=SAPIG1267	0,00427	SAPIG1267	n	S	protein LMOf2365_1338	Unknown
ID=SAPIG1232	0,00427	acpP	y	I, Q		
ID=SAPIG1898	0,00433	hemH	n	H	ferrochelatase	Metabolic pathways
ID=SAPIG1588	0,00433	xseB	y	L		
ID=SAPIG1112	0,00439	SAPIG1112	n	O	cytochrome aa3-controlling protein	Posttranslational modification
ID=SAPIG2123	0,00442	SAPIG2123	y	M		
ID=SAPIG1625	0,00442	SAPIG1625	y	R		
ID=SAPIG2153	0,00443	SAPIG2153	y	M		
ID=SAPIG0559	0,00447	ksgA	n	J	dimethyladenosine transferase	
ID=SAPIG1090	0,00449	pdhA	y	no related COG		
ID=SAPIG1087	0,00457	SAPIG1087	y	S		
ID=SAPIG2281	0,00457	infA	y	no related COG		
ID=SAPIG0567	0,00459	SAPIG0567	n	J	ribosomal protein L25, Ctc-form	
ID=SAPIG2572	0,00460	SAPIG2572	n	K	transcriptional regulator MarR family	Antibiotic resistance
ID=SAPIG0822	0,00467	SAPIG0822	y	T		
ID=SAPIG2084	0,00472	SAPIG2084	n	R	redox-sensing transcriptional repress	Transcription regulator
ID=SAPIG1712	0,00474	mreC	y	M		
ID=SAPIG1352	0,00480	acnA	n	C	aconitate hydratase 1	TCA cycle
ID=SAPIG1742	0,00481	coaE	y	H		
ID=SAPIG1480	0,00481	SAPIG1480	y	no related COG		
ID=SAPIG0620	0,00483	rpsL	y	J		
ID=SAPIG0548	0,00485	tmk	y	F		
ID=SAPIG0916	0,00494	SAPIG0916	y	M		
ID=SAPIG1002	0,00494	SAPIG1002	y	G		
ID=SAPIG1705	0,00498	ruvA	y	L		
ID=SAPIG2750	0,00498	SAPIG2750	n	K	hypothetical protein	Unknown

ID=SAPIG1135	0,00499	pheT	y	J		
ID=SAPIG0546	0,00503	recR	n	L	recombination protein RecR	DNA repair
ID=SAPIG0161	0,00510	SAPIG0161	n	E	putative DNA-binding protein	Transcription regulator
ID=SAPIG2465	0,00513	SAPIG2465	n	no related COG	L-cystine-binding protein TcyA	Tranport
ID=SAPIG1245	0,00514	SAPIG1245	y	C		
ID=SAPIG1257	0,00517	codY	n	K	GTP-sensing transcriptional pleiotrop	Transcriptional repressor
ID=SAPIG1907	0,00519	SAPIG1907	y	O		
ID=SAPIG1258	0,00521	rpsB	y	J		
ID=SAPIG2122	0,00526	SAPIG2122	n	J, K, L	cold-shock deAd box protein a (atp-d	DNA replication, recombination, and repair
ID=SAPIG0002	0,00529	dnaN	y	L		
ID=SAPIG2276	0,00542	rpæQ	y	J		
ID=SAPIG1194	0,00545	rluD	n	J	ribosomal large subunit pseudouridin	Translation (found in stress proteins)
ID=SAPIG0021	0,00547	SAPIG0021	y	T		
ID=SAPIG1113	0,00548	cyoE	n	O	protoheme IX farnesyltransferase	Metabolic pathways
ID=SAPIG1781	0,00554	tyrS	y	J		
ID=SAPIG0629	0,00557	ilvE	n	E, H	branched-chain amino acid aminotra	Amino acid biosynthesis
ID=SAPIG0556	0,00557	metG	y	J		
ID=SAPIG1733	0,00560	rplT	y	J		
ID=SAPIG1711	0,00565	mreD	n	M	rod shape-determining protein MreD	Cell envelope biogenesis
ID=SAPIG1259	0,00567	tsf	y	J		
ID=SAPIG1695	0,00571	lytH	y	M		
ID=SAPIG0605	0,00571	cysS	y	J		
ID=SAPIG0718	0,00571	SAPIG0718	y	no related COG		
ID=SAPIG0901	0,00572	sufB	y	O		
ID=SAPIG2146	0,00577	atpG	n	C	ATP synthase F1, gamma subunit	Metabolic pathways
ID=SAPIG2016	0,00583	SAPIG2016	n	G	YkgB	Metabolic pathways
ID=SAPIG2108	0,00599	SAPIG2108	n	K, T	phosphoserine phosphatase RsbU (Si	Signal transduction mechanisms
ID=SAPIG0006	0,00599	gyrA	y	L		
ID=SAPIG2078	0,00606	SAPIG2078	n	G	sucrose-6-phosphate hydrolase (Suc	Carbohydrate metabolism
ID=SAPIG0865	0,00606		n	no related COG	hypothetical protien	unknown
ID=SAPIG1043	0,00608	menB	y	H		
ID=SAPIG1214	0,00613	def	y	J		
ID=SAPIG0603	0,00619	gltX	y	J		

ID=SAPIG0931	0,00625	SAPIG0931	n	P	Na(+)/H(+) antiporter subunit E (Mu	Na/H transport	
ID=SAPIG1991	0,00633	SAPIG1991	y	I, R			
ID=SAPIG0621	0,00637	rpsG	y	J			
ID=SAPIG1093	0,00640	lpdA	n	C	dihydrolipoyl dehydrogenase	Metabolic pathways	
ID=SAPIG0836	0,00643	hprK	y	T			
ID=SAPIG1913	0,00645	SAPIG1913	y	K			
ID=SAPIG2014	0,00645	SAPIG2014	y	C			
ID=SAPIG1309	0,00646	SAPIG1309	n	P	aluminium resistance protein	Inorganic ion transport and metabolism	
ID=SAPIG1744	0,00646	SAPIG1744	n	L	DNA polymerase I (POL I)	DNA repair, recombination and replication	
ID=SAPIG1779	0,00647	SAPIG1779	y	I			
ID=SAPIG2105	0,00649	sigB	n	K	RNA polymerase sigma-B factor	Alternative sigma factor (stress response)	
ID=SAPIG1006	0,00649	SAPIG1006	y	I			
ID=SAPIG0597	0,00649	SAPIG0597	n	K	transcriptional regulator CtsR	Transcriptional repressor, stress genes	
ID=SAPIG2124	0,00654	SAPIG2124	y	M			
ID=SAPIG0914	0,00654	SAPIG0914	y	no related COG			
ID=SAPIG0661	0,00664	SAPIG0661	n	S	chlorite dismutase		
ID=SAPIG0434	0,00677	SAPIG0434	y	I			
ID=SAPIG1596	0,00687	SAPIG1596	y	no related COG			
ID=SAPIG1660	0,00687	SAPIG1660	y	J			
ID=SAPIG1397	0,00692	dapB	y	E			
ID=SAPIG2549	0,00692	galU	n	no related COG	UTP-glucose-1-phosphate uridylyltra	Cell envelope biogenesis	
ID=SAPIG0343	0,00694		n	no related COG	hypothetical protien	unknown	
ID=SAPIG1548	0,00694		n	no related COG	hypothetical protien	unknown	
ID=SAPIG2488	0,00702	SAPIG2488	n	S	hypothetical protein	Unknown	**Cut-off 0.02**
ID=SAPIG1700	0,00702	SAPIG1700	y	U			
ID=SAPIG2046	0,00702		n	no related COG	hypothetical protien	unknown	
ID=SAPIG2144	0,00741	atpC	n	C	ATP synthase F1, epsilon subunit	Metabolic pathways	
ID=SAPIG1724	0,00747	hemD	n	H	uroporphyrinogen-III synthase	Metabolic pathways	
ID=SAPIG0993	0,00758	SAPIG0993	y	P			
ID=SAPIG1496	0,00758		n	no related COG	hypothetical protien	unknown	
ID=SAPIG1545	0,00758		n	no related COG	phi PVL ORF 37 analogue	unknown	
ID=SAPIG1590	0,00769	nusB	n	K	transcription antitermination factor N	Transcription	
ID=SAPIG1246	0,00770	sucD	n	C	succinate-CoA ligase, alpha subunit s	Metabolic pathways	

ID=SAPIG0269	0,00780	SAPIG0269	y	E, R		
ID=SAPIG0715	0,00784	SAPIG0715	y	M		
ID=SAPIG0526	0,00787	SAPIG0526	n	E	cystathionine gamma-lyase (Gamma	
ID=SAPIG0850	0,00789	SAPIG0850	n	K	glycolytic operon regulator	Transcription
ID=SAPIG0935	0,00790	SAPIG0935	y	C, P		
ID=SAPIG1185	0,00794	SAPIG1185	n	no related COG	hypothetical protein	Unknown
ID=SAPIG0843	0,00803	SAPIG0843	n	S	protein YvcK	
ID=SAPIG1233	0,00820	rnc	y	K		
ID=SAPIG0872	0,00833	SAPIG0872	n	no related COG	hypothetical protein	Unknown
ID=SAPIG1054	0,00838	SAPIG1054	n	V	beta-lactamase	Defense mechanisms
ID=SAPIG0929	0,00840	SAPIG0929	n	P	MrpG	Transport
ID=SAPIG0948	0,00868	lepB	y	U		
ID=SAPIG1911	0,00870	SAPIG1911	n	S	protein BLi01058/	
ID=SAPIG1623	0,00891	SAPIG1623	n	J, K, L	ATP-dependent RNA helicase	DNA and RNA unwinding
ID=SAPIG1572	0,00923	SAPIG1572	n	K	transcriptional regulator, AraC family	
ID=SAPIG1247	0,00939	SAPIG1247	n	no related COG	LytN protein	
ID=SAPIG1765	0,00949	SAPIG1765	n	L	hypothetical protein	Unknown
ID=SAPIG1341	0,00962	lexA	y	K, T		
ID=SAPIG2254	0,00969	lacB	n	G	galactose-6-phosphate isomerase, La	Metabolic pathways
ID=SAPIG1679	0,00971	SAPIG1679	n	S	hypothetical protein	Defense mechanisms
ID=SAPIG1273	0,00980	truB	n	J	tRNA pseudouridine synthase B	
ID=SAPIG0429	0,00980	SAPIG0429	n	K	hypothetical protein	Recombination regulator
ID=SAPIG1734	0,00995	rpmI	y	J		
ID=SAPIG2272	0,00995	truA	n	J	tRNA pseudouridine synthase A	
ID=SAPIG1199	0,00999	carA	n	E, F	carbamoyl-phosphate synthase, sma	
ID=SAPIG1370	0,01001	trpD	n	E	anthranilate phosphoribosyltransfera	Metabolic pathways
ID=SAPIG2002	0,01003	purB	n	F	adenylosuccinate lyase	Nucleotide transport and metabolism
ID=SAPIG0731	0,01004	SAPIG0731	n	no related COG	hypothetical protein	Unknown
ID=SAPIG1198	0,01017	SAPIG1198	n	F	dihydroorotase (DHOase)	Nucleotide metabolism
ID=SAPIG1466	0,01028	aroC	n	E	chorismate synthase	Metabolic pathways
ID=SAPIG1534	0,01042		n	no related COG	hypothetical protien	unknown
ID=SAPIG0887	0,01050	gcvH	n	E	glycine cleavage system H protein	
ID=SAPIG0458	0,01053	ahpC	n	O	peroxiredoxin	

ID=SAPIG0270	0,01066	SAPIG0270	y	M		
ID=SAPIG1114	0,01082	SAPIG1114	n	S	hypothetical protein	Cell envelope biogenesis
ID=SAPIG1281	0,01088	SAPIG1281	n	R	peptidase, M16 family	Transcriptional regulator
ID=SAPIG1966	0,01099	SAPIG1966	n	R	regulatory protein RecX	Unknown
ID=SAPIG2275	0,01111	SAPIG2275	n	no related COG	cobalt import ATP-binding protein Cb	Transport
ID=SAPIG1550	0,01126	SAPIG1550	n	no related COG	77ORF023	Cell envelope biogenesis
ID=SAPIG1587	0,01134	SAPIG1587	y	H		
ID=SAPIG2141	0,01134	fabZ	y	I		
ID=SAPIG2598	0,01138	SAPIG2598	n	O	ATP-dependent Clp protease ATP-bin	Metabolic pathways
ID=SAPIG0341	0,01158	SAPIG0341	n	no related COG	phage regulatory protein, Rha family	purine metabolism
ID=SAPIG1390	0,01165	SAPIG1390	n	P	phosphate-binding protein PstS (PBP	Unknown
ID=SAPIG1796	0,01166	SAPIG1796	n	S	protein BCE33L4443	Phage associated
ID=SAPIG1176	0,01175	mraW	n	M	S-adenosyl-methyltransferase MraW	Unknown
ID=SAPIG1748	0,01182	icd	n	C	isocitrate dehydrogenase, NADP-dep	Unknown
ID=SAPIG1186	0,01183	SAPIG1186	y	R		
ID=SAPIG0153	0,01187	deoB	n	G	phosphopentomutase	pirimidine metabolism
ID=SAPIG1651	0,01190	rpsT	y	J		
ID=SAPIG1658	0,01197	SAPIG1658	n	H	hypothetical protein	
ID=SAPIG1631	0,01205	recO	n	L	DNA repair protein RecO	Adhesion
ID=SAPIG1561	0,01230	SAPIG1561	n	S	ScpA/B protein	Metabolic pathways
ID=SAPIG0919	0,01235	SAPIG0919	n	O	nitrogen fixation protein NifU	Cell division
ID=SAPIG1544	0,01235		n	no related COG	hypothetical protien	unknown
ID=SAPIG1469	0,01240	SAPIG1469	y	H		
ID=SAPIG0749	0,01240	SAPIG0749	n	S	hypothetical protein	
ID=SAPIG1394	0,01244	SAPIG1394	n	E	aspartokinase 2 (Aspartokinase II) (A	Unknown
ID=SAPIG1197	0,01247	pyrB	n	F	aspartate carbamoyltransferase	Metabolic pathways
ID=SAPIG0720	0,01253	tagD	y	M		
ID=SAPIG1674	0,01258	SAPIG1674	y	K		
ID=SAPIG2619	0,01282	SAPIG2619	n	no related COG	immunodominant staphylococcal antigen A	
ID=SAPIG0742	0,01290	SAPIG0742	n	P	low-affinity inorganic phosphate tran	Transport
ID=SAPIG0900	0,01290	SAPIG0900	y	C		
ID=SAPIG0926	0,01291	SAPIG0926	n	R	Na+/H+ antiporter family protein	Unknown
ID=SAPIG0967	0,01292	SAPIG0967	n	no related COG	conjugative transposon protein	

ID=SAPIG1201	0,01293	pyrF	n	F	orotidine 5'-phosphate decarboxylase	Unknown
ID=SAPIG0741	0,01294	SAPIG0741	n	P	protein YkaA	Inorganic ion transport and metabolism
ID=SAPIG2552	0,01316	SAPIG2552	n	no related COG	hypothetical protein	Unknown
ID=SAPIG1939	0,01316	SAPIG1939	n	no related COG	tRNA	Unknown
ID=SAPIG2223	0,01316	SAPIG2223	n	no related COG	tRNA	
ID=SAPIG1950	0,01316		n	P	tRNA	
ID=SAPIG1951	0,01316		n	no related COG	tRNA	
ID=SAPIG1618	0,01333	SAPIG1618	n	P	superoxide dismutase [Mn] (General	Unknown
ID=SAPIG2101	0,01351	SAPIG2101	n	no related COG	tRNA	Vitamin B6 metabolism
ID=SAPIG1844	0,01352	menE	y	I, Q		
ID=SAPIG1200	0,01355	carB	n	E, F	carbamoyl-phosphate synthase, large	Recombination
ID=SAPIG0669	0,01361	SAPIG0669	n	K	probable DNA-binding protein	Transport
ID=SAPIG1893	0,01370	SAPIG1893	n	no related COG	tRNA	Unknown
ID=SAPIG2755	0,01389	gidB	n	no related COG	methyltransferase GidB	TCA cycle
ID=SAPIG1380	0,01421	SAPIG1380	n	E, P	oligopeptide transport ATPase	Unknown
ID=SAPIG1062	0,01449	purE	n	F	phosphoribosylaminoimidazole carbo	Metabolic pathways
ID=SAPIG2126	0,01449	SAPIG2126	n	no related COG	hypothetical protein	Unknown
ID=SAPIG0594	0,01464	SAPIG0594	n	H	pyridoxine biosynthesis protein	
ID=SAPIG0922	0,01471	SAPIG0922	n	S	hypothetical protein	
ID=SAPIG1970	0,01476	SAPIG1970	n	R	conserved protein YfkA	Unknown
ID=SAPIG1053	0,01478	SAPIG1053	n	K	cell envelope-related transcriptional attenuator domain family	
ID=SAPIG2072	0,01481	SAPIG2072	y	no related COG		
ID=SAPIG1559	0,01491	rluB	y	J		
ID=SAPIG2274	0,01510	SAPIG2274	n	P	cobalt import ATP-binding protein Cb	Metabolic pathways
ID=SAPIG0572	0,01515	SAPIG0572	n	J	heat shock protein 15	Transport
ID=SAPIG1621	0,01527	SAPIG1621	n	P	ABC transporter, ATP-binding protein	RNA or DNA unwinding
ID=SAPIG2107	0,01529	SAPIG2107	n	T	anti-sigma-B factor antagonist (Anti-	regulators of anti-sigma factors
ID=SAPIG0747	0,01534	SAPIG0747	n	S	hypothetical protein	
ID=SAPIG1499	0,01538		n	no related COG	hypothetical protien	unknown
ID=SAPIG2446	0,01545	SAPIG2446	n	no related COG	NreA	
ID=SAPIG0738	0,01563	SAPIG0738	n	no related COG	sensor protein BceS	
ID=SAPIG0563	0,01575	SAPIG0563	n	J	endoribonuclease L-PSP, putative	DNA-binding repressors and activators
ID=SAPIG0281	0,01587	SAPIG0281	n	no related COG	hypothetical protein	Protease

ID=SAPIG0598	0,01587	SAPIG0598	n	S	UvrB/UvrC motif domain protein	Transport
ID=SAPIG1364	0,01587	SAPIG1364	y	R		
ID=SAPIG1842	0,01597	menC	y	no related COG		
ID=SAPIG0595	0,01604	SAPIG0595	n	H	glutamine amidotransferase subunit	PdxT (Glutamineamidotransferase glutaminase subu
ID=SAPIG1211	0,01613	SAPIG1211	n	no related COG	lipoprotein, putative	Unknown
ID=SAPIG0103	0,01621	SAPIG0103	n	J	tRNA-dihydrouridine synthase	Cell division
ID=SAPIG1351	0,01626	SAPIG1351	n	no related COG	hypothetical protein	Unknown
ID=SAPIG1497	0,01634		n	no related COG	hypothetical phage-related protein	unknown
ID=SAPIG1296	0,01639	SAPIG1296	n	S	hypothetical protein	Surface antigen
ID=SAPIG0562	0,01641	purR	n	F	pur operon repressor	
ID=SAPIG0859	0,01643	rnr	y	K		
ID=SAPIG1343	0,01667	SAPIG1343	n	S	protein Stu0508	Pentose-phosphate pathway
ID=SAPIG0558	0,01676	SAPIG0558	n	L	putative primase homolog	Unknown
ID=SAPIG0582	0,01677	folK	y	H		
ID=SAPIG1784	0,01679	fhs	n	F	formate--tetrahydrofolate ligase	Signal transduction
ID=SAPIG0633	0,01699	SAPIG0633	n	J, F	tRNA-specific adenosine deaminase	Unknown
ID=SAPIG1868	0,01707	SAPIG1868	n	no related COG	hypothetical protein	Vitamin B6 metabolism
ID=SAPIG1151	0,01728	SAPIG1151	n	no related COG	fprl1 inhibitory protein (flipr)	Unknown
ID=SAPIG0551	0,01741	SAPIG0551	n	S	stage 0 sporulation protein YaaT	Transcriptional regulator
ID=SAPIG1177	0,017412935	SAPIG1177	y	D		
ID=SAPIG2154	0,01746	upp	n	F	uracil phosphoribosyltransferase	
ID=SAPIG1715	0,01747	SAPIG1715	n	L	DNA repair protein RadC	
ID=SAPIG2110	0,01754	SAPIG2110	n	no related COG	hypothetical protein	stress
ID=SAPIG1017	0,01754	SAPIG1017	y	P		
ID=SAPIG1369	0,01754	SAPIG1369	n	E, H	anthranilate synthase glutamine amidotransferase	
ID=SAPIG0867	0,01767	SAPIG0867	n	no related COG	lpxtg-motif cell wall anchor domain	Transport
ID=SAPIG2206	0,01778	SAPIG2206	n	no related COG	hypothetical protein	cell wall
ID=SAPIG1149	0,01786	SAPIG1149	n	no related COG	hypothetical protein	
ID=SAPIG2165	0,01786	SAPIG2165	n	K	putative transcriptional regulator sup	Unknown
ID=SAPIG1975	0,01812	SAPIG1975	n	no related COG	hypothetical protein	Metabolic pathways
ID=SAPIG2640	0,01812	SAPIG2640	n	no related COG	hypothetical protein	
ID=SAPIG1125	0,01823	SAPIG1125	n	no related COG	iron-regulated heme-iron binding pro	Unknown
ID=SAPIG2055	0,01837	SAPIG2055	n	K	GntR family regulatory protein	Unknown

ID=SAPIG0974	0,01839	clpB	n	O	ATP-dependent chaperone ClpB	DNA replication, recombination, and repair
ID=SAPIG1320	0,01844	SAPIG1320	n	no related COG	hypothetical protein	Unknown
ID=SAPIG1046	0,01846	SAPIG1046	n	E	glutamyl endopeptidase (Staphyloco	Unknown
ID=SAPIG1488	0,01852	SAPIG1488	n	no related COG	hypothetical protein	
ID=SAPIG1650	0,01864	lepA	n	M	GTP-binding protein LepA	
ID=SAPIG2255	0,01865	lacA	n	G	galactose-6-phosphate isomerase, LacA subunit	
ID=SAPIG1302	0,01870	glpK	n	C	glycerol kinase	
ID=SAPIG0786	0,01880	SAPIG0786	n	no related COG	hypothetical protein	
ID=SAPIG1925	0,01883	SAPIG1925	n	R	regulatory protein	Regulation
ID=SAPIG1507	0,01887		n	no related COG	hypothetical protien	unknown
ID=SAPIG1786	0,01896	SAPIG1786	n	no related COG	acetoin utilization protein AcuA	Unknown
ID=SAPIG0965	0,01907	SAPIG0965	n	no related COG	transcriptional regulator, Cro/CI family	
ID=SAPIG1019	0,01913	SAPIG1019	n	P	Na+ transporting ATP synthase	
ID=SAPIG1396	0,01914	dapA	y	M, E		
ID=SAPIG1556	0,01935	SAPIG1556	n	no related COG	lipoprotein, putative	Unknown
ID=SAPIG0635	0,01940	SAPIG0635	n	R	fmn-dependent NADPH-azoreductase	
ID=SAPIG1190	0,01942	SAPIG1190	n	D	cell-divisio initiation protein	Unknown
ID=SAPIG1589	0,01943	xseA	n	L	exodeoxyribonuclease VII, large subunit	
ID=SAPIG1654	0,01948	SAPIG1654	n	F	ComE operon protein 2	Translation, ribosomal structure
ID=SAPIG1142	0,01964	uvrC	n	L	excinuclease ABC, C subunit	Unknown
ID=SAPIG0616	0,01970	SAPIG0616	n	J	methyltransferase small domain supe	pyrimidine metabolism
ID=SAPIG0236	0,01984	pflA	n	O	pyruvate formate-lyase 1-activating	transcriptional regulators
ID=SAPIG1404	0,01990	SAPIG1404	n	K	cold shock protein, CSD family	
ID=SAPIG2765	0,00205	rRNA				**508 protein encoding genes**
ID=SAPIG2773	0,00870	rRNA				
ID=SAPIG2766	0,00200	rRNA				
ID=SAPIG2768	0,00137	rRNA				
ID=SAPIG2771	0,00128	rRNA				
ID=SAPIG2770	0,00103	rRNA				
ID=SAPIG2776	0,00870	rRNA				
ID=SAPIG2763	0,00103	rRNA				
ID=SAPIG2761	0,00320	rRNA				
ID=SAPIG2760	0,00137	rRNA				

ID=SAPIG2769	0,00506	rRNA
ID=SAPIG2777	0,00870	rRNA
ID=SAPIG2764	0,00068	rRNA
ID=SAPIG2767	0,00320	rRNA
ID=SAPIG2757	0,00072	tRNA
ID=SAPIG0585	0,01316	tRNA
ID=SAPIG1022	0,01333	tRNA
ID=SAPIG1930	0,01389	tRNA

Table S3. Comparison of essential gene lists of s. aureus

Description								
Essential in S. aureus Fey et al. 2013	JE2 (USA300)	y	y	y	y	y	y	y
Essential in S. aureus Bae et al. 2004	Newman	y	y	y	y	y	y	y
Essential by Automated TMDH Chaudhuri et al. 2009	NCTC8325 (SH1000)	y	y	y	y	y	y	y
Essential or advantageous by TraDIS This study	S0385	y	y	y	n	y	y	y
ST398 genes		ID=SAPIG0001	ID=SAPIG0002	ID=SAPIG0016	ID=SAPIG2139	ID=SAPIG0544	ID=SAPIG0550	ID=SAPIG1210	ID=SAPIG1652
NCTC8325 genes		SAOUHSC_00001	SAOUHSC_00002	SAOUHSC_00018	SAOUHSC_00349	SAOUHSC_00442	SAOUHSC_00454	SAOUHSC_01179	SAOUHSC_01241
Category	**S. aureus strains:**	**DNA metabolism**							
DNA replication		dnaA	dnaN	dnaB	single-strand binding protein family,	DNA polymerase III gamma/tau	DNA polymerase III delta subunit	primosomal protein N'	holA

DNA replication	SAOUHSC_01470	ID=SAPIG1452	n	y	y	y	nth
DNA replication	SAOUHSC_01663	ID=SAPIG1627	y	y	y	y	DNA primase
DNA replication	SAOUHSC_01690	ID=SAPIG0550	y	y	y	y	DNA polymerase III delta' subunit
DNA replication	SAOUHSC_01791	ID=SAPIG1738	y	y	y	y	dnaI
DNA replication	SAOUHSC_01792	ID=SAPIG1739	y	y	y	y	chromosome replication initiation/membrane attachment protein
DNA replication	SAOUHSC_01811	ID=SAPIG1756	y	y	y	y	DNA polymerase III subunit alpha
DNA replication	SAOUHSC_02122	ID=SAPIG1998	y	y	y	n	DNA ligase, NAD-dependent
DNA replication	SAOUHSC_02123	ID=SAPIG1999	y	y	y	y	ATP-dependent DNA helicase PcrA
DNA packaging and segregation	SAOUHSC_00005	ID=SAPIG0005	y	y	y	y	gyrB
DNA packaging and segregation	SAOUHSC_00006	ID=SAPIG0006	y	y	n	n	gyrA
DNA packaging and segregation	SAOUHSC_01204	ID=SAPIG1234	y	n	n		smc
DNA packaging and segregation	SAOUHSC_01222	ID=SAPIG1252	y	y	n	y	DNA topoisomerase I
DNA packaging and segregation	SAOUHSC_01351	ID=SAPIG1356	y	y	y	y	DNA topoisomerase IV, B subunit
DNA packaging and segregation	SAOUHSC_01352	ID=SAPIG1357	y	y	n	n	DNA topoisomerase IV, A subunit
DNA packaging and segregation	SAOUHSC_01490	ID=SAPIG1471	y	y	y	y	DNA-binding protein HU
DNA packaging and segregation	SAOUHSC_01588	ID=SAPIG1559	y	n	n		rluB
DNA packaging and segregation	SAOUHSC_01589	ID=SAPIG1560	y	n	n		scpB
DNA packaging and segregation	SAOUHSC_01750	ID=SAPIG1704	y	y	y	y	Holliday junction DNA helicase RuvB
DNA packaging and segregation	SAOUHSC_01751	ID=SAPIG1705	y	y	y	y	Holliday junction DNA helicase RuvA
DNA packaging and segregation	SAOUHSC_01466	ID=SAPIG1449.p0	y	y	y	y	recU
DNA packaging and segregation	SAOUHSC_01720	ID=SAPIG1680	y	n	n	n	hypothetical protein
DNA packaging and segregation	SAOUHSC_02791	ID=SAPIG1459	n	n	n	n	pyrophosphohydrolase, putative
RNA metabolism							
Basic transcription machinery	SAOUHSC_00524	ID=SAPIG0617	y	y	y	y	rpoB
Basic transcription machinery	SAOUHSC_00525	ID=SAPIG0618	y	y	y	y	rpoC
Basic transcription machinery	SAOUHSC_01662	ID=SAPIG1626	y	y	y	y	rpoD
Basic transcription machinery	SAOUHSC_02485	ID=SAPIG2277	y	y	y	y	rpoA
RNA modification	SAOUHSC_01035	ID=SAPIG1086	y	y	n		protein_0989
RNA modification	SAOUHSC_01203	ID=SAPIG1233	y	n	n		rnc

RNA modification	SAOUHSC_01209	ID=SAPIG1239	y	y	y	y	rimM
RNA modification	SAOUHSC_01210	ID=SAPIG1240	y	y	n	y	trmD
RNA modification	SAOUHSC_01252	ID=SAPIG1277	y	y	y	y	RNA-metabolising
RNA modification	SAOUHSC_01474	ID=SAPIG1457	y	y	y	y	cca-adding enzyme
RNA modification	SAOUHSC_01598	ID=SAPIG1570	y	y	y	y	ribonuclease Z
RNA modification	SAOUHSC_01725	?		y	y	y	tRNA methyl transferase, putative (SAOUHSC_01725)
RNA modification	SAOUHSC_01726	ID=SAPIG1685	y	y	y	y	trmU
RNA modification	SAOUHSC_01988	ID=SAPIG1921	n	n	n		trmH
RNA modification	SAOUHSC_03053	ID=SAPIG2757	y	y	y	n	trmE
RNA modification	SAOUHSC_03054	ID=SAPIG2758	y	y	y	y	rnpA
RNA regulation	SAOUHSC_00020	ID=SAPIG0020	y	y	y	y	transcriptional regulatory protein
RNA regulation	SAOUHSC_00021	ID=SAPIG0021	y	y	y	y	YycG
RNA regulation	SAOUHSC_01243	ID=SAPIG1268	y	y	y	y	nusA
RNA regulation	SAOUHSC_01592	ID=SAPIG1564	y	n	y	n	transcriptional regulator
RNA regulation	SAOUHSC_00620	ID=SAPIG0697	y	n	n	n	staphylococcal accessory regulator A
RNA regulation	SAOUHSC_00803	ID=SAPIG0859	y	n	y	n	ribonulease R
RNA regulation	SAOUHSC_00934	ID=SAPIG0993	y	n	y	y	regulatory protein spx
RNA regulation	SAOUHSC_01285	ID=SAPIG1310	y	y	y	y	glutamine synthetase repressor
RNA regulation	SAOUHSC_01361	ID=SAPIG1363	y	y	n	y	regulatory protein MsrR
RNA regulation	SAOUHSC_01333	ID=SAPIG1341	y	y	y	y	LexA repressor
RNA regulation	SAOUHSC_01714	ID=SAPIG1674	y	n	n	y	greA
Protein synthesis							
Ribosomal proteins	SAOUHSC_00017	ID=SAPIG0015	y	n	y	y	ribosomal protein L9
Ribosomal proteins	SAOUHSC_00348	ID=SAPIG0445	y	y	y	y	ribosomal protein S6
Ribosomal proteins	SAOUHSC_00350	ID=SAPIG0447.p0	y	n	y	y	ribosomal protein S18
Ribosomal proteins	SAOUHSC_00518	ID=SAPIG0612	y	y	y	y	ribosomal protein L11
Ribosomal proteins	SAOUHSC_00519	ID=SAPIG0613	y	y	y	y	ribosomal protein L1
Ribosomal proteins	SAOUHSC_00520	ID=SAPIG0614	y	y	y	y	ribosomal protein L10
Ribosomal proteins	SAOUHSC_00521	ID=SAPIG0615	y	y	y	y	ribosomal protein L7/L12
Ribosomal proteins	SAOUHSC_00527	ID=SAPIG0620	y	y	y	y	ribosomal protein S12

Ribosomal proteins	SAOUHSC_00528	ID=SAPIG0621	y	y	y	y	ribosomal protein S7
Ribosomal proteins	SAOUHSC_01078	ID=SAPIG1124	y	y	y	y	ribosomal protein L32
Ribosomal proteins	SAOUHSC_01191	ID=SAPIG1223	n	n	n	y	ribosomal protein L28
Ribosomal proteins	SAOUHSC_01208	ID=SAPIG1238	y	n	y	y	ribosomal protein S16
Ribosomal proteins	SAOUHSC_01211	ID=SAPIG1241	y	y	n	y	ribosomal protein L19
Ribosomal proteins	SAOUHSC_01232	ID=SAPIG1258	y	n	y	y	ribosomal protein S2
Ribosomal proteins	SAOUHSC_01250	ID=SAPIG1275	y	n	y	y	ribosomal protein S15
Ribosomal proteins	SAOUHSC_01328	ID=SAPIG1337	n	n	n	y	ribosomal protein L33
Ribosomal proteins	SAOUHSC_01651	ID=SAPIG1616	n	n	y	y	ribosomal protein L33
Ribosomal proteins	SAOUHSC_01678	ID=SAPIG1641	y	n	y	y	ribosomal protein S21
Ribosomal proteins	SAOUHSC_01689	ID=SAPIG1651	y	n	y	y	ribosomal protein S20
Ribosomal proteins	SAOUHSC_01755	ID=SAPIG1708	y	y	n	y	ribosomal protein L27
Ribosomal proteins	SAOUHSC_01757	ID=SAPIG1710	y	y	y	y	ribosomal protein L21
Ribosomal proteins	SAOUHSC_01784	ID=SAPIG1733	y	y	y	y	ribosomal protein L20
Ribosomal proteins	SAOUHSC_01785	ID=SAPIG1734	y	n	y	y	ribosomal protein L35
Ribosomal proteins	SAOUHSC_01829	ID=SAPIG1772	y	y	y	y	ribosomal protein S4
Ribosomal proteins	SAOUHSC_02361	ID=SAPIG2162	y	y	y	y	ribosomal protein L31
Ribosomal proteins	SAOUHSC_02477	ID=SAPIG2270	y	y	y	y	ribosomal protein S9
Ribosomal proteins	SAOUHSC_02478	ID=SAPIG2271	y	y	y	y	ribosomal protein L13
Ribosomal proteins	SAOUHSC_02484	ID=SAPIG2276	y	y	y	y	ribosomal protein L17
Ribosomal proteins	SAOUHSC_02486	ID=SAPIG2278	y	y	y	y	ribosomal protein S11
Ribosomal proteins	SAOUHSC_02487	ID=SAPIG2279	y	y	y	y	rpsM
Ribosomal proteins	SAOUHSC_02488	ID=SAPIG2280	y	y	y	y	ribosomal protein L36
Ribosomal proteins	SAOUHSC_02492	ID=SAPIG2284	y	y	y	y	ribosomal protein L15
Ribosomal proteins	SAOUHSC_02493	ID=SAPIG2285	y	y	y	y	ribosomal protein L30
Ribosomal proteins	SAOUHSC_02494	ID=SAPIG2286	y	y	y	y	ribosomal protein S5
Ribosomal proteins	SAOUHSC_02495	ID=SAPIG2287	y	y	y	y	ribosomal protein L18
Ribosomal proteins	SAOUHSC_02496	ID=SAPIG2288	y	y	y	y	ribosomal protein L6
Ribosomal proteins	SAOUHSC_02498	ID=SAPIG2289	y	y	y	y	ribosomal protein S8
Ribosomal proteins	SAOUHSC_02499	ID=SAPIG2290	y	y	y	y	rpsN
Ribosomal proteins	SAOUHSC_02500	ID=SAPIG2291	y	y	y	y	50S ribosomal protein L5
Ribosomal proteins	SAOUHSC_02501	ID=SAPIG2292	n	y	y	y	ribosomal protein L24
Ribosomal proteins	SAOUHSC_02502	ID=SAPIG2293	y	y	y	y	ribosomal protein L14

Ribosomal proteins	SAOUHSC_02503	ID=SAPIG2294	y	y	y	y	30S ribosomal protein S17
Ribosomal proteins	SAOUHSC_02504	ID=SAPIG2295	y	y	y	y	ribosomal protein L29
Ribosomal proteins	SAOUHSC_02505	ID=SAPIG2296	y	y	y	y	ribosomal protein L16
Ribosomal proteins	SAOUHSC_02506	ID=SAPIG2297	y	y	y	y	ribosomal protein S3
Ribosomal proteins	SAOUHSC_02507	ID=SAPIG2298	y	y	y	y	ribosomal protein L22
Ribosomal proteins	SAOUHSC_02508	ID=SAPIG2299	y	y	y	y	ribosomal protein S19
Ribosomal proteins	SAOUHSC_02509	ID=SAPIG2300	y	y	y	y	ribosomal protein L2
Ribosomal proteins	SAOUHSC_02510	ID=SAPIG2301	y	y	y	y	ribosomal protein L23
Ribosomal proteins	SAOUHSC_02511	ID=SAPIG2302	y	y	y	y	rplD
Ribosomal proteins	SAOUHSC_02512	ID=SAPIG2303	y	y	y	y	ribosomal protein L3
Ribosomal proteins	SAOUHSC_03055	ID=SAPIG2759	y	n	y	y	ribosomal protein L34
Ribosomal proteins	SAOUHSC_00474	?		n	n	y	ribosomal 5S rRNA E-loop binding
tRNA synthetase	SAOUHSC_00009	ID=SAPIG0009	y	y	y	y	serS
tRNA synthetase	SAOUHSC_00461	ID=SAPIG0556	y	y	n	y	metG
tRNA synthetase	SAOUHSC_00493	ID=SAPIG0583	y	y	y	y	lysS
tRNA synthetase	SAOUHSC_00509	ID=SAPIG0603	y	y	n	n	gltX
tRNA synthetase	SAOUHSC_00511	ID=SAPIG0605	y	y	n	n	cysS
tRNA synthetase	SAOUHSC_00611	ID=SAPIG0688	y	y	y	y	argS
tRNA synthetase	SAOUHSC_00933	ID=SAPIG0992	y	y	y	y	trpS
tRNA synthetase	SAOUHSC_01092	ID=SAPIG1134	y	y	y	y	pheS
tRNA synthetase	SAOUHSC_01093	ID=SAPIG1135	y	y	y	y	pheT
tRNA synthetase	SAOUHSC_01159	ID=SAPIG1191	y	y	y	y	ileS
tRNA synthetase	SAOUHSC_01240	ID=SAPIG1265	y	y	y	y	proS
tRNA synthetase	SAOUHSC_01471	ID=SAPIG1454	n	n	n		asnS
tRNA synthetase	SAOUHSC_01666	ID=SAPIG1630	y	y	y	y	glyS
tRNA synthetase	SAOUHSC_01722	ID=SAPIG1682	y	y	y	y	alaS
tRNA synthetase	SAOUHSC_01737	ID=SAPIG1693	y	y	y	y	aspS
tRNA synthetase	SAOUHSC_01738	ID=SAPIG1694	y	y	y	y	hisS
tRNA synthetase	SAOUHSC_01767	ID=SAPIG1718	y	y	y	y	valS
tRNA synthetase	SAOUHSC_01788	ID=SAPIG1737	y	y	y	n	thrS
tRNA synthetase	SAOUHSC_01839	ID=SAPIG1781	y	y	y	y	tyrS
tRNA synthetase	SAOUHSC_01875	ID=SAPIG1811	y	y	y	y	leuS

tRNA synthetase	SAOUHSC_02116	ID=SAPIG1993	y	y	n	y	amidotransferase subunit B
tRNA synthetase	SAOUHSC_02117	ID=SAPIG1994	y	y	y	y	amidotransferase subunit A
tRNA synthetase	SAOUHSC_02118	ID=SAPIG1995	y	n	y	y	glutamyl-tRNA(Gln) amidotransferase, C subunit
tRNA Ile modification	SAOUHSC_00484	ID=SAPIG0575	y	y	y	y	tRNA(Ile)-lysidine synthase
tRNA met modification	SAOUHSC_01183	ID=SAPIG1215	y	y	y	y	fmt methionyl-tRNA formyltransf.
Translation factors	SAOUHSC_00475	ID=SAPIG0568	y	n	y	y	pth
Translation factors	SAOUHSC_00529	ID=SAPIG0622	y	y	y	y	fusA
Translation factors	SAOUHSC_00530	ID=SAPIG0623	y	y	y	y	tuf
Translation factors	SAOUHSC_00771	ID=SAPIG0830	y	y	y	y	prfB
Translation factors	SAOUHSC_01234	ID=SAPIG1259	y	y	y	y	tsf
Translation factors	SAOUHSC_01236	ID=SAPIG1261	y	y	y	y	frr
Translation factors	SAOUHSC_01246	ID=SAPIG1271	y	y	y	y	infB
Translation factors	SAOUHSC_01625	ID=SAPIG1594	y	y	y	y	efp
Translation factors	SAOUHSC_01698	ID=SAPIG1660	y	y	y	y	conserved hypothetical protein
Translation factors	SAOUHSC_01741	ID=SAPIG1696	y	y	y	y	dtd
Translation factors	SAOUHSC_01786	ID=SAPIG1735	y	y	y	y	infC
Translation factors	SAOUHSC_02359	ID=SAPIG2160	y	y	y	y	prfA
Translation factors	SAOUHSC_02489	ID=SAPIG2281	y	y	y	y	infA
Translation factors	SAOUHSC_00804	ID=SAPIG0860	y	n	y	y	smpB
Protein folding	SAOUHSC_01682	ID=SAPIG1645	y	y	y	y	dnaJ
Protein folding	SAOUHSC_01683	ID=SAPIG1646	y	y	y	y	dnaK
Protein folding	SAOUHSC_01684	ID=SAPIG1647	y	y	y	y	grpE
Protein folding	SAOUHSC_02254	ID=SAPIG2066	y	y	y	y	groL
Protein folding	SAOUHSC_02255	ID=SAPIG2067	y	y	y	y	groS
Protein modification	SAOUHSC_00790	ID=SAPIG0846	y	y	n	n	clpP
Protein modification	SAOUHSC_02102	ID=SAPIG1981	y	y	y	y	map
Protein modification	SAOUHSC_01038	ID=SAPIG1088	y	n	y	y	def peptide deformylase

Protein modification	?	ID=SAPIG1214	y				def polypeptide deformylase
Protein translocation	SAOUHSC_00516	ID=SAPIG0610	y	n	y	y	secE
Protein translocation	SAOUHSC_00769	ID=SAPIG0829	y	y	y	y	secA
Protein translocation	SAOUHSC_00903	ID=SAPIG0948	y	y	y	y	lepB
Protein translocation	SAOUHSC_01205	ID=SAPIG1235	y	y	y	y	cell division protein FtsY
Protein translocation	SAOUHSC_01207	ID=SAPIG1237	y	y	y	y	ffh
Protein translocation	SAOUHSC_01746	ID=SAPIG1700	y	y	y	y	export membrane protein SecDF
Protein translocation	SAOUHSC_01972	ID=SAPIG1907	n	n	n		foldase protein PrsA
Protein translocation	SAOUHSC_02327	ID=SAPIG2133	y	y	y	y	membrane protein OxaA
Protein translocation	SAOUHSC_02491	ID=SAPIG2283	y	y	y	y	preprotein translocase, SecY subunit
Cell envelope/Cell wall and associated proteins							
Lipids	SAOUHSC_00920	ID=SAPIG0979	y	y	y	y	3-oxoacyl-(acyl carrier protein) synthase III
Lipids	SAOUHSC_00921	ID=SAPIG0980	y	y	y	y	fabF
Lipids	SAOUHSC_00947	ID=SAPIG1006	y	y	y	y	enoyl-acyl-carrier-protein reductase
Lipids	SAOUHSC_01197	ID=SAPIG1228	y	y	y	y	plsX
Lipids	SAOUHSC_01198	ID=SAPIG1229	y	y	y	y	fabD
Lipids	SAOUHSC_01199	ID=SAPIG1230	y	y	y	y	fabG
Lipids	SAOUHSC_01201	ID=SAPIG1232	y	n	y	y	acpP
Lipids	SAOUHSC_01473	ID=SAPIG1456	y	y	y	y	biotin-acetyl-CoA-carboxylase ligase
Lipids	SAOUHSC_01623	ID=SAPIG1592	y	y	y	y	accC
Lipids	SAOUHSC_01624	ID=SAPIG1593	y	y	y	y	accB
Lipids	SAOUHSC_01808	ID=SAPIG1753	y	y	y	n	accA
Lipids	SAOUHSC_01809	ID=SAPIG1754	y	y	y	y	accD
Lipids	SAOUHSC_02306	ID=SAPIG2112	y	y	y	y	acpS
Lipids	SAOUHSC_02336	ID=SAPIG2141	y	y	y	y	fabZ
Lipids	SAOUHSC_00881	ID=SAPIG0927	y	n	n	n	thioesterase family protein
Phospholipids	SAOUHSC_01238	ID=SAPIG1263	y	y	y	y	cdsA
Phospholipids	SAOUHSC_01260	ID=SAPIG1285	y	y	y	y	pgsA
Phospholipids	SAOUHSC_01491	ID=SAPIG1472	n	n	y	y	glycerol-3-phosphate dehydrogenase, NAD-dependent

Phospholipids	SAOUHSC_01837	ID=SAPIG1779	y	n	y	y	1-acylglycerol-3-phosphate O-acyltransferase
Phospholipids	SAOUHSC_02114	ID=SAPIG1991	y	y	y	n	conserved hypothetical protein
Cell wall/amino sugar	SAOUHSC_00120	ID=SAPIG0169	n	n	n		UDP-N-acetylglucosamine 2-epimer.
Cell wall/amino sugar	SAOUHSC_00129	ID=SAPIG0178	n	n	n		UDP-N-acetylglucosamine 2-epimer.
Cell wall/amino sugar	SAOUHSC_00471	ID=SAPIG0565	y	y	n		glmU
Cell wall/amino sugar	SAOUHSC_02352	ID=SAPIG2153	n	n	n		UDP-N-acetylglucosamine 2-epimer.
Cell wall/amino sugar	SAOUHSC_02399	ID=SAPIG2209	y	y	y	y	glmS
Cell wall/amino sugar	SAOUHSC_02405	ID=SAPIG2216	y	y	y	n	glmM
Diaminopimelate biosynthesis	SAOUHSC_01395	ID=SAPIG1395	n	n	n		asd
Diaminopimelate biosynthesis	SAOUHSC_01396	ID=SAPIG1396	y	n	n		dapA
Diaminopimelate biosynthesis	SAOUHSC_01397	ID=SAPIG1397	y	n	n		dapB
Diaminopimelate biosynthesis	SAOUHSC_01398	ID=SAPIG1398	n	n	n		tetrahydrodipicolinate acetyltransf.
Peptidoglycan biosynthesis	SAOUHSC_00752	ID=SAPIG0814	y	y	y	y	murB
Peptidoglycan biosynthesis	SAOUHSC_00954	ID=SAPIG1985	y	y	y	y	Mur ligase family protein
Peptidoglycan biosynthesis	SAOUHSC_01106	ID=SAPIG1146	y	y	y	y	murI
Peptidoglycan biosynthesis	SAOUHSC_01146	ID=SAPIG1179	y	y	y	y	mraY
Peptidoglycan biosynthesis	SAOUHSC_01147	ID=SAPIG1180	y	y	y	y	murD
Peptidoglycan biosynthesis	SAOUHSC_01237	ID=SAPIG1262	y	y	y	y	uppS
Peptidoglycan biosynthesis	SAOUHSC_01400	ID=SAPIG1400	n	n	n	y	alanine racemase 2
Peptidoglycan biosynthesis	SAOUHSC_01424	ID=SAPIG1419	y	y	y	y	murG
Peptidoglycan biosynthesis	SAOUHSC_01467	ID=SAPIG1178	y	y	y	y	penicillin-binding protein 2B (PBP-2B)
Peptidoglycan biosynthesis	SAOUHSC_01739	ID=SAPIG1695	y	y	n	n	lytH
Peptidoglycan biosynthesis	SAOUHSC_01856	ID=SAPIG1793	y	y	y	y	murC
Peptidoglycan biosynthesis	SAOUHSC_02107	ID=SAPIG2123	y	y	y	y	UDP-N-acetylmuramoyl-tripeptide--D-alanyl-D-alanine ligase
Peptidoglycan biosynthesis	SAOUHSC_02305	ID=SAPIG2111	n	n	n	y	alr
Peptidoglycan biosynthesis	SAOUHSC_02317	ID=SAPIG1014	y	y	y	y	ligase
Peptidoglycan biosynthesis	SAOUHSC_02318	ID=SAPIG2124	y	y	y	y	D-alanine--D-alanine ligase
Peptidoglycan biosynthesis	SAOUHSC_02337	ID=SAPIG2142	y	y	y	n	murA
Peptidoglycan biosynthesis	SAOUHSC_02527	ID=SAPIG2314	y	y	y	y	femX

Peptidoglycan biosynthesis	SAOUHSC_01373	ID=SAPIG1375	y	y	y	y	femA
Peptidoglycan biosynthesis	SAOUHSC_01374	ID=SAPIG1376	y	y	y	y	femB
Teichoic acid biosynthesis	SAOUHSC_00223	ID=SAPIG0270	y	y	y	y	teichoic acid biosynthesis protein B
Teichoic acid biosynthesis	SAOUHSC_00227	?	n	y	y	y	
Teichoic acid biosynthesis	SAOUHSC_00640	ID=SAPIG0715	y	n	y	y	UDP-N-acetyl-D-mannosamine transferase
Teichoic acid biosynthesis	SAOUHSC_00641	ID=SAPIG0716	y	y	y	y	teichoic acids export ATP-binding protein TagH
Teichoic acid biosynthesis	SAOUHSC_00642	ID=SAPIG0717	y	y	y	y	teichoic acid translocation permease protein TagG
Teichoic acid biosynthesis	SAOUHSC_00643	ID=SAPIG0718	y	y	y	n	teichoic acid biosynthesis protein
Teichoic acid biosynthesis	SAOUHSC_00645	ID=SAPIG0720	y	n	y	y	glycerol-3-phosphate cytidylyltransf.
Teichoic acid biosynthesis	SAOUHSC_00762	ID=SAPIG0419	n	y	y	n	bacterial luciferase family protein
Teichoic acid biosynthesis	SAOUHSC_00869	ID=SAPIG0915	y	y	y	y	dltA
Teichoic acid biosynthesis	SAOUHSC_00870	ID=SAPIG0916	y	y	n	y	dltB
Teichoic acid biosynthesis	SAOUHSC_00871	ID=SAPIG0917	y	y	y	y	dltC
Teichoic acid biosynthesis	SAOUHSC_00872	ID=SAPIG0918	y	y	n	y	poly D-alanine transfer protein
Cell division	SAOUHSC_00482	ID=SAPIG0573	y	n	y	y	cell-division initiation protein
Cell division	SAOUHSC_01144	ID=SAPIG1177	y	n	y	y	cell division protein
Cell division	?	ID=SAPIG1186	y				ylmE
Cell division	SAOUHSC_01154	ID=SAPIG1187	y	y	y	y	ylmF
Cell division	SAOUHSC_01149	ID=SAPIG1182	y	y	y	y	ftsA
Cell division	SAOUHSC_01150	ID=SAPIG1183	y	y	y	y	ftsZ
Cell division	SAOUHSC_01827	ID=SAPIG1770	y	y	y	y	EzrA
Cell division	SAOUHSC_03049	ID=SAPIG2754	y	y	y	n	protein YyaA
Cell division	SAOUHSC_01148	ID=SAPIG1445	y	y	y	y	cell division protein diviva
Cell division	SAOUHSC_01462	?		n	y	n	conserved hypothetical protein
Cell division	SAOUHSC_01063	ID=SAPIG1110	y	y	y	n	cell cycle protein
Cell division	SAOUHSC_01145	ID=SAPIG1450	n	y	y	y	penicillin-binding protein (PBP1)
Cell division	SAOUHSC_03052	ID=SAPIG2756	y	y	n	n	gidA
Cell shape	SAOUHSC_01759	ID=SAPIG1712	y	n	y		mreC

Cell shape	SAOUHSC_02319	ID=SAPIG2125	y	n	y		rod shape determining protein RodA
Cell shape	SAOUHSC_01062	ID=SAPIG1109	n	n	n		protein BLi01701
Na/H transporter	SAOUHSC_00625	ID=SAPIG0702	n	n	n		Na(+)/H(+) antiporter subunit A
Na/H transporter	SAOUHSC_00626	ID=SAPIG0703	n	n	y		Na(+)/H(+) antiporter subunit B
Na/H transporter	SAOUHSC_00627	ID=SAPIG0704	n	n	y		Na(+)/H(+) antiporter subunit C
Na/H transporter	SAOUHSC_00628	ID=SAPIG0706	n	n	n		Na+/H+ antiporter
Na/H transporter	SAOUHSC_00884	ID=SAPIG0707	n	n	y		multisubunit Na+/H+ antiporter, MnhF subunit
Na/H transporter	SAOUHSC_00883	ID=SAPIG0930	y	n			Na(+)/H(+) antiporter subunit F
Na/H transporter	SAOUHSC_00886	ID=SAPIG0932	y	n	y		Na(+)/H(+) antiporter subunit D
Na/H transporter	SAOUHSC_00887	ID=SAPIG0933	y	n	y		Na(+)/H(+) antiporter subunit C
Na/H transporter	SAOUHSC_00888	ID=SAPIG0934	y	n	y		Na(+)/H(+) antiporter subunit B
Na/H transporter	SAOUHSC_00889	ID=SAPIG0935	y	n	y		Na(+)/H(+) antiporter subunit A
Cell envelope - other	SAOUHSC_00998	ID=SAPIG1215	y	y	n	n	fmt
Cell envelope - other	SAOUHSC_01359	ID=SAPIG1361	y	y	y	n	fmtC
Cell envelope - other	SAOUHSC_01501	ID=SAPIG1480	y	y	y	n	elastin-binding protein EbpS
Carbon metabolism							
Glycolysis	SAOUHSC_00472	ID=SAPIG0566	y	y	y	y	prsA
Glycolysis	SAOUHSC_00796	ID=SAPIG0852	y	y	y	y	pgk
Glycolysis	SAOUHSC_00797	ID=SAPIG0853	y	y	y	y	tpiA
Glycolysis	SAOUHSC_00798	ID=SAPIG0854	y	n	n	n	pgmI
Glycolysis	SAOUHSC_00799	ID=SAPIG0855	y	y	y	y	eno
Glycolysis	SAOUHSC_00900	ID=SAPIG0945	y	y	n	y	pgi
Glycolysis	SAOUHSC_01337	ID=SAPIG1344	y	y	y	y	tkt
Glycolysis	SAOUHSC_01806	ID=SAPIG1751	n	y	n	y	pyk
Glycolysis	SAOUHSC_01807	ID=SAPIG1752	y	n	y	n	pfkA
Glycolysis	SAOUHSC_02366	ID=SAPIG2167	y	y	y	n	fba
Glycolysis	SAOUHSC_00795	ID=SAPIG0851	y	y	y	n	gap
Pentose phosphate	SAOUHSC_01189	ID=SAPIG1221	y	y	y	y	rpe

Pentose phosphate	SAOUHSC_01599	ID=SAPIG1571	y	n	y	y	zwf
Pentose phosphate	SAOUHSC_01605	ID=SAPIG1576	y	y	n	y	gnd
Pentose phosphate	SAOUHSC_02612	ID=SAPIG2388	y	y	n	n	rpiA
Intermediary metabolism	SAOUHSC_01216	ID=SAPIG1245	y	y	y	n	succinyl-CoA synthetase beta chain
Intermediary metabolism	SAOUHSC_01287	ID=SAPIG1311	y	y	y	y	glnA
Intermediary metabolism	SAOUHSC_00788	?		y	y	y	conserved hypothetical protein
Regulation	SAOUHSC_00781	ID=SAPIG0836	y	n	n	y	hprK
Regulation	SAOUHSC_01028	ID=SAPIG1081	y	n	y	y	phosphocarrier protein HPr
Respiratory pathways							
Isoprenoid/Mevalonate biosynth	SAOUHSC_00225	ID=SAPIG0268	y	y	y	y	protein YvcK
Isoprenoid/Mevalonate biosynth	SAOUHSC_00336	ID=SAPIG0434	y	y	n	n	acetyl-CoA acetyltransferase
Isoprenoid/Mevalonate biosynth	SAOUHSC_00466	ID=SAPIG0561	n	n	n		ispE
Isoprenoid/Mevalonate biosynth	SAOUHSC_00577	ID=SAPIG0664	y	y	y	y	mvk
Isoprenoid/Mevalonate biosynth	SAOUHSC_00578	ID=SAPIG0665	y	y	y	y	mvaD
Isoprenoid/Mevalonate biosynth	SAOUHSC_00579	ID=SAPIG0666	y	y	y	y	phosphomevalonate kinase
Isoprenoid/Mevalonate biosynth	SAOUHSC_01618	ID=SAPIG1587	y	n	n		geranyltranstransferase
	?	ID=SAPIG1588	y				xseB
Isoprenoid/Mevalonate biosynth	SAOUHSC_02623	ID=SAPIG2398	y	y	y	n	fni
Isoprenoid/Mevalonate biosynth	SAOUHSC_02859	ID=SAPIG2595	y	y	n	y	hydroxymethylglutaryl-CoA reductase, degradative
Isoprenoid/Mevalonate biosynth	SAOUHSC_02860	ID=SAPIG2596	y	y	n	y	hydroxymethylglutaryl-CoA synthase
Menaquinone biosynthesis	SAOUHSC_00980	ID=SAPIG1039	y	n	y	y	menA
Menaquinone biosynthesis	SAOUHSC_00983	ID=SAPIG1041	n	n	n		menD
Menaquinone biosynthesis	SAOUHSC_00985	ID=SAPIG1043	y	n	y		menB
Menaquinone biosynthesis	SAOUHSC_01486	ID=SAPIG1468	y	y	y		heptaprenyl diphosphate synthase component II
Menaquinone biosynthesis	SAOUHSC_01487	ID=SAPIG1469	y	n	y		menaquinone biosynthesis methyltransferase UbiE
Menaquinone biosynthesis	SAOUHSC_01488	ID=SAPIG1470	n	n	y		hyothetical protein
Menaquinone biosynthesis	SAOUHSC_01915	ID=SAPIG1842	y	n	n		menC

Menaquinone biosynthesis	SAOUHSC_01916	ID=SAPIG1844	y	n	y		menE
Thioredoxin	SAOUHSC_00785	ID=SAPIG0840	y	y	y	y	trxB
Thioredoxin	SAOUHSC_01100	ID=SAPIG1141	y	n	y	y	trx
Nucleotides							
Purine biosynthesis	SAOUHSC_00374	ID=SAPIG0467	y	n	n		guaB
Purine biosynthesis	SAOUHSC_00375	ID=SAPIG0468	y	y	n	y	GMP synthase
Purine biosynthesis	SAOUHSC_00485	ID=SAPIG0576	y	n	y		hpt
Purine biosynthesis	SAOUHSC_01176	ID=SAPIG1207	y	y	y	y	gmk
Purine biosynthesis	SAOUHSC_02490	ID=SAPIG2282	y	y	y	y	adenylate kinase
Purine metabolism	SAOUHSC_01742	ID=SAPIG1697	y	y	y	n	GTP pyrophosphokinase
Purine/Pyrimidine biosynthesis	SAOUHSC_00741	ID=SAPIG0807	y	y	y	y	nrdI
Purine/Pyrimidine biosynthesis	SAOUHSC_00742	ID=SAPIG0808	y	y	y	y	ribonucleoside-diphosphate reductase, alpha subunit
Purine/Pyrimidine biosynthesis	SAOUHSC_00743	ID=SAPIG0809	y	y	y	y	ribonucleoside-diphosphate reductase, beta subunit
Pyrimidine biosynthesis	SAOUHSC_00451	ID=SAPIG0548	y	y	y	y	tmk
Pyrimidine biosynthesis	SAOUHSC_01235	ID=SAPIG1260	y	y	y	y	pyrH
Pyrimidine biosynthesis	SAOUHSC_01435	ID=SAPIG1428	y	y	y	y	thyA
Pyrimidine biosynthesis	SAOUHSC_01496	ID=SAPIG1476	y	y	n	y	cmk
Pyrimidine biosynthesis	SAOUHSC_02368	ID=SAPIG2169	y	y	y	y	pyrG
Cofactors							
Acetyl CoA/CoA	SAOUHSC_00574	ID=SAPIG0662	y	y	n	y	pta
Acetyl CoA/CoA	SAOUHSC_01075	ID=SAPIG1121	y	y	y	y	coaD
Acetyl CoA/CoA	SAOUHSC_01178	ID=SAPIG1209	y	y	y	y	coaBC
Acetyl CoA/CoA	SAOUHSC_01795	ID=SAPIG1742	y	n	y	n	coaE
Acetyl CoA/CoA	SAOUHSC_02371	ID=SAPIG2172	y	y	y	y	pantothenate kinase
Folate	SAOUHSC_00490	ID=SAPIG0581	y	y	y	y	folB

Folate	SAOUHSC_00549	ID=SAPIG0641	y	y	y	n	conserved hypothetical protein
Folate	SAOUHSC_01007	ID=SAPIG1061	y	n	y		bifunctional protein FolD
Folate	SAOUHSC_01434	ID=SAPIG1427	y	y	y	y	folA
Folate	SAOUHSC_01766	ID=SAPIG1717	y	y	y	y	folylpolyglutamate synthase
Folate	SAOUHSC_02354	ID=SAPIG2155	n	n	n		glyA
Folate	SAOUHSC_00491	ID=SAPIG0582	y	n	n	y	folK
Folate	SAOUHSC_00489	ID=SAPIG0580	y	n	n	y	folP
NAD biosynthesis	SAOUHSC_00943	ID=SAPIG1002	y	n	y	y	ATP-NAD kinase
NAD biosynthesis	SAOUHSC_01697	ID=SAPIG1659	y	n	y	y	nadD
NAD biosynthesis	SAOUHSC_02132	ID=SAPIG2007	y	y	y	y	nadE
NAD biosynthesis	SAOUHSC_02133	ID=SAPIG2008	y	y	y	y	pncB
SAM	SAOUHSC_01909	ID=SAPIG1838	y	y	y	y	metK
Fe-sulpate cluster	SAOUHSC_00847	ID=SAPIG0897	y	y	y	y	FeS assembly ATPase SufC
Fe-sulpate cluster	SAOUHSC_00848	ID=SAPIG0898	y	y	y	y	FeS assembly protein SufD
Fe-sulpate cluster	SAOUHSC_00849	ID=SAPIG0899	y	y	y	y	cysteine desulfurase
Fe-sulpate cluster	SAOUHSC_00850	ID=SAPIG0900	y	y	n	y	SUF system FeS assembly protein, NifU family
Fe-sulpate cluster	SAOUHSC_00851	ID=SAPIG0901	y	y	y	y	FeS assembly protein SufB
Fe-sulpate cluster	SAOUHSC_01727	ID=SAPIG1686	y	y	y	y	cysteine desulfurase
Fe-sulpate cluster	SAOUHSC_01504	ID=SAPIG1483	y	n	n	y	ferrodoxin
Riboflavin biosynthesis	SAOUHSC_01249	ID=SAPIG1274	y	y	y	y	ribF
Other/Unknown							
Amino acid transporter	SAOUHSC_01787	ID=SAPIG1736	y	y	y	n	lysine-specific permease
GTP binding	SAOUHSC_01214	ID=SAPIG1243	y	n	y	y	GTPase family
GTP binding	SAOUHSC_01668	ID=SAPIG1632	y	n	y	y	era
GTP binding	SAOUHSC_01753	ID=SAPIG1473	y	y	y	y	engA
GTP binding	SAOUHSC_01777	ID=SAPIG1728	y	y	y	n	EngB

GTP binding	SAOUHSC_01700	ID=SAPIG1662	y	y	y	y	GTPase family protein
GTP binding	SAOUHSC_01492	?		y	y	y	conserved hypothetical protein
Other	SAOUHSC_00510	ID=SAPIG0604	y	y	y	y	cysE
Other	SAOUHSC_02140	ID=SAPIG2014	y	y	y	y	manganese-dependent inorganic pyrophosphatase
Other	SAOUHSC_01416	ID=SAPIG1414	n	n	n		sucB
Other	SAOUHSC_01040	ID=SAPIG1090	y	n	y	n	pdhA (alpha subunit)
		ID=SAPIG1091	y				pdhB (beta subunit)
Other	SAOUHSC_02277	ID=SAPIG2088	y	y	y	y	O-sialoglycoprotein endopeptidase
Unknown	SAOUHSC_00015	ID=SAPIG0014	y	y	y	n	DHH subfamily 1 protein
Unknown	SAOUHSC_00226	ID=SAPIG0269	y	y	y	y	alcohol dehydrogenase
Unknown	SAOUHSC_00728	ID=SAPIG0796	y	y	y	n	anion-binding protein
Unknown	SAOUHSC_00760	ID=SAPIG0822	y	y	n	n	conserved hypothetical protein
Unknown	SAOUHSC_A01041	?		y		y	hypothetical protein
Unknown	SAOUHSC_01188	ID=SAPIG1220	y	y	n	n	ribosome small subunit-dependent GTPase A
Unknown	SAOUHSC_01263	ID=SAPIG1288	y	n	y		conserved hypothetical protein
Unknown	SAOUHSC_01350	ID=SAPIG1355	y	y	y	y	conserved hypothetical protein
Unknown	SAOUHSC_01477	ID=SAPIG1460	y	y	y	y	Zn-dependent protease
Unknown	SAOUHSC_01661	ID=SAPIG1625	y	y	y	y	Bcl-2 family protein
Unknown	SAOUHSC_01701	ID=SAPIG1663	y	y	y	y	yqeG
Unknown		ID=SAPIG1664	y				mtnN
Unknown	SAOUHSC_01782	ID=SAPIG1732	y	y	n	y	MutT/nudix family protein
Unknown	SAOUHSC_01871	ID=SAPIG1807	y	y	y	y	polysaccharide biosynthesis protein
Unknown	SAOUHSC_01908	ID=SAPIG1837	n	y	y	y	conserved hypothetical protein
Unknown	SAOUHSC_01979	ID=SAPIG1913	y	y	y	n	helix-turn-helix domain protein
Unknown	SAOUHSC_02106	ID=SAPIG1984	y	y	y	y	conserved hypothetical protein
Unknown	SAOUHSC_02151	ID=SAPIG2051	n	y	y	y	membrane protein, putative
Unknown	SAOUHSC_02152	ID=SAPIG2052	n	y	y	n	ABC transporter
Unknown	SAOUHSC_02279	ID=SAPIG2088	y	y	y	y	O-sialoglycoprotein endopeptidase
Unknown	SAOUHSC_02280	ID=SAPIG2090	y	y	y	y	peptidase M22, glycoprotease
Unknown	SAOUHSC_02357	ID=SAPIG2159	y	y	y	y	hemK

Unknown	SAOUHSC_02407	ID=SAPIG2218	y	y	y	y	conserved hypothetical protein
Unknown	SAOUHSC_02571	ID=SAPIG2594	n	y	y	n	secretory antigen SsaA
Unknown	SAOUHSC_00003	ID=SAPIG0003	y	n	y	y	S4 domain protein YaaA
Unknown	SAOUHSC_00444	ID=SAPIG0545	y	n	y	y	conserved hypothetical protein
Unknown	SAOUHSC_00575	ID=SAPIG0663	y	n	y	y	lipoate-protein ligase A protein
Unknown	SAOUHSC_00793	ID=SAPIG0849	n	n	n	n	conserved hypothetical protein
Unknown	SAOUHSC_00868	ID=SAPIG0914	y	n	y		conserved hypothetical protein
Unknown	SAOUHSC_00892	ID=SAPIG0938	y	n	y	y	general stress protein 13
Unknown	SAOUHSC_00922	ID=SAPIG0981	n	n	n		conserved hypothetical protein
Unknown	SAOUHSC_00957	ID=SAPIG1017	y	y	y	y	toxic anion resistance protein
Unknown	SAOUHSC_01036	ID=SAPIG1087	y	y	y	y	conserved domain protein
Unknown	SAOUHSC_01050	ID=SAPIG1100	n	n	y	n	conserved hypothetical protein
Unknown	SAOUHSC_01077	ID=SAPIG1123	y	n	n	y	conserved hypothetical protein
Unknown	SAOUHSC_01119	ID=SAPIG1156	n	n	y		conserved hypothetical protein
Unknown	SAOUHSC_01190	ID=SAPIG1222	y	n	y	n	thiamine pyrophosphokinase
Unknown	SAOUHSC_01244	ID=SAPIG1269	y	y	y	y	conserved hypothetical protein
Unknown	SAOUHSC_01245	ID=SAPIG1270	y	y	y	y	ribosomal protein L7AE family
Unknown	SAOUHSC_01627	ID=SAPIG1596	y	n	y	n	lipoprotein, putative
Unknown	SAOUHSC_01672	ID=SAPIG1635	y	y	y	y	conserved hypothetical protein
Unknown	SAOUHSC_01721	ID=SAPIG1681	n	n	n	y	protein Stu1959
Unknown	SAOUHSC_01756	ID=SAPIG1709	y	y	y	y	conserved hypothetical protein
Unknown	SAOUHSC_01770	ID=SAPIG1721	n	n	y		hypothetical protein
Unknown	SAOUHSC_01866	ID=SAPIG1802	y	n	y	n	aminoglycoside phosphotransferase family protein
Unknown	SAOUHSC_01928	ID=SAPIG1876	n	n	y	y	transposase
Unknown	SAOUHSC_01930	ID=SAPIG1881	n	n	n	n	probable exported protein
Unknown	SAOUHSC_02757	?		n	n	n	conserved hypothetical protein
Unknown	SAOUHSC_02805	ID=SAPIG2554	y	y	y		conserved hypothetical protein
Unknown	SAOUHSC_00345	ID=SAPIG0442	y	n	y	y	conserved domain protein
Unknown	SAOUHSC_01362	ID=SAPIG1364	y	n	y	n	conserved domain protein
Unknown	SAOUHSC_02260	ID=SAPIG2072	y	n	y		hypothetical protein
Unknown	SAOUHSC_02572	?		n	n		conserved hypothetical protein
Unknown	SAOUHSC_02575	ID=SAPIG2353	n	n	y	n	conserved hypothetical protein

Unknown	SAOUHSC_02720	?	n	conserved hypothetical protein

Essential/advantegeous in S. aureus (y)
Non-essential in S. aureus (n)
Inconsistancy between different S. aureus strains and/or methods

Title: Genes important for survival of livestock-associated methicillin-resistant *Staphylococcus aureus* Sequence Type 398 in the porcine reservoir.

Running title: LA-MRSA ST398 survival in the pig reservoir.

Keywords: LA-MRSA ST398, livestock, porcine reservoir, high-throughput methods, adhesion and survival.

Authors: Mette T. Christiansen[a (#)], J. Inhong Wang[b], Dan Tucker[b], Henrik Hasman[a] and Frank M. Aarestrup[a (#)]

National Food Institute, Technical University of Denmark, DK[a]; Department of Veterinary Medicine, University of Cambridge, UK[b]

Corresponding authors: Mette T. Christiansen email: mtso@food.dtu.dk and Frank M. Aarestrup fmaa@food.dtu.dk

Abstract

Staphylococcus aureus is an important opportunistic pathogen that colonizes the upper respiratory tract and the skin surface of several animal species, including humans. Comparative genomic studies have identified a few phage associated genes that appears to be correlated with virulence in humans, but have not been able to identify genes of importance for successful colonization or infection in livestock or other animals. The porcine reservoir is important for the spread of *S. aureus* sequence type 398 (ST398) and the identification of genes important for survival of ST398 in pigs could contribute to a better understanding of transmission and adaptation. In this study we screened a transposon mutant library consisting of approximately one million livestock-associated methicillin-resistant *S. aureus* (LA-MRSA) ST398 mutants to identify genes important for porcine survival. Seventeen genes were identified as important for porcine skin adhesion and survival. Ten genes represent mutants with reduced fitness and they primarily encode transporters and enzymes involved in metabolic pathways. In addition four mutants with increased fitness were identified and they encode DNA binding proteins involved in regulation. Sixteen genes were identified as important for nasal epithelial survival, encoding proteins involved in regulation, metabolic enzymes, cell wall components and hypothetical proteins. The genes identified here can constitute targets for MRSA decolonization in pigs, which could prevent further spread of the ST398 linage. Additional investigations into the specific function of the genes identified in this study as important for porcine survival are needed.

Introduction

Staphylococcus aureus is an important opportunistic pathogen that colonizes the upper respiratory tract and the skin surface of several animal species, including humans (1–3). During the past decade a livestock-associated methicillin-resistant *Staphylococcus aureus* (LA-MRSA) linage, belonging to clonal complex 398 (CC398), has become of increasing concern. CC398 is the predominant clonal complex in pigs (4). It has been suggested that sequence type 398, belonging to CC398, originated as MSSA in humans and from there transmitted to livestock, where it acquired mobile genetic elements like SCC*mec* and Tn916 facilitating methicillin and tetracycline resistance and in addition lost a prophage, carrying genes potentially important for survival in the human host (5).

Pigs constitute a large reservoir for LA-MRSA ST398 and contribute to an ongoing spread and genetic adaptation. However the bacterial mechanisms underlying successful colonization and survival in pigs are poorly understood. To better understand the adaptation and interspecies transmission potential of LA-MRSA ST398, genes important for porcine survival needs to be determined.

Previous studies have applied porcine nasal epithelial cells, porcine and human skin corneocytes and keratinocytes, porcine nasal mucosa explants and live pigs to study *S. aureus* nasal and skin colonization (1, 6–8). All studies used wild type strains or single knockout mutants. Corrigan *et al.* (2009) concluded that the ability of *S. aureus* to adhere to human desquamated nasal epithelial cells was multifactorial and involved the serine-aspartic acid repeat proteins SdrC and SdrD as well as iron regulated surface determine protein A (IsdA) and clumping factor B (ClfB) (6). IsdA and clfB have also been suggested as important for nasal adhesion in other studies (9–11). Tulinski *et al.* (2013) demonstrated that a beta-toxin (hlb) *S. aureus* ST398 mutant, showing a different hemolysis pattern, had reduced colonization properties to porcine nasal epithelial explants compared to wild type (1).

Different *S. aureus* clonal lineages show different adhesion patterns. Some lineages demonstrate preferred adhesion to corneocytes isolated from pigs and some to skin corneocytes isolated from humans. These patterns were confirmed by *in vivo* colonization experiments in piglets (12). LA-MRSA ST398 did not show preferred binding to corneocytes from either pigs or humans (7), but human associated methicillin-sensitive *S. aureus* (MSSA) ST398 showed enhanced adhesion to human isolated skin keratinocytes and keratin (8).

Mutants applied in previous adhesion and colonization studies were generated based on previous knowledge about *S. aureus*, but this approach cannot help to highlight other potential gene candidates, which has not previously been associated with *S. aureus* colonization. In this study we use a transposon mutant library consisting of approximately one million LA-MRSA ST398 mutants to identify genes important for porcine survival. The mutant library was generated and previously verified by us (manuscript I) and was used in a comprehensive screening of genes important for adhesion to and survival on porcine skin explants and survival on porcine nasal epithelial tissue.

Materials and methods

Bacterial strains and culture conditions. A mariner transposon mutant library was generated in a previous study in the whole genome sequenced LA-MRSA ST398 isolate S0385 (manuscript I). The transposon mutant library consisted of approximately 1 million mutants with around 140,000 unique insertion sites and the average number of unique inserts per gene was calculated to 44.8. $>10^6$ mutant cells from frozen aliquots were inoculated into BHI broth (Oxoid) supplemented with 5 mg/l erythromycin (Sigma) and incubated at 37°C with aeration overnight. Mutants were harvested, washed twice in phosphate buffered saline (PBS) and re-suspended in PBS. To obtain mutant from exponential growth phase, 2 ml from the o/n culture was re-inoculated into fresh BHI supplemented with 5 mg/l erythromycin and

grown to mid-exponential phase OD_{600} 0.5 before the cells were harvested, washed and re-suspended in PBS.

Ethics statement. The study protocol was submitted to the ethical review committee at the University of Cambridge, Department of Veterinary Medicine, who reported that post mortem collection of tissue following the slaughter of male pigs, surplus to a breeding program, is not a regulated procedure and provided ethical approval. The UK Animals (Scientific Procedures) Act 1986 allows for the use of animal tissues and blood in research that come from animals not regulated by the Act. These animals were slaughtered by a method of killing identified in Schedule 1 of the Act. In this case, two 6-month-old pigs, a male (Pig_1) and a female (Pig_2), were collected at different days with two weeks in between. They were euthanized by intravenous overdose of pentobarbitone and the tissue was collected immediately postmortem after obtaining the farm owner's permission for the use of their pigs in this study.

Preparation of porcine skin. The pig skin was prepared as described previously (13). The skin areas behind the ears were washed with chlorhexidine soap and disinfected with 70 % ethanol before epilation with a sterile razor. A squared skin piece of around 8 x 8 cm was removed from the pigs and the adipose tissue beneath the dermis was removed with a scalpel. The skin was dissected under sterile conditions into 2 cm^2 pieces, placed in 6-well plates (NUNC) and embedded in Hepes agar (145 mM NaCl, 5 mM KCl, 1 mM $MgSO_4$, 10 mM Hepes, 10 mM glucose, 5 % Agarose) leaving the skin surface uncovered (see Figure 1). The skin pieces were disinfected with 70 % ethanol for 5 min at room temperature followed by washing with PBS three times. Swabs were taken from the washed skin surface, streaked on blood agar plates and incubated overnight at 37°C to test for surface contamination.

Ex vivo porcine skin survival. Genomic DNA (gDNA) was extracted from a transposon mutant overnight culture (~10^9 cells) using MasterPure Gram Positive Purification Kit (Epicentre) and stored at -20°C as input pool. 10 µl of up-concentrated mutant culture (~10^{11} cells) were inoculated onto the porcine skin surface and incubated under atmospheric conditions at 32°C for ~24 or ~48 hours (duplicates were generated for each incubation period from Pig_1 and Pig_2). After incubation the skin explants were homogenized (with a ball bearing and 2 x 5 min, 20 Hz) in 1 ml PBS. 9 x 10^7 - 2.5 x 10^8 CFU/ml was recovered after ~24 hours and 2.1 x 10^8 – 4.1 x 10^8 CFU/ml were recovered after ~48 hours incubation on the skin explants. The cell suspensions from each tissue explants were re-inoculated into 10 ml fresh BHI supplemented 5 mg/l erythromycin (to select for transposon mutant and reduce growth of the natural porcine skin microbiota) and incubated overnight at 37°C with aeration. From the overnight cultures gDNA was extracted from ~10^9 cells and stored at -20°C as output pools.

Ex vivo porcine skin adhesion. *S. aureus* expresses different surface proteins depending on growth phase (3) and therefore both exponentially and stationary grown cell were used in the skin adhesion assay. The exponentially grown cells were harvested at OD_{600} 0.5-0.8. gDNA was extracted from an exponentially and stationary grown transposon mutant culture (~10^9 cells from each growth phase) and stored at -20°C as input pools. 10 µl of up-concentrated exponentially grown and stationary grown cells (~10^{11} from each growth phase) were inoculated onto the porcine skin surface and incubated under atmospheric conditions at 32°C for ~20 hours (four replicates for each growth phase were performed on tissue from Pig_1). After the tissue pieces were inoculated they were placed into a clean and sterile Eppendorf tube with 1 ml PBS and washed once by vigorously vortexing for 5 seconds (to remove

loosely adhered/attached bacterial cells). The washed tissue was transferred to a clean and sterile Eppendorf tube with 1 ml of 0.1 % Triton-X (to facilitate detachment of adhered bacterial cells) and the skin tissue was homogenized (with a ball bearing and 2 x 5 min, 20 Hz). 3 x 10^7 – 8 x 10^7 CFU/ml were recovered of the stationary cells and 1.1 x 10^7 – 1.5 x 10^7 CFU/ml were recovered of the exponential cells. The cell suspensions from each tissue explants were re-inoculated into 10 ml fresh BHI supplemented 5 mg/l erythromycin (to select for transposon mutant and reduce growth of the natural porcine skin microbiota) and incubated overnight at 37°C with aeration. From the overnight cultures gDNA was extracted from ~10^9 cells and stored at -20°C as output pools.

Preparation of nasal epithelial tissue. For isolation of nasal epithelial tissue, the pig head was removed from the carcass and immediately used for isolation of the nasal septum, leaving the lining nasal epithelial tissue intact. The tissue was washed in Dulbecco's Modified Eagle Medium (DMEM - Sigma) supplemented with 2 µg/ml enrofloxacin, 50 µg/ml streptomycin, 100 U/ml penicillin, and 2.5 µg/ml Fungizone for 15 min. at 37°C at 80 rpm followed by a 2 hours wash in DMEM with 2 µg/ml enrofloxacin at the same incubation conditions. The antibiotic wash was followed by an antibiotic free wash using 500 ml DMEM 2 x 15 min. plus 4 x 30 min. in 250 ml DMEM (changing to fresh media 6 times) at 37°C at 80 rpm. After washing, the tissue was kept in 50:50 ratio of DMEM and Roswell Park Memorial Institute medium (RPMI) (Sigma) supplemented with 50µM glutamine. The nasal epithelium was dissected from the underlying cartilage of the nasal septum and divided into pieces of approximate 0.5 x 0.5 cm^2, in a sterile environment. Antibiotic residual test was performed on a bacterial lawn of LA-MRSA ST398 S0385. The tissue pieces were placed on filter-paper overlying agar-plugs with the external side facing up-wards. The agar-plugs were

arranged in 6-Well plates (NUNC) with a DMEM reservoir, moistening the filter paper and in that way nourishing the tissue (see Figure 1).

Porcine nasal epithelial survival. Genomic DNA was extracted from a transposon mutant overnight culture (~10^9 cells) and stored at -20°C as input pool. Five to ten µl of up-concentrated mutants (~10^{11} cells) were inoculated onto the prepared nasal epithelium and incubated at 37°C plus 5 % CO_2 for ~24 hours (duplicates from Pig_1 and Pig_2). After incubation the epithelial tissue was homogenized (with a ball bearing and 2 x 5 min, 20 Hz) in 1 ml PBS. 2.7 x 10^8 – 4.2 x 10^{10} CFU/ml was recovered after ~24 hours incubation on the nasal epithelium explants. The cell suspensions were re-inoculated into 10 ml fresh BHI supplemented 5 mg/l erythromycin and incubated overnight at 37°C with aeration. From the overnight cultures gDNA was extracted from ~10^9 cells and stored at -20°C as output pools.

Library preparation for Illumina sequencing. The approach, Transposon directed insertion-site sequencing (TraDIS) described previously (14) was used for identification of genomic transposon insertions sites. The library preparations were performed as described in manuscript I. The libraries were pooled in a 1:1 molar ratio with 7 or 8 samples per flow cell lane. The samples were sequenced on an Illumina Hiseq2000 platform for 43 cycles plus index read using a custom sequencing primer (5-GACACTATAGAAGAGACCGGGGACTTATCAGC-3) resulting in reads with 10 transposon insert specific nucleotides (Tn sequence) followed by the junction region.

Sequence analysis and statistics. Sequence reads from the Illumina FASTQ files were sorted by index and by using the program Sabre (https://github.com/najoshi/sabre), evaluated for the Tn sequence (CAACCTGTTA) allowing 1 mismatch. The Tn sequence and adapter

sequences were stripped using Cutadapt (15) in addition to short reads (<10 nucleotides) and nucleotides with poor base call quality (<Q15). The junction regions were extracted and mapped to the reference genome (accession no. AM990992) using Bowtie 2.0 (16).

The number of reads corresponding to each transposon insertion site in the input pools was compared to the number of reads mapping to the equivalent position in the output pools using the DESeq package in R (17, 18). The read counts corresponding to transposon insertion sites were normalized to account for variation in the total number of reads obtained from each samples. The ratio of input:output reads counts were determined and referred to as a $\log_2$ fold change, which will be referred to as a fitness score. A negative fitness score reflected an attenuated mutant. An attenuated mutant was determined when the number of read counts from input pool to output pool decreased and thereby illustrated a decrease in mutant clones after selection. For strongly attenuated mutants zero clones will be present in the output pools and the $\log_2$ fold change was defined as minus infinity and a fitness-score of -12 was assigned to such mutants. Like done by Chaudhuri *et al.* (2013) for each individual mutant, the hypothesis that the fitness score was equal to zero and thereby that the mutant was present at equivalent levels in the input and output pools was tested for using a negative binomial distribution as implemented in DESeq (19). DESeq models variance under the assumption that the mutants with comparable levels of sequence coverage exhibit similar levels of dispersion. The model was fitted only from those mutants from which replicate data was available which was in this case primarily sequence read counts from output pools, as no biological replicates were available from input pools. The resultant model was then applied to data derived from all mutants to estimate *P* values.

Results

Porcine skin adhesion and survival. An *ex vivo* porcine skin model was generated using freshly isolated porcine skin from the rear of the ears of two different pigs. The skin surface was washed and disinfected before inoculation to remove dirt and the surface associated natural microbiota. Skin swabs were taken to test the sterilization approach and all the tissue samples tested negative for surface contamination. The transposon mutant library was screened in the porcine skin model and the skin samples were incubated for approximately 20 (adhesion assay), 24 (survival assay) or 48 (survival assay) hours. DNA was isolated from the transposon mutant library input sample and from output samples. The mutant compositions in the input and output pools were quantified and compared based on number of sequence reads mapping to open reading frames encoded in the reference genome.

When the transposon mutant library was selected on porcine skin explants a decrease in cell counts (from ~10^{11} to an average of ~10^8 CFU/ml) were observed suggesting an initial selection on the mutant pool. In the skin survival assay a slight increase in cell counts were observed between 24 and 48 hours incubation (from an average of ~2 x 10^8 to ~3 x 10^8 CFU/ml) which propose that the mutants that are present on the skin explants are viable. A decrease in cell counts was observed between the stationary cells recovered in the adhesion assay (an average of ~5.4 x 10^7 CFU/ml) compared to the cells recovered in the survival assay after 24 hours incubation (an average of ~2 x 10^8 CFU/ml). This indicates that some mutants were lost in the washing step preformed in the adhesion assay. A lower number of mutants were recovered after porcine skin adhesion with exponential cells compared to stationary cells (an average of ~1.3 x 10^7 and ~5.4 x 10^7 CFU/ml respectively), which could point to that the stationary grown mutants adhere better to the porcine skin explants.

In the porcine skin survival assay 27 genes were identified to be associated with alteration in fitness and therefore defined as important for LA-MRSA ST398 isolate S0385 survival on

porcine skin. The genes represent mutants that had a significant (*P* level ≤0.01) change in fitness when screened on skin explants isolated from both pigs (two replicates from Pig_1 and Pig_2). The genes are listed in Table S2 in supplementary materials. Twenty-two mutants were identified as attenuated whereas five mutants were hypercompetitive within the specific environment. Fourteen of the attenuated mutants also showed a reduction in fitness when grown under laboratory conditions (genes important for LA-MRSA ST398 isolate S0385 survival under laboratory conditions can be found in manuscript I supplementary Table S1 and S2).

The transposon mutant library was also screened in a porcine skin adhesion assay to identify genes that were important for skin surface attachment. The mutant pool was screened both in the stationary and exponential growth phase as it is known that *S. aureus* displays a different set of surface proteins in the different growth phases. As only the genes specifically important for skin attachment were of interest, genes identified as important for survival under laboratory conditions (data not shown) and survival in porcine skin assay were removed. Only the genes representing mutants with a significant (P level ≤ 0.05) reduction in fitness were selected. Sixty-eight genes were identified as representing mutants with reduced fitness when the transposon mutant library was selected in the adhesion assay as stationary grown cells (see Table S2). These genes correspond to genes encoding adhesion factors displayed by *S. aureus* when grown to a stationary phase and screened in the porcine skin adhesion model in the study. Twenty-nine genes representing mutants with attenuated fitness were identified based on the same criteria as stated above and when screening the transposon mutant library as exponentially grown culture in the adhesion assay (see Table S3).

When comparing the list of genes obtained in the porcine skin survival and adhesion assay screened with stationary grown cells, eight genes were evaluated as important for both

adhesion and survival in the skin model and of these six genes represent attenuated mutants and two genes hypercompetitive mutants (see Table 1). Eight genes were identified as important for adhesion in the exponential growth phase and survival in the porcine skin model, two of which showed hypercompetitive mutants with increased fitness in both assays and six attenuated mutants with reduced fitness in both assays (see Table 2). Three genes showed inconsistency within the two assays. However the genes were only identified with fitness changes in skin survival assay for Pig_1 after 1 day of incubation.

Porcine nasal epithelial survival. Another *ex vivo* model based on porcine nasal epithelial tissue was generated to screen for *S. aureus* genes important for nasal survival. The tissue was collected from two different pigs and was washed extensively with antibiotics to remove the natural bacterial microbiota. The absence of residual antibiotics in the tissue was confirmed. The transposon mutant library was screened on the porcine nasal explants for 1 day and DNA was isolated from input and output samples. The mutant composition in the input and output pools were quantified and compared.

When the transposon mutant library was selected on porcine nasal epithelium explants a decrease in cell counts (from ~10^{11} to an average of ~10^{9} CFU/ml) were observed like in the porcine skin model, suggesting an initial selection on the mutants.

Four genes with specific importance for nasal epithelium survival were found in this study, two of which showed decrease in fitness and two with increased fitness. Table 3 shows the genes that were identified with a significant change in fitness score (P level ≤ 0.05) on both pigs.

Discussion

The purpose of this study was to identify genetic factors that are important for LA-MRSA ST398 survival in the pig reservoir. *S. aureus* porcine colonization studies have been performed previously using different wild type *S. aureus* strains or mutants that were generated based on already know colonization factors (1, 6–8). These types of studies will however not be able to identify unknown colonization factors utilized by *S. aureus* and will not give an estimate of the relative importance of the different genes. In this study, a genome-wide screening of a previously generated transposon mutant library in the LA-MRSA ST398 S0398 isolate, was performed in an *ex vivo* porcine skin and nasal epithelial model. The models were based on freshly isolated porcine tissue to mimic an *in vivo* environment where host factors play important parts for bacterial attachment and persistence (3). Such models can be studied under controlled conditions and requires fewer pigs to be sacrificed compared to *in vivo* studies. The combination of *ex vivo* models and high complexity transposon mutant libraries constitutes strong screening tools for identification of unknown genetic factors important for bacterial survival in various environments.

Some consistency was found between the genes identified as important for porcine nasal epithelial survival when comparing the results obtained from the two pigs. It is know that many host factors are involved in *S. aureus* colonization because only around 20-40% of the human population are persistent carries of *S. aureus* (3, 20). The differences seen between the two pigs used in this study could be related to genetic variation, immune status of the host, gender or simple differences between pig replicates when using this model system. To get more conclusive data the screen should have been repeated on more explants isolated from other pigs but this was unfortunately not possible in this study. However, as the genes presented here only illustrate genes which were identified as important for survival on several replicate explants isolated from both pigs, they should be considered as genes relevant for

survival in the porcine reservoir. They constitute good gene candidates for generation of single knockout mutants which should be tested within the same assays for a complete definition of gene essentiality.

Genes identified with a significant change in fitness when screened in porcine skin survival model on tissue samples from both pigs are illustrated in Table S1. Some of the genes identified as important for skin survival were also identified as important for survival under laboratory conditions. Eight genes representing mutant with a reduced fitness score were defined as important for porcine skin survival only and they are described as hypothetical proteins, regulators and transporters mainly. The S0385 strain contains 3 circular plasmids (21) and after two days incubation on the porcine skin explants, mutants with transposon insert into the replication protein Rep located in plasmid 3 (PSAPIG030001) showed a drop in fitness. The plasmid is annotated to encode two different genes, the replication protein and a transcriptional regulator (SAPIG030002), one of which might be important for porcine skin survival. In addition five genes representing hypercompetitive mutants were defined as important for porcine skin survival only. Two of which encode a reductase, one phage integrase and two repressors. These functions might not be important in porcine skin survival *ex vivo* but could be essential in other more natural environments were competition and selection, are important factors for bacterial survival.

The adhesion assays identified a large number of attenuated mutants which are presented in supplementary Table S2 and Table S3. Overall only smaller changes in fitness score were identified in the adhesion assays which could be due to low selection pressure in these specific models. The washing to remove non-adherent or loosely attached cells was only performed once and repetition of this step could increase the selective pressure. Generally

various enzymatic encoding genes and genes encoding secreted protein and surface proteins represented the mutants with the most profound loss in fitness in the skin adhesion assays. Clumping factor B (ClfB) and another fibrinogen-binding protein (SAPIG1154) were evaluated as important for skin adhesion when screening stationary grown cells. ClfB has previous been evaluated to be involved in human nasal adhesion and carriage (6). ClfB is predominantly expressed in the exponential growth phase and clumping factor A (ClfA) is mainly expressed on the surface of cells from the stationary growth phase (3). In this study *clfB* mutants are identified with attenuated fitness when screening stationary grown cells for skin adhesion *in vitro* and *clfA* (SAPIG0866) was not identified with attenuated fitness. The cells were incubated for ~24 hours on the skin surface before washing. During incubation, the mutants could have continued to grow at an unknown rate, resulting in a switch from stationary to exponential growth at some point.

Immunoglobulin G binding protein A and staphylococcal secretory antigen ssA1 and ssA2 were identified as important for skin adhesion. In addition *cap5A* and *cap5D* (both involved in capsular polysaccharide biosynthesis) mutants were identified with attenuated fitness in the skin adhesion assay using exponentially grown cells. Protein A and capsular polysaccharide inhibit phagocytosis (3) and the staphylococcal secretory antigens have predicted immunogenic function. This indicates that immune evasion and modulation are important features for the initial *S. aureus* ST398 colonization of porcine skin.

Genes with either a fitness reduction or increase (negative or positive fitness score) in both the porcine skin adhesion and survival models are defined as the genes of interest (see Table 1 and 2). The genes were selected if they showed a significant reduction or increase in read counts at the *P* level ≤ 0.01 in both the adhesion and the survival assay after 1 or 2 days of infection. Table 3 shows an overview and descriptions of the genes evaluated as important for

porcine skin adhesion and survival. SAPIG0737 and SAPIG0740, encoding a DNA-binding response regulator and an ABC transporter respectively, represent mutants with significant attenuated fitness in the adhesion assay, using either stationary or exponentially grown cells, and the skin survival assay after 1 and 2 days of infection. In addition SAPIG0739, encoding the export ATP-binding protein BceA known to be involved in *S. aureus* infection, showed a reduced fitness in the adhesion assay screened with a stationary culture and in the skin survival assay. These three proteins are encoded just next to each other in the S0385 genome with just one gene, SAPIG0738, in between them. BceA (SAPIG0739) and BceB (SAPIG0740) make up an ABC transporter whereas the BceR (SAPIG0737) and BceS (SAPIG0738) are the regulatory and sensing part located just upstream of the transporter genes (22). BceAB transporter shows similarity with *Bacillus subtilis* ABC transporter with the same annotation and was previously defined as responsible for bacitracin efflux in Bacillus (23). Bacitracin is a polypeptide antibiotic produced by *B. subtilis* and *B. licheniformis* (24, 25) and these polypeptides disrupt cell wall and peptidoglycan synthesis in Gram positive and Negative bacteria. It has been shown previously that mutation in *bce*RS and *bce*AB reduced the resistance to bacitracin and in addition inactivation of *bce*AB reduced oxacillin resistance slightly, indicating that the ABC transporter might be involved in cell wall biosynthesis (22). As *bce*S (SAPIG0738) mutants were not identified as relevant for porcine skin adhesion and survival and as the model environment was supposedly bacitracin free, this ABC transporter system may have other functions relating to the survival on porcine skin. In general ABC transporters constitute a large family of membrane transporters contributing to import and export of various substances such as proteins, peptides, polysaccharides and antibiotics (26). Therefore the specific bceAB transporter system identified here as important for porcine skin survival could have several functions relevant for bacterial survival. Another gene that showed importance for adhesion and survival on

porcine skin, though only evaluated as important for survival on one of the pigs, was EsaB, which is a negative regulator of EsaC. EsaC production and secretion is increased when Staphylococci replicate in serum or infected hosts (27). EsaB and EsaC are defined as being involved in *S. aureus* virulence and are required for persistent infection, EsaB mutants fail to repress EsaC and bacteria lacking EsaB function will overproduce EsaC. The over-expression of EsaC is also the natural response when *S. aureus* is replicating in host tissue. Animals and humans mount however an immune response to EsaC during infection (27), which could explain why a constitutive over-expression of EsaC, in the EsaB mutants, might not be in the favour of the pathogen in the long run. Enzymes involved in membrane lipid metabolism and galactose metabolism were also identified as important for porcine skin adhesion and survival. Tn916 integrase mutants were evaluated to be hypercompetitive in the porcine skin screening assays. Tn916 encodes tetracycline resistance which most likely is responsible for selection of ST398 in the pig reservoir as tetracycline often is used for therapeutic treatment in pigs. The Tn916 integrase is part of the conjugation transfer system of the transposon (21). The mobility of the Tn916 is disrupted in the integrase mutant, which might give the mutant a competitive advantage under the experimental conditions applied.

A previous study using porcine nasal explants identified beta-toxin gene *hlb* as a *S. aureus* S0385 nasal colonization factor, based on CFU quantifications of S0385 wild type and S0385 beta haemolysin mutant (1). SAPIG2471 encoding beta haemolysin was not among the genes identified as most significant for nasal epithelium survival in this study. When inspecting the raw count data a decrease in read count from input to output for the beta haemolysin genes was indentified in three of the four replicates (data not shown), but this reduction was not defined as significant using the DESeq package in R. To verify the significance of the gene in nasal survival more tissue replicates should have been included in the analysis. Only four genes with specific importance for nasal epithelium survival were found in this study, two of

which showed decrease in fitness and two with increased fitness. SAPIG1248 encoding an aminoacyltransferase FemA, which is essential for expression of *mec*A, was identified with reduced fitness in the nasal survival model. S0385 genome is annotated with four different aminoacyltransferase FemA encoding genes, one of which (SAPIG1375) was identified by us previous as essential for growth under laboratory conditions (manuscript 1). This essential *fem*A gene is 1263 nucleotides whereas the *fem*A gene found as important for nasal survival in this study is only 486 and they show 88 % identity. FemA is involved in methicillin resistance and *fem*A mutants have shown a reduced glycine content in the peptidoglycan layer, a reduced cell wall turnover in growing cells, reduced whole-cell autolysis under non-growing conditions and increased methicillin sensitivity (28). The alteration in the cell wall could reduce bacterial resistance to the host immune response. Two genes representing hypercompetitive mutants when screened in the nasal survival model, encode a cell wall anchor domain and a transpeptidase which anchors surface proteins to the cell wall. The *S. aureus* sortase attaches surface proteins to the cell wall and the lpxtg-mediated anchoring domain also mediates attachment of proteins to the cell wall. These could be important for bacterial adhesion, but also facilitate a immune recognition site and as the nasal tissues were not washed post infection, genes important for adhesion to these tissue samples were not tested in this study.

Conclusion

Comparative genomic studies have identified a few phage associated genes that appears to be correlated with virulence in humans, but have not been able to identify genes of importance for successful colonization or infection in livestock or other animals (8). The pig reservoir is important for the spread of ST398 and the identification of genes important for survival of

ST398 in pigs could contribute to a better understanding of transmission and adaptation. In this study we screened a previously generated genome saturated LA-MRSA ST398 transposon mutant library, in *ex vivo* porcine skin and nasal epithelium survival and adhesion assays. Seventeen genes were identified as important for porcine skin adhesion and survival. Ten genes represent attenuated mutants with reduced fitness. They primarily encode transporters and enzymes involved in metabolic pathways. In addition four hypercompetitive mutants with increased fitness were identified and they encode DNA binding proteins involved in regulation. Sixteen genes were identified as important for nasal epithelial survival, encoding proteins involved in regulation, metabolic enzymes, cell wall components and hypothetical proteins. The genes identified in this study could constitute targets for MRSA decolonization in pigs and thereby prevent further spread and the potential adaption within the ST398 lineage which takes place in the pig reservoir. However the genes need further investigation to understand the specific function in porcine survival.

Acknowledgements

The work funded in part by the Danish Ministry of Food, Agriculture and Fisheries (Grant no.: 3304-FVFP-09-F-002-1). We thank the staff at the Sequencing Centre at Copenhagen University, DK for generating the sequences and DTU Multi-Assay Core (DMAC) for assistance with the sequence library preparations.

1. **Tulinski P, Fluit AC, van Putten JPM, de Bruin A, Glorieux S, Wagenaar J a, Duim B**. 2013. An ex vivo porcine nasal mucosa explants model to study MRSA colonization. PloS one **8**:e53783.

2. **Fluit a C**. 2012. Livestock-associated Staphylococcus aureus. Clinical Microbiology and Infection **18**:735–44.

3. **Foster TJ**. 2009. Colonization and infection of the human host by staphylococci: adhesion, survival and immune evasion. Veterinary dermatology **20**:456–70.

4. **Hasman H, Moodley a, Guardabassi L, Stegger M, Skov RL, Aarestrup FM**. 2010. Spa type distribution in Staphylococcus aureus originating from pigs, cattle and poultry. Veterinary microbiology **141**:326–31.

5. **Price L, Stegger M, Hasman H, Aziz M**. 2012. Staphylococcus aureus CC398: host adaptation and emergence of methicillin resistance in livestock. MBio **3**:e00305–11.

6. **Corrigan RM, Miajlovic H, Foster TJ**. 2009. Surface proteins that promote adherence of Staphylococcus aureus to human desquamated nasal epithelial cells. BMC microbiology **9**:22.

7. **Moodley A, Espinosa-Gongora C**. 2012. Comparative host specificity of human-and pig-associated Staphylococcus aureus clonal lineages. PloS one **7**:1–7.

8. **Uhlemann A, Porcella S, Trivedi S**. 2012. Identification of a highly transmissible animal-independent Staphylococcus aureus ST398 clone with distinct genomic and cell adhesion properties. MBio **3**:e00027–12.

9. **Clarke SR, Wiltshire MD, Foster SJ**. 2004. IsdA of Staphylococcus aureus is a broad spectrum, iron-regulated adhesin. Molecular microbiology **51**:1509–19.

10. **Schaffer AC, Solinga RM, Cocchiaro J, Portoles M, Kiser KB, Risley A, Randall SM, Valtulina V, Speziale P, Walsh E, Foster T, Lee JC**. 2006. Immunization with Staphylococcus aureus clumping factor B, a major determinant in nasal carriage, reduces nasal colonization in a murine model. Infection and immunity **74**:2145–53.

11. **Wertheim HFL, Walsh E, Choudhurry R, Melles DC, Boelens HAM, Miajlovic H, Verbrugh HA, Foster T, van Belkum A**. 2008. Key role for clumping factor B in Staphylococcus aureus nasal colonization of humans. PLoS medicine **5**:e17.

12. **Moodley A, Espinosa-Gongora C, Nielsen SS, McCarthy AJ, Lindsay J a, Guardabassi L**. 2012. Comparative host specificity of human- and pig- associated Staphylococcus aureus clonal lineages. PloS one **7**:e49344.

13. **Maisch T, Bosl C, Szeimies R-M, Love B, Abels C**. 2007. Determination of the antibacterial efficacy of a new porphyrin-based photosensitizer against MRSA ex vivo. Photochemical & photobiological sciences: Official journal of the European Photochemistry Association and the European Society for Photobiology **6**:545–51.

14. **Langridge GC, Phan M, Turner DJ, Perkins TT, Parts L, Haase J, Charles I, Maskell DJ, Peters SE, Dougan G, Wain J, Parkhill J, Turner AK**. 2009. Simultaneous assay of every Salmonella Typhi gene using one million transposon mutants. Genome research 2308–2316.

15. **Martin M**. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet. journal **17**:10–12.

16. **Langmead B, Salzberg SL**. 2012. Fast gapped-read alignment with Bowtie 2. Nature methods **9**:357–9.

17. **Anders S, Huber W**. 2010. Differential expression analysis for sequence count data. Genome biology **11**:R106.

18. **Anders S, Huber W**. 2012. Differential expression of RNA-Seq data at the gene level – the DESeq package. U.S. Patent DESeq version 1.13.0European Molecular Biology Laboratory (EMBL), Heidelberg, Germany. European Molecular Biology Laboratory (EMBL), Heidelberg, Germany.

19. **Chaudhuri RR, Morgan E, Peters SE, Pleasance SJ, Hudson DL, Davies HM, Wang J, van Diemen PM, Buckley AM, Bowen AJ, Pullinger GD, Turner DJ, Langridge GC, Turner a. K, Parkhill J, Charles IG, Maskell DJ, Stevens MP**. 2013. Comprehensive Assignment of Roles for Salmonella Typhimurium Genes in Intestinal Colonization of Food-Producing Animals. PLoS Genetics **9**:e1003456.

20. **Williams R**. 1963. Healthy carriage of Staphylococcus aureus: its prevalence and importance. Bacteriological Reviews **27**:56–71.

21. **Schijffelen MJ, Boel CHE, van Strijp JAG, Fluit AC**. 2010. Whole genome analysis of a livestock-associated methicillin-resistant Staphylococcus aureus ST398 isolate from a case of human endocarditis. BMC genomics **11**:376.

22. **Yoshida Y, Matsuo M, Oogai Y, Kato F, Nakamura N, Sugai M, Komatsuzawa H**. 2011. Bacitracin sensing and resistance in Staphylococcus aureus. FEMS microbiology letters **320**:33–9.

23. **Ohki R, Tateno K, Masuyama W, Moriya S, Kobayashi K, Ogasawara N**. 2003. The BceRS two-component regulatory system induces expression of the bacitracin transporter, BceAB, in Bacillus subtilis. Molecular Microbiology **49**:1135–1144.

24. **Azevedo E, Rios E, Fukushima K, GM C-T**. 1993. Bacitracin production by a new strain of Bacillus subtilis. Applied biochemistry and biotechnology **42**.

25. **Johnson B, Anker H, Meleney F**. 1945. Bacitracin: A new antibiotic produced by a member of the B. subilis group. Science **102**:376–377.

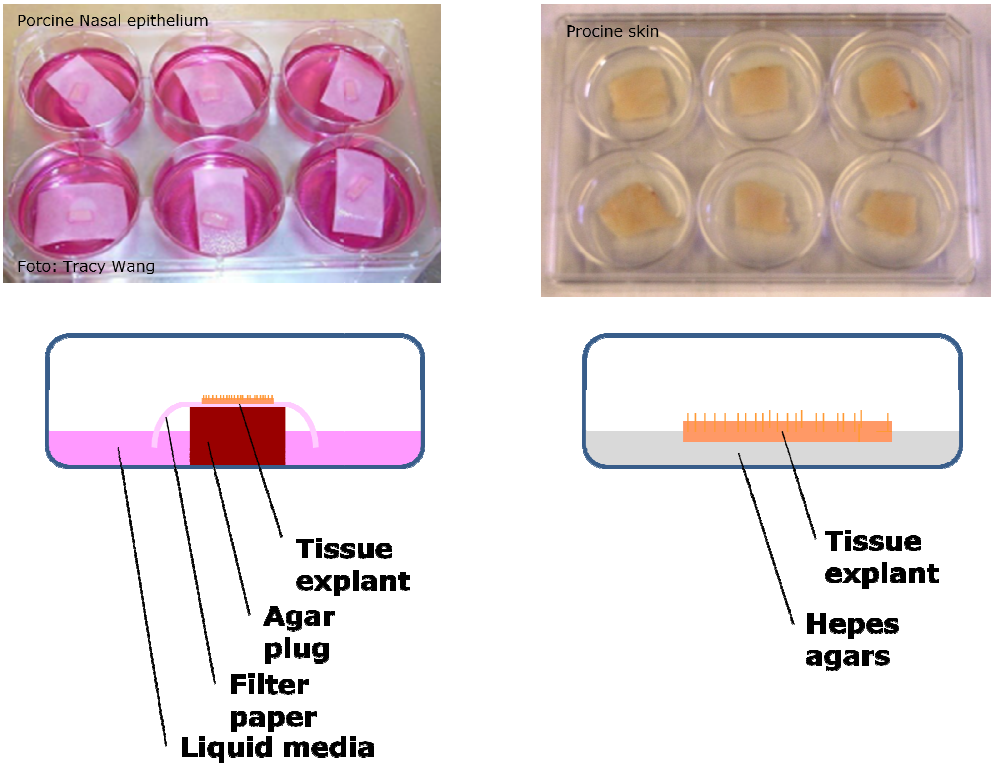
26. **Higgins CF**. 1992. ABC transporters: from microorganisms to man. Annual review of cell biology **8**:67–113.

27. **Burts M, DeDent A, Missiakas D**. 2008. EsaC substrate for the ESAT-6 secretion pathway and its role in persistent infections of Staphylococcus aureus. Molecular microbiology **69**:736–46.

28. **Maidhof H, Reinicke B, Blümel P, Berger-Bächi B, Labischinski H**. 1991. femA, which encodes a factor essential for expression of methicillin resistance, affects glycine content of peptidoglycan in methicillin-resistant and methicillin-susceptible Staphylococcus aureus strains. Journal of bacteriology **173**:3507–13.

Figures and Tables

Figure 1. Illustration of the porcine e*x vivo* nasal epithelial and skin models.



The figure shows photos and schematic presentations of the porcine *ex vivo* nasal epithelial model and the porcine *ex vivo* skin model.

Table 1. Genes important for porcine skin adhesion (stationary cells) and survival.

Gene ID	Pig_1 Adhesion (stationary cells)	Pig_1 Survival	Pig_2 Survival	Fitness Score	*P* value
SAPIG0287	Yes	Day 1+2	Day 2	-3.8 to -12 (4/5)	0.001 to <0.0001
SAPIG0300	Yes	Day 1	-	-4.3 to -5.3 (2/5)	<0.0001
SAPIG0737	Yes	Day 1+2	Day 1	-2.7 to -3.7 (4/5)	0.009 to <0.0001
SAPIG0739	Yes	Day 1	Day 1	-2.5 to -2.9 (3/5)	0.0009 to <0.0001
SAPIG0740	Yes	Day 1+2	Day 1	-1.8 to -2.7 (4/5)	0.001 to < 0.0001
SAPIG1303	Yes	Day 1+2	Day 1	-1.7 to -3.1 (4/5)	0.007 to <0.0001
SAPIG1425	Yes	Day 1+2	Day 1	2.0 to 3.2 (4/5)	0.005 to < 0.0001
SAPIG2410	Yes	Day 1	Day 1	2.6 to 3.5 (3/5)	0.002 to <0.0001

The table shows the genes which represent mutants with altered fitness after selection on porcine skin explants. Only mutant found with altered fitness in both the adhesion assay using stationary grown cells and the survival assay (survival after day 1 and/or day 2) are represented. A negative fitness score correspond to mutants with attenuated fitness and a positive fitness score correspond to mutants with increase fitness. All together 5 assays were conducted with 2-4 biological replicates in each assay. The genes selected showed a significant change in mutant clones from input to output at *P* level ≤0.01.

Table 2. Genes important for porcine skin adhesion (exponential cells) and survival.

Gene ID	Pig_1 Adhesion (exponential cells)	Pig_1 Survival	Pig_2 Survival	Fitness Score	*P* value
SAPIG0737	Yes	Day 1+2	Day 1	-2.5 to -4.6 (4/5)	0.009 to < 0.0001
SAPIG0740	Yes	Day 1+2	Day 1	-1.8 to -2.5 (4/5)	0.004 to < 0.0001
SAPIG0837	Yes	-	Day 1	-2.7 to -3.3 (2/5)	0.0003 to 0.0007
SAPIG1193	Yes	Day 1+2	-	-1.7 to -2.6 (3/5)	0.0003 to 0.0002
SAPIG1300	Yes	Day 1+2	Day 1	-2.0 to -3.3 (4/5)	0.01 to > 0.0001
SAPIG1964	Yes	Day 1+2	-	-2.2 to -2.9 (3/5)	0.009 to < 0.0001
SAPIG1825	Yes	Day 1	-	-1.6 to 1.4 (2/5)	0.003 to 0.0006
SAPIG0721	Yes	Day 1	-	-1.4 to 2.3	0.004 to 0.0008

				(2/5)	
	Yes			-2.3 to 3.3	0.002 to 0.0004
SAPIG1418		Day 1	-	(2/5)	
SAPIG0953	Yes	-	Day 1	2.1 to 3.2 (2/5)	0.008 to 0.0006
SAPIG1586	Yes	Day 1	Day 1	1.7 to 3.4 (3/5)	0.006 to 0.003

The table shows the genes which represent mutants with altered fitness after selection on porcine skin explants. Only mutant found with altered fitness in both the adhesion assay using exponentially grown cells and the survival assay (survival after day 1 and/or day 2) are represented. A negative fitness score correspond to mutants with attenuated fitness and a positive fitness score correspond to mutants with increase fitness. All together 5 assays were conducted with 2-4 biological replicates in each assay. The genes selected showed a significant change in mutant clones from input to output at *P* level $\leq$0.01.

Table 3. Description of the genes identified as important for porcine skin adhesion and survival.

Gene ID	Relative fitness	Description	KEGG
Attachment (stationary cells) and Survival			
SAPIG0287	Attenuated	Hypothetical protein	-
SAPIG0300	Attenuated	Protein EsaB	Virulence protein/Secretion system
SAPIG0737	Attenuated	DNA-binding response regulator	Bacitracin transport
SAPIG0739	Attenuated	Bacitracin export ATP-binding protein BceA	Bacitracin transport/*S. aureus* infection
SAPIG0740	Attenuated	ABC transporter, permease protein	*S. aureus* infection
SAPIG1303	Attenuated	Aerobic glycerol-3-phosphate dehydrogenase	Glycerophospolipid metabolism
SAPIG1425	Hypercompetitive	Methionine-S-sulfoxide reductase MsrA	-
SAPIG2410	Hypercompetitive	HTH-type transcriptional regulator TcaR	Arsenical Resistance Operon Repressor and similar prokaryotic, metal regulated homodimeric repressors
Attachment (exponential cells) and Survival			
SAPIG0737	Attenuated	DNA-binding response regulator	Bacitracin transport
SAPIG0740	Attenuated	ABC transporter, permease protein	*S. aureus* infection
SAPIG0837	Attenuated	Prolipoprotein diacylglyceryl transferase	-
SAPIG1193	Attenuated	Phosphoenolpyruvate-dependent sugar phosphotransferase system, eiia 2, putative	Galactose metabolism/metabolic pathways
SAPIG1300	Attenuated	Glycerol uptake operon antiterminator regulatory protein	-
SAPIG1964	Attenuated	Teichoic acid translocation ATP-binding protein TagH	Transporter
SAPIG0721	Inconsistancy	Penicillin binding protein 4	Peptidoglycan biosynthesis
SAPIG1418	Inconsistancy	Hypothetical protein	-
SAPIG1825	Inconsistancy	Glucosaminidase	-
SAPIG0953	Hypercompetitive	Transposase from transposon (Integrase)	DNA binding domain of tn916 integrase
SAPIG1586	Hypercompetitive	Arginine repressor	DNA binding domain

The genes found to represent mutants with altered fitness when screened in the porcine skin survival and adhesion assay are shown in Table 2. Gene ID corresponding to the NCBI gene database, relative fitness, gene description, and KEGG are illustrated. The genes marked in

purple were identified with attenuated fitness in the adhesion assay regardless of growth phase.

Table 4. Porcine nasal epithelium survival.

Gene ID	Nasal	BHI	Skin	Fitness Score	*P* value	Description
SAPIG2163	Yes	Yes	Yes	-5.7 and -12	0.005 to < 0.0001	Transcription termination factor Rho
SAPIG0562	Yes	Yes	Yes	-3.0 and -5.2	0.04 to 0.005	Pur operon repressor
SAPIG2016	Yes	Yes	Yes	-3.8 and -4.9	0.03 to 0.01	YkgB
SAPIG1363	Yes	Yes	Yes	-3.2 and -4.7	0.02 to 0.01	Regulatory protein MsrR
SAPIG1248	Yes	No	No	-3.7 and 4.2	0.04 to 0.02	Aminoacyltransferase FemA (Factor essential for expression of methicillin resistance A)
SAPIG2147	Yes	Yes	Yes	-4.0 and -12	0.02 to 0.0006	ATP synthase F1, alpha subunit
SAPIG2568	Yes	No	Yes	-3.2 and -3.9	0.05 to 0.0002	Fructose-1,6-bisphosphatase
SAPIG1302	Yes	Yes	Yes	-3.9 and -4.5	0.04 to 0.004	Glycerol kinase
SAPIG1833	Yes	Yes	Yes	-3.9 and -4.9	0.003 to 0.002	Hypothetical protein
SAPIG0738	Yes	Yes	Yes	-2.6 and -3.8	0.007 to 0.0001	Sensor protein BceS
SAPIG2002	Yes	Yes	No	-3.5 and -7.8	0.04 to 0.03	Adenylosuccinate lyase
SAPIG0814	Yes	Yes	No	-3.3 and -6.2	0.02 to 0.0003	UDP-N-acetylenolpyruvoylglucosamine reductase
SAPIG0287	Yes	No	No	-2.9 and -5.8	0.02	Hypothetical protein
SAPIG0786	Yes	Yes	Yes	-2.3 and -2.6	0.04 to 0.02	Hypothetical protein
SAPIG1809	Yes	No	No	2.5 and 3.2	0.04 to 0.02	lpxtg-motif cell wall anchor domain
SAPIG2578	Yes	No	No	4.4 and 6.4	0.03 to 0.0003	Sortases are cysteine transpeptidases, found in gram-positive bacteria, that anchor surface proteins to peptidoglycans of the bacterial cell wall envelope

Table 4 illustrates the genes representing mutants identified with significant altered fitness in the nasal epithelial survival assay (*P* level ≤0.05). Information about whether these genes were essential/beneficial for growth under laboratory conditions (manuscript I Table S1 and S2) and for porcine skin survival was included in the table (*P* level ≤0.05).

Supplementary materials

Table S1. Porcine skin survival.

Gene ID	BHI	Pig_1	Pig_2	Fitness Score	*P* value	Description
PSAPIG030001	No	Day 2	Day 2	-3.5 to -4.2 (2/4)	<0.0001	replication protein Rep
SAPIG0004	Yes	Day 2	Day 2	-2.5 to -3.6 (2/4)	<0.0001	DNA replication and repair protein RecF
SAPIG0287	No	Day 1+2	Day 2	-3.8 to -12 (3/4)	0.001 to <0.0001	hypothetical protein
SAPIG0737	No	Day 1+2	Day 1	-2.5 to 3.7 (3/4)	0.01 to <0.0001	DNA-binding response regulator
SAPIG0738	Yes	Day 1+2	Day 1	-2.0 to -3.6 (3/4)	0.005 to <0.0001	sensor protein BceS
SAPIG0739	No	Day 1	Day 1	-2.9 (2/4)	0.001 to <0.0001	bacitracin export ATP-binding protein BceA
SAPIG0740	No	Day 1+2	Day 1	-1.6 to -2.8 (3/4)	0.001 to >0.0001	ABC transporter, permease protein
SAPIG0786	Yes	Day 1+2	Day 1	-2.2 to -4.5 (3/4)	0.001 to <0.0001	hypothetical protein
SAPIG0814	Yes	Day 1+2	Day 2	-3.2 to -5.1 (3/4)	0.002 to <0.0001	UDP-N-acetylenolpyruvoylglucosamine reductase
SAPIG1198	Yes	Day 1	Day 1	-4.5 to -4.6 (2/4)	0.0004 to <0.0001	dihydroorotase (DHOase)
SAPIG1300	No	Day 1+2	Day 1	-2.4 to -3.3 (3/4)	0.002 to <0.0001	glycerol uptake operon antiterminator regulatory protein
SAPIG1302	Yes	Day 1	Day 1+2	-2.4 to -5.3 (3/4)	0.005 to 0.004	glycerol kinase
SAPIG1303	No	Day 1+2	Day 1	-1.7 to -3.1 (3/4)	0.007 to <0.0001	aerobic glycerol-3-phosphate dehydrogenase
SAPIG1309	Yes	Day 1+2	Day 2	-1.8 to -4.0 (3/4)	0.007 to 0.0004	aluminium resistance protein
SAPIG1464	No	Day 1+2	Day 2	-3.0 to -5.0 (3/4)	0.02 to <0.0001	3-phosphoshikimate 1-carboxyvinyltransferase
SAPIG1756	Yes	Day	Day	-4.1 to -	0.0003 to	DNA polymerase III subunit

		1	2	7.3 (2/4)	<0.0001	alpha
SAPIG1788	Yes	Day 2	Day 1	-5.3 to -5.5 (2/4)	<0.0001	catabolite control protein A
SAPIG1833	Yes	Day 1	Day 1	-2.3 to -4.7 (2/4)	0.0004 to <0.0001	hypothetical protein
SAPIG2016	Yes	Day 1+2	Day 1+2	-5.2 to -7.9 (4/4)	0.005 to <0.0001	YkgB
SAPIG2090	Yes	Day 2	Day 1	-1.2 to -4.5 (2/4)	0.005 to 0.002	peptidase M22, glycoprotease
SAPIG2147	Yes	Day 1+2	Day 1+2	-4.2 to -6.6 (4/4)	0.0002 to <0.0001	ATP synthase F1, alpha subunit
SAPIG2163	Yes	Day 1+2	Day 1+2	-4.8 to -8.6 (4/4)	0.001 to <0.0001	transcription termination factor Rho
SAPIG0457	No	Day 1	Day 1	1.4 to 2.0 (2/4)	0.01 to 0.006	alkyl hydroperoxide reductase, F subunit
SAPIG0701	No	Day 2	Day 1	1.5 to 2.5 (2/4)	0.008 to 0.0003	phage integrase family protein
SAPIG1425	No	Day 1+2	Day 1	2.1 to 4.9 (3/4)	0.005 to 0.002	methionine-S-sulfoxide reductase
SAPIG1586	No	Day 1	Day 1	1.7 to 3.4 (2/4)	0.005 to 0.004	arginine repressor
SAPIG2410	No	Day 1	Day 1	3.2 to 3.5 (2/4)	0.002 to <0.0001	HTH-type transcriptional regulator TcaR

The table shows genes representing mutants that were identified with a change in fitness in the porcine skin survival assay on Pig_1 and Pig_2 for 1 and/or 2 days incubation. Only genes with a significant change in fitness score at the *P* level ≤0.01 were included. The BHI column represents genes evaluated previously as essential/beneficial for survival under laboratory conditions (manuscript I Table S1 and S2).

Table S2. Adhesion assay with stationary grown mutants (Pig_1).

Gene ID	Fitness Score	*P* value	Description
SAPIG1372	-3,02	0,016	N-(5'phosphoribosyl)anthranilate isomerase
SAPIG1342	-2,74	0,037	hypothetical protein
SAPIG2137	-2,55	0,019	tena/thi-4 family
SAPIG1154	-2,52	0,025	fibrinogen-binding protein
SAPIG2731	-2,37	0,001	ATP phosphoribosyltransferase regulatory subunit
SAPIG0199	-2,22	0,001	N-acetyl-gamma-glutamyl-phosphate reductase
SAPIG2510	-2,14	0,000	glutamate synthase-ferredoxin large subunit
SAPIG1800	-2,12	0,037	metallo-beta-lactamase superfamily protein
SAPIG0659	-2,12	0,006	hypothetical protein
SAPIG0135	-2,10	0,010	pyridoxal-dependent decarboxylase decarboxylase
SAPIG2262	-1,99	0,003	hyaluronate lyase (Hyaluronidase) (HYase)
SAPIG2477	-1,90	0,005	aminotransferase, class II
SAPIG0971	-1,84	0,006	Hydrolase
SAPIG1383	-1,82	0,026	hypothetical protein
SAPIG0489	-1,77	0,004	3-beta hydroxysteroid dehydrogenase/isomerase
SAPIG1111	-1,65	0,011	pyruvate carboxylase
SAPIG0643	-1,56	0,020	hypothetical protein
SAPIG2500	-1,56	0,010	glycine betaine/carnitine/choline transport ATP-binding protein opuCA
SAPIG0243	-1,56	0,016	acyl-CoA dehydrogenase family protein
SAPIG2506	-1,55	0,029	hypothetical protein
SAPIG2268	-1,54	0,013	acetolactate synthase, catabolic
SAPIG0849	-1,50	0,009	hypothetical protein
SAPIG2631	-1,49	0,026	hydrolase, alpha/beta hydrolase fold family
SAPIG0437	-1,49	0,013	bifunctional homocysteine S-methyltransferase/5,10-methylenetetrahydrofolate reductase protein
SAPIG2426	-1,48	0,028	pyridine nucleotide-disulphide oxidoreductase family protein
SAPIG0606	-1,47	0,018	cysteinyl-tRNA synthetase
SAPIG1973	-1,44	0,017	hypothetical protein
SAPIG0578	-1,43	0,043	chaperonin HslO
SAPIG1482	-1,42	0,002	hypothetical protein
SAPIG2231	-1,39	0,002	probable uridylyltransferase
SAPIG0541	-1,37	0,041	alpha,alpha-phosphotrehalase
SAPIG2563	-1,35	0,010	DedA family protein
SAPIG0748	-1,32	0,001	hypothetical protein
SAPIG2350	-1,30	0,013	staphylococcal secretory antigen ssaA2
SAPIG0752	-1,29	0,029	Surface antigen
SAPIG0010	-1,25	0,036	AzlC family protein
SAPIG2238	-1,25	0,043	alanine racemase, N-domain family
SAPIG2357	-1,24	0,004	bifunctional autolysin
SAPIG1703	-1,22	0,028	S-adenosylmethionine:tRNA ribosyltransferase-isomerase
SAPIG1075	-1,21	0,004	hypothetical protein
SAPIG2343	-1,19	0,005	urease accessory protein UreG
SAPIG2564	-1,13	0,050	multidrug-efflux transporter
SAPIG2111	-1,12	0,008	alanine racemase
SAPIG1996	-1,10	0,029	sodium/proline symporter

SAPIG2491	-1,10	0,004	sodium/hydrogen exchanger family protein
SAPIG2352	-1,08	0,001	NAD/nadp octopine/nopaline dehydrogenase family protein
SAPIG2485	-1,08	0,004	hypothetical protein
SAPIG0441	-1,07	0,004	transporter, small conductance mechanosensitive ion channel (MscS) family
SAPIG2703	-1,05	0,034	translocase, putative
SAPIG2617	-1,04	0,048	staphylococcal secretory antigen ssaA1
SAPIG2348	-1,02	0,045	transcriptional regulator, AraC family
SAPIG0431	-1,01	0,020	ABC transporter ATP-binding protein
SAPIG0800	-1,00	0,027	ABC transporter permease protein
SAPIG2237	-0,92	0,010	ferrichrome ABC transporter lipoprotein
SAPIG0456	-0,92	0,025	hypothetical protein
SAPIG2679	-0,89	0,010	clumping factor B (Fibrinogen-binding protein B) (Fibrinogenreceptor B)
SAPIG1100	-0,89	0,003	hypothetical protein
SAPIG2264	-0,86	0,019	hypothetical protein
SAPIG0122	-0,85	0,034	immunoGlobulin g binding protein a
SAPIG2569	-0,84	0,049	hypothetical protein
SAPIG0858	-0,82	0,037	Carboxylesterase
SAPIG0795	-0,80	0,049	allophanate hydrolase subunit 2
SAPIG0782	-0,79	0,009	Amino acid transport and metabolism
SAPIG2335	-0,72	0,016	ferric hydroxamate receptor 1
SAPIG2215	-0,71	0,016	truncated FmtB protein
SAPIG0271	-0,65	0,013	glycosyl transferase, group 2 family protein
SAPIG0046	-0,59	0,045	hypothetical protein
SAPIG2589	-0,57	0,040	pyruvate oxidase

The table shows the genes representing mutants identified with a change in fitness in the adhesion assay with stationary grown mutants. Only mutants, that showed a significant attenuation in fitness at *P* level ≤ 0.05, were included. These genes were not identified as essential/beneficial for growth under laboratory conditions (manuscript I) or with significant altered fitness in the skin survival assay.

Table S3. Adhesion assay with exponentially grown mutants (Pig_1).

Gene ID	Fitness Score	*P* value	Description
SAPIG0883	-5,19	0,004	3-dehydroquinate dehydratase, type I
PSAPIG030002	-4,57	0,014	transcriptional regulator
SAPIG0772	-3,92	0,002	multidrug resistance protein 1 (Multidrug-efflux transporter 1)
SAPIG2627	-3,84	0,006	hypothetical protein
SAPIG1493	-3,72	0,005	Holin
SAPIG1094	-3,42	0,045	hypothetical protein
SAPIG0017	-3,27	<0,001	adenylosuccinate synthetase
SAPIG1163	-3,07	0,039	ornithine carbamoyltransferase
SAPIG2138	-3,02	0,035	SceD
SAPIG0163	-2,96	0,026	capsular polysaccharide type 5 biosynthesis protein cap5A
SAPIG2049	-2,76	0,050	hypothetical protein
SAPIG0257	-2,62	0,036	phosphoenolpyruvate-dependent sugar phosphotransferase system, eiia 2, putative
SAPIG2098	-2,55	0,023	3-isopropylmalate dehydratase, large subunit
SAPIG1305	-2,53	0,048	tRNA delta(2)-isopentenylpyrophosphate transferase
SAPIG1982	-2,46	0,014	protein in map 5'region
SAPIG0166	-2,07	0,005	capsular polysaccharide biosynthesis protein Cap5D
SAPIG0405	-2,05	0,023	protein in Tap1-dppD intergenic region
SAPIG0105	-1,90	0,016	transport protein
SAPIG2633	-1,74	0,040	ferrous iron transporter protein B
SAPIG2248	-1,54	0,017	cell surface hydrolase
SAPIG2651	-1,51	0,019	amino acid permease family protein
SAPIG2671	-1,46	0,037	ABC transporter
SAPIG1782	-1,45	0,017	penicillin-binding protein 1A
SAPIG0062	-1,44	0,050	coenzyme A disulfide reductase/ disulfide bond regulator domain
SAPIG0289	-1,38	0,035	transmembrane efflux pump protein
SAPIG2168	-1,25	0,039	hypothetical protein
SAPIG2386	-1,09	0,030	hypothetical protein
SAPIG2419	-1,07	0,048	L-lactate permease

The table shows the genes representing mutants identified with a change in fitness in the adhesion assay with exponentially grown mutants. Only mutants, that showed a significant attenuation in fitness at *P* level ≤ 0.05, were included. These genes were not identified as essential/beneficial for growth under laboratory conditions (manuscript I) or with significant altered fitness in the skin survival assay.

Title: Identification of virulence genes in whole genome sequenced *Staphylococcus aureus*.

Running title: *Staphylococcus aureus* VirulenceFinder.

Keywords: *Staphylococcus aureus*, virulence genes identification, database, virulence profile, whole genome sequencing, genotype.

Authors: Mette T. Christiansen[1(#)], Henrik Hasman[1] and Frank M. Aarestrup[1(#)]

[1]National Food Institute, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark

Corresponding authors: Mette T. Christiansen, email: mtso@food.dtu.dk and Frank M. Aarestrup, email: ffma@food.dtu.dk

Abstract

Staphylococcus aureus is an opportunistic pathogen that colonizes various animal hosts, including humans. Between 20-40 % of the human population are permanent or intermediate carriers of *S. aureus* (1, 2). It can cause a variety of infections ranging from minor soft tissue and skin infections to life-threatening systemic infections. Successful infection in a giving host depends on virulence factors produced by the bacterium, which can promote adhesion, immune evasion as well as damage to host cells. Different *S. aureus* strains encode different sets of virulence genes that somewhat reflect the environment within the host.

Tool for identification of specific virulence genes are important in diagnostics and surveillance. New approaches within diagnostics and surveillance for species identification, evolutionary clustering, and identification of resistance and virulence markers are based on whole genome sequencing (WGS). The biggest challenge with the appliance of WGS is, however, to interpret the large amount of data retrieved with this technology. We constructed the *S. aureus* VirulenceFinder, a web server that can utilise WGS data from *S. aureus* genomes to extract a virulence profile. The database will be freely available through The Centre for Genomic Epidemiology (CGE) (www.genomicepidemiology.org) web services.

Introduction

Staphylococcus aureus is a well know opportunistic pathogen that colonizes mucous membranes and skin surfaces of various animal species including humans (1, 3–5). *S. aureus* can cause a variety of infections ranging from minor soft tissue and skin infections to life-threatening systemic infections (6, 7). Successful infection in a giving host depends on virulence factors produced by *S. aureus* (4). Both secreted and cell surface associated proteins can promote adhesion to host extracellular matrices, damage host cells and evade the host immune system (5). *S. aureus* infections have become of increasing concern with the development of resistance to multiple antimicrobial drugs, including methicillin (methicillin-resistant *S. aureus* - MRSA). Different *S. aureus* strains encode different sets of virulence genes that somewhat reflect the environment within the host. Health care-associated MRSA (HA-MRSA) and community-associated MRSA (CA-MRSA) are genetically distinct as their environmental niches differ and they display different virulence profiles. HA-MRSA infects immune compromised individuals often under antibiotic selective pressure, whereas CA-MRSA infects healthy individuals requiring a different set of virulence factors such as the Panton-Valentine leukocidin (PVL), which is thought to be an important toxin in the CA strains (8). In the early 2000s a new emerging CA linage was identified in livestock in Europe. The linage belongs to the Sequence Type 398 (ST398) and is referred to as a livestock-associated (LA) *S. aureus* (9, 10). ST398 is genetically distinct from other CA *S. aureus* and display a broader host spectrum compared to HA *S. aureus* and can colonize and infect livestock as well as humans (11). Additionally evidence has been shown, suggesting that ST398 originated in humans, transmitted and adapted to pigs and are now transferring back to humans (12). All together *S. aureus* display a large set of virulence factors which is required for bacterial colonization and infection in different host under various conditions.

Identification of specific virulence genes as well as antibiotic resistance markers and bacterial typing are important tool in diagnostics and surveillance. Conventional typing is based on phenotypic methods some of which are verified by genotypic tests. However in recent years, whole genome sequencing (WGS) has become increasingly available. There have been huge improvements in sequencing technologies and the cost has gone down significantly. This gives rise to a new approach within diagnostics and surveillance, where WGS can be utilized for species identification, evolutionary clustering, identification of resistance and virulence markers just to mention a few of the many applications. The biggest challenge with the appliance of WGS is, however, to interpret the large amount of data retrieved with this technology. To translate large amounts of DNA sequences into functional information requires bioinformatics tools that are standardized and simple to use. The Centre for Genomic Epidemiology (CGE) (www.genomicepidemiology.org) aims at generating bioinformatic tools for handling WGS information, useful for outbreak investigation, epidemiological surveillance, source tracking and diagnostics. The service is publically available through web servers.

In this study we present the construction of the *S. aureus* VirulenceFinder, a web server that utilises WGS data from *S. aureus* genomes to extract a virulence profile. The database will be freely available through the CGE web services.

Methods

Building the database. Data on virulence genes were retrieved from the publically available virulence database (http://www.mgc.ac.cn/VFs/) and published papers (1, 4, 5, 13). All DNA sequences were collected from the NCBI nucleotide database (http://www.ncbi.nlm.nih.gov/nuccore/). The virulence gene sequences included in the database were selected based on the annotation and gene description from 31 different *S. aureus* genomes

(supplementary material Table S1). The genes were selected based on the annotation and gene description defined in the NCBI gene database and the definition of the virulence genes included in the VirulenceFinder database are therefore dependent on the annotations and gene description accuracy found within the NCBI gene database. When two genes with the same annotation or gene description showed 100 % nucleotide identity when using NCBI's nucleotide BLAST and had the same gene length, only one of the genes was included in the VirulenceFinder database. If two genes with the same annotation or gene description showed 100 % nucleotide identity using BLAST but were of different length, both genes were included in the database. If two genes with the same annotation or gene description showed less than 100 % nucleotide identity using BLAST, both genes were included in VirulenceFinder. No pseudo-genes were included in the database and minority variance found within the same *S. aureus* sample were also not included as the virulence genes were selected at the consensus level. Genotyping cannot be performed with *S. aureus* VirulenceFinder but other tools are available for typing on the CGE website (http://www.genomicepidemiology.org/).

Prior to submission to the *S. aureus* VirulenceFinder, draft assembly of sequence reads need to be performed. The server can assemble short reads sequences to draft genomes or it can be done as described previously by Larsen *et al.* (2012) (15). Once the sequence is submitted for a run, the VirulenceFinder uses BLAST to identify virulence genes matching any gene sequence found within the database. It is possible to select a threshold of sequence identity (% ID) between 85 % - 100 %, where 100 % ID is default. Here, any gene found within the VirulenceFinder database must show a minimum nucleotide identity corresponding to the selected threshold over the full gene length to be included in the output. A gene will not be reported if the submitted sequence contains less than 60 % of the full gene length of the matching virulence gene found within the database. If a virulence

gene is identified in a submitted sample the best-matching genes are given as output with corresponding GenBank accession number, correlating to the genome in which the gene matches.

Initial evaluation of method. For an initial evaluation of the database the genome of the assembled LA-MRSA ST398 S0385 isolate (GenBank accession AM990992) was screened for virulence genes. This genome was one of the 31 genomes used for building the database and was therefore selected as a preliminary evaluation of the database. The fasta file of the completed S0385 genome was submitted to the *S. aureus* VirulenceFinder alpha version (http://cge.cbs.dtu.dk/services/VirulenceFinder/index2.php) and the threshold for nucleotide identity was set to ID=98 %. The virulence profile of S0385 was evaluated and compared to previous findings in ST398 strains (4, 11, 14).

Further evaluation - identification of virulence genes in whole genome sequenced *S. aureus* strains. For further evaluation of *S. aureus* VirulenceFinder 89 previously sequenced ST398 genomes originating from various hosts, were screened for virulence genes using the database (12). Fourteen representatives of these, originating from humans and pigs were subjects for further analysis. The 14 isolates were selected to represent, the three most dominant spa types in the Lance B. Price *et al.* (2012) study, different host origin, different resistance pattern and different country of isolation. In addition two strains from each of the seven clades given by the Minimum-parsimony tree generated by Lance B. Price et al. (2012) were included (12). All genes within the VirulenceFinder database were BLASTed against the assembled genomes and the best matching genes were given as output. The threshold was set to 95 % identity (ID=95 %) for evaluation of a less stringent nucleotide identity threshold.

Results

The database accepts input as complete or partial, preassembled genomes. VirulenceFinder consists of different configurations and gives the option to select one or several species for a run and uses BLAST to identify virulence genes included in the database. It is possible to select a threshold of sequence identity (% ID) between 85 % - 100 % and the best-matching genes are given as output. If a virulence gene is identified in a submitted sample the output gives a GenBank accession number, correlating to the genome in which the gene matches.

For initial evaluation of the output the annotated genome of *S. aureus* ST398 S0385 (GenBank accession AM990992) was tested with the VirulenceFinder. The whole genome sequence dataset of the S0385 isolate was used, amongst 30 other genomes, to build the VirulenceFinder database. The output consists of genes matching with 98 % identity to the 1053 GenBank files of which the database was created. The list of virulence genes identified in the ST398 S0385 genome can be found in Table 1. 63 genes defined as virulence genes in the database were identified in the ST398 S0385 genome, of which 17 associated to adhesion, 7 exoenzymes, 20 genes involved in host immune evasion, 6 genes related to secretion systems and 13 toxins. Sixteen of the identified virulence genes showed <100 % identity to sequences from the database.

Eighty-eight *S. aureus* ST398 genomes have been assembled and published previous by Lance B. Price et al. (2012) (12). 14 of these were selected and tested in the *S. aureus* VirulenceFinder. The isolates were selected to represent all the seven clades generated by a Maximum-parsimony tree. Isolates from different countries, different host origins and with different resistance pattern were included in this study. An overview of the virulence profiles at the 95 % identity level can be found in table 2. One isolate showed a somewhat different profile with fewer identified virulence genes

compared to the other 13 strains including in the comparison. Overall the virulence profiles were similar, but different virulence patterns in isolates from pig origin and human origin were identified, correlating to what was found by Price *et al.* (2012) (12).

Discussion

Resistance and virulence profiles can help elucidate the approach for optimal treatment and define the virulence capacity of the infectious agent. Such information is crucial at hospitals in diagnostics and such profiles can as well be applied in local and global surveillance studies.

The *S. aureus* VirulenceFinder database generated and evaluated in this study comprises a bioinformatic tool for identifying virulence genes in *S. aureus* genomes using WGS data.

The *S. aureus* ST398 S0385 genome, which has been sequenced and annotated previously, was evaluated with the *S. aureus* VirulenceFinder. Sixty-three different genes defined as virulence genes in the database were identified. The whole genome sequence dataset of S0385 was used to build the database and it was therefore expected that all the identified virulence genes would show 100 % identity to a sequence within database. However 16 of the identified virulence genes showed <100 % identity. This indicates that even though not all gene variants are included in the database one can still identify the genes by lowering the default identity threshold. The NCBI nucleotide database contains many variants of the same gene and the gene annotation of the different genes does not always correlate, which makes it difficult to include all gene variants found within the NCBI nucleotide database. In addition the GenBank databases are continuously increasing and the addition of new complete annotated genomes can contribute to new gene annotations within older genomes.

Correlating with previous findings in ST398 isolates the VirulenceFinder identified *fnbA*, *clfA*, *cna*, *cap5A* and *eap/map* in the ST398 S0385 genome (4, 14). Enterotoxin P (*sep*) and exfoliative toxin type A (*eta*) were identified in the S0385 genome by the database, contradicting that ST398 does not contain enterotoxins and exfoliative toxin genes (14, 16). Both toxins are annotated in the ST398 S0385 genome in the GenBank gene database, which explains why these genes are identified when screening the S0385 genome using the VirulenceFinder.

Virulence profiles of 14 selected ST398 strains from a previous study (12) were generated using the VirulenceFinder with the threshold for nucleotide identity set to 95 %. The assembled genomes were BLASTed against all gene sequences within the database and hits with at least 95 % nucleotide identity constituted the output. The ID threshold can be set by the user and a less stringent threshold is recommended as the alpha version of the database only contains the number variants of each virulence gene, originating from 31 different *S. aureus* genomes used in building the database. A stringent threshold may result in some variation missed when using the database. However a less stringent threshold will result in a considerably larger output that requires more analysis. Overall the profiles were similar except for one isolate (13349_6) which also was found to be an outlier by Price *et al.* (2012) (12). The four isolates originating from a human host were positive for the *scn* gene whereas none of the isolates originating from pigs contain the staphylococcal complement inhibitor. The *scn* and *sak* genes are both markers for strains of human origin (12, 17). Even though the *scn* was not identified in the isolates originating from pigs a staphylococcal complement inhibitor variant has been identified on a pathogenicity island in ST398 S0385 which is considered a porcine originating strain (11). Two of the human originating isolates contained both the sak gene and the two Panton-Valentine leukocidin encoding genes *lukF-PV* and

lukS-PV. This is in agreement the finding in Price *et al.* (2012) showing that some LA *S. aureus* strains are highly virulent, as is common for many of the CA *S. aureus* strains (12).

Different patterns in adhesins like the Ser-Asp rich fibrinogen-binding proteins *SdrC* and *SdrD* and the fibronectin binding protein B (*fnbB* gene) were observed. SdrC and SdrD have been shown to promote adhesion to human desquamated nasal epithelial cells together with other factor (18). The four isolates, originating from human hosts, contain *SdrC* and *SdrD* which could indicate that both genes are contributing to adhesion to the human nares (18), whereas only *SdrC* seems to be essential for adhesion to the porcine nares as all the porcine isolates contain *SdrC* and only some the *SdrD* gene. All strains contain fibronectin binding protein A (*fnbA* gene) in agreement with a previous study (14). However some of the isolates also harboured fibronectin binding protein B (*fnbB* gene). A previous study showed that the *fnbA* gene product was more important in *in vitro* and *in vivo* infections, but cooperation between fibronectin binding proteins A and B is necessary for the induction of severe infections resulting in septic death (19).

Previous studies have emphasized that *S. aureus* ST398 do not contain any enterotoxins and exfoliative toxins (4, 14, 16). However all the ST398 strains tested, expect the outlier strain 13349_6, showed a positive result for enterotoxins A and P (*sea* and *sep* gene respectively) and exfiliative toxin A (*eta* gene) when using the VirulenceFinder. The enterotoxin P gene identified in the ST398 isolates show 100% sequence identity to the annotated enterotoxin P gene found in ST398 S0385 genome (GenBank gene no. SAPIG1666). The definition of the genes, in the VirulenceFinder database, is dependent on the annotations within the genomes applied for building the database. The Enterotoxin P was originally defined after the full genome sequencing of *S. aureus* N315 (20). The sequence of the annotated enterotoxin P encoded in the ST398 S0385 genome (gene SAPIG1666) was BLASTed against sequences in the NCBI nucleotide database and similarities with two different genes within the N315 genome were identified (GenBank gene

SA1429 and SA1430). These genes encode an enterotoxin homolog and a protein similar to enterotoxin A precursor. This indicates that the enterotoxin P (*sep* gene) identified in the ST398 stains tested here is not the same as the original *sep* gene defined in *S. aureus* N315 (GenBank gene SA1761) and that the definition given by the VirulenceFinder is a consequence of the annotations given to the reference genomes used for building the database. The *sep* gene identified here in the ST398 show however similarity to other enterotoxins and might indicate that *S. aureus* ST398 strains can contain enterotoxin like proteins that can be identified when using WGS data. This emphasises that as outputs from the *S. aureus* VirulenceFinder database relates to the annotations in the NCBI nucleotide database and for details beyond these annotations further investigation might be needed.

Here we demonstrated how an informative tool for WGS data can be generated. The *S. aureus* VirulenceFinder database is part of the tool package found on the CGE webpage (www.genomicepidemiology.org). Here are tools like MLST and ResFinder already available (15, 21) and additional tools for phylogenetic studies are under development.

Acknowledgements

The work funded in part by the Danish Ministry of Food, Agriculture and Fisheries (Grant no.: 3304-FVFP-09-F-002-1) and The Technical University of Denmark. We acknowledge Ea Zankari, Rolf Sommer Kaas and Martin Vestergaard from the National Food Institute and the Center for Biological Sequence analysis, Technical University of Denmark for assistance with generation of the VirulenceFinder database.

References

1. **Foster TJ**. 2009. Colonization and infection of the human host by staphylococci: adhesion, survival and immune evasion. Veterinary dermatology **20**:456–70.

2. **Williams R**. 1963. Healthy carriage of Staphylococcus aureus: its prevalence and importance. Bacteriological Reviews **27**:56–71.

3. **Tulinski P, Fluit AC, van Putten JPM, de Bruin A, Glorieux S, Wagenaar J a, Duim B**. 2013. An ex vivo porcine nasal mucosa explants model to study MRSA colonization. PloS one **8**:e53783.

4. **Fluit a C**. 2012. Livestock-associated Staphylococcus aureus. Clinical Microbiology and Infection **18**:735–44.

5. **Foster TJ**. 2005. Immune evasion by staphylococci. Nature reviews. Microbiology **3**:948–58.

6. **Huijsdens XW, Dijke BJ Van, Spalburg E, Van MG, Heck MEOC, Pluister GN, Voss A, Wannet WJB, Neeling AJ De**. 2006. Community-acquired MRSA and pig-farming. Annals of clinical microbiology and antimicrobials **5**:1–4.

7. **Hasman H, Moodley a, Guardabassi L, Stegger M, Skov RL, Aarestrup FM**. 2010. Spa type distribution in Staphylococcus aureus originating from pigs, cattle and poultry. Veterinary microbiology **141**:326–31.

8. **Diep BA, Palazzolo-Ballance AM, Tattevin P, Basuino L, Braughton KR, Whitney AR, Chen L, Kreiswirth BN, Otto M, DeLeo FR, Chambers HF**. 2008. Contribution of Panton-Valentine leukocidin in community-associated methicillin-resistant Staphylococcus aureus pathogenesis. PloS one **3**:e3198.

9. **Armand-Lefevre L**. 2005. Clonal comparison of Staphylococcus aureus isolates from healthy pig farmers, human controls, and pigs. Emerging infectious diseases **11**:711–714.

10. **Voss A, Loeffen F, Bakker J**. 2005. Methicillin-resistant Staphylococcus aureus in pig farming. Emerging infectious diseases **11**:2004–2005.

11. **Schijffelen MJ, Boel CHE, van Strijp JAG, Fluit AC**. 2010. Whole genome analysis of a livestock-associated methicillin-resistant Staphylococcus aureus ST398 isolate from a case of human endocarditis. BMC genomics **11**:376.

12. **Price L, Stegger M, Hasman H, Aziz M**. 2012. Staphylococcus aureus CC398: host adaptation and emergence of methicillin resistance in livestock. MBio **3**:e00305–11.

13. **Cheung AL, Bayer AS, Zhang G, Gresham H, Xiong Y-Q**. 2004. Regulation of virulence determinants in vitro and in vivo in Staphylococcus aureus. FEMS Immunology & Medical Microbiology **40**:1–9.

14. **Hallin M, De Mendonça R, Denis O, Lefort A, El Garch F, Butaye P, Hermans K, Struelens MJ**. 2011. Diversity of accessory genome of human and livestock-associated ST398 methicillin resistant Staphylococcus aureus strains. Infection, genetics and evolution **11**:290–9.

15. **Larsen M, Cosentino S**. 2012. Multilocus sequence typing of total-genome-sequenced bacteria. Journal of clinical microbiology **50**:1355–61.

16. **Golding GR, Bryden L, Levett PN, McDonald RR, Wong A, Graham MR, Tyler S, Van Domselaar G, Mabon P, Kent H, Butaye P, Smith TC, Kadlec K, Schwarz S, Weese SJ, Mulvey MR**. 2012. whole-genome sequence of livestock-associated st398 methicillin-resistant staphylococcus aureus Isolated from Humans in Canada. Journal of bacteriology **194**:6627–8.

17. **Sung JM-L, Lloyd DH, Lindsay J a**. 2008. Staphylococcus aureus host specificity: comparative genomics of human versus animal isolates by multi-strain microarray. Microbiology (Reading, England) **154**:1949–59.

18. **Corrigan RM, Miajlovic H, Foster TJ**. 2009. Surface proteins that promote adherence of Staphylococcus aureus to human desquamated nasal epithelial cells. BMC microbiology **9**:22.

19. **Shinji H, Yosizawa Y, Tajima A, Iwase T, Sugimoto S, Seki K, Mizunoe Y**. 2011. Role of fibronectin-binding proteins A and B in in vitro cellular infections and in vivo septic infections by Staphylococcus aureus. Infection and immunity **79**:2215–23.

20. **Omoe K, Imanishi K, Hu D, Kato H, Fugane Y, Abe Y, Hamaoka S, Watanabe Y, Nakane A, Uchiyama T, Shinagawa K**. 2005. Characterization of Novel Staphylococcal Enterotoxin-Like Toxin Type P. Infection and immunity **73**:5540–5546.

21. **Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, Aarestrup FM, Larsen MV**. 2012. Identification of acquired antimicrobial resistance genes. The Journal of antimicrobial chemotherapy **67**:2640–4.

Figures and Tables

Table 1. Test run of the ST398 S0385 genome in the *S. aureus* VirulenceFinder.

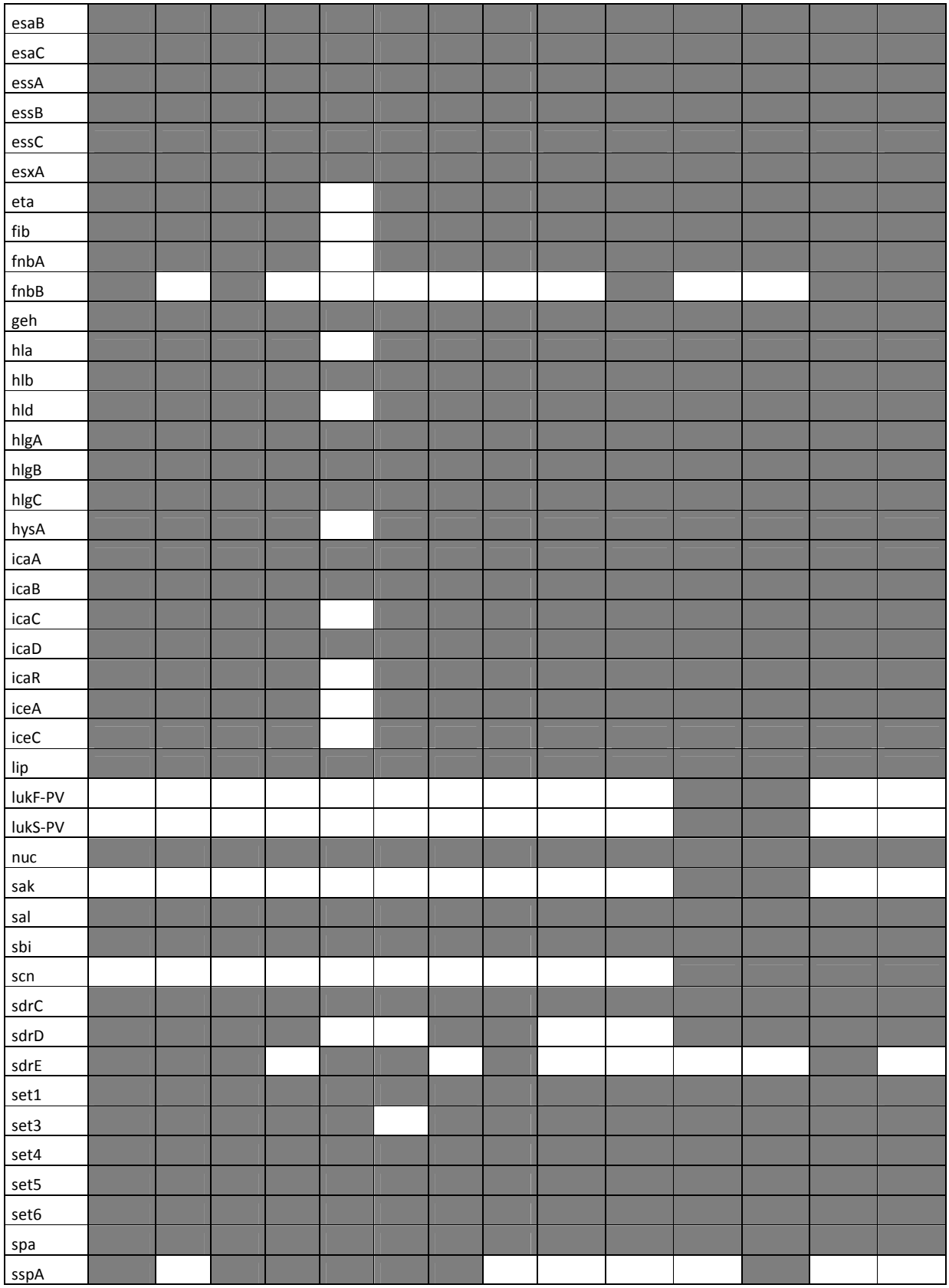
ID	Gene	Virulence factor	
100.00%	atl	**Autolysin**	***Adherence (17 genes)***
100.00%	clfA	**Clumping factor A**	
100.00%	clfB	**Clumping factor B**	
100.00%	cna	**Collagen adhesion**	
100.00%	ebpS	**Elastin binding protein**	
98.83%	eap/map	**Extracellular adherence protein/MHC analogous protein**	
100.00%	fib	**Fibrinogen binding protein**	
100.00%	efb	**Extracullelar Fibrinogen binding protein**	
100.00%	fnbA	**Fibronectin binding proteins**	
100.00%	fnbB		
100.00%	icaR	**Intercellular adhesin**	
100.00%	icaA		
99.67%	icaD		
100.00%	icaB		
100.00%	icaC		
100.00%	spa	**Staphylococcal protein A**	
100.00%	vwb	**von Willebrand factor**	
ID	**Gene**	**Virulence factor**	
99.14%	sspB	**Cysteine protease**	***Exoenzyme (7 genes)***
100.00%	sspC		
100.00%	hysA	**Hyaluronate lyase**	
100.00%	lip	**Lipase**	
100.00%	geh		
100.00%	coa	**Staphylocoagulase**	
100.00%	nuc	**Thermonuclease**	
ID	**Gene**	**Virulence factor**	
100.00%	isb	**IgG-binding protein**	***Host Immune evasion (20 genes)***
100.00%	cap5A	**Capsule**	
99.28%	cap1B		
98.84%	cap5B		
99.48%	cap5C		
100.00%	cap5D		
100.00%	cap8E		
98.75%	cap5F		
99.29%	cap5G		
100.00%	cap5H		
100.00%	cap5I		
100.00%	cap5J		
100.00%	cap5K		
100.00%	cap8L		
99.64%	cap5M		
98.76%	cap5N		
99.29%	cap5O		
98.95%	cap5P		
100.00%	capA		
100.00%	capC		
ID	**Gene**	**Virulence factor**	

100.00%	esaA	**Type VII secretion system**	***Secretion system (6 genes)***
100.00%	esaB		
99.33%	esaC		
100.00%	essA		
100.00%	essB		
100.00%	essC		
100.00%	esxA		
100.00%	**Gene**	**Virulence factor**	
100.00%	hla	**Alpha hemolysin**	***Toxin (13 genes)***
100.00%	hld	**Delta hemolysin**	
100.00%	sep	**Enterotoxin P (SEntP)**	
100.00%	eta	**Exfoliative toxin type A**	
100.00%	set1	**Exotoxin/superantigen-like proteins**	
100.00%	set3		
100.00%	set4		
100.00%	set5		
100.00%	set6		
99.45%	sal		
100.00%	SExo	**Superantigen-like**	
99.59%	hlgA	**Exotoxin**	
99.68%	hlgC	**Gamma hemolysin**	

The table illustrates the output from the ST398 S0385 genome (accession no. AM990992.1) test run in with the VirulenceFinder. The first column shows the sequence identity when all genes within the database were BLASTed against the assembled genomes and the best matching genes are given as output. The threshold was set to 98 % ID. Second column and third give the gene name and the encoding virulence factor. Fourth column defines the virulence group.

Table 2. Virulence profiles of 14 selected *S. aureus* ST398 strains.

Country:	FI	DE	IT	PL	IT	US	CA	US	US	US	DK	CN	US	FR
Spa type:	t034	t034	t011	t011	t011	t034	t034	t034	t034	t034	t034	t571	t571	t571
Host:	pig	pig	pig	pig	pig	pig	pig	pig	pig	pig	human	human	human	human
MET	R	S	R	R	R	S	R	R	S	S	R	S	S	S
Clade	IIa1i	IIa1i	IIa1ii	IIa1ii	IIa2	IIa2	IIa	IIa	IIa-GOI	IIa-GOI	II-GOI	II-GOI	I	I
Genes/Strains	2008-60-1662-5	1061	29139	6919/08	13349_6	F20	7-109	P23-02_SW62.1	F38	F10	50148	P23-9_WZ-1	1953	ST20091526
SEnt														
sea														
sep														
SExo														
atl														
cap1A														
cap1B														
cap1C														
cap5A														
cap5B														
cap5C														
cap5D														
cap5F														
cap5G														
cap5H														
cap5I														
cap5J														
cap5K														
cap5M														
cap5N														
cap5O														
cap5P														
cap8E														
cap8L														
clfA														
clfB														
cna														
coa														
eap/map														
ebh														
ebpS														
efb														
esaA														


esaB
esaC
essA
essB
essC
esxA
eta
fib
fnbA
fnbB
geh
hla
hlb
hld
hlgA
hlgB
hlgC
hysA
icaA
icaB
icaC
icaD
icaR
iceA
iceC
lip
lukF-PV
lukS-PV
nuc
sak
sal
sbi
scn
sdrC
sdrD
sdrE
set1
set3
set4
set5
set6
spa
sspA

sspB														
sspC														
vwb														

The table illustrates the virulence profiles of 14 selected *S. aureus* ST398 strains given by the *S. aureus* VirulenceFinder. The threshold was set to 95 % ID. The strains have been published previous by Lance B. Price et al. (2012). A grey box indicates the presence of a gene at the 95% identity level and a white box illustrates that the gene is not present in the genome at the 95% identity level. The top five rows show country of isolation, spa type, hhost origin, resistant or sensitive to methicillin and clade in which the isolate cluster according to Lance B. Price et al. (2012).

Supporting figures

Table S1. List of strains included in building of the database.

Strain:	GenBank accession no.
Staphylocoocus aures subsp. aureus N315	BA000018.3
Staphylocoocus aures subsp. aureus NCTC8325	CP000253.1
Staphylocoocus aures subsp. aureus TW20	FN433596.1
Staphylocoocus aures subsp. aureus 04-02981	CP001844.2
Staphylocoocus aures subsp. aureus 08BA02176	CP003808.1
Staphylocoocus aures subsp. aureus RF122	AJ938182.1
Staphylocoocus aures subsp. aureus 11819-97	CP003194.1
Staphylocoocus aures subsp. aureus 71193	CP003045.1
Staphylocoocus aures subsp. aureus COL	CP000046.1
Staphylocoocus aures subsp. aureus ECT-R2	FR714927.1
Staphylocoocus aures subsp. aureus ED133	CP001996.1
Staphylocoocus aures subsp. aureus ED98	CP001781.1
Staphylocoocus aures subsp. aureus HO 5096 0412	HE681097.1
Staphylocoocus aures subsp. aureus JH1	CP000736.1
Staphylocoocus aures subsp. aureus JH9	CP000703.1
Staphylocoocus aures subsp. aureus JKD6159	CP002114.2
Staphylocoocus aures subsp. aureus LGA251	FR821771.1
Staphylocoocus aures subsp. aureus MSHR1132	FR821777.2
Staphylocoocus aures subsp. aureus MSSA476	BX571857.1
Staphylocoocus aures subsp. aureus MW2	BA000032.2
Staphylocoocus aures subsp. aureus MU3	CP009324.1
Staphylocoocus aures subsp. aureus MU50	BA000017.4
Staphylocoocus aures subsp. aureus ST398	AM990992.1
Staphylocoocus aures subsp. aureus T0131	CP002643.1
Staphylocoocus aures subsp. aureus TCH60	CP002110.1
Staphylocoocus aures subsp. aureus USA_300_FPR3757	CP000255.1
Staphylocoocus aures subsp. aureus USA_300_TCH1516	CP000730.1
Staphylocoocus aures subsp. aureus VC40	CP003033.1
Staphylocoocus aures subsp. aureus JKD6008	CP002120.1
Staphylocoocus aures subsp. aureus str. Newman	AP009351.1
Staphylocoocus aures subsp. aureus MRSA252	BX571856.1

Table S2. Virulence genes included in the *S. aureus* VirulenceFinder database.

Virulence factors	Related genes
Adherence (22 genes)	
Autolysin	atl
Cell wall associated fibronectin binding protein	ebh
Clumping factor A	clfA
Clumping factor B	clfB
Collagen adhesion	cna
Elastin binding protein	ebpS
Extracellular adherence protein/MHC analogous protein	eap/map
Fibrinogen binding protein	fib
Extracullelar Fibrinogen binding protein	efb
Fibronectin binding proteins	fnbA
	fnbB
Intercellular adhesin	icaR
	icaA
	icaD
	icaB
	icaC
Ser-Asp rich fibrinogen-binding proteins	sdrC
	sdrD
	sdrE
	sdrH
Staphylococcal protein A	spa
von Willebrand factor	vwb
Exoenzyme (16 genes)	
Cysteine protease	sspB
	sspB2
	sspC
Hyaluronate lyase	hysA
Lipase	lip
	geh
Serine protease	splA
	splB
	splC
	splD
	splE
	splF

22

Serine V8 protease	sspA	
Staphylocoagulase	coa	
Staphylokinase	sak	
Thermonuclease	nuc	***16***
Host Immune evasion (52 genes)		
Exoprotein SCIN	scn	
IgG-binding protein	isb	***2***
Capsule Type 1(A-C), 5(A-P) and 8(A-M and P)	capA	
	capB	
	capC	
	capD	
	capE	
	capF	
	capG	
	capH	
	capI	
	capJ	
	capK	
	capL	
	capM	
	capN	
	capO	
	capP	***50***
Secretion system (8 genes)		
Type VII secretion system	esxA	
	esaA	
	essA	
	esaB	
	essB	
	essC	
	esaC	
	esxB	***8***
Toxins (59 genes)		
Alpha hemolysin	hla	
Beta hemolysin	hlb	
Delta hemolysin	hld	
Enterotoxin A (SEntA)	sea	
Enterotoxin B (SEntB)	seb	
Enterotoxin C (SEntC)	sec	
Enterotoxin G (SEntG)	seg	
Enterotoxin H (SEntH)	seh	
Enterotoxin I (SEntH)	sei	
Enterotoxin K (SEntK)	see	

Enterotoxin L (SEntL)	sel
Enterotoxin M (SEntM)	sem
Enterotoxin N (SEntN)	sen
Enterotoxin O (SEntO)	seo
Enterotoxin P (SEntP)	sep
Enterotoxin Q (SEntQ)	seq
General Enterotoxin	SEnt
Enterotoxin Yent1	yent1
Enterotoxin Yent2	yent2
Enterotoxin-like	SEnt-like
Exfoliative toxin type A	eta
Exotoxin/superantigen-like proteins	set1
	set2
	set3
	set4
	set5
	set6
	set7
	set8
	set9
	set10
	set11
	set12
	set13
	set14
	set15
	set16
	set17
	set18
	set19
	set20
	set21
	set22
	set23
	set24
	set25
	set26
	set30
Superantigen-like	sal
Exotoxin	SExo
Gamma hemolysin	hlgA
	hlgB
	hlgC

Leukocidin M	lukM
Leukotoxin D	lukD
Leukotoxin E	lukE
Panton-Valentine leukocidin	lukS-PV
	lukF-PV
Toxic shock syndrome toxin	tsst

59
151

National Food Institute
Technical University of Denmark
Mørkhøj Bygade 19
DK - 2860 Søborg

Tel. 35 88 70 00
Fax 35 88 70 01

www.food.dtu.dk

ISBN: 978-87-93109-06-3