UiT The Arctic University of Norway

Faculty of Health Sciences
Department of Clinical Medicine

Exploring the pangenome of Staphylococcus haemolyticus

Colonisation, hospital adaption, pathogenicity and novel species identification

--

Maria Pain

A dissertation for the degree of Philosophiae Doctor

A dissertation for the degree of Philosophiae Doctor

Exploring the pangenome of *Staphylococcus haemolyticus*

Colonization, hospital adaptation, pathogenicity and novel species identification

MARIA CHARLENE RONESSEN PAIN

xxxx 2020

Research Group of Paediatric Infections

Department of Clinical Medicine

Faculty of Health Sciences

UiT – The Arctic University of Norway

Contents

Acknowledgements

This PhD has been a challenging, yet extremely rewarding and exciting journey. I am deeply grateful to everyone who has been involved and contributed to any aspect of this process.

First and foremost, I would like to thank all of my supervisors, Pauline Cavanagh, Erik Hjerde, Claus Klingenberg, Elizabeth Fredheim and Trond Flægstad. Thank you for giving me the opportunity to be part of your research team and for letting me take part in this highly interesting project. I am truly grateful for all your guidance and support during the last few years.

I would like to express my sincerest thanks to my main supervisor Pauline Cavanagh. Thank you Pauline for your infectious enthusiasm, immense patience and most importantly for all your help. I really do appreciate that you have always kept your door open for me both professionally and personally, whether that be something small or a larger discussion. I will always be thankful for you giving me the freedom to explore my ideas and for helping me develop into an independent researcher. I am profoundly grateful for all your contributions during my PhD and especially the insight you were able to provide for my writing.

I would like to thank my co-supervisor Erik Hjerde for sharing his knowledge in bioinformatics with me. I felt like a total novice in this area at the beginning of the project but now feel a sense of competence and this is largely thanks to you Erik. Your refreshing approach in teaching me was important in improving my understanding in this area. Thank you for never making me feel stupid even when I had the silliest of questions, and for taking your time with me while explaining things.

I would like to thank Elizabeth for the very warm welcome she gave me when I first started my PhD. Thank you Elizabeth for providing me with a great introduction to the project, for sharing your knowledge with me and for all your contributions and guidance in both the manuscript and thesis writing.

My gratitude goes to co-supervisors Trond and Claus. Trond, thank you for your encouragement and support. Claus, thank you for the incredibly useful feedback and massive help with writing: your experience was invaluable.

I would like to offer my thanks to Jack Bruun and Ilona Urbarova for giving me their time and for providing their excellent feedback for my proteomics inquiries, it really improved my understanding of the subject.

I am so thankful for the girls in my office; Dagny, Runa, Inger Therese and Aline – You have all provided me with a wonderful working environment. Dagny, my biggest cheerleader, I am so grateful you came into “my” office and my life! You are just such a thoughtful, interesting and funny person – and equally as messy as me. Runa, thank you so much for all your help - your lab and organizational skills are impeccable, and for being such a marvelous colleague and friend! Inger Therese, our own cheerful and sweet encyclopedia, thank you for always being willing to help and forever having the patience to explain, and for being so wonderfully

quirky. Aline, our newest office inhabitant, thank you for all of our pleasant chats and for your support, already eagerly anticipating our stays at the Mortenhals resort!

I might not have spent a huge amount of time in the lab, still, I know that the people there are great, and I really want to extend my thanks to the 9th floor "lab gang"- getting to know you guys has been amazing. All the social events have been an essential part of enjoying this journey. Clement, thank you for being my friend from the very beginning, seeing past my shyness, and for all of our interesting discussions – those properly scientific, and those much less so... Adriana, my dear friend, my sister from another mister, getting to know you has really been instrumental in getting me to where I am today. Thank you so much for your positivity, your heartfelt help, your honesty, the very best hugs and for your friendship. Hermoine, I am delighted that I have gotten to know you over the past year. You have been a fantastic addition to our group, and I find you so inspiring. Thank you so much for your generousity, and for your valuable input to both manuscripts and thesis. Jessin, I am so grateful for your help on bash scripts – your kind help and explanation is what made me want to learn it and helped spark my passion. It made a huge impact when working with 170 genomes.

To all the fellow contestants in FGP 2019, to the mentors Audun Hetland and Malcolm Jobling, and to Stine Holand and Irene Mercer; thank you all for the most amazing and inspiring experience. It made me see both myself and my project with renewed eyes and it gave me the extra fuel I needed for the rest of the writing process.

To my wonderful friends, both near and far, for always believing in me and for being so understanding. Susanne, Kjerstin, Lena, Siri, Pia, Julie – Thank you for being such amazing people.

There is one friend that needs an extra special mention: my dear Sunniva. You have been absolutely invaluable to me during my PhD, as a discussion partner, as a life coach and as a true and honest friend - your support in all aspects of life have been crucial. I don't know what I would have done without you.

To my family, both here and in the UK, thank you for all your encouragement and support. A very special thanks goes to my silly, somewhat annoying, sometimes clever and always loving husband. Thank you so much, dear Jonathan, for always being understanding, for letting me talk about my studies when you'd rather do anything else, for giving me freedom, and for never trying to change me. I love you, always.

And finally, my sweetest daughter Annie, I am so grateful to have such an understanding and loving child, and for thinking your mum has the most important job in the world.

Summary

Staphylococcus haemolyticus is a ubiquitous bacterium recognised as a significant cause of nosocomial infections. Today it represents one of the most clinically relevant staphylococcal species, hallmarked by its extreme genome plasticity and multiresistant phenotype. Despite its relevance, little is known about its virulence repertoire or how it causes disease. The main objective of this PhD thesis was to achieve a better understanding of S. *haemolyticus* colonization, virulence and hospital adaption by using a combination of comparative genomics and phenotypic assays.

In **paper I**, by using whole genome sequencing and pangenome analysis, we showed a clear phylogenetic separation between clinical and commensal isolates, and a distinct genomic signature of clinical *S. haemolyticus* isolates. Eighty-eight percent of the clinical isolates were multidrug resistant, compared to only 11 % of the commensal isolates. Clinical isolates typically carried the resistance genes *aacA-aphD* and *mecA*, in addition to the associated IS element IS256. These three genes were absent in most of the commensal isolates and were therefore good candidates for markers to differentiate the two groups. Additionally, *sraP* and the polysaccharide capsule operon, important virulence and immune evasion factors in other staphylococcal species, were more often detected in clinical isolates.

In **paper II**, we compared the adhesive and biofilm forming properties of clinical and commensal *S. haemolyticus* isolates, and showed that clinical isolates formed a thicker biofilm. By developing a novel method for investigating surface proteins expressed during human host colonization, we identified several surface proteins, with potential roles in colonization (sdr-like proteins, SceD), biofilm formation (Atl, Ebh) and immune evasion (TirS and SasH-like).

In **paper III**, we described a new species of the *Staphylococcus* genus; *Staphylococcus borealis*. The novel species was closely related to *S. haemolyticus*, but compared to *S. haemolyticus*, *S. borealis* showed considerable phylogenetic distance, yellow pigmented phenotype and the ability to produce urease.

In conclusion, these studies have greatly advanced our knowledge of *S. haemolyticus* and its potential as a nosocomial pathogen. We have uncovered several potential markers which can distinguish clinical and commensal isolates and potentially be used as diagnostic markers of invasive disease. We also identified several important colonisation,

virulence and immune evasion factors in *S. haemolyticus*, - some of which may be possible future targets for therapy. Finally, we demonstrate the high discriminatory power of whole genome sequencing by identifying a new staphylococcal species, now described as *Staphylococcus borealis*.

Abbreviations

ANI	Average Nucleotide Identity
ARG	Antibiotic Resistance Gene
CC	Clonal Complex
CFU	Colony Forming Unit
CONS	Coagulase Negative Staphylococci
CDS	Coding Sequence
CWA	Cell Wall Anchored (proteins)
DDH	DNA-DNA hybridisation
DNA	Deoxyribonucleic acid
eDNA	Extracellular deoxyribonucleic acid
GI	Genetic Island
HGT	Horizontal Gene Transfer
IS	Insertion Sequence
MDR	Multidrug Resistant
MGE	Mobile Genetic Element
MIC	Minimum Inhibitory Concentration
MSCRAMMs	Microbial Surface Components Recognizing Adhesive Matrix Molecules
MLST	Multi Locus Sequence Typing
OGRI	Overall Genome Related Index
Orf	Open reading frame
PFGE	Pulsed field gel electrophoresis
PIA	Polysaccharide Intercellular adhesin
PSM	Phenol Soluble Modulin
SCC	Staphylococcal cassette chromosome
SE	Staphylococcal Enterotoxins
SraP	Serine rich adhesin for platelets
TIR	Toll/Interleukin 1 Receptor
TirS	TIR-domain containing protein of *S. aureus*?
WGS	Whole Genome Sequencing

List of papers

This thesis is based on the following three manuscripts:

Paper I

Maria Pain, Erik Hjerde, Claus Klingenberg, Jorunn Pauline Cavanagh (2019).

Comparative genomic analysis of *Staphylococcus haemolyticus* reveals keys to hospital adaptation and pathogenicity. *Front Microbiol*. 10: 2096.

Paper II

Runa Wolden, Maria Pain, Roger Karlsson, Anders Karlsson, Elizabeth G. Aarag Fredheim, Jorunn Pauline Cavanagh (2020).

Identification of surface proteins in a clinical *Staphylococcus haemolyticus* isolate by bacterial surface shaving. Manuscript submitted January 2020

Paper III

Maria Pain, Runa Wolden, Daniel Jaén-Luchoro, Francisco Salvà-Serra, Roger Karlsson, Claus Klingenberg, Jorunn Pauline Cavanagh (2020).

***Staphylococcus borealis* sp.nov. – a novel member of the *Staphylococcaceae* family isolated from skin and blood in humans.** Manuscript in preparation.

1 Introduction

Improvements and development in technology over the past few decades have changed the way we live our lives. Demographic and medical developments – such as electronic documentation, faster and better diagnostic tools and improved medical equipment – have allowed us to live longer and to survive conditions that previously had a deadly outcome. As a result, these advances have created more elderly, multimorbid and immunocompromised patients - a growing patient group more vulnerable than their healthy counterparts. These patients and the increased use of inserted medical devices have contributed to the progressive increase of nosocomial infections[1,2]. Perhaps the most important and significant discovery in the medical field goes back to 1928 when Alexander Flemming discovered penicillin[3]. Antibiotics, which have played a crucial role in preventing premature death and other complications, have in addition to treating bacterial infection, been important drugs to prevent infection during operations and other medical procedures.

Today we are on a fast track towards the feared post-antibiotic era. In short, this means that there are already many bacteria resistant to all available antibiotic drug classes, and this resistance is spreading fast through the bacterial populations[4–6]. At this speed we may in the near future find ourselves in a position where previously easily curable infections can no longer be treated[7]. The development of new antibiotics is time consuming and expensive[8]. Since the 1980s only two new antibiotic classes (lipopeptides and oxazolidinones) have been developed and approved by regulatory authorities[9]. In addition, inappropriate use of especially broad-spectrum antibiotics eradicates beneficial bacterial species of the normal flora, enabling other more harmful bacteria (e.g *Clostridium difficile*) to blossom[10]. Advances made in technology have increased our knowledge about microbes, and much focus has been given to those living in and on us, our microbiome. Research on the microbiome is a new and rapidly expanding field and novel findings have changed the way we look at medicine. The knowledge that not all microbes are bad has long been accepted, but how important bacteria are for us, and the plethora of important tasks they perform is a new and extremely interesting field. We find bacteria in all environments, from hot springs to space stations[11,12]. Bacteria are able to adapt extremely rapidly, and some bacteria have adapted to the harsh environment of hospitals. To survive in the clinical environments the bacteria need to overcome all the potential threats they encounter, including antibiotics and antiseptics. Problems arise when nosocomial bacteria, especially those adapted to the hospital settings, cause infections. Hospital

adapted bacteria are often extensively resistant and can therefore be difficult to treat[13,14]. The commensal bacterium *Staphylococcus haemolyticus,* on which we have focused our research on, has the ability to adopt to and thrive in the hospital environment.

In this project, we investigated the genomic composition and surface proteome of *S. haemolyticus*. By performing comparative analysis of commensal and clinical isolates of *S. haemolyticus* we identified factors likely to be involved in *S. haemolyticus* hospital adaption. The identified factors included mobile genetic elements (MGEs), genes involved in virulence, immune evasion and antibiotic resistance which will be described in more detail below. Initial genomic analysis also revealed a new staphylococcal species, for which we have proposed the name *Staphylococcus borealis*. The bioinformatical and proteomic rationale and tools used for species determination are introduced in this thesis. Our findings will contribute to the understanding of the virulence potential, the surface proteins and the population structure of *S. haemolyticus*.

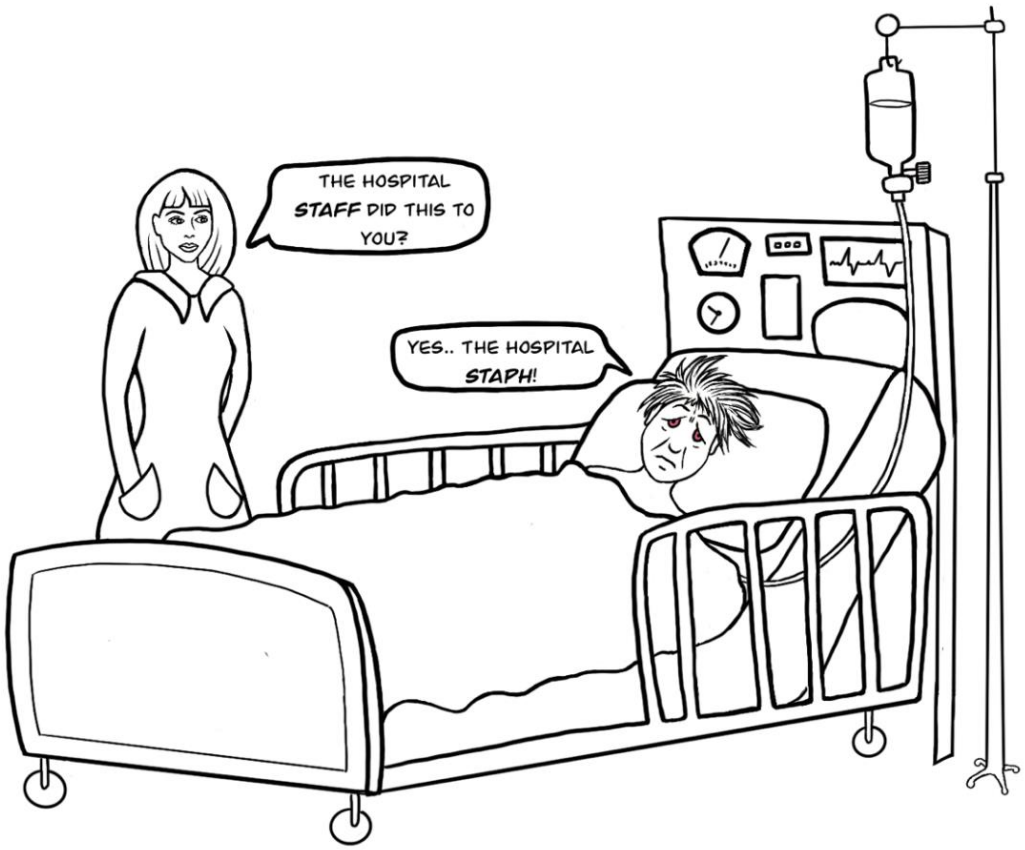


Figure 1: The hospital staph. Redrawn and adapted based on image from[15].

1.1 *Staphylococcus*

Staphylococcus is a genus of Gram-positive cocci with the following taxonomic classification:

- Domain: Bacteria
- Phylum: Firmicutes
- Class: Bacilli
- Order: Bacilliales
- Family: Staphylococcaceae
- Genus: *Staphylococcus*

The *Staphylococcus* genus was first defined by Rosenbach in 1884[16]. Today the genus includes 54 validated species and 22 validated subspecies[1]. Three recent publications propose three new coagulase-negative staphylococcal species; *Staphylococcus caeli,* isolated from air in an industrial rabbit holding, *Staphylococcus debuckii*, isolated from bovine milk, and *Staphylococcus pseudoxylosus,* isolates from bovine mastitis. These three staphylococcal strains are not yet on the updated list from bacterionet, but included here in to total number of strains[17–19]. Under the microscope staphylococci appear spherical, 0.5-1.5 um in diameter, and characteristically form irregular grape-like clusters. They are non-motile. The cell wall contains peptidoglycan and teichoic acid. Staphylococci are usually un-encapsulated or have limited capsule formation. Most staphylococci are facultative anaerobe (exceptions; *S. aureus subsp. anarobius* and *Staphylococcus saccharolyticus*). Staphylococci are usually catalase positive (exceptions; *S. aureus* subsp. *anaerobius* and *S. saccharolyticus*, in addition, some strains of catalase negative of *S. epidermidis* and *S. aureus* have been reported) and oxidase negative (exception; *S. sciuri*). Most strains grow in the presence of 10 % NaCl and between 18 and 40 °C[20]. Host or niche range may be narrow or wide, depending on the particular species or subspecies. Some species are commonly isolated from environmental sources (fomites, soil, air, water) and animal products (meat, milk, cheese). Some species are opportunistic pathogens of humans and animals[20].

1 Based on Staphylococcus species from bacterio.net. The following species and subspecies were removed as they have been either moved or combined with other species: S. *caselyticus* (to *Macrococcus caseolyticus*) *S. jettensis* (to *S. petrasii* subsp. *jettensis*), *S. pulveri* (combined with *S. vitulinus)*, *S. hyicus* subsp. *chromogenes* (to *S. chromogenes*), *S. hyicus* subsp. *hyicus* (to *S. hyicus*) and all subspecies of *S. sciuri* combined into *S. sciuri* (with no subspecies).

The staphylococci are often divided into two main groups; the coagulase negative and the coagulase positive, depending on the presence of the enzyme coagulase causing the fibrin of the blood plasma to clot. The coagulase positive staphylococci comprise a small group of species (*S. aureus*, *S. intermedius*, *S. pseudointermedius*, *S. delphini*, *S. lutrae*, *S. schleiferi* subsp. *coagulans*, and coagulase-variable *S. hyicus*), with *S. aureus* being the most clinically important member. The remaining staphylococci are coagulase negative (CoNS), and the most clinically relevant species are *S. haemolyticus*, *S. epidermidis*, *S. lugdunensis* and *S. saprophyticus*[21].

Staphylococci are bacterial species with a low G+C content, ranging from 27 to 41 %, and the average genome size is between 2 and 3 Mbp. Takeuchi and colleagues identified a unique region in the staphylococcal chromosome just downstream of the origin of replication that showed little homology among the *S. aureus*, *S. epidermidis* and *S. haemolyticus* species, and named this the *oriC* environ[22,23]. The *oriC* environ contains integrated copies of the staphylococcal cassette chromosome (SCC) in its left part, and genes encoded on MGEs typically carrying virulence factors, antibiotic resistance genes (ARG) and specie specific genes, in its right part[22,24,25]. Genes encoding protein A (*spa*), coagulase (*coa*), and the capsule operon (*cap*5/8 A to P), all important virulence determinants of *S. aureus*, were all contained in the right part of the *S. aureus oriC* environ[22]. In *S. haemolyticus,* genes found in the right part of *oriC* were capsule operon homologs. The region does not contain genes essential for bacterial viability, since most of the region can be deleted without affecting growth. The diversifying power of the *oriC* environ has likely been a driving force for the generation of staphylococcal species capable of survival within the human host. In addition, Takeuchi *et al.* proposed that the SCCs serve as efficient vehicles for the introduction of exogenous genes into the *oriC* environ and that abundant insertion sequences (IS) and other recombinases within the region serve as the machinery for excision of genes that are no longer beneficial to the bacterium[22].

1.1.1 Clinical significance

Staphylococci, and mainly *S. aureus*, have long been considered important human pathogens. CoNS species on the other hand were previously regarded as harmless skin flora inhabitants, but over the last decades CoNS have emerged as important opportunistic pathogens. The mode of infection differs between the different species; *S.*

aureus has a wide arsenal of virulence factors, including wide range of different toxins, capsule and enzymes such as coagulase, staphylokinase, and nuclease[26–28]. Some of the major human infections caused by *S. aureus* are pneumonia, acute endocarditis, enterocolitis, meningitis, bacteraemia, toxic shock syndrome, and abscesses of the muscle, skin, urogenital tract, central nervous system, and various intraabdominal organs[20,29]. The other staphylococci are believed to largely depend on adhesion and biofilm production for infection[20]. Among the CoNS, the following have been associated with infections in humans; *S. epidermidis*, *S. haemolyticus*, *S. lugdunensis*, *S. schleiferi*, *S. saprophyticus, S. simulans*, *S. capitis*, *S. cohnii* and *S. sciuri*. Infections commonly cause by CoNS include; bacteremia, septicaemia, native and prosthetic valve endocarditis, urinary tract infections (UTI), prosthetic joint infection, and also wound, bone and joint infections[20,21,30–32].

During the last decades, an increasing number of susceptible patient groups (premature neonates, elderly, multimorbid, chronically ill, and often immunocompromised patients), has led to an increased recognition of the large variety of infections caused by CoNS[21]. CoNS are today the most prevalent pathogens causing foreign body related infections[21,33] and are often associated with infections in neonates, especially those born preterm[34–38].

1.1.2 *Staphylococcus haemolyticus*

This thesis mainly focuses on one important, and perhaps somewhat "neglected" member of the staphylococci; the ubiquitous *S. haemolyticus*. Primarily considered a skin commensal, *S. haemolyticus* has also gained increased attention as a significant pathogen of nosocomial infections today[39–43]. *S. haemolyticus* is normally considered the second most prevalent cause of infections by CoNS, after *S. epidermidis*[21,44]. Predominantly associated with bloodstream and device-associated infections, *S. haemolyticus* particularly affect individuals with a compromised host defence. Nosocomial isolates of *S. haemolyticus* are characterized by their high levels of resistance to several antimicrobial agents, and is ranked the most antibiotic resistant CoNS species[37,44–49]. The development and spread of ARGs result in limited treatment options in *S. haemolyticus* infections[39,50]. *S. haemolyticus* has been implicated in native valve endocarditis, septicaemia, peritonitis, and UTI, and is occasionally associated with wound, bone, and joint infections[21,38,40,44]. Several studies show *S. haemolyticus* to be frequently recovered from both the hospital and community environment[48,51–55]. *S. haemolyticus* is reported to be preferentially

located on skin sites where apocrine glands are numerous (e.g. in the axillae and inguinal and perineal areas)[21,56]. *S. haemolyticus* has been shown to be a more prevalent part of the microbiome in neonates and children compared to adults, interpreted to be due to hospital stay[57–59].

How *S. haemolyticus* causes infection is still largely unknown, but biofilm production by *S. haemolyticus* has been identified as an important virulence factor, protecting the cells against antibiotics and the immune system[22,60–62]. The first fully sequenced *S. haemolyticus* (JCSC1435) revealed as many as 82 insertion sequences, likely causing the frequent rearrangements observed in this isolate, which probably facilitate uptake of resistance genes and/or other genes[22,47].

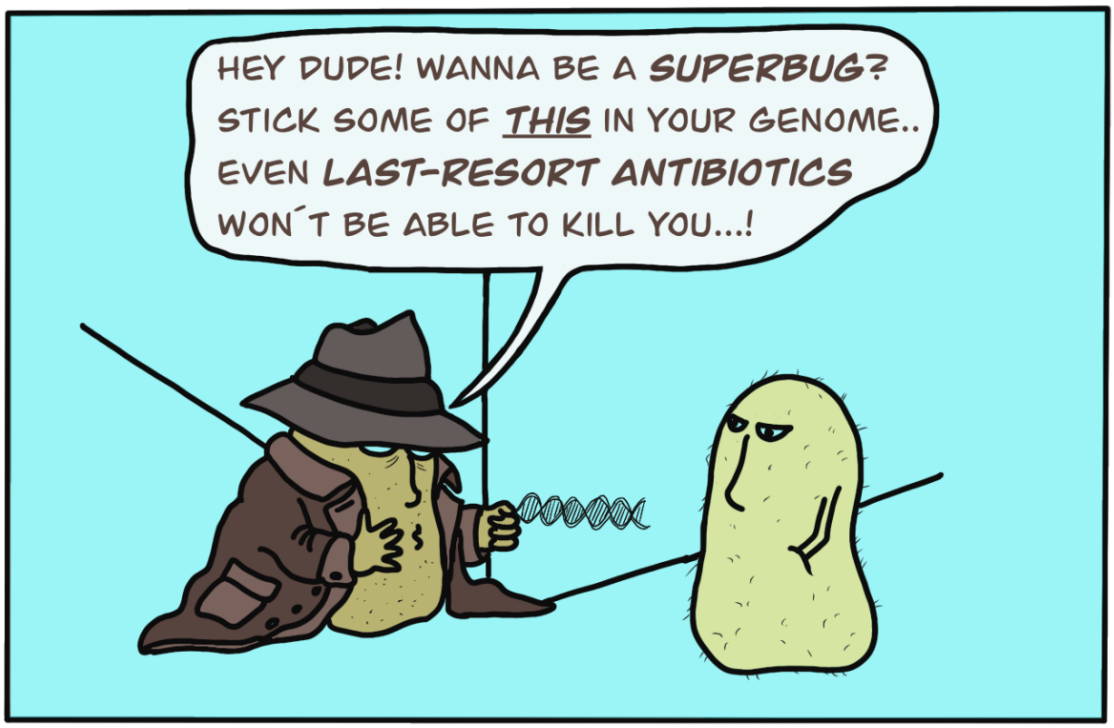


Figure 2: **Multidrug-resistant bacteria.** Redrawn and adapted based on cartoon by Nick D Kim[111].

1.2 Virulence factors and interaction with the host

Despite its rising clinical relevance, little is known about the virulence-associated properties in *S. haemolyticus*. Takeuchi *et al.* analysed the first fully sequenced genome of a human pathogenic *S. haemolyticus* strain in 2005 and reported 57 open reading frames (ORFs) associated with virulence[22]. They identified numerous genes encoding putative surface adhesins, enzymes and toxins. Some virulence properties, such as biofilm formation and haemolysis, have been explored, but often not linked to specific genes[60,62–64]. In the following sections, factors contributing to the disease burden of *S. haemolyticus* are presented.

1.2.1 Antibiotic resistance

Antibiotic resistance is not in itself a virulence factor. However, in certain situations it can be a key factor in development and progress of infection by preventing successful bacterial eradication. High rates of antimicrobial resistance in clinical *S. haemolyticus* have been consistently reported for decades[44,46,47,65,66]. The type of antimicrobial resistance depends on the country or region and the prescription regiments followed in that specific area. Still, multidrug resistant *S. haemolyticus* are reported worldwide[38,47,67,68]. In studies of clinical *S. haemolyticus* isolates, multidrug resistance (MDR – defined as resistance to at least three classes of antimicrobial agents) is generally identified in over 70 % of the strains[47,66,68].

S. haemolyticus shows particularly high resistance rates towards beta-lactams, aminoglycosides and macrolides[46,47,49,66]. Even "last resort" antibiotics such as vancomycin, teicoplanin and linezolid, are not secure treatment options for clearing an *S. haemolyticus* infection[69–73]. *S. haemolyticus* was the first Gram-positive pathogen to acquire glycopeptide resistance, earlier than any other staphylococcal or enterococcal species[66,74–76].

The few studies investigating resistance patterns in commensal CoNS report higher resistance towards penicillin and erythromycin than towards other antimicrobials[77–80]. Our research group analysed the antibiotic resistance patterns of different CoNS isolated from the skin of healthy volunteers[56]. Among the 48 *S. haemolyticus* isolates included in this study, MDR was identified in only 6.3%, which was slightly lower than what was observed for *S. hominis* (10.8%) and *S. epidermidis* (6.4%). However, as many as 60.4% of

these commensal *S. haemolyticus* were resistant to erythromycin[56], higher than reported by others[77,79,80]. A few studies have investigated the antimicrobial resistance pattern of CoNS from environmental samples. *S. haemolyticus* isolated from polluted water in Brazil showed generally low antibiotic resistance, with erythromycin and penicillin resistance being the highest, 27.8% and 33.3%, respectively[81]. Environmental samples of *S. haemolyticus* from a university in Thailand showed high resistance towards erythromycin (76.5%), oxacillin (70.6%) and cefoxitin (70.6%)[48]. In another study looking at bacterial contamination on inanimate surfaces of hotels in London, 86% of *S. haemolyticus* were MDR, being most resistant to fusidic acid (67.7%) and erythromycin (59%). The same study also reported 29.4 % of *S. haemolyticus* to be vancomycin resistant[54]. A trend amongst environmental and commensal *S. haemolyticus* isolates is elevated resistance towards erythromycin compared to other antibiotics.

1.2.2 "Last resort" antibiotics

The glycopeptide antibiotic vancomycin was introduced in clinical use in 1958 for the treatment of Gram-positive bacteria[82]. Shortly after being introduced, vancomycin was eclipsed by antibiotics that were considered to be less toxic and equally or more effective[83]. The use of this agent then has had a dramatic increase in the last 30 years, in large part due to the increasing prevalence of methicillin resistance in both CoNS and *S. aureus*[84–88]. Vancomycin resistance in CoNS was first observed in *S. epidermidis* 40 years ago[89,90]. In 1987 the first report of vancomycin resistance in *S. haemolyticus* was published[91]. Since then several reports on decreased susceptibility to vancomycin have been published[70,92–94]. *In vitro* selection for vancomycin resistant *S. haemolyticus* has also been reported[95,96], and Biavasco *et al.* showed that higher minimum inhibitory concentration (MIC) towards vancomycin were found in *S. haemolyticus* compared to *S. aureus* and *S. epidermidis*[97]. Vancomycin resistant *S. haemolyticus* exhibit typically a thickened cell wall, decreased cell growth phenotype and decreased autolysis activity[98]. The mechanism for decreased susceptibility to vancomycin in staphylococci has been reported to be mainly due to mutations in transcription regulatory genes, including *walkR*, *vraSR*, and *rpoB* genes, which are involved in the regulation of cell wall synthesis and cell envelope stress response. Mutations is these genes alter their activity[99–102]. Complete vancomycin resistance conferred by the *vanA* operon (encoded on transposon Tn*1546*, originally a part of a vancomycin-resistant enterococci conjugative plasmid[103]), has been shown in *S.*

aureus. Resistance due to *vanA* operon in *S. aureus* is rare[104], and to date this operon has not been reported in CoNS.

Linezolid was presented as a valuable agent for the treatment of serious infections caused by antibiotic resistant Gram-positive cocci in 2000[105]. However, emergence of linezolid resistant staphylococci questions the efficacy of this antibiotic[73,106,107]. Over the last decade, several reports have emerged showing linezolid resistant *S. haemolyticus* in India, Spain, Brazil and Italy, and the majority of studies show point mutations in 23S and L3 and L4 ribosomal proteins[106–110]. Although linezolid resistant strains of *S. haemolyticus* are still rare, they are found significantly more often in *S. haemolyticus* compared to *S. epidermidis*[106] and there have been reports of *S. haemolyticus* becoming resistant to linezolid during treatment[73]. Additionally, the developed resistance was high and relatively stable, being unchanged after thirty passages in drug free medium[73].

1.2.3 Capsule

The capsule is the outmost structures of certain bacteria and protects the bacteria from immune cell recognition during infection of mammalian hosts. With the exception of the poly-γ-glutamate capsule of *Bacillus anthracis*, all other known bacterial capsules are composed of polysaccharides[112]. The polysaccharide capsule (CP) of *S. aureus* is extensively described, and its protective properties are well documented[113–117]. The *S. aureus* capsule has been divided into 11 serotypes. Serotype 5 (CP5) and serotype 8 (CP8) are the most common[118], with CP5 being more associated with virulence[115].

Several publications before 1988 described the phenotypic presence of capsule in *S. epidermidis,* including different capsule types, but the specific capsular genes have not yet been described for the species[119–121]. Since then, reports on staphylococcal capsule have been largely limited to *S. aureus*, with a few exceptions; capsule genes have been detected in *S. haemolyticus* and *S. saprophyticus*[122,123].

The presence of a potential capsule in *S. haemolyticus* was first reported by Poutrel *et al.* in 1990. From a collection of 19 bovine *S. haemolyticus* isolates, 13 reacted with monoclonal antibodies of *S. aureus* CP5[124]. Takeuchi *et al.* analysed the genome structure of *S. haemolyticus* JCSC1435, and reported a capsule operon within the *oriC* environ consisting of 13 genes, termed *capA* to *capM*, where the first seven genes, *capA-capG*, showed similarity to *S. aureus capA-capG*. The remaining capsule genes, *capH-capM*,

however, showed no homology to other capsule genes[22]. The presence and function of *S. haemolyticus* capsule operon was further explored by Flahaut *et al*. where they confirmed the presence and also showed its protective features against phagocytosis[123].

1.2.4 Toxins

Many bacteria produce toxins, which play an important role in pathogenicity. Bacterial toxins are categorized into two groups; i) endotoxins, such as lipopolysaccharide (LPS), which is part of the cell wall of Gram-negative bacteria, and ii) exotoxins, which are secreted by the bacteria and are primarily composed of proteins. This chapter focus on different types of exotoxins.

1.2.4.1 Staphylococcal enterotoxins

Staphylococcal enterotoxins (SE) constitute a family of more than 20 different biologically and structurally related staphylococcal and streptococcal exotoxins. These bacterial toxins are pyrogenic (fever-inducing) and are linked to significant human debilities such as food poisoning and toxic shock syndrome[125]. It has been shown that genes encoding SEs often are located on MGEs such as plasmids, bacteriophages or pathogenicity islands[126,127], indicating horizontal transfer of SE genes between strains. SEs are well characterized virulence factors in *S. aureus*[125], but less is known regarding their role in CONS. One study, analysing the prevalence of SE genes among Staphylococci isolated from clinical samples, reported a higher frequency of *S. epidermidis* isolates producing detectable amounts of toxins compared to other CoNS species[128].

In a Brazilian study, Pinheiro and co-workers studied the presence of SE in a collection of *S. epidermidis* and *S. haemolyticus* blood culture isolates[64]. Quite surprisingly, they reported relatively high frequency of SE genes in *S. haemolyticus* (*sea*, *seb*, *sec*, *seg* and *sei;* identified in between 24 % and 61 % of the isolates) in contrast to what had been reported previously[128]. Of note, Brazilian isolates of *S. haemolyticus* are different to isolates from other countries, which will be further remarked upon later in this thesis. Hence, the presence of SE genes in *S. haemolyticus* could be limited to this region, and not reflect the overall toxin profile of this species.

1.2.4.2 Haemolysins and phenol soluble modulins

Haemolysins (another type of exotoxin) are pore-forming lipids or proteins that can cause lysis of red blood cells by destroying their cell membrane. Haemolysins are important molecules involved in the pathogenesis of *S. aureus*[129]. Their role in CoNS infections is still largely unknown, but *S. haemolyticus* haemolytic capacity has been described[64,130].

The haemolysins of staphylococci are classified in four different types including alpha (α)-toxin, beta (ß)-toxin, gamma (γ)-toxin and delta (δ)-toxin. α-toxin, also called α-hemolysin (*hla*), is an important virulence factor of *S. aureus* and is involved in cellular lysis. It has been demonstrated to affect a wide range of human cell types, including epithelial cells, endothelial cells, T-cells, monocytes and macrophages[131–135]. Studies have shown that the expression of Hla may be required for the pathogenesis of invasive disease in healthy individuals[135]. ß-toxin (ß-haemolysin, Hlb) is cytotoxic towards human keratinocytes, polymorphonuclear leukocytes, monocytes and T-cells and inhibits interleukin-8 (IL-8) expression by endothelial cells. These have been shown to contribute to phagosomal escape of *S. aureus* and induction of biofilm formation[136,137]. Beta-toxins are produced in large quantities in chronic human skin infections[138], and several studies demonstrate their importance in *S. aureus* pathogenicity[136,138,139]. γ-toxin, also called γ-haemolysin (HlgA, HlgC, HlgB), is a leukotoxin that exhibits cytolytic activity towards human leukocytes[131,140]. The role of HlgACB in virulence is not well understood, but it has been shown to be required for *S. aureus* survival and proliferation during blood stream infection (BSI)[141,142]. δ-toxin (PSM-γ), is a member of the phenol soluble modulin (PSM) family[143]. Pinheiro *et al.* reported the genotypic and phenotypic presence of haemolysins in *S. haemolyticus*[64].

The phenol soluble modulins (PSM) family are good candidates contributing to the poorly understood virulence characteristics of CoNS. PSMs are a family of pro-inflammatory and cytolytic staphylococcal peptide toxins[144]. Studies in *S. epidermidis* and *S. aureus* have shown that PSMs have multiple functions in staphylococcal physiology and pathogenesis, including sepsis[143–146]. In addition, PSMs have been linked to playing a role in the structure of biofilms and biofilm dispersal[147,148]. The identification of PSM in *S. haemolyticus* dates all the way back to 1984, then described as anti-gonococcal peptides (or inhibitors), due to its toxic effect on gonococcal species[149]. A few years later the structure of these peptides was reported and the toxigenic effect was shown to be a result of the destruction of the gonococcal membrane[150,151]. The structure, role and function of PSMs in *S. haemolyticus* were thoroughly investigated recently by Da *et al*[63]. In addition

to identifying the three already known PSMs, they also identified a new PSM not previously described, which they called PSMα. They found that the haemolytic capacity was paired with pronounced cytolytic capacity toward human neutrophils, indicating a role of particularly PSMα in the immune evasion properties of *S. haemolyticus*[63]. In a recent paper describing for the first time the secretome of *S. haemolyticus*, both PSMα and PSMß were reported as the most prevalent secreted molecules[152].

1.2.5 Staphylococcal surface proteins

Colonisation is often the prerequisite for infection. Adhesion is an essential step in colonization, and also the first step in bacterial biofilm formation. Staphylococci express several surface proteins that can interact with host cell receptors, abiotic surfaces and soluble macromolecules. Many of these surface proteins have been well characterized in *S. aureus*, including function, interaction partners and pathogenicity potential[153]. Surface proteins are covalently attached to the peptidoglycan and are referred to as cell wall anchored (CWA) proteins. These proteins offer vital opportunities for bacteria to interact with the host and are crucial for survival in the commensal state and during invasive infections. Surface proteins carry out a broad range of functions that are essential for the colonisation of, and survival in, the host[153]. The number of CWA proteins on the surface of staphylococcal species varies among species and strains, but *S. aureus* can express up to 25 different CWA proteins[154], while CONS such as *S. epidermidis* and *S. lugdunensis* express a smaller number, 11 and 13 CWA, respectively[155,156]. All ORFs of CWA proteins contain a secretory Sec-dependent signal sequence at the amino terminus and a sorting signal — which comprises an LPXTG sortase cleavage motif, a hydrophobic domain and a stretch of positively charged residues — at the C terminus[153].

Foster *et al*. proposed classifying the CWA proteins into four groups based on the presence of motifs that have been defined by structure–function analysis. The most prevalent group is the microbial surface component recognizing adhesive matrix molecule (MSCRAMM) family, which includes serine-aspartate (sdr) genes, clumping factor, collagen binding and fibronectin binding genes, responsible for adhesion to different types of cells and tissue (Table 1). The second group is the near iron transporter (NEAT) motif family of proteins, which are involved in haem capture from haemolglobin and help bacteria survive in the iron-restricted host milieu. The NEAT family includes Isd proteins, which have not been identified in *S. epidermidis* or *S. haemolyticus*. The third group is the Three-helical bundle

family consisting of protein A, only identified in *S. aureus*. Finally, the last group is the G5-E repeat family which consists of SasG and Pls (a SasG homologue in MRSA) in *S. aureus* which is closely related to the accumulation protein (Aap) in *S. epidermidis*. The expression of CWA proteins can be altered by growth conditions; some proteins are expressed only under iron-limited conditions, while others are found predominantly in cells during exponential or stationary growth phase[157–159].

A collection of the most important surface proteins and their function for *S. aureus* and *S. epidermidis* is listed in tables 1 and 2. Takeuchi *et al.* presented a list of potential virulence ORFs in *S. haemolyticus* which included 13 proposed surface proteins (three of which were truncated)[22]. Neither their prevalence nor their role has to date been further investigated.

TABLE 1: *S. aureus* cell wall anchored proteins. Uncharacterized proteins with no known function are not included. Table adapted from and references can be found in [153].

***Staphylococcus aureus* cell wall anchored proteins**		
Protein class	**Protein group**	**Function**
MSCRAMMS	ClfA	Adhesion to fibrinogen, immune evasion
	ClfB	Adhesion epithelial cells, nasal colonization
	SdrC, SdrD	Nasal colonization?
	SdrE	Immune evasion
	FnBPA, FnBPB	Adhesion to extracellular matrix (ECM), invasion
	Cna	Adhesion to collagen, prevents compliment activation
NEAT motif family	IsdA, IsdB, IsdH	Iron acquisition, adhesion epithelial cells, resistant to lactoferrin, bactericidal lipids and antimicrobial peptides, survival in neutrophils, invasion of non-phagocytic cells
Three-helical bundle	Protein A	Inhibition of opsonophagocytosis, inflammation, endovascular infection, endocarditis
G5-E repeat family	SasG/pls, glycoprotein	Adhesion to epithelial cells, biofilm formation
Structurally uncharacterized	AdsA/SasH	Promotion of survival in neutrophils
	SasX	Biofilm formation, cell aggregation, and adhesion
	SasC	Attachment and biofilm formation
	SraP/SasA	Endocarditis, endovascular infection
	Bap	Biofilm, aggregation on epithelial cells (only in bovine)

TABLE 2: *S. epidermidis* cell wall anchored proteins. Only proteins with proposed functions are included. Table and protein classes adapted from[156,160].

***Staphylococcus epidermidis* cell wall anchored proteins**			
Protein class	**Protein**	**Function**	**Reference**
MMSCRAMs	sdrF	Adhesion to collagen Adhesion to keratinocytes Adhesion to plastic and medical devices	161 162 163
	sdrG	Adhesion to fibrinogen	164
	SesJ	Involved in biofilm formation	166
G5-E repeat	SesF (Aap)	Cellular aggregation Involved in biofilm formation	167 168
Unclassified proteins	SesD (Bhp)	Proposed to have similar function as the *S. aureus* homolog Bap	169
Repeat-containing proteins	SesE	Proposed to be involved in cell aggregation and biofilm formation	170
Non-repeat-containing proteins	SesC	Involved in biofilm formation	171
	SesI	Allows bacterial adherence and colonization	172

1.2.5.1 Accessory sec and *sraP*

One predicted surface protein, streptococcal hemagglutin-like protein (SH0326), has by Takeuchi been listed as one of 57 predicted virulence factors - with loss of this protein leading to loss of agglutination[22]. SH0326 is a homolog of the serine-rich adhesion for platelets for protein, SraP (alternatively termed SasA), and was first described in *S. aureus* in 2005[173]. SraP is part of the uncharacterised group of CWA in *S. aureus*. This protein belongs to a highly conserved family of serine-rich surface glycoproteins of Gram-positive bacteria, including the first serine-rich protein identified, Fab1 in *Streptococcus parasanguinis*[174] and the currently best characterized GspB in *Streptococcus gordonii*[175]. These are large glycosylated proteins that are not efficiently transported by the canonical (the traditional system responsible for translocation of most secreted proteins) secretion

system[176]. Instead, a dedicated transport system, called the accessory secretion system (accessory sec) is required to efficiently translocate these proteins to the bacterial surface. The accessory sec system was first identified in mycobacteria less than 20 years ago[177] and has been most extensively studied in Streptococcus[178–181]. In figure 3 the different accessory sec systems are described. The expression of the serine rich glycoproteins has been linked to adhesion to different types of cells, including human platelets[173,180], epithelial cells[182], salivary components[183], pneumocytes[184,185], in addition to being implemented in biofilm formation[186,187].

In a study published in 2009, all the sequenced *S. aureus* genomes at the time contained SraP-like serine-rich repeats, highlighting their importance[178]. To date no papers dedicated to describing the SraP homolog in *S. haemolyticus* or any other staphylococci than *S. aureus* have been published, but a review from 2009 describes and compares different Gram positive SraP homologs and their accessory sec systems, including that of *S. haemolyticus,* but only based on genomic information[178].

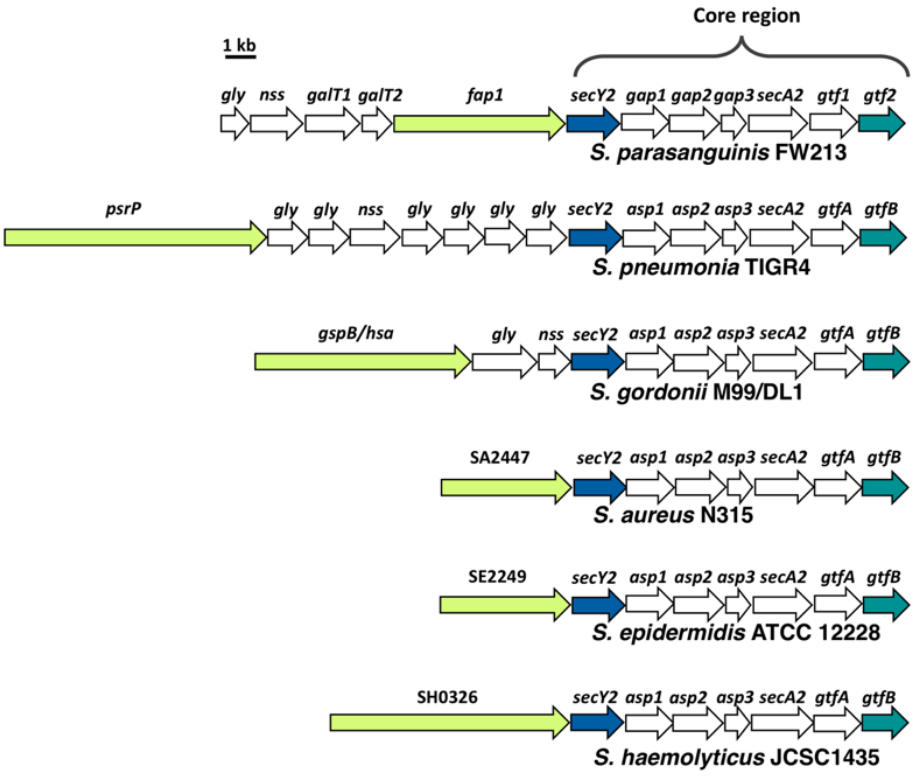


Figure 3: Accessory secretion system. Conservation of serine-rich glycoproteins and proteins contributing to their biosynthesis and secretion in streptococci and staphylococci, adapted from[178].

1.2.6 Biofilm

Biofilm has been defined as "aggregates of microorganisms in which cells that are frequently embedded within a self-produced matrix of extracellular polymeric substances (EPS) adhere to each other and/or to a surface"[188]. The biofilm matrix consists of extracellular polysaccharide, proteins, DNA and lipids, and biofilms from different bacterial species have different compositions. Biofilms are not uniform cultures of physiological identical cells, and different mechanisms can protect subpopulations of cells in the biofilm[189]. Within the biofilm, cells with different genotypes and phenotypes coexist. This heterogeneity is a result of distinct metabolic pathways being expressed based on the local environmental conditions within the biofilm. The metabolic activity of bacteria is higher in the outer parts of the biofilm and lower in the inner parts[189]. The physiological conditions in biofilms can be stressful due to the scarcity of nutrients, excess waste products, hypoxia and antimicrobials, and hence antimicrobial resistance development in biofilm may partly reflect cells responding to stress which promotes mutations[190]. In addition, the high density of bacterial cells in biofilm increases the spread of plasmids by conjugation[191].

Infections caused by biofilm-forming bacteria can be difficult to eradicate with antibiotics, and a 10 to 1000 fold increase in antimicrobial tolerance compared to planktonic cells have been reported[192]. Antibiotic tolerance mechanisms in biofilms includes failure of antibiotics to penetrate biofilms and slower growth rate of bacteria embedded in the biofilm[193,194]. The ability to produce an adherent multi-layered biofilm on implanted devices is considered an important virulence factor of staphylococci, and biofilm production has been well studied in *S. aureus* and *S. epidermidis*[195]. The best studied and well characterized type of staphylococcal biofilms is the polysaccharide intercellular adhesin (PIA), encoded for by the *ica* genes operon[195]. Other *S. epidermidis* genes associated with biofilm formation includes *aap/sesF, embp/ebh*, and *bhp/bap/sesD*[168,169,195–197]. *S. epidermidis* AtlE (Atl in *S. aureus*) is important in cell wall turnover and binding to both unmodified and vitronectin covered surfaces[196,198] and has been shown to have a significant role in eDNA mediated biofilm[199,200]. Several CWA proteins have also been implicated in biofilm formation, as show in table 1 and 2.

Many studies have looked at *S. haemolyticus* biofilm production, and most have demonstrated that *S. haemolyticus* has the ability to produces biofilm[38,60,201]. However, in contrast to *S. aureus* and *S. epidermidis,* the genetic mechanisms and triggers for biofilm formation has not yet been clearly identified. Fredheim *et al.* showed that *S. haemolyticus*

produces a biofilm primarily made of DNA and proteins, and not polysaccharides as seen in *S. epidermidis*[60]. A simplified representation of bacterial biofilm is shown in figure 4.

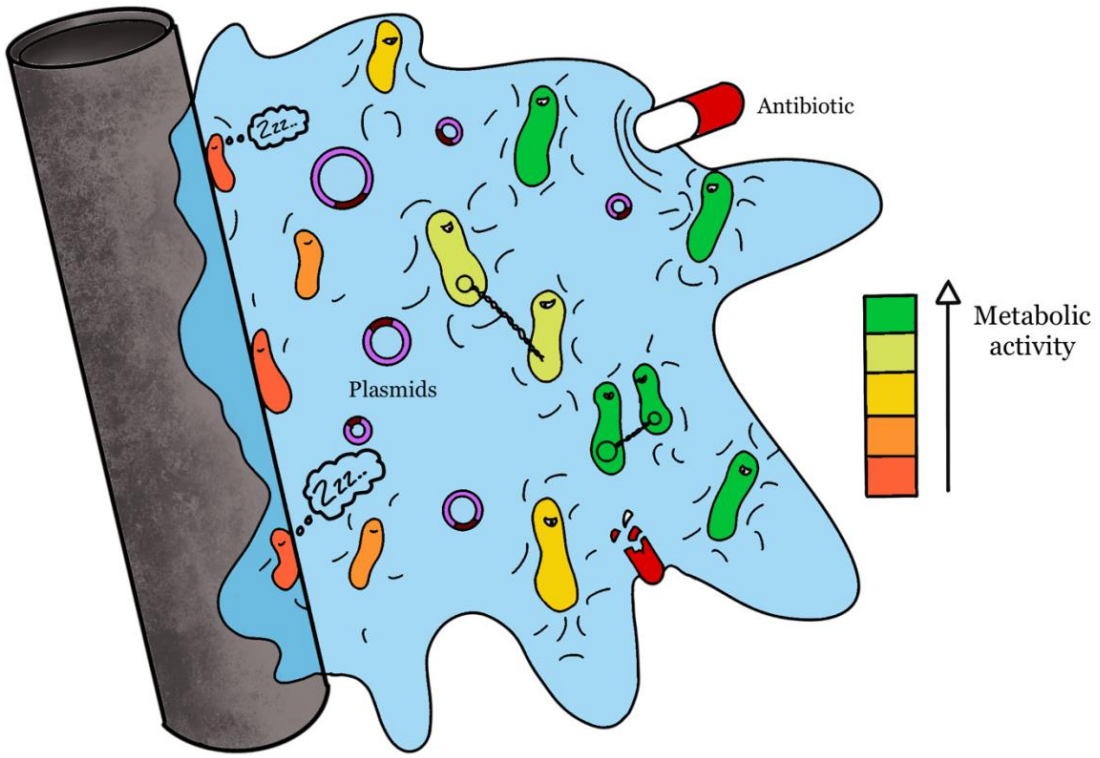


Figure 4: Life in biofilm. Bacteria in biofilm have extra protection against antibiotics. The antibiotic may fail to penetrate beyond the surface layer of the biofilm and there might be zones of nutrient depletion or waste product accumulation preventing the antibiotics mode of action. The biofilm is also ideal for horizontal gene transfers, as the bacteria are in close proximity to one another. In addition, some bacteria might enter a different metabolic state, like the small colony variants (SCV - indicated in deep orange colour)[61].

1.3 Pangenome

Comparative studies in microbiology seek to describe and explain population structures. Can strains be separated by ecology, geographical boundaries, pathogenicity, type of resistance or by other traits?

The first definition of the pangenome was proposed by Tettelin *et al.* in 2005, shortly after the beginning of the high-throughput sequencing era[203]. A pangenome can be defined as being the entire gene content belonging to a study group. It includes genes present in all

strains (core genome) and genes present only in some strains of a species (accessory genome)[204,205]. The applications of pangenome analysis are vast, including the study of pathogenicity MGE, ARGs, bacterial lifestyle and taxonomy[206–209]. Pangenome studies have been suggested as a tool for classification of novel species as well as redefining known species, by looking at the ratio between core and pangenome. A break in this ratio means there is no transition from one species to another, leading to the definition of different species[209,210]. A pangenome can be defined as open or closed[204], according to the species' capacity to acquire exogenous DNA and to have the machinery to use it[211,212]. The open or closed nature of a pangenome is often bound to the lifestyle of the bacterial species[211–213]. Bacterial species living in a narrow niche usually have small genomes and a closed pangenome, because they are specialised[212,213]. Bacterial species living in a community tend to have larger genomes and an open pangenome and a high HGT rate[214].

As an extension of pangenomics, bacterial pangenome-wide association studies (panGWAS) pairing phenotypes with SNPs, presence and absence of genes and other regions like regulatory sequences have begun to emerge[215–217].

1.3.1 The core genome

There is a high degree of conservation of genes among staphylococci, as shown by genome comparisons of *S. haemolyticus*, *S. aureus* and *S. epidermidis*[22]. Comparing the orfs of these three species revealed that 1158 orfs were present in all three species[22]. A recent, more comprehensive comparative study, focused on the genomes of 182 *S. aureus* and 143 *S. epidermidis,* and showed a conservation of 1478 genes, found in all these isolates. The genes shared between the species comprised 52 % of total genome for *S. aureus* and 56 % for *S. epidermidis*[218].

The core genome is defined as the pool of genes common to all the studied genomes of a given species[214]. In general, the core genome includes all genes responsible for the basic aspects of the biology of a species and its major phenotypic traits[211]. The core genome of *S. epidermidis* has been reported to be approximately 80 % [207,218], while the core genome of *S. aureus* has been shown to be smaller (55-70 %)[218,219].

1.3.2 The accessory genome

The accessory genome (also called the variable/flexible/dispensable genome) contains the 'dispensable' genes present in a subset of the strains. The accessory genome encodes non-essential genes, typically providing selective advantages to their host in a particular environmental niche, for example genes encoding virulence factors and/or resistance determinants. Many of these are probably acquired by HGT. HGT enables bacteria to disseminate genes among related and unrelated bacterial species, which is important for the adaption to new niches or challenges such as antimicrobial pressure and virulence genes[220]. HGT is dependent upon interactions between MGE, such as insertion sequences (IS) and transposons – mediating intracellular movement of DNA- and plasmids, bacteriophage and integrative and conjugative elements that promote intercellular DNA mobility.

Comparative analysis of WGS of different staphylococcal species has shown that the accessory genes constitute approximately 22-45% of the total genome for *S. aureus*[218,219,221], 20-22% for *S. epidermidis*[207,218] and 19% for *S. lugdunensis*[222]. Acquisition, maintenance and dissemination of accessory genes have been central in the ongoing success of staphylococci as pathogens, and *S. aureus* is reported to carry most of their virulence genes on MGE[25,223,224].

DNA can be introduced into staphylococci by each of the three classical bacterial gene transfer mechanisms; transformation, transduction and conjugation. Transformation is thought to be limited by extracellular nucleases and is usually very inefficient[220], although more recent research suggests that natural competence might arise under suitable growth conditions and/or in subpopulations of cells [225]. Besides the three main mechanisms of HGT, genetic material can also be transmitted through membrane vesicles (MV)[226–229].

Identical or nearly identical (homologues) accessory genes, elements, and plasmids have been detected in different staphylococcal species and other bacterial genera, like enterococci and streptococci[230,231]. These observations suggest that gene transfer mechanisms operating in staphylococci aid not only gene transfer among staphylococci, but also interspecies exchange and hence gives bacteria access to an extended and shared reservoir of genes. *S. aureus* is reported to carry most of their virulence genes on MGE[25,223,224].

1.3.2.1 Plasmids

Plasmids are double-stranded circular or linear DNA molecules capable of autonomous replication and can be transferred between different bacterial species and clones. Many of the known plasmids have been identified as they confer phenotypes that are subject to positive selection in the recipient bacteria, such as the presence of virulence or ARGs[232,233]. These features promote the successful spread of different plasmid types among bacteria from different sources and geographical origins[233].

Transfer of plasmid DNA in staphylococci occurs mainly through the process of conjugation; a process where a donor cell makes contact with a recipient cell and directly transfer DNA into the recipient cell. For plasmid translocation, a conjugation apparatus is required and conjugative plasmids carry all the genes needed for plasmids translocation. In staphylococci three distinct families of conjugative plasmids have been reported, namely the pSK41, pWBG749 and the pWBG4 family[234]. Another type of plasmids, mobilizable plasmids, carry the DNA-transfer genes required for mobilization, but lack genes required for translocation[235]. Mobilizable plasmids can take advantage of conjugative plasmids for horizontal transfer and dissemination, but are non-mobile in cells that lack mobile elements carrying compatible mating-pore genes[234]. A genome survey of staphylococcal plasmids isolated since the 1940s has shown that most staphylococcal species carry at least one plasmid over >20 kb, for which the most common families were pMW2, pIB485 and pUSA300HOUMR, representing 43% of all plasmids in the 20-30 kb size range. These plasmids all lack conjugation or mobilization genes[235]. In fact, only around 5–6% of *S. aureus* plasmids are conjugative, but it appears that the majority of non-conjugative plasmids, including most large MDR-plasmids, are potentially mobilizable[234], taking advantage of conjugative plasmids for horizontal transfer and dissemination. Recent research has demonstrated a novel method of mobilizing non-conjugative staphylococcal plasmids without mobilization genes. Instead these plasmids carry one or several sequences (*oriT* mimics) similar to recognition sequences for mobilization *(oriT)* in conjugative plasmids[234]. These mimic sequences can be recognised by the conjugative plasmids' conjugation machinery and further be translocated. While the prevalence of conjugative plasmids in isolates of *S. aureus* is low, the presence of *mob* and *oriT* sequences by most non-conjugative plasmids demonstrate that conjugative mobilization is an event frequent enough for most *S. aureus* plasmids to have evolved to take advantage of. Carriage of an *oriT* mimic likely has even a smaller

impact on plasmid size than *mob*-gene carriage and the accumulation of multiple *oriT* mimics likely increases the opportunity for transfer[234].

Takeuchi *et al.* described three extrachromosomal plasmids (assigned pSHaeA, pSHaeB and pSHaeC) in *S. haemolyticus* JCSC1435[22]. pSHaeA (2,300 bp) and pSHaeB (2,366 bp) only encoded the replication apparatus and resistance genes *fosB* and *ermC,* respectively. pSHaeC (8,180 bp) carried two detoxification related genes. These findings showed that accumulation of MDR-plasmids contributes to MDR in *S. haemolyticus* JCSC1435. To date, no studies have identified conjugative plasmids in *S. haemolyticus*.

1.3.2.2 Transposable elements

Transposable elements are DNA sequences that can change their position within a genome, occasionally creating or reversing mutations and changing the cell's genetic identity and genome size. I will here focus on insertion sequences (IS) (also known as IS elements) and transposons.

IS elements are the simplest transposable element. The main characteristics of IS elements are that they are small in comparison to other transposable elements and only code for proteins involved in the transposition activity. The ends of the IS elements are usually inverted repeat sequences. Two IS elements inserting relatively near each other allow the entire region to be transposable and thus further promoting the potential for genetic exchange in a bacterial population. IS elements are different from composite transposons, which also carry accessory genes such as ARGs[220,236]. In *S. haemolyticus* JCSC1435, as many as 82 IS elements were detected, of which 60 were intact. Three IS groups within *S. haemolyticus*, namely ISSha1, IS1272 and IS256, comprised 85 % of the IS elements found[22]. This number is larger than what was seen in *S. epidermidis* and *S. aureus*[22], although certain strains of *S. aureus* of sequence type 247 (ST247), a MDR sub-lineage of clonal complex 8 (CC8), has revealed a high number of IS256[237]. IS256 was first described as a part of the transposon Tn*4001*, which harbours the ARG *aacA-aphD*[238] mediating aminoglycoside-resistance in staphylococci[238,239]. IS256 transposes by a "copy and paste" mechanism, leading to an accumulation of copies in the genome, as every transposition increases the copy number[240]. The integration of an IS element into a gene or its promoter may result in inactivation or overexpression of the affected gene[237]. In *S. aureus* it has been shown to be involved in antibiotic resistance modulation, increased virulence (by insertion into the promoter of *rot*, a global virulence gene receptor), and formation of small colony variants (SCV)[237,241–244].

Bouchami *et al*. recently studied the impact of IS1272 on the population structure of *S. haemolyticus*, and observed that IS movement and/or chromosomal alterations during stability assays promoted phenotypic changes. The changes observed, namely in mannitol fermentation, haemolysis and biofilm formation, could be clinically significant. In addition, most of these changes were reversible[50].

Transposons are larger, more complex elements, which encode multiple genes. These elements can change its position within a genome, which can sometimes create or reverse mutations and alter the genetic identity and genome size of the cell. Transposons can be divided into two classes based on their structure; simple transposon and composite transposon, where two copies of identical IS elements flank certain genes, often ARGs. Transposition can occur from genomic DNA to plasmid and vice versa, in addition to plasmid to plasmid[245].

1.3.2.3 Bacteriophages

Transduction involves the transfer of genetic material between bacteria through infection with a bacteriophage. Bacteriophages (phages) are viruses that infect and replicate within bacteria. During the process of bacteriophage replication bacterial DNA might erroneously be packaged into the virus head, called the "transduction particle", which can attach to and transfer DNA into a recipient cell. In order to be stably inherited and expressed the DNA must be incorporated into the genome by homologues recombination. Plasmid DNA may also be transduced and expressed in a recipient without recombination. A bacteriophage integrated into the genome is referred to as a prophage[245]. Most bacteria contain prophages, integrated either into their chromosome or as extra-chromosomal elements, contributing to substantial genetic variability. These MGEs can be responsible for gene disruption and provide docking regions for genomic rearrangements[246]. In addition to shaping the bacterial genome architecture, phages also constitute major tools for HGT[246,247], contributing to virulence by encoding several virulence or fitness factors, and by their movements within genomes[248–250]. Hence, it is clear that phages play essential roles in bacterial evolution and adaptation[249].

Bacteriophages in *S. aureus* have been shown to carry known toxins such as enterotoxin A (*sea*, food poisoning superantigen and allergy inducer), Panton-Valentine leukocidin (*PV-luk*, implicated in necrotic pneumonia and severe skin infection), complement inhibitory protein (SCIN), chemotaxis inhibitory protein (CHIP) and staphylokinase

(*sak*)[251–254]. Phage-encoded virulence factors responsible for *S. aureus* pathogenesis appear to be absent in CoNS[249]. Takeuchi *et al*. reported the identification of two prophages in *S. haemolyticus* JCSC1435[22].

1.3.2.4 Staphylococcal cassette chromosome

The staphylococcal cassette chromosome (SCC) is a family of MGE first described in staphylococci[255]. SCC operates as instruments transporting genes such as ARGs and also larger elements like transposon and plasmids. The emergence of methicillin resistant *S. aureus* (MRSA) originated from the acquisition of SCC carrying *mecA*. To date 11 different SCC types have been described[256]. In a recent review on SCC in CoNS, Saber *et al*. showed that *S. haemolyticus* carry SCC type I to V[257]. However, Silva *et al.* have also reported the presence of SCC type VII and VIII in clinical *S. haemolyticus* isolates[258]. In their study, SCC type I was the most prevalent. This stands in contrast to the majority of studies on SCC in *S. haemolyticus,* which have reported SCC type V to be the most prevalent in the typeable isolates[46,50,62,67]. Several studies show that *S. haemolyticus* are often non-typeable[67,259] and Bouchami *et al.* reported a high number of non-typeable SCC*mec* types (65.4 %) in their *S. haemolyticus* collection. They speculated that the high number of non-typeable SCC*mec* found in *S. haemolyticus* results from SCC*mec* rearrangements promoted by recombination and IS-induced genetic rearrangements[50]. In a study from 2001, Wielders *et al*. demonstrated the transfer of *SCCmec* from *S. epidermidis* to *S. aureus in vivo* during infection[260]. In 2018, a web-based tool for identification of SCC*mec*, based on whole genome sequences, was launched. However, due to the SCC diversity in *S. haemolyticus*, and that fact that the majority of strains are non-typeable, it offers limited insights into *S. haemolyticus* SCC[261].

1.3.2.5 Pathogenicity islands

A special class of MGEs is called pathogenicity islands (PIs), first described in the human pathogens of *Escherichia coli*. PIs has since been found in the genomes of various pathogens of humans, animals, and plants[262]. PIs contain groups of co-ordinately controlled virulence genes, often with IS elements at their ends[245]. *S. aureus* PIs (SaPIs) are a family of 14–27 kb genetic elements that usually stably reside in the *S. aureus* genome, similar to prophages, and contain phage-like repressor, integrase and terminase

genes, but do not contain genes encoding for phage structural proteins. Translocation of SaPIs is dependent on a helper phage for both excision and replication[263]. They usually carry two or more superantigens and are responsible for most superantigen-related human diseases, especially staphylococcal toxic shock syndrome[127,264].

Pathogenicity islands have been considered to be limited to *S. aureus*, but evidence of their presence has been reported in CoNS as well. Madhusoodanan *et al*. reported the finding of an *S. epidermidis* PI (SePI), with toxin genes[265]. The structure of this PI suggests that it might have originated from *S. aureus*. At present, the variety and the extent of such genetic transfers remain unclear[266]. Takeuchi *et al*. reported the presence of three predicted genomic islands in *S. haemolyticus*, however most of the genes were hypothetical, and the impact of these PIs in *S. haemolyticus* remain unknown[22].

1.4 Molecular epidemiology and population identification

Bacteria evolve extremely fast, and more rapidly than eukaryotic species. In addition to vertical gene transfer (parental cell to offspring), they can also share genetic material by HGT. Due to this ability, bacteria can rapidly make changes to their genome, which subsequently often leads to altered phenotype including altered antibiotic susceptibility. As a result, species identification of the infecting bacteria is not sufficient, we also need to know about the genetic background to ensure effective antibiotic treatment.

Within bacterial species we often have bacterial lineages, where bacterial species of a specific lineage have developed certain characteristics that might not be present in the same species from a different lineage[267,268]. Close examination and comparative analysis of different linages allows an understanding of bacterial evolution, caused by a combination of mutation, recombination and the acquisition of foreign DNA by horizontal gene exchange.

Bacterial typing - identifying bacteria at the species or clonal level - is of importance for diagnosis, treatment and epidemiological surveillance of bacterial infections. This is especially useful for identifying bacteria exhibiting high levels of antibiotic resistance or virulence, and those involved in nosocomial or pandemic infections[269]. In the following section, some typing methods used to identify intra-species variation in different staphylococcal species are presented.

Pulse-Field Gel Electrophoresis (PFGE)

PFGE has been a widely used method in molecular typing of bacteria for the past few decades[270–272]. The total genome is fragmented by enzymatic restriction cutting and separated on an agarose gel by applying alternating electrical fields, ideally fragmenting the chromosome into 20-30 fragments. Acquisition or loss of genes is reflected in variations in the band patterns, and can be directly compared to other isolate sample runs on the same gel. From there, comparison between bands (number of matches and relative size of presence/absence of mismatches) are made to determine the relatedness by defined criteria[273]. This method has a high sensitivity towards rapid genomic variation and is therefore highly valuable in outbreak situations. On the other hand, PFGE is time consuming and there is poor portability of results between laboratories, in addition to difficulty assessing levels of differences between strains.

Multi locus sequence typing (MLST)

MLST, a procedure for determining population structure of bacteria using internal fragments of normally seven housekeeping genes, was proposed by Maiden *et al.* in 1998 as a method for overcoming data exchange problems between different laboratories[274]. MLST is a sequenced based method that has been widely used for a number of different bacterial species to study bacterial evolution and pathogenicity[275]. MLST creates an allelic profile based on combination of single nucleotide polymorphisms found in a section of usually seven housekeeping genes present in all isolates. The analysed isolates are assigned a sequence type (ST) after comparing their sequences with known alleles. By selecting slowly evolving housekeeping genes, the evolutionary relationship can be studied, as well as the establishment and spread of specific bacterial clones. The eBURST algorithm groups isolates sharing similar allele SNPs into clonal complexes[276]. Isolates that have six out of seven identical loci are grouped together. A clonal complex is defined by the ST believed to be the founder, from which the other STs have evolved. For each gene (allele) the combination of different SNPs is assigned a number, and if the SNP combination is different to those already present in the database it will be assigned a new number. The allele combination combined creates a ST. As increasing numbers of clinical isolates are sequenced, and additional genes can now be included to these schemes, offering higher resolution. MLST is a typing method widely used in *S. aureus* and *S. epidermidis*[277,278], and schemes are also developed for *S. haemolyticus*, *S. hominis* and *S.*

pseudointermedius[279–281]. Currently the MLST databases of *S. aureus*, *S. pseudointermedius* and *S. epidermidis* consists of 5533, 1523 and 897 different STs, respectively. Both *S. haemolyticus* and *S. hominis* have more modest MLST databases, with only 69 and 57 ST at the time of writing (checked 02.01.2020)[282].

Extended MLST

The traditional MLST became both successful and globally used for typing of many bacterial species[282], but the method unfortunately lacks the discriminatory capability to differentiate tightly linked isolates[279,283]. Now, WGS (see section 1.6.1 for more details) has become a powerful and attractive tool for rapid typing of bacteria and has gradually replaced the traditional MLST as a new "gold standard" in molecular epidemiology for surveillance of infectious disease, on a local as well as on a global level[284]. Extended MLST analyses the alleles of several hundreds or even thousands of genes (coreMLST or whole-genomeMLST), and thereby offers a much stronger discriminatory power, that the traditional MLST method is lacking. Today there are several online resources used for bacterial typing (Enterobase, PubMLST, BacWGSTdb)[282,285,286]. The advantages of some of these databases are that, in addition to the genotypic data, epidemiological data (such as host, source, associated disease, collection time and geographical origin) accompany the specific strain. Some databases also check for additional factors such as antibiotic resistance and virulence genes (BacWGSTdb)[285,287]. Both source tracing and surveillance is made possible based on this information. Currently several of these tools are limited to certain pathogenic bacteria.

1.5 Species and subspecies identification

The emergence of new technologies in genomics and proteomics has been shifting traditional phenotypic techniques for bacterial classification, identification, and characterization towards methods based on the elucidation of specific gene sequences or molecular components of a cell[288,289]. Characterization of bacteria has many practical applications in addition to the fundamental questions of bacterial systematics, taxonomy, and evolution. Rapid identification and discrimination of pathogenic bacteria has a major impact on public health, in relation to correct diagnosis and timely disease treatment. Bacterial characterization can also aid in interpreting the mechanisms of bacterial

pathogenesis, and allows for the discovery of important molecular targets essential for the development of vaccines, diagnostic kits, and therapeutics for infectious diseases. These types of applications make the continued development of techniques for bacterial identification important both in basic science research and for the maintenance of human health[288].

Species and subspecies identification (Lower resolution)

16S rRNA: 16S ribosomal RNA is a component of the 30S subunit of the prokaryotic ribosome. It is universally present in all bacteria and useful for reconstruction of phylogenies due to the slow rate of evolution of this gene. 16S rRNA is very often used for conferring relatedness of different bacteria and for species determination. The percentage similarity of 16S rRNA between species to define them as the same species was previously set to 97%, but is now increased to 98.7%. For staphylococci, however, the 16S rRNA gene is very similar across species, and two different species can have identical 16S rRNA sequence[290,291]. As an example, *S. caprae* and *S. capitis* cannot be distinguished from one another based on 16S rRNA analysis[292]. The inaccuracy related to the use of 16S rRNA for species determination is reported for several bacterial species[293,294].

Species and subspecies identification (Medium resolution)

MALDI-TOF: In contrast to the other genotypic identification methods mentioned here, matrix-assisted laser desorption ionization–time of flight (MALDI-TOF) mass spectrometry (MS) is a phenotypic molecular typing technique that identifies bacterial isolates based on unique protein profiles[295]. For detection and identification, a protein spectrum is obtained and compared to a reference database of bacterial protein spectra. According to the manufacturer's criteria a score of >2.000 represent species level, a score of 1.700-1.999 a genus level, and < 1.700 no identification[296]. MALDI-TOF MS is currently established for routine identification of bacterial pathogens in microbiological laboratories[297,298], and for the distinction of bacterial strains during nosocomial outbreaks in intensive care units[299].

Species and subspecies identification (Higher resolution)

DNA–DNA hybridization (DDH) experiments have been considered as the "gold standard" since the 1960s in determining relatedness between bacterial species, as this has been one of the few universally applicable techniques available that could offer truly genome-wide comparisons between organisms. A value of 70% DDH has been the recommended standard for delineating species[300,301]. Because the DDH value reflects relatedness or similarity between two genomes, there has been a series of efforts to develop a bioinformatic method to replace DDH for differentiating species. Devising values analogous to the DDH values, on the basis of similarity or distance, has been termed the overall genome related index (OGRI)[302]. OGRI represents any measurements indicating how similar two genome sequences are.

Among the different OGRIs, average nucleotide identity (ANI) is most widely used. ANI is a category of computational analysis that can be used to define bacterial species boundaries. Calculating ANI usually involves the fragmentation of genome sequences followed by nucleotide sequence search, alignment, and identity calculation. The final ANI value is the mean of identity values of all fragments of the query genome[300]. ANI has been widely used to compare bacterial genome sequences when classifying and identifying bacteria. There are currently several software tools and online calculators available for calculating the ANI, and improved versions are developed for taxonomic purposes[303–306].

An alternative to ANI is digital DDH (Genome-to-Genome Distance Calculator; GGDC), which is also widely used for taxonomic purposes. The GGDC is a state-of-the-art in silico method for genome-to-genome comparison, reliably mimicking conventional DDH[307]. It has been recommended that authors proposing new species should provide OGRI values between the type strain of proposed species and type strains of related species that show above 98.7% 16S rRNA sequence similarity[289,308]. The workflow describing species identification using OGRI is shown in figure 5.

Because OGRI does not have a taxonomic resolution above the species level, a multigene-based phylogenomic treeing approach could be used for defining genera or higher taxa. Application of genome data to phylogenetic analysis, called phylogenomic treeing, can be achieved by inferring phylogenetic trees on the basis of multiple genes, instead of a single gene such as 16S rRNA. It uses the same approach as MLST analysis, but can be differentiated by the use of substantially higher number of orthologous genes, rationally

selected using large scale comparative genomics[309,310]. Functionally important genes, for example genes encoding ribosomal proteins or housekeeping genes, are usually selected for phylogenomic treeing. Currently, there are several software tools that can be used for phylogenomic treeing, using sets of 31 to 400 house-keeping core genes, are available[282,289,311–314].

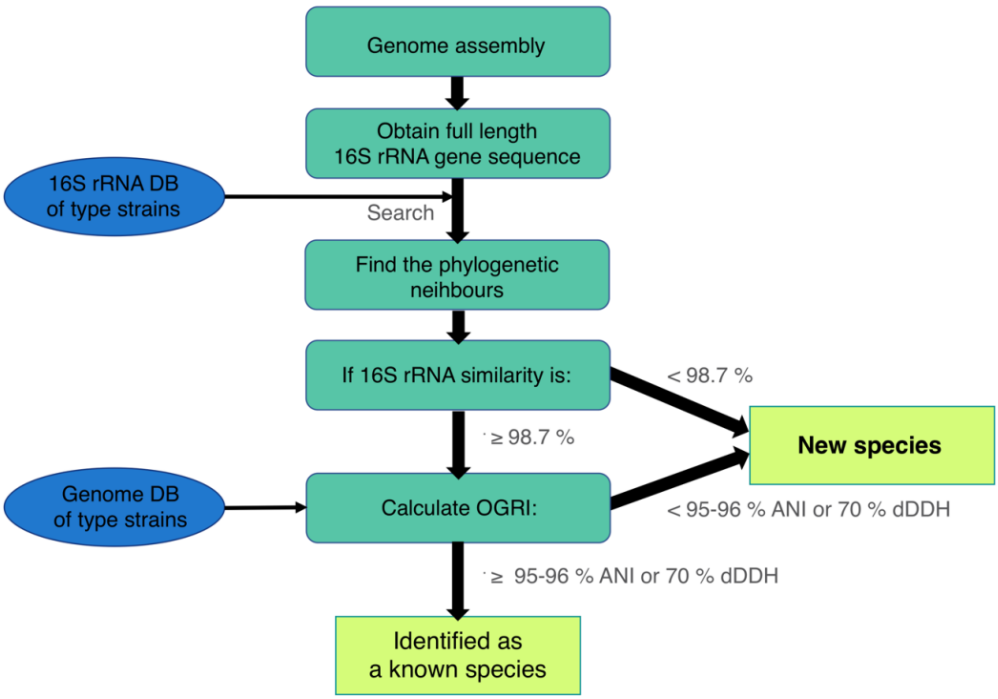


Figure 5: Workflow of genome-based classification of species, figure adapted from[289].

1.6 Bioinformatics and computational biology

Bioinformatics is the application of techniques derived from disciplines such as applied mathematics, computer science, and statistics to analyse, store, retrieve and interpret biological data. Computational tools are routinely used for many types of analysis such as characterization of genes, determining structural and physiochemical properties of proteins and phylogenetic analysis, to name a few. The following section will focus on the general steps in obtaining a readable genome, from isolating the bacterial DNA to "building" the genome and predicting the function of the proteins. I will also describe the sequential steps to learning more about the bacterium being analysed with the help of different databases and programs, with special emphasis on the methods chosen in this thesis.

1.6.1 Whole Genome Sequencing

Over the last two decades, exceptional progress has been made in DNA sequencing technologies, which has led to substantially decreased cost per megabase and a drastic increase in the number and the diversity of sequenced genomes[315]. Whole genome sequencing (WGS) is the process of determining the complete DNA sequence of an organism. WGS of bacteria has, until recently, mainly been used as a research tool, but is currently being introduced to clinical medicine aiding in diagnostics, and particularly useful in surveillance of bacterial outbreaks[285–287,316–318]. Using WGS has also significantly increased the understanding of several conditions and diseases[319–321]. In the future of personalized medicine WGS data may be an important tool to guide therapeutic intervention[322,323]. There are two main types of WGS technologies on the market; short-read and long-read sequencers. I will next (section 1.6.1.1-2) give a short introduction to the current technologies, with focus on their strengths and limitations.

1.6.1.1 Short read WGS

Short-read sequencing are further divided into two broad categories: sequencing by ligation and sequencing by synthesis. The sequencing by ligation uses the enzyme DNA ligase to identify the nucleotide present at a given position in a DNA sequence. These technologies, including SOLiD and Complete Genomics (BGIseq), offer very short reads (75bp and 28-100bp, respectively) which limits their use for genome assembly applications, and will not be described further.

Sequencing by synthesis uses a DNA polymerase to incorporate nucleotides in order to determine the DNA sequence. This approach can be further classified as i) cyclic reversible termination or ii) single-nucleotide addition. Illumina® sequencing technology, using the cyclic reversible termination approach, dominates the short-read sequencing market in part due to its experience as a technology, high level of cross-platform compatibilities and its wide range of platforms. Briefly explained, after library preparation (generating the collection of DNA fragments), attachment and amplification (forming clusters) on the flow cell, Illumina utilizes modified deoxyribose nucleoside triphosphates (dNTPs) acting as chain terminators. These modified dNTPs are incorporated into the DNA and prevents extension of the DNA molecule until the chemical "block" signal has been removed. A mixture containing all the four dNTPs, each with distinct chemical fluorescent block, is added to the flow cell. Each cluster will emit one of

four fluorescent signals, corresponding to the incorporation event. Unincorporated dNTPs are removed and the flow cell is imaged using a camera recording the signal. The terminators are then removed and the cycle can start again[315,324].

Illumina's technology has some limitations. The short reads length (50 – 300 bp) can create problems with the determination and assembly of complex genomic regions, such as repeat sequences and AT- and GC-rich regions[325–327]. One way to optimize the performance in the problematic repeat regions is to obtain sequences from both ends of the randomly selected DNA fragment, and to have an algorithm keeping track of these "paired-ends" and the distance between them. This improves the chances of being able to anchor the sequence data from each fragment onto the genome. However, what Illumina lacks in read length it makes up for in high-quality bases and very high sequencing capacity with low error rates[315,324,328].

The single nucleotide approach is used by the Ion Torrent® systems offering longer read lengths compared to other short-read sequencers with reads up to an average of 400 bp, and thereby provides some advantages for applications that focus on repetitive or complex DNA[329]. As for Illumina, the genome is broken down to fragments and each fragment is amplified. In Ion Torrent technology the fragmented DNA molecules are attached to a bead and then amplified. dNTPs are added to the flow cell sequentially and each dNTPs incorporation results in release of H+ and a chip detects the resulting change in acidity (pH). In essence, tiny pH meters convert chemical information (H^+) into digital sequence data[329]. Ion Torrent is faster than most other current platforms which makes the device well suited for clinical applications[330]. The drawbacks of this technology are insertion and deletion errors, although the overall error rate is similar with other NGS platforms in non-homopolymer regions[331].

1.6.1.2 Long read WGS

Bacterial genomes are highly complex with several long repetitive elements, copy number alterations and other structural variations relevant to the evolution, adaptation and disease of the species[332,333]. However, many of these complex elements are so long that short-read paired-end technologies are incapable of resolving them. Thus, the main limitations with short-read WGS are i) the short read length and ii) the need for amplification. Amplification is time consuming and has the potential of introducing bias, such as over- or underrepresenting certain regions.

In contrast to short-read WGS, long-read WGS allows for the resolution of these large structural features. These long reads can span complex or repetitive regions with a single continuous read, and therefore eliminate ambiguity in the positions or size of genomic elements[315]. Currently, the most widely used long-read platform is the single-molecule real-time (SMRT) sequencing approach used by Pacific Biosciences (PacBio®)[334–336]. This device is capable of generating single reads up to 50 kb with average read lengths of 10–20 kb[328]. The SMRT sequencing device by PacBio uses a flow cell covered with tiny wells, where each well has a single, specially engineered DNA polymerase attached to its bottom. Single molecule sequencing is possible because the polymerase is stationary. The instrument camera is able to focus on what is happening at the bottom of each well, recording the fluorescence each time the DNA polymerase incorporate a dNTP into the DNA chain (each dNTP has its own fluorescent tag). These properties are ideal for *de novo* genome assembly applications and for sequencing long complex genomic structures[337]. The error rate for long reads is as high as 15%, however, these errors are distributed randomly within each read and thus sufficiently high coverage can overcome the high error rate[338,339].

A different long read sequencer, the MinION from Oxford Nanopore Technologies (ONT), became commercially available in 2014[340]. The ONT MinION® is a small USB-based device that runs on a personal computer, making it the smallest of any current sequencing platform. This gives the MinION superior portability, highlighting its utility for rapid clinical responses and remote field locations. Like PacBio, ONT technologies sequences single molecule DNA in real time. The procedure begins with applying an electric current to a synthetic membrane that can have up to several thousand pores inserted into it. A specialised motor protein delivers the double stranded DNA molecule to the nanopore, unzips the DNA and passes them through the pore one base at the time. The sequence of each DNA molecule is inferred by recording the disruptions in current that occur at each nanopore as the nucleotide pass through[340]. One of the great advantages with ONT technologies is the superior read lengths. Indeed, read lengths approaching 1 Mbp has been reported. On the other side, ONT MinION still has a higher error rate, currently decreased from 40 to 15 %, and is dominated by indel errors[341]. In addition, effective homopolymer (difficulties inferring the precise number when a stretch of identical nucleotides are encountered) sequencing also remains a challenge for ONT MinION[342]. Fortunately, recent improvements in the chemistry and the base calling algorithms are improving accuracy[343]. Total sequence output was initially modest, but increasing steadily as the technology matures. Despite its massive potential, error rates for currently

available single molecule sequencers are higher than for Illumina, and the total number of reads generated per instrument is typically lower[315,328].

At present time, SMRT platforms have proven to be particularly effective when used in combination with more traditional technologies like Illumina. The idea is to combine the high quality and output of Illumina with the larger read length of SMRT platforms[315]. The ability to analyse data in real time presents a major advantage in clinical scenarios, where fast detection of specific mutations can provide epidemiological information, such as the relatedness of outbreak strains or ARGs and virulence genes, that directly impacts initial management decisions during hospital outbreaks[344,345].

1.6.2 From millions of read to a draft genome

Depending on the application, once a bacterial species has been sequenced the reads need further processing to build a draft genome. The majority of bioinformatical tools are available as command line tools that can be used free of charge. In order to use these types of tools, some bioinformatical skills and in many cases access to Unix-based computers are needed. In the following section the steps of building a draft genome are briefly discussed including some the commonly used software.

After having received the reads from sequencing the first step is to check the quality[346]. After assuring satisfactory quality of the sequence reads it is often useful to trim the reads, again depending on the planned downstream analysis. Different tools are available for processing of the reads, which include the trimming of adapter sequences and low-quality reads. The trimming of sequence reads has been shown to increase the quality and reliability while decreasing the computational requirements of downstream analyses[347]. Developing tools for read trimming is an active area of bioinformatics research and there are currently many options to choose from, such as Trimmomatic and PRINSEQ[348,349].

The next step in building a genome is to combine the short reads to build larger stretches of DNA, a process called sequence assembly. An assembly algorithm is then implemented to compile reads into larger sequences (contigs) that eventually represent a genome. There are two different types of assembly methods; mapping the reads to a reference and *de novo* assembly. If a reference genome of that particular strain exists, the reads can be mapped directly on that reference. Most often, it is necessary to assemble the genome with no reference available, which is what we call *de novo* assembly.

It is often a good practice to test a few different tools before deciding on which one is most suitable for the data being analysed. A helpful tool to assess assembly quality and compare genome assemblies is QUAST[350]. N50, defined as the minimum contig length needed to cover 50 % of the genome, is often used as a standard metric to evaluate the quality of an assembly in terms of contiguity[346]. There are several tools available, and certain tools might work better for your specific genome. Once the assembly is complete, it is possible to map the contigs onto a closely related strain. This will order and orientate them along chromosomes, which can be useful for later genome comparisons[351].

After a genome assembly, it is often of interest to perform functional genome annotation, which is the process of identifying the locations of genes and all of the coding regions in a genome, as well as predicting gene function. There are several tools available, and it might be important to use the annotation tools that have the best compatibility for planned downstream analysis. RAST® (Rapid Annotation Using Subsystem Technology) is a fully automated Web-based tool that can be used to annotate contigs[352]. A major disadvantage with RAST is the long processing time, which can be days depending on the current server traffic. PROKKA® (Rapid Prokaryotic Genome Annotation) coordinates a suite of software tools to achieve a rich and reliable annotation of genomic bacterial genomes. It is a fast and easy-to-use command line tool, with its main drawback being decreased annotation performance for understudied genomes[328,353]. A benchmark test showed that PROKKA outperforms RAST in terms of the number of predicted elements[353]. Figure 6 shows the workflow from sequencing to annotation.

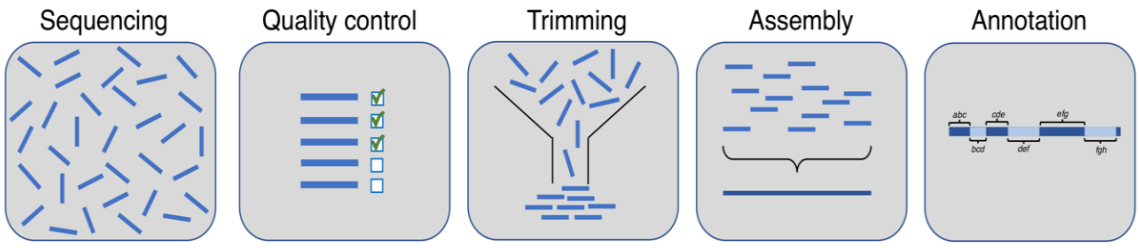


Figure 6: Overview of recommended workflow from sequencing to annotation.

1.6.3 Genome mining using online tools

The amount of available genetic information on a vast number of microorganisms has increased rapidly over the past few decades, and as a result a plethora of different databases containing specific types of genomic information are available today. By obtaining the WGS of an organism you can harvest a lot of information about that specific organism by the use of online databases. There are databases on species determination, virulence and resistance detection, identification of MGE, genome classification, protein localization and secondary metabolites, etc. It is important to keep in mind that a predicted gene in the genome is not a determination of function. The following section will give a brief introduction to some widely used databases, including the ones used in this thesis.

Virulence

There are two main web-based tools for prediction of virulence factors in bacteria; VirulenceFinder® 2.0[344] VFDB®[354,355] (virulence finder data base). VirulenceFinder is easy to use and takes both raw sequence reads and assembled contigs as input[344]. However, a significant limitation of VirulenceFinder is that it only contains virulence genes for a few bacteria (*Listeria*, *S. aureus*, *E. coli* and *Enterococcus)*. VFDB contains data from 74 distinct genera, 951 bacterial strains, and a total of 1,080 identified virulence factors[354]. The VFDB platform has also recently released an interactive web interface[355], making it easier to use the database. Currently, in the VFDB it is only possible to upload one genome sequence at a time, and the processing time is long. Compared to VirulenceFinder, this database contains more markers associated with virulence, and several genes involved in housekeeping functions. However, the content of VFDB is not as well validated as VirulenceFinder[328]. When dealing with a large number of genomes, using online tools can be very time consuming. The command line tool ABRicate[356] is a great option for mass screening of contigs for ARGs or virulence genes. It comes bundled with multiple databases: NCBI, CARD, ARG-ANNOT, Resfinder, NCBI, EcOH, PlasmidFinder, Ecoli_VF and VFDB.

Antimicrobial resistance (AMR)

ResFinder®[357] is an easy-to-use web-based tool used to detect acquired ARGs in WGS data. The database has regular updates, and as for VirulenceFinder, inputs for ResFinder can be either raw sequence reads or contigs[357]. CARD (The Comprehensive Antibiotic Resistance Database) is a manually curated resource containing high quality reference data on the molecular basis of AMR, with an emphasis on the genes, proteins and mutations involved in AMR[358]. CARD spans the range of AMR drug classes and resistance mechanisms, including intrinsic, mutation-driven and acquired resistance[358]. CARD curation is updated monthly, based on an interplay of manual literature curation, computational text mining, and genome analysis.

Mobile genetic elements (MGE)

PlasmidFinder is a user-friendly web-based tool for the detection of bacterial plasmids[233]. Most plasmids include specific regions, called replicons, encoding functions that are able to activate and control replication, and this tool is based on the detection of these replicons[232,233]. The PlasmidFinder output gives an indication of which plasmids are present, but further investigation is needed for confirmation. PHASTER (PHAge Search Tool Enhanced Release) is a user-friendly web-based tool for the rapid and accurate identification and annotation of prophage sequences within bacterial genomes and plasmids[359]. It accepts either raw DNA sequence data or partially annotated GenBank formatted data. PHASTER also generates downloadable, interactive graphics that display all identified prophage components in both circular and linear genomic views[359,360].

ISsaga is a web application pipeline that provides computational tools and methods for high-quality IS annotation[361]. It uses the established ISfinder annotation standards[362], and rapidly processes single or multiple prokaryote genomes. ISsaga provides information on genome context of individual IS elements and a graphical overview of IS distribution around the genome of interest[361].

It is important to emphasize that these tools state whether genes (or similar genes) are present, not if they are expressed and related to a phenotype. All these tools and methods need to be validated by functional analysis.

1.7 Proteomics

Most of the functional information of genes is characterized by the proteome. Proteomics is the characterization of proteome, including expression, structure, functions, interactions and modifications of proteins at any stage[363]. Proteomics involves technologies for the identification and/or quantification of the protein content of a cell, tissue or an organism. It supplements the other "omics" technologies such as genomics (DNA) and transcriptomics (RNA) to clarify the identity/structure and adds functional information of the proteins of an organism. Proteomics-based technologies are utilized in various capacities for different research settings, such as detection of various diagnostic markers, finding candidates for vaccine production, understanding pathogenicity mechanisms, investigating alteration of expression patterns in response to different signals and interpretation of functional protein pathways in different diseases[364]. Mass spectrometry, with LC-MS/MS and MALDI-TOF MS being widely used equipment, is central among current proteomics[364].

Surface-exposed proteins form the first-line of contact between bacteria and host. These proteins are major constituents in this complex interplay as they directly interact with epithelial and immune cells[365]. The in-depth analysis of the proteosurfaceome, defined as "the proteinaceous subset of the surfaceome found at the cell wall and totally or partially exposed on the external side of the cell membrane", represents an important point to elucidate molecular mechanisms underlying host-microbe crosstalk[366,367]. Surfaceome investigation uses sophisticated strategies that selectively target surface proteins by either protease digestion (shaving procedures) or biotin labelling (protein biotinylation procedures) of intact cells[368].

2 Objectives and aim of this thesis

The main objective of this PhD thesis was to use a combination of comparative genomics and phenotypic assays to achieve a better understanding of *Staphylococcus haemolyticus* colonization, virulence and hospital adaption.

I aimed to answer the following three main research questions:

1. Are there genotypic and/or phenotypic differences between clinical and commensal isolates of *S. haemolyticus*?
2. What are the genes predominantly found in clinical (and not commensal) *S. haemolyticus* isolates, and can these genes explain their adaption and ability to cause disease and be used as future clinical markers of *S. haemolyticus* infection?
3. Which proteins are expressed on the surface of *S. haemolyticus,* and can these explain the adhesion/biofilm potential of this species*?*

Objectives per paper

Paper I: In this paper, we aimed to identify genomic differences between clinical and commensal *S. haemolyticus* isolates, and to find for specific markers that can differentiate between the two.

Paper II: In this paper, we aimed to identify *S. haemolyticus* surface proteins using a novel method for identifying surface proteins expressed during human host colonization.

Paper III: In this paper, we aimed to describe a new species of the *Staphylococcus* genus, detected during our work on comparative genomics and phenotypic assays in isolates previously identified as *S. haemolyticus*.

3 Methodology

Research should be performed with the aim of creating reproducible and sustainable knowledge. The chosen methods should meet the requirements of modern scientific principles and deliver credible data. When possible, different approaches illustrating the same phenomenon should be used to confirm results. In addition, the methods should highlight the research questions from different angles. However, most methods will have some limitations, which must be kept in mind when interpreting the data.

Detailed descriptions of the methods are presented in the three manuscripts that forms the basis for this thesis. The following section lists a general presentation of the methods used in the manuscripts. Some of these methods needed optimisation, and are thus described more in detail.

3.1 Bacterial culture collection

The bacterial collection used in this thesis (**paper I-III**) included 123 clinical and 46 commensal *S. haemolyticus* isolates. The clinical isolates were collected between 1988-2010, and the majority were obtained from blood cultures[47]. Moreover, the invasive isolates were from the following geographical locations: Norway (n=93), Switzerland (n=42), Japan (n=13), Germany (n=9), United Kingdom (n=9), Spain (n=2) and USA (n=1). The commensal isolates were mainly collected in Tromsø, Norway (n=41) in 2014, being part of skin sampling of CONS from healthy volunteers[56]. The body locations of these commensal *S. haemolyticus* isolates were as follows: From the groin (n=27), hamstring (n=6), armpit (n=5), nasal cavity (n=1) and unknown (n=7).

3.2 Phenotypic assays

Antibiotic resistance determination (paper I & III)

In **paper I and III** we determined both the phenotypic and genotypic antibiotic resistance pattern. We tested susceptibility to antibiotics commonly used in the treatment of Staphylococcal infections. Phenotypic antibiotic testing was performed according to EUCAST guidelines[369]. Combining both genotypic and phenotypic data on antibiotic resistance is of importance and great value when interpreting the antimicrobial resistance patterns. It is of interest identifying the genes responsible for phenotypic antimicrobial resistance, as well as identifying putative ARGs genes where no resistance phenotype is yet observed. For identification of ARGs we used the previously mentioned CARD database[358], as CARD provides more extensive results than the other ARG databases.

Adhesion, Cell adhesion and Biofilm assays (paper II)

Colonization of biotic and abiotic surfaces is predicted to be the first step in *S. haemolyticus* infections, and is thus regarded as a virulence factor. In order to look at the adhesion and biofilm potential of the isolates we performed different bioassays, to investigate whether there was a difference in the adhesion properties between clinical and commensal isolates.

Adhesion and biofilm assays: Phenotypic variations were observed between the biological replicates when performing the different bioassays (adhesion to plastic, collagen, fibronectin and HaCat cells, and the biofilm assay). These assays are sensitive to external factors, and we tried to minimize these by doing the following modifications; the assays for all included isolates were performed on the same day, the same equipment and batches of reagents and medium were used and the experiments were performed by the same person. For biofilm detection we used the modified Christensen method[60,370].

Cell adhesion assay: We chose to use human keratinocytes (HaCaT) to study host microbe interaction, as *S. haemolyticus* is skin colonizer. The cell adhesion assays were performed as described in the original protocol developed in our lab, based on Edwards and Masseys protocol[371], with a few modifications to improve reproducibility. First, adding a PBS-EDTA step prior to the trypsin treatment of HaCaT cells allowed a shorter trypsination and thereby healthier HaCat cells. Secondly, centrifuging the plates after inoculating the

bacteria onto the HaCat cells allowed a more rapid and uniform contact between bacteria and HaCat cells. Both these steps improved the consistency and reproducibility between the experiments. In the cell adhesion assay, we also investigated internalization and survival of the bacteria within the HaCaT cells. After the bacteria and HaCat cells had co-incubated for 60 minutes, antibiotics were added. For the internalization experiment antibiotics were left for 1 hour, and for the survival experiment the antibiotics were left for 24 hours in order to kill extracellular bacteria. The antibiotic most commonly used in this assay is gentamicin. As several of the clinical isolates were resistance to gentamicin, we added lysostaphin and vancomycin. In order to ensure efficient killing of extracellular bacteria, we added an additional control step by plating out the supernatant to check for bacterial growth.

Phenotypic methods for describing a new species (paper III)

The genomic analysis of our *S. haemolyticus* collection indicated that five of these isolates were potentially a new species. We therefore proceeded to characterize these isolates using various phenotypic methods, following the recommendations of Freney *et al.*[372]. The experiments were performed on all the five isolates in addition to the *S. haemolyticus* type strain CCUG 7323 and JCSC1435. The experiments required for describing a new species were performed using the recommended media and reagents, with a few exceptions. When investigating the haemolytic activity of the bacteria it is recommended to use agar plates containing bovine, sheep or human blood, instead we used horse blood for our experiments as this is the standard protocol at the University hospital of North Norway, and has also been described by others[17]. For motility testing we used a lower percentage of agar (2-3 %) and used a motile *E. coli* as positive control. To determine pigment production, milk or egg yolk agar is recommended and we made P agar, a non-selective medium recommended for staphylococci cultivation[373], with full fat milk for this purpose. The isolates grew poorly, and normal P agar and P agar with horse blood had to be used instead.

3.3 Surface shaving

Surface proteins promoting adhesion are predicted to be one of the most important virulence factors of *S. haemolyticus*. Knowledge of *S. haemolyticus* surface proteins are scarce, and for **paper II** we developed a method to investigate expressed surface proteins

after co-incubation with human keratinocytes (HaCaT). Bacteria cultured in cell-culture media supplemented with foetal bovine serum were used as controls. Briefly, *S. haemolyticus* was co-incubated with HaCaT cells for 60 minutes, prior to harvesting. Harvesting of bacteria adhering to HaCaT cells was done by mechanical cell scraping of the tissue culture plate. Regular trypsin-treatment was avoided as this would have affected the surface proteins expressed on the bacterial surface. Bacteria were separated from the HaCaT cells by flow cytometry (FACS), followed by two centrifugation steps in order to concentrate the bacteria.

The bacteria were then added to a lipid-based flow cell, where surface expressed proteins were "shaved" of by trypsin. Surface proteins were then collected and labelled using tandem mass tags (TMT), to enable easier quantification, and to compare the abundance of different proteins by mass spectrometry. This Lipid-based Proteins Immobilization (LPI) technology was developed by our collaborators in Nanoxis®. It allows for immobilization of intact bacterial cells within a flow cell via membrane-gold interaction. The surface expressed bacterial proteins are subjected to enzymatic digestion within the flow cell and proteins are subsequently identified using LC-MS[374].

3.4 Bioinformatical and statistical analyses

Whole genome sequencing, assembly and annotation (paper I & III)

WGS **(paper I & III)** was performed by using the Illumina technology (Genome Analyzer GAII and MiSeq) on index-tagged library using paired-end reads. The clinical isolates were sequenced earlier than the commensal isolates and the sequence reads were shorter. The clinical isolates were assembled, annotated, analysed and published in 2014[47]. To avoid biases, all clinical isolates were re-assembled and annotated identical to the commensal isolates (**paper I**). Several assembly tools (Velvet[375], ray[376], SPAdes[377], IDBA-UB[378]were tested on a few isolates of both clinical and commensal origin to find the best suitable tool for the *S. haemolyticus* genomes. We used Quast to evaluate the quality of assembly results, and SPAdes provided in general longer and fewer contigs and higher N50[350]. All *S. haemolyticus* isolates were then finally assembled using SPAdes[377].

When analysing the large collection of *S. haemolyticus* isolates we identified that five isolates were genetically distant from the rest of isolates. These five isolates, later proposed as the novel species *Staphylococcus borealis*, were initially assembled using

SPAdes[377]. While preparing **paper III**, the reads were reassembled with Shovill® due to a high number of contigs, this yielded a drastically improved assembly[379]. Shovill is a pipeline that uses SPAdes at its core, but alters the steps prior to the primary assembly step. Shovill downsamples the FASTQ file to 100x depth, avoiding a poor assembly due to too much data.

There are several annotation services available, and for **paper I & III** we chose to use PROKKA[353], as it coordinates a suite of software tools to achieve a rich and reliable annotation of genomic bacterial genomes. It is a fast and easy-to-use command line tool and it is good for prokaryotic genomes in addition to providing several different file formats of output which is an advantage for downstream analysis[328,353].

Phylogeny and Pangenome (paper I & III)

A pangenome analysis was performed to uncover potential genetic differences between commensal and clinical isolates of *S. haemolyticus*. For performing pangenome analysis there are several available tools. We tested three different tools with different settings; Get_homologues, BPGA and Roary[380–382]. Which tool and setting to choose depends on the type of application and planned downstream analysis. Other factors which influence the choice of methods are; processing time, flexibility, output and ease-to-use.

The cut-off percentage for sequence similarities decides when genes are to be considered of the same cluster. Lower sequence similarity leads to fewer clusters, where homologues CDS are placed in the same cluster even if there are some changes in the sequence, which would be an appropriate setting for looking at the presence and absence of genes. Choosing a higher sequence similarity produces a bigger pangenome where homologues genes are separated into different clusters if there is some variation in the protein sequence. This can be useful when looking at different variants of the same gene. We chose to use Roary as it is very easy to use and was compatible with the downstream analysis we wanted to do. As an extension of pangenomics, bacterial pangenome-wide association studies (panGWAS) pairing phenotypes with SNPs, presence and absence of genes and other regions like regulatory sequences have begun to emerge[215–217]. For this purpose we used the program Scoary[383].

Use of different databases and local blast (paper I-III)

By obtaining the WGS of an organism, a lot of information on the specific species can be gathered by using online databases. A local blast database with all isolates genomes was set up, and interesting findings reported by the different online databases were always checked and confirmed by local blast searches.

Genomic species determination (paper III)

In the past, species determination was largely dependent upon phenotypic characterisation. In the genomic era, more emphasis is put on genomic characterisation, and less phenotypic tests are necessary to prove species delineation. Recent paper describing new species focuses largely on the genetic differences between the newly proposed species and its closest relatives[17,18,384].

There are several bioinformatic tools available for comparing the relatedness between bacterial species. The uncharacterised novel species was compared with *S. haemolyticus* type strain CCUG 7323 using ANI and dDDH calculators[306,307]. In addition, relatedness between species can also be conferred by phylogenomic tree reconstruction, of which several tools are available. We used UBCG pipeline to infer phylogenomic treeing, which looks at 92 conserved proteins in the bacterial genome (taking draft genomes as input).

Statistical analyses (paper I)

In **paper I** selected data (ARGs and virulence factors) were also analysed using IBM-SPSS statistical software (IBM Corp. Released 2015. IBM SPSS Statistics for Windows, Version 23.0. Armonk, NY: IBM Corp). Categorical data are displayed as ratios and frequency (%), and analysed using the chi square test. We used biological knowledge to identify known pathogenicity-associated traits enriched in clinical isolates. In particular, we focused on ARGs and previously established virulence factors. We developed different scores including the following four traits; *aacA-aphD*, *mecA*, *folP* and phenotypic biofilm production. In order to find a pragmatic score that could differentiate between a clinical and a commensal isolate we calculated the area under receiver operating characteristic (ROC) curves, and its 95% confidence interval.

4 Summary of main results

Paper I: Comparative genomic analysis of *Staphylococcus haemolyticus* reveals key to hospital adaptation and pathogenicity

In this paper we described the genetic differences found between clinical and commensal isolates of *S. haemolyticus*

- We show a clear separation of clinical and commensal isolates of *S. haemolyticus* based on a phylogenetic tree using core genome SNPs
- The clinical isolates were overall more resistant to antibiotics with 88% being MDR compared to 11% of the commensal isolates (**Figure 7**)
- By performing a pangenome analysis we identified several genetic traits that differed between clinical and commensal isolates
    - The ARGs *aacA-aphD* and *mecA* were found significantly more common in clinical vs. commensal isolates (85 % vs 4 % and 85 % vs 13 %, respectively).
    - The IS-element IS256 was also found significantly more common in clinical vs. commensal isolates (86 % vs 11% in commensals).
    - Two different versions of both *folP and folB* were identified in the commensal vs the clinical isolates (78 % of the clinical isolates had a different version).
    - A large phage denoted StauST387-2/vB_saus_phi2, associated with *aacA-aphD, dfrC* and the IS elements IS256 and IS431, was only found only in clinical isolates (51 %).
    - The polysaccharide capsule operon was more prevalent in the clinical isolates (48 % vs 22 %). We also detected in total three alternative capsule operon versions, in addition to the already described *S. haemolyticus* capsule operon.
    - The accessory sec system and the *SraP*-homolog, shown to be important for virulence in other bacterial species, were also more prevalent among the clinical isolates (97% vs 74%).

In conclusion, there was a clear segregation of isolates of commensal origin, and specific genetic signatures distinguishing the clinical isolates from the commensal isolates.

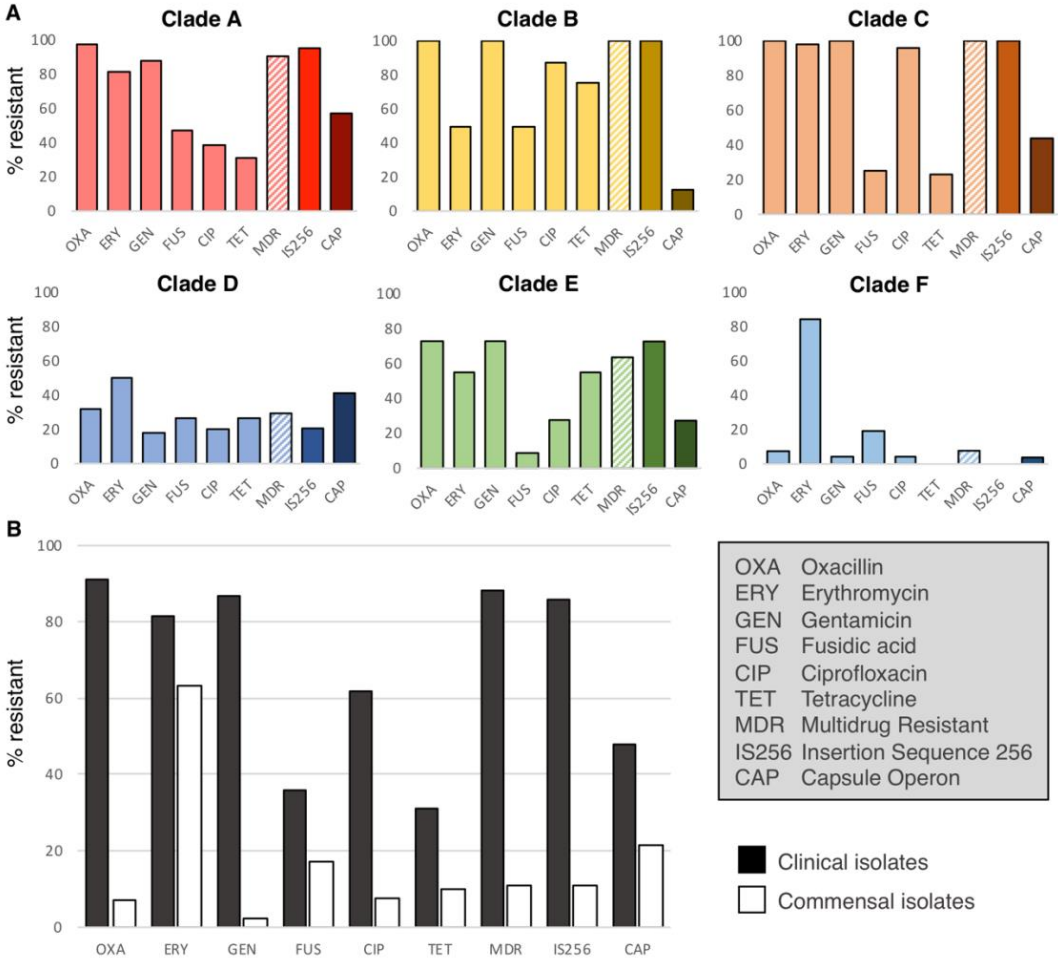


Figure 7: Phenotypic antibiotic resistance rates (%) in 169 *S. haemolyticus* isolates (123 clinical and 46 commensal). Phenotypic resistance distribution among the six different clades (A) and between clinical and commensal isolates (B).

Paper II: Identification of surface proteins in a clinical *Staphylococcus haemolyticus* isolate by bacterial surface shaving

In this paper, we compared the adhesive and biofilm forming properties of 20 *S. haemolyticus* isolates. We also described a novel method for investigating expression of surface proteins in *S. haemolyticus* after human host colonization compared to protein expression in cell culture medium supplemented with serum (**figure 8**).

- Adherence to fibronectin, collagen and plastic was low in all tested isolates, but with significantly higher adhesion to fibronectin ($p = 0.041$) and collagen ($p = 0.001$) in the commensal isolates compared to the clinical isolates.
- There was a trend towards a higher degree of biofilm formation in the clinical isolates ($p = 0.059$), where 5/10 clinical isolates formed substantial amounts of biofilm assay ($OD_{570} \geq 3$) compared to 0/10 commensal strains.
- After surface shaving of a clinical *S. haemolyticus* isolate we identified 324 proteins, of which 65 were classified as surface proteins.
- Among the surface proteins, five were LPXTG-containing proteins, of which two had serine rich repeats previously characterised in several staphylococcal CWA proteins.
- The SasH-like protein has been shown to be important in *S. aureus* immune evasion.
- Ebh/Embp implicated in adhesion and biofilm formation in *S. epidermidis* and *S. aureus* was identified in this study.
- The TIR-domain protein, Atl and SceD were significantly upregulated following HaCaT co-incubation.
    - TirS was 100 % identical to the *S. aureus* TirS protein where it has been shown to have immune evasive properties.
    - The Autolysin Atl has been shown to be involved in biofilm formation.
    - The transglycosylase SceD has been shown to be important for cell wall remodeling contributing to resistance to antimicrobial peptides, adhesion, *S. aureus* pathogenicity, and colonisation of nares in cotton rats.

In conclusion we have identified surface proteins and immune evasive proteins previously only functionally described in other staphylococcal species. We have also identified hypothetical surface proteins, which need to be further characterized. The majority of the identified proteins were equally abundantly identified in both conditions, indicating that serum is a strong inducer of protein expression.

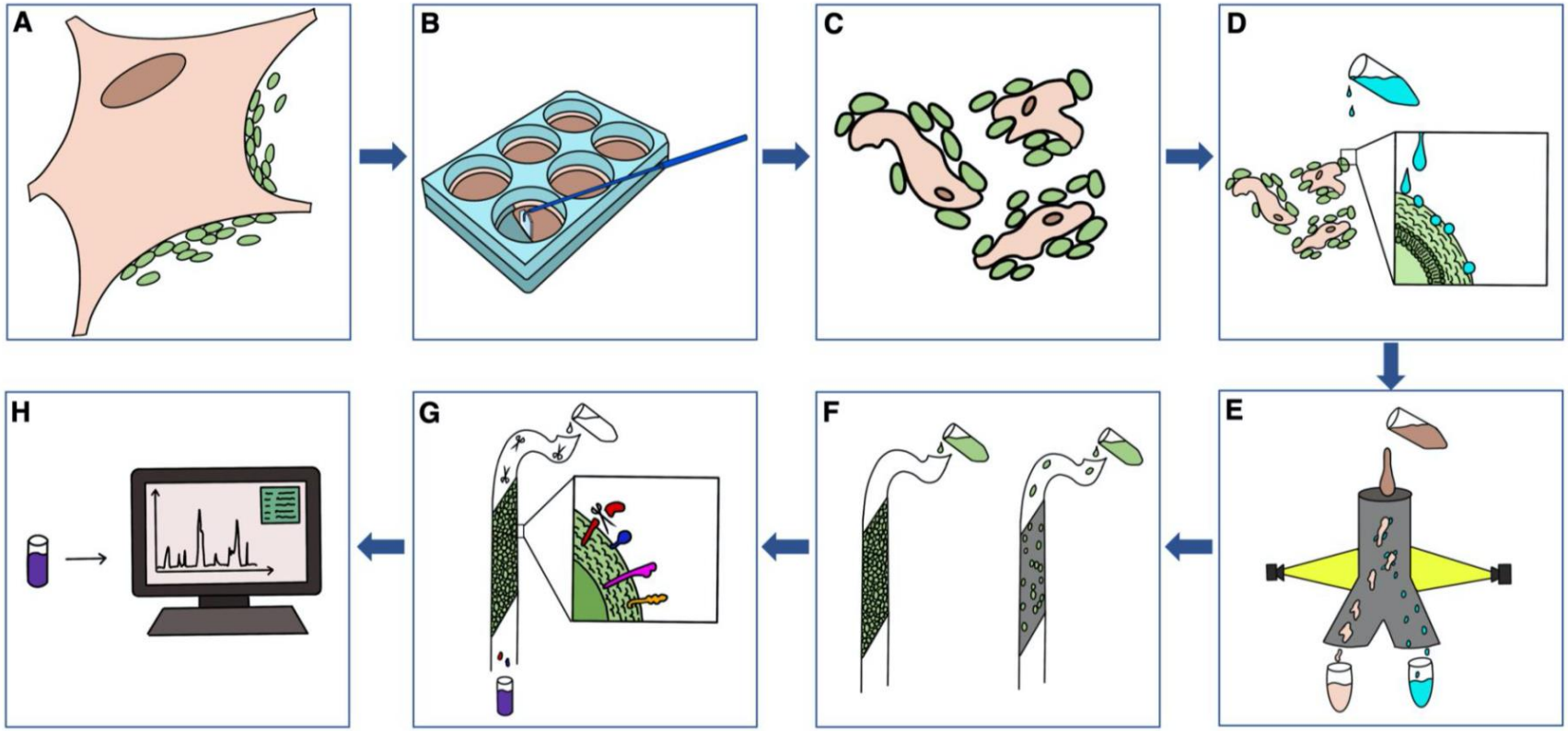


Figure 8: Simplistic step-by-step representation of the surface-shaving process.
Bacteria are grown with human cells to allow for adhesion to the cells (**A**). After co-incubation wells are washed several times, and the remaining cells with the bacteria attached is collected into a tube using a cell scraper (**B,C**). The bacteria are labeled with a dye that only binds to features of the bacteria (**D**), allowing the human cells and bacteria to be separated on FACS (**E**), based on the signal from the dye and size. The collected bacteria are added to a flow cell (**F**), which will become fully saturated with bound bacteria. Digestive enzymes are added to the flow cell and allowed to work for a limited amount of time so that only the proteins on the surface is cut of and collected (**G**). The collected peptides are then subjected to Mass Spectrometry analysis (**H**).

Paper III: *Staphylococcus borealis* sp.nov. – a novel member of the *Staphylococcaceae* family isolated from skin and blood in humans

In this paper we describe the characterisation of five strains belonging to a novel staphylococcal species, *Staphylococcus borealis*. The five strains were noticed as they presented aberrant genotypic traits when analysing a large cohort of *S. haemolyticus* by WGS.

- Based on the traditional MLST scheme the new species revealed large nucleotide diversities in all internal MLST genes compared to *S. haemolyticus*. The difference was 23 to 79 bp difference across 6 genes (none had the gene Ribose ABC – which is also true a few *S. haemolyticus* isolates).
- Phylogenomic treeing, based on 92 housekeeping gene, placed all five *S. borealis* on a separate branch
- Genotypic methods confirm that these strains can be proposed as a new species
  - The average nucleotide identity (ANI) showed the five isolates to be between 87.74 and 87.99 % similar to the type-strain of *S. haemolyticus,* which is well below the threshold value of 95-96% for isolates of the same species.
  - The Genome-to-genome Distance Calculator (GGDC) showed an identity of 34.1 to 34.3 % with the *S. haemolyticus* type strain, which is also well below the suggested threshold of 70% for isolates of same species
- Performing a pan-genome analysis revealed 944 genes only found in these 5 isolates. Several of the unique genes are homologs to *S. haemolyticus* genes.
- Compared to the 169 *S. haemolyticus* isolates all five *S. borealis* isolates displayed yellow colonies, where *S. haemolyticus* has white-gray colonies (**figure 9 and 10**)
- The five isolates showed a stronger haemolytic capacity on p-agar compared to *S. haemolyticus* (**figure 11**)
- *S. borealis* harboured a different capsule gene organisation compared to *S. haemolyticus*
  (**figure 12**)
- All five isolates produced urease and contained the urease operon

In conclusion the phenotypic and genotypic results support the description of a new staphylococcal species (suggested name *Staphylococcus borealis),* closely related to *S. haemolyticus*. WGS analysis is a powerful tool for identification of novel species.

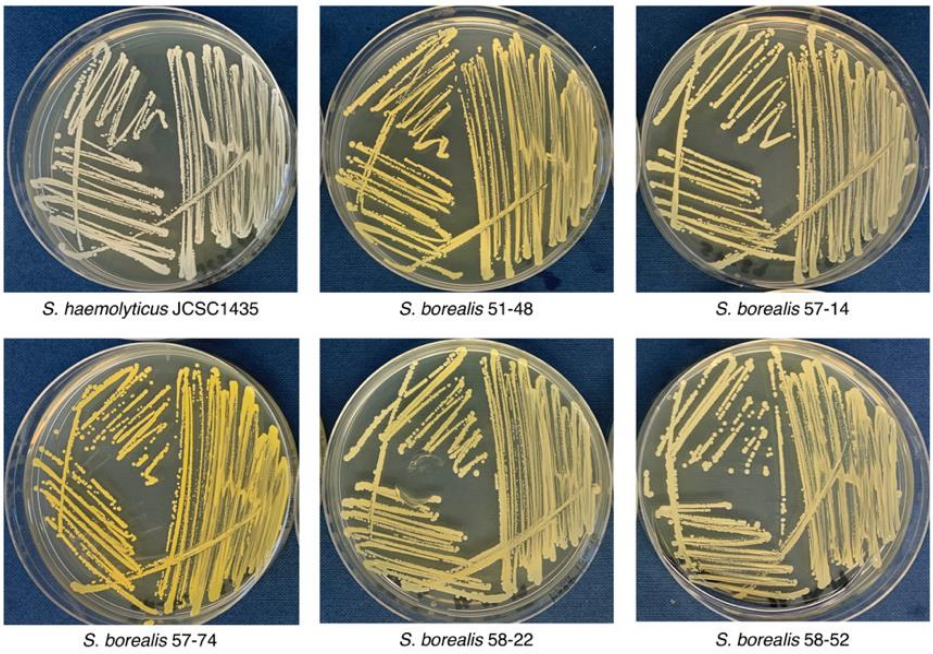


Figure 9: Pigmentation production of *S. haemolyticus* JCSC1435 and the five *S. borealis* isolates on P-agar plates.

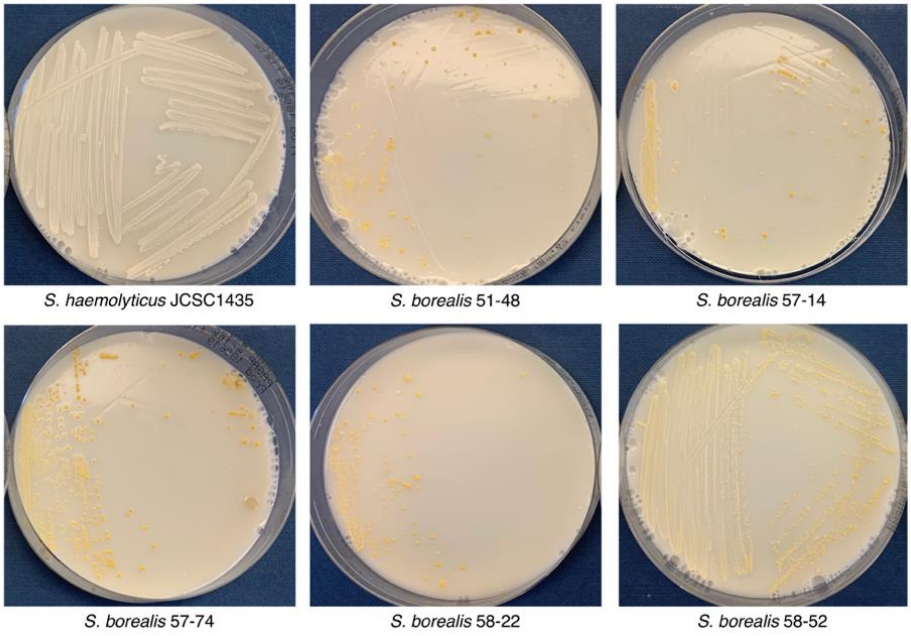


Figure 10: Pigmentation and growth of *S. haemolyticus* JCSC1435 and the five *S. borealis* isolates on P-agar with full-fat milk.

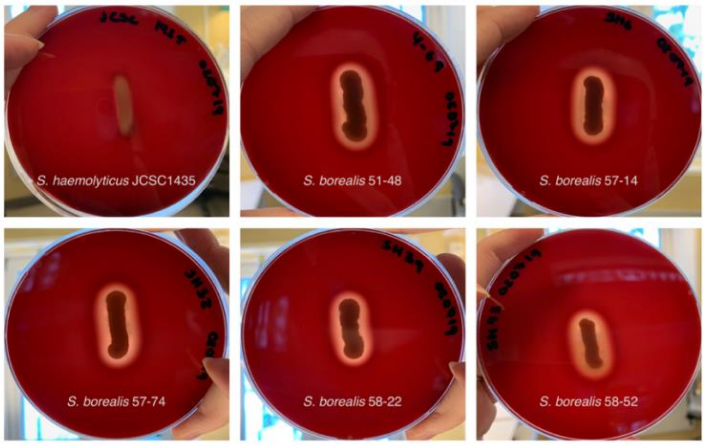


Figure 11: Haemolysis of *S. haemolyticus* JCSC1435 and the five *S. borealis* isolates on P-agar with 5 % horse blood.

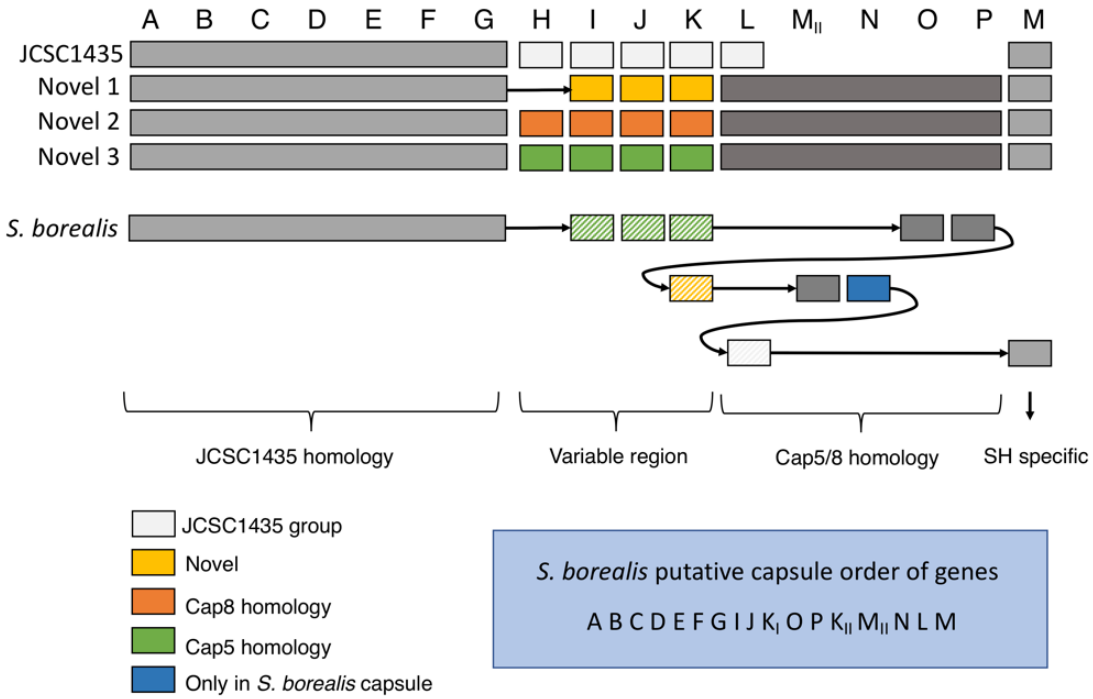


Figure 12: Organization of the different polysaccharide capsule operon identified in our *S. haemolyticus* and *S. borealis* isolates. The three novel *S. haemolyticus* operons and the *S. borealis* operon varies in their homology to JSCS1435. *capA-G* is homologues among all versions. The region *capH-K* is different among all *S. haemolyticus* versions, while the *S. borealis* (green stripes) shows most similarity to the novel version 3. *capH-K* novel 2 shows similarities to *S. aureus* Cap8, and novel 3 to *S. aureus* Cap5. *capL-P* is similar between the three *S. haemolyticus* versions, and homologues to *S. aureus capL-P*. These genes are absent in the JCSC1435 group. The organization after *capK* in *S. borealis* is very different to the *S. haemolyticus* capsule versions.

5 General Discussion

Staphylococcus haemolyticus has long been regarded an important nosocomial pathogen[66,78,385–387]. Yet information about this species is scarce compared to *S. aureus* and *S. epidermidis*. Typically, *S. haemolyticus* is referred to "as the second most frequently isolated CoNS, after *S. epidermidis*"[21], often referring to studies conducted more than a decade ago[22,44]. However, recent publications suggest this might not be a true picture today. Several studies now report *S. haemolyticus* as the most frequently isolated CoNS-species from urine[388,389], blood culture[43,58], puerperal (childbed fever) infection[390] and other clinically relevant conditions[40,42,49,391,392]. The majority of these studies were performed in high-income countries, and it is difficult to predict whether the prevalence is similar in other parts of the world. A few recent studies from western Europe on CoNS-bacteraemia still show a higher prevalence of *S. epidermidis* compared to *S. haemolyticus*[393,394]. *S. haemolyticus* has been considered to have few important virulence attributes[39], and perhaps for that reason has been placed in the shadow of *S. epidermidis*. However, the increased prevalence of MDR *S. haemolyticus* infections truly highlights the importance of comprehensive research and surveillance of this nosocomial pathogen.

The large collection of both clinical and commensal *S. haemolyticus* isolates that were investigated in this thesis allowed the combined in-depth analysis on the genetic composition and the phenotypic traits in order to uncover more information on *S. haemolyticus*.

5.1 Population structure

In **paper I**, one of our main research aims was to uncover genetic mechanisms for hospital adaption and pathogenicity. We therefore excluded isolates that could "cloud" our analysis, which comprised isolates of animal origin, isolates where the genome sequences were of poor quality after assembly, and also five isolates that were genetically very distant from the rest of the isolates in our collection (we termed these five isolates the "out-group", and they were further characterized in **paper III)**. Phylogenetic analysis by comparing SNPs in the core genome divided the isolates into six clades, where the majority of commensal isolates (85 %) were restricted to two clades and displayed high diversity compared to the clinical isolates.

The clinical isolates grouped tightly together independent of country of isolation, showing a clonal population structure, where successful hospital adapted clones have expanded and persisted[47]. This observation is different to what was recently shown for *S. epidermidis*, where pathogenic clones can emerge seemingly independent of lineage[208]. The *S. haemolyticus* population structure is more similar to *S. aureus*, where the rise of successful pathogenic lineages is seen[5,218,395].

MLST was performed for all *S. haemolyticus* isolates in our collection using an established MLST scheme[279]. The majority (37/49) of the identified STs were new. The clinical isolates were divided into 31 STs, and 92 % of the clinical isolates belonged to the major CC1. The commensal isolates were divided to 25 different STs, of which only 4 were previously described. However, only 50 % of the commensal belonged to CC1. Additionally, we performed MLST on the five "out-group" isolates, but all alleles differed significantly from the ones already described. Based on the combined findings of MLST and SNP-based phylogeny we speculated that these five isolates could represent a new species. We therefore in **paper III** characterised the "out-group" isolates in detail, and proposed that these isolates belong to a new staphylococcal species. For the rest of the isolates both MLST and SNP-based core phylogeny showed some of the same trends; a significantly higher diversity among the commensal isolates than the clinical isolates. The combined results from these analyses suggest that due to the clonal structure of the hospital isolates, traditional MLST with only 7 gene loci has very limited discriminatory ability in *S. haemolyticus*. The MLST scheme still did discriminate quite well between clinical and commensal isolates, but it does not provide sufficient resolution for outbreak investigation. An alternative MLST scheme for *S. haemolyticus* was published in 2011[396], but the paper was written in Russian and therefore of limited use for the overall public. Recently a study compared these two MLST schemes and suggested a new MLST scheme using some loci from both previous schemes in addition to one new loci[397]. However, the new scheme also did not appear to produce significantly higher resolution. This indicates that it would be more appropriate to develop a whole genome MLST scheme for *S. haemolyticus* rather than trying to optimize the current MLST schemes.

One reason for the differences in population structure between *S. haemolyticus* and *S. epidermidis* could be different niches and roles for the two staphylococcal species. Both are inhabitants of the skin, but while *S. epidermidis* is a prominent and prevalent species on dry, moist and sebaceous body sites[398], less is known about *S. haemolyticus* preferred/natural niche. A few older studies have reported *S. haemolyticus* to be

preferentially isolated from arms and legs[399,400], and less frequently from nares and axillae[400]. In contrast, in a study by our group *S. haemolyticus* was preferentially cultures from the axillae and groins[56]. In a recent publication on the human skin microbiome, *S. haemolyticus* was only listed among the top ten abundant bacterial species on the foot[398]. A different study by the same author showed that *S. haemolyticus* accounted for 1 % of the total staphylococcal species, independent of disease state of the skin microbiome in paediatric atopic dermatitis compared to healthy controls[401]. In the healthy population the most commonly isolated staphylococci were *S. epidermidis* and *S. hominis*[18,393,401]. There is a clear discrepancy in literature regarding *S. haemolyticus* preferred body site, however it seems clear that its relative prevalence is lower than that for *S. epidermidis* and *S. hominis*. A new study found that the two major lineages of *S. epidermidis* (the A/C and B-lineage, as defined by[207,218]) appeared to occupy different skin niches and perform different biological functions on the skin; lineage A/C had a broader and more superficial skin niche, while lineage B dominated hair follicles and sebaceous glands. The authors found that linage A/C isolates were more likely to cause infections[402]. Hence, understanding the species-specific niche could be an important step in understanding the roles and functions of the bacteria, and can also potentially provide information about pathogenic potential.

In **paper I** we showed that the population structure of the clinical isolates differed from the commensal isolates - presenting a more clonal population, rich in ARGs associated with certain IS elements, indicating a successful genotype. Combining these findings with previous reports on *S. haemolyticus* suggest that established hospital clones outcompetes commensal clones. This is exemplified by Mannion *et al.* who showed a 5-fold increase of *S. haemolyticus* post-operatively compared to before operation, with oxacillin resistant isolates increasing from 9% to 95 % [403]. Additionally, Larson *et al.* also found that *S. haemolyticus* was significantly more often isolated from patients compared to healthy non-hospitalised individuals[78].

Whereas the prevalence of *S. haemolyticus* on skin of healthy adults appear to be low, studies show *S. haemolyticus* to be a prevalent part of the microbiota of neonates[57,59,404,405]. It has been proposed that neonates might play an important role in the spread of MDR *S. haemolyticus*[59,405]. It might not be surprising then that *S. haemolyticus* is also retrieved from surfaces, material and air within the hospital as the most frequently isolated species[51,52].

The relative low abundance of *S. haemolyticus* on the skin of healthy adults, the higher prevalence on neonates and in patients compared to healthy counterparts, and the clonal population structure observed for the clinical isolates of *S. haemolyticus*, suggest that infections caused by this species are less likely to come from the patient's own flora, and more likely to have been acquired from the surroundings and from the hospital upon admission. S. *haemolyticus* also appears to possess a unique ability to adapt to a range of different, and often very harsh, environments. For *S. epidermidis* it has been suggested that infections can often result from translocation of the hosts own bacteria[208,406,407]. However, it appears that this is not a common transmission route for *S. haemolyticus*, rather, infections caused by *S. haemolyticus* are most likely to originate from already established hospital clones. The diversity we observed among the commensal isolates, which mainly originated from the same geographic location (**paper I)** is strongly suggestive of a different adaption strategy than for the clinical isolates.

5.2 Antibiotic resistance

Ever since *S. haemolyticus* was first described in 1975[408], its resistance to several antimicrobials has been described, and is referred to as the most extensively resistant CoNS[39,44]. However, most of the studies have investigated isolates of clinical origins, and the commensal counterparts have been given little attention. In **paper I** we added genotypic resistance patterns to the already presented phenotypic antimicrobial patterns of the commensal isolates [56], and combined it with previous published results on the clinical isolates[47]. Phenotypically we showed that the clinical isolates were resistant to multiple antimicrobials, consistent with what has been reported in literature over the last 40 years. In contrast, the commensal isolates showed significantly less antimicrobial resistance. A study conducted in Portugal investigated resistance pattern in 170 commensal CoNS isolates from the nares of healthy humans and found 19 % to be MDR.[80] Interestingly, among the commensal isolates 65 % were resistant to erythromycin. The commensal isolates carried mainly *mphC*, while the clinical carried either *ermC* or both *ermC* and *mphC*. A study, investigating staphylococcal species isolated from indoor air of flats in Poland reported that 63.6 % of the staphylococci were resistant to erythromycin with *mphC* being the most detected gene[409]. In other studies of commensal CoNS, resistance towards penicillin and erythromycin is the most prevalent[77–80,410]. However, none of these studies showed such a high prevalence of erythromycin resistance as

observed in our study. In Norway until 2013, the macrolide azithromycin was prescribed as a first-line antibiotic to treat genital chlamydia infections. Since then guidelines for genital chlamydia infection therapy have changed and now recommend doxycycline due to resistance development towards macrolides[411,412]. Genital chlamydia infections have a high prevalence in Tromsø[413,414] and as 72 % of the commensal *S. haemolyticus* with macrolide resistance were isolated from the groin, it could be possible that these resistant isolates have been spreading through human contact.

In **paper II** we studied the ability of *S. haemolyticus* to adhere to human keratinocytes (HaCaT cells). In the same assay we also investigated the isolates ability to invade and internalize the HaCaT cells. As the clinical isolates were MDR, we added vancomycin and lysostaphin in addition to gentamicin in the experiment, to ensure that all extracellular bacteria were killed after co-incubation of bacteria with HaCaT cells. In theory, none of the bacteria should be able to survive as all were susceptible to both vancomycin and lysostaphin. However, we found that after 60 minutes, 6/10 of the clinical isolates survived antimicrobial treatment despite being susceptible to the aforementioned treatment. Additionally, we observed that several clinical isolates produced small colony variants (SCV). The explanation for this observation could be multifaceted. The vancomycin MIC, as measured with MIC test in standard medium in the laboratory, may not be the same in a different medium. Indeed, studies have shown that MIC testing performed in host-mimicking media added serum provides different MIC values compared to the standard bacterial medium (Mueller-Hinton Broth) for antimicrobial susceptibility testing[415–417].

Kim *et al.* showed that vancomycin treated *S. haemolyticus* demonstrated cell wall thickening, slower growth and decreased autolysis. Increased resistance to vancomycin also resulted in decreased susceptibility to lysostaphin, teicoplanin, daptomycin and tigecycline[98]. Furthermore, studies on *S. aureus* shows that vancomycin induce and positively selects for SCV[418]. This could provide a potential explanation for our observation of an increase in SCV's of *S. haemolyticus* after vancomycin treatment (**paper II)**. SCV are slow-growing subpopulations of bacteria that has atypical colony morphology and unusual biochemical characteristics, often deficient in electron transport or thymidine biosynthesis[419]. They are often isolated from chronic infections and have reduced susceptibility to antibiotics[237]. Clinical SCV isolates are often formed during intracellular growth or prolonged antibiotic therapy[420,421]. Moreover, SCV has been associated with increased biofilm production in *S. aureus*[421,422]. The insertion element

IS256 is also associated with the formation of SCV in *S. aureus*[237]. In our genome comparison (**paper I**) between commensal and clinical *S. haemolyticus*, IS256 was one of the strongest differentiating findings separating the two groups, found in 86 % of clinical isolates and 11 % of commensals. It is therefore possible that IS256 was involved in the observed SCV phenotype in the cell assays performed. Kleinert *et al*. also demonstrated that unlike other IS elements (IS1272, IS1181, and IS431mec), which were mostly located between the coding sequences, 46% of the IS256 insertions were intragenic and inducing SCV formation[237].

A case report describing a patient with endocarditis, showed that the infection could not be cleared with treatment using glycopeptides, even though the infecting *S. haemolyticus* was classified as susceptible to both vancomycin and teicoplanin[69]. In this report the authors speculated that the finding of SCV, which they attributed to early rifampicin treatment, reduced the *in vivo* susceptibility to glycopeptides. They concluded that *in vitro* susceptibility towards glycopeptides was not predictive of *in vivo* eradication of *S. haemolyticus* from endocarditis and that recovery from endocarditis by CoNS without surgery is unlikely[69].

Neither vancomycin nor linezolid resistance was detected in the *S. haemolyticus* isolates in our collection by conventional phenotypic susceptibility methods. Resistance to glycopeptide antibiotics in CoNS is largely dependent on point mutations, and studies have shown resistance to occur in *S. haemolyticus* during infection, as well as easily developing *in vitro*. We have shown in this thesis that clinical isolates of *S. haemolyticus* have the ability to form SCV, produces a stronger biofilm and contains several copies of IS256, all factors that can influence antibiotic susceptibility **Paper I** and **paper II**).

5.3 Horizontal gene transfer and mobile genetic elements

For the "last resort" antibiotics (e.g. vancomycin and linezolid), the most common mechanism for decreased susceptibility and resistance to *S. haemolyticus* are point mutations. However, many of the other antibiotic resistance mechanisms found in *S. haemolyticus* are based on ARGs located on MGE. We showed in **paper I** that a large portion of the accessory genome of *S. haemolyticus* belonged to the COG categories associated with MGE and antimicrobial resistance. It is believed that *S. haemolyticus* easily acquires MGE [39], but so far successful HGT, with *S. haemolyticus* as recipient, has

not been reported – neither *in vitro* or *in vivo*. A few studies have suggested transfer of MGE from *S. haemolyticus* to other species[423–425].

One reason why genetic engineering of *S. haemolyticus* is so difficult lies in the restriction modification (RM) systems found in this species. RM systems are important components of prokaryotic defence mechanism against invading gene. Restriction is achieved by the cleavage of the foreign DNA, which is unmethylated, while the genome of the host remains protected due to methylation by the cognate methyltransferase[426]. A study from our research group has previously shown four different types of RM systems in our clinical isolates, and some isolates contained more than one[47].

Revealing the mechanisms behind HGT for *S. haemolyticus* is important in order to understand, and in turn, hopefully try to limit the spread of resistance. Even though the exact mechanism of transfer and uptake of genes has not yet been fully explained, it seems clear that *S. haemolyticus* does rely on HGT. In **paper I** we showed that the majority of genes separating clinical isolates from commensal isolates are typical MGE-related genes. The aminoglycoside-modifying gene *aacA-aphD,* associated with transposon Tn*4001* that includes IS256[238] was found almost exclusively in clinical isolates. Several plasmids and transposons, carrying ARGs, also found in *S. aureus* were identified in many *S. haemolyticus* isolates. Overall, the clinical isolates had a higher number of MGE's than the commensal counterparts. The few studies that have looked at plasmids in *S. haemolyticus* have not identified conjugative plasmids, and exactly how transfer of resistance genes in this species is mediated is presently unknown.

Several studies have reported a high genome plasticity and a very high number of IS elements in *S. haemolyticus*[22,47,427]. In our *S. haemolyticus* collection, IS1272 was detected in all isolates in several copies. However, due to the fragmented genome, the exact copy number was hard to predict, and we could not determine whether the number of this element differed between clinical and commensal isolates. IS256 on the other hand was identified almost exclusively in clinical isolates.

5.4 Biofilm

Research on *S. haemolyticus* has predominantly focused on antibiotic resistance and biofilm formation, as these have been considered the most important traits for clinical significance. To date most of the research has focused on presence and prevalence of these

factors rather than the genes and specific proteins responsible for these traits. Despite some variations in the methods used and the results, the overall consensus is that this bacterium has the ability to produce biofilm[60,201,428]. A comprehensive study on *S. haemolyticus* biofilm analysed the type of biofilm formed by this species[60]. In contrast to *S. epidermidis* which produces a polysaccharide biofilm, the major constituents of *S. haemolyticus* biofilm were proteins and extracellular DNA[60]. In the same study it was shown that some genes, namely bifunctional autolysin *atlE* and *fbe*, previously described in biofilm formation in other staphylococcal species were also present in the genome of *S. haemolyticus*[195]. However, their role in biofilm formation was not determined[60]. In our collection *atl* (*atlE* homolog) was found in all isolates and in **paper II** Atl was also one of the proteins found to be more abundant on the *S. haemolyticus* surface after co-incubation with keratinocytes, indicating, at least, a potential role in adhesion to host cells.

The best studied biofilm associated genes in *S. epidermidis* are the *ica* genes. Generally, the prevalence of *ica* genes in *S. haemolyticus* is low[51,60,387,428]. However, a few studies from Greece and Brazil showed a high prevalence of *ica* containing *S. haemolyticus* strains, this suggests that they provide certain advantages in these environments[38,201,429]. In a previous study by our group only 3/72 isolates were positive for *icaD* by PCR, confirmed by southern blot in 2 of these 3 isolates. Two of these three isolates were also part of the collection used in this thesis, but *ica* genes were not found it any of our isolates. It has previously been reported that DNA segments were deleted in strains of *S. haemolyticus* JSCS1435, an observation assigned to IS-encoded transposase acticity[22], and a similar deletion event may have occurred in the previously *ica*-positive isolates.

We performed phenotypic biofilm assays and showed that the clinical isolates formed more biofilm compared to commensal isolates. However, the difference was only found to be statistically significant in **paper I**, and not in **paper II**, and the interpretations of the results should be handled with caution. Biofilm formation is a dynamic process and for *S. haemolyticus* the genetic mechanisms regulating this process are still largely unknown. Therefore, a number of factors might influence the results of biofilm assays. In static biofilm assays several researchers agree that glucose is a trigger for biofilm formation[60,430,431], while NaCl lowers the amount of biofilm produced[60,432]. Others have showed that while NaCl on its own might inhibit biofilm formation, a combination of NaCl and glucose promotes more biofilm than glucose on its own[431]. As part of the required tests for species determination in **paper III**, we observed that growing JSCS1435 in p-broth at pH 5 produced a visible biofilm, which was not observed on higher pH, indicating that pH

could also be an important contributor in inducing biofilm. Most biofilm assays are performed in tryptic soy broth (TSB), but some studies have been performed in Mueller Hinton Broth (MHB) or Brain-Heart Infusion broth (BHI). Barros *et al.* tested the biofilm formation of *S. haemolyticus* JSCS1435 in both TSB and BHI, with and without glucose and NaCl, and found that BHI with 1% glucose yielded much more biofilm compared to the other conditions[428]. We also experienced that the type of microtiter plate used can have a large impact on the amount of biofilm the bacteria produce. Lastly, how the optical density (OD) read outs are interpreted, including the cut-off values for biofilm positivity, can have a huge impact on the reported results. In the original Christensen *et al.* study from 1985 an OD > 0.12 was considered as positive for biofilm formation. This value was chosen because it was three standard deviations (SD) above the mean of a clean tissue plate stained in the same way as the bacterial samples[370]. A modified and new version of this calculation to determine cut-off for biofilm positive samples has been used, such as, three times SD above mean of negative control (the negative control being a known biofilm negative bacterium)[433,434] and above two times mean of negative control[435], while others have used OD above 0.1 and/or 0.25 to be considered biofilm positive[51,60]. Whether a bacterial isolate is biofilm positive or not is hugely dependent on medium, equipment and OD interpretation, and in my opinion a true comparison between isolates can only be made if all above factors are exactly the same for the strains compared (e.g. comparison of strains in a single assay). Biofilm formation is likely a highly regulated process, with different host and environmental signals influencing and regulating the different stages in biofilm formation [436,437]. We decided to choose a biological relevant model when investigating the surface proteins of a clinical *S. haemolyticus* isolate. In **paper II** we first investigated the biofilm formation ability of 10 clinical and 10 commensal isolates, and although all isolates were considered biofilm positive, the clinical isolates were on average stronger biofilm producers. To make a more generalised conclusion on whether clinical isolates are stronger biofilm producers more isolates would need to be tested, and such a result would then only be valid in those specific conditions. We chose one of the clinical isolates, that showed superior adhesion to keratinocytes in addition to biofilm formation for further investigation of the surface proteome (**paper II**).

In the surface shaving experiment (**paper II**) we detected an Embp (Ebh in *S. aureus*) homolog. Embp has been described as a giant fibronectin-binding protein in *S. epidermidis* sufficient and important for biofilm formation[197]. Christner *et al.* also showed that Empb is a multifunctional cell surface protein that mediates attachment to host tissue (ECM), biofilm escape phagocytosis[197]. Linnes *et al.* suggested that the central role

of Embp lies in commensal skin colonization, rather than as a virulence factor[438]. In the isolate used for surface shaving, three different *embp* homologs were found in the genome. Two of the three Empb proteins were detected in the surface shaving experiment, which proves that in *S. haemolyticus* at least two out of the three genes are expressed and/or functional. The specific role of these proteins in *S. haemolyticus* needs to be further investigated. In agreement with other studies, we show in **paper I and II** that *S. haemolyticus* isolates have the ability to produce biofilm, and that different external factors are likely to be involved in biofilm regulation. Biofilm production was not dependent on *ica* genes in our collection of *S. haemolyticus* isolates.

5.5 Staphylococcal surface proteins

As part of this thesis we aimed to characterize the surface proteins of *S. haemolyticus*, and for this we used two different approaches; WGS and surface shaving. WGS reveals the blueprint and the genetic potential of bacteria, but the presence of a gene does not mean it is expressed or even functional. In contrast, in the surface shaving approach, the proteins identified are indeed functional and expressed in the setting tested. In order to perform proteomic analysis, like surface shaving, knowledge about the genome, as obtained from WGS, is a prerequisite.

Surface proteins in staphylococcal species often contain long stretches of repeated DNA, which is difficult to assemble correctly with only short sequence reads at hand. Despite these limitations we found some interesting surface proteins (**paper I** and **paper II)**, partly with hitherto unknown function. In **paper I** we identified SH0326 described as "streptococcal hemagglutinin-like protein[22] in the majority of the clinical isolates. Takeuchi *et al*. predicted SH0326 to play a role in bacterial cell agglutination as a mutant strain with a deleted region of 21 orfs, including SH0326 led to loss of agglutination[22]. Upon further investigation we identified this gene to be a homolog of the serine-rich adhesin for platelets, *sraP,* in *S. aureus*. SraP may be associated with or play a part in biofilm formation, as has been shown for the streptococcal homolog, Fab1[186]. Homologs have also been identified in *Lactococcus* species[439], but is to date best studied in several streptococcal species[178,179,181]. The protein appears to have its own secretion apparatus; the accessory secretion system. In JCSC1435 SH0326 (*sraP)* is 10.827 bp (3.608 amino acids) long, including a long repeat sequence in its central part. In our isolates, we were only able to detect the extreme N and C terminal of this gene. However, based on detection

of the C-terminal of *sraP* directly upstream of the accessory sec operon, in addition to the N-terminal of *sraP* somewhere in the genome, we were able to predict the presence and potential function of this gene. We found the accessory sec system to be present in 96% of the clinical isolates, and 74% of the commensal isolates, suggesting an important role in the clinical isolates, perhaps essential in the niche they occupy. In *S. aureus,* SraP has been shown to bind to platelets, and has been associated with infective endocarditis[173,440]. In *S. epidermidis* SraP has yet to be functionally described but Zhou and Wu described the structure based purely based on genome mining[178]. Moreover, Meric *et al.* showed that the accessory sec gene *asp3* and *sraP* were both significantly associated with virulence in *S. epidermidis*[208].

We were not able to detect SraP after surface shaving of the clinical isolate co-incubated with HaCat cells **(paper II);** either it was not expressed under these conditions or it was expressed but not identified. There are several potential explanations for this. Biologically, homologs of this protein have been shown to adhere to blood platelets[173]. SraP might not be expressed under the conditions used in the surface shaving experiment. Additionally, we showed in **paper I** that several commensal lacks *sraP* and *accessory sec*, strongly suggesting it is not needed on the skin. Technically, glycoproteins are also more difficult to detect by mass spectrometry than other proteins. Even though SraP is a huge protein, the majority of the protein consists of repeat sequences without any trypsin cut sites. Second, due to the variety of glycans that can be present on the same glycosylation site, different protein glycoforms exist for each protein resulting in a lower abundance for each glycoform compared with non-glycosylated peptides[441].

Higher prevalence of *S. haemolyticus* among neonates compared to adults has been reported by several studies[57,59,442]. An association has also been shown between the hospital adapted CoNS colonization and the risk of developing neonatal sepsis[59,443]. In *S. aureus* glycoproteins have been shown to play a role in sepsis, by adhesion and disruption of platelets[444]. *Accessory sec* and *sraP* were highly prevalent among the clinical *S. haemolyticus* isolates (97 %), while this system was absent in several commensal isolates (26%, mainly absent in isolates from the diverse clade F). This suggests that SraP might provide an important function for the clinical isolates, and it is possible it also includes binding to human platelets as seen for *S. aureus* SraP, and hence also plays a role in sepsis development.

In the surface shaving experiment performed as part of **paper II** we detected five surface proteins containing either or both LPXTG and YSIRK motifs. The two proteins with both

motifs were serine-aspartate repeat-containing (sdr-like) proteins. Sdr-protein have comparable structural organization, however they are not closely related, with only 20-30 % identical amino acid residues[445]. Different sdr proteins have been shown to have different roles in pathogenicity. Hence, the role of the two sdr-like proteins identified on *S. haemolyticus* surface needs to be further investigated before a function can be proposed.

We conclude that exploring and defining the surface proteins of pathogenic bacteria is an important step in increasing the understanding of the species, and the structure of surface proteins are important targets for new treatment, such as virulence inhibitors and vaccine candidates.

5.6 Immune evasion

Microbial immune evasion strategies may not be generally categorized as virulence-factors, but their importance in pathogenicity is unquestionable. The polysaccharide capsule has been associated with biofilm formation in bacteria[446–448]. The staphylococcal capsule is important in the pathogenesis of *S. aureus* infections, as the capsule enhances virulence by evading phagocytosis which results in bacterial persistence in the bloodstream of infected hosts[118,449,450]. The polysaccharide capsule modulates *S. aureus* adherence to endothelial surfaces *in vitro* and animal studies suggests that it also promotes bacterial colonization and persistence on mucosal surfaces[118]. The protective properties of the *S. haemolyticus* capsule was demonstrated by Flahaut *et al*[123]. In **paper I** we investigated the presence and prevalence of this capsule in our strain collection. We identified a capsule operon identical to the one described by Takeuchi *et al.* in 17 of the 169 isolates, all of which were clinical isolates[22]. In addition, we identified three novel capsule operons. All four putative operons were homologs in *capA-capG*, while *capH-capK* were unique for each capsule type. This observed variable region is also the variable region in *S. aureus* capsule, and it is what separates CP5 from CP8[118]. Furthermore, the three novel capsule operons contained the region *capL-capP* which the JSCS1435 capsule lacked, and these genes were homologs to the *S. aureus* CP5/8. The GC content of this variable region *capH-capK* was significantly lower than the surrounding cap genes, indicating the variable region were acquired by HGT. The novel putative capsule operons appeared to be clade specific, and the commensal isolates containing a capsule operon all belonged to one type. We also identified a capsule operon in isolates forming the new species (**paper III**). The putative capsule operon of the novel specie displayed a different

organization and order of the capsule genes compared to the four types identified in *S. haemolyticus*. All five *S. borealis* isolates displayed the identical capsule operon structure. The structure of the different capsule operons can be viewed in figure 12.

In **paper II**, a Toll/interleukin-1 like (TIR) domain protein was significantly upregulated after *S. haemolyticus* co-incubation with HaCaT cells. TIR domain containing proteins have been reported in several pathogenic bacteria[451–453]. Upon further investigation we identified the TIR domain protein to be identical to the TirS identified in *S. aureus* MSSA476. TirS in *S. aureus* has been shown to increase survival in the host[452]. TIR proteins have not been described in *S. haemolyticus* previously, however the gene encoding the TIR protein was only identified in three *S. haemolyticus* isolates, and in one *S. borealis* isolate. A recent study investigated the presence of *tirS* in a *S. aureus* collection of 226 isolates and found the gene in 12.4 % of the isolates restricted to only three CCs (CC1, CC5 and CC8). The authors showed that the *tirS*-mutant conferred superior virulence compared to wildtype (WT) underlining the role of TirS in immune modulation[454]. Additionally, they showed that *tirS* was located in SCC in close proximity of *fusC*, and that sub-inhibitory concentrations of fusidic acid increased *tirS* expression[454]. The fact that *tirS/fusC* in our isolates were identical to *S. aureus* strongly suggest a transfer event between *S. aureus* and *S. haemolyticus*. Our four isolates with *tirS/fusC* were not closely related (different clades/different countries) suggesting independent horizontal events.

Another protein found on the surface of *S. haemolyticus* (**paper II**) shared homology with the SasH/AdsA protein in *S. aureus,* a protein shown to promote survival of *S. aureus* within neutrophils[455]. SasH has also been positively associated with disease isolates of *S. aureus*[456].

5.7 Other virulence determinants

Biofilm, surface adhesins and capsule can all contribute to the virulence potential of a bacterium. Another important virulence factors for staphylococci, and perhaps especially in *S. aureus*, are toxins[131,457]. CoNS are generally portrayed as a group with little toxin-production, however few studies have addressed this topic[21,458].

Toxins

All *S. haemolyticus* isolates (n= 169) displayed haemolysis on TSA blood agar plates. In **paper III**, haemolysis testing of the isolates proposed as the new species *S. borealis* was performed. The *S. borealis* isolates, in addition to JSCS1435 were plated on p-agar with 5 % horse blood. All *S. borealis* isolates displayed a more pronounced haemolysis compared to *S. haemolyticus* (figure 11). However, performing the same experiment on TSA with 5 % horse blood revealed the opposite, showing the *S. haemolyticus* strain to induce the strongest haemolysis. Clearly, the medium plays a big role, and this observation suggests that regulation of the genes responsible for haemolysis is different between the proposed *S. borealis* strains and *S. haemolyticus*. The genetic background responsible for haemolysis in *S. haemolyticus* or the new species *S. borealis* is not clear. To date no homologs of beta or delta toxin gene contributing to haemolysis in *S. aureus* have been reported for *S. haemolyticus*.

The phenol soluble modulins (PSM) provide a better explanation for genes responsible for haemolysis[63]. It was suggested by Da *et al*. that the haemolysis observed in *S. haemolyticus* is a result, at least in part, due to PSMα, as the amount of PSMα produced by each strain tested correlated well with the degree of haemolysis observed. The authors also found that the haemolytic capacity was paired with pronounced cytolytic capacity toward human neutrophils, indicating a role of particularly PSMα in the immune evasion properties of *S. haemolyticus*[63]. In addition, the authors proposed, based on the conserved structure of PSMs, and the fact that all PSM types have been shown to activate immune cells, that other PSMs may also promote sepsis[63]. As part of **paper I**, we compared the presence and prevalence of PSMß, and the more recently identified *S. haemolyticus* specific PSM. In *S. haemolyticus* five PSMß genes are found, with the structure ß3-ß2-ß3-ß2-ß1. This cluster was found in all isolates, but some commensal isolates in the diverse clade F had additional copies of some of these genes. One can speculate that perhaps replication of some of these genes could provide benefit in competition against other bacterial species on the skin. PSMα was identified in all our strains, suggesting this gene to have a potentially important function for this species. We did not detect any other toxins than the PSMs in any of our isolates.

Staphyloxanthin

In **paper III** we showed that all five isolates of the proposed new species *S. borealis* had yellow pigmentation, a feature we did not observe for any of the other *S. haemolyticus* isolates which all displayed an off-white/grey colour. The yellow pigment of some staphylococcal species comes from carotenoid pigment staphyloxanthin, which is an important virulence factor in *S. aureus*[459]. The pigment acts as an antioxidant, enabling the detoxification of host immune system-generated reactive oxygen species (ROS)[460,461]. *S. aureus* species lacking staphyloxanthin grow normally but they are rapidly killed by ROS from host neutrophils and are deficient in skin abscess formation[461]. We identified the staphyloxanthin gene cluster in both the *S. haemolyticus* isolates and the five *S. borealis* isolates. However, *S. borealis* isolates had a *tetR*-like transcriptional regulator three genes upstream of the staphyloxanthin gene cluster. This *tetR*-like gene was not detected in any of the *S. haemolyticus* isolates, and could provide a potential explanation for the observed yellow pigmentation in *S. borealis*.

Urease

In **paper III** we also showed that the five *S. borealis* isolates were all urease positive by biochemical testing, and the urease operon was also confirmed genetically. One of the main biochemical traits separating *S. haemolyticus* from *S. warneri* is the production of urease, whereas *S. haemolyticus* is urease negative[462]. Urease is crucial for niche adaption of many bacterial pathogens[463], and is an important virulence factor for several uropathogenic bacterial species, including *S. saprophyticus*[464–466]. The majority of *S. aureus* strains are urease producing[467], and a recent study established that urease is not only the primary component of the acid response network but also an important factor required for persistent murine renal infection[463]. By increasing the pH, it also counteracts the slightly acidic pH caused by lactic acid secreted in the human skin, enabling increased survival on the skin[463]. Four out of five *S. borealis* isolates were of commensal origin, colonizing the skin, thus counteracting the acidic pH by producing urease might be an important colonizing strategy.

6 Concluding remarks and future aspects

With the rise of nosocomial infections and the fast-approaching post-antibiotic era soon upon us, we are in dire needs of faster and accurate diagnosis in addition to new or alternative treatment options. Finding a good clinical marker that easily determines whether the isolated bacteria is the infecting agent or an innocent contamination, can aid in accurate and fast diagnosis. Additionally, new targets for treatments, like anti-virulence factors – where the bacteria are disarmed instead of killed, and thereby less likely to quickly develop resistance - is sorely needed to counteract the growing antibiotic resistance crisis.

Whole genome sequencing offers the most comprehensive view of genomic information and associated biological implications. A large part of this thesis has focused on the genomic content of *S. haemolyticus* and previously unknown information has emerged as a result. We have confirmed the MDR nature of *S. haemolyticus*, revealing the ARGs responsible for the observed phenotype, but also several ARGs where phenotype has not been investigated. MDR is a common trait of clinically isolated *S. haemolyticus*, a feature not common among the commensal counterparts. Additionally, we found several other pathogenicity related genes that were positively associated with clinical isolates of *S. haemolyticus*. These findings include, but are not limited to; IS256 (previously associated with ARG and nosocomial infections in both staphylococci and enterococci), a capsule operon (shown to be protective for phagocytosis in both *S. haemolyticus* and *S. aureus*), *sraP* (a glycoprotein implicated as important in infection for both staphylococci and streptococci), and the ARGs *aacA-aphD* and *mecA*.

Through genomic inspection we found five isolates differing from the other *S. haemolyticus* to such an extent that we suggested classification of a new species. We have named this species *Staphylococcus borealis*. Pangenome analysis shows that, while related, *S. borealis* differs substantially from *S. haemolyticus*, with over a third of their genome being unique to *S. borealis*. There are also observable phenotypic and morphological differences between the new species and *S. haemolyticus,* seen in pigmentation, haemolysis, and urease production.

We also developed a method to investigate the *S. haemolyticus* surface proteome after colonisation of human keratinocytes. We show, for the first time in *S. haemolyticus*, several interesting surface proteins with potential roles in adhesion, biofilm and immune

evasion. These detected surface proteins will need to be further explored to characterize their precise function in host microbe interaction.

We are currently developing a molecular toolbox for genetic manipulation in *S. haemolyticus*, and we plan to follow up our findings with functional studies of the most interesting genetic markers of virulence that emerged during the course of this thesis.

When we embarked on this study, there was a substantial knowledge gap regarding *S. haemolyticus*. It was known that it was a nosocomial pathogen, but data relating to **how** it caused infection was extremely sparse. At the end of this study, we have done the largest and most comprehensive analysis of a *S. haemolyticus* collection to date. We have shown that clinical *S. haemolyticus* isolates show a clear genetic signature and possesses several genes described in other species as virulence factors. This study has enabled the identification of proteins involved in adhesion and colonisation, and the identification of genes potentially responsible for the traits that make this species well adapted in hospital settings. We have proposed several potential **clinical markers** and **targets for therapy**, and described a new species, *Staphylococcus borealis*. Our contributions provide new knowledge on both *S. haemolyticus* and *S. borealis*, and lay the foundation for more work to be done.

References

1. Darouiche, R. O. Device-Associated Infections: A macroproblem that starts with microadherence. *Clin. Infect. Dis.* **33**, 1567–1572 (2001).

2. Khan, H. A., Baig, F. K. & Mehboob, R. Nosocomial infections: Epidemiology, prevention, control and surveillance. *Asian Pac. J. Trop. Biomed.* **7**, 478–482 (2017).

3. Flemming, A. On the antibacterial action of cultures of a penicillum, with special reference to their use in the isolation of B. influenzae. *Br. J. Exp. Pathol.* **10**, (1929).

4. O'Brien, T. F. Emergence, spread, and environmental effect of antimicrobial resistance: How use of an antimicrobial anywhere can increase resistance to any antimicrobial anywhere else. *Clin. Infect. Dis.* **34**, S78–S84 (2002).

5. Holden, M. T. G. *et al.* A genomic portrait of the emergences, evolution, and global spread of methicillin-resistant *Staphylococcus aureus* pandemic. *Genome Res.* **23**, 653–664 (2013).

6. Baker, S., Thomson, N., Weill, F.-X. & Holt, K. E. Genomic insight into the emergence and spread of antimicrobial-resistant bacterial pathogens. *Science.* **260**, 733–738 (2017).

7. Cassini, A. *et al.* Attributable deaths and disability-adjusted life-years caused by infections with antibiotic-resistant bacteria in the EU and the European Economic Area in 2015: a population-level modelling analysis. *Lancet Infect. Dis.* **19**, 56–66 (2019).

8. Tacconelli, E. *et al.* Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis. *Lancet Infect. Dis.* **18**, 318–327 (2018).

9. Luepke, K. H. *et al.* Past, present, and future of antibacterial economics: increasing bacterial resistance, limited antibiotic pipeline, and societal implications. *Pharmacotherapy* **37**, 71–84 (2017).

10. Hensgens, M. P. M., Goorhuis, A., Dekkers, O. M. & Kuijper, E. J. Time interval of increased risk for *Clostridium difficile* infection after exposure to antibiotics. *J.*

Antimicrob. Chemother. **67**, 742–748 (2012).

11. Singh, N. K. *et al.* Multi-drug resistant *Enterobacter bugandensis* species isolated from the International Space Station and comparative genomic analyses with human pathogenic strains. *BMC Microbiol.* **18**, 1–13 (2018).

12. Ghilamicael, A. M., Boga, H. I., Anami, S. E., Mehari, T. & Budambula, N. L. M. Potential human pathogenic bacteria in five hot springs in Eritrea revealed by next generation sequencing. *PLoS One* **13**, 1–12 (2018).

13. Dancer, S. J. Controlling hospital-acquired infection: Focus on the role of the environment and new technologies for decontamination. *Clin. Microbiol. Rev.* **27**, 665–690 (2014).

14. Prestinaci, F., Pezzotti, P. & Pantosti, A. Antimicrobial resistance: A global multifaceted phenomenon. *Pathog. Glob. Health* **109**, 309–318 (2015).

15. Chip Bok's Editorial Cartoons. Available at: http://www.theeditorialcartoons.com/store/add.php?iid=19031.

16. Rosenbach, F. J. Mikroorganismen bei den Wundinfections-Krank- heiten des Menschen Wiesbaden. *Germany* (1884).

17. MacFadyen, A. C. *et al. Staphylococcus caeli* sp. nov., isolated from air sampling in an industrial rabbit holding. *Int. J. Syst. Evol. Microbiol.* **69**, 82–86 (2019).

18. Naushad, S. *et al. Staphylococcus debuckii* sp. nov., a coagulase-negative species from bovine milk. *Int. J. Syst. Evol. Microbiol.* **69**, 2239–2249 (2019).

19. MacFadyen, A. C. *et al. Staphylococcus pseudoxylosus* sp. nov., isolated from bovine mastitis. *Int. J. Syst. Evol. Microbiol.* **69**, 2208-2213 (2019).

20. Schleifer, K.-H. & Bell, J. A. *Staphylococcus. In: Bergey DH, Harrison FC, Breed RS, Hammer BW, Huntoon FM (eds) Bergeys manual of systematic archaea and bacteria* (In Bergey's Manual of Systematics of Archaea and Bacteria (eds W. B. Whitman, F. Rainey, P. Kämpfer, M. Trujillo, J. Chun, P. DeVos, B. Hedlund and S. Dedysh), 2015).

21. Becker, K., Heilmann, C. & Peters, G. Coagulase-negative staphylococci. *Clin. Microbiol. Rev.* **27**, 870–926 (2014).

22. Takeuchi, F. *et al.* Whole-genome sequencing of *Staphylococcus haemolyticus*

uncovers the extreme plasticity of its genome and the evolution of human-colonizing Staphylococcal species. *J. Bacteriol.* **187**, 7292–7308 (2005).

23. Hiramatsu, K. *et al.* Multi-drug-resistant *Staphylococcus aureus* and future chemotherapy. *J. Infect. Chemother.* **20**, 593–601 (2014).

24. Ito, T. *et al.* Structural comparison of three types of staphylococcal cassette chromosome mec integrated in the chromosome in methicillin-resistant *Staphylococcus aureus*. *Antimicrob. Agents Chemother.* **45**, 1323–1336 (2001).

25. Holden, M. T. G. *et al.* Complete genomes of two clinical *Staphylococcus aureus* strains: Evidence for the evolution of virulence and drug resistance. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 9786–9791 (2004).

26. Bien, J., Sokolova, O. & Bozko, P. Characterization of virulence factors of *Staphylococcus aureus* : Novel function of known virulence factors that are implicated in activation of airway epithelial proinflammatory response . *J. Pathog.* **2011**, 1–13 (2011).

27. Powers, M. E. & Wardenburg, J. B. Igniting the fire: *Staphylococcus aureus* virulence factors in the pathogenesis of sepsis. *PLoS Pathog.* **10**, 10–13 (2014).

28. Lacey, K. A., Geoghegan, J. A. & McLoughlin, R. M. The role of *Staphylococcus aureus* virulence factors in skin infection and their potential as vaccine antigens. *Pathogens* **5**, (2016).

29. Tong, S. Y. C., Davis, J. S., Eichenberger, E., Holland, T. L. & Fowler, V. G. *Staphylococcus aureus* infections: Epidemiology, pathophysiology, clinical manifestations, and management. *Clin. Microbiol. Rev.* **28**, 603–661 (2015).

30. Piette, A. & Verschraegen, G. Role of coagulase-negative staphylococci in human disease. *Vet. Microbiol.* **134**, 45–54 (2009).

31. Von Eiff, C., Peters, G. & Heilmann, C. Pathogenesis of infections due to coagulase-negative staphylococci. *Lancet Infect. Dis.* **2**, 677–685 (2002).

32. Klingenberg, C. *et al.* Persistent strains of coagulase-negative staphylococci in a neonatal intensive care unit: Virulence factors and invasiveness. *Clin. Microbiol. Infect.* **13**, 1100–1111 (2007).

33. Rogers, K., Fey, P. & Rupp, M. Coagulase-negative staphylococcal infections.

Infect. Dis. Clin. North Am. **23**, 73–98 (2009).

34. Didier, C. *et al.* Late-onset neonatal infections: Incidences and pathogens in the era of antenatal antibiotics. *Eur. J. Pediatr.* **171**, 681–687 (2012).

35. D'Angio, C. T., McGowan, K. L., Baumgart, S., Geme, J. S. & Harris, M. C. Surface colonization with coagulase-negative staphylococci in premature neonates. *J. Pediatr.* **114**, 1029–1034 (1989).

36. Wisplinghoff, H. *et al.* Nosocomial bloodstream infections in US hospitals: Analysis of 24,179 cases from a prospective nationwide surveillance study. *Clin. Infect. Dis.* **39**, 309–317 (2004).

37. Dong, H., Cao, H. & Zheng, H. Pathogenic bacteria distributions and drug resistance analysis in 96 cases of neonatal sepsis. *BMC Pediatr.* **17**, 1–6 (2017).

38. Pereira, P. M. A. *et al. Staphylococcus haemolyticus* disseminated among neonates with bacteremia in a neonatal intensive care unit in Rio de Janeiro, Brazil. *Diagn. Microbiol. Infect. Dis.* **78**, 85–92 (2014).

39. Czekaj, T., Ciszewski, M. & Szewczyk, E. M. *Staphylococcus haemolyticus* – An emerging threat in the twilight of the antibiotics age. *Microbiol. (United Kingdom)* **161**, 2061–2068 (2015).

40. Chang, P. H. *et al.* Clinical features, outcomes, and molecular characteristics of an outbreak of *Staphylococcus haemolyticu*s infection, among a mass-burn casualty patient group, in a tertiary center in northern Taiwan. *J. Microbiol. Immunol. Infect.* **51**, 847-855 (2018).

41. Singh, S. *et al.* Coagulase-negative staphylococci causing blood stream infection at an Indian tertiary care hospital: Prevalence, antimicrobial resistance and molecular characterisation. *Indian J. Med. Microbiol.* **34**, 500–505 (2016).

42. Nanoukon, C. *et al.* Pathogenic features of clinically significant coagulase-negative staphylococci in hospital and community infections in Benin. *Int. J. Med. Microbiol.* **307**, 75–82 (2017).

43. Pedroso, S. H. S. P. *et al.* Biofilm and toxin profile: A phenotypic and genotypic characterization of coagulase-negative staphylococci isolated from human bloodstream infections. *Microb. Pathog.* **100**, 312–318 (2016).

44. Hope, R., Livermore, D. M., Brick, G., Lillie, M. & Reynolds, R. Non-susceptibility trends among staphylococci from bacteraemias in the UK and Ireland, 2001-06. *J. Antimicrob. Chemother.* **62**, ii65–ii74 (2008).

45. Kresken, M. *et al.* Resistance trends and in vitro activity of tigecycline and 17 other antimicrobial agents against Gram-positive and Gram-negative organisms, including multidrug-resistant pathogens, in Germany. *Eur. J. Clin. Microbiol. Infect. Dis.* **30**, 1095–1103 (2011).

46. Barros, E. M., Ceotto, H., Bastos, M. C. F., Dos Santos, K. R. N. & Giambiagi-deMarval, M. *Staphylococcus haemolyticus* as an important hospital pathogen and carrier of methicillin resistance genes. *J. Clin. Microbiol.* **50**, 166–168 (2012).

47. Cavanagh, J. P. *et al.* Whole-genome sequencing reveals clonal expansion of multiresistant *Staphylococcus haemolyticus* in European hospitals. *J. Antimicrob. Chemother.* **69**, 2920–2927 (2014).

48. Seng, R., Leungtongkam, U., Thummeepak, R., Chatdumrong, W. & Sitthisak, S. High prevalence of methicillin-resistant coagulase-negative staphylococci isolated from a university environment in Thailand. *Int. Microbiol.* **20**, 65–73 (2017).

49. Rajkumar, S. *et al.* Prevalence and genetic mechanisms of antimicrobial resistance in Staphylococcus species: A multicentre report of the indian council of medical research antimicrobial resistance surveillance network. *Indian J. Med. Microbiol.* **35**, 53–60 (2017).

50. Bouchami, O., De Lencastre, H. & Miragaia, M. Impact of insertion sequences and recombination on the population structure of *Staphylococcus haemolyticus. PLoS One* **11**, 1–17 (2016).

51. Seng, R. *et al.* Biofilm formation of methicillin-resistant coagulase negative staphylococci (MR-CoNS) isolated from community and hospital environments. *PLoS One* **12**, (2017).

52. Dziri, R. *et al.* High prevalence of Staphylococcus haemolyticus and *Staphylococcus saprophyticus* in environmental samples of a Tunisian hospital. *Diagn. Microbiol. Infect. Dis.* **85**, 136–140 (2016).

53. Alharbi, S. A. *et al.* Assessment of the bacterial contamination of hand air dryer in washrooms. *Saudi J. Biol. Sci.* **23**, 268–271 (2016).

54. Xu, Z., Mkrtchyan, H. V. & Cutler, R. R. Antibiotic resistance and *mecA* characterization of coagulase-negative staphylococci isolated from three hotels in London, UK. *Front. Microbiol.* **6**, 1–6 (2015).

55. Wojtyczka, R. D. *et al.* Biofilm formation and antimicrobial susceptibility of *Staphylococcus epidermidis* strains from a hospital environment. *Int. J. Environ. Res. Public Health* **11**, 4619–4633 (2014).

56. Cavanagh, J. P. *et al.* Antimicrobial susceptibility and body site distribution of community isolates of coagulase-negative staphylococci. *Apmis* **124**, 973–978 (2016).

57. Hira, V. *et al.* Colonization dynamics of antibiotic-resistant coagulase-negative staphylococci in neonates. *J. Clin. Microbiol.* **51**, 595–597 (2013).

58. Jain, A., Agarwal, A., Verma, R., Awasthi, S. & Shingh, K. Intavenous device associated blood stream staphylococcal infection in paediatric patients. *Indian J. Med. Reserach* **134**, 193–199 (2011).

59. Ahmed, D. M., Abel Messih, M. A. W., Ibrahim, N. H., Meabed, M. H. & Abdel-Salam, S. M. Frequency of *icaA* and *icaD* determinants and biofilm formation among coagulase-negative staphylococci associated with nasal carriage in neonatal intensive care units. *Germs* **9**, 61–70 (2019).

60. Fredheim, E. G. A. *et al.* Biofilm formation by *staphylococcus haemolyticus*. *J. Clin. Microbiol.* **47**, 1172–1180 (2009).

61. Stewart, P. S. & Costerton, J. W. Antibiotic resistance of bacteria in biofilms. *Lancet* **358**, 135–138 (2001).

62. Panda, S. & Singh, D. V. Biofilm Formation by ica-Negative Ocular Isolates of *Staphylococcus haemolyticus*. *Front. Microbiol.* **9**, 1–9 (2018).

63. Da, F. *et al.* Phenol-Soluble Modulin toxins of *Staphylococcus haemolyticus*. *Front. Cell. Infect. Microbiol.* **7**, 1–11 (2017).

64. Pinheiro, L. *et al. Staphylococcus epidermidis* and *Staphylococcus haemolyticus*: Molecular detection of cytotoxinand enterotoxin genes. *Toxins (Basel).* **7**, 3688–3699 (2015).

65. Price, S. B. & Flournoy, D. J. Comparison of antimicrobial susceptibility patterns

among coagulase-negative staphylococci. *Antimicrob. Agents Chemother.* **21**, 436–440 (1982).

66. Froggatt, J. W., Johnston, J. L., Galetto, D. W. & Archer, G. L. Antimicrobial resistance in nosocomial isolates of *Staphylococcus haemolyticus*. *Antimicrob. Agents Chemother.* **33**, 460–466 (1989).

67. Hosseinkhani, F., Buirs, M. T., Jabalameli, F., Emaneini, M. & Leeuwen, W. B. Van. High diversity in SCC mec elements among multidrug-resistant *Staphylococcus haemolyticus* strains originating from paediatric patients ; characterization of a new composite island. *J. Med. Microbiol.* **7**, 915-921 (2018).

68. Lerbech, A. M. *et al.* Antibiotic exposure in a low-income country: Screening urine samples for presence of antibiotics and antibiotic resistance in coagulase negative staphylococcal contaminants. *PLoS One* **9**, 1–18 (2014).

69. Louagie, H. *et al.* Problems in diagnosis and treatment of *Staphylococcus haemolyticus* endocarditis in a hemodialysis patient. *Clin. Microbiol. Infect.* **4**, 44–48 (1998).

70. Veach, L. A., Pfaller, M. A., Barrett, M., Koontz, F. P. & Wenzel, R. P. Vancomycin resistance in *Staphylococcus haemolyticus* causing colonization and bloodstream infection. *J. Clin. Microbiol.* **28**, 2064–2068 (1990).

71. Cunningham, R., Gurnell, M., Bayston, R., Cockayne, A. & Shelton, A. Teicoplanin resistance in *Staphylococcus haemolyticus*, developing during treatment. *J. Antimicrob. Chemother.* **39**, 438–439 (1997).

72. Spanik, S., Trupl, J., Studena, M. & Krcmery Jr, V. Breakthrough nosocomial bacteraemia due to teicoplanin-resistant *Staphylococcus haemolyticus* in five patients with acute leukaemia. *J. Hosp. Infect.* **35**, 155–164 (1997).

73. Mazzariol, A. *et al.* Outbreak of linezolid-resistant *Staphylococcus haemolyticus* in an Italian intensive care unit. *Eur. J. Clin. Microbiol. Infect. Dis.* **31**, 523–527 (2012).

74. Biavasco, F., Vignaroli, C. & Varaldo, P. E. Glycopeptide resistance in coagulase-negative staphylococci. *Eur. J. Clin. Microbiol. Infect. Dis.* **19**, 403–417 (2000).

75. Kristóf, K. *et al.* Significance of methicillin-teicoplanin resistant *Staphylococcus*

haemolyticus in bloodstream infections in patients of the Semmelweis University hospitals in Hungary. *Eur. J. Clin. Microbiol. Infect. Dis.* **30**, 691–699 (2011).

76. Rodriguez-Aranda, A. *et al.* Nosocomial spread of linezolid-resistant *Staphylococcus haemolyticus* infections in an intensive care unit. *Diagn. Microbiol. Infect. Dis.* **63**, 398–402 (2009).

77. Cherifi, S. *et al.* Comparative epidemiology of *Staphylococcus epidermidis* isolates from patients with catheter-related bacteremia and from healthy volunteers. *J. Clin. Microbiol.* **51**, 1541–1547 (2013).

78. Larson, E. L., McGinley, K. J., Foglia, A. R., Talbot, G. H. & Leyden, J. J. Composition and antimicrobic resistance of skin flora in hospitalized and healthy adults. *J. Clin. Microbiol.* **23**, 604–608 (1986).

79. Oliveira, F. & Cerca, N. Antibiotic resistance and biofilm formation ability among coagulase-negative staphylococci in healthy individuals from portugal. *J. Antibiot. (Tokyo).* **66**, 739–741 (2013).

80. Rolo, J., de Lencastre, H. & Miragaia, M. Strategies of adaptation of *Staphylococcus epidermidis* to hospital and community: Amplification and diversification of SCCmec. *J. Antimicrob. Chemother.* **67**, 1333–1341 (2012).

81. Basso, A. P. *et al.* Antibiotic resistance and enterotoxin genes in *Staphylococcus* sp. Isolates from polluted water in southern Brazil. *An. Acad. Bras. Cienc.* **86**, 1813–1820 (2014).

82. Srinivasan, A., Dick, J. D. & Perl, T. M. Vancomycin resistance in Staphylococci. *Clin. Microbiol. Rev.* **15**, 430–438 (2002).

83. Levine, D. P. Vancomycin : A history. *Clin. Infect. Dis.* **42**, s5-12 (2006).

84. Asbell, P. A., Sahm, D. F., Shaw, M., Draghi, D. C. & Brown, N. P. Increasing prevalence of methicillin resistance on serious ocular infections caused by *Staphylococcus aureus* in the United States: 2000 to 2005. *J. Cataract Refract. Surg.* **34**, 814–818 (2008).

85. MacMorran, E. *et al.* The rise of methicillin resistant *Staphylococcus aureus*: Now the dominant cause of skin and soft tissue infection in Central Australia. *Epidemiol. Infect.* **145**, 2817–2826 (2017).

86. Moxnes, J. F., de Blasio, B. F., Leegaard, T. M. & Moen, A. E. F. Methicillin-resistant *Staphylococcus aureus* (MRSA) is increasing in Norway: A time series analysis of reported MRSA and methicillin-sensitive *S. aureus* cases, 1997-2010. *PLoS One* **8**, (2013).

87. Lutro, O. *et al.* Increasing resistance of coagulase-negative staphylococci in total hip arthroplasty infections: 278 THA-revisions due to infection reported to the norwegian arthroplasty register from 1993 to 2007. *Adv. Orthop.* **2014**, (2014).

88. Antibiotic Use in Hospitals, 2017 - CDC. Available at: www.cdc.gov/antibiotic-use/stewardship-report/hospital.html. (Accessed: 16th December 2019)

89. Siebert, W. T., Moreland, N. & Williams Jr, T. W. Synergy of vancomycin plus cefazolin or cephalothin against methicillin-resistant *Staphylococcus epidermidis*. *J. Infect. Dis.* **139**, 1979 (1979).

90. Tuazon, C. U. & Miller, H. Clinical and microbiologic aspects of serious infections cause by *Staphylococcus epidermidis*. *Scand. Jourrnal Infect. Dis.* **15**, 347–60 (1983).

91. Schwalbe, R., Stapleton, J. & Gilligan, P. Emergence of vancomycin resistance in coagulase-negative staphylococci. *N. Engl. J. Med.* **316**, 927–931 (1987).

92. Pinheiro, L. *et al.* Susceptibility profile of *Staphylococcus epidermidis* and *Staphylococcus haemolyticus* isolated from blood cultures to vancomycin and novel antimicrobial drugs over a period of 12 years. *Microb. Drug Resist.* **22**, 283–293 (2016).

93. Olivares, M. F. *et al.* Activity of vancomycin, ciprofloxacin, daptomycin and linezolid against coagulase-negative staphylococci bacteremia. *Rev. Esp. Quimioter.* **24**, 74–78 (2011).

94. Tabe, Y., Nakamura, A., Oguri, T. & Igari, J. Molecular characterization of epidemic multiresistant *Staphylococcus haemolyticus* isolates. *Diagn. Microbiol. Infect. Dis.* **32**, 177–183 (1998).

95. Herwaldt, L., Boyken, L. & Pfaller, M. In vitro selection of resistance to vancomycin in bloodstream isolates of *Staphylococcus haemolyticus* and *Staphylococcus aureus*. *Eur. J. Clin. Microbiol. Infect. Dis.* **10**, 1007–12 (1991).

96. Schwalbe, R. S., Ritz, W. T., Verma, P. R., Barranco, E. A. & Gilligan, P. H. Selection for vancomycin resistance in clinical isolates of *Staphylococcus haemolyticus*. *J. Infect. Dis.* **161**, 45–51 (1990).

97. Biavasco, F., Giovanetti, E., Montanari, M. P., Lupidi, R. & Varaldo, P. E. Development of in-vitro resistance to glycopeptide antibiotics: assessment in staphylococci of different species. *J. Antimicrob. Chemother.* **27**, 71–79 (1991).

98. Kim, J. W., Chung, G. T., Yoo, J. S., Lee, Y. S. & Yoo, J. Il. Autolytic activity and molecular characteristics of *Staphylococcus haemolyticus* strains with induced vancomycin resistance. *J. Med. Microbiol.* **61**, 1428–1434 (2012).

99. Hafer, C., Lin, Y., Kornblum, J., Lowy, F. D. & Uhlemann, A. C. Contribution of selected gene mutations to resistance in clinical isolates of vancomycin-intermediate *Staphylococcus aureus*. *Antimicrob. Agents Chemother.* **56**, 5845–5851 (2012).

100. Kato, Y., Suzuki, T., Ida, T. & Maebashi, K. Genetic changes associated with glycopeptide resistance in *Staphylococcus aureus*: Predominance of amino acid substitutions in YvqF/VraSR. *J. Antimicrob. Chemother.* **65**, 37–45 (2009).

101. Watanabe, Y., Cui, L., Katayama, Y., Kozue, K. & Hiramatsu, K. Impact of rpoB mutations on reduced vancomycin susceptibility in *Staphylococcus aureus*. *J. Clin. Microbiol.* **49**, 2680–2684 (2011).

102. Hu, Q., Peng, H. & Rao, X. Molecular events for promotion of vancomycin resistance in vancomycin intermediate *Staphylococcus aureus*. *Front. Microbiol.* **7**, (2016).

103. Arthur, M., Molinas, C., Depardieu, F. & Courvalin, P. Characterization of Tn*1546*, a Tn3-related transposon conferring glycopeptide resistance by synthesis of depsipeptide peptidoglycan precursors in *Enterococcus faecium* BM4147. *J. Bacteriol.* **175**, 117–127 (1993).

104. Shajari, G., Khorshidi, A. & Moosavi, G. Vancomycin resistance in *Staphylococcus aureus* strains. *Arch. Razi Inst.* **90**, 107–110 (2017).

105. Henwood, C. J. *et al.* Susceptibility of Gram-positive cocci from 25 UK hospitals to antimicrobial agents including linezolid. *J. Antimicrob. Chemother.* **46**, 931–940 (2000).

106. De Almeida, L. M., Lincopan, N., De Araújo, M. R. E. & Mamizuka, E. M. Clonal dissemination of linezolid-resistant *Staphylococcus haemolyticus* exhibiting the G2576T mutation in the 23S rRNA gene in a tertiary care hospital in Brazil. *Antimicrob. Agents Chemother.* **56**, 2792–2793 (2012).

107. Gupta, V., Garg, S., Jain, R., Garg, S. & Chander, J. Linezolid resistant *Staphylococcus haemolyticus*: First case report from India. *Asian Pac. J. Trop. Med.* **5**, 837–838 (2012).

108. Tarazona, R. E. R., Padilla, T. P., Gomez, J. C. S. & Sanchez, J. E. G. Firs report in Spain of linezolid non-suceptibility in a clinical isolate of *Staphylococcus haemolyticus*. *Int. J. Antimicrob. Agents* **30**, 277–278 (2007).

109. Rajan, V., Kumar, V. G. S. & Gopal, S. A cfr-positive clinical staphylococcal isolate from India with multiple mechanisms of linezolid-resistance. *Indian J. Med. Res.* **139**, 463–467 (2014).

110. Lozano, C. *et al.* Caracterización de cepas de *Staphylococcus epidermidis* y *S. haemolyticus* resistentes a meticilina y linezolid en un hospital español. *Enferm. Infecc. Microbiol. Clin.* **31**, 136–141 (2013).

111. Kim, N. D. Science and Ink. Available at: https://www.lab-initio.com/screen_res/nz083.jpg.

112. Wen, Z. & Zhang, J.-R. Bacterial capsules. in *Molecular medical biology* (ed. Sails, A.) 33–53 (Academic Press, 2014).

113. Bhasin, N. *et al.* Identification of a gene essential for O-acetylation of the *Staphylococcus aureus* type 5 capsular polysaccharide. *Mol. Microbiol.* **27**, 9–21 (1998).

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

115. Watts, A. *et al. Staphylococcus aureus* strains that express serotype 5 or serotype 8 capsular polysaccharides differ in virulence. *Infect. Immun.* **73**, 3502–3511 (2005).

116. Kuipers, A. *et al.* The *Staphylococcus aureus* polysaccharide capsule and Efb-

dependent fibrinogen shield act in concert to protect against phagocytosis. *Microbiol. (United Kingdom)* **162**, 1185–1194 (2016).

117. Mohamed, N. *et al.* Molecular epidemiology and expression of capsular polysaccharides in *Staphylococcus aureus* clinical isolates in the United States. *PLoS ONE.* **14**, (2019).

118. Riordan, K. O. & Lee, J. C. *Staphylococcus aureus* capsular polysaccharides. *Clin. Microbiol. Rev.* **17**, 218–234 (2004).

119. Yoshida, K. & Minegishi, Y. Isolation of an encapsulated strain of *Staphylococcus epidermidis*. *J. Appl. Bacteriol.* **47**, 299–301 (1979).

120. Ichiman, Y. & Yoshida, K. The relationshop of capsular-type of *Staphylococcus epidermidis* to virulence and induction of resistance in the mouse. *J. Appl. Bacteriol.* **51**, 229–241 (1981).

121. Yamada, T., Ichiman, Y. & Yoshida, K. Possible common biological and immunological properties for detecting encapsulated strains of *Staphylococcus epidermidis*. *J. Clin. Microbiol.* **26**, 2167–2172 (1988).

122. Park, S. *et al.* Characterization of the structure and biological functions of a capsular polysaccharide produced by *Staphylococcus saprophyticus*. *J. Bacteriol.* **192**, 4618–4626 (2010).

123. Flahaut, S. *et al.* Structural and biological characterization of a capsular polysaccharide produced by *Staphylococcus haemolyticus*. *J. Bacteriol.* **190**, 1649–1657 (2008).

124. Poutrel, B., Mendolia, C., Sutra, L. & Fournier, J. M. Reactivity of coagulase-negative staphylococci isolated from cow and goat milk with monoclonal antibodies to *Staphylococcus aureus* capsular polysaccharide types 5 and 8. *J. Clin. Microbiol.* **28**, 358–360 (1990).

125. Pinchuk, I. V., Beswick, E. J. & Reyes, V. E. Staphylococcal enterotoxins. *Toxins (Basel).* **2**, 2177–2197 (2010).

126. Zhang, S., Iandolo, J. J. & Stewart, G. C. The enterotoxin D plasmid of *Staphylococcus aureus* encodes a second enterotoxin determinant (sej). *FEMS Microbiol. Lett.* **168**, 227–233 (1998).

127. Lindsay, J. A., Ruzin, A., Ross, H. F., Kurepina, N. & Novick, R. P. The gene for toxic shock toxin is carried by a family of mobile pathogenicity islands in *Staphylococcus aureus*. *Mol. Microbiol.* **29**, 527–543 (1998).

128. da Cunha, M. de L. R. ., Calsolari, R. A. O. & Júnior, J. P. A. Detection of enterotoxin and toxic shock syndrome toxin 1 genes in *Staphylococcus* , with emphasis on coagulase-negative staphylococci. *Microbiol. Immunol.* **51**, 381–390 (2007).

129. Vandenesch, F., Lina, G. & Henry, T. *Staphylococcus aureus* hemolysins, bi-component leukocidins, and cytolytic peptides: A redundant arsenal of membrane-damaging virulence factors? *Front. Cell. Infect. Microbiol.* **2**, 1–15 (2012).

130. Molnar, C., Hevessy, Z., Rozgonyi, F., Hevessy, Z. & Gemmell, C. G. Pathogenicity and virulence of coagulase negative staphylococci in relation to adherence, hydrophobicity, and toxin production in vitro. *J. Clin. Pathol.* **47**, 743–748 (1994).

131. Oliveira, D., Borges, A. & Simões, M. *Staphylococcus aureus* toxins and their molecular activity in infectious diseases. *Toxins (Basel).* **10**, (2018).

132. Bhakdil, S. & Tranum-Jensen, J. Alpha-toxin of *Staphylococcus aureus*. *Microbiol. Rev.* **55**, 733–751 (1991).

133. Bhakdi, S. *et al.* Staphylococcal alpha toxin promotes blood coagulation via attack on human platelets. *J. Exp. Med.* **168**, 527–542 (1988).

134. Nygaard, T. K. *et al.* Alpha-toxin induces programmed cell death of human T cells, B cells, and monocytes during USA300 infection. *PLoS One* **7**, (2012).

135. Berube, B. J. & Wardenburg, J. B. *Staphylococcus aureus* α-toxin: Nearly a century of intrigue. *Toxins (Basel).* **5**, 1140–1166 (2013).

136. Huseby, M. J. *et al.* Beta toxin catalyzes formation of nucleoprotein matrix in staphylococcal biofilms. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 14407–14412 (2010).

137. Tajima, A., Iwase, T., Shinji, H., Seki, K. & Mizunoe, Y. Inhibition of endothelial interleukin-8 production and neutrophil transmigration by *Staphylococcus aureus* beta-hemolysin. *Infect. Immun.* **77**, 327–334 (2009).

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

1203 (2013).

139. Hayashida, A., Bartlett, A. H., Foster, T. J. & Park, P. W. *Staphylococcus aureus* beta-toxin induces lung injury through syndecan-1. *Am. J. Pathol.* **174**, 509–518 (2009).

140. Prevost, G. *et al.* Panton-valentine leucocidin and gamma-hemolysin from *Staphylococcus aureus* ATCC 49775 are encoded by distinct genetic loci and have different biological activities. *Infect. Immun.* **63**, 4121–4129 (1995).

141. Spaan, A. N. *et al. Staphylococcus aureus* targets the duffy antigen receptor for chemokines (DARC) to lyse erythrocytes. *Cell Host Microbe* **18**, 363–370 (2015).

142. Spaan, A. N. *et al.* The staphylococcal toxin γ-hemolysin AB and CB differentially target phagocytes by employing specific chemokine receptors. *Nat. Commun.* **5**, (2014).

143. Peschel, A. & Otto, M. Phenol-soluble modulins and staphylococcal infection. *Nat. Rev. Microbiol.* **11**, 667–673 (2013).

144. Cheung, G. Y. C., Joo, H.-S., Chatterjee, S. S. & Otto, M. Phenol-soluble modulins - critical determinants of staphylococcal virulence. *FEMS Microbiol. Rev.* **38**, 698–719 (2014).

145. Wang, R. *et al.* Identification of novel cytolytic peptides as key virulence determinants for community-associated MRSA. *Nat. Med.* **13**, 1510–1514 (2007).

146. Qin, L. *et al.* Toxin mediates sepsis caused by methicillin-resistant *Staphylococcus epidermidis*. *PLoS Pathog.* **13**, 1–16 (2017).

147. Wang, R. *et al. Staphylococcus epidermidis* surfactant peptides promote biofilm maturation and dissemination of biofilm-associated infection in mice. *J. Clin. Invest.* **121**, 238–248 (2011).

148. Periasamy, S. *et al.* How S*taphylococcus aureus* biofilms develop their characteristic structure. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 1281–1286 (2012).

149. Frenette, M., Beaudet, R., Bisaillon, J. G., Sylvestre, M. & Portelance, V. Chemical and biological characterization of a gonococcal growth inhibitor produced by *Staphylococcus haemolyticus* isolated from urogenital flora. *Infect. Immun.* **46**, 340–345 (1984).

150. Watson, D. C., Yaguchi, M., Bisaillon, J. G., Beaudet, R. & Morosoli, R. The amino acid sequence of a gonococcal growth inhibitor from *Staphylococcus haemolyticus*. *Biochem. J.* **252**, 87–93 (1988).

151. Frenette, M., Beaudet, R., Bisaillon, J. G. & Portelance, V. Site of action of a gonococcal growth inhibitor produced by *Staphylococcus haemolyticus*. *J. Med. Microbiol.* **26**, 199–204 (1988).

152. Cavanagh, J. P. *et al.* Comparative exoproteome profiling of an invasive and a commensal *Staphylococcus haemolyticus isolate*. *J. Proteomics* **197**, 106–114 (2019).

153. Foster, T. J., Geoghegan, J. A., Ganesh, V. K. & Höök, M. Adhesion, invasion and evasion: The many functions of the surface proteins of *Staphylococcus aureus*. *Nat. Rev. Microbiol.* **12**, 49–62 (2014).

154. Lacey, K. A. *et al.* The *Staphylococcus aureus* cell wall-anchored protein clumping factor A is an important T cell antigen. *Infect. Immun.* **85**, 1–12 (2017).

155. Heilbronner, S. *et al.* Genome sequence of *Staphylococcus lugdunensis* N920143 allows identification of putative colonization and virulence factors. *FEMS Microbiol. Lett.* **322**, 60–67 (2011).

156. Bowden, M. G. *et al.* Identification and preliminary characterization of cell-wall-anchored proteins of *Staphylococcus epidermidis*. *Microbiology* **151**, 1453–1464 (2005).

157. Bischoff, M. *et al.* Microarray-basd analysis of the Staphylococcus sigmaB regulon. **186**, 4085–4099 (2004).

158. Hammer, N. D. & Skaar, E. P. Molecular mechanisms of *Staphylococcus aureus* iron acquisition . *Annu. Rev. Microbiol.* **65**, 129–147 (2011).

159. McAleese, F. M., Walsh, E. J., Sieprawska, M., Potempa, J. & Foster, T. J. Loss of clumping factor B fibrinogen binding activity by *Staphylococcus aureus* involves cessation of transcription, shedding and cleavage by metalloprotease. *J. Biol. Chem.* **276**, 29969–29978 (2001).

160. Ortega-Peña, S., Martínez-García, S., Rodríguez-Martínez, S., Cancino-Diaz, M. E. & Cancino-Diaz, J. C. Overview of *Staphylococcus epidermidis* cell wall-

anchored proteins: potential targets to inhibit biofilm formation. *Mol. Biol. Rep.* **47**, 771-784 (2020).

161. Arrecubieta, C., Lee, M. H., Macey, A., Foster, T. J. & Lowy, F. D. SdrF, a *Staphylococcus epidermidis* surface protein, binds type I collagen. *J. Biol. Chem.* **282**, 18767–18776 (2007).

162. Trivedi, S. *et al.* The surface protein SdrF mediates *Staphylococcus epidermidis* adherence to keratin. *J. Infect. Dis.* **215**, 1846–1854 (2017).

163. Arrecubieta, C. *et al.* SdrF, a *Staphylococcus epidermidi*s surface protein, contributes to the initiation of ventricular assist device driveline-related infections. *PLoS Pathog.* **5**, (2009).

164. Davis, S. L., Gurusiddappa, S., McCrea, K. W., Perkins, S. & Höök, M. SdrG, a Fibrinogen-binding bacterial adhesin of the microbial surface components recognizing adhesive matrix molecules subfamily from *Staphylococcus epidermidis*, targets the thrombin cleavage site in the Bß chain. *J. Biol. Chem.* **276**, 27799–27805 (2001).

165. Ortega-Peña, S., Martínez-García, S., Rodríguez-Martínez, S., Cancino-Diaz, M. E. & Cancino-Diaz, J. C. Overview of Staphylococcus epidermidis cell wall-anchored proteins: potential targets to inhibit biofilm formation. *Mol. Biol. Rep.* (2019).

166. Arora, S., Uhlemann, A. C., Lowy, F. D. & Hook, M. A novel MSCRAMM subfamily in coagulase negative staphylococcal species. *Front. Microbiol.* **7**, 1–9 (2016).

167. Hussain, M., Herrmann, M., Von Eiff, C., Perdreau-Remington, F. & Peters, G. A 140-kilodalton extracellular protein is essential for the accumulation of *Staphylococcus epidermidis* strains on surfaces. *Infect. Immun.* **65**, 519–524 (1997).

168. Schaeffer, C. R. *et al.* Accumulation-associated protein enhances *Staphylococcus epidermidis* biofilm formation under dynamic conditions and is required for infection in a rat catheter model. *Infect. Immun.* **83**, 214–226 (2015).

169. Tormo, M. Á., Knecht, E., Götz, F., Lasa, I. & Penadés, J. R. Bap-dependent biofilm formation by pathogenic species of Staphylococcus: Evidence of horizontal

gene transfer? *Microbiology* **151**, 2465–2475 (2005).

170. Harris, L. G. *et al.* Biofilm morphotypes and population structure among *Staphylococcus epidermidis* from commensal and clinical samples. *PLoS One* **11**, 1–15 (2016).

171. Khodaparast, L. *et al.* The possible role of *staphylococcus epidermidis* LPxTG surface protein SesC in biofilm formation. *PLoS One* **11**, (2016).

172. Qi, X. *et al.* SesI may be associated with the invasiveness of *Staphylococcus epidermidis*. *Front. Microbiol.* **8**, 1–10 (2018).

173. Siboo, I. R., Chambers, H. F. & Sullam, P. M. Role of SraP, a serine-rich surface protein of *Staphylococcus aureus*, in binding to human platelets. *Infect. Immun.* **73**, 2273–2280 (2005).

174. Wu, H., Mintz, K. P., Ladha, M. & Fives-Taylor, P. M. Isolation and characterization of Fap1, a fimbriae-associated adhesin of *Streptococcus parasanguis* FW213. *Mol. Microbiol.* **28**, 487–500 (1998).

175. Bensing, B. A., Gibson, B. W. & Sullam, P. M. The *Streptococcus gordonii* platelet binding protein GspB undergoes glycosylation independently of export. *J. Bacteriol.* **186**, 638–645 (2004).

176. Prabudiansyah, I. & Driessen, A. J. M. The Canoncial and accessory sec system of Gram-positive bacteria. *Curr. Top. Microbiol. Immunol.* **404**, 45-67 (2017).

177. Braunstein, M., Brown, A. M., Kurtz, S. & Jacobs, J. Two nonredundant SecA homologues function in mycobacteria. *J. Bacteriol.* **183**, 6979–6990 (2001).

178. Zhou, M. & Wu, H. Glycosylation and biogenesis of a family of serine-rich bacterial adhesins. *Microbiology* **155**, 317–327 (2009).

179. Bandara, M. *et al.* The accessory Sec system (SecY2A2) in *Streptococcus pneumoniae* is involved in export of pneumolysin toxin, adhesion and biofilm formation. *Microbes Infect.* **19**, 402–412 (2017).

180. Bensing, B. A. & Sullam, P. M. An accessory sec locus of *Streptococcus gordonii* is required for export of the surface protein GspB and for normal levels of binding to human platelets. *Mol. Microbiol.* **44**, 1081–1094 (2002).

181. Spencer, C., Bensing, B. A., Mishra, N. N. & Sullam, P. M. Membrane trafficking

of the bacterial adhesin GspB and the accessory Sec transport machinery. *J. Biol. Chem.* **294**, 1502–1515 (2018).

182. Jakubovics, N. S. *et al.* Functions of cell surface-anchored antigen I/II family and Hsa polypeptides in interactions of *Streptococcus gordonii* with host receptors. *Infect. Immun.* **73**, 6629–6638 (2005).

183. Takamatsu, D., Bensing, B. A., Prakobphol, A., Fisher, S. J. & Sullam, P. M. Binding of the streptococcal surface glycoproteins GspB and Hsa to human salivary proteins. *Infect. Immun.* **74**, 1933–1940 (2006).

184. Obert, C. *et al.* Identification of a candidate *Streptococcus pneumoniae* core genome and regions of diversity correlated with invasive pneumococcal disease. *Infect. Immun.* **74**, 4766–4777 (2006).

185. Rose, L. *et al.* Antibodies against PsrP, a Novel *Streptococcus pneumoniae* adhesin, block adhesion and protect mice against pneumococcal challenge . *J. Infect. Dis.* **198**, 375–383 (2008).

186. Froeliger, E. H. & Fives-Taylor, P. *Streptococcus parasanguis* fimbria-associated adhesin fap1 is required for biofilm formation. *Infect. Immun.* **69**, 2512–2519 (2001).

187. Wu, H., Zeng, M. & Fives-Taylor, P. The glycan moieties and the N-terminal polypeptide backbone of a fimbria-associated adhesin, Fap1, play distinct roles in the biofilm development of *Streptococcus parasanguinis*. *Infect. Immun.* **75**, 2181–2188 (2007).

188. Vert, M. *et al.* Terminology for biorelated polymers and application (IUPAC Recommendation 2012)*. *Pure Appl. Chem.* **84**, 377–410 (2012).

189. Olsen, I. Biofilm-specific antibiotic tolerance and resistance. *Eur. J. Clin. Microbiol. Infect. Dis.* **34**, 877–886 (2015).

190. Williamson, K. S. *et al.* Heterogeneity in *Pseudomonas aeruginosa* biofilms includes expression of ribosome hibernation factors in the antibiotic-tolerant subpopulation and hypoxia-induced stress response in the metabolically active population. *J. Bacteriol.* **194**, 2062–2073 (2012).

191. Olsen, I., Tribble, G. D., Fiehn, N. E. & Wang, B. Y. Bacterial sex in dental

plaque. *J. Oral Microbiol.* **5**, (2013).

192. Mah, T. F. Regulating antibiotic tolerance within biofilm microcolonies. *J. Bacteriol.* **194**, 4791–4792 (2012).

193. Römling, U. & Balsalobre, C. Biofilm infections, their resilience to therapy and innovative treatment strategies. *J. Intern. Med.* **272**, 541–561 (2012).

194. Stewart, P. S. Antimicrobial tolerance in biofilms. *Microbiol. Spectr.* **3**, 1–30 (2015).

195. Otto, M. Staphylococcal Biofilms. *Microbiol. Spectr.* **6**, 1–17 (2018).

196. Büttner, H., Mack, D. & Rohde, H. Structural basis of *Staphylococcus epidermidis* biofilm formation: Mechanisms and molecular interactions. *Front. Cell. Infect. Microbiol.* **5**, 1–15 (2015).

197. Christner, M. *et al.* The giant extracellular matrix-binding protein of *Staphylococcus epidermidis* mediates biofilm accumulation and attachment to fibronectin. *Mol. Microbiol.* **75**, 187–207 (2010).

198. Heilmann, C., Hussain, M., Peters, G. & Götz, F. Evidence for autolysin-mediated primary attachment of *Staphylococcus epidermidis* to a polystyrene surface. *Mol. Microbiol.* **24**, 1013–1024 (1997).

199. Qin, Z. *et al.* Role of autolysin-mediated DNA release in biofilm formation of *Staphylococcus epidermidis*. *Microbiology* **153**, 2083–2092 (2007).

200. Christner, M. *et al.* SarA negatively regulates *Staphylococcus epidermidis* biofilm formation by modulating expression of 1MDa extracellular matrix binding protein and autolysis-dependent release of eDNA. *Mol. Microbiol.* **86**, 394–410 (2012).

201. Giormezis, N. *et al.* Coagulase-negative staphylococcal bloodstream and prosthetic-device-associated infections: The role of biofilm formation and distribution of adhesin and toxin genes. *J. Med. Microbiol.* **63**, 1500–1508 (2014).

202. Pereira-Ribeiro, P. M. *et al.* Influence of antibiotics on biofilm formation by different clones of *Staphylococcus haemolyticus*. *Future Microbiol.* **14**, (2019).

203. Tettelin, H. *et al.* Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: Implications for the microbial 'pan-genome'. *Proc. Natl. Acad. Sci.* **102**, 13950–13955 (2005).

204. Tettelin, H., Riley, D., Cattuto, C. & Medini, D. Comparative genomics: the bacterial pan-genome. *Curr. Opin. Microbiol.* **11**, 472–477 (2008).

205. Mira, A., Martín-Cuadrado, A. B., D'Auria, G. & Rodríguez-Valera, F. The bacterial pan-genome: A new paradigm in microbiology. *Int. Microbiol.* **13**, 45–57 (2010).

206. Freschi, L. *et al.* The *Pseudomonas aeruginosa* pan-genome provides new insights on its population structure, horizontal gene transfer, and pathogenicity. *Genome Biol. Evol.* **11**, 109–120 (2019).

207. Conlan, S. *et al. Staphylococcus epidermidis* pan-genome sequence analysis reveals diversity of skin commensal and hospital infection-associated isolates. *Genome Biol.* **13**, R64 (2012).

208. Méric, G. *et al.* Disease-associated genotypes of the commensal skin bacterium *Staphylococcus epidermidis*. *Nat. Commun.* **9**, 5034 (2018).

209. Caputo, A. *et al.* Pan-genomic analysis to redefine species and subspecies based on quantum discontinuous variation: The Klebsiella paradigm. *Biol. Direct* **10**, 1–12 (2015).

210. Caputo, A., Fournier, P.-E. & Raoult, D. Genome and pan-genome analysis to classify emerging bacteria. *Biol. Direct* **14**, 1–9 (2019).

211. Medini, D., Donati, C., Tettelin, H., Masignani, V. & Rappuoli, R. The microbial pan-genome. *Curr. Opin. Genet. Dev.* **15**, 589–594 (2005).

212. Diene, S. M. *et al.* The rhizome of the multidrug-resistant enterobacter aerogenes genome reveals how new 'Killer Bugs' are created because of a sympatric lifestyle. *Mol. Biol. Evol.* **30**, 369–383 (2013).

213. Georgiades, K. & Raoult, D. Defining pathogenic bacterial species in the genomic era. *Front. Microbiol.* **1**, 1–13 (2011).

214. Rouli, L., Merhej, V., Fournier, P. E. & Raoult, D. The bacterial pangenome as a new tool for analysing pathogenic bacteria. *New Microbes New Infect.* **7**, 72–85 (2015).

215. Hall, B. G. SNP-associations and phenotype predictions from hundreds of microbial genomes without genome alignments. *PLoS One* **9**, (2014).

216. Lees, J. A. *et al.* Sequence element enrichment analysis to determine the genetic basis of bacterial phenotypes. *Nat. Commun.* **7**, (2016).

217. Earle, S. G. *et al.* Identifying lineage effects when controlling for population structure improves power in bacterial association studies. *Nat. Microbiol.* **1**, (2016).

218. Méric, G. *et al.* Ecological overlap and horizontal gene transfer in *Staphylococcus aureus* and *Staphylococcus epidermidis*. *Genome Biol. Evol.* **7**, 1313–1328 (2015).

219. Bosi, E. *et al.* Comparative genome-scale modelling of *Staphylococcus aureus* strains identifies strain-specific metabolic capabilities linked to pathogenicity. *Proc. Natl. Acad. Sci.* **113**, E3801–E3809 (2016).

220. Firth, N., Jensen, S. O., Kwong, S. M., Skurray, R. A. & Ramsay, J. P. Staphylococcal plasmids, transposable and integrative elements. *Microbiol. Spectr.* **6**, 1–26 (2018).

221. Lindsay, J. A. & Holden, M. T. G. *Staphylococcus aureus:* Superbug, super genome? *Trends Microbiol.* **12**, 378–385 (2004).

222. Argemi, X. *et al.* Comparative genomic analysis of *Staphylococcus lugdunensis* shows a closed pan-genome and multiple barriers to horizontal gene transfer. *BMC Genomics* **19**, (2018).

223. Feng, Y. *et al.* Evolution and pathogenesis of *Staphylococcus aureus*: Lessons learned from genotyping and comparative genomics. *FEMS Microbiol. Rev.* **32**, 23–37 (2008).

224. Highlander, S. K. *et al.* Subtle genetic changes enhance virulence of methicillin resistant and sensitive *Staphylococcus aureus*. *BMC Microbiol.* **7**, 1–14 (2007).

225. Morikawa, K. *et al.* Expression of a Cryptic Secondary Sigma Factor Gene Unveils Natural Competence for DNA Transformation in *Staphylococcus aureus*. *PLoS Pathog.* **8**, (2012).

226. Liu, Y., Defourny, K. A. Y., Smid, E. J. & Abee, T. Gram-positive bacterial extracellular vesicles and their impact on health and disease. *Front. Microbiol.* **9**, 1–8 (2018).

227. Klieve, A. V. *et al.* Naturally occurring DNA transfer system associated with

membrane vesicles in cellulolytic *Ruminococcus* spp. of ruminal origin. *Appl. Environ. Microbiol.* **71**, 4248–4253 (2005).

228. Rumbo, C. *et al.* Horizontal transfer of the OXA-24 carbapenemase gene via outer membrane vesicles: A new mechanism of dissemination of carbapenem resistance genes in *Acinetobacter baumannii*. *Antimicrob. Agents Chemother.* **55**, 3084–3090 (2011).

229. Domingues, S. & Nielsen, K. M. Membrane vesicles and horizontal gene transfer in prokaryotes. *Curr. Opin. Microbiol.* **38**, 16–21 (2017).

230. Schwarz, S. *et al.* Plasmid-mediated antimicrobial resistance in Staphylococci and other Firmicutes. *Microbiol. Spectr.* **2**, (2014).

231. Weigel, L. M. *et al.* Genetic analysis of a high-level vancomycin-resistant isolate of *staphylococcus aureus*. *Science.* **302**, 1569–1571 (2003).

232. Couturier, M., Bex, F., Bergquist, P. L. & Maas, W. K. Identification and classification of bacterial plasmids. *Microbiol. Rev.* **52**, 375–395 (1988).

233. Carattoli, A. *et al. In silico* detection and typing of plasmids using plasmidfinder and plasmid multilocus sequence typing. *Antimicrob. Agents Chemother.* **58**, 3895–3903 (2014).

234. Ramsay, J. P. *et al.* An updated view of plasmid conjugation and mobilization in *Staphylococcus* . *Mob. Genet. Elements* **6**, e1208317 (2016).

235. Shearer, J. E. S. *et al.* Major families of multiresistant plasmids from geographically and epidemiologically diverse staphylococci. *G3 Genes, Genomes, Genet.* **1**, 581–591 (2011).

236. Vandecraen, J., Chandler, M., Aertsen, A. & Van Houdt, R. The impact of insertion sequences on bacterial genome plasticity and adaptability. *Crit. Rev. Microbiol.* **43**, 709–730 (2017).

237. Kleinert, F. *et al.* Influence of IS 256 on genome variability and formation of small-colony variants in *Staphylococcus aureus*. *Antimicrob. Agents Chemother.* **61**, e00144-17 (2017).

238. Byrne, M. E., Rouch, D. A. & Skurray, R. A. Nucleotide sequence analysis of IS256 from the *Staphylococcus aureus* gentamicin-tobramycin-kanamycin-

resistance transposon Tn*4001*. *Gene* **81**, 361–367 (1989).

239. Klingenberg, C. *et al.* Phenotypic and genotypic aminoglycoside resistance in blood culture isolates of coagulase-negative staphylococci from a single neonatal intensive care unit, 1989-2000. *J. Antimicrob. Chemother.* **54**, 889–896 (2004).

240. Loessner, I., Dietrich, K., Dittrich, D., Hacker, J. & Ziebuhr, W. Transposase-dependent formation of circular IS256 derivatives in *Staphylococcus epidermidis* and *Staphylococcus aureus*. *J. Bacteriol.* **184**, 4709–4714 (2002).

241. McEvoy, C. R. E. *et al.* Decreased vancomycin susceptibility in *Staphylococcus aureus* caused by IS256 tempering of walKR expression. *Antimicrob. Agents Chemother.* **57**, 3240–3249 (2013).

242. Benson, M. A. *et al.* Evolution of hypervirulence by MRSA clone through acquisition of a transposable element. *Mol. Microbiol.* **93**, 664–682 (2014).

243. Jansen, A. *et al.* Role of insertion elements and yycFG in the development of decreased susceptibility to vancomycin in *Staphylococcus aureus*. *Int. J. Med. Microbiol.* **297**, 205–215 (2007).

244. Maki, H., McCallum, N., Bischoff, M., Wada, A. & Berger-Bachi, B. tcaA inactivation increases glycopeptide resistance in *Staphylococcus aureus*. *Antimicrob. Agents Chemother.* **48**, 1953–1959 (2004).

245. Goering, R. V *et al. Mims' Medical Microbiology*. (Mosby, 2008).

246. Brüssow, H., Canchaya, C. & Hardt, W.-D. Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. *Microbiol. Mol. Biol. Rev.* **68**, 560–602 (2004).

247. Lindsay, J. A. Genomic variation and evolution of *Staphylococcus aureus*. *Int. J. Med. Microbiol.* **300**, 98–103 (2010).

248. Kwan, T., Liu, J., DuBow, M., Gros, P. & Pelletier, J. The complete genomes and proteomes of 27 *Staphylococcus aureus* bacteriophages. *Proc. Natl. Acad. Sci.* **102**, 5174–5179 (2005).

249. Deghorain, M. & Van Melderen, L. The staphylococci phages family: An overview. *Viruses* **4**, 3316–3335 (2012).

250. Goerke, C., Wirtz, C., Flückiger, U. & Wolz, C. Extensive phage dynamics in

Staphylococcus aureus contributes to adaptation to the human host during infection. *Mol. Microbiol.* **61**, 1673–1685 (2006).

251. Betley, M. & Mekalanos, J. Staphylococcal enterotoxin A is encoded by a phage. *Science.* **229**, 185–187 (1985).

252. Kaneko, J., Kimura, T., Narita, S., Tomita, T. & Kamio, Y. Complete nucleotide sequence and molecular characterization of the temperate staphylococcal bacteriophage φPVL carrying Panton-Valentine leukocidin genes. *Gene* **215**, 57–67 (1998).

253. Wamel, W. J. B. Van, Rooijakkers, S. H. M., Kessel, K. P. M. Van, Strijp, J. a G. Van & Ruyken, M. The innate immune modulators staphylococcal complement inhibitor and chemotaxis inhibitory protein of *Staphylococcus aureus* are located on beta-hemolysin-converting bacteriophages. *J. Bacteriol.* **188**, 1310–1315 (2006).

254. Sørensen, M. *et al. Staphylococcus aureus* enterotoxin sensitization is associated with allergic poly-sensitization and allergic multimorbidity in adolescents. *Allergy Eur. J. Allergy Clin. Immunol.* **72**, 1548–1555 (2017).

255. Ito, T., Katayama, Y. & Hiramatsu, K. Cloning and nucleotide sequence determination of the entire mec DNA of pre-methicillin-resistant *Staphylococcus aureus* N315. *Antimicrob. Agents Chemother.* **43**, 1449–1458 (1999).

256. Liu, J. *et al.* Staphylococcal chromosomal cassettes mec (SCC*mec*): A mobile genetic element in methicillin-resistant *Staphylococcus aureus*. *Microb. Pathog.* **101**, 56–67 (2016).

257. Saber, H., Jasni, A. S., Jamaluddin, T. Z. M. T. & Ibrahim, R. A review of staphylococcal cassette chromosome mec (SCCmec) types in coagulase-negative staphylococci (CoNS) species. *Malaysian J. Med. Sci.* **24**, 7–18 (2017).

258. Silva, P. V. *et al.* The antimicrobial susceptibility, biofilm formation and genotypic profiles of *Staphylococcus haemolyticus* from bloodstream infections. *Mem. Inst. Oswaldo Cruz* **108**, 812–816 (2013).

259. Hanssen, A. M. & Ericson Sollid, J. U. Multiple staphylococcal cassette chromosomes and allelic variants of cassette chromosome recombinases in *Staphylococcus aureus* and coagulase-negative staphylococci from Norway. *Antimicrob. Agents Chemother.* **51**, 1671–1677 (2007).

260. Wielders, C. L. C. *et al.* Evidence for in-vivo transfer of *mecA* DNA between strains of *Staphylococcus aureus*. *Lancet* **357**, 1674–1675 (2001).

261. Kaya, H. *et al.* SSCmecFinder, a web-based tool for typing of staphylococcal cassette chromosome mec in *Staphylococcus aureus* using whole-genome sequence data. *mSphere* **3**, e00612-17 (2018).

262. Hacker, J. & Kaper, J. B. Pathogenicity islands and the evolution of microbes. *Annu. Rev. Microbiol.* **54**, 641–679 (2000).

263. Xia, G. & Wolz, C. Phages of *Staphylococcus aureus* and their impact on host evolution. *Infect. Genet. Evol.* **21**, 593–601 (2014).

264. Novick, R. & Subedi, A. The SaPIs: Mobile pathogenicity islands of *Staphylococcus*. in *Chemical Immunology and Allergy* **93**, 42–57 (2007).

265. Madhusoodanan, J. *et al.* An enterotoxin-bearing pathogenicity island in *Staphylococcus epidermidis*. *J. Bacteriol.* **193**, 1854–1862 (2011).

266. Argemi, X. *et al.* Comparative genomics and identification of an enterotoxin-bearing pathogenicity island, SEPI-1/SECI-1, in *Staphylococcus epidermidis* pathogenic strains. *Toxins (Basel).* **10**, 1–12 (2018).

267. King, J. M., Kulhankova, K., Stach, C. S., Vu, B. G. & Salgado-pabón, W. Phenotypes of virulence amonf *Staphylococcus aureus* USA100, USA200, USA300, USA400 and USA600 clonal lineages. *mSphere* **1**, e00071-16 (2016).

268. Recker, M. *et al.* Clonal differences in *Staphylococcus aureus* bacteraemia-associated mortality. *Nat. Microbiol.* **2**, 1381–1388 (2017).

269. Fournier, P. E., Dubourg, G. & Raoult, D. Clinical detection and characterization of bacterial pathogens in the genomics era. *Genome Med.* **6**, 1–15 (2014).

270. Schwartz, D. C. & Cantor, C. R. Seperation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. *Cell* **37**, 67–75 (1984).

271. Goering, R. V. Pulsed field gel electrophoresis: A review of application and interpretation in the molecular epidemiology of infectious disease. *Infect. Genet. Evol.* **10**, 866–875 (2010).

272. Neoh, H. min, Tan, X. E., Sapri, H. F. & Tan, T. L. Pulsed-field gel electrophoresis (PFGE): A review of the “gold standard” for bacteria typing and current

alternatives. *Infect. Genet. Evol.* **74**, 103935 (2019).

273. Tenover, F. C. *et al.* Interpreting chromosomal DNA restriction patterns produced by pulsed- field gel electrophoresis: Criteria for bacterial strain typing. *J. Clin. Microbiol.* **33**, 2233–2239 (1995).

274. Maiden, M. *et al.* Multilocus sequence typing : A portable approach to the identification of clones within populations of. *Proc. Natl. Acad. Sci. USA* **95**, 3140–3145 (1998).

275. Maiden, M. C. J. Multilocus Sequence Typing of Bacteria. *Annu. Rev. Microbiol.* **60**, 561–588 (2006).

276. Feil, E. J., Li, B. C., Aanensen, D. M., Hanage, W. P. & Spratt, B. G. eBURST: Inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. *J. Bacteriol.* **186**, 1518–1530 (2004).

277. Liu, Y. & Ji, Y. Multilocus Sequence Typing of *Staphylococcus aureus*. in *Methicillin-Resistant Staphylococcus aureus (MRSA) Protocols* (ed. Ji, Y.) 95–102 (Humana, 2020).

278. Thomas, J. C. & Robinson, A. Multilocus Sequence Typing of *Staphylococcus epidermidis*. in *Staphylococcus Epidermidis. Methods in Molecular Biology (Methods and Protocols)* (ed. Fey, P.) 61–69 (Humana Press, 2014).

279. Cavanagh, J. P. *et al.* Core genome conservation of *Staphylococcus haemolyticus* limits sequence based population structure analysis. *J. Microbiol. Methods* **89**, 159–166 (2012).

280. Zhang, L. *et al.* Multilocus sequence typing and further genetic characterization of the enigmatic pathogen, *Staphylococcus hominis*. *PLoS One* **8**, (2013).

281. Solyman, S. M. *et al.* Multilocus sequence typing for characterization of *Staphylococcus pseudintermedius*. *J. Clin. Microbiol.* **51**, 306–310 (2013).

282. Jolley, K. A., Bray, J. E. & Maiden, M. C. J. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. *Wellcome Open Res.* **3**, 1–20 (2018).

283. Fakhr, M. K., Nolan, L. K. & Logue, C. M. Multilocus sequence typing lacks the

discriminatory ability of pulsed-field gel electrophoresis for typing *Salmonella enterica* serovar Typhimurium. *J. Clin. Microbiol.* **43**, 2215–2219 (2005).

284. Rumore, J. *et al.* Evaluation of whole-genome sequencing for outbreak detection of Verotoxigenic *Escherichia coli* O157:H7 from the Canadian perspective. *BMC Genomics* **19**, 1–13 (2018).

285. Ruan, Z. & Feng, Y. BacWGSTdb, a database for genotyping and source tracking bacterial pathogens. *Nucleic Acids Res.* **44**, D682–D687 (2016).

286. Alikhan, N. F., Zhou, Z., Sergeant, M. J. & Achtman, M. A genomic overview of the population structure of *Salmonella. PLoS Genet.* **14**, 1–13 (2018).

287. Ruan, Z., Yu, Y. & Feng, Y. The global dissemination of bacterial infections necessitates the study of reverse genomic epidemiology. *Brief. Bioinform.* **00**, 1–10 (2019).

288. Emerson, D., Agulto, L., Liu, H. & Liu, L. Identifying and characterizing bacteria in an era of genomics and proteomics. *Bioscience* **58**, 925–936 (2008).

289. Chun, J. *et al.* Proposed minimal standards for the use of genome data for the taxonomy of prokaryotes. *Int. J. Syst. Evol. Microbiol.* **68**, 461–466 (2018).

290. Ash, C., Farrow, J. A. E., Dorsch, M., Stackebrandt, E. & Collins, M. D. Comparative analysis of *Bacillus anthracis*, B*acillus cereus*, and related species on the basis of reverse transcriptase sequencing of 16S rRNA. *Int. J. Syst. Bacteriol.* **41**, 343–346 (1991).

291. Trebesius, K., Harmsen, D., Rakin, A., Schmelz, J. & Heesemann, J. Development of rRNA-targeted PCR and in situ hybridization with fluorescently labelled oligonucleotides for detection of *Yersinia species*. *J. Clin. Microbiol.* **36**, 2557–2564 (1998).

292. Ghebremedhin, B., Layer, F., König, W. & König, B. Genetic classification and distinguishing of *Staphylococcus* species based on different partial gap, 16S rRNA, hsp60, rpoB, sodA, and tuf gene sequences. *J. Clin. Microbiol.* **46**, 1019–1025 (2008).

293. Mignard, S. & Flandrois, J. P. 16S rRNA sequencing in routine bacterial identification: A 30-month experiment. *J. Microbiol. Methods* **67**, 574–581 (2006).

294. Beye, M., Fahsi, N., Raoult, D. & Fournier, P. E. Careful use of 16S rRNA gene sequence similarity values for the identification of *Mycobacterium* species. *New Microbes New Infect.* **22**, 24–29 (2018).

295. Caprioli, R. M., Farmer, T. B. & Gile, J. Molecular imaging of biological samples: Localization of peptides and proteins using MALDI-TOF MS. *Anal. Chem.* **69**, 4751–4760 (1997).

296. Patel, R. MALDI-TOF MS for the diagnosis of infectious diseases. *Clin. Chem.* **61**, 100–111 (2015).

297. Winkler, M. A., Uher, J. & Cepa, S. Direct analysis and identification of *Helicobacter* and *Campylobacter* species by MALDI-TOF mass spectrometry. *Anal. Chem.* **71**, 3416–3419 (1999).

298. Dieckmann, R. & Malorny, B. Rapid screening of epidemiologically important *Salmonella enterica* subsp. enterica serovars by whole-cell matrix-assisted laser desorption ionization-time of flight mass spectrometry. *Appl. Environ. Microbiol.* **77**, 4136–4146 (2011).

299. Berrazeg, M. *et al.* Biotyping of multidrug-resistant *Klebsiella pneumoniae* clinical isolates from France and Algeria using MALDI-TOF MS. *PLoS One* **8**, (2013).

300. Goris, J. *et al.* DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int. J. Syst. Evol. Microbiol.* **57**, 81–91 (2007).

301. Wayne, L. G. *et al.* Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. *Int. J. Syst. Evol. Microbiol.* **37**, 463–464 (1987).

302. Chun, J. & Rainey, F. A. Integrating genomics into the taxonomy and systematics of the Bacteria and Archaea. *Int. J. Syst. Evol. Microbiol.* **64**, 316–324 (2014).

303. Lee, I., Kim, Y. O., Park, S. C. & Chun, J. OrthoANI: An improved algorithm and software for calculating average nucleotide identity. *Int. J. Syst. Evol. Microbiol.* **66**, 1100–1103 (2016).

304. Richter, M. & Rosselló-Móra, R. Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 19126–19131 (2009).

305. Yoon, S. H., Ha, S. min, Lim, J., Kwon, S. & Chun, J. A large-scale evaluation of

algorithms to calculate average nucleotide identity. *Antonie van Leeuwenhoek.* **110**, 1281–1286 (2017).

306. Richter, M., Rosselló-Móra, R., Oliver Glöckner, F. & Peplies, J. JSpeciesWS: A web server for prokaryotic species circumscription based on pairwise genome comparison. *Bioinformatics* **32**, 929–931 (2015).

307. Meier-Kolthoff, J. P., Auch, A. F., Klenk, H. P. & Göker, M. Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics* **14**, (2013).

308. Stackebrandt, E. & Goebel, B. M. Taxonomic note: A place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. *Int. J. Syst. Bacteriol.* **44**, 846–849 (1994).

309. Glaeser, S. P. & Kämpfer, P. Multilocus sequence analysis (MLSA) in prokaryotic taxonomy. *Syst. Appl. Microbiol.* **38**, 237–245 (2015).

310. Hahnke, R. L. *et al.* Genome-based taxonomic classification of Bacteroidetes. *Front. Microbiol.* **7**, (2016).

311. Na, S. I. *et al.* UBCG: Up-to-date bacterial core gene set and pipeline for phylogenomic tree reconstruction. *J. Microbiol.* **56**, 281–285 (2018).

312. Wu, M. & Scott, A. J. Phylogenomic analysis of bacterial and archaeal sequences with AMPHORA2. *Bioinformatics* **28**, 1033–1034 (2012).

313. Segata, N., Börnigen, D., Morgan, X. C. & Huttenhower, C. PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. *Nat. Commun.* **4**, (2013).

314. Ankenbrand, M. J. & Keller, A. BcgTree: Automatized phylogenetic tree building from bacterial core genomes. *Genome* **59**, 783–791 (2016).

315. Goodwin, S., McPherson, J. D. & McCombie, W. R. Coming of age: Ten years of next-generation sequencing technologies. *Nat. Rev. Genet.* **17**, 333–351 (2016).

316. Didelot, X., Bowden, R., Wilson, D. J., Peto, T. E. A. & Crook, D. W. Transforming clinical microbiology with bacterial genome sequencing. *Nat. Rev. Genet.* **13**, 601–612 (2012).

317. Bertelli, C. & Greub, G. Rapid bacterial genome sequencing: Methods and

applications in clinical microbiology. *Clin. Microbiol. Infect.* **19**, 803–813 (2013).

318. Ellington, M. J. *et al.* The role of whole genome sequencing in antimicrobial susceptibility testing of bacteria: report from the EUCAST Subcommittee. *Clin. Microbiol. Infect.* **23**, 2–22 (2017).

319. Gilissen, C. *et al.* Genome sequencing identifies major causes of severe intellectual disability. *Nature* **511**, 344–347 (2014).

320. Clark, M. M. *et al.* Meta-analysis of the diagnostic and clinical utility of genome and exome sequencing and chromosomal microarray in children with suspected genetic diseases. *NPJ Genomic Med.* **3**, 1–10 (2018).

321. Tarailo-Graovac, M. *et al.* Exome sequencing and the management of neurometabolic disorders. *N. Engl. J. Med.* **374**, 2246–2255 (2016).

322. Perkins, B. A. *et al.* Precision medicine screening using whole-genome sequencing and advanced imaging to identify disease risk in adults. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 3686–3691 (2018).

323. Esplin, E. D., Oei, L. & Snyder, M. P. Personalized sequencing and the future of medicine: Discovery, diagnosis and defeat of disease. *Pharmacogenomics* **15**, 1771–1790 (2014).

324. Bentley, D. R. *et al.* Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456**, 53–59 (2008).

325. Harismendy, O. *et al.* Evaluation of next generation sequencing platforms for population targeted sequencing studies. *Genome Biol.* **10**, (2009).

326. Dohm, J. C., Lottaz, C., Borodina, T. & Himmelbauer, H. Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucleic Acids Res.* **36**, (2008).

327. Nakamura, K. *et al.* Sequence-specific error profile of Illumina sequencers. *Nucleic Acids Res.* **39**, (2011).

328. Quainoo, S. *et al.* Whole-genome sequencing of bacterial pathogens: the future of nosocomial outbreak analysis. *Clin. Microbiol. Rev.* **30**, 1015–1064 (2017).

329. Rothberg, J. M. *et al.* An integrated semiconductor device enabling non-optical genome sequencing. *Nature* **475**, 348–352 (2011).

330. Malapelle, U. *et al.* Ion Torrent next-generation sequencing for routine identification of clinically relevant mutations in colorectal cancer patients. *J. Clin. Pathol.* **68**, 64–68 (2015).

331. Loman, N. J. *et al.* Performance comparison of benchtop high-throughput sequencing platforms. *Nat. Biotechnol.* **30**, 434–439 (2012).

332. Heilbronner, S. *et al.* Competing for iron: duplication and amplification of the *isd* locus in *Staphylococcus lugdunensis* HKU09-01 provides a competitive advantage to overcome nutritional limitation. *PLoS Genet.* **12**, 1–25 (2016).

333. Romero, D., Martínez-Salazar, J., Ortiz, E., Rodríguez, C. & Valencia-Morales, E. Repeated sequences in bacterial chromosomes and plasmids: A glimpse from sequenced genomes. *Res. Microbiol.* **150**, 735–743 (1999).

334. Eid, J. *et al.* Single Polymerase Molecules. *Science.* **323**, 133–138 (2009).

335. Weirather, J. L. *et al.* Comprehensive comparison of Pacific Biosciences and Oxford Nanopore Technologies and their applications to transcriptome analysis. *F1000Research* **6**, 100 (2017).

336. Mendez-Garcia, C., Bargiela, R., Martinez-Martinez, M. & Ferrer, M. Chapter 2 - Metagenomic Protocols and Strategies. in *Metagenomics - Perspective, Methods, and Applications* 15–54 (Academic Press, 2018).

337. Schatz, M. C., Delcher, A. L. & Salzberg, S. L. Assembly of large genomes using second-generation sequencing. *Genome Res.* **20**, 1165–1173 (2010).

338. Carneiro, M. O. *et al.* Pacific biosciences sequencing technology for genotyping and variation discovery in human data. *BMC Genomics* **13**, (2012).

339. Koren, S. *et al.* Hybrid error correction and de novo assembly of single-molecule sequencing reads. *Nat. Biotechnol.* **30**, 693–700 (2012).

340. Clarke, J. *et al.* Continuous base identification for single-molecule nanopore DNA sequencing. *Nat. Nanotechnol.* **4**, 265–270 (2009).

341. Rang, F. J., Kloosterman, W. P. & de Ridder, J. From squiggle to basepair: Computational approaches for improving nanopore sequencing read accuracy. *Genome Biol.* **19**, 1–11 (2018).

342. Goodwin, S. *et al.* Oxford Nanopore sequencing, hybrid error correction, and de

novo assembly of a eukaryotic genome. *Genome Res.* **25**, 1750–1756 (2015).

343. Jain, M. *et al.* Improved data analysis for the MinION nanopore sequencer. *Nat. Methods* **12**, 351–356 (2015).

344. Joensen, K. G. *et al.* Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic *Escherichia coli*. *J. Clin. Microbiol.* **52**, 1501–1510 (2014).

345. Flusberg, B. A. *et al.* Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat. Methods* **7**, 461–465 (2010).

346. Dominguez Del Angel, V. *et al.* Ten steps to get started in Genome Assembly and Annotation. *F1000Research* **7**, (2018).

347. Didion, J. P., Martin, M. & Collins, F. S. Atropos: Specific, sensitive, and speedy trimming of sequencing reads. *PeerJ* **2017**, 1–19 (2017).

348. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

349. Schmieder, R. & Edwards, R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* **27**, 863–864 (2011).

350. Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. QUAST: Quality assessment tool for genome assemblies. *Bioinformatics* **29**, 1072–1075 (2013).

351. Lischer, H. E. L. & Shimizu, K. K. Reference-guided de novo assembly approach improves genome reconstruction for related species. *BMC Bioinformatics* **18**, 1–12 (2017).

352. Aziz, R. K. *et al.* The RAST Server: Rapid annotations using subsystems technology. *BMC Genomics* **9**, 1–15 (2008).

353. Seemann, T. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).

354. Chen, L., Zheng, D., Liu, B., Yang, J. & Jin, Q. VFDB 2016: Hierarchical and refined dataset for big data analysis - 10 years on. *Nucleic Acids Res.* **44**, D694–D697 (2016).

355. Liu, B., Zheng, D., Jin, Q., Chen, L. & Yang, J. VFDB 2019: A comparative

pathogenomic platform with an interactive web interface. *Nucleic Acids Res.* **47**, D687–D692 (2019).

356. Seemann, T. ABRicate, Github. Available at: https://github.com/tseemann/abricate.

357. Zankari, E. *et al.* Identification of acquired antimicrobial resistance genes. *J. Antimicrob. Chemother.* **67**, 2640–2644 (2012).

358. Jia, B. *et al.* CARD 2017: Expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Res.* **45**, D566–D573 (2017).

359. Arndt, D. *et al.* PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res.* **44**, W16–W21 (2016).

360. Zhou, Y., Liang, Y., Lynch, K. H., Dennis, J. J. & Wishart, D. S. PHAST: A Fast Phage Search Tool. *Nucleic Acids Res.* **39**, 347–352 (2011).

361. Varani, A. M., Siguier, P., Gourbeyre, E., Charneau, V. & Chandler, M. ISsaga is an ensemble of web-based methods for high throughput identification and semi-automatic annotation of insertion sequences in prokaryotic genomes. *Genome Biol.* **12**, R30 (2011).

362. Siguier, P. ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Res.* **34**, D32–D36 (2006).

363. Domon, B. & Aebersold, R. Mass spectrometry and protein analysis. *Science (80-.).* **312**, 212–217 (2006).

364. Aslam, B., Basit, M., Nisar, M. A., Khurshid, M. & Rasool, M. H. Proteomics: Technologies and their applications. *J. Chromatogr. Sci.* **55**, 182–196 (2017).

365. Siciliano, R. A., Lippolis, R. & Mazzeo, M. F. Proteomics for the investigation of surface-exposed proteins in probiotics. *Front. Nutr.* **6**, 1–8 (2019).

366. Desvaux, M., Dumas, E., Chafsey, I. & Hébraud, M. Protein cell surface display in Gram-positive bacteria: From single protein to macromolecular protein structure. *FEMS Microbiol. Lett.* **256**, 1–15 (2006).

367. Desvaux, M., Candela, T. & Serror, P. Surfaceome and proteosurfaceome in parietal monoderm bacteria: Focus on protein cell-surface display. *Front.*

Microbiol. **9**, 1–28 (2018).

368. Solis, N. & Cordwell, S. J. Current methodologies for proteomics of bacterial surface-exposed and cell envelope proteins. *Proteomics* **11**, 3169–3189 (2011).

369. The European Committee on Antimicrobial Suceptibility Testing. Breakpoint tables for interpretation of MICs and zone diameters, version 9.0, 2019.

370. Christensen, G. D. *et al.* Adherence of coagulase-negative staphylococci to plastic tissue culture plates: A quantitative model for the adherence of staphylococci to medical devices. *J. Clin. Microbiol.* **22**, 996–1006 (1985).

371. Edwards, A. M. & Massey, R. C. Invasion of human cells by a bacterial pathogen. *J. Vis. Exp.* **49**, 1–4 (2011).

372. Freney, J. *et al.* Recommended minimal standards for description of new staphylococcal species. *Int. J. Syst. Bacteriol.* **49**, 489–502 (2009).

373. Kloos, W. E., Tornabene, T. G. & Schleifer, K. H. Isolation and characterization of micrococci from human skin, including two new species: *Micrococcus lylae* and *Micrococcus kristinae*. *Int. J. Syst. Bacteriol.* **24**, 79–101 (1974).

374. Nanoxis Consulting AB - LPI Technology. Available at: http://nanoxisconsulting.com/services-2/lpi-technology-2.html.

375. Zerbino, D. R. & Birney, E. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* **18**, 821–829 (2008).

376. Boisvert, S., Laviolette, F. & Corbeil, J. Ray: Simultaneous assembly of reads from a mix of high-throughput sequencing technologies. *J. Comput. Biol.* **17**, 1401–1415 (2010).

377. Bankevich, A. *et al.* SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).

378. Peng, Y., Leung, H. C. M., Yiu, S. M. & Chin, F. Y. L. IDBA-UD: A de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* **28**, 1420–1428 (2012).

379. Seemann, T. Shovill, Github. Available at: https://github.com/tseemann/shovill.

380. Contreras-Moreira, B. & Vinuesa, P. GET_HOMOLOGUES, a versatile software

package for scalable and robust microbial pangenome analysis. *Appl. Environ. Microbiol.* **79**, 7696–7701 (2013).

381. Chaudhari, N. M., Gupta, V. K. & Dutta, C. BPGA-an ultra-fast pan-genome analysis pipeline. *Sci. Rep.* **6**, 1–10 (2016).

382. Page, A. J. *et al.* Roary: Rapid large-scale prokaryote pan genome analysis. *Bioinformatics* **31**, 3691–3693 (2015).

383. Brynildsrud, O., Bohlin, J., Scheffer, L. & Eldholm, V. Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. *Genome Biol.* **17**, 238 (2016).

384. Murray, A. K. *et al. Staphylococcus cornubiensis* sp. nov., a member of the staphylococcus intermedius group (SIG). *Int. J. Syst. Evol. Microbiol.* **68**, 3404–3408 (2018).

385. Gill, V. J., Selepak, S. T. & Williams, E. C. Species identification and antibiotic susceptibilities of coagulase-negative staphylococci isolated from clinical specimens. *J. Clin. Microbiol.* **18**, 1314–1319 (1983).

386. Low, D. E. *et al.* An endemic strain of *Staphylococcus haemolyticus* colonizing and causing bacteremia in neonatal intensive care unit patients. *Pediatrics* **89**, 696–700 (1992).

387. De Silva, G. D. I. *et al.* The ica operon and biofilm production in coagulase-negative staphylococci associated with carriage and disease in a neonatal intensive care unit. *J. Clin. Microbiol.* **40**, 382–388 (2002).

388. Barberis, I. L., Pájaro, M. C., Godino, S. D., Pascual, L. & Daniele, M. D. Antimicrobial sensitivity and adherence study in strains of coagulase-negative *Staphylococcus* spp. *Rev. Latinoam. Microbiol.* **43**, 109–113 (2001).

389. Al-Naqshbandi, A. A., Chawsheen, M. A. & Abdulqader, H. H. Prevalence and antimicrobial susceptibility of bacterial pathogens isolated from urine specimens received in rizgary hospital — Erbil. *J. Infect. Public Health* **12**, 330–336 (2019).

390. Ahmed, S. *et al.* Drug resistance and molecular epidemiology of aerobic bacteria isolated from puerperal infections in Bangladesh. *Microb. Drug Resist.* **21**, 297–306 (2015).

391. Shittu, A. *et al.* Characterization of methicillin-susceptible and -resistant staphylococci in the clinical setting: A multicentre study in Nigeria. *BMC Infect. Dis.* **12**, (2012).

392. Azih, A. & Enabulele, I. Species distribution and virulence factors of coagulase negative staphylococci isolated from clinical samples from the university of Benin teaching hospital, Edo state; nigeria. *J. Nat. Sci. Res.* **3**, 38–44 (2013).

393. Ehlersson, G., Hellmark, B., Svartström, O., Stenmark, B. & Söderquist, B. Phenotypic characterisation of coagulase-negative staphylococci isolated from blood cultures in newborn infants, with a special focus on *Staphylococcus capitis*. *Acta Paediatr. Int. J. Paediatr.* **106**, 1576–1582 (2017).

394. Hitzenbichler, F., Simon, M., Salzberger, B. & Hanses, F. Clinical significance of coagulase-negative staphylococci other than *S. epidermidis* blood stream isolates at a tertiary care hospital. *Infection* **45**, 179–186 (2017).

395. Knight, G. M. *et al.* Shift in dominant hospital-associated methicillin-resistant *Staphylococcus aureus* (HA-MRSA) clones over time. *J. Antimicrob. Chemother.* **67**, 2514–2522 (2012).

396. Voronina, O. L., Kunda, M. S., Dmitrenko, O. A., Lunin, V. G. & Gintsburg, A. L. Development of *Staphylococcus haemolyticus* multilocus sequencing scheme and its use for molecular-epidemiologic analysis of strains isolated in hospitals in Russian federation in 2009-2010. *Zhurnal Mikrobiol. Epidemiol. Immunobiol.* **5**, 62–67 (2011).

397. Kornienko, M. *et al.* Analysis of nosocomial *Staphylococcus haemolyticus* by MLST and MALDI-TOF mass spectrometry. *Infect. Genet. Evol.* **39**, (2016).

398. Byrd, A. L., Belkaid, Y. & Segre, J. A. The human skin microbiome. *Nat. Rev. Microbiol.* **16**, 143–155 (2018).

399. Higaki, S., Kltagawa, T., Morohashi, M. & Yamagishi, T. Distribution and antimicrobial susceptibility of coagulase-negative staphylococci from skin lesions. *J. Int. Med. Res.* **27**, 191–195 (1999).

400. Kloos, W. E. & Musselwhite, M. S. Distribution and persistence of *Staphylococcus* and *Micrococcus* species and other aerobic bacteria on human skin. *Appl. Microbiol.* **30**, 381–395 (1975).

401. Byrd, A. L. *et al. Staphylococcus aureus* and *Staphylococcus epidermidis* strain diversity underlying pediatric atopic dermatitis. *Sci. Transl. Med.* **9**, (2017).

402. Espadinha, D. *et al.* Distinct phenotypic and genomic Signatures underlie contrasting pathogenic potential of *Staphylococcus epidermidis* clonal lineages. *Front. Microbiol.* **10**, 1–17 (2019).

403. Mannion, P. T., Thom, B. T., Reynolds, C. S. & Strachan, C. J. L. The acquisition of antibiotic resistant coagulase-negative staphylococci by aortic graft recipients. *J. Hosp. Infect.* **14**, 313–323 (1989).

404. Carr, D. L. & Kloos, W. E. Temporal study of the staphylococci and micrococci of normal infant skin. *Appl. Environ. Microbiol.* **34**, 673–680 (1977).

405. Mehta, G., Singh, S. & Kumari, S. Observations on coagulase-negative staphylococci in a neonatal unit in India. *J. Hosp. Infect.* **19**, 273–281 (1991).

406. Rohde, H. *et al.* Detection of virulence-associated genes not useful for discriminating between invasive and commensal *Staphylococcus epidermidis* strains from a bone marrow transplant unit. *J. Clin. Microbiol.* **42**, 5614–5619 (2004).

407. Kozitskaya, S. *et al.* Clonal analysis of *Staphylococcus epidermidis* isolates carrying or lacking biofilm-mediating genes by multilocus sequence typing. *J. Clin. Microbiol.* **43**, 4751–4757 (2005).

408. Schleifer, K. H. & Kloos, W. E. Isolation and characterization of staphylococci from human Skin I. Amended descriptions of *Staphylococcus epidermidis* and *Staphylococcus saprophyticus* and descriptions of three new species: *Staphylococcus cohnii*, *Staphylococcus haemolyticus*, and *Staphylococcus xylosus*. *Int. J. Syst. Bacteriol.* **25**, 50–61 (1975).

409. Lenart-Boroń, A., Wolny-Koładka, K., Juraszek, K. & Kasprowicz, A. Phenotypic and molecular assessment of antimicrobial resistance profile of airborne *Staphylococcus* spp. isolated from flats in Kraków. *Aerobiologia (Bologna).* **33**, 435–444 (2017).

410. Widerström, M., Wiström, J., Ek, E., Edebro, H. & Monsen, T. Near absence of methicillin-resistance and pronounced genetic diversity among *Staphylococcus epidermidis* isolated from healthy persons in northern Sweden. *Apmis* **119**, 505–

512 (2011).

411. Antibiotikaalmenprasis.no. Antibiotikabruk i primærhelsetjenesten - Klamydiainfeksjon. (2016). doi:http://www.antibiotikaiallmennpraksis.no/index.php?action=showtopic&topic=nnsCcA7Q

412. Lindbæk, M. *et al.* Nye retningslinjer for antibiotikabruk i primærhelsetjenesten. *Tidsskr. den Nor. Laegeforening* **133**, 1052–1053 (2013).

413. Folkehelseinstituttet. Klamydia og lymfogranuloma venerum (LGV) i Norge 2013. (2014). Available at: https://www.fhi.no/globalassets/dokumenterfiler/trykksaker/gonore-syfilis-hiv-klamydia/klamydia-og-lgv-arsrapport-2013-pdf.pdf. (Accessed: 10th January 2020)

414. Folkehelseinstituttet. Klamydiastatistikk 2018. (2019). Available at: https://www.fhi.no/nyheter/2019/klamydiastatistikk-2018/. (Accessed: 10th January 2020)

415. Ersoy, S. C. *et al.* Correcting a fundamental flaw in the paradigm for antimicrobial susceptibility testing. *EBioMedicine* **20**, 173–181 (2017).

416. Choi, H., Chakraborty, S., Liu, R., Gellman, S. H. & Weisshaar, J. C. Medium effects on minimum inhibitory concentrations of nylon-3 polymers against E. coli. *PLoS One* **9**, (2014).

417. Kavanagh, A., Ramu, S., Gong, Y., Cooper, M. A. & Blaskovich, M. A. T. Effects of microplate type and broth additives on microdilution MIC susceptibility assays. *Antimicrob. Agents Chemother.* **63**, (2019).

418. Lenhard, J. R. *et al.* Evolution of *Staphylococcus aureus* under vancomycin selective pressure: The role of the small-colony variant phenotype. *Antimicrob. Agents Chemother.* **59**, 1347–1351 (2015).

419. Proctor, R. A. *et al.* Small colony variants: A pathogenic form of bacteria that facilitates persistent and recurrent infections. *Nat. Rev. Microbiol.* **4**, 295–305 (2006).

420. Tuchscherr, L. *et al.* Sigma factor SigB is crucial to mediate *Staphylococcus*

aureus adaptation during chronic infections. *PLoS Pathog.* **11**, (2015).

421. Kahl, B. C. & Becker, K. Clinical significance and pathogenesis of staphylococcal small colony. *Am. Soc. Microbiol.* **29**, 401–427 (2016).

422. Mirani, Z. A., Aziz, M. & Khan, S. I. Small colony variants have a major role in stability and persistence of *Staphylococcus aureus* biofilms. *J. Antibiot. (Tokyo).* **68**, 98–105 (2015).

423. Ciusa, M. L. *et al.* A novel resistance mechanism to triclosan that suggests horizontal gene transfer and demonstrates a potential selective pressure for reduced biocide susceptibility in clinical strains of *Staphylococcus aureus*. *Int. J. Antimicrob. Agents* **40**, 210–220 (2012).

424. Berglund, C. & Söderquist, B. The origin of a methicillin-resistant *Staphylococcus aureus* isolate at a neonatal ward in Sweden - Possible horizontal transfer of a staphylococcal cassette chromosome mec between methicillin-resistant *Staphylococcus haemolyticus* and *Staphylococcus aureus*. *Clin. Microbiol. Infect.* **14**, 1048–1056 (2008).

425. Rossi, C. C. *et al.* Transfer of mupirocin resistance from *Staphylococcus haemolyticus* clinical strains to *Staphylococcus aureus* through conjugative and mobilizable plasmids. *FEMS Microbiol. Lett.* **363**, 1–5 (2016).

426. Vasu, K. & Nagaraja, V. Diverse functions of restriction-modification systems in addition to cellular defense. *Microbiol. Mol. Biol. Rev.* **77**, 53–72 (2013).

427. Watanabe, S., Ito, T., Morimoto, Y., Takeuchi, F. & Hiramatsu, K. Precise excision and self-integration of a composite transposon as a model for spontaneous large-scale chromosome inversion/deletion of the *Staphylococcus haemolyticus* clinical strain JCSC1435. *J. Bacteriol.* **189**, 2921–2925 (2007).

428. Barros, E. M., Lemos, M., Souto-Padrón, T. & Giambiagi-deMarval, M. Phenotypic and genotypic characterization of biofilm formation in *Staphylococcus haemolyticus*. *Curr. Microbiol.* **70**, 829–834 (2015).

429. Pinheiro, L., Brito, C. I., Oliveira, A. de, Pereira, V. C. & Cunha, M. de L. R. de S. da. *Staphylococcus epidermidis* and *Staphylococcus haemolyticus:* detection of biofilm genes and biofilm formation in blood culture isolates from patients in a Brazilian teaching hospital. *Diagn. Microbiol. Infect. Dis.* **86**, 11–14 (2016).

430. Waldrop, R., McLaren, A., Calara, F. & McLemore, R. Biofilm growth has a threshold response to glucose in vitro. *Clin. Orthop. Relat. Res.* **472**, 3305–3310 (2014).

431. Agarwal, A. & Jain, A. Glucose & sodium chloride induced biofilm production & ica operon in clinical isolates of staphylococci. *Indian J. Med. Reserach* **138**, 262–266 (2013).

432. Potter, A. *et al.* The gene bap, involved in biofilm production, is present in *Staphylococcus* spp. strains from nosocomial infections. *J. Microbiol.* **47**, 319–326 (2009).

433. Pfaller, M. *et al.* Development of the quantitative micro-test for slime production by coagulase-negative staphylococci. *Eur. J. Clin. Microbiol. Infect. Dis.* **7**, 30–33 (1988).

434. Stepanovic, S. *et al.* Quantification of biofilm in microiter plates: overview of testing conditions and practical recommendations for assessment if biofilm production by staphylococci. *Apmis* 891–899 (2007).

435. de Araujo, G. L. *et al.* Commensal isolates of methicillin-resistant *Staphylococcus epidermidis* are also well equipped to produce biofilm on polystyrene surfaces. *J. Antimicrob. Chemother.* **57**, 855–864 (2006).

436. Karatan, E. & Watnick, P. Signals, Regulatory networks, and materials that Build and break bacterial biofilms. *Microbiol. Mol. Biol. Rev.* **73**, 310–347 (2009).

437. Kostakioti, M., Hadjifrangiskou, M. & Hultgren, S. J. Bacterial biofilms: Development, dispersal, and therapeutic strategies in the dawn of the postantibiotic era. *Cold Spring Harb. Perspect. Med.* **3**, 1–23 (2013).

438. Linnes, J. C., Ma, H. & Bryers, J. D. Giant extracellular matrix binding protein expression in *Staphylococcus epidermidis* is regulated by biofilm formation and osmotic pressure. *Curr. Microbiol.* **66**, 627–633 (2013).

439. Sequeira, S. *et al.* Structural basis for the role of serine-rich repeat proteins from *Lactobacillus reuteri* in gut microbe–host interactions. *Proc. Natl. Acad. Sci.* **115**, E2706-E2715 (2018).

440. Yang, Y. H. *et al.* Structural insights into SraP-mediated *Staphylococcus aureus*

adhesion to host cells. *PLoS Pathog.* **10**, (2014).

441. Zhu, Z. & Desaire, H. Carbohydrates on proteins: Site-specific glycosylation analysis by mass spectrometry. *Annu. Rev. Anal. Chem.* **8**, 463–483 (2015).

442. Ternes, Y. M. *et al.* Molecular epidemiology of coagulase-negative Staphylococcus carriage in neonates admitted to an intensive care unit in Brazil. *BMC Infect. Dis.* **13**, (2013).

443. Hira, V. *et al.* Coagulase-negative staphylococcal skin carriage among neonatal intensive care unit personnel: From population to infection. *J. Clin. Microbiol.* **48**, 3876–3881 (2010).

444. Thomer, L. *et al.* N-Acetylglucosaminylation of serine-aspartate repeat proteins promotes *Staphylococcus aureus* bloodstream infection. *J. Biol. Chem.* **289**, 3478–3486 (2014).

445. Sabat, A. *et al.* Distribution of the serine-aspartate repeat protein-encoding sdr genes among nasal-carriage and invasive *Staphylococcus aureus* strain. *J. Clin. Microbiol.* **44**, 1135–1138 (2006).

446. Salimena, A. P. S. *et al.* Genotypic and phenotypic detection of capsular polysaccharide and biofilm formation in *Staphylococcus aureus* isolated from bovine milk collected from Brazilian dairy farms. *Vet. Res. Commun.* **40**, 97–106 (2016).

447. Matysik, A. & Kline, K. A. *Streptococcus pyogenes* capsule promotes microcolony-independent biofilm formation. *J. Bacteriol.* **201**, 1–14 (2019).

448. Slany, M., Oppelt, J. & Cincarova, L. Formation of *Staphylococcus aureus* biofilm in the presence of sublethal concentrations of disinfectants studied via a transcriptomic analysis using transcriptome sequencing (RNA-seq). *Appl. Environ. Microbiol.* **83**, 1–13 (2017).

449. Nanra, J. S. *et al.* Capsular polysaccharides are an important immune evasion mechanism for *Staphylococcus aureus*. *Hum. Vaccines Immunother.* **9**, 480–487 (2013).

450. Foster, T. J. Immune evasion by staphylococci. *Nat. Rev. Microbiol.* **3**, 948–958 (2005).

451. Rana, R. R., Zhang, M., Spear, A. M., Atkins, H. S. & Byrne, B. Bacterial TIR-containing proteins and host innate immune system evasion. *Med. Microbiol. Immunol.* **202**, 1–10 (2013).

452. Askarian, F. A Staphylococcus aureus TIR-domain protein virulence factor vlocks TLR2-mediated NF-kB signaling. *J. Innate Immunol.* **6**, 485–498 (2014).

453. Wagner, T. M. *et al. Enterococcus faecium* TIR-domain genes are part of a gene cluster which promotes bacterial survival in blood. *Int. J. Microbiol.* **2018**, (2018).

454. Patot, S. *et al.* The TIR homologue lies near resistance genes in *Staphylococcus aureus*, coupling modulation of virulence and antimicrobial susceptibility. *PLoS Pathog.* **13**, 1–23 (2017).

455. Thammavongsa, V., Missiakas, D. & Schneewind, O. *Staphylococcus aureus* degrades neutrophil extracellualr traps to promote immune cell death. *Science.* **342**, 863–866 (2013).

456. Roche, F. M. *et al.* Characterization of novel LPXTG-containing proteins of *Staphylococcus aureus* identified from genome sequences. *Microbiology* **149**, 643–654 (2003).

457. Otto, M. *Staphylococcus aureus* toxins. *Curr. Opin. Microbiol.* **0**, 32–37 (2014).

458. Cunha & Calsolari, R. A. O. Toxigenicity in *Staphylococcus aureus* and coagulase-negative staphylococci: Epidemiological and molecular aspects . *Microbiol. Insights* **1**, 13-24 (2008).

459. Song, Y. *et al.* Inhibition of staphyloxanthin virulence factor biosynthesis in *Staphylococcus aureus*: In vitro, in vivo and crystallographic results. *J. Med. Chem.* **52**, 3869–3880 (2009).

460. Liu, G. Y. *et al. Staphylococcus aureus* golden pigment impairs neutrophil killing and promotes virulence through its antioxidant activity. *J. Exp. Med.* **202**, 209–215 (2005).

461. Clauditz, A., Resch, A., Wieland, K. P., Peschel, A. & Götz, F. Staphyloxanthin plays a role in the fitness of *Staphylococcus aureus* and its ability to cope with oxidative stress. *Infect. Immun.* **74**, 4950–4953 (2006).

462. Cunha, M. D. L. R. S., Sinzato, Y. K. & Silveira, L. V. A. Comparison of methods

for the identification of coagulase-negative staphylococci. *Mem. Inst. Oswaldo Cruz* **99**, 855–860 (2004).

463. Zhou, C. *et al.* Urease is an essential component of the acid response network of *Staphylococcus aureus* and is required for a persistent murine kidney infection. *PLoS Pathog.* **15**, 1–23 (2019).

464. Loes, A. N. *et al.* Inhibition of urease activity in the urinary tract pathogen *Staphylococcus saprophyticus*. *Lett. Appl. Microbiol.* **58**, 31–41 (2014).

465. Mora, D. & Arioli, S. Microbial urease in health and disease. *PLoS Pathog.* **10**, 10–13 (2014).

466. Burne, R. A. & Chen, Y. Y. M. Bacterial ureases in infectious diseases. *Microbes Infect.* **2**, 533–542 (2000).

467. Murchan, S., Aucken, H. M., Ganner, M. & Cookson, B. D. Emergence, spread, and characterization of phage variants of epidemic methicillin-resistant *Staphylococcus aureus* 16 in England and Wales. *J. Clin. Microbiol.* **42**, 5154–5160 (2004).

frontiers in Microbiology

ORIGINAL RESEARCH
published: 10 September 2019
doi: 10.3389/fmicb.2019.02096

Comparative Genomic Analysis of *Staphylococcus haemolyticus* Reveals Key to Hospital Adaptation and Pathogenicity

Maria Pain[1], *Erik Hjerde*[2], *Claus Klingenberg*[1,3] *and Jorunn Pauline Cavanagh*[1,3]*

[1] Pediatric Infections Research Group, Department of Clinical Medicine, UiT The Arctic University of Norway, Tromsø, Norway, [2] Department of Chemistry, Norstruct, UiT The Arcic University of Norway, Tromsø, Norway, [3] Department of Paediatrics, University Hospital of North Norway, Tromsø, Norway

Staphylococcus haemolyticus is a skin commensal gaining increased attention as an emerging pathogen of nosocomial infections. However, knowledge about the transition from a commensal to an invasive lifestyle remains sparse and there is a paucity of studies comparing pathogenicity traits between commensal and clinical isolates. In this study, we used a pan-genomic approach to identify factors important for infection and hospital adaptation by exploring the genomic variability of 123 clinical isolates and 46 commensal *S. haemolyticus* isolates. Phylogenetic reconstruction grouped the 169 isolates into six clades with a distinct distribution of clinical and commensal isolates in the different clades. Phenotypically, multi-drug antibiotic resistance was detected in 108/123 (88%) of the clinical isolates and 5/46 (11%) of the commensal isolates ($p < 0.05$). In the clinical isolates, we commonly identified a homolog of the serine-rich repeat glycoproteins *sraP*. Additionally, three novel capsular polysaccharide operons were detected, with a potential role in *S. haemolyticus* virulence. Clinical *S. haemolyticus* isolates showed specific signatures associated with successful hospital adaption. Biofilm forming *S. haemolyticus* isolates that are resistant to oxacillin (*mecA*) and aminoglycosides (*aacA-aphD)* are most likely invasive isolates whereas absence of these traits strongly indicates a commensal isolate. We conclude that our data show a clear segregation of isolates of commensal origin, and specific genetic signatures distinguishing the clinical isolates from the commensal isolates. The widespread use of antimicrobial agents has probably promoted the development of successful hospital adapted clones of *S. haemolyticus* clones through acquisition of mobile genetic elements or beneficial point mutations and rearrangements in surface associated genes.

Keywords: *Staphylococcus haemolyticus*, pangenome, multidrug resistance, bacterial genomics, pathogenicity, antibiotic resistance genes

OPEN ACCESS

Edited by:
David W. Ussery, University of Arkansas for Medical Sciences, United States

Reviewed by:
Zhi Ruan, Zhejiang University, China
Robin Anderson, Agricultural Research Service, United States Department of Agriculture, United States

***Correspondence:**
Jorunn Pauline Cavanagh
Pauline.cavanagh@uit.no

Specialty section:
This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal Frontiers in Microbiology

Received: *16 April 2019*
Accepted: *26 August 2019*
Published: *10 September 2019*

Citation:
Pain M, Hjerde E, Klingenberg C and Cavanagh JP (2019) Comparative Genomic Analysis of Staphylococcus haemolyticus Reveals Key to Hospital Adaptation and Pathogenicity. Front. Microbiol. 10:2096. doi: 10.3389/fmicb.2019.02096

Abbreviations: ARG, antibiotic resistance genes; aSec, accessory secretion system; CARD, comprehensive antibiotic resistance database; CDS, coding sequence; COG, cluster of orthologous groups; CoNS, coagulase negative staphylococci; CP, capsule polysaccharide; ENA, European Nucleotide Archive; IS, insertion sequences; MDR, multi drug resistant; Orf, open reading frame; WGS, whole genome sequence.

INTRODUCTION

Staphylococcus haemolyticus is an emerging pathogen of nosocomial infections, and the most frequently isolated coagulase-negative staphylococcal (CoNS) species alongside *Staphylococcus epidermidis* (Hope et al., 2008; Pereira et al., 2014; Nanoukon et al., 2017; Teeraputon et al., 2017). *S. haemolyticus* infections particularly affect immunocompromised patients and mainly occur as bloodstream and device-associated infections. Nosocomial *S. haemolyticus* isolates are ranked as the most antibiotic resistant species among the CoNS, and antibiotic therapy choices are therefore very limited (Hope et al., 2008; Kresken et al., 2011; Barros et al., 2012).

Compared to the more virulent *Staphylococcus aureus*, *S. haemolyticus* possesses few typical virulence factors (Takeuchi et al., 2005). Formation of biofilm (Fredheim et al., 2009; Giormezis et al., 2014; Pereira et al., 2014), production of phenol-soluble modulins (Da et al., 2017) and frequent phenotypic rearrangements due to a large number of insertion sequences (IS) (Takeuchi et al., 2005) have been suggested as important *S. haemolyticus* virulence determinants. However, these traits have not yet been linked explicitly to strains of clinical origin. The *oriC* environ is a chromosomal region of staphylococci proposed to be important for the evolution and differentiation of each staphylococcal species. There is little homology between the *oriC* environ of the different staphylococcal species, and the region does not contain genes essential for viability. The o*riC* environ is significantly larger in *S. haemolyticus* compared to that of *S. aureus* and *Staphylococcus epidermidis*. Moreover, almost half of the candidate coding sequences (CDS) for virulence are located within the *oriC* environ, encoding e.g., surface adhesins and capsular polysaccharides, factors that can modulate adherence and contribute to phagocytosis resistance (Takeuchi et al., 2005; Flahaut et al., 2008).

Despite the advancing clinical relevance of *S. haemolyticus*, knowledge about the transition from a commensal to an invasive lifestyle remains sparse. Moreover, there is a paucity of studies comparing pathogenicity traits between commensal and clinical invasive isolates. Predicting invasiveness of staphylococcal strains by use of marker genes is one approach to differentiate isolates with different pathogenicity. For *S. epidermidis*, the *ica* operon encoding biofilm formation and the insertion sequence element IS256; associated with aminoglycoside resistance, have been proposed as markers for invasive strains of *S. epidermidis* (Kozitskaya et al., 2004; Rohde et al., 2004). More recently, Méric et al. (2018) presented how calculation of a genotype risk score can predict pathogenicity in *S. epidermidis* isolates with 80% accuracy. With the advances in sequencing technologies and more available bacterial genomes new methods for analysis have emerged. The total number of genes in a bacterial population, collectively called the pan-genome, allows identification of genes more present and important in pathogenic strains (Tettelin et al., 2005). Studies of *S. haemolyticus* to date have focused predominantly on clinical isolates. The aim of this study was to identify factors important for infection and hospital adaptation by exploring the genomic variability of clinical and commensal *S. haemolyticus* using a pan-genomic approach.

MATERIALS AND METHODS

Bacterial Isolates, Species Identification, Antibiotic Susceptibility and Biofilm Testing

This study includes 169 *S. haemolyticus* isolates; 123 clinical isolates and 46 commensal isolates. The clinical isolates (mainly from blood cultures, invasive catheters and wounds), in addition to five of the commensal isolates, were from different hospital units and had different geographical origins. They were collected between 1988 and 2010, and have been described previously (Cavanagh et al., 2014). The remaining 41 commensal isolates were skin samples isolated from healthy volunteers with no antibiotic exposure and no hospitalization or health care affiliation during the three previous months (Cavanagh et al., 2016). This was a separate collection from one geographical location (Tromsø, Norway) and collected between 2013 and 2014 (Cavanagh et al., 2016). An overview of all the isolates used and their characteristics can be found in **Supplementary Data S1**. Species identification was done by 16s rRNA sequencing and/or matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS) using a Microflex LT instrument (Bruker Daltonics, Bremen, Germany), Flex Control software and the MALDI Biotyper 3.1 software (Bruker Daltonics, Bremen, Germany). Antibiotic susceptibility testing was performed as previously described (Cavanagh et al., 2014, 2016), and interpreted according to the 9th version of the EUCAST guidelines[1]. Isolates resistant to three or more classes of antibiotics were classified as multidrug resistant (MDR). Semi-quantitative determination of biofilm formation was performed with the modified Christensen assay, as described previously (Fredheim et al., 2009). Isolates were considered biofilm positive if they had an optical density (OD) value >0.2 above the negative control.

Whole Genome Sequencing (WGS), Assembly and Annotation

The WGS procedure for all clinical isolates and five of the 46 commensal isolates is described in a previous study (Cavanagh et al., 2014). For the remaining 41 commensal isolates, bacterial DNA was extracted and prepared for WGS using the Wizard Genomic DNA purification kit (Promega, Madison, United States) according to the manufacturer's instructions. WGS was performed on index-tagged libraries for each *S. haemolyticus* strain by paired-end sequencing at the Norwegian Sequencing Centre on an Illumina MiSeq (Illumina Inc., San Diego, California, United States). Subsequently, all 169 genomes were (re-assembled using SPAdes version 3.7 software (Bankevich et al., 2012), with some modification to the default parameters. Contigs >500 bp were reordered relative to the only complete fully annotated closed reference genome (JCSC 1435) (Takeuchi et al., 2005) using ABACAS (version 1.3.1). Protein CDSs were predicted using Prokka v1.12 (Seemann, 2014) using default

[1]http://www.eucast.org/

settings. The sequences are deposited in the European Nucleotide Archive[2]; study accession no ERP000943 and ERP114853.

Phylogeny and Molecular Typing

Subtyping of the 169 isolates was performed using kSNP3 to identify single-nucleotide polymorphisms (SNPs) in the core genomes and to reconstruct a parsimony phylogenomic tree (Gardner et al., 2015). Phylogenetic trees were visualized using the online tool iTol[3].

Pan-Genome and Pan-Genome-Wide Association Study (Pan-GWAS) of *S. haemolyticus*

Pan-genome analyses of all 169 isolates were performed using the Roary software package with default settings (Page et al., 2015). The program generates a file containing all the predicted gene clusters and the sequence identifier of each isolate containing said gene. Based on this file, core, accessory and unique genes were extracted and saved as individual lists. The accessory list was subdivided based on clusters common for both clinical and commensal isolates, in addition to clusters unique to each group. The unique list was further subdivided based on genes identified in clinical and commensal isolates. These lists were uploaded to eggNOG to get cluster of orthologs groups (COG) identifications (Huerta-Cepas et al., 2016).

Files created by Roary, were used as input for Scoary, a microbial pan-GWAS tool that calculates the association between all genes in the accessory genome and traits defined by the users. For our purpose we only used one trait; whether a given isolate were of clinical or commensal origin. Based on this information Scoary reported a list of genes sorted by strength of association per gene (Brynildsrud et al., 2016).

In silico Analysis and Statistics

The resistance gene identifier in the comprehensive antibiotic resistance database (CARD; version 1.1.1; Department of Biochemistry and Biomedical Science; McMaster University, Canada (Jia et al., 2017) was used to predict genes presumed to confer antibiotic resistance, and the results were further compared with the phenotypic susceptibility test results. Potential virulence factors were identified by homology searches against the virulence factor database (VFDB) together with putative virulence factors previously predicted by Takeuchi et al. (2005) and Chen et al. (2016). Identification of IS elements was performed using the ISsaga program (Varani et al., 2011). The presence of IS elements were confirmed by sequence search against the complete *S. haemolyticus* sequence collection (both on contigs and CDS). Sequence coverage of contigs harboring IS elements was used to quantify the copy numbers: the coverage of contigs with IS elements divided by the overall average coverage of its respective genome. Identification of putative plasmids was performed by screening the genome assemblies for plasmid replicon (*rep*) genes using the PlasmidFinder database (Carattoli et al., 2014) with coverage settings set to default of 75%. Identification of prophages, potentially important for horizontal gene transfer (HGT), was performed by using PHASTER (Arndt et al., 2016). Plasmid replicon sequences with more than 80% coverage and predicted intact phages identified in more than 10 isolates were further investigated by sequence search in order to confirm the presence of plasmid and phage genes, respectively.

Data were also analyzed using IBM-SPSS statistical software (IBM Corp. Released 2015. IBM SPSS Statistics for Windows, Version 23.0. Armonk, NY: IBM Corp.). Categorical data are displayed as ratios and frequency (%), and analyzed using the Chi square test. Pathogenicity is a complex multifactorial property. We used biological knowledge to identify known pathogenicity-associated traits enriched in clinical isolates. In particular, we focused on antibiotic resistance genes (ARGs) and previously established virulence factors. We developed different scores including the following four traits; *aacA-aphD*, *mecA*, *folP* and phenotypic biofilm production. In order to find a pragmatic score that could differentiate between a clinical and a commensal isolate we calculated the area under receiver operating characteristic (ROC) curves, and its 95% confidence interval.

RESULTS

Genome Composition and Genetic Variability

The average size of the assembled genomes was 2.52 Mb (2.32–2.86 Mb), with an average of 120 (32–415) contigs per genome. Each genome had on average 2,457 (2,239–2,816) predicted protein sequences (CDSs) with an average GC content of 32%.

The pan-genome of the 169 *S. haemolyticus* isolate dataset comprised 9,092 Cluster of Orthologous groups (COGs). We divided the pan-genome into core genes (genes shared by all strains), accessory genes (genes shared by some but not all strains) and unique genes only present in one genome (**Figure 1A**). The pan-genome sub-groups were annotated and sorted into COG categories (**Figure 1B**). Gene accumulation curves showed that the core genome plateaued at 1,522 genes reflecting a stable core and an open pan-genome where the addition of each new genome increases the total gene pool (**Figure 1C**).

Thirty-two percent of the annotated genes were categorized as function unknown, and this class was excluded from the graphical representation of the COG categories. The most abundant categories in the core genome were genes involved in housekeeping functions, like transcription and translation, and different metabolism categories. The accessory genome had a larger portion of genes associated with mobile genetic elements (MGE) such as transposons and bacteriophages (transcription, replication, recombination and repair (24%), which was also the most enriched category amongst the unique genes (25.7%). Genes involved in transcription (12%), genes associated with cell wall and membrane biogenesis (9.2%) and defense mechanisms (6.8%) were also considerably more prevalent in the accessory and unique gene pool, compared to the core gene pool. No

[2]www.ebi.ac.uk/ena

[3]https://itol.embl.de/

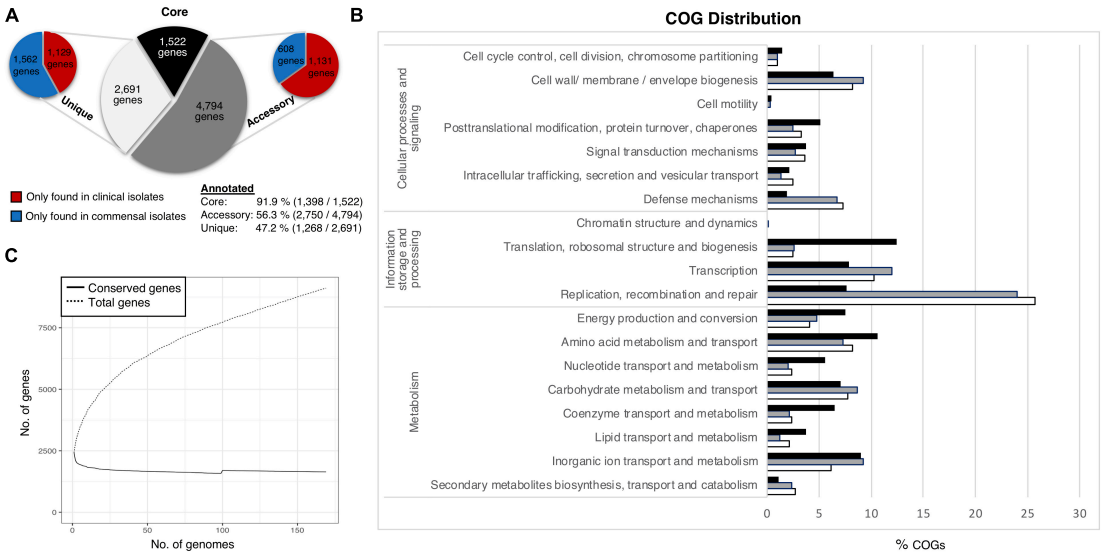


FIGURE 1 | *Staphylococcus haemolyticus* pangenome statistics. **(A)** The size and distribution of the pangenome into the subgroups; core (shared by all isolates), accessory (shared by some isolates) and unique (found only in one isolate). **(B)** Collective distribution of core (black), accessory (gray) and unique (white) genes in COG. **(C)** Pangenome curve generated by plotting total number of gene families in the pan and core genome.

significant differences of the COG distribution were found between the clinical and commensal group (data not shown).

Phylogeny and Population Structure

The phylogenetic reconstruction grouped the 169 isolates into six clades (A–F). There was a distinct distribution of clinical and commensal isolates into the different clades (**Figure 2**). The two largest clades, A and C, consisted almost exclusively of clinical isolates (88/90; 98%), while the majority of the commensal isolates (39/46; 85%) were found in clades D and F. The clinical isolates in clade A–C were more closely related compared to the commensal isolates. Twenty-eight of 123 (22.8%) clinical isolates were grouped in the three clades (D–F) predominantly consisting of commensal isolates. A long branch separating the "commensal heavy" clade F (78% commensal isolates) from the rest of the isolates underlines the high variability observed in the commensal isolates, compared to the clustered clinical isolates (**Figure 2**). We observed that with the exception of clade F, the isolates in all the other clades were closely related, independent of country of origin. The commensal isolates in clade F, mainly originating from the same geographic location in Norway, were in contrast very diverse.

Antibiotic Resistance and Mobile Genetic Elements (MGE)

Phenotypically, 113/169 (68%) of the isolates were classified as MDR; 108/123 (88%) of the clinical isolates (Cavanagh et al., 2014) and 5/46 (11%) of the commensal isolates (Cavanagh et al., 2016) ($p < 0.05$). Using a genotypic approach, we identified the ARGs for most of the observed phenotypes, and overall, phenotypic and genotypic resistance correlated well. However, ARGs toward eight additional antibiotic classes were also detected, but not phenotypically tested (**Figure 3**).

ARGs were among the most common genes among all clinical isolates. In particular *mecA*, *aacA/aphD, blaZ* were predominantly found in invasive isolates (**Table 1**). Phenotypic macrolide resistance was common in both the commensal (30/46; 65%) and the clinical (100/123; 81%) isolates. However, the majority of commensal isolates carried the macrolide resistance gene *mphC* (28/46; 61%) and few carried *ermA* (5/46; 11%), while the clinical isolates had both *mphC* (75/123; 61%) and *ermC* (52/123; 42.3%). The antiseptic resistance genes *qacA* and *qacB* were found in both clinical and commensal isolates, however *qacB* was identified predominantly in the commensal isolates whereas *qacA* was detected predominantly in clinical isolates (**Figure 3**).

For the plasmid analysis, we found a total of 51 replicon sequences, with an average of 5.5 replicon per clinical isolates and 7.6 average replicon in commensal isolates. When a replication sequence was identified in more than 10 isolates, we examined the presence of the entire plasmid. For smaller plasmids presence was easily determined, and numbers in clinical and commensal isolates were calculated and presented in **Table 2**. Larger plasmid sequences were often split over several contigs, thus not all of the plasmid genes were identified. If the whole plasmid could not be identified, we looked for the presence of its cargo genes instead. The replicon sequence of *S. haemolyticus* plasmid pSHaeB, carrying *ermC*, was only found in the clinical isolates (55/123; 45%). The *S. aureus* plasmid replicon for pUB110, carrying kanamycin and bleomycin resistance genes, was identified in 23/123 (19%) clinical and 2/46 (4%) commensal isolates. Chloramphenicol resistance plasmid pC221 was predicted based on the replicon sequence, and

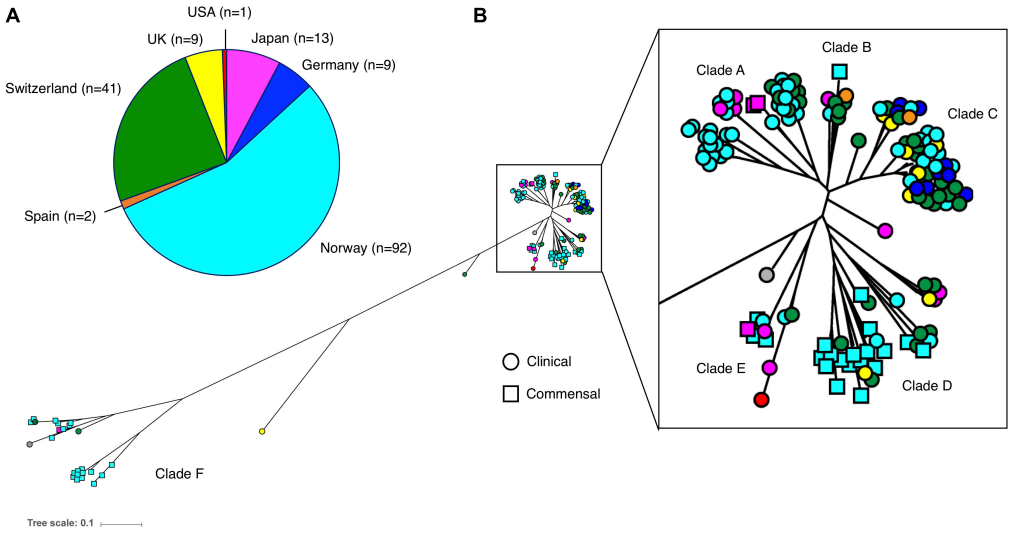
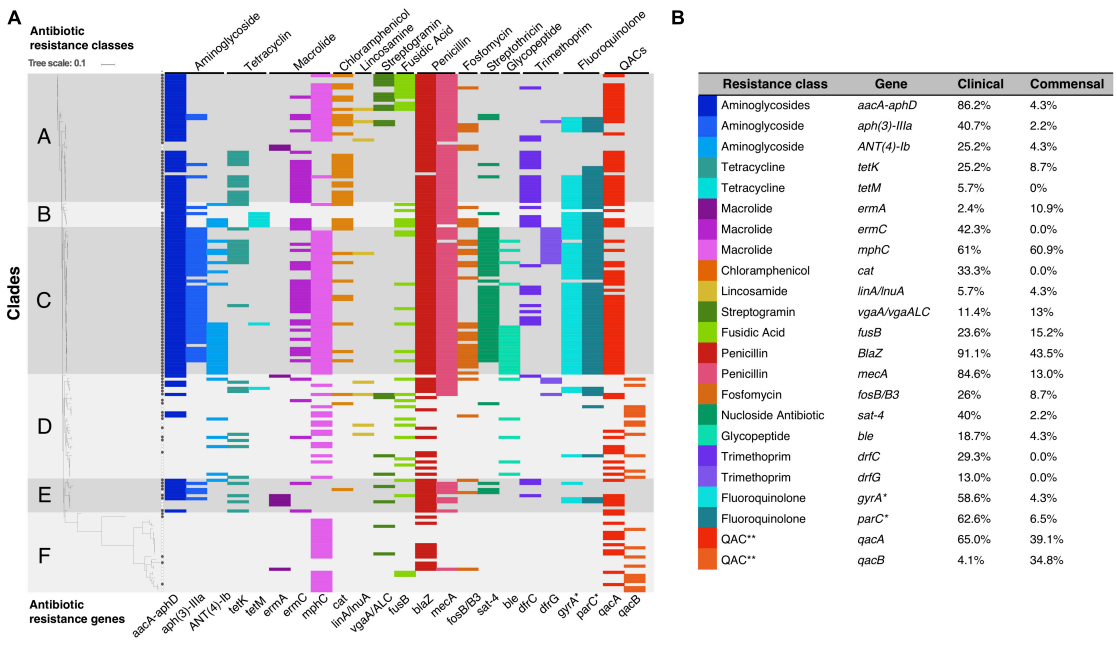


FIGURE 2 | (A) Number and origin of isolates included in this study. **(B)** Phylogenetic tree of the 169 *S. haemolyticus* isolates, based on SNPs in the core genome. Each isolate is color coded based on country of origin, as demonstrated in the pie chart. Clinical isolates are displayed as circles and commensal isolates as squares.



Resistance class	Gene	Clinical	Commensal
Aminoglycosides	*aacA-aphD*	86.2%	4.3%
Aminoglycoside	*aph(3)-IIIa*	40.7%	2.2%
Aminoglycoside	*ANT(4)-Ib*	25.2%	4.3%
Tetracycline	*tetK*	25.2%	8.7%
Tetracycline	*tetM*	5.7%	0%
Macrolide	*ermA*	2.4%	10.9%
Macrolide	*ermC*	42.3%	0.0%
Macrolide	*mphC*	61%	60.9%
Chloramphenicol	*cat*	33.3%	0.0%
Lincosamide	*linA/lnuA*	5.7%	4.3%
Streptogramin	*vgaA/vgaALC*	11.4%	13%
Fusidic Acid	*fusB*	23.6%	15.2%
Penicillin	*BlaZ*	91.1%	43.5%
Penicillin	*mecA*	84.6%	13.0%
Fosfomycin	*fosB/B3*	26%	8.7%
Nucloside Antibiotic	*sat-4*	40%	2.2%
Glycopeptide	*ble*	18.7%	4.3%
Trimethoprim	*drfC*	29.3%	0.0%
Trimethoprim	*drfG*	13.0%	0.0%
Fluoroquinolone	*gyrA**	58.6%	4.3%
Fluoroquinolone	*parC**	62.6%	6.5%
QAC**	*qacA*	65.0%	39.1%
QAC**	*qacB*	4.1%	34.8%

FIGURE 3 | (A) Graphical representation of *S. haemolyticus* antibiotic resistance genotype plotted onto the phylogenetic SNP-based core tree. The figure shows the presence and absence of ARGs and classes across the different clades. *Mutations in *gyrA* and *parC* gives resistance to fluoroquinolones. ** QACs: quaternary ammonium compounds. **(B)** The percentage of each ARG present in the clinical and commensal subgroups.

the chloramphenicol gene (*cat*) was identified in 39 isolates, all of clinical origin. In addition, the plasmids pSHaeA and pSHAaeC, the former carrying the fosfomycin resistance gene, were identified in our collection but in less than 1/3 of all isolates, and with no clear difference between clinical and commensal isolates. Several larger plasmids were predicted based on replicon sequence but none of these were identified in their entirety. We extracted what we considered interesting cargo genes from these plasmids and determined the prevalence of these (**Table 3**).

We predicted 13 different intact prophages in our collection, identified in 68% (114/169) of the isolates (clinical 94/123; 76%

TABLE 1 | Genes and gene versions enriched in 123 clinical isolates, sorted by significance.

Gene name	SH_ref	Annotation	Clinical (%)	Commensal (%)
NA	SH1612	Acetyltransferase GNAT	86.2	4.4
aacA-aphD	SH1611	Bifunctional AAC/APH	85.4	4.4
folP^X	SH2496	Dihydropteroate synthase	78.9	0.0
folB^X	SH2495	Dihydroneopterin aldolase	78.1	0.0
NA	SH2607	AtP-binding protein (plasmid)	76.4	2.2
paaZ	SH0090	Bifunctional protein PaaZ	85.4	13.0
mecA	SH0091	Methicillin resistance gene	85.4	13.0
ugpQ	SH0089	Glycerophosphodiester phosphodiesterase	82.1	13.0
blaR1^X	SH1763	Regulatory protein BlaR1	84.6	19.6
tagE	SH2252	Truncated glycosyltransferase	86.2	21.7
blaZ^X	SH1764	Beta-lactamase	78.1	15.2
blaI^X	SH1762	Penicillinase repressor	60.2	2.2
ansA^X	SH1433	Putative L-asparaginase	55.3	0.0
NA^X	SH2156	Ferrichrome ABC transporter truncated	88.6	30.4
nikA^X	SH0292	Nickel-binding periplasmic protein	94.3	41.3
fabG^X	SH0438	FabG	89.4	32.6
NA	SH0155	MFS transporter	96.8	47.8
sraP^X	SH0326	SraP	96.8	47.8
secA2^X	SH0331	Protein translocase subunit SecA2	96.8	47.8
yjjG^X	SH0452	Pyrimidine 5'-nucleotidase YjjG	84.6	26.1
PHAGE		StauST398-2/vB_saus_phi2	51.2	0.0

All these genes/gene versions were significantly enriched in clinical isolates ($p < 0.001$), the order of presentation from top of the table indicates those with the lowest p values. ^{X} These genes exist in different conserved versions in the clinical and commensal isolates. SH_ref; S. haemolyticus reference gene numbers from Takeuchi et al. (2005), accession number AP006716.1. The complete list of genes can be found in ***Supplementary Data S2****.*

TABLE 2 | Smaller plasmids and their distribution in clinical (n = 123) and commensal (n = 46) isolates.

Plasmid	Origin	Size	Associated genes	Clinical (%)	Commensal (%)
pSHaeA	*S. haemolyticus*	2,300 bp	*fosB* (fosfomycin resistance)	26.8	8.7
pSHaeB	*S. haemolyticus*	2,366 bp	*ermC* (erythromycin resistance)	44.7	0
pSHaeC	*S. haemolyticus*	8,180 bp	MFS transporter, *mobC*, *marR* transcriptional regulator	8.9	6.5
pWBG1773	*S. aureus*	2,916 bp		25.2	28.3
pKH21	*S. aureus*	2,531 bp	Lincosamin resistance	5.7	4.3
pWBG754	*S. aureus*	2,241 bp	*qacC*	29.3	17.4
pC221	*S. aureus*	4,555 bp	Chloramphenicol resistance	31.7	0
pUB110	*S. aureus*	4,548 bp	Bleomycin and kanamycin resistance	18.7	4.3

TABLE 3 | Distribution of genes associated with larger plasmids in the isolates.

Associated plasmids	Origin	Gene	Clinical (%)	Commensal (%)
pSE-12228-05, pWBG753	*S. epidermidis* *S. aureus*	Tn552	72.4	6.5
pSE-12228-05, pWBG753	*S. epidermidis* *S. aureus*	blaR1	91.9	43.5
pSE-12228-05	*S. epidermidis*	Penicillinase repressor	94.3	45.7
pWBG753	*S. aureus*	blaZ	91.9	43.5
VRSAp	*S. aureus*	aacA-aphD	85.3	4.3
VRSAp	*S. aureus*	qacA/B	69.1	73.9
VRSAp	*S. aureus*	qacR	67.5	76.1
pWBG753	*S. aureus*	tetK	26	8.7
SAP099B, pLEW6932	*S. aureus*, *Staphylococcus* sp.	cadC	17.1	17.4
SAP099B, pLEW6932	*S. aureus*, *Staphylococcus* sp.	arsR	4.1	63

and commensal 17/46; 37%). Five of these prophages were found in more than 10 isolates. The most prevalent phage, exclusively found in only clinical isolates (63/123; 51%), was predicted as staphylococcal phage vB_Saus_phi2 and *S. aureus* phage StauST398-2. Sequence searches of the two phage genes matched the same *S. haemolyticus* genes. Staphylococcus phage SPbeta-like is a phage of 127,726 bp with 156 CDS, and was identified in 21 clinical isolates and in one commensal isolate. This phage carries the genes *aacA-aphA* and *dfrC*, encoding resistance to gentamicin and trimethoprim, in addition to five IS256 and one IS431 element (CDS 146–156). Of all the isolates carrying this phage, we could identify the first 145 genes in the same order as in the reference phage.

We identified a large number of IS elements, ranging from 15 to 88 per each isolate. ISSha1 and IS1272 were found as several copies in all isolates. IS256 was found almost exclusively in clinical isolates (106/123; 86%) and rarely (5/46; 11%) in commensal isolates. The transposon Tn552/IS481 was also identified predominantly in clinical isolates (89/123; 72%) compared to commensal isolates (6/46; 13%).

Biofilm Production and Biofilm Encoding Genes

Phenotypic biofilm production was significantly more prevalent among clinical isolates (83/123; 67%) versus commensal isolates

(16/46; 35%), $p < 0.001$. However, we did not detect *ica* genes which typically encode biofilm production in other CoNS species.

Putative Surface Proteins

The C-terminal part of a hypothetical serine-rich surface protein (SH0326) named *sraP* (serine rich adhesin for platelets) was present in 119/123 (97%) of the clinical isolates and in 22/46 (48%) of commensal isolates. The partial *sraP* gene was located up-streams of *secY2*, the first gene in the accessory sec system (aSec); a structure dedicated to transport and modification of this surface protein. The N-terminal was located on a different contig which was, in most cases, correctly assembled directly upstream of the *aSec* contig. The central part of *SraP*, known to contain long stretches of repeated patterns was lost during assembly. Mapping sequence reads on assemblies of isolates, predicted to contain *sraP*, against the full length SH0326 showed that these strains did indeed have the full sequence. However, there appeared to be individual differences in the length of these repeat sequences. The complete *aSec* operon was identified in 119 clinical isolates (97%) and 34 (74%) of commensal isolates by performing sequence read mapping. Isolates without *sraP* and *aSec* belonged to the commensal clades D (12 isolates) and F (3 isolates), and one commensal isolate in clade B.

Putative Capsule Polysaccharide (CP) Operons

We identified four different structures of the CP operons among the 169 isolates. Seventeen isolates, all of clinical origin, had the *capA-M* operon structure similar to that described in *S. haemolyticus* JCSC 1435 (Takeuchi et al., 2005). Another 52 isolates had three potential novel CP operons (**Figure 4**). These three novel CP operons were homologs to the *capA-G* of *S. haemolyticus* JCSC 1435 (65–100% CDS identity). Each novel CP operon contained a *capH-K* region unique to its group, followed by the region $capM_{II}$-*P* (CDS 51–78% identity) which was common to all the three novel versions. The last gene in the operon, named *capM* in JCSC1435 was present in all four CP operon versions. *capM* identified in the three novel versions were named $capM_{II}$ to distinguish it from the already annotated *capM* found in JCSC 1435. $capL\text{-}M_{II}$ shared homology with *S. aureus* cap5/8. The novel CP version 1 was lacking the *capH* gene and *capI-K* did not show considerable homology to any other known CP genes. However, the novel CP version 2 *capH-K* showed homology to *S. aureus* cap8 (CDS 55–72% identity) and novel CP version 3 to *S. aureus* cap5 (CDS 57–68% identity).

Genes and MGE Significantly Different Between Clinical and Commensal Isolates

In order to identify a specific signature in the gene repertoire of the clinical isolates important for hospital adaptation, a pan-GWAS analysis was performed on the accessory genome. 1887 predicted genes were statistically different between the two groups ($p < 0.05$), and the genes found most enriched for each group are listed in **Tables 1**, **2**. As the population of *S. haemolyticus* is diverse, the gene sequence of genes with the same function varied significantly between the isolates, resulting in several orthologs clusters for one gene. Two enzymes in the folate pathway, dihydropteroate synthase (*folP*) and dihydroneopterin aldolase (*folB*) are a representative example of this. One version was only present in clinical isolates (96/123 isolates, 78%), and not found in any commensal isolates.

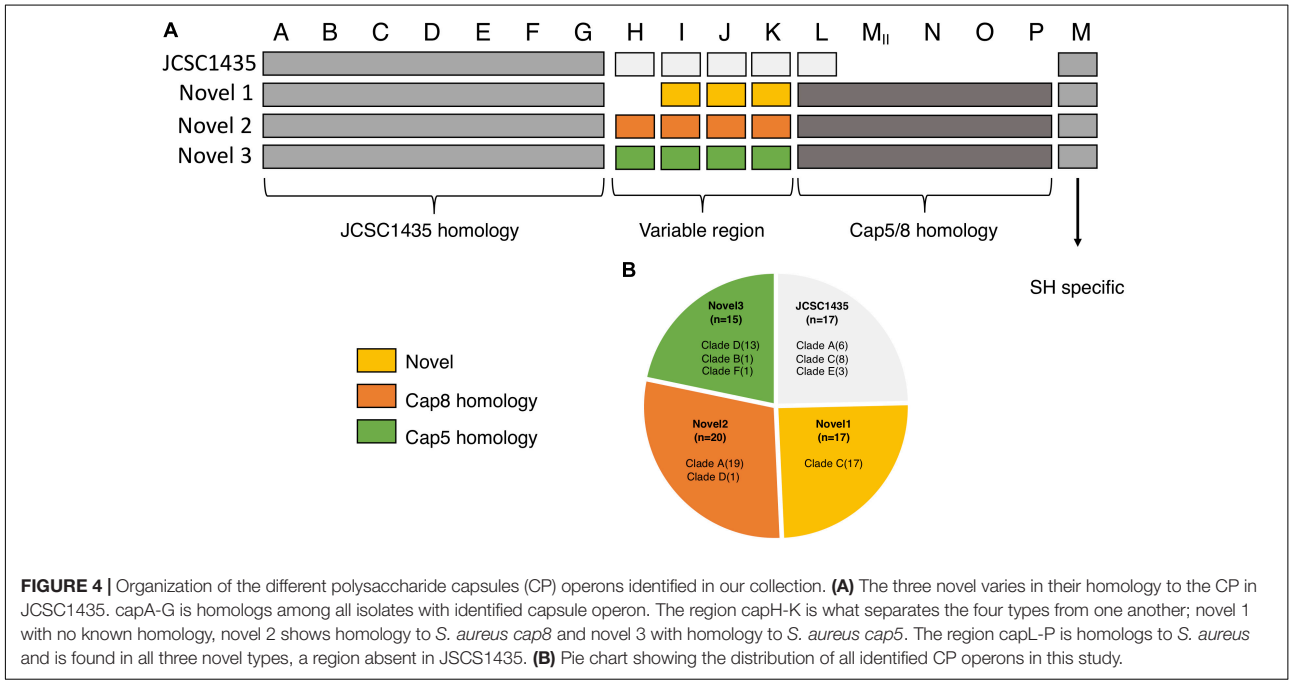


FIGURE 4 | Organization of the different polysaccharide capsules (CP) operons identified in our collection. **(A)** The three novel varies in their homology to the CP in JCSC1435. capA-G is homologs among all isolates with identified capsule operon. The region capH-K is what separates the four types from one another; novel 1 with no known homology, novel 2 shows homology to *S. aureus cap8* and novel 3 with homology to *S. aureus cap5*. The region capL-P is homologs to *S. aureus* and is found in all three novel types, a region absent in JSCS1435. **(B)** Pie chart showing the distribution of all identified CP operons in this study.

A different version was found in all 46 commensal isolates, and 27/123 (22%) of clinical isolates. The majority of the clinical isolates of the latter group belonged to the commensal clade D and F (19 isolates), with 8 isolates scattered amongst clade A, B and E. Extracting these CDSs from each isolate and aligning them displays conserved differences between the genes in the two groups.

The most prevalent genes with a known function enriched in the clinical isolates included several ARGs (**Table 1**). The most prevalent genes with a known function enriched in the commensal isolates included several genes involved in metal control, transposases, and generally different versions of genes found in clinical isolates (**Table 4**). Using scores including one or more ARGs and phenotypic biofilm formation resulted in ROC curves with area under the curve (AUC) values >0.9, indicating a high discriminatory capacity to differentiate between clinical and commensal isolates (**Figure 5**).

DISCUSSION

There is a significant lack of knowledge about *S. haemolyticus* pathogenicity. In the present comparative genome study of clinical and commensal isolates we identified several genetic determinants and genotypes associated with the pathogenicity and the success of *S. haemolyticus* in the hospital environment. Resistance to commonly used antimicrobial agents and disinfectants, in addition to genes encoding proteins involved in adhesion and human immune defense escape predominates in the clinical isolates.

Population Structure – Distinct Signatures of Clinical and Commensal Isolates

The establishment of hospital adapted *S. haemolyticus* clones was previously reported by our group (Cavanagh et al., 2014). In the current study the majority of the clinical and commensal isolates form distinct phylogenetic groups when phylogeny is reconstructed based on SNPs in the core genome. In clades where clinical isolates are grouped with commensal isolates, they display a signature more similar to the commensals. In *S. epidermidis* there has been more research on phylogenetic relationship between clinical isolates, but data reported are conflicting. Conlan et al. (2012) showed that commensal and invasive isolates were grouped as distinct groups. In contrast, Méric et al. (2018) recently demonstrated that pathogenic clones of *S. epidermidis* can arise from various commensal backgrounds. This was shown by the absence of specific disease associated clades in a comparative study of invasive and commensal isolates. *Staphylococcus haemolyticus* and *S. epidermidis* are the two most recovered CoNS species from infection. Abundance of *S. epidermidis* on skin enhances the probability of contamination of indwelling devices. While *S. haemolyticus* also is a skin commensal, the population structure observed among our isolates suggests that specific persistent hospital adapted clones of

TABLE 4 | Genes and gene versions enriched in 46 commensal isolates, sorted by significance.

Gene name	SH_ref	Annotation	Clinical (%)	Commensal (%)
folB^{X}	SH2495	Dihydroneopterin aldolase	22.0	100.0
folP^{X}	SH2496	Dihydropteroate synthase	23.6	100.0
group_2286		Major Facilitator (MFS)	0.0	56.5
csoR		Copper-sensing transcriptional repressor CsoR	8.1	76.1
copZ		Copper chaperone CopZ	8.1	73.9
copA		Copper-exporting P-type ATPase	8.1	71.7
czcD		Cadmium, cobalt and zinc/H(+)-K(+) antiporter	8.1	71.7
NA^{X}	SH0287	Caax amino terminal protease family protein	4.1	63.0
arsR		ArsR family transcpriptional regulator	4.1	63.0
ISSha1^{X}	SH2073	Transposase	0.0	47.8
grxC		Glutaredoxin	8.1	69.6
nikA^{X}	SH0292	Nickel-binding periplasmic protein	4.1	60.9
ydaF^{X}	SH2658	Putative ribosomal *N*-acetyltransferase YdaF	7.3	65.2
recQ	SH1430	ATP-dependent DNA helicase RecQ	44.7	100.0
ansA^{X}	SH1433	Putative L-asparaginase	45.5	100.0
IS30		IS30 family transposase	1.6	47.8
yjaB		Putative N-acetyltransferase YjaB	2.4	50.0
group_1785		Putative Insertion element/transposase	2.4	50.0
IS1272^{X}	SH2041	IS1272	0.0	39.1
yodJ ^{X}	SH0186	Putative carboxypeptidase YodJ	2.4	47.8

All these genes/gene versions were significantly enriched in commensal isolates ($p < 0.001$), the order of presentation from top of the table indicates those with the lowest p values. ^{X} These genes exist in different conserved versions in the clinical and commensal isolates. SH_ref; S. haemolyticus reference gene numbers from Takeuchi et al. (2005), accession number AP006716.1. The complete list of genes can be found in ***Supplementary Data S2****.*

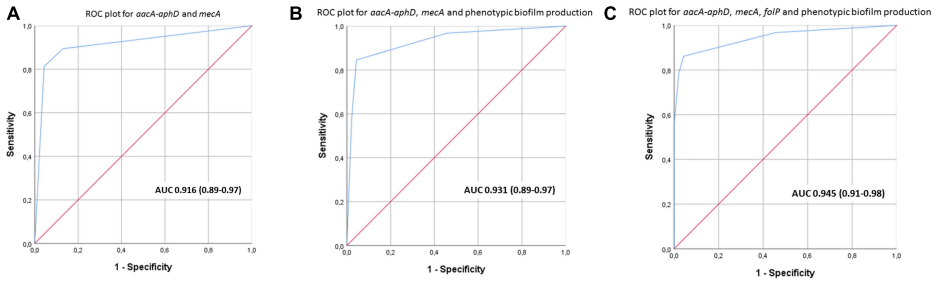


FIGURE 5 | (A–C) Receiver operating characteristics (ROC) curves with area under curve (95% confidence interval) for scores using different combinations of *aacA-aphD*, *mecA*, *folP* and phenotypic biofilm production in order to discriminate between clinical and commensal isolates.

S. haemolyticus are the major sources of infections, and that skin contamination is less likely (Tognetti et al., 2012; Kaspar et al., 2016; Byrd et al., 2018).

Surveillance and molecular typing are important in order to monitor the epidemiology of established and developing *S. haemolyticus* hospital clones. The availability of online resources, such as PubMLST[4], Enterobase[5], and BacWGSTdb[6], for bacterial typing offers rapid classification and source tracing, which is increasingly important in a globalized community (Ruan and Feng, 2016; Alikhan et al., 2018; Jolley et al., 2018; Ruan et al., 2019). Currently, it is possible to investigate *S. haemolyticus* epidemiology using the traditional MLST database[7], but extended MLST schemes are unavailable. As *S. haemolyticus* is an emerging nosocomial pathogen with extended antibiotic resistance, an online resource offering rapid typing and phylogenetic relatedness linked to antibiotic resistance genes and clinical data would be very useful.

The accessory gene repertoire of the clinical isolates is characterized by high prevalence of ARGs. This was expected as *S. haemolyticus* infections are commonly caused by MDR isolates (Barros et al., 2012; Chang et al., 2018). Creating a score with ARGs and phenotypic biofilm formation clearly separated the invasive from the commensal isolates. This can be useful in clinical microbiology when conveying results of a positive *S. haemolyticus* culture back to the treating clinician. Biofilm forming *S. haemolyticus* isolates that are resistant to oxacillin (*mecA*) and aminoglycosides (*aacA-aphD)* are most likely invasive isolates whereas absence of these traits strongly indicates a commensal isolate.

The interpretation of the *folB/folP* results is complex. We found distinct conserved differences in *folB* and *folP* clearly separating clinical from commensal *S. haemolyticus* isolates. In *S. aureus* five mutations in *folP* have been shown to directly contribute to sulfonamide resistance (Griffith et al., 2018). Combination treatment of trimetoprim-sulfamethoxazole has been used in infections caused by methicillin resistant staphylococci, and the rates of trimetoprim-sulfamethoxazole resistance is increasing (Khamash et al., 2018). Neither of the two versions of *folP* that we identified in our *S. haemolyticus* collection had the mutations shown to increase resistance to sulfonamides in *S. aureus*. However, there is substantial sequence variations between the *S. haemolyticus* and *S. aureus folB* and *folP* genes, and other mutations could also lead to decreased susceptibility. The importance of the *folP* versions in *S. haemolyticus* needs further evaluation, and extensive testing on sulfonamide resistance is needed.

In the successful epidemic *S. aureus* EMRSA-15, *mecA* acquisition, mutations in *gyrA* and *glrA;* resulting in fluoroquinolone resistance, and uptake of plasmids carrying *ermC,* was shown to be strong drivers of evolution (Holden et al., 2013). Uptake of plasmid pSHaeB carrying *ermC* was only observed for the clinical *S. haemolyticus* isolates, and has also been shown to be prevalent in several clinically important staphylococcal species (Águila-Arcos et al., 2017). The observed population structure in *S. haemolyticus* appears more similar to *S. aureus* than to *S. epidermidis.* In *S. aureus* defined pathogenic clones have adapted to the specific challenges of the hospital environment (Rasmussen et al., 2013), and our data support that similar phenomenon occurs in *S. haemolyticus*.

IS256, found in the majority of clinical *S. haemolyticus* isolates, has also been reported to be more common in nosocomial isolates of *S. epidermidis*, and has been associated with gentamicin resistance due to co-localization on the transposon Tn4001, in addition to biofilm formation (Kozitskaya et al., 2004; Conlan et al., 2012; Sharma et al., 2018). IS256 can shape the genome by affecting gene expression. In several successful virulent MRSA STs, such as ST 247, the presence of IS256 has been linked to increased virulence, vancomycin resistance and formation of small-colony variants (Kleinert et al., 2017). The opposite was demonstrated in Brazilian MRSA ST 239 isolates where IS256 was integrated near global regulatory genes (*agr* and *mgr*) likely leading to rapid changes of bacterial traits, resulting in reduced virulence (Botelho et al., 2019). IS256 has been suggested as a marker for molecular typing and identification of nosocomial, invasive *S. epidermidis* isolates (Gu et al., 2005; Yao et al., 2005; Montanaro et al., 2007). There is also some evidence indicating that IS256 is not only enriched within invasive isolates, but also more prevalent in isolates with poor treatment outcome (Post et al., 2017).

The transposon Tn552/IS481, was also predominantly identified in clinical *S. haemolyticus* isolates, and this MGE often

[4]http://pubmlst.org
[5]http://enterobase.warwick.ac.uk
[6]http://bacdb.org/BacWGSTdb
[7]https://pubmlst.org/shaemolyticus/

carry antiseptic resistance and beta lactamase genes (Anthonisen et al., 2002). The pandemic *S. aureus* ST239 has acquired resistance to multiple antibiotics and antiseptics, among them *qac A* and *B* (Chang et al., 2017). Qac proteins are efflux pumps that protect bacteria not only from a variety of toxic substances but also from fluorquinolones and beta-lactams (Prag et al., 2014; Wassenaar et al., 2015; Taheri et al., 2016). In our study, *qacA* was detected more often in clinical isolates, while *qacB* was almost exclusive to the commensal isolates. *qacA* has been reported to have a broader spectrum of resistance than *qacB,* and this might be the reason why we found *qacA* more often in the clinical samples (in 65% clinical isolates and 39% commensal isolates), obtained from an environment with higher antibiotic pressure (Wassenaar et al., 2015).

The Adhesion Protein SraP Is Highly Prevalent in Clinical Isolate

In the clinical isolates we commonly identified a homolog (SH0326) of the serine-rich repeat glycoproteins SraP of *S. aureus* and the accessory sec system, dedicated for the export of SraP. SraP belongs to a highly conserved family of serine-rich surface glycoproteins of Gram-positive bacteria. Expression of SraP has been linked to adhesion to different types of cells, including human platelets and is associated with infective endocarditis (Siboo et al., 2005; Yang et al., 2014; Bensing et al., 2016). The higher prevalence of the *aSec* system and *sraP* is likely advantageous for the clinical *S. haemolyticus* isolates as they are mainly isolated from blood. In *S. epidermidis,* genome signatures linked to pathogenicity identified the aSec gene *asp3* as one of four strong virulence predictors. In their full list of pathogenicity-associated signature genes *sraP* was also identified (Méric et al., 2018).

A Novel Capsular Polysaccharide Operon

The polysaccharide capsule (CP) in *S. haemolyticus* has been shown to play a role in the protection against uptake and killing by human neutrophils (Flahaut et al., 2008). In *S. aureus* it was demonstrated to modulate adherence to endothelial surfaces *in vitro*, and to promote bacterial colonization and persistence on mucosal surfaces in animal models (Riordan and Lee, 2004).

In this study we found four different capsule operons of which only the first type has been described in *S. haemolyticus* JCSC1435 previously (Takeuchi et al., 2005). Two of the three novel CP (*capNOP*) structures have homology to the *S. aureus* cap 5/8 genes, and identification of *capO* in *S. haemolyticus* has not previously been reported. The third type has no homology to any previously described cap version. The new CP versions in *S. haemolyticus* have a variable region, capH-K which is unique to each of the four versions with little or no homology between them. The GC content of this variable region is also significantly lower than the surrounding cap genes, indicating the variable region was acquired by HGT. Both CP genes and SCC are located in the *oriC* environ, a region of high genomic flexibility, allowing staphylococci to maintain or acquire genes needed for the adaption to on-going environmental changes (Hiramatsu et al., 2014). The new CP structures were mainly found in isolates belonging to two clinical clades. The four different CP types were shown to be clade specific, which has also been shown for *S. aureus*. In *S. aureus* 13 putative capsular operons have been reported, but only CP type 5 and 8 have been associated with disease (Mohamed et al., 2019), thus the three novel CP types now identified in *S. haemolyticus* need to be further investigated for their role in virulence.

Taken together, our findings point toward HGT as a driving force in *S. haemolyticus* evolution and in response to the selective pressure of broad-spectrum antibiotics used in hospitals.

The *S. haemolyticus* Pan-Genome Distribution Reflects the High Variation of Commensal Isolates

The pan-genome analysis reflected a relatively stable core genome which is comparable to what is observed in *S. epidermidis*, *S. aureus* and *Staphylococcus lugdunensis*. In *S. haemolyticus* the core genome is slightly smaller which could be explained by the higher number of unique genes in the commensal isolates (Méric et al., 2015, 2018; Argemi et al., 2018). Similar to *S. haemolyticus*, *S. aureus* and *S. epidermidis* also have an open pan-genome in contrast to *S. lugdunensis* (Bosi et al., 2016; Argemi et al., 2018). However, the pan-genome accumulation curves for *S. epidermidis* and *S. aureus* are not as steep as we see for *S. haemolyticus* (Conlan et al., 2012; Méric et al., 2015; Sharma et al., 2018). The relatively large pan-genome of *S. haemolyticus* is, at least in part, due to the variation seen in genes of the same function, resulting in two or more clusters for one gene. The most abundant categories in the core genome were genes involved in housekeeping functions, like transcription and translation, and different metabolism categories, a result similar to reports on *S. aureus* and *S. epidermidis* (Bosi et al., 2016; Sharma et al., 2018). The large repertoire of genes in the accessory genome confer advantages in highly variable environmental conditions.

Strengths and Limitation With the Study

This study is the largest comparative study of clinical and commensal *S. haemolyticus* isolates to date, and the clinical isolates have wide spatial and temporal distribution. We have used state of the art technology to analyze pathogenicity traits and the genetic signatures of clinical and invasive isolates. The study also has limitations. First, the majority of the commensal isolates are from one geographic location, and more recently collected than the invasive isolates. Still, the very diverse commensal population indicates that commensal isolates may lack the pathogenic traits of hospital-adapted clones. Second, the invasive isolates are collected over a wide time-span and some of the isolates originated from infections two decades ago. Still we see a very clear phylogenetic clustering, and invasive isolates from different geographical origin cluster together. We believe this clearly indicates the emergence of disease-causing isolates with a homogenous genetic signature. Third, it would have been of interest to collect skin samples from hospitalized patients without infection, in order to see whether colonization of more pathogenic isolates emerges after hospitalization. We did not have the opportunity to collect such samples in this study, but we

will pursue this in the future. Finally, we do not know whether some of the clonal groups are still circulating or might have been replaced by new clones. This phenomenon was observed in *S. aureus* where the pandemic clone ST239-MRSA-III that circulated for several decades (Monecke et al., 2018) was replaced with clones with increased fitness (Hsu et al., 2015). Still, we believe that our current study has identified important features of hospital adapted *S. haemolyticus* clones. Future sampling of both commensal and invasive isolates is needed in order to monitor the evolution of *S. haemolyticus*.

CONCLUSION

In this study we have gained a deeper understanding of the mechanism of adaption of *S. haemolyticus* in the hospital environment by phylogenetic and pan-genome analysis. We have found a clear segregation of isolates of commensal origin, and specific genetic signatures distinguishing the clinical isolates from the commensal isolates. It is highly likely that the widespread use of antimicrobial agents has promoted the development of MDR clones of *S. haemolyticus* persisting in the hospital environment, and that these isolates have responded through acquisition of mobile genetic elements or beneficial point mutations and rearrangements in surface associated genes. Defining pathogen-associated signatures is an important step in infection control. Continuous surveillance and molecular typing are important in order to monitor the spread and evolution of the *S. haemolyticus* hospital clones in the future.

DATA AVAILABILITY

The datasets generated for this study can be found in the European Nuclotide Archive (www.ebi.ac.uk/ena) study accession no. ERP000943 for the clinical isolates and ERP114853 for the commensal isolates. The DNA accession number for the individual sequences are found in **Supplementary Data S2**.

AUTHOR CONTRIBUTIONS

MP organized and performed the bioinformatic analysis, took part in the study design, wrote the first version of the manuscript, and revised the manuscript. EH participated in bioinformatical analyzes and revised the manuscript. JC and CK conceptualized and designed the study and revised the final manuscript. All authors approved the final manuscript as submitted and agreed to be accountable for all aspects of the work.

FUNDING

MP was funded by UiT The Arctic University of North Norway. The publication charges for this article have been funded by a grant from the publication fund of UiT The Arctic University of Norway.

ACKNOWLEDGMENTS

We thank Elizabeth Aarag Fredheim and Runa Wolden for sampling of volunteers for the collection of commensal *S. haemolyticus* isolates.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2019.02096/full#supplementary-material

REFERENCES

Águila-Arcos, S., Álvarez-Rodríguez, I., Garaiyurrebaso, O., Garbisu, C., Grohmann, E., and Alkorta, I. (2017). Biofilm-forming clinical *Staphylococcus* isolates harbor horizontal transfer and antibiotic resistance genes. *Front. Microbiol.* 8:2018. doi: 10.3389/fmicb.2017.02018

Alikhan, N. F., Zhou, Z., Sergeant, M. J., and Achtman, M. (2018). A genomic overview of the population structure of *Salmonella*. *PLoS Genet.* 14:e1007261. doi: 10.1371/journal.pgen.1007261

Anthonisen, I.-L., Sunde, M., Steinum, T. M., Sidhu, M. S., and Sørum, H. (2002). Organization of the antiseptic resistance gene qacA and Tn552-related beta-lactamase genes in multidrug- resistant *Staphylococcus haemolyticus* strains of animal and human origins. *Antimicrob. Agents Chemother.* 46, 3606–3612. doi: 10.1128/AAC.46.11.3606

Argemi, X., Matelska, D., Ginalski, K., Riegel, P., Hansmann, Y., Bloom, J., et al. (2018). Comparative genomic analysis of *Staphylococcus lugdunensis* shows a closed pan-genome and multiple barriers to horizontal gene transfer. *BMC Genomics* 19:621. doi: 10.1186/s12864-018-4978-4971

Arndt, D., Grant, J. R., Marcu, A., Sajed, T., Pon, A., Liang, Y., et al. (2016). PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res.* 44, W16–W21. doi: 10.1093/nar/gkw387

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. doi: 10.1089/cmb.2012.0021

Barros, E. M., Ceotto, H., Bastos, M. C. F., Dos Santos, K. R. N., and Giambiagi-deMarval, M. (2012). *Staphylococcus haemolyticus* as an important hospital pathogen and carrier of methicillin resistance genes. *J. Clin. Microbiol.* 50, 166–168. doi: 10.1128/JCM.05563-5511

Bensing, B. A., Loukachevitch, L. V., Mcculloch, K. M., Yu, H., Vann, K. R., Wawrzak, Z., et al. (2016). Structural basis for sialoglycan binding by the *Streptococcus sanguinis* SrpA adhesin. *J. Biol. Chem.* 291, 7230–7240. doi: 10.1074/jbc.M115.701425

Bosi, E., Monk, J. M., Aziz, R. K., Fondi, M., Nizet, V., and Palsson, B. Ø. (2016). Comparative genome-scale modelling of *Staphylococcus aureus* strains identifies strain-specific metabolic capabilities linked to pathogenicity. *Proc. Natl. Acad. Sci. U.S.A.* 113, E3801–E3809. doi: 10.1073/pnas.1523199113

Botelho, A. M. N., Cerqueira e Costa, M. O., Moustafa, A. M., Beltrame, C. O., Ferreira, F. A., Côrtes, M. F., et al. (2019). Local diversification of methicillin-resistant *Staphylococcus aureus* ST239 in South America after its rapid worldwide dissemination. *Front. Microbiol.* 10:82. doi: 10.3389/fmicb.2019.00082

Brynildsrud, O., Bohlin, J., Scheffer, L., and Eldholm, V. (2016). Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. *Genome Biol.* 17:238. doi: 10.1186/s13059-016-1108-1108

Byrd, A. L., Belkaid, Y., and Segre, J. A. (2018). The human skin microbiome. *Nat. Rev. Microbiol.* 16, 143–155. doi: 10.1038/nrmicro.2017.157

Carattoli, A., Zankari, E., Garciá-Fernández, A., Larsen, M. V., Lund, O., Villa, L., et al. (2014). In Silico detection and typing of plasmids using plasmidfinder and plasmid multilocus sequence typing. *Antimicrob. Agents Chemother.* 58, 3895–3903. doi: 10.1128/AAC.02412-14

Cavanagh, J. P., Hjerde, E., Holden, M. T. G., Kahlke, T., Klingenberg, C., Flaegstad, T., et al. (2014). Whole-genome sequencing reveals clonal expansion of multiresistant *Staphylococcus haemolyticus* in European hospitals. *J. Antimicrob. Chemother.* 69, 2920–2927. doi: 10.1093/jac/dku271

Cavanagh, J. P., Wolden, R., Heise, P., Esaiassen, E., Klingenberg, C., and Aarag Fredheim, E. G. (2016). Antimicrobial susceptibility and body site distribution of community isolates of coagulase-negative staphylococci. *APMIS* 124, 973–978. doi: 10.1111/apm.12591

Chang, P. H., Liu, T. P., Huang, P. Y., Lin, S. Y., Lin, J. F., Yeh, C. F., et al. (2018).). Clinical features, outcomes, and molecular characteristics of an outbreak of *Staphylococcus haemolyticus* infection, among a mass-burn casualty patient group, in a tertiary center in northern Taiwan. *J. Microbiol. Immunol. Infect* 51, 847–855. doi: 10.1016/j.jmii.2018.07.004

Chang, S.-C., Lee, M.-H., Yeh, C.-F., Liu, T.-P., Lin, J.-F., Ho, C.-M., et al. (2017). Characterization of two novel variants of staphylococcal cassette chromosome mec elements in oxacillin-resistant *Staphylococcus lugdunensis*. *J. Antimicrob. Chemother.* 72, 3258–3262. doi: 10.1093/jac/dkx291

Chen, L., Zheng, D., Liu, B., Yang, J., and Jin, Q. (2016). VFDB 2016: hierarchical and refined dataset for big data analysis - 10 years on. *Nucleic Acids Res.* 44, D694–D697. doi: 10.1093/nar/gkv1239

Conlan, S., Mijares, L. A., Becker, J., Blakesley, R. W., Bouffard, G. G., Brooks, S., et al. (2012). *Staphylococcus epidermidis* pan-genome sequence analysis reveals diversity of skin commensal and hospital infection-associated isolates. *Genome Biol.* 13:R64. doi: 10.1186/gb-2012-13-7-r64

Da, F., Joo, H.-S., Cheung, G. Y. C., Villaruz, A. E., Rohde, H., Luo, X., et al. (2017). Phenol-soluble modulin toxins of *Staphylococcus haemolyticus*. *Front. Cell. Infect. Microbiol.* 7:206. doi: 10.3389/fcimb.2017.00206

Flahaut, S., Vinogradov, E., Kelley, K. A., Brennan, S., Hiramatsu, K., and Lee, J. C. (2008). Structural and biological characterization of a capsular polysaccharide produced by *Staphylococcus haemolyticus*. *J. Bacteriol.* 190, 1649–1657. doi: 10.1128/JB.01648-1647

Fredheim, E. G. A., Klingenberg, C., Rohde, H., Frankenberger, S., Gaustad, P., Flaegstad, T., et al. (2009). Biofilm formation by *Staphylococcus haemolyticus*. *J. Clin. Microbiol.* 47, 1172–1180. doi: 10.1128/JCM.01891-1898

Gardner, S. N., Slezak, T., and Hall, B. G. (2015). kSNP3.0: SNP detection and phylogenetic analysis of genomes without genome alignment or reference genome. *Bioinformatics* 31, 2877–2878. doi: 10.1093/bioinformatics/btv271

Giormezis, N., Kolonitsiou, F., Foka, A., Drougka, E., Liakopoulos, A., Makri, A., et al. (2014). Coagulase-negative staphylococcal bloodstream and prosthetic-device-associated infections: the role of biofilm formation and distribution of adhesin and toxin genes. *J. Med. Microbiol.* 63, 1500–1508. doi: 10.1099/jmm.0.075259-0

Griffith, E. C., Wallace, M. J., Wu, Y., Kumar, G., Gajewski, S., Jackson, P., et al. (2018). The structural and functional basis for recurring sulfa drug resistance mutations in *Staphylococcus aureus* dihydropteroate synthase. *Front. Microbiol.* 9:1369. doi: 10.3389/fmicb.2018.01369

Gu, J., Li, H., Li, M., Vuong, C., Otto, M., Wen, Y., et al. (2005). Bacterial insertion sequence IS256 as a potential molecular marker to discriminate invasive strains from commensal strains of *Staphylococcus epidermidis*. *J. Hosp. Infect* 61, 342–348. doi: 10.1016/j.jhin.2005.04.017

Hiramatsu, K., Katayama, Y., Matsuo, M., Sasaki, T., Morimoto, Y., Sekiguchi, A., et al. (2014). Multi-drug-resistant *Staphylococcus aureus* and future chemotherapy. *J. Infect. Chemother.* 20, 593–601. doi: 10.1016/j.jiac.2014.08.001

Holden, M. T. G., Hsu, L., Kurt, K., Weinert, L. A., Mather, A. E., Harris, S. R., et al. (2013). A genomic portrait of the emergences, evolution, and global spread of methicillin-resistant *Staphylococcus aureus* pandemic. *Genome Res.* 23, 653–664. doi: 10.1101/gr.147710.112

Hope, R., Livermore, D. M., Brick, G., Lillie, M., and Reynolds, R. (2008). Non-susceptibility trends among staphylococci from bacteraemias in the UK and Ireland, 2001-06. *J. Antimicrob. Chemother.* 62, ii65–ii74. doi: 10.1093/jac/dkn353

Hsu, L.-Y., Harris, S. R., Chlebowicz, M. A., Lindsay, J. A., Koh, T.-H., Krishnan, P., et al. (2015). Evolutionary dynamics of methicillin-resistant *Staphylococcus aureus* within a healthcare system. *Genome Biol.* 16, 1–13. doi: 10.1186/s13059-015-0643-z

Huerta-Cepas, J., Szklarczyk, D., Forslund, K., Cook, H., Heller, D., Walter, M. C., et al. (2016). EGGNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res.* 44, D286–D293. doi: 10.1093/nar/gkv1248

Jia, B., Raphenya, A. R., Alcock, B., Waglechner, N., Guo, P., Tsang, K. K., et al. (2017). CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Res.* 45, D566–D573. doi: 10.1093/nar/gkw1004

Jolley, K. A., Bray, J. E., and Maiden, M. C. J. (2018). Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications [version 1; referees: 2 approved]. *Wellcome Open Res.* 3, 1–20. doi: 10.12688/wellcomeopenres.14826.1

Kaspar, U., Kriegeskorte, A., Schubert, T., Peters, G., Rudack, C., Pieper, D. H., et al. (2016). The culturome of the human nose habitats reveals individual bacterial fingerprint patterns. *Environ. Microbiol.* 18, 2130–2142. doi: 10.1111/1462-2920.12891

Khamash, D. F., Voskertchian, A., Tamma, P. D., Akinboyo, I. C., Carroll, K. C., and Milstone, A. M. (2018). Increasing clindamycin and trimethoprim-sulfamethoxazole resistance in pediatric *Staphylococcus aureus* infections. *J. Pediatric Infect. Dis. Soc.* doi: 10.1093/jpids/piy062 [Epub ahead of print].

Kleinert, F., Kallies, R., Hort, M., Zweynert, A., Szekat, C., Nagel, M., et al. (2017). Influence of IS 256 on genome variability and formation of small-colony variants in *Staphylococcus aureus*. *Antimicrob. Agents Chemother.* 61:e00144-17. doi: 10.1128/AAC.00144-117

Kozitskaya, S., Cho, S.-H., Dietrich, K., Marre, R., Naber, K., and Ziebuhr, W. (2004). The bacterial insertion sequence element IS256 occurs preferentially in nosocomial *Staphylococcus epidermidis* isolates: association with biofilm formation and resistance to aminoglycosides. *Infect. Immun.* 72, 1210–1215. doi: 10.1128/IAI.72.2.1210-1215.2004

Kresken, M., Becker, K., Seifert, H., Leitner, E., Körber-Irrgang, B., Von Eiff, C., et al. (2011). Resistance trends and in vitro activity of tigecycline and 17 other antimicrobial agents against Gram-positive and Gram-negative organisms, including multidrug-resistant pathogens, in Germany. *Eur. J. Clin. Microbiol. Infect. Dis.* 30, 1095–1103. doi: 10.1007/s10096-011-1197-y

Méric, G., Mageiros, L., Pensar, J., Laabei, M., Yahara, K., Pascoe, B., et al. (2018). Disease-associated genotypes of the commensal skin bacterium *Staphylococcus epidermidis*. *Nat. Commun.* 9:5034. doi: 10.1038/s41467-018-07368-7367

Méric, G., Miragaia, M., de Been, M., Yahara, K., Pascoe, B., Mageiros, L., et al. (2015). Ecological overlap and horizontal gene transfer in *Staphylococcus aureus* and *Staphylococcus epidermidis*. *Genome Biol. Evol.* 7, 1313–1328. doi: 10.1093/gbe/evv066

Mohamed, N., Timofeyeva, Y., Jamrozy, D., Rojas, E., Hao, L., Silmon de Monerri, N. C., et al. (2019). Molecular epidemiology and expression of capsular polysaccharides in *Staphylococcus aureus* clinical isolates in the United States. *PLoS One* 14:e0208356. doi: 10.1371/journal.pone.0208356

Monecke, S., Slickers, P., Gawlik, D., Müller, E., Reissig, A., Ruppelt-Lorz, A., et al. (2018). Molecular typing of ST239-MRSA-III from diverse geographic locations and the evolution of the SCCmec III element during its intercontinental spread. *Front. Microbiol.* 9:1436. doi: 10.3389/fmicb.2018.01436

Montanaro, L., Campoccia, D., Pirini, V., Ravaioli, S., Otto, M., and Aricola, C. R. (2007). Antibiotic multiresistance strictly associated with IS256 and ica genes in *Staphylococcus epidermidis* strains from implant orthopedic infections. *J. Biomed. Mater. Res.* 83A, 813–818. doi: 10.1002/jbm.a.31399

Nanoukon, C., Argemi, X., Sogbo, F., Orekan, J., Keller, D., Affolabi, D., et al. (2017). Pathogenic features of clinically significant coagulase-negative staphylococci in hospital and community infections in Benin. *Int. J. Med. Microbiol.* 307, 75–82. doi: 10.1016/j.ijmm.2016.11.001

Page, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T. G., et al. (2015). Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 31, 3691–3693. doi: 10.1093/bioinformatics/btv421

Pereira, P. M. A., Binatti, V. B., Sued, B. P. R., Ramos, J. N., Peixoto, R. S., Simões, C., et al. (2014). *Staphylococcus haemolyticus* disseminated among neonates with bacteremia in a neonatal intensive care unit in Rio de Janeiro. *Brazil.*

Diagn. Microbiol. Infect. Dis. 78, 85–92. doi: 10.1016/j.diagmicrobio.2013.06.026

Post, V., Harris, L. G., Morgenstern, M., Mageiros, L., Hitchings, M. D., Méric, G., et al. (2017). Comparative genomics study of *Staphylococcus epidermidis* isolates from orthopedic-device-related infections correlated with patient outcome. *J. Clin. Microbiol.* 55, 3089–3103. doi: 10.1128/JCM.00881-17

Prag, G., Falk-Brynhildsen, K., Jacobsson, S., Hellmark, B., Unemo, M., and Söderquist, B. (2014). Decreased susceptibility to chlorhexidine and prevalence of disinfectant resistance genes among clinical isolates of *Staphylococcus epidermidis*. *APMIS* 122, 961–967. doi: 10.1111/apm.12239

Rasmussen, G., Monecke, S., Ehricht, R., and Söderquist, B. (2013). Prevalence of clonal complexes and virulence genes among commensal and invasive *Staphylococcus aureus* isolates in Sweden. *PLoS One* 8:e7747. doi: 10.1371/journal.pone.0077477

Riordan, K. O., and Lee, J. C. (2004). *Staphylococcus aureus* capsular polysaccharides. *Clin. Microbiol. Rev.* 17, 218–234. doi: 10.1128/CMR.17.1.218

Rohde, H., Kalitzky, M., Kröger, N., Scherpe, S., Horstkotte, M. A., Knobloch, J. K. M., et al. (2004). Detection of virulence-associated genes not useful for discriminating between invasive and commensal *Staphylococcus epidermidis* strains from a bone marrow transplant unit. *J. Clin. Microbiol.* 42, 5614–5619. doi: 10.1128/JCM.42.12.5614-5619.2004

Ruan, Z., and Feng, Y. (2016). BacWGSTdb, a database for genotyping and source tracking bacterial pathogens. *Nucleic Acids Res.* 44, D682–D687. doi: 10.1093/nar/gkv1004

Ruan, Z., Yu, Y., and Feng, Y. (2019). The global dissemination of bacterial infections necessitates the study of reverse genomic epidemiology. *Brief. Bioinform.* 00, 1–10. doi: 10.1093/bib/bbz010

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153

Sharma, S., Chaudhry, V., Kumar, S., and Patil, P. B. (2018). Phylogenomic based comparative studies on Indian and American commensal *Staphylococcus epidermidis* isolates. *Front. Microbiol* 9:333. doi: 10.3389/fmicb.2018.00333

Siboo, I. R., Chambers, H. F., and Sullam, P. M. (2005). Role of SraP, a serine-rich surface protein of *Staphylococcus aureus*, in binding to human platelets. *Infect. Immun.* 73, 2273–2280. doi: 10.1128/IAI.73.4.2273

Taheri, N., Ardebili, A., Amouzandeh-Nobaveh, A., and Ghaznavi-Rad, E. (2016). Frequency of antiseptic resistance among *Staphylococcus aureus* and coagulase-negative staphylococci isolated from a university hospital in Central Iran. *Oman Med. J.* 31, 426–432. doi: 10.5001/omj.2016.86

Takeuchi, F., Watanabe, S., Baba, T., Yuzawa, H., Morimoto, Y., Kuroda, M., et al. (2005). Whole-genome sequencing of *Staphylococcus haemolyticus* uncovers the extreme plasticity of its genome and the evolution of human-colonizing staphylococcal species. *J. Bacteriol.* 187, 7292–7308. doi: 10.1128/JB.187.21.7292

Teeraputon, S., Santanirand, P., Wongchai, T., Songjang, W., Lapsomthob, N., Jaikrasun, D., et al. (2017). Prevalence of methicillin resistance and macrolide–lincosamide–streptogramin B resistance in *Staphylococcus haemolyticus* among clinical strains at a tertiary-care hospital in Thailand. *New Microbes New Infect.* 19, 28–33. doi: 10.1016/j.nmni.2017.05.007

Tettelin, H., Masignani, V., Cieslewicz, M. J., Donati, C., Medini, D., Ward, N. L., et al. (2005). Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome.". *Proc. Natl. Acad. Sci. U.S.A.* 102, 13950–13955. doi: 10.1073/pnas.0506758102

Tognetti, L., Martinelli, C., Berti, S., Hercogova, J., Lotti, T., Leoncini, F., et al. (2012). Bacterial skin and soft tissue infections: review of the epidemiology, microbiology, aetiopathogenesis and treatment. *J. Eur. Acad. Dermatol. Venereol.* 26, 931–941. doi: 10.1111/j.1468-3083.2011.04416.x

Varani, A. M., Siguier, P., Gourbeyre, E., Charneau, V., and Chandler, M. (2011). ISsaga is an ensemble of web-based methods for high throughput identification and semi-automatic annotation of insertion sequences in prokaryotic genomes. *Genome Biol.* 12:R30. doi: 10.1186/gb-2011-12-3-r30

Wassenaar, T. M., Ussery, D., Nielsen, L. N., and Ingmer, H. (2015). Review and phylogenetic analysis of qac genes that reduce susceptibility to quaternary ammonium compounds in *Staphylococcus* species. *Eur. J. Microbiol. Immunol.* 5, 44–61. doi: 10.1556/eujmi-d-14-00038

Yang, Y. H., Jiang, Y. L., Zhang, J., Wang, L., Bai, X. H., Zhang, S. J., et al. (2014). Structural insights into SraP-mediated *Staphylococcus aureus* adhesion to host cells. *PLoS Pathog* 10:e1004169. doi: 10.1371/journal.ppat.1004169

Yao, Y., Sturdevant, D. E., Villaruz, A., Xu, L., Gao, Q., and Otto, M. (2005). Factors characterizing *Staphylococcus epidermidis* invasiveness determined by comparative genomics. *Infect. Immun.* 73, 1856–1860. doi: 10.1128/IAI.73.3.1856-1860.2005

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Identification of surface proteins in a clinical *Staphylococcus haemolyticus* isolate by bacterial surface shaving

Runa Wolden[1], Maria Pain[1], Roger Karlsson[2, 3, 4], Anders Karlsson[2], Elizabeth G Aarag Fredheim[5], Jorunn Pauline Cavanagh[1, 6*]

[1]Pediatric Research group, Department of Clinical Medicine, Faculty of Health Sciences, UiT The Arctic University of Norway, Tromsø, Norway

[2]Nanoxis Consulting AB, Gothenburg, Sweden

[3]Department of Infectious Diseases, Institute of Biomedicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden

[4]Department of Clinical Microbiology, Sahlgrenska University Hospital, SE-413 46 Gothenburg, Region Västra Götaland, Sweden

[5]Microbial Pharmacology and Population Biology (MicroPop), Department of Pharmacy, Faculty of Health Sciences, UiT The Arctic University of Norway, Tromsø, Norway

[6]Department of Pediatrics, The University Hospital of North Norway, Tromsø, Norway

***Correspondence:**

Jorunn Pauline Cavanagh

Pauline.cavanagh@uit.no

Running title: *Staphylococcus haemolyticus* surface proteins

***Staphylococcus haemolyticus* surface proteins**

Abstract

Background

Staphylococcus haemolyticus is a skin commensal that has gained increased attention as an emerging nosocomial pathogen. Despite its clinical relevance, published information about *S. haemolyticus* virulence factors is scarce compared to literature published on other staphylococcal species. In this study, the adhesive and biofilm forming properties of ten clinical and ten commensal *S. haemolyticus* strains were examined using standard adhesion and biofilm assays. One of the clinical strains for bacterial surface shaving was used to identify expressed surface proteins. Protein abundance was examined by a comparative analysis between bacterial protein expression after human keratinocyte (HaCaT) colonization and growth in cell culture media supplemented with serum. Relative protein quantification was performed by labeling peptides with tandem mass tags (TMT) prior to Mass Spectrometry analysis.

Results

Adherence to fibronectin, collagen and plastic was low in all tested strains, but with significantly higher adhesion to fibronectin ($p = 0.041$) and collagen ($p = 0.001$) in the commensal strains compared to the clinical strains. There was a trend towards higher degree of biofilm formation in the clinical strains ($p = 0.059$).

By using the surface shaving approach, three hundred and twenty-five proteins were detected in total, of which 65 proteins were classified as surface proteins. The comparative analyses showed that the abundance of nineteen (5.8 %) proteins were significantly changed following HaCaT colonization. The bacterial Toll/interleukin-1 like (TIRs) domain containing protein (p=0.04), the transglycosylase SceD (p=0.01), and the bifunctional autolysin Atl (p=0.04) showed a 1.4, 1.6- and 1.5-fold increased abundance respectively. The staphylococcal secretory antigen (SsaA) (p=0.04) was significantly downregulated (-1.5 fold change) following HaCaT colonization.

Among the 65 surface proteins the elastin binding protein (Ebps), LPXAG and LPXSG domain containing proteins and five LPXTG domain containing proteins were identified; three Sdr-like proteins (Serine Aspartate repeat containing proteins), the extracellular matrix binding protein Embp, a SasH-like protein and two hypothetical proteins.

Conclusions

The identification of expressed proteins after host-microbe interaction offers a tool for the discovery and design of novel targets for antimicrobial treatment.

Keywords: *Staphylococcus haemolyticus*[1], surface protein[2], surface shaving[3], biofilm[4], adhesion[5], virulence[6], keratinocytes[7], host-microbe interaction[8]

1 Background

Staphylococcus haemolyticus is a coagulase-negative staphylococcus (CoNS) and a member of the skin microbiome. It is an increasing cause of nosocomial infections associated with indwelling medical devices, particularly affecting immunocompromised patients and premature babies [1-3]. A distinct characteristic of clinical *S. haemolyticus* strains is the ability to acquire resistance to several classes of antimicrobial agents [2]. The ability to colonize and form biofilms is regarded as the most important virulence trait for CoNS [4]. Adhesion is the first step to form biofilm on surfaces [5] and staphylococci express several adhesive surface molecules that interact with eukaryotic host cell receptors, abiotic surfaces or soluble macromolecules. The number of adhesive surface proteins varies among different staphylococcal species. In *Staphylococcus aureus*, 24 different cell wall anchored proteins have been identified, while CoNS express a smaller number [6]. Cell wall anchored (CWA) proteins are covalently attached to the peptidoglycan layer. The most prevalent CWA proteins are the microbial surface component recognizing adhesive matrix molecule (MSCRAMM) family. All CWA proteins contain an LPXTG motif (Leu-Pro-X-Thr-Gly; where X can be any amino acid) that anchor the protein to the cell wall [6]. The Sdr protein subfamily of MSCRAMMs contains a serine-aspartate repeat region [1, 6] and a signal peptide with an YSIRK motif. In *S. aureus* the majority (13/21) surface proteins harbors the YSIRK/GS signal sequence, allowing delivery of surface proteins to unique locations in the cell wall [7]. *Sdr*-like genes have previously been described in *S. haemolyticus* [8].

Another family of the CWA proteins is the Serine Rich Repeats Proteins family. Like the Sdr proteins, they have a serine repeat region, but with alanine, valine or threonine instead of aspartate [9]. Bacterial surface proteins can act as new targets in treatment and prevention of infections in multiresistant bacteria. One method to examine bacterial surface proteins is by surface shaving. Surface-shaving is a

technique where peptides from bacterial surface proteins are cleaved off when protease treatment is applied followed by a Liquid Chromatography tandem Mass Spectrometry (LC-MS/MS) analysis [10]. The Lipid-based Protein Immobilization (LPI^{TM}) technology enables surface shaving of intact bacterial cells in a flow cell, and thus promotes detection of proteins expressed in the surface proteome over the highly abundant cytosolic proteins. The flow cell channels, binds intact cells by a passive process. As the surface is similar in each channel, the same number of cells are bound. Thus, combining the surface shaving approach with protocols for relative quantification, such as tandem mass tags (TMT), makes studies of low abundant virulence factors possible [11-17].

Several studies on surface proteins and their relevance in host-pathogen interactions and virulence have been performed after bacterial growth in standard laboratory medium [18-21]. In order to mimic a more biological relevant host-microbe interaction, we developed a novel method to investigate expressed surface proteins of a clinical *S. haemolyticus* isolate after colonization of human keratinocytes (HaCaT) before bacterial surface shaving was performed (Figure 1). To our knowledge surface protein shaving of bacteria subsequent to colonization of mammalian skin cells has never been described before.

In this study, we aimed to investigate the adhesive and biofilm forming abilities of ten commensal and ten clinical strains. We have previously shown that there are specific genetic signatures associated with clinical *S. haemolyticus* strains compared to commensal strains [22], thus we wanted to investigate if any functional differences in adhesive properties between commensal and clinical isolates could be observed. Furthermore, the expression of surface-associated proteins of one clinical *S. haemolyticus* strain was investigated by mass spectrometry and proteomics. The LPI surface shaving approach and relative quantification proteomics using TMT labels was employed to identify possible novel targets for treatment, prevention and biofilm formation.

2 Results

We wanted to examine if commensal and clinical strains had different ability to interact and adhere to selected host proteins. The adhesive ability of ten commensal and ten clinical strains to both uncoated plastic and plastic coated with fibronectin and collagen was examined to determine if binding to fibronectin or collagen would enhance binding to plastic, as we observed that binding to plastic in its native form was generally low. Further the biofilm forming capacity was examined.

Staphylococcus haemolyticus surface proteins

2.1 Adhesion to plastic and host matrix proteins

Both clinical and commensal strains adhered to plastic but no significant difference was observed between the two groups. Fibronectin and collagen binding were low for all strains, but still significantly higher for the commensal strains compared to clinical strains, p=0.041 and p=0.001 respectively (Figure 3A-C).

2.2 Semi-quantitative determination of biofilm formation

The biofilm-forming ability of the strains was determined using a semi-quantitative assay. All strains formed biofilms and a trend towards higher biofilm formation was observed for the clinical strains (p=0.059) where 5/10 clinical strains formed substantial amounts of biofilm in this assay ($OD_{570} \geq 3$) compared to 0/10 commensal strains (Figure 3D).

2.3 Adhesion to human keratinocytes

The strains were screened for their ability to adhere to human keratinocytes. In three clinical and one commensal strain >60% of the inoculum adhered to the keratinocytes, while seven strains showed an adhesion of ~10-20% of the inoculum, which was in the same range as the *S. aureus* (NCTC 8325-4) control strain (Figure 3E). On average, the clinical strains adhered better to the keratinocytes compared to the commensal strains, although the findings were not statistically significant (p = 0.4). One strain, displaying high adhesion to HaCaT cells in addition to being a strong biofilm producer, was chosen for further analyzes.

2.4 Bacterial surface protein shaving

Expressed surface proteins of a clinical *S. haemolyticus* isolate either colonizing HaCaT cells or grown in cell culture medium supplemented with serum, was examined by surface shaving using a Lipid-

based Protein Immobilization flow cell. Relative quantification of protein abundance was performed by labelling proteins with tandem mass tags (protein markers) prior to LC-MS/MS.

2.4.1 Protein identification and subcellular localization of *S. haemolyticus* proteins detected by surface shaving

Cell surface shaving of bacteria colonizing HaCaT cells or incubated in cell culture media supplemented with serum resulted in identification of 436 proteins by LC-MS/MS analysis. Only proteins with ≥ #2 peptide-spectrum matches (PSMs) were included for further analysis, resulting in 325 proteins (Supplementary Table 3 and 4).

Subcellular localization analysis of the 325 proteins *in silico* and functional annotation predicted 249/325 (76.6%) cytoplasmic proteins, 65/325 (20.0%) surface proteins (i.e. proteins predicted to originate from the cytoplasmic membrane, cell wall or extracellular origin), and 11/325 (3.4%) as undefined proteins.

2.4.2 Clusters of Orthologous Groups

The 325 identified proteins were distributed in Clusters of Orthologous Groups (COG). A higher percentage of proteins in COG groups J (translation, ribosomal structure and biogenesis), D (cell cycle control, cell division and chromosomal partitioning), M (cell wall/membrane/envelope biogenesis), O (post translational modification, protein turnover and chaperones), C (energy production and conversion), F (nucleotide transport and metabolism) and G (Carbohydrate transport and Metabolism) was found when we compared the COG distribution of the identified proteins to the COG distribution of the total number of predicted proteins (2,539) encoded in the *S. haemolyticus* genome (Figure 4).

2.4.3 *S. haemolyticus* surface proteins

Characteristic motifs of surface proteins such as signal peptides and LPXTG motifs were identified by bioinformatic tools. The covalently anchored cell wall proteins classified as MSCRAMMs are

characterized by the C-terminal LPXTG sorting signal. In total 19 proteins were predicted to have LPXTG motifs based on *in silico* analysis of the whole genome sequence of *S. haemolyticus* 53-38, of these seven were annotated as adhesion proteins, four were hypothetical proteins and two were DUF 402 and 368.

Of the 325 proteins identified after surface shaving, 65 were annotated as surface proteins (Table 2). Three of the LPXTG proteins identified as adhesins by the *in silico* analysis were expressed on the *S. haemolyticus* surface. Five LPXTG, one LPXSG and one LPXAG domain containing surface proteins were identified. Three Serin-Aspartate-Repeat (Sdr-like) proteins, the extracellular matrix binding protein (Embp), one Mannosylglucosyl-3-phosphoglycerate phosphatase (SasH-like), and two uncharacterized surface proteins were identified. Other well characterized proteins identified surface proteins were the lytic transglycosylase immunodominant staphylococcal antigen A (IsaA), the Immunodominant staphylococcal antigen B (IsaB) and the elastin binding protein (EbpS).

2.4.4 HaCaT colonisation causes changes in abundance of proteins

We wanted to explore if protein abundance differed when *S. haemolyticus* colonized HaCaT cells compared to when grown in cell culture media supplemented with serum. The large majority of proteins were found similarly abundant when comparing the two conditions, this included EbpS, IsaB and cytoplasmic proteins (Supplementary Table 3).

Only nineteen of 325 proteins (5.8%) showed a significant change in abundance (≥± 1.2 fold change) following HaCaT colonization (Table 3). The lytic transglycosylase *Staphylococcus epidermidis* D protein (SceD) (p=0.01) and the autolysin Atl (p=0.04) showed significantly increased abundance with a fold increase of 1.6 and 1.5 respectively when *S. haemolyticus* colonized keratinocytes. The Toll/interleukin-1 like (TIRs) domain protein (p=0.04) also had an increase in abundance (1.4-fold) after HaCaT co-incubation, while the Staphylococcal secretory antigen (SsaA) was significantly (p=0.04) less abundant following keratinocyte colonization, showing a 1.5-fold reduced abundance.

2.4.5 Moonlighting proteins identified by surface shaving

Several proteins that have previously been shown in other bacteria to have moonlighting functions - proteins dually engaged intracellularly and with important adhesive functions extracellularly - were found among the predicted cytoplasmic proteins. These are the moonlighting proteins glyceraldehyde-3- phosphate dehydrogenase (GAPDH), [23-25], enolase [26], aldolase (ALDA) [25], triose phosphate isomerase (TPI) [27], fructose-bisphosphate aldolase (FBA) [28], ornithine carbamoyl transferase (ARGF) [29], pyruvate kinase (PYK) [30], Inosine 5′-monophosphate dehydrogenase (IMPDH) [31], Clp [32], DNaK [33] and (Atl) [34].

3 Discussion

The ability to adhere to and colonize implanted biomaterials in addition to biofilm formation is considered the main virulence factors of *S. haemolyticus* and other coagulase-negative staphylococci

[1-3]. Despite the clinical relevance of *S. haemolyticus*, published information about virulence factors is scarce compared to literature published on other staphylococcal species. We recently published a comparative analysis of clinical and commensal *S. haemolyticus* isolates [22]. We identified distinct differences in the population structure, where the clinical isolates clustered together separately from the commensal isolates. Clinical isolates were more antibiotic resistant and had different versions of genes encoding surface proteins. [22]. In this study adhesive properties and biofilm formation was compared between clinical and commensal isolates, while the expressed surface proteins were characterized in one clinical isolate after keratinocyte colonization or incubation in cell culture medium supplemented with serum.

3.1 Solid phase host matrix protein binding assay

We found that both fibronectin and collagen binding was low for all *S. haemolyticus* strains. However, fibronectin and collagen binding were significantly higher for commensal compared to the clinical strains. Fibronectin is a glycoprotein found in substantial amounts in blood and loose connective tissue [35] while collagen is an abundant class of proteins in humans, offering structural support to connective tissues and the extracellular matrix [36]. In *S. aureus*, fibronectin binding is described as a crucial step

in host cell adhesion. Adhesion mainly involves binding by bacterial fibronectin binding proteins (FNBPs) to fibronectin which forms a bridge between $(\alpha_5)\beta_1$ integrin on mammalian cells [37]. Low fibronectin binding in *S. haemolyticus* was previously shown when compared to *S. aureus* [38], while a varying capacity of fibronectin binding in clinical *S. haemolyticus* and other CoNS was demonstrated by Switalski *et al*. [39]. FnBPA and FnBPB involved in *S. aureus* fibronectin binding have not been identified in CoNS so far, but fibronectin binding by the extracellular matrix binding protein (Embp) has been shown in *S. epidermidis*. Expression of embp in *S. epidermidis* was shown to be induced by supplementation of serum in the growth media [40]. Embp mediates adhesion to fibronectin and biofilm accumulation in *S. epidermidis* [41], and is present in 90% of clinical *S. epidermidis* strains [42]. Cell culture media supplemented with serum was also used in the adhesion assays in this study, where low binding was observed for all strains tested. We identified Embp on the surface of *S. haemolyticus* in the presence of serum. However, if Embp mediates fibronectin binding in *S. haemolyticus*, this did not result in good fibronectin binding in the adhesion assay in this study. Our findings reflect that the role of Embp in fibronectin binding of *S. haemolyticus* needs to be further investigated.

Cooperative binding of collagen in the presence of vitronectin has previously been demonstrated for *S. haemolyticus* [43]. Paulsson *et al*. used different bacterial growth media to induce optimal binding to both collagen and vitronectin. Thus, the type of media used in our experiments might not have been optimal for expression of proteins conferring collagen and fibronectin binding, which also could explain the low binding capacity observed in our experiments.

3.2 Adherence to plastic and semi-quantitative determination of biofilm formation

When we examined the ability to form biofilm we found trends towards more biofilm formation in the clinical strains compared to the commensal strains. However, all strains had the ability to form biofilm. In *S. epidermidis,* similar biofilm forming abilities were observed for both clinical and commensal strains, despite differences in population structure. Rather, different biofilm morphotypes and biofilm encoding genes were found among distinct genetic lineages indicating that biofilm formation is an important property of both commensal and clinical strains [44, 45].

***Staphylococcus haemolyticus* surface proteins**

We did not find any correlation between adherence to plastic and the degree of biofilm formation. As adherence is the first step in biofilm formation, one could expect an observed correlation between adhesion to plastic and biofilm formation. The discrepancy in these results can be explained by the use of different media when performing the two assays. It has previously been shown that the amount of biofilm varies depending on the media [46], making comparisons of results from different methods difficult.

3.3 Adhesion to human keratinocytes and bacterial surface protein shaving

We found a trend towards higher adhesion to keratinocytes for the clinical strains compared to the commensal strains. We selected one clinical strain with good adhesive and biofilm forming properties, and performed bacterial surface shaving after *S. haemolyticus* colonized with keratinocytes, in order to identify surface proteins important for adhesion to mammalian cells.

To date, most surface protein expression analyses are performed on bacteria incubated in bacterial growth medium [18-21]. As *S. haemolyticus* constitute a significant proportion of the skin microbiota of humans [1, 47, 48], we decided to choose a more biological relevant condition to study protein expression; incubation of *S. haemolyticus* with keratinocytes prior to bacterial surface shaving. Abundance of proteins following keratinocyte colonization was compared to protein abundance following growth in cell culture medium supplemented with bovine serum.

We identified 65 surface proteins in total, of which SceD and Atl were significantly more abundant when *S. haemolyticus* was colonizing keratinocytes. Transglycosylases cleave the β-1,4 glycosidic bond between *N*-acetylmuramic acid and *N*-acetylglucosamine residues of peptidoglycan, accompanied with formation of 1,6-anhydromuramic acid residues [49]. In *S. aureus* the transglycosylases SceD and IsaA are well described virulence factors involved in cell wall remodeling, contributing to resistance to antimicrobial peptides, adhesion and pathogenicity, shown in a murine septic arthritis model [50]. SceD has also been shown to have a pronounced upregulation upon nasal colonization of humans and rats [50, 51].

Biofilm formation is an important virulence factor in *S. haemolyticus*, and in this study we showed a trend towards stronger biofilm formation in clinical *S. haemolyticus* isolates. The bifunctional autolysin Atl was significantly more abundant in *S. haemolyticus* colonizing HaCaT cells. Atl homologs are

described in several staphylococcal species [1]. In *S. epidermidis* and *S. aureus*, Atl is important for initial adhesion and biofilm formation [52], and has in *S. epidermidis* been demonstrated to mediate adhesion to vitronectin [53]. In *S. aureus* IsaA is involved in biofilm formation and *isaA* mutants form significantly less biofilm [54]. In this study we identified IsaA when *S. haemolyticus* was grown in the presence of serum. The *S. haemolyticus* biofilm is mainly composed of environmental DNA (eDNA) and proteins [46]. As Atl also mediates adhesion indirectly by hydrolysis of the bacterial cell wall causing the release of proteins and eDNA [1], it is likely that Atl and IsaA expression also in *S. haemolyticus* have similar functions as observed in *S. epidermidis* and *S. aureus* in both adhesion and biofilm formation.

Mannosylglucosyl-3-phosphoglycerate phosphatase was one of the identified surface proteins, and has homologies to SasH in *S. aureus*. SasH has in *S. aureus* been shown to be positively correlated with invasive disease isolates, and is found in serum from convalescent patient´s sera [55]. If the SasH-like protein has a role in infection in *S. haemolyticus*, needs to be further elucidated.

In silico analysis of the genome sequence of the clinical *S. haemolyticus* isolate used for HaCaT colonization identified 19 LPXTG containing genes. Seven of these genes were annotated as genes encoding proteins involved in adhesion, while six had unknown function. These findings resemble what is found in *S. aureus*, where 21 LPXTG genes were predicted *in silico*, of which eleven had unknown function [56]. In this study, five LPXTG and two LPXSG, LPXAG containing proteins were identified after surface shaving. We identified three Sdr-like proteins which were expressed both when *S. haemolyticus* were co-incubated with HaCaT cells, and when grown in media containing serum. In *S. aureus,* transcription of SdrD and SdrG is increased in the presence of blood and serum [57, 58]. As both tested conditions contained media supplemented with serum, this could explain the expression of the Sdr-like proteins under both conditions.

In *S. epidermidis,* three Sdr proteins have been identified; SdrF, SdrG (Fbe) and SdrH. SdrF has been shown to mediate strong binding to keratins, keratinocytes and nasal epithelial cells [59]. In *S. aureus,* SdrD has been shown to mediate adhesion to keratinocytes through binding to desmoglein1, expressed in human epidermis [60]. The expression of Sdr-like proteins in S. haemolyticus after HaCaT colonization and grown in the presence of serum suggests that it might exert similar functions in keratinocyte binding, as found in *S. epidermidis* and *S. aureus*.

***Staphylococcus haemolyticus* surface proteins**

HaCaT colonization resulted in the significant upregulation of a TIR protein. TIR domain containing proteins have been shown in several pathogenic bacteria [61], but has not previously been described in *S. haemolyticus*. TirS in *S. aureus* increases survival in the host by blocking the cascade reaction leading to activation of the nuclear factor–κB (NF-κB), which regulates the expression of a pro-inflammatory immune response [62]. Bacterial circumvention of the host immune defense is an important mechanism in bacterial host colonization.

The secretory antigen SsaA was significantly less abundant when bacteria were co-incubated with HaCaT cells. SsaA is a highly antigenic protein identified in sera from patients infected with *S. aureus* [63], and was shown to be secreted during infective endocarditis infection in *S. epidermidis* [64]. In the same study, anti SsaA Immunoglobulin G (IgG) antibody in serum from *S. haemolyticus* infections was not found, however only one sample was tested. The finding of a reduced abundance of this protein during HaCaT colonization, could indicate that it is expressed during infection and not colonization. Further studies should be conducted with serum samples from *S. haemolyticus* infections in order to examine the antigenic capacity of SsaA in *S. haemolyticus.*

3.3.1 Cytoplasmic proteins

Several of the proteins identified in this experiment were predicted as cytoplasmic proteins. Detection of some cytoplasmic proteins are inevitable when performing surface shaving [10, 65]. The presence of predicted cytoplasmic proteins after bacterial surface shaving can be due to cellular lysis, moonlighting proteins or protein containing membrane-vesicles (MV) [10, 65].

We recently showed that *S. haemolyticus* produces MVs [66]. The *S. haemolyticus* MV cargo mainly contained cytoplasmic proteins, amongst them several moonlighting proteins, which are proteins that express more than one function when transported to a different cellular location [23]. Release of MVs in incubation buffer after culturing and washing of cells might add to the identification of predicted cytoplasmic proteins [10]. The moonlighting proteins GAPDH, enolase and triose phosphate isomerase were identified in this study, and have demonstrated binding to plasminogen in *S. aureus* and *S. epidermidis* [26, 31, 67]. Ornithine carbamoyltransferase has been shown to contribute to fibronectin binding in *S. epidermidis*, when localized on the bacterial cell surface [29], thus these proteins might also exert similar adhesive functions in *S. haemolyticus.*

3.3.2 Strengths and limitations of the study

The main advantage of the developed method is the direct contact between bacteria and mammalian cells before bacterial surface shaving, mimicking a more relevant host-microbe interaction compared to other protein expression systems. *S. haemolyticus* surface shaving subsequent to colonization of human keratinocytes has to our knowledge not been described before. By using the LPI™ approach for bacterial surface shaving, whole cells are immobilized by a passive process (personal communication Nanoxis Consulting AB) within a flow cell prior to digestion, allowing binding of intact cells only. In this study we only used one clinical isolate. In order to find surface proteins that are present only in clinical vs. commensal isolates, several isolates from different commensal and clinical lineages need to be compared.

The separation of bacteria from the mammalian cells by FACS is time consuming, leading to a low throughput of samples. The individual sorting of samples before being concentrated and subsequently subjected to surface shaving in individual LPI flow cell channels, might have led to slight variations in the concentration of cells or even slight differences in expression due to slight differences in handling time. However, we kept all samples on ice and in PBS throughout the experiment in order to minimize potential alteration of gene expression.

4 Conclusion

This is to our knowledge the first described study using surface shaving of expressed staphylococcal proteins after direct contact with eukaryotic cells and in cell culture media supplemented with serum. Gaining information about surface exposed proteins is important in order to better understand host-pathogen interactions, biofilm formation and for the discovery and design of novel targets for antimicrobial and anti-biofilm treatment. Thus, this method is transferable to other bacterial species and mammalian cell types. The method has provided novel knowledge about the *S. haemolyticus* surface proteins in a clinical isolate. We have identified surface proteins and immune evasive proteins previously only functionally described in other staphylococcal species. We have also identified hypothetical surface proteins, of which future analysis should be undertaken in order to describe function. Further functional assays should be performed to determine the importance of the different identified proteins in host microbe interactions and biofilm formation.

5 Methods

5.1 Bacterial strains and mammalian cell lines

Ten clinical and ten commensal *S. haemolyticus* strains were included in the study (Table 1). The clinical strains are a subset of a larger collection, isolated from blood, catheters and wounds [2]. The commensal strains are a subset of a collection of strains from the skin of healthy adults [48]. HaCaT cells were from a human keratinocyte cell line [68] (Cell Lines Service (CLS), Germany, no. 300493).

5.2 Solid phase host matrix protein binding assay

The ability of *S. haemolyticus* to adhere to collagen, fibronectin and plastic was determined using a protocol based on Edwards *et al.* [69]. Bacterial cultures were grown for 10 hours (Optical density $(OD)_{600}$ 0.7-1.0) in Dulbecco's Modified Eagle's Medium (DMEM) (Merck, Germany) with 10% heat inactivated Fetal Bovine serum (FBS) (Thermo Fisher Scientific, MA, USA), pelleted and re-suspended to a concentration of 10^8 colony forming units (CFU)/mL. Microtiterplates (96 well) pre-coated with collagen (Thermo Fisher Scientific, MA, USA) or fibronectin, 1 μg/well (R&D Systems, MN, USA) were blocked with 150 μl 3% Bovine Serum Albumin (BSA) (Merck, Germany) for 1 h at room temperature and then washed 2x with Phosphate Buffered Saline **(**PBS) (Merck, Germany). Inoculum was added to plastic (CAT.NO 163320, Thermo Fisher Scientific, MA, USA), collagen and fibronectin plates and incubated for 1 h at 37 °C followed by 1x wash with PBS. The plates were fixed at 55 ºC for 1 h and stained with 0.25% crystal violet (Merck, Germany) for five minutes. Biomass of adherent bacteria was determined by solubilization of crystal violet with 150 μL 70% EtOH. Absorbance (Abs) was measured at 590 nm (Versamax, Molecular Devices, CA, USA). Values from bacterial binding to wells coated with BSA only were subtracted.

5.3 Semi-quantitative determination of biofilm formation

We performed semi-quantitative determination of biofilm production as described previously [46, 70]. Biofilm formation was induced in Tryptic Soy Broth (TSB) (BD, NJ, USA / Merck, Germany) with 1% glucose (Merck, Germany) in 96-well microtiter plates (Thermo Fisher Scientific, MA, USA). All

strains were tested in eight wells with three parallel runs and controls were included on each plate. After 24 h, wells were washed, fixed and stained with 0.1% crystal violet (Merck, Germany). Crystal violet was dissolved from the biofilm with 70% ethanol for 10 min and Abs_{570} was determined (Versamax, Molecular Devices, CA, USA). We removed the highest and lowest outlier for each parallel and the remaining six values were averaged. Based on the distribution of the tested strains, strains with average OD values over 1 were considered strong biofilm-producers.

5.4 Adhesion to human keratinocytes

S. haemolyticus adhesion to human keratinocytes (HaCaT) was determined. HaCaT ($2x10^5$ cells/ml) were added to 24-well plates (Thermo Fisher Scientific, MA, USA) and allowed to attach for 16 h (37°C, 5% CO_2) in DMEM with 10% FBS. Bacterial cultures were grown at 37 °C to late exponential phase (OD_{600} 0.7-1.0) in DMEM with 10% FBS, and then washed twice in Dulbecco´s Phosphate Buffered Saline (DPBS) (Merck, Germany). Approximately 2×10^6 CFU in DMEM with 10% FBS were added to each well of a cell culture plate to achieve a multiplicity of infection dose (MOI) of 10:1. The plates were centrifuged at 900xG (Eppendorf 5430R, Germany) for 10 min at 37 °C and incubated for 30 min. at 37 °C in 5% CO_2 [71]. After incubation, the plates were thoroughly washed to remove all unbound bacterial cells. To enumerate the number of adhered bacteria, 0.25 mg/mL Trypsin-EDTA (Merck, Germany) and 0.1% mg/mL Triton X-100 (Merck, Germany) were added, and the suspension was pipetted in order to fully lyse the HaCaT cells. CFU/mL was determined by plating on blood agar plates (Thermo Fisher Scientific, MA, USA) and incubated at 37 °C overnight. Three biological replicates were performed.

5.5 Bacterial surface protein shaving

5.5.1 Preparation of bacteria for cell surface shaving

To explore the expression of surface proteins in *S. haemolyticus* when colonizing HaCaT cells, one clinical bacterial strain (53-38) with strong adhesive and biofilm-forming properties (Table 1) was co-incubated with HaCaT cells. We wanted to further explore this isolate as adhesion and biofilm formation is regarded as important virulence traits in the coagulase negative staphylococci. A bacterial

control sample (same bacterial isolate) grown in cell culture media supplemented with serum but without HaCaT cultivation was included. Three biological replicates were performed for all samples and both conditions. The workflow of the bacterial surface shaving experiment is summarized in Figure 1 and Supplementary Table 1.

HaCaT cells were seeded in 6-well plates, and bacterial cultures were grown to late exponential phase (OD_{600} 0.6±0.1) in DMEM with 10% FBS, washed twice in DPBS and resuspended in DMEM with 10% FBS and further handled as previously described for the HaCaT adhesion assay. A MOI of 100:1 was used and bacteria were centrifuged with HaCaT cells for 10 min, and further incubated for 50 min. After incubation, tissue culture plates were washed 4 times with DPBS to remove free-floating bacteria. Mechanical detachment of eukaryotic and bacterial cells from the tissue culture plates was performed with a cell scraper (VWR, PA, USA) followed by pipetting in DPBS. Cells were transferred to polystyrene tubes with a cell strainer cap (Thermo Fisher Scientific, MA, USA). Twelve wells from two tissue culture plates were used for each replicate. The samples were prepared for Fluorescence-activated cell sorting (FACS), in order to separate bacteria from HaCaT cells, by labelling with the Vancomycin BODIPY™ FL Conjugate (Thermo Fisher Scientific, MA, USA) (0.6 µg/mL), targeting the Gram-positive bacterial cell wall [72].

The bacterial control samples that were not co-cultivated with HaCaT cells were grown to late exponential phase in DMEM with 10% FBS (OD_{600} 0.6±0.1) and resuspended in DPBS after centrifugation and washing and further stored on ice. Samples were then prepared for FACS by Vancomycin BODIPY ™ labelling, in order to treat the bacterial control samples in a similar manner to the test samples.

5.5.2 Fluorescence-activated cell sorting system (FACS)

S. haemolyticus was sorted from HaCaT cells by using FACS Aria III (BD, NJ, USA) (Software BD FACSDiva 8.0.1). Fluorescent beads (Polystyrene Particle, Flow Cytometry grade PPS-6K and Nano Blank Polystyrene NFPPS-52-4K (Spherotech, IL, USA)) were used for calibration. Vancomycin BODIPY™ was excited with a 488 nm blue laser. A FITC-detector was used to read the emitted green, fluorescent light. Normal density filter 1.0 was used in front of the FSC detector. After FACS all samples were stored on ice.

5.5.3 Surface shaving - Sample processing and generation of peptides by LPI™ HexaLane flow cell

In order to concentrate the bacterial samples after FACS (≈230 mL), samples were centrifuged twice, both steps at 10000xG for 30 min at 4 °C in swing bucket rotors (Beckman Coulter, CA, USA), resulting in samples containing approximately 2.8 x 10^7 CFU/mL. The concentrated samples were resuspended in ice cold PBS, kept on ice and immediately loaded into the LPI™ HexaLane Flow Cell (Nanoxis Consulting AB, Sweden), as seen in Figure 2, step 1. To allow bacterial attachment, the flow cell was incubated for 35 min at room temperature. The cells attach to the gold coated channels in the Flow Cell by a passive process (personal communication Nanoxis Consulting AB). Unbound bacteria were removed by washing the channels with 200 µL PBS using a syringe pump (Harvard Apparatus, MA, USA) at a flow rate of 50 µL/min. Enzymatic digestion of bacterial surface proteins was performed by injecting 100 µL of trypsin (Promega, WI, USA) (40 µg/mL) into the LPI HexaLane Flow Cell channels and further incubated for 20 min at room temperature. After digestion, peptides were eluted in 200 µL PBS and the digestion was terminated by adding 4 µL formic acid (neat) (Merck, Germany). The peptide samples were centrifuged for 10 min at 10000xG, in order to remove any cell debris and the supernatants were subsequently dried using a SpeedVac (Eppendorf, Germany) and stored at -20 °C.

5.5.4 Protein identification and relative quantitation

The proteomic analysis was performed at The Proteomics Core Facility at Sahlgrenska Academy, Gothenburg University. Digested peptides were dissolved in 100 µL triethylammonium Bicarbonate (TEAB) (350 mM, Thermo Fisher Scientific, MA, USA) and labelled using TMT 10-plex isobaric mass tagging reagents (Thermo Fisher Scientific, MA, USA) according to the manufacturer's instructions. The TMT-set were fractionated into twelve fractions using Pierce High pH Reversed-Phase Peptide Fractionation Kit (Thermo Fisher Scientific, MA, USA) according to the manufacturer's protocol, but with a modified gradient (Supplementary Table 2).

The fractions were analyzed on a QExactive HF mass spectrometer (MS) interfaced with Easy-nLC1200 liquid chromatography system (LC-MS/MS) (Thermo Fisher Scientific, MA, USA). Peptides were trapped on an Acclaim Pepmap 100 C18 trap column (100 µm x 2 cm, particle size 5 µm, Thermo

Fisher Scientific, MA, USA) and separated on an in-house packed analytical column (75 μm x 300 mm, particle size 3 μm, Reprosil-Pur C18, Dr. Maisch, Germany) using a gradient from 7% to 35% B over 70 min followed by an increase to 100% B for 5 min at a flow of 300 nL/min. Solvent A was 0.2% formic acid and solvent B was 80% acetonitrile, 0.2% formic acid. The instrument operated in data-dependent mode where the precursor ion mass spectra were acquired at a resolution of 60 000, the 10 most intense ions were isolated in a 0.8 Da isolation window and fragmented using collision energy HCD settings at either 28 or 50. MS2 spectra were recorded at a resolution of 60 000, charge states 2 to 4 were selected for fragmentation and dynamic exclusion was set to 20 s with 10 ppm tolerance.

MS raw data files for the TMT set were merged for identification and relative quantification using Proteome Discoverer version 1.4 (Thermo Fisher Scientific, MA, USA). *S. haemolyticus* 53-38 with European Nucleotide Archive (ENA) accession number GCA_001226325.1 (Illumina sequence) and ENA accession number PRJEB36042 (PacBio sequence) [2] were aligned using BWA-MEM [73] and further used as reference proteome (2539 coding sequences). Structural and functional annotations were performed using Prokka [74]. Mascot 2.5 (Matrix Science Ltd., UK) was used as a search engine with precursor mass tolerance of 5 ppm and fragment mass tolerance of 200 mmu. Tryptic peptides were accepted with one missed cleavage and variable modifications of methionine oxidation, cysteine alkylation and fixed modifications of N-terminal TMT-label and lysine TMT-label were selected. Fixed Value of 13 was used for identification and the quantified proteins were filtered at 1% False Discovery Rate (FDR) resulting in a mascot score of at least 20. No missing values were present in the data set at Threshold of 2000. Proteins were grouped by sharing the same sequences to minimize redundancy. The resulting ratios were normalized in the Proteome Discoverer 1.4 and the sum of the samples cultivated with HaCaT was used as denominator. Only unique peptides were used for comparison between groups.

The mass spectrometry proteomics data has been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifiers PXD014450.

5.6 Statistical analyses

For the results from biofilm-, solid phase host matrix protein and HaCaT adhesion assays the data were analyzed using IBM SPSS software, version 25.0. The non-parametric Mann-Whitney U-test was used to compare two groups, a p value < 0.05 was considered statistically significant.

As the technical variation for the identified proteins was assumed to be 20%, only proteins displaying a higher degree of fold change (FC) than ±1.2 were considered as biologically significant regarding increased or reduced abundance of proteins. The most changed abundance of proteins had a threshold of at least ±1.5. Welch's t-test was performed (3 parallels vs. 3 parallels) and only proteins passing filter $p < 0.05$ were considered statistically significant.

5.7 Bioinformatic analyses

LPXTG motifs were predicted *in silico* from the whole genome sequence of *S. haemolyticus* 53-38 using a manual sequence search. Prediction of the subcellular localization of proteins was done using PSORTb v.3.0 algorithms [75], CELLO v.2.5 [76] and LocateP v.2.0 [77]. Positive prediction of subcellular localization was determined by a two out of three or greater concurrent results between the databases. Surface proteins were defined as proteins predicted from cytoplasmic membrane, cell wall or extracellular origin.

Functional annotation of proteins was done with the EggNOG v.5.0 database with HMMER and Diamond mapping mode; i.e. functional description, seed orthologues, predicted name, KEGG KO and categorization of proteins into Clusters of Orthologous Groups of proteins (COG) [78], PHMMER v.3.3 [79, 80] and protein BLAST [81].

Moonlighting proteins were identified by using the MoonProt database and by manual searches based on published literature [82, 83].

6 Abbreviations

Abs - Absorbance

ALDA - Aldolase

ARGF - Ornithine carbamoyl transferase

Atl - Autolysin

BSA - Bovine Serum Albumin

CFU - Colony Forming Units

CLS - Cell Lines Service

COG - Clusters of Orthologous Groups

CoNS - Coagulase Negative Staphylococci

CWA - Cell Wall Anchored

DMEM - Dulbecco's Modified Eagle's Medium

DPBS - Dulbecco´s Phosphate Buffered Saline

DUF - Domain of Unknown Function

Ebps - Elastin binding protein

eDNA - Extracellular DNA

Embp - Extracellular matrix binding protein

ENA - European Nucleotide Archive

FACS - Fluorescence-activated cell sorting

FC - Fold Change

FBA - Fructose-Bisphosphate Aldolase

FBS - Fetal Bovine Serum

FDR - False Discovery Rate

FNBPs - Bacterial Fibronectin binding proteins

GAPDH - Glyceraldehyde-3- phosphate dehydrogenase

HaCaT - Human Keratinocytes

IgG - Immunoglobulin G

***Staphylococcus haemolyticus* surface proteins**

IMPDH - Inosine 5′-monophosphate dehydrogenase

IsaA - Immunodominant staphylococcal antigen A

IsaB - Immunodominant staphylococcal antigen B

LC-MS/MS - Liquid Chromatography tandem Mass Spectrometry

LPI - Lipid-based Protein Immobilization

LPXTG - Leu-Pro-X-Thr-Gly; where X can be any amino acid

MOI - Multiplicity Of Infection

MS - Mass Spectrometry

MSCRAMM - Microbial Surface Component Recognizing Adhesive Matrix Molecule

MV - Membrane Vesicle

NF-κB - Nuclear Factor–κB

OD - Optical Density

PBS - Phosphate Buffered Saline

PRIDE - PRoteomics IDEntifications database

PSM - Peptide Spectrum match

PYK - PYruvate Kinase

SasH - Mannosylglucosyl-3-phosphoglycerate phosphatase

SceD - Lytic transglycosylase *Staphylococcus epidermidis* D protein

Sdr - Serine Aspartate repeat containing protein

SsaA - Staphylococcal secretory antigen

TEAB - Triethylammonium Bicarbonate

TIR - Toll/interleukin-1 receptor

TMT - Tandem Mass Tags

TPI - Triose Phosphate Isomerase

TSB - Tryptic Soy Broth

7 Declarations

7.1 Ethics approval and consent to participate

Not applicable

7.2 Consent for publication

Not applicable

7.3 Availability of data and materials

The datasets supporting the conclusions of this article are available in the ProteomeXchange Consortium via the PRIDE partner repository [84] with the dataset identifiers PXD014450, the European Nucleotide Archive with the unique identifier ERS066380 [85] and PRJEB36042 [86]. The dataset supporting the conclusions of this article is included within the article (and its additional file S1-4).

7.4 Competing interests

The authors declare that the submitted work was not carried out in the presence of any personal, professional or financial relationships that could potentially be construed as a conflict of interest.
Authors AK and RK are affiliated to a company, Nanoxis Consulting AB. The Company did not have influence on the collection, analysis, or interpretation of data, the writing of the paper, or the decision to submit for publication.

7.5 Funding

The study was supported by grants from the Northern Norway Regional Health Authority (HNF1344-17). The publication charges for this article have been funded by a grant from the publication fund of UiT The Arctic University of Norway. The funders were not involved in the design, data collection or interpretation of the results obtained in this study.

7.6 Authors` contributions

RW: Participated in experimental design and planning. Performed the surface shaving experiments, analyzed the results and wrote the first draft of the manuscript.
MP: Participated in experimental design and planning. Performed the adhesion experiments, analyzed the results and participated in writing the first draft of the manuscript.
RK: Participated in experimental design and performance of the cell surface shaving experiment. Performed the mass spectrometry analysis and read through the final version of the manuscript.
AK: Participated in experimental design and performance of the cell surface shaving experiment. Performed the mass spectrometry analysis and read through the final version of the manuscript.
EF: Participated in experimental design and planning. Performed the biofilm and adhesion experiments, analyzed the results and participated in writing the first draft of the manuscript.
JPC: Conceptualized the experimental design. Participated in the surface shaving and adhesion experiments, analysis of results and in writing the manuscript.

All authors have read and approved the manuscript.

7.7 Acknowledgements

We would like to acknowledge Annika Thorsell at The Proteomics Core Facility at Sahlgrenska Academy, Gothenburg University, for performing the proteomic analysis. We would also like to acknowledge Roy André Lyså at The Bio-imaging Platform at UiT, The Arctic University of Norway, for performing the cell sorting.

8 References

1. Becker K, Heilmann C, Peters G. Coagulase-negative staphylococci. Clinical microbiology reviews. 2014;27(4):870-926.DOI: 10.1128/cmr.00109-13.

2. Cavanagh JP, Hjerde E, Holden MT, Kahlke T, Klingenberg C, Flaegstad T, et al. Whole-genome sequencing reveals clonal expansion of multiresistant Staphylococcus haemolyticus in European hospitals. The Journal of antimicrobial chemotherapy. 2014;69(11):2920-7.DOI: 10.1093/jac/dku271.

3. Czekaj T, Ciszewski M, Szewczyk EM. Staphylococcus haemolyticus - an emerging threat in the twilight of the antibiotics age. Microbiology (Reading, England). 2015;161(11):2061-8.DOI: 10.1099/mic.0.000178.

4. Heilmann C, Ziebuhr W, Becker K. Are coagulase-negative staphylococci virulent? Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases. 2018.DOI: 10.1016/j.cmi.2018.11.012.

5. Kline KA, Falker S, Dahlberg S, Normark S, Henriques-Normark B. Bacterial adhesins in host-microbe interactions. Cell host & microbe. 2009;5(6):580-92.DOI: 10.1016/j.chom.2009.05.011.

6. Foster TJ, Geoghegan JA, Ganesh VK, Hook M. Adhesion, invasion and evasion: the many functions of the surface proteins of Staphylococcus aureus. Nature reviews Microbiology. 2014;12(1):49-62.DOI: 10.1038/nrmicro3161.

7. DeDent A, Bae T, Missiakas DM, Schneewind O. Signal peptides direct surface proteins to two distinct envelope locations of Staphylococcus aureus. The EMBO journal. 2008;27(20):2656-68.DOI: 10.1038/emboj.2008.185.

8. Takeuchi F, Watanabe S, Baba T, Yuzawa H, Ito T, Morimoto Y, et al. Whole-genome sequencing of staphylococcus haemolyticus uncovers the extreme plasticity of its genome and the evolution of human-colonizing staphylococcal species. Journal of bacteriology. 2005;187(21):7292-308.DOI: 10.1128/jb.187.21.7292-7308.2005.

9. Arora S, Uhlemann AC, Lowy FD, Hook M. A Novel MSCRAMM Subfamily in Coagulase Negative Staphylococcal Species. Frontiers in microbiology. 2016;7:540.DOI: 10.3389/fmicb.2016.00540.

10. Olaya-Abril A, Jimenez-Munguia I, Gomez-Gascon L, Rodriguez-Ortega MJ. Surfomics: shaving live organisms for a fast proteomic identification of surface proteins. Journal of proteomics. 2014;97:164-76.DOI: 10.1016/j.jprot.2013.03.035.

11. Karlsson R, Karlsson A, Backman O, Johansson BR, Hulth S. Identification of key proteins involved in the anammox reaction. FEMS microbiology letters. 2009;297(1):87-94.DOI: 10.1111/j.1574-6968.2009.01677.x.

12. Karlsson R, Davidson M, Svensson-Stadler L, Karlsson A, Olesen K, Carlsohn E, et al. Strain-level typing and identification of bacteria using mass spectrometry-based proteomics. Journal of proteome research. 2012;11(5):2710-20.DOI: 10.1021/pr2010633.

13. Karlsson R, Karlsson A, Backman O, Johansson BR, Hulth S. Subcellular localization of an ATPase in anammox bacteria using proteomics and immunogold electron microscopy. FEMS microbiology letters. 2014;354(1):10-8.DOI: 10.1111/1574-6968.12425.

14. Karlsson R, Thorell K, Hosseini S, Kenny D, Sihlbom C, Sjoling A, et al. Comparative Analysis of Two Helicobacter pylori Strains using Genomics and Mass Spectrometry-Based Proteomics. Frontiers in microbiology. 2016;7:1757.DOI: 10.3389/fmicb.2016.01757.

15. Jansson ET, Trkulja CL, Olofsson J, Millingen M, Wikstrom J, Jesorka A, et al. Microfluidic flow cell for sequential digestion of immobilized proteoliposomes. Analytical chemistry. 2012;84(13):5582-8.DOI: 10.1021/ac300519q.

16. Gonzales-Siles L, Karlsson R, Kenny D, Karlsson A, Sjoling A. Proteomic analysis of enterotoxigenic Escherichia coli (ETEC) in neutral and alkaline conditions. BMC microbiology. 2017;17(1):11.DOI: 10.1186/s12866-016-0914-1.

17. Thompson A, Schafer J, Kuhn K, Kienle S, Schwarz J, Schmidt G, et al. Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. Analytical chemistry. 2003;75(8):1895-904.DOI.

18. Marin E, Haesaert A, Padilla L, Adan J, Hernaez ML, Monteoliva L, et al. Unraveling Gardnerella vaginalis Surface Proteins Using Cell Shaving Proteomics. Frontiers in microbiology. 2018;9:975.DOI: 10.3389/fmicb.2018.00975.

19. Solis N, Cain JA, Cordwell SJ. Comparative analysis of Staphylococcus epidermidis strains utilizing quantitative and cell surface shaving proteomics. Journal of proteomics. 2016;130:190-9.DOI: 10.1016/j.jprot.2015.09.011.

20. Bowden MG, Chen W, Singvall J, Xu Y, Peacock SJ, Valtulina V, et al. Identification and preliminary characterization of cell-wall-anchored proteins of Staphylococcus epidermidis. Microbiology (Reading, England). 2005;151(Pt 5):1453-64.DOI: 10.1099/mic.0.27534-0.

21. Gomez-Gascon L, Luque I, Olaya-Abril A, Jimenez-Munguia I, Orbegozo-Medina RA, Peralbo E, et al. Exploring the pan-surfome of Streptococcus suis: looking for common protein antigens. Journal of proteomics. 2012;75(18):5654-66.DOI: 10.1016/j.jprot.2012.07.025.

22. Pain M, Hjerde E, Klingenberg C, Cavanagh JP. Comparative Genomic Analysis of Staphylococcus haemolyticus Reveals Key to Hospital Adaptation and Pathogenicity. 2019;10(2096).DOI: 10.3389/fmicb.2019.02096.

23. Kainulainen V, Korhonen TK. Dancing to another tune-adhesive moonlighting proteins in bacteria. Biology. 2014;3(1):178-204.DOI: 10.3390/biology3010178.

24. Glenting J, Beck HC, Vrang A, Riemann H, Ravn P, Hansen AM, et al. Anchorless surface associated glycolytic enzymes from Lactobacillus plantarum 299v bind to epithelial cells and extracellular matrix proteins. Microbiol Res. 2013;168(5):245-53.DOI: 10.1016/j.micres.2013.01.003.

25. Henderson B, Martin A. Bacterial virulence in the moonlight: multitasking bacterial moonlighting proteins are virulence determinants in infectious disease. Infection and immunity. 2011;79(9):3476-91.DOI: 10.1128/iai.00179-11.

26. Antikainen J, Kuparinen V, Lahteenmaki K, Korhonen TK. Enolases from Gram-positive bacterial pathogens and commensal lactobacilli share functional similarity in virulence-associated traits. FEMS Immunol Med Microbiol. 2007;51(3):526-34.DOI: 10.1111/j.1574-695X.2007.00330.x.

27. Yamaguchi M, Ikeda R, Nishimura M, Kawamoto S. Localization by scanning immunoelectron microscopy of triosephosphate isomerase, the molecules responsible for contact-mediated killing of Cryptococcus, on the surface of Staphylococcus. Microbiol Immunol. 2010;54(6):368-70.DOI: 10.1111/j.1348-0421.2010.00225.x.

28. Blau K, Portnoi M, Shagan M, Kaganovich A, Rom S, Kafka D, et al. Flamingo cadherin: a putative host receptor for Streptococcus pneumoniae. The Journal of infectious diseases. 2007;195(12):1828-37.DOI: 10.1086/518038.

29. Hussain M, Peters G, Chhatwal GS, Herrmann M. A lithium chloride-extracted, broad-spectrum-adhesive 42-kilodalton protein of Staphylococcus epidermidis is ornithine carbamoyltransferase. Infection and immunity. 1999;67(12):6688-90.DOI.

30. Katakura Y, Sano R, Hashimoto T, Ninomiya K, Shioya S. Lactic acid bacteria display on the cell surface cytosolic proteins that recognize yeast mannan. Applied microbiology and biotechnology. 2010;86(1):319-26.DOI: 10.1007/s00253-009-2295-y.

31. Molkanen T, Tyynela J, Helin J, Kalkkinen N, Kuusela P. Enhanced activation of bound plasminogen on Staphylococcus aureus by staphylokinase. FEBS Lett. 2002;517(1-3):72-8.DOI.

32. Wang W, Jeffery CJ. An analysis of surface proteomics results reveals novel candidates for intracellular/surface moonlighting proteins in bacteria. Molecular bioSystems. 2016;12(5):1420-31.DOI: 10.1039/c5mb00550g.

33. Amblee V, Jeffery CJ. Physical Features of Intracellular Proteins that Moonlight on the Cell Surface. PloS one. 2015;10(6):e0130575.DOI: 10.1371/journal.pone.0130575.

34. Allignet J, England P, Old I, El Solh N. Several regions of the repeat domain of the Staphylococcus caprae autolysin, AtlC, are involved in fibronectin binding. FEMS microbiology letters. 2002;213(2):193-7.DOI: 10.1111/j.1574-6968.2002.tb11305.x.

35. Pankov R, Yamada KM. Fibronectin at a glance. J Cell Sci. 2002;115(Pt 20):3861-3.DOI: 10.1242/jcs.00059.

36. Madani A, Garakani K, Mofrad MRK. Molecular mechanics of Staphylococcus aureus adhesin, CNA, and the inhibition of bacterial adhesion by stretching collagen. PloS one. 2017;12(6):e0179601.DOI: 10.1371/journal.pone.0179601.

37. Grundmeier M, Hussain M, Becker P, Heilmann C, Peters G, Sinha B. Truncation of fibronectin-binding proteins in Staphylococcus aureus strain Newman leads to deficient adherence and host cell invasion due to loss of the cell wall anchor function. Infection and immunity. 2004;72(12):7155-63.DOI: 10.1128/iai.72.12.7155-7163.2004.

38. Maali Y, Martins-Simoes P, Valour F, Bouvard D, Rasigade JP, Bes M, et al. Pathophysiological Mechanisms of Staphylococcus Non-aureus Bone and Joint Infection: Interspecies Homogeneity and Specific Behavior of S. pseudintermedius. Frontiers in microbiology. 2016;7:1063.DOI: 10.3389/fmicb.2016.01063.

39. Switalski LM, Ryden C, Rubin K, Ljungh A, Hook M, Wadstrom T. Binding of fibronectin to Staphylococcus strains. Infection and immunity. 1983;42(2):628-33.DOI.

40. Christner M, Heinze C, Busch M, Franke G, Hentschke M, Bayard Duhring S, et al. sarA negatively regulates Staphylococcus epidermidis biofilm formation by modulating expression of 1 MDa extracellular matrix binding protein and autolysis-dependent release of eDNA. Molecular microbiology. 2012;86(2):394-410.DOI: 10.1111/j.1365-2958.2012.08203.x.

41. Christner M, Franke GC, Schommer NN, Wendt U, Wegert K, Pehle P, et al. The giant extracellular matrix-binding protein of Staphylococcus epidermidis mediates biofilm accumulation and attachment to fibronectin. Molecular microbiology. 2010;75(1):187-207.DOI: 10.1111/j.1365-2958.2009.06981.x.

42. Buttner H, Mack D, Rohde H. Structural basis of Staphylococcus epidermidis biofilm formation: mechanisms and molecular interactions. Frontiers in cellular and infection microbiology. 2015;5:14.DOI: 10.3389/fcimb.2015.00014.

43. Paulsson M, Wadstrom T. Vitronectin and type-I collagen binding by Staphylococcus aureus and coagulase-negative staphylococci. FEMS Microbiol Immunol. 1990;2(1):55-62.DOI: 10.1111/j.1574-6968.1990.tb03479.x.

44. Harris LG, Murray S, Pascoe B, Bray J, Meric G, Mageiros L, et al. Biofilm Morphotypes and Population Structure among Staphylococcus epidermidis from Commensal and Clinical Samples. PloS one. 2016;11(3):e0151240.DOI: 10.1371/journal.pone.0151240.

45. Freitas AI, Lopes N, Oliveira F, Bras S, Franca A, Vasconcelos C, et al. Comparative analysis between biofilm formation and gene expression in Staphylococcus epidermidis isolates. Future microbiology. 2018;13:415-27.DOI: 10.2217/fmb-2017-0140.

46. Fredheim EG, Klingenberg C, Rohde H, Frankenberger S, Gaustad P, Flaegstad T, et al. Biofilm formation by Staphylococcus haemolyticus. Journal of clinical microbiology. 2009;47(4):1172-80.DOI: 10.1128/jcm.01891-08.

47. Grice EA, Segre JA. The skin microbiome. Nature reviews Microbiology. 2011;9(4):244-53.DOI: 10.1038/nrmicro2537.

48. Cavanagh JP, Wolden R, Heise P, Esaiassen E, Klingenberg C, Aarag Fredheim EG. Antimicrobial susceptibility and body site distribution of community isolates of coagulase-negative

staphylococci. APMIS : acta pathologica, microbiologica, et immunologica Scandinavica. 2016;124(11):973-8.DOI: 10.1111/apm.12591.

49. Holtje JV, Mirelman D, Sharon N, Schwarz U. Novel type of murein transglycosylase in Escherichia coli. Journal of bacteriology. 1975;124(3):1067-76.DOI.

50. Stapleton MR, Horsburgh MJ, Hayhurst EJ, Wright L, Jonsson IM, Tarkowski A, et al. Characterization of IsaA and SceD, two putative lytic transglycosylases of Staphylococcus aureus. Journal of bacteriology. 2007;189(20):7316-25.DOI: 10.1128/jb.00734-07.

51. Burian M, Wolz C, Goerke C. Regulatory adaptation of Staphylococcus aureus during nasal colonization of humans. PloS one. 2010;5(4):e10040.DOI: 10.1371/journal.pone.0010040.

52. Biswas R, Voggu L, Simon UK, Hentschel P, Thumm G, Gotz F. Activity of the major staphylococcal autolysin Atl. FEMS microbiology letters. 2006;259(2):260-8.DOI: 10.1111/j.1574-6968.2006.00281.x.

53. Heilmann C, Hussain M, Peters G, Gotz F. Evidence for autolysin-mediated primary attachment of Staphylococcus epidermidis to a polystyrene surface. Molecular microbiology. 1997;24(5):1013-24.DOI.

54. Lopes AA, Yoshii Y, Yamada S, Nagakura M, Kinjo Y, Mizunoe Y, et al. Roles of lytic transglycosylases in biofilm formation and beta-lactam resistance in methicillin-resistant Staphylococcus aureus. Antimicrob Agents Chemother. 2019.DOI: 10.1128/aac.01277-19.

55. Roche FM, Massey R, Peacock SJ, Day NP, Visai L, Speziale P, et al. Characterization of novel LPXTG-containing proteins of Staphylococcus aureus identified from genome sequences. Microbiology (Reading, England). 2003;149(Pt 3):643-54.DOI: 10.1099/mic.0.25996-0.

56. Casas V, Vadillo S, San Juan C, Carrascal M, Abian J. The Exposed Proteomes of Brachyspira hyodysenteriae and B. pilosicoli. Frontiers in microbiology. 2016;7:1103.DOI: 10.3389/fmicb.2016.01103.

57. Sitkiewicz I, Babiak I, Hryniewicz W. Characterization of transcription within sdr region of Staphylococcus aureus. Antonie van Leeuwenhoek. 2011;99(2):409-16.DOI: 10.1007/s10482-010-9476-7.

58. Askarian F, Uchiyama S, Valderrama JA, Ajayi C, Sollid JUE, van Sorge NM, et al. Serine-Aspartate Repeat Protein D Increases Staphylococcus aureus Virulence and Survival in Blood. Infection and immunity. 2017;85(1).DOI: 10.1128/iai.00559-16.

59. Trivedi S, Uhlemann AC, Herman-Bausier P, Sullivan SB, Sowash MG, Flores EY, et al. The Surface Protein SdrF Mediates Staphylococcus epidermidis Adherence to Keratin. The Journal of infectious diseases. 2017;215(12):1846-54.DOI: 10.1093/infdis/jix213.

60. Askarian F, Ajayi C, Hanssen AM, van Sorge NM, Pettersen I, Diep DB, et al. The interaction between Staphylococcus aureus SdrD and desmoglein 1 is important for adhesion to host cells. Scientific reports. 2016;6:22134.DOI: 10.1038/srep22134.

61. Rana RR, Zhang M, Spear AM, Atkins HS, Byrne B. Bacterial TIR-containing proteins and host innate immune system evasion. Med Microbiol Immunol. 2013;202(1):1-10.DOI: 10.1007/s00430-012-0253-2.

62. Askarian F, van Sorge NM, Sangvik M, Beasley FC, Henriksen JR, Sollid JU, et al. A Staphylococcus aureus TIR domain protein virulence factor blocks TLR2-mediated NF-kappaB signaling. Journal of innate immunity. 2014;6(4):485-98.DOI: 10.1159/000357618.

63. Pourmand MR, Clarke SR, Schuman RF, Mond JJ, Foster SJ. Identification of antigenic components of Staphylococcus epidermidis expressed during human infection. Infection and immunity. 2006;74(8):4644-54.DOI: 10.1128/iai.00521-06.

64. Lang S, Livesley MA, Lambert PA, Littler WA, Elliott TS. Identification of a novel antigen from Staphylococcus epidermidis. FEMS Immunol Med Microbiol. 2000;29(3):213-20.DOI: 10.1111/j.1574-695X.2000.tb01525.x.

65. Solis N, Cordwell SJ. Cell Shaving and False-Positive Control Strategies Coupled to Novel Statistical Tools to Profile Gram-Positive Bacterial Surface Proteomes. Methods in molecular biology (Clifton, NJ). 2016;1440:47-55.DOI: 10.1007/978-1-4939-3676-2_4.

66. Cavanagh JP, Pain M, Askarian F, Bruun JA, Urbarova I, Wai SN, et al. Comparative exoproteome profiling of an invasive and a commensal Staphylococcus haemolyticus isolate. Journal of proteomics. 2019;197:106-14.DOI: 10.1016/j.jprot.2018.11.013.

67. Modun B, Williams P. The staphylococcal transferrin-binding protein is a cell wall glyceraldehyde-3-phosphate dehydrogenase. Infection and immunity. 1999;67(3):1086-92.DOI.

68. Boukamp P, Petrussevska RT, Breitkreutz D, Hornung J, Markham A, Fusenig NE. Normal keratinization in a spontaneously immortalized aneuploid human keratinocyte cell line. The Journal of cell biology. 1988;106(3):761-71.DOI.

69. Edwards AM, Potts JR, Josefsson E, Massey RC. Staphylococcus aureus host cell invasion and virulence in sepsis is facilitated by the multiple repeats within FnBPA. PLoS pathogens. 2010;6(6):e1000964.DOI: 10.1371/journal.ppat.1000964.

70. Christensen GD, Simpson WA, Younger JJ, Baddour LM, Barrett FF, Melton DM, et al. Adherence of coagulase-negative staphylococci to plastic tissue culture plates: a quantitative model for the adherence of staphylococci to medical devices. Journal of clinical microbiology. 1985;22(6):996-1006.DOI.

71. Edwards AM, Massey RC. Invasion of human cells by a bacterial pathogen. Journal of visualized experiments : JoVE. 2011(49).DOI: 10.3791/2693.

72. Hildebrandt P, Surmann K, Salazar MG, Normann N, Volker U, Schmidt F. Alternative fluorescent labeling strategies for characterizing gram-positive pathogenic bacteria: Flow cytometry supported counting, sorting, and proteome analysis of Staphylococcus aureus retrieved from infected host cells. Cytometry Part A : the journal of the International Society for Analytical Cytology. 2016;89(10):932-40.DOI: 10.1002/cyto.a.22981.

73. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXivorg. 2013.DOI.

74. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics (Oxford, England). 2014;30(14):2068-9.DOI: 10.1093/bioinformatics/btu153.

75. Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R, et al. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics (Oxford, England). 2010;26(13):1608-15.DOI: 10.1093/bioinformatics/btq249.

76. Yu CS, Chen YC, Lu CH, Hwang JK. Prediction of protein subcellular localization. Proteins. 2006;64(3):643-51.DOI: 10.1002/prot.21018.

77. Zhou M, Boekhorst J, Francke C, Siezen RJ. LocateP: genome-scale subcellular-location predictor for bacterial proteins. BMC bioinformatics. 2008;9:173.DOI: 10.1186/1471-2105-9-173.

78. Huerta-Cepas J, Szklarczyk D, Heller D, Hernandez-Plaza A, Forslund SK, Cook H, et al. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic acids research. 2019;47(D1):D309-d14.DOI: 10.1093/nar/gky1085.

79. Eddy SR. Accelerated Profile HMM Searches. PLoS computational biology. 2011;7(10):e1002195.DOI: 10.1371/journal.pcbi.1002195.

80. Potter SC, Luciani A, Eddy SR, Park Y, Lopez R, Finn RD. HMMER web server: 2018 update. Nucleic acids research. 2018;46(W1):W200-w4.DOI: 10.1093/nar/gky448.

81. Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL. NCBI BLAST: a better web interface. Nucleic acids research. 2008;36(Web Server issue):W5-9.DOI: 10.1093/nar/gkn201.

82. Mani M, Chen C, Amblee V, Liu H, Mathur T, Zwicke G, et al. MoonProt: a database for proteins that are known to moonlight. Nucleic acids research. 2015;43(Database issue):D277-82.DOI: 10.1093/nar/gku954.

83. Chen C, Zabad S, Liu H, Wang W, Jeffery C. MoonProt 2.0: an expansion and update of the moonlighting proteins database. Nucleic acids research. 2018;46(D1):D640-d4.DOI: 10.1093/nar/gkx1043.

84. PRoteomics IDEntifications Database [Internet]. 2020. Available from: https://www.ebi.ac.uk/pride/archive.

85. European Nucleotide Archive, ERS066380 [Internet]. 2020. Available from: https://www.ebi.ac.uk/ena/data/search?query=ERS066380.

86. European Nucleotide Archive, [Internet]. 2020. Available from: https://www.ebi.ac.uk/.

9 Figure Legends

Figure 1– Bacterial surface protein shaving, graphical abstract.

Figure 1 - Comparison of S. haemolyticus surface protein expression after HaCaT colonization (top) and the control group (bottom). Bacterial surface proteins (multicolored) are degraded by the protease Trypsin (scissors).

Figure 2 - The use of LPI™ methodology together with TMT labelling when performing surface shaving.

Figure 2 - Three lanes were filled with bacterial cells after exposure to HaCaT cells (A) and three lanes were filled with bacterial cells only exposed to media (B). After surface shaving, the eluted peptides were tagged with TMT labels, pooled and subsequently analyzed by LC-MS/MS.

Figure 3 - Adhesion and biofilm assays of S. haemolyticus.
Figure 3 - Columns with black bars are clinical isolates and white bars are commensal isolates. Sample no. 6 was chosen for bacterial surface shaving (marked with asterisk). A-C) Solid phase host matrix binding assay; A) Adhesion to fibronectin; B) Adhesion to plastic; C) Adhesion to collagen; D) Semi-quantitative determination of biofilm formation; E) Adhesion to human keratinocytes.

Figure 4 – Clusters of Orthologous Groups of proteins (COG).

Figure 4 - Comparison of Clusters of Orthologous Groups of proteins (COG) between the total proteins of the strain (#2539) and proteins found after HaCaT.

Supplementary Table 1 - Workflow for bacterial protein surface shaving samples. X = performed, - = not performed

***Staphylococcus haemolyticus* surface proteins**

Supplementary Table 2 - Manufacturer's and modified gradient using the Pierce High pH Reversed-Phase Peptide Fractionation Kit.

Supplementary Table 3 - The cell surface shaving and LC-MS/MS analysis results identified 325 proteins with ≥#2 peptide-spectrum matches (PSMs).

Supplementary Table 4- FASTA sequences of the proteins from the cell surface shaving and LC-MS/MS analysis.

Table 1 - S. haemolyticus strains included in the study.

Sample	Country	Isolated from	Year of isolation	ENA ID*	Lab. ID
1	Norway	Blood	1995	ERS066267	25-12
2	Norway	Blood	2004	ERS066284	51-11
3	Norway	Blood	2002	ERS066281	51-08
4	Switzerland	Blood	2001	ERS066398	53-18
5	Germany	Blood	2008	ERS066335	53-73
6**	Switzerland	Wound	2004	ERS066380	53-38
7	Norway	Blood	2004	ERS066295	51-29
8	Switzerland	Blood	2004	ERS066370	53-35
9	Switzerland	Unknown	2006	ERS066381	53-49
10	Switzerland	Blood	2005	ERS066386	53-48
11	Norway	Nasal Swab	2010	ERS066315	54-64
12	Norway	Armpit	2013	ERS3370776	57-01
13	Norway	Groin	2013	ERS3370780	57-12
14	Norway	Armpit	2014	ERS3370802	57-66
15	Norway	Groin	2014	ERS3370809	58-28
16	Norway	Hamstring	2013	ERS3370784	57-22
17	Norway	Groin	2014	ERS3370790	57-33
18	Norway	Groin	2014	ERS3370800	57-61
19	Norway	Groin	2014	ERS3370806	58-08
20	Norway	Unknown	2013	ERS3370815	58-62

Table 2 - Ten clinical and ten commensal S. haemolyticus strains were included in the study. Samples 1-10 are clinical strains and 11-20 are commensal strains. *ENA = European Nucleotide Archive. ** Strain no. 6 was chosen for bacterial surface protein shaving

Table 2 – Predicted surface proteins after bacterial surface protein shaving of S. haemolyticus.

Accession	# PSM	# Unique Peptides	Fold change HaCaT vs Control	p-value HaCaT vs Control	LPxTG Cell-wall anchored	Preferred name, EggNOG	Annotation summary
ACAKHAOO_00208	8	7	1,75	**0.046**	-	ymaC	DUF867 type protein
ACAKHAOO_02015	7	2	1,60	**0.014**	-	sceD	Putative transglycosylase SceD
ACAKHAOO_00540	433	60	1,56	0.123	LPDTG	pelX	Serine-aspartate repeat-containing protein I / YSIRK-type signal peptide-containing protein
ACAKHAOO_01033	54	25	1,46	**0.039**	-	atl	Bifunctional autolysin
ACAKHAOO_00522	6	3	1,35	0.054	LPNAG	sasH	Mannosylglucosyl-3-phosphoglycerate phosphatase
ACAKHAOO_00546	188	38	1,34	0.200	LPDTG	-	Serine-aspartate repeat-containing protein I / C protein alpha-antigen
ACAKHAOO_00080	6	5	1,24	0.143	LPKSG	-	Serine-aspartate repeat-containing protein D / YSIRK-type signal peptide-containing protein
ACAKHAOO_02469	10	3	1,24	0.380	-	isaA	Putative transglycosylase IsaA
ACAKHAOO_00631	5	2	1,18	0.213	-	-	Hypothetical protein
ACAKHAOO_02587	2	2	1,06	0.815	-	isaB	Immunodominant staphylococcal antigen B
ACAKHAOO_00744	2	2	1,01	0.916	-	dtpT	Di-tripepride ABC transporter
ACAKHAOO_02598	8	5	-1,01	0.997	-	proX	ABC transporter substrate-binding protein / Glycine betaine/carnitine transport binding protein GbuC
ACAKHAOO_02593	81	28	-1,01	0.936	LPNTG	-	Cell wall anchor protein / hypothetical protein
ACAKHAOO_01810	3	3	-1,02	0.960	-	yhaN	Putative protein YhaN
ACAKHAOO_01224	2	2	-1,04	0.907	-	rseP	Putative zinc metalloprotease
ACAKHAOO_02549	5	3	-1,05	0.766	-	brpA	Polyisoprenyl-teichoic acid--peptidoglycan teichoic acid transferase TagU / transcriptional regulator
ACAKHAOO_01770	2	2	-1,06	0.763	-	lytD	Bifunctional autolysin
ACAKHAOO_01453	7	3	-1,07	0.734	-	lapA	Extracellular matrix-binding protein ebh / YSIRK-type signal peptide-containing protein
ACAKHAOO_01492	13	5	-1,07	0.915	-	ebpS	Elastin-binding protein EbpS
ACAKHAOO_01541	2	2	-1,09	0.689	-	yqhL	Putative protein YibN / sulfurtransferase
ACAKHAOO_00323	5	4	-1,09	0.573	-	ykuT	Small-conductance mechanosensitive channel
ACAKHAOO_01042	4	2	-1,10	0.582	-	cyoA	Putative quinol oxidase subunit 2
ACAKHAOO_02168	8	5	-1,10	0.528	-	cusA	Swarming motility protein SwrC
ACAKHAOO_01077	3	3	-1,10	0.596	-	recN	Cell-wall vinding protein / hypothetical protein
ACAKHAOO_01640	2	2	-1,11	0.230	-	secF	Protein translocase subunit SecDF
ACAKHAOO_01808	9	5	-1,12	0.545	-	prsA	Foldase protein PrsA
ACAKHAOO_00719	2	1	-1,14	0.487	-	corC1	UPF0053 protein / HlyC/CorC family transporter
ACAKHAOO_02236	3	1	-1,14	0.633	-	lyrA	Lysostaphin resistance protein A
ACAKHAOO_01582	14	7	-1,16	0.360	-	yqfA	UPF0365 protein / hypothetical protein
ACAKHAOO_01462	6	6	-1,17	0.399	-	ponA	Penicillin-binding protein

ACAKHAOO_01561	3	3	-1,17	0.634	-	sodA	Superoxide dismutase [Mn/Fe]
ACAKHAOO_01806	5	3	-1,18	0.235	-	yhaH	Hypothetical protein
ACAKHAOO_00464	28	12	-1,19	0.327	-	ftsH	ATP-dependent zinc metalloprotease FtsH
ACAKHAOO_01734	49	8	-1,19	0.394	-	ytxG	DUF948 domain-containing protein
ACAKHAOO_02191	7	6	-1,20	0.452	-	fhuD	Ferrichrome ABC transporter substrate-binding protein
ACAKHAOO_01088	5	1	-1,20	0.591	-	-	DUF4064 hypothetical protein
ACAKHAOO_01722	15	8	-1,21	0.377	-	htrA	Serine protease Do-like HtrA/HtrB
ACAKHAOO_02068	2	2	-1,21	0.607	-	-	Hypothetical protein
ACAKHAOO_01403	2	1	-1,21	0.660	-	pstB	Phosphate import ATP-binding protein PstB 3
ACAKHAOO_00958	5	4	-1,22	0.357	-	spsB	Signal peptidase IB
ACAKHAOO_00494	2	2	-1,23	0.243	-	yacL	Putative PIN and TRAM-domain containing protein YacL
ACAKHAOO_02026	2	1	-1,23	0.314	-	atpF	ATP synthase subunit b
ACAKHAOO_01924	10	3	-1,24	0.066	-	-	Hypothetical protein
ACAKHAOO_01062	3	1	-1,25	0.454	-	-	Hypothetical protein
ACAKHAOO_00182	40	16	-1,28	0.265	-	sitA	Metal ABC transporter substrate-binding protein / Manganese-binding lipoprotein MntA
ACAKHAOO_01347	8	5	-1,29	0.151	-	-	Hypothetical protein
ACAKHAOO_00718	4	3	-1,31	0.223	-	fruA	PTS system fructose-specific EIIABC component
ACAKHAOO_00753	25	10	-1,32	0.226	-	fatB	Putative ABC transporter solute-binding protein YclQ
ACAKHAOO_01747	12	6	-1,34	0.096	-	-	Hypothetical protein
ACAKHAOO_00561	8	7	-1,34	0.209	-	murF	Capsule biosynthesis protein CapA
ACAKHAOO_02597	8	7	-1,35	0.204	-	ydfJ	Membrane protein YdfJ
ACAKHAOO_01736	2	2	-1,38	0.169	-	sftA	DNA translocase FtsK/SftA
ACAKHAOO_02099	2	2	-1,38	0.121	-	fecB	Iron citrate ABC transporter substrate-binding protein YfmC
ACAKHAOO_00362	6	5	-1,39	0.108	-	penP	Beta-lactamase
ACAKHAOO_00976	9	5	-1,40	0.051	-	oppA	Oligopeptide ABC transporter / Dipeptide-binding protein DppE
ACAKHAOO_00003	25	7	-1,41	0.500	LPMTG		Hypothetical protein
ACAKHAOO_01406	17	7	-1,42	0.107	-	pstS	Phosphate-binding protein PstS
ACAKHAOO_02108	18	7	-1,43	0.261	-	-	Hypothetical protein
ACAKHAOO_00974	3	3	-1,47	0.139	-	oppD	ABC transporter / nickel transport system / Oligopeptide transport ATP-binding protein OppD
ACAKHAOO_00701	11	4	-1,47	0.338	-	-	Hypothetical protein
ACAKHAOO_00229	25	14	-1,49	0.130	-	pbpC	Beta-lactam-inducible penicillin-binding protein
ACAKHAOO_01885	2	1	-1,53	0.146	-	yihY	UPF0761 protein
ACAKHAOO_02197	5	2	-1,54	**0.038**	-	ssaA	Staphylococcal secretory antigen SsaA / CHAP domain-containing protein
ACAKHAOO_01752	3	2	-1,67	0.102	LPNTG	-	Extracellular matrix-binding protein ebh / Signal peptide protein, YSIRK family / DUF1542
ACAKHAOO_00904	9	4	-1,74	**0.026**	-	metQ	Methionine-binding lipoprotein MetQ

Staphylococcus haemolyticus surface proteins

Table 3 – Proteins with statistically significant altered abundance after surface shaving of S. haemolyticus incubated with human keratinocytes.

Accession	# PSM	# Unique Peptides	Fold change HaCaT vs Control	p-value HaCaT vs Control	Summarized prediction of subcellular localization	Preferred name, EggNOG	Annotation summary
ACAKHAOO_01782	3	2	1,90	0.015	Cytoplasmic	metK	S-adenosylmethionine synthase
ACAKHAOO_00208	8	7	1,75	0.046	Surface	ymaC	DUF867 type protein
ACAKHAOO_02015	7	2	1,60	0.014	Surface	sceD	Putative transglycosylase SceD
ACAKHAOO_02031	2	2	1,57	0.016	Cytoplasmic	upp	Uracil phosphoribosyltransferase
ACAKHAOO_00454	6	3	1,55	0.027	Cytoplasmic	ctc	50S ribosomal protein L25
ACAKHAOO_01033	54	25	1,46	0.039	Surface	atl	Bifunctional autolysin
ACAKHAOO_00250	4	3	1,40	0.044	Cytoplasmic	-	TIR domain-containing protein
ACAKHAOO_00947	2	1	1,39	0.032	Cytoplasmic	ppiB	Putative peptidyl-prolyl cis-trans isomerase
ACAKHAOO_02231	2	2	1,35	0.031	Cytoplasmic	-	Putative oxidoreductase YghA
ACAKHAOO_01626	2	1	1,33	0.012	Cytoplasmic	mnmA	tRNA-specific 2-thiouridylase MnmA
ACAKHAOO_01821	4	3	1,31	0.001	Cytoplasmic	nagB	Glucosamine-6-phosphate deaminase
ACAKHAOO_00516	112	20	1,22	0.017	Cytoplasmic	tuf	Elongation factor Tu
ACAKHAOO_00797	45	14	-1,31	0.048	Cytoplasmic	pgk	Phosphoglycerate kinase
ACAKHAOO_01712	7	5	-1,44	0.026	Cytoplasmic	ezrA	Septation ring formation regulator EzrA
ACAKHAOO_01065	2	1	-1,51	0.004	Cytoplasmic	-	DUF697 domain-containing protein
ACAKHAOO_02197	5	2	-1,54	0.038	Surface	ssaA	Staphylococcal secretory antigen SsaA / CHAP domain-containing protein
ACAKHAOO_01875	14	5	-1,65	0.034	Cytoplasmic	yhbO	Uncharacterized protein SH1084
ACAKHAOO_00904	9	4	-1,74	0.026	Surface	metQ	Methionine-binding lipoprotein MetQ
ACAKHAOO_01422	2	2	-1,78	0.000	Cytoplasmic	yaaN	TelA-like protein

Figure 1

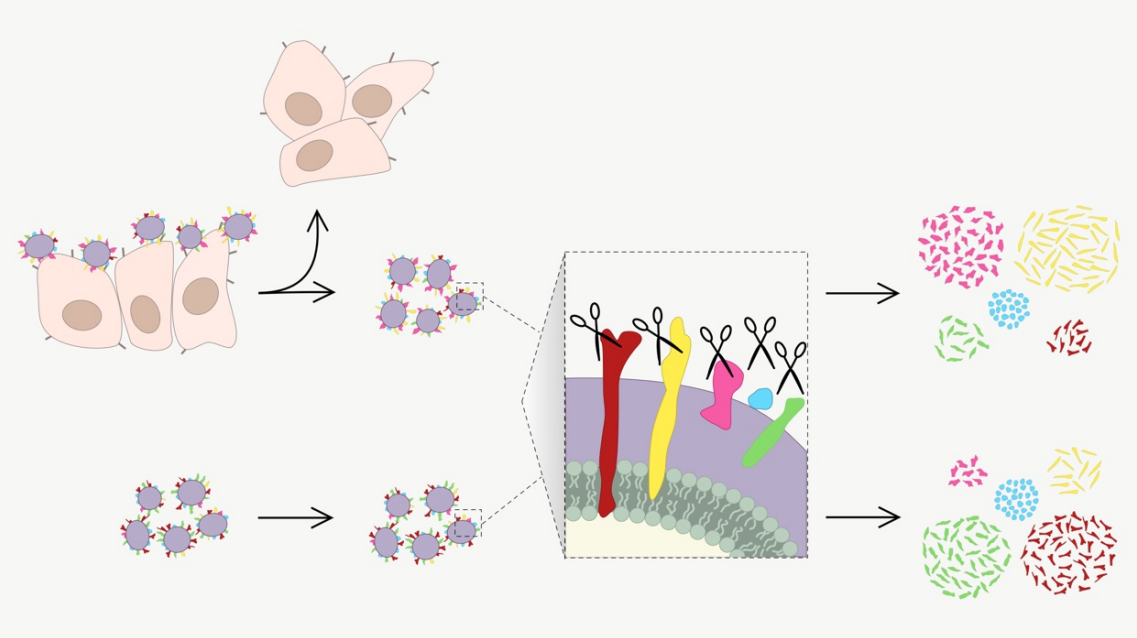


Figure 2

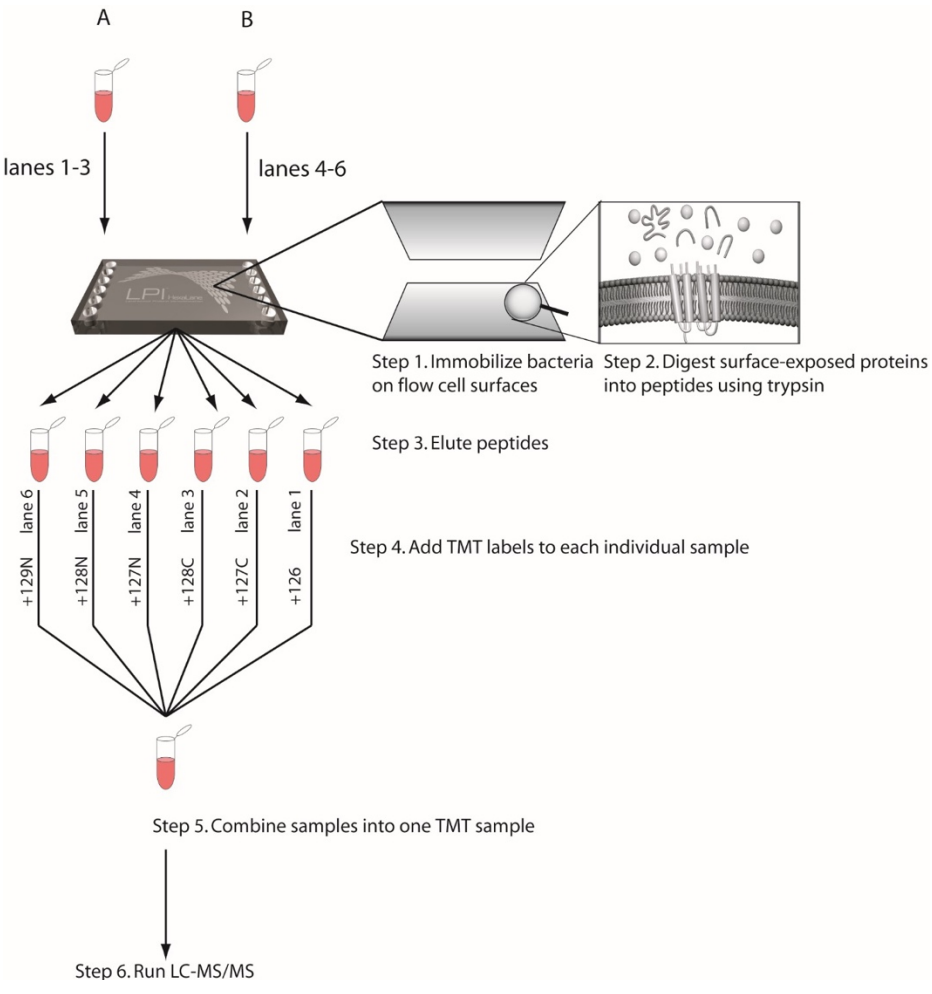


Figure 3

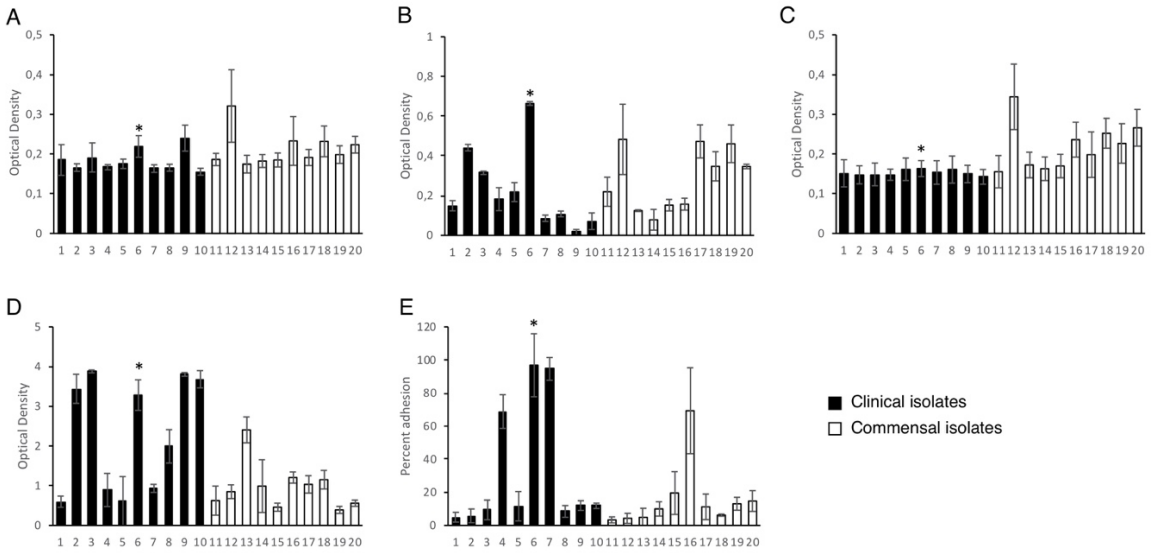
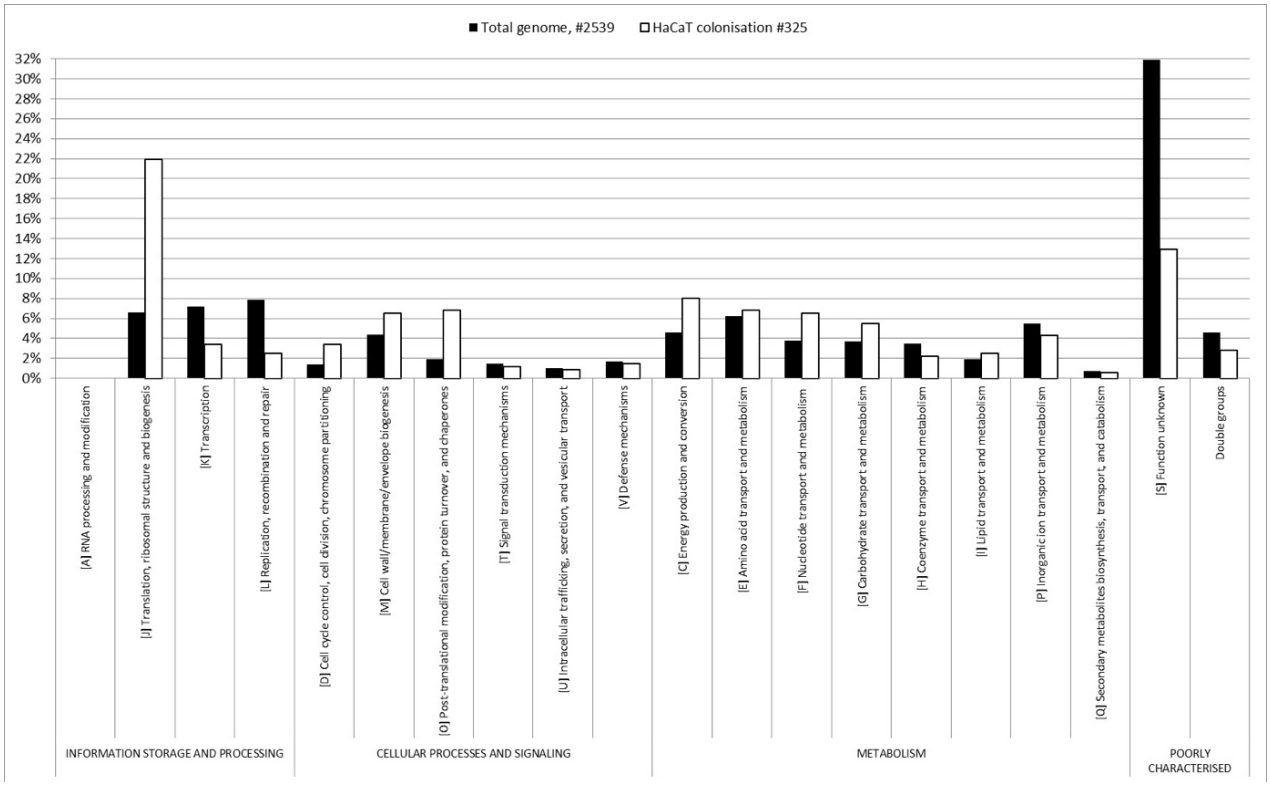


Figure 4



Supplementary table 1 - Overview of workflow for bacterial protein surface shaving samples. X = performed, - = not performed

Workflow in chronological order	**HaCaT colonization**	**Control group**
Primary TSB culture	**X**	**X**
Subculture in DMEM with 10% FBS	**X**	**X**
Wash twice with DPBS	**X**	**X**
Resuspended in DMEM with 10% FBS	**X**	**-**
Incubate *S. haemolyticus* with HaCaT cells in tissue culture plates	**X**	**-**
Wash four times with DPBS	**X**	**-**
Resuspended in DPBS	**X**	**X**
FACS	**X**	**X**
Ultra centrifugation after FACS	**X**	**X**
LPI^{TM} Flow Cell	**X**	**X**

Supplementary table 2 – Manufacturer's and modified gradient using the Pierce High pH Reversed-Phase Peptide Fractionation Kit.

Protocol according to manufacturer

Elution enough for 1 sample = 300 µl

Fraction No.	Acetonitrile (%)	Acetonitrile (µL)	Triethylamine (0.1%) (µL)
1	10.0%	100	900
2	12.5%	125	875
3	15.0%	150	850
4	17.5%	175	825
5	20.0%	200	800
6	22.5%	225	775
7	25.0%	250	750
8	50.0%	500	500

Modified protocol

TMT gradient enough for 3 samples = 1000 µl, wash 3%

Fraction No.	Acetonitrile (%)	Acetonitrile (µL)	Triethylamine (0.1%) (µL)
1	7.0%	70	930
2	9.0%	90	910
3	10.0%	100	900
4	11.0%	110	890
5	12.0%	120	880
6	14.0%	140	860
7	16.0%	160	840
8	18.0%	180	820
9	20.0%	200	800
10	22.0%	220	780
11	25.0%	250	750
12	50.0%	500	500
13=extra	75.0%	750	250

Ratio Calculation:

==> Minimum Quan Value Detected = 599.1
==> Minimum Quan Value Used = 599.1

Experimental Bias:

Normalize On Protein Median
- Minimum Protein Count: = 20

==> Resulting Normalization Values:
126/(126+127_C+128_C): 0.235
127_C/(126+127_C+128_C): 0.390
127_N/(126+127_C+128_C): 0.250
128_C/(126+127_C+128_C): 0.361
128_N/(126+127_C+128_C): 0.241
129_N/(126+127_C+128_C): 0.490

Ratio Reporting:

126/(126+127_C+128_C)
127_C/(126+127_C+128_C)
127_N/(126+127_C+128_C)
128_C/(126+127_C+128_C)
128_N/(126+127_C+128_C)
129_N/(126+127_C+128_C)

SET1	1	2	5	3	4	7
sample	1783	1784	1785	1786	1787	1788
label	126	127C	128C	127N	128N	129N
volume	41/124	41/124	41/124	41/124	41/124	41/124
	01:01	01:02	01:03	02:01	02:02	02:03

Sample ID	Notes	
1783	Exposure human cells	1 HaCaT + FACS
1784	Exposure human cells	2 HaCaT + FACS
1785	Exposure human cells	3 HaCaT + FACS
1786	Control	FACS
1787	Control	FACS
1788	Control	FACS

Staphylococcus borealis sp. nov., a novel member of the *Staphylococcaceae* family isolated from skin and blood in humans

Maria Pain[1], Runa Wolden[1], Daniel Jaén-Luchoro[2,3], Francisco Salvà-Serra[3,4,5,6], Roger Karlsson[4], Claus Klingenberg[1,5] Jorunn Pauline Cavanagh[1,5]

[1] Pediatric Infection Group, Department of Clinical Medicine, UiT The Arctic University of Norway, Tromsø, Norway
[2] Department of Infectious Diseases, Institute of Biomedicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden
[3] Centre for Antibiotic Resistance Research (CARe), University of Gothenburg, Gothenburg, Sweden
[4] Department of Clinical Microbiology, Sahlgrenska University Hospital, Region Västra Götaland, Gothenburg, Sweden
[5] Culture Collection University of Gothenburg (CCUG), Sahlgrenska Academy, University of Gothenburg, Sweden
[6] Microbiology, Department of Biology, University of the Balearic Islands, Palma de Mallorca, Spain
[7] Department of Paediatrics, University Hospital of North Norway, Tromsø, Norway

Abstract

When analysing a large cohort of *Staphylococcus haemolyticus*, five isolates obtained from different sources (four from skin and one from blood culture) with aberrant phenotypic and genotypic traits were identified. The five isolates were initially speciated as *S. haemolyticus* based on 16S rRNA gene sequence and MALDI-TOF MS. All five isolates were phenotypically similar to and had nearly identical 16S rRNA gene sequences compared to *S. haemolyticus*. However, compared to *S. haemolyticus,* these five isolates demonstrate i) considerable phylogenetic distance with an average nucleotide identity < 95 % and inferred DNA-DNA hybridization < 70 %, ii) 1,185 specific genes not found in *S. haemolyticus*. iii) a pigmented phenotype and iv) urease production. Based on the phenotypic and genotypic results, we conclude that these isolates represent a novel species, for which the name *Staphylococcus borealis* sp. nov. is proposed. The whole genome sequence of type strain 51-48^T (= CCUG 73747^T = CECT 30011^T) is deposited to the European nucleotide archive under accession number GCA_001224225.1 and PRJEB36042.

Introduction

Members of the genus *Staphylococcus*, currently consisting of 54 species and 22 subspecies[1] with validly published names are most often found on the skin and mucus membranes of mammals and birds (Götz et al., 2006). Staphylococci, and in particular the coagulase-positive *Staphylococcus aureus*, are a major cause of clinical disease in both humans and animals. The coagulase-negative staphylococci (CoNS) colonise different niches of the human skin (Cavanagh et al., 2016) and are part of the commensal host microbiota. However, over the last decades some CoNS-species such as *Staphylococcus epidermidis, Staphylococcus hominis* and *Staphylococcus haemolyticus* have emerged as important opportunistic pathogens causing primarily disease in patients with foreign body implants or impaired immunity (Becker et al., 2014).

As part of a previous study analysing a large cohort of *S. haemolyticus* (Pain et al., 2019), we detected five bacterial isolates with aberrant phenotypic and genotypic traits. All five isolates originated from the same geographic location, Tromsø, in North Norway, Norway. Four strains were isolated from skin swabs from the groin and armpit of healthy volunteers (Cavanagh et al., 2016), and one strain was isolated from blood culture in 1997 at the University Hospital of North Norway, Norway (Cavanagh et al., 2014, 2016). The five isolates were all initially identified as *S. haemolyticus* based on 16S rRNA gene sequencing and/or matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS) using a Microflex LT instrument (Bruker Daltonics, Bremen, Germany), Flex Control software and the MALDI Biotyper 3.1 software (Bruker Daltonics, Bremen, Germany).

The five isolates were further characterized, and based on established phenotypic and genotypic methods for species identification (Chun et al., 2018; Freney et al., 2009), we propose that they belong to a new staphylococcal species hereafter designated *Staphylococcus borealis*.

[1] Based on *Staphylococcus* species from www.bacterio.net. The following species and subspecies were removed as they have been either moved or combined with other species: *S. caselyticus* (to *Macrococcus caseolyticus*) *S. jettensis* (to *S. petrasii* subsp. *jettensis*), *S. pulveri* (combined with *S. vitulinus)*, *S. hyicus* subsp. *chromogenes* (to *S. chromogenes*), *S. hyicus* subsp. *hyicus* (to *S. hyicus*) and all subspecies of *S. sciuri* combined into *S. sciuri* (with no subspecies). The newly validated species *S. caeli*, *S. debuckii,* and *S. pseudoxylosus* not yet listed on bacterionet was also included in the total number.

Whole genome sequencing (WGS) and phylogenetic analysis

DNA from the five novel strains were extracted using the Wizard® Genomic DNA kit (Promega, Durham, USA) and the MasterPure™ Gram Positive DNA Purification Kit (Lucigen, Middleton, USA) for isolating DNA for Pac Bio sequencing of the type strain. WGS was performed using Illumina MiSeq and Genome Analyzer II sequencing, as described previously (Cavanagh et al., 2014; Pain et al., 2019), and strain 51-48T was additionally sequenced with PacBio. Assembly was done using Shovill for the Illumina sequences (https://github.com/tseemann/shovill). For the Pacbio sequence, circular consensus sequences generated from demultiplexed sequencing reads were assembled with Canu v1.8 (Koren et al., 2017), and circularized using Circlator (Hunt et al., 2015). The assembly was polished with Pilon (Walker et al., 2014), using Illumina sequences (GCA_001224225.1) generated previously by Cavanagh *et al.* (Cavanagh et al., 2014). Alignments of Illumina sequences and the PacBio assembly were done using BWA-MEM (Li, 2013). The resultant draft genomes were deposited in GenBank under the accession number (GCA_001224225, PRJEB36042, XX). The genome size range was is 2,621,226 – 2,760,219 bp, with 2,486-2,556 CDSs. The G+C content of the novel species ranged from 33.57 % to 33.72 % (**Table 1**), which is in the range of 33-40% expected for species of the genus *Staphylococcus* (Götz et al., 2006). The G+C content is 0.69-0.84% higher than for the *S. haemolyticus* type strain CCUG 7323T.

The 16S rRNA gene sequences of the five isolates were analysed using the EzBioCloud online tool 16S-based ID (Yoon et al., 2017). These results showed that the highest similarities were found with *S. haemolyticus* MTCC 3383T (99.86% for isolates 51-48T, 57-14 and 57-74; 99.93% for 58-22 and 58-52) and *Staphylococcus petrasii* subsp. *jettensis* (99.51% for isolates 51-48T, 57-14 and 57-74; 99.44% for 58-22 and 58-52) (**Table 2**). In total 16 staphylococcal species and subspecies showed 16S rRNA gene identity > 98.7%. Considering the previously described cut-off value for species identification based on 16S rRNA gene sequence (98.7%) the five isolates were not designed to any of the previously described staphylococcal species (Stackebrandt and Ebers, 2006). **Figure 1** shows the reconstructed phylogenetic trees generated from alignment of the 16S rRNA gene sequence of the staphylococci type strains.

Genome-based phylogeny plays a central role in future taxonomy and phylogenetics of bacteria, and provides higher resolution than 16S rRNA gene phylogeny (Na et al., 2018). WGS comparisons were performed according to the recommended minimal standards for description of new staphylococcal species (Freney et al., 2009). We used the up-to-date bacterial core gene

set (UBCG)(Na et al., 2018), which produces an alignment based on 92 single-copy core genes extracted from whole genome sequences (WGS) of staphylococcal type strains available in GenBank (01.01.2010) (**Table 3**).The results from the phylogenomic tree confirmed that the five novel isolates are most closely related to *S. haemolyticus, S. petrasii*, *Staphylococcus devriesei* and *Staphylococcus hominis*. This further supports that the five *S. borealis* isolates belonging to a novel CONS-species forming their own well-supported branch (**Figure 2**).

We used overall genome related index (OGRI) methods (Chun et al., 2018) to calculate average nucleotide identity (ANI) and tetra using the online service JSpecies WS (Richter et al., 2015), The *in silico* DNA-DNA hybridization (dDDH) was calculated using the genome-to-genome distance calculator (GGDC) version 2.1 (Meier-Kolthoff et al., 2013). The GGDC results were based on the recommended formula 2 (sum of all identities found in high-scoring segment pairs (HSPs), divided by the overall HSP length), which is independent of genome size. A threshold value of 70 % DNA-DNA hybridization has long been established for defining bacterial species (Wayne et al., 1987). The results from all the ORGI methods confirm that the five *S. borealis* isolates belong to a novel species, which is related to, but distinctly different from *S. haemolyticus.* The OGRI values between the closest related staphylococcal type strains and the *S. borealis* strains is summarised in **Table 2, 4** and **5**.

Based on the WGS data we constructed core-genome SNP-based maximum likelihood (ML) trees of 169 *S. haemolyticus* isolates and the five *S. borealis* isolates, using KSNP3 package (Gardner et al., 2015). The ML tree, visualised with iTol, clearly demonstrates that the five *S. borealis* isolates form a distinct cluster separated from *S. haemolyticus* (**Figure 3**).

We also performed multilocus sequence typing (MLST) with the *S. haemolyticus* specific MLST-scheme (Cavanagh et al., 2012). We observed variations for each of the seven alleles between 23 and 79 SNPs, including gaps and insertions, further supporting the identification of a new species.

A pan-genome analysis was performed using 169 *S. haemolyticus* genome sequences (Pain et al., 2019) and the five proposed *S. borealis* isolates using Roary version 3.11.2 with default settings (Page et al., 2015). We identified 1,185 genes solely detected in the *S. borealis* isolates, from which 944 genes were identified in all five *S. borealis* isolates.

We identified antibiotic resistance genes (ARGs) using the Comprehensive Antibiotic Resistance Database (CARD)(http://arpcard.mcmaster.ca). All five isolates contained the vgaALC gene (98.53 % identity), a streptogramin A resistance variant first identified in *S. haemolyticus* (Novotna and Janata, 2006). Strains 51-48T (= CCUG 73747T = CECT 30011T) and 58-52 (= CCUG 73751) had the *ermC* gene while 58-52 (= CCUG 73751) also carried *fusC* which correlated with the observed phenotypic resistance profiles. No other antibiotic resistance genes were identified in the five *S. borealis* isolates by CARD.

Phenotypic tests and metabolic profiling

All five *S. borealis* isolates were Gram-stain positive cocci growing in clusters. They were non-motile on motility agar. All were oxidase negative, DNAse negative, coagulase negative, clumping factor negative, catalase positive and facultative anaerobic, determined by using the Brewer thioglycollate medium.

Scanning electron microscopy analyses were performed using a Zeiss Zigma scanning electronmicroscope (SEM) (Zeiss, Oberkochen, Germany). Whole cells were fixed with 2,5% glutaraldehyde and 4% formaldehyde in PHEM-buffer, before sedimentation onto poly-L-lysin coated coverslips. Samples were further processed according to the protocol of Cocchiaro using the Pelco Biowave (Ted Pella, Redding, USA)(Cocchiaro et al., 2008). Following dehydration samples were dried in a Leica EM CPD300 (Leica, Wetzlar, Germany) and mounted on SEM-stubs, gold/palladium was applied with a Polaron Range Sputter Coater (Newhaven, UK).

We tested temperature (4, 15, 30, 37, 42 and 45 ºC), NaCl tolerance (0, 0.5, 1.5, 3, 5, 7.5, 10 and 15%) according to the protocol by Freney *et al.* (Freney et al., 2009), using P agar plates (Götz et al., 2006). The haemolysis assay was performed on blood agar plates (Oxoid,Basingstoke, UK). *S. haemolyticus* CCUG 7323T was included as a reference strain in all tests. All five *S. borealis* isolates were able to grow at 30 to 42 ºC, and showed tolerance to NaCl up to 15%. All five *S. borealis* isolates displayed yellow pigmentation on P-agar plates (**Figure 4**). After 24 h of aerobic incubation on horse blood agar at 37°C, the *S. borealis* isolates formed smooth, circular, raised or slightly convex colonies reaching 3-5 mm in diameter. A clear β-haemolysis (2 mm) was observed in the *S. borealis* isolates and *S. haemolyticus* CCUG 7323T (1.5 mm) on horse blood agar plates.

Metabolic profiling of all *S. borealis* strains and *S. haemolyticus* CCUG 7323T reference strain was performed. The CCUG STX phenotypic worksheet was followed using the API bacterial identification systems API®Staph, API® 32 Staph and API® Coryne test (bioMérieux, Marcy-

l'Étoile, France)(www.ccug.se/identification/worksheets), following the instructions of the manufacturer. The metabolic profiles are summarized in **Table 6.** Biochemically, the five *S. borealis* isolates differed in four tests when compared with the *S. haemolyticus* CCUG 7323T reference strain. Briefly, the *S. borealis* strains were positive for fermentation of fructose and mannitol, as well as urease positive. Only the strain CCUG 73751 was positive for utilization of mannose.

Cell fatty acid-fatty acid methyl ester (CFA-FAME) analysis was done for the *S. borealis* strains. The strains were cultivated on Columbia Blood Agar Base plus 5% defibrinated horse blood (prepared by the Substrate Unit, Department of Clinical Microbiological, Sahlgrenska University Hospital), at 37°C, aerobically, for 24 h. The CFA-FAME profile was determined, using an HP 5890 gas chromatograph (Hewlett-Packard, Palo Alto, CA, USA) and a standardized protocol similar to the MIDI Sherlock MIS system (Sasser, 1990) as described previously (Zamora et al., 2012). CFA-FAME analysis of the five *S. borealis* strains determined the major CFAs to be the saturated fatty acids, C15:0 ISO (11%), C15:0 ANTEISO (63%) and C17:0 ANTEISO (13%), while other CFAs observed included C17:0 ISO (5%) and C18:0 (2,5%) (**Table 7**).

Antimicrobial resistance testing was performed using the disc diffusion method according the EUCAST guidelines (The European Committee on Antimicrobial Suceptibility Testing. Breakpoint tables for interpretation of MICs and zone diameters, version 9.0, 2019). Briefly, a 0,5 McFarland bacterial suspension was inoculated on Mueller Hinton agar plates(Oxoid, Basingstoke; UK). The plates were incubated at 37 ºC for 16-18 hours, and zones of inhibition were measured. All strains were susceptible to antimicrobial agents cefoxitin (30 ug), ciprofloxacin (5 ug), clindamycin (2 ug), gentamycin (10 ug), novobiocin (XX) linezolid (10 ug), rifampicin, tetracycline (30 ug), trimethoprim-sulfamethoxazole (1.25-23.75) and vancomycin (0.015-256ug). 51-48T (= CCUG 73747T = CECT 30011T) and 58-52 (= CCUG 73751) were resistant to erythromycin while 58-52 (= CCUG 73751) was also resistant to fucidic acid.

In conclusion, although the five *S. borealis* isolates share near identical 16S rRNA gene sequences to *S. haemolyticus* NCTC 11042T, and are phylogenetically closest related to *S. haemolyticus,* there are strong phenotypic and genomic justifications for assigning the novel isolates to a novel species of the genus *Staphylococcus,* for which the name *Staphylococcus borealis* sp. nov is proposed.

These justifications are:

1) Phylogenetic distance, ANI < 95% and inferred DDH < 70 %.
2) Pigmented phenotype
3) Production of urease
4) Pangenome comparison

Description of *Staphylococcus borealis* sp. nov.

S. borealis (bo.re.a'.lis. L. masc. adj. *borealis* related to the North, boreal)

Colonies are 3-5 mm in diameter, round, smooth and have a yellow tint. The difference in pigmentation between typical *S. haemolyticus* and *S. borealis* is particularly evident on different supplemented p-agars (non-supplemented, full fat milk and horse blood) after 48 h at 37ºC. Cells are Gram-stain positive, coccoid, 650 nm to 1.23 µm in diameter and form clusters. Facultative anaerobic. Cells are coagulase negative and catalase positive. Biochemically negative for fructose, mannitol and positive for production of urease. The major fatty acids are C15:0 ISO, C15:0 ANTEISO and C17:0 ANTEISO, while C17:0 ISO and C18:0 are present in lower amount.

The type strain, 51-48^T (= CCUG 73737^T = CECT 30011^T), was isolated from blood culture in Tromsø, Norway in 1997.

Figure legends

Figure 1. Phylogenetic relationship of Staphylococcal type strains based on 16S rRNA phylogenetic tree. The phylogenetic tree was generated from MUSCLE alignment of the 16S rRNA sequences of Staphylococcal type strains and the type strain of *Macrococcus caseolyticus*. The maximum likelihood method was used and bootstrapping based on 500 replicated was performed using the RAxML software (Stamatakis, 2014). rRNA sequences were obtained from ezbiocloud(Yoon et al., 2017) (Type strain ID listed in tree, accession listed in table 1).

Figure 2. Phylogenetic relationship of Staphylococcal type strains based on core genes. A maximum-likelihood phylogenetic tree was produced based on the alignment of 92 single-copy core genes, utilized by UBCG. *Macrococcus caseolyticus* was used for rooting the tree.

Figure 3. SNP-based phylogenetic tree of 169 *S. haemolyticus* isolates and the five newly proposed *S. borealis* isolates, using the KSNP3 program.

Figure 4. Yellow pigmentation shown on P agar of the five *S. borealis* isolates in comparison to *S. haemolyticus* CCUG 7323T.

References

Becker, K., Heilmann, C., and Peters, G. (2014). Coagulase-negative staphylococci. *Clin. Microbiol. Rev.* 27, 870–926. doi:10.1128/CMR.00109-13.

Cavanagh, J. P., Hjerde, E., Holden, M. T. G., Kahlke, T., Klingenberg, C., Flægstad, T., et al. (2014). Whole-genome sequencing reveals clonal expansion of multiresistant Staphylococcus haemolyticus in European hospitals. *J. Antimicrob. Chemother.* 69, 2920–2927. doi:10.1093/jac/dku271.

Cavanagh, J. P., Klingenberg, C., Hanssen, A. M., Fredheim, E. A., Francois, P., Schrenzel, J., et al. (2012). Core genome conservation of Staphylococcus haemolyticus limits sequence based population structure analysis. *J. Microbiol. Methods* 89, 159–166. doi:10.1016/j.mimet.2012.03.014.

Cavanagh, J. P., Wolden, R., Heise, P., Esaiassen, E., Klingenberg, C., and Aarag Fredheim, E. G. (2016). Antimicrobial susceptibility and body site distribution of community isolates of coagulase-negative staphylococci. *Apmis* 124, 973–978. doi:10.1111/apm.12591.

Chun, J., Oren, A., Ventosa, A., Christensen, H., Arahal, D. R., da Costa, M. S., et al. (2018). Proposed minimal standards for the use of genome data for the taxonomy of prokaryotes. *Int. J. Syst. Evol. Microbiol.* 68, 461–466. doi:10.1099/ijsem.0.002516.

Cocchiaro, J. L., Kumar, Y., Fischer, E. R., Hackstadt, T., and Valdivia, R. H. (2008). Cytoplasmic lipid droplets are translocated into the lumen of the Chlamydia trachomatis parasitophorous vacuole. *Proc. Natl. Acad. Sci. U. S. A.* 105, 9379–9384. doi:10.1073/pnas.0712241105.

Freney, J., Kloos, W. E., Hajek, V., Webster, J. A., Bes, M., Brun, Y., et al. (2009). Recommended minimal standards for description of new staphylococcal species. *Int. J. Syst. Bacteriol.* 49, 489–502. doi:10.1099/00207713-49-2-489.

Gardner, S. N., Slezak, T., and Hall, B. G. (2015). kSNP3.0: SNP detection and phylogenetic analysis of genomes without genome alignment or reference genome. *Bioinformatics* 31, 2877–2878. doi:10.1093/bioinformatics/btv271.

Götz, F., Bannerman, T., and Schleifer, K. (2006). "The genera Staphylococcus and Macrococcus," in *The Prokayotes, vol 4: Bacteria: Firmicutes, Cyanobacteria*, eds. M. Dworkin, S. Falkow, E. Rosenberg, K. Schleifer, and E. Stackebrandt (New York, NY: Springer US), 5–75.

Hunt, M., Silva, N. De, Otto, T. D., Parkhill, J., Keane, J. A., and Harris, S. R. (2015). Circlator: Automated circularization of genome assemblies using long sequencing reads.

Genome Biol. 16, 1–10. doi:10.1186/s13059-015-0849-0.

Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., and Phillippy, A. M. (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* 27, 722–736. doi:10.1101/gr.215087.116.

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *Preprint* 00, 1–3. Available at: http://arxiv.org/abs/1303.3997.

Meier-Kolthoff, J. P., Auch, A. F., Klenk, H. P., and Göker, M. (2013). Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics* 14. doi:10.1186/1471-2105-14-60.

Na, S. I., Kim, Y. O., Yoon, S. H., Ha, S. min, Baek, I., and Chun, J. (2018). UBCG: Up-to-date bacterial core gene set and pipeline for phylogenomic tree reconstruction. *J. Microbiol.* 56, 281–285. doi:10.1007/s12275-018-8014-6.

Novotna, G., and Janata, J. (2006). A new evolutionary variant of the streptogramin A resistance protein, Vga(A)LC, from Staphylococcus haemolyticus with shifted substrate specificity towards lincosamides. *Antimicrob. Agents Chemother.* 50, 4070–4076. doi:10.1128/AAC.00799-06.

Page, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T. G., et al. (2015). Roary: Rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 31, 3691–3693. doi:10.1093/bioinformatics/btv421.

Pain, M., Hjerde, E., Klingenberg, C., and Cavanagh, J. P. (2019). Comparative genomic analysis of Staphylococcus haemolyticus reveals key to hospital adaptation and pathogenicity. *Front. Microbiol.* 10, 1–13. doi:10.3389/fmicb.2019.02096.

Richter, M., Rosselló-Móra, R., Oliver Glöckner, F., and Peplies, J. (2015). JSpeciesWS: A web server for prokaryotic species circumscription based on pairwise genome comparison. *Bioinformatics* 32, 929–931. doi:10.1093/bioinformatics/btv681.

Sasser, M. (1990). Identification of bacteria by gas chromatography of cellular fatty acids. *MIDI - Tech. note #101*, 1–6. Available at: http://natasha.eng.usf.edu/gilbert/courses/Biotransport Phenomena/pdf/bacteria_gc_1.pdf.

Stackebrandt, E., and Ebers, J. (2006). Taxonomic parameter revised: tarnishes gols standards. *Microbiol. Today*, 153–155.

Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi:10.1093/bioinformatics/btu033.

The European Committee on Antimicrobial Suceptibility Testing. Breakpoint tables for

interpretation of MICs and zone diameters, version 9.0, 2019.

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., et al. (2014). Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9. doi:10.1371/journal.pone.0112963.

Wayne, L. G., Brenner, D. J., Colwell, R. R., Grimont, P. A. D., Kandler, O., Krichevsky, M. I., et al. (1987). Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. *Int. J. Syst. Evol. Microbiol.* 37, 463–464. doi:10.1016/S0723-2020(88)80020-1.

Yoon, S. H., Ha, S. M., Kwon, S., Lim, J., Kim, Y., Seo, H., et al. (2017). Introducing EzBioCloud: A taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. *Int. J. Syst. Evol. Microbiol.* 67, 1613–1617. doi:10.1099/ijsem.0.001755.

Zamora, L., Fernández-Garayzábal, J. F., Svensson-Stadler, L. A., Palacios, M. A., Domínguez, L., Moore, E. R. B., et al. (2012). Flavobacterium oncorhynchi sp. nov., a new species isolated from rainbow trout (Oncorhynchus mykiss). *Syst. Appl. Microbiol.* 35, 86–91. doi:10.1016/j.syapm.2011.11.007.

Table 1. Overview of genomic information for all five *S. borealis* strains.

Isolate ID	Genome size	Contigs	CDS	rRNA	tmRNA	tRNA	GC%
51-48T **(=CCUG 73747T = CECT 30011T)**	2,664,964 bp	60	2,556	2	1	44	33.64 %
57-14**(=CCUG 73748 = CECT 30010)**	2,631,744 bp	37	2,495	10	1	60	33.68 %
57-74 **(=CCUG 73749)**	2,621,226 bp	42	2,486	10	1	61	33.69 %
58-22 **(=CCUG 73750)**	2,671,705 bp	41	2,553	10	1	61	33.72 %
58-52 **(=CCUG 73751)**	2,760,219 bp	31	2,509	10	1	60	33.57 %

Table 2. Overview of result for species identity of the closest related *Staphylococcus* species and subspecies, compared to the proposed type strain of *S. borealis* 51-48T.

Staphylococcal type strains	16S rRNA (98.7%)	ANIb (<95%)	ANIm (<95%)	Tetra (<0.99)	dDDH (<70 %)
***S. devriesei* NCTC 13828T**	99.25	79.88	84.90	0.95471	23.5
***S. petrasii* subsp. *petrasii* CCM8418T**	99.39	79.92	85.46	0.96338	23.3
***S. petrasii* subsp *croceolyticus* CCM8421T**	99.39	80.24	85.68	0.96603	23.5
***S. petrasii* subsp. *jettensis* SEQ110T**	99.51	80.14	85.59	0.96258	23.6
***S. petrasii* subsp. *pragensis* NRL/St 12/356T**	99.46	80.32	85.70	0.96993	23.6
***S. hominis* subsp. *hominis* DSM 20328T**	99.25	78.25	85.10	0.95276	22.3
***S. hominis* subsp. *novobiosepticus* GTC 1228T**	98.83	78.51	85.30	0.95520	22.6

Table 3: Overview of the accession numbers of staphylococcal type strains used in the genomic analysis

Species	Subpecies	Type strain used	WGS accession	16S accession
S. agnetis		DSM 23656		HM484980
S. argensis		DSM 29875		PPPX01000013
S. argenteus		MSHR1132		FR821777
S. arlettae		NCTC 12413		AB009933
S. aureus	*aureus*	NCTC 8325		AMYL01000007
	anaerobius	DSM 20714		D83355
S. auricularis		DSM 20609		L37598
S. caeli		82B		MH431939
S.capitis	*capitis*	NCTC 11045		L37599
	urealyticus	DSM 6717		AB233325
S. caprae		NCTC 12196		AB009935
S. carnosus	*carnosus*	-		UHCT01000001
	utilis	-		AB233329
S. chromogenes		NCTC 10530		D83360
S. cohnii	*cohnii*	NCTC 11041		D83361
	urealyticus	DSM 6718		AB009936
S. condimenti		DSM 11674		CP015114
S. cornubiensis		NW1		-
S. debuckii		-		MK121623
S. delphini		NCTC 12225		AB009938
S. devriesei		CCUG 58238		UHCZ01000002
S. edaphicus		CMM 8730		KY315825
S. epidermidis		ATCC 14990		UHDF01000003
S. equorum	*equorum*	NCTC 12414		AB009939
	linens	DSM 15097		AF527483
S. felis		ATCC 49168		D83364
S. fleuretti		NCTC 13829??		UHDL01000001
S. gallinarum		DSM 20610		D83366
S. haemolyticus		NCTC 11042		LILF01000056
S. hominis	*hominis*	NCTC 11320		X66101
	novobiosepticus	CCUG 42399		AB233326
S. hyicus		ATCC 11249		CP008747
S. intermedius		NCTC 11048		CAIB01000045
S. kloosi		ATCC 43959		AB009940
S. lentus		NCTC 12102		D83370
S. lugdunensis		NCTC 12217		AB009941
S. lutrae		ATCC 700373		CP020773
S. massiliensis		CCUG 55927		EU707796
S. microti		DSM 22147		UHDT01000001
S. muscae		ATCC 49910		FR733703
S. nepalensis		DSM 15150		UHDS01000001
S. pasteuri		-		AF041361
S. petrasii	*petrasii*	CCM 8418		JX139845
	croceilyticus	CCM 8421		AY953148
	jettensis	CCM 8494		JN092118
	pragensis	CCM 8529		KM873669
S. pettenkoferi		CCUG 51270		AF322002
S. piscifermentans		NCTC 13836		AB009943
S. pseudintermedius		CCUG 49543		AJ780976
S. pseudoxylosus				MH643903
S. rostri		DSM 21968		FM242137
S. saccharolyticus		NCTC 11807		L37602
S. saprophyticus	*saprophyticus*	ATCC 15306		AP008934
	bovis	CCUG 38042		AB233327
S. schleiferi	*schleiferi*	-		AB009945
	coagulans	-		AB233334
S. schweitzeri		DSM 28300		CCEL01000025
S. sciuri		NCTC 12103		AJ421446
S. simiae		CCUG 51256		LT906460
S. simulans		NCTC 11046		D83373
S. stepanovicii		DSM 26319		LT906462
S. succinus	*succinus*	DSM 14617		AF004220
	casei	DSM 15096		AJ320272
S. vitulinus		DSM 15615		AB009946
S. warneri		NCTC 11044		L37603
S. xylous		ATCC 29971		MRZO01000018

Table 4: Result for species delineation of *S. borealis* against *S. haemolyticus* NCTC 11042T using 16S RNA gene sequence, ANI, tetra and dDDH comparisons. Value in [] is percent alignment between *S. haemolyticus* NCTC 11042T and the different *S. borealis* isolates.

Isolate	16s rRNA gene (98.7%)	ANIb (<95-96%)	ANIm (<95-96%)	Tetra (<0.99)	dDDH (<70 %)
51-48T	99.86 %	87.37 % [78.70]	88.53 % [79.92]	0.98451	34.1 %
57-14	99.86 %	87.55 % [80.45]	88.58 % [81.95]	0.98428	34.2 %
57-74	99.86 %	87.58 % [80.05]	88.60 % [82.31]	0.98446	34.2 %
58-22	99.93 %	87.50 % [78.52]	88.57 % [80.43]	0.98413	34.2 %
58-52	99.93 %	87.52 % [79.24]	88.61 % [81.35]	0.98451	34.3 %

Table 5: Results for species delineation between the *S. borealis* isolates. Percent ANIb (based on blast+) scores and [aligned nucleotides].

	***S. haemolyticus* NCTC11042T**	***S. borealis* 51-48T**	***S. borealis* 57-14**	***S. borealis* 57-74**	***S. borealis* 58-22**
***S. borealis* 51-48T**	87.37 [78.70]	--			
***S. borealis* 57-14**	87. 55 [80.45]	99.59 [93.90]	--		
S. borealis 57-74	87.58 [80.05]	99.60 [94.97]	99.83 [95.81]	--	
***S. borealis* 58-22**	87.50 [78.52]	99.54 [94.32]	99.75 [96.15]	99.84 [94.99]	--
***S. borealis* 58-52**	87.52 [79.24]	99.54 [93.79]	99.67 [94.90]	99.71 [95.26]	99.75 [95.66]

Table 6. Biochemical tests, based on API 32 STAP, API STAPH and API CORYNE.

Culture Collection University of Gothenburg (CCUG)		**73747T**	**73748**	**73749**	**73750**	**73751**	**7323T**
Local strain identification number		**51-48**	**57-14**	**57-74**	**58-22**	**58-52**	**63-42**
Glucose	**GLU**	+	+	+	+	+	+
Fructose	**FRU**	+	+	+	+	+	-
Arabinose	**ARA**	-	-	-	-	-	-
Ribose	**RIB**	+	+	+	+	+	+
Mannose	**MNE**	-	-	-	-	+	-
Xylose	**XYL**	-	-	-	-	-	-
Sucrose	**SAC**	+	+	+	+	+	+
Lactose	**LAC**	-	-	-	-	-	-
Turanose	**TUR**	-	+	+	-	-	-
Cellobiose	**CEL**	-	-	-	-	-	-
Maltose	**MAL**	+	+	+	+	+	+
Trehalose	**TRE**	+	+	+	+	+	+
Melibiose	**MEL**	-	-	-	-	-	-
Raffinose	**RAF**	-	-	-	-	-	-
Glycogen	**GLYG**	-	-	-	-	-	-
N-acetyl-glucosamine	**NAG**	+	-	+	+	-	+
MethyL-α-D-glucopiranoside	**MDG**	-	+	+	+	+	-
Mannitol	**MAN**	+	+	+	+	+	-
Xylitol	**XLT**	-	-	-	-	-	-
Nitrate	**NIT**	+	+	+	+	+	+
Acetoin Production	**VP**	-	-	+	-	+	+
Novobiocin	**NOVO**	-	-	-	-	-	-
Gelatine	**GEL**	-	-	-	-	-	-
Esculin	**ESC**	+	-	+	-	+	-
Catalase	**CAT**	+	+	+	+	+	+
Urease	**URE**	+	+	+	+	+	-
N-acetyl-β-Glucosaminidase	**βNAG**	-	-	-	-	-	-
α-glucosidase	**αGLU**	-	-	-	-	-	-
β-galactosidase	**βGAL**	-	-	-	-	-	-
β-glucuronidase	**βGUR**	+	-	-	+	-	+
Alkaline phosphatase	**PAL**	+	-	-	+	-	+
Pyrazinamidase	**PYZ**	+	+	+	+	+	+
Arginine arylamidase	**ArgA**	-	-	-	-	-	-
Pyrrolidonyl arylamidase	**PyrA**	+	+	+	+	+	+
Ornithine decarboxilase	**ODC**	-	-	-	-	-	-
Arginine dihydrolase	**ADH**	+	+	+	+	+	+

Table 7. Cell fatty acid-fatty acid methyl ester (CFA-FAME) analysis of *S. borealis* strains (CCUG 73747T, CCUG 73748, CCUG 73749, CCUG 73750, CCUG 73751), showing the peak ID, name of CFAs and the Area per peak (%). Also shown are the corresponding CFA profiles of *S. aureus* (CCUG 1800T), *S. devriesei* (CCUG 58238T) and *haemolyticus* (CCUG 7323T). tr denotes "trace" which means a peak has been recorded, but too small to be integrated.

Peak ID	**38**	**59**	**60**	**77**	**86**	**101**	**102**	**124**	**125**	**126**	**131**	**144**	**145**	**164**
Peak name of CFA	**C14:0 ISO**	**C15:0 ISO**	**C15:0 ANTESIO**	**C16:0 ISO**	**C16:0**	**C17:0 ISO**	**C17:0 ANTESIO**	**C18:0 ISO**	**Summed feature***	**18:1 ω9c**	**C18:0**	**C19:0**	**C19:0 ANTESIO**	**C20:0**
Spp	**Area % per Peak**													
***S. aureus* CCUG1800T**	0,5	7,5	55,0	2,0	2,0	4,0	19,5	tr	3,5	0,0	3,0	1,0	1,5	0,5
***S. devriesei* CCUG 58238T**	0,0	3,0	52,0	1,0	1,0	5,0	26,0	0,0	2,5	1,5	2,5	1,0	4,0	0,5
***S. haemolyticus* CCUG 7323T**	1,0	7,0	42,7	1,0	1,8	6,8	19,0	0,0	1,0	0,0	8,5	3,0	5,7	2,5
***S. borealis* CCUG 73747T**	1,0	10,0	62,0	1,0	1,0	5,0	13,0	0,0	0,5	1,0	2,0	1,0	2,0	0,5
***S. borealis* CCUG 73748**	tr	12,0	64,3	tr	1,2	5,6	13,1	0,0	0,0	tr	2,4	tr	1,6	0,0
***S. borealis* CCUG 73749**	1,1	10,9	65,6	tr	tr	4,6	12,2	0,0	tr	tr	2,1	1,3	2,3	tr
***S. borealis* CCUG 73750**	1,2	11,0	61,0	1,0	1,2	5,5	13,3	0,0	0,7	tr	2,5	1,0	1,7	tr
***S. borealis* CCUG 73751**	1,3	11,7	59,0	1,0	2,0	5,3	13,8	0,0	1,0	tr	4,0	0,0	1,0	tr

Figure 1

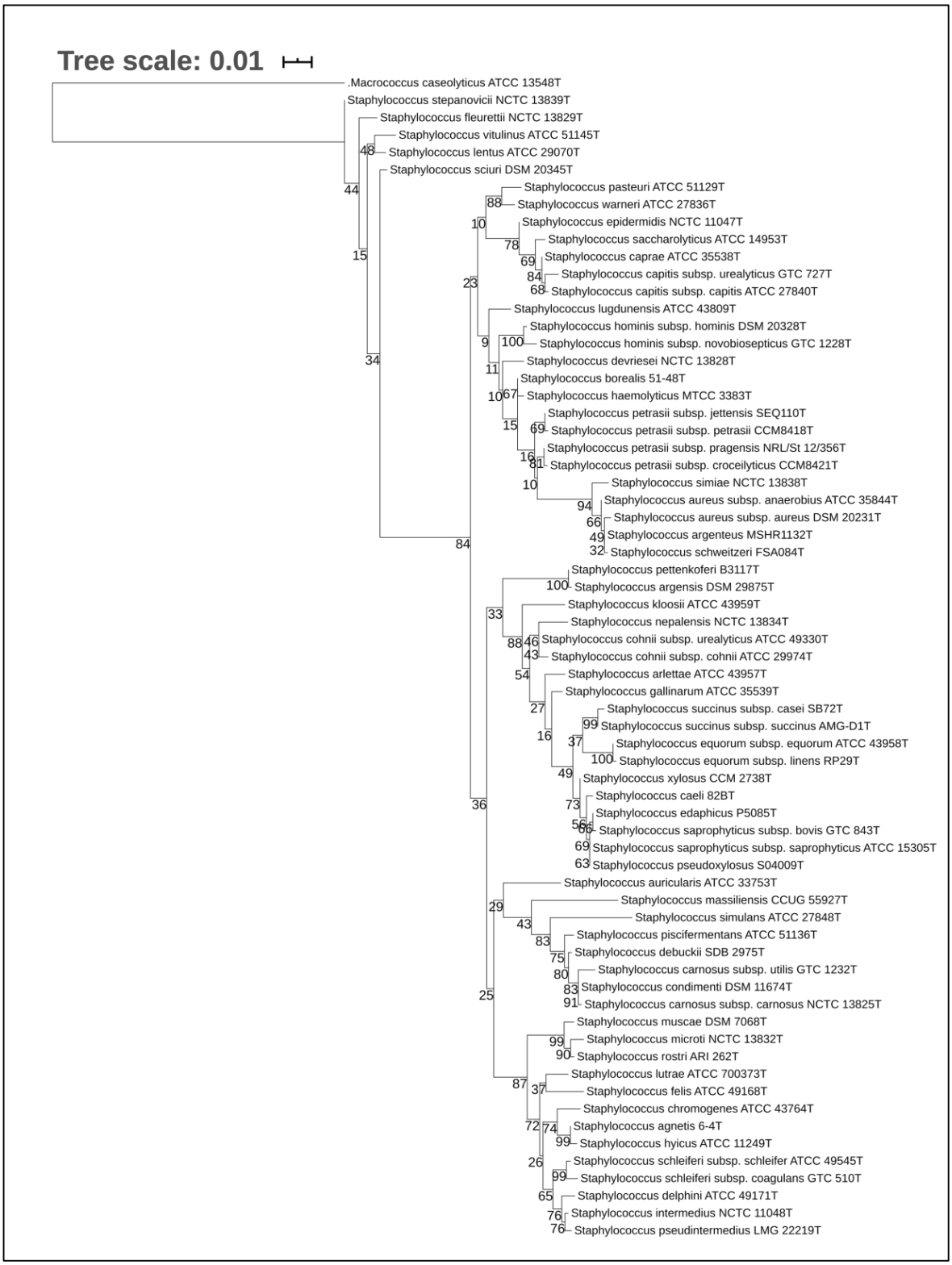

Tree scale: 0.01
.Macrococcus caseolyticus ATCC 13548T
Staphylococcus stepanovicii NCTC 13839T
Staphylococcus fleurettii NCTC 13829T
Staphylococcus vitulinus ATCC 51145T
Staphylococcus lentus ATCC 29070T
Staphylococcus sciuri DSM 20345T
Staphylococcus pasteuri ATCC 51129T
Staphylococcus warneri ATCC 27836T
Staphylococcus epidermidis NCTC 11047T
Staphylococcus saccharolyticus ATCC 14953T
Staphylococcus caprae ATCC 35538T
Staphylococcus capitis subsp. urealyticus GTC 727T
Staphylococcus capitis subsp. capitis ATCC 27840T
Staphylococcus lugdunensis ATCC 43809T
Staphylococcus hominis subsp. hominis DSM 20328T
Staphylococcus hominis subsp. novobiosepticus GTC 1228T
Staphylococcus devriesei NCTC 13828T
Staphylococcus borealis 51-48T
Staphylococcus haemolyticus MTCC 3383T
Staphylococcus petrasii subsp. jettensis SEQ110T
Staphylococcus petrasii subsp. petrasii CCM8418T
Staphylococcus petrasii subsp. pragensis NRL/St 12/356T
Staphylococcus petrasii subsp. croceilyticus CCM8421T
Staphylococcus simiae NCTC 13838T
Staphylococcus aureus subsp. anaerobius ATCC 35844T
Staphylococcus aureus subsp. aureus DSM 20231T
Staphylococcus argenteus MSHR1132T
Staphylococcus schweitzeri FSA084T
Staphylococcus pettenkoferi B3117T
Staphylococcus argensis DSM 29875T
Staphylococcus kloosii ATCC 43959T
Staphylococcus nepalensis NCTC 13834T
Staphylococcus cohnii subsp. urealyticus ATCC 49330T
Staphylococcus cohnii subsp. cohnii ATCC 29974T
Staphylococcus arlettae ATCC 43957T
Staphylococcus gallinarum ATCC 35539T
Staphylococcus succinus subsp. casei SB72T
Staphylococcus succinus subsp. succinus AMG-D1T
Staphylococcus equorum subsp. equorum ATCC 43958T
Staphylococcus equorum subsp. linens RP29T
Staphylococcus xylosus CCM 2738T
Staphylococcus caeli 82BT
Staphylococcus edaphicus P5085T
Staphylococcus saprophyticus subsp. bovis GTC 843T
Staphylococcus saprophyticus subsp. saprophyticus ATCC 15305T
Staphylococcus pseudoxylosus S04009T
Staphylococcus auricularis ATCC 33753T
Staphylococcus massiliensis CCUG 55927T
Staphylococcus simulans ATCC 27848T
Staphylococcus piscifermentans ATCC 51136T
Staphylococcus debuckii SDB 2975T
Staphylococcus carnosus subsp. utilis GTC 1232T
Staphylococcus condimenti DSM 11674T
Staphylococcus carnosus subsp. carnosus NCTC 13825T
Staphylococcus muscae DSM 7068T
Staphylococcus microti NCTC 13832T
Staphylococcus rostri ARI 262T
Staphylococcus lutrae ATCC 700373T
Staphylococcus felis ATCC 49168T
Staphylococcus chromogenes ATCC 43764T
Staphylococcus agnetis 6-4T
Staphylococcus hyicus ATCC 11249T
Staphylococcus schleiferi subsp. schleifer ATCC 49545T
Staphylococcus schleiferi subsp. coagulans GTC 510T
Staphylococcus delphini ATCC 49171T
Staphylococcus intermedius NCTC 11048T
Staphylococcus pseudintermedius LMG 22219T

Figure 2

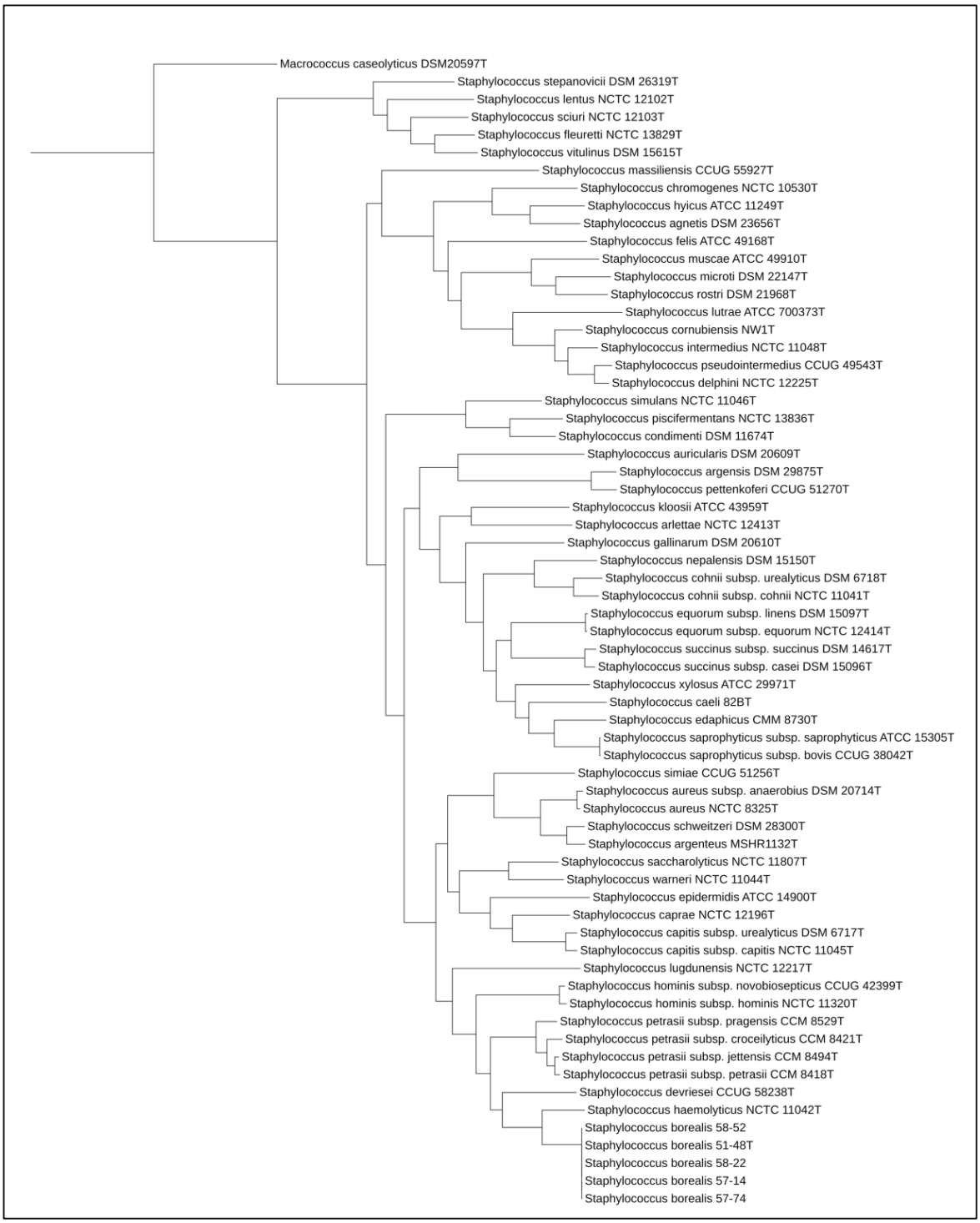

Macrococcus caseolyticus DSM20597T
Staphylococcus stepanovicii DSM 26319T
Staphylococcus lentus NCTC 12102T
Staphylococcus sciuri NCTC 12103T
Staphylococcus fleuretti NCTC 13829T
Staphylococcus vitulinus DSM 15615T
Staphylococcus massiliensis CCUG 55927T
Staphylococcus chromogenes NCTC 10530T
Staphylococcus hyicus ATCC 11249T
Staphylococcus agnetis DSM 23656T
Staphylococcus felis ATCC 49168T
Staphylococcus muscae ATCC 49910T
Staphylococcus microti DSM 22147T
Staphylococcus rostri DSM 21968T
Staphylococcus lutrae ATCC 700373T
Staphylococcus cornubiensis NW1T
Staphylococcus intermedius NCTC 11048T
Staphylococcus pseudointermedius CCUG 49543T
Staphylococcus delphini NCTC 12225T
Staphylococcus simulans NCTC 11046T
Staphylococcus piscifermentans NCTC 13836T
Staphylococcus condimenti DSM 11674T
Staphylococcus auricularis DSM 20609T
Staphylococcus argensis DSM 29875T
Staphylococcus pettenkoferi CCUG 51270T
Staphylococcus kloosii ATCC 43959T
Staphylococcus arlettae NCTC 12413T
Staphylococcus gallinarum DSM 20610T
Staphylococcus nepalensis DSM 15150T
Staphylococcus cohnii subsp. urealyticus DSM 6718T
Staphylococcus cohnii subsp. cohnii NCTC 11041T
Staphylococcus equorum subsp. linens DSM 15097T
Staphylococcus equorum subsp. equorum NCTC 12414T
Staphylococcus succinus subsp. succinus DSM 14617T
Staphylococcus succinus subsp. casei DSM 15096T
Staphylococcus xylosus ATCC 29971T
Staphylococcus caeli 82BT
Staphylococcus edaphicus CMM 8730T
Staphylococcus saprophyticus subsp. saprophyticus ATCC 15305T
Staphylococcus saprophyticus subsp. bovis CCUG 38042T
Staphylococcus simiae CCUG 51256T
Staphylococcus aureus subsp. anaerobius DSM 20714T
Staphylococcus aureus NCTC 8325T
Staphylococcus schweitzeri DSM 28300T
Staphylococcus argenteus MSHR1132T
Staphylococcus saccharolyticus NCTC 11807T
Staphylococcus warneri NCTC 11044T
Staphylococcus epidermidis ATCC 14900T
Staphylococcus caprae NCTC 12196T
Staphylococcus capitis subsp. urealyticus DSM 6717T
Staphylococcus capitis subsp. capitis NCTC 11045T
Staphylococcus lugdunensis NCTC 12217T
Staphylococcus hominis subsp. novobiosepticus CCUG 42399T
Staphylococcus hominis subsp. hominis NCTC 11320T
Staphylococcus petrasii subsp. pragensis CCM 8529T
Staphylococcus petrasii subsp. croceilyticus CCM 8421T
Staphylococcus petrasii subsp. jettensis CCM 8494T
Staphylococcus petrasii subsp. petrasii CCM 8418T
Staphylococcus devriesei CCUG 58238T
Staphylococcus haemolyticus NCTC 11042T
Staphylococcus borealis 58-52
Staphylococcus borealis 51-48T
Staphylococcus borealis 58-22
Staphylococcus borealis 57-14
Staphylococcus borealis 57-74

Figure 3

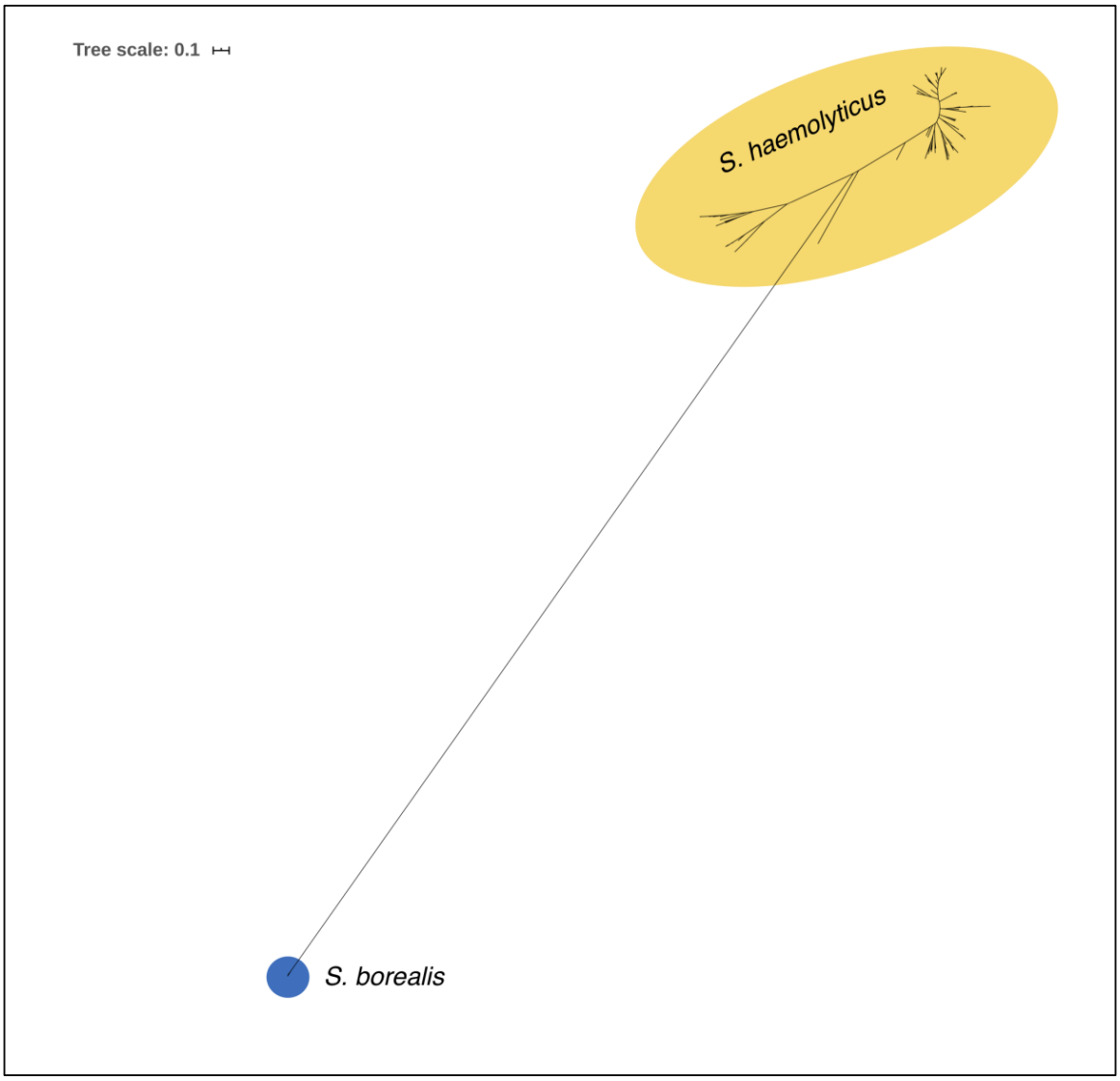
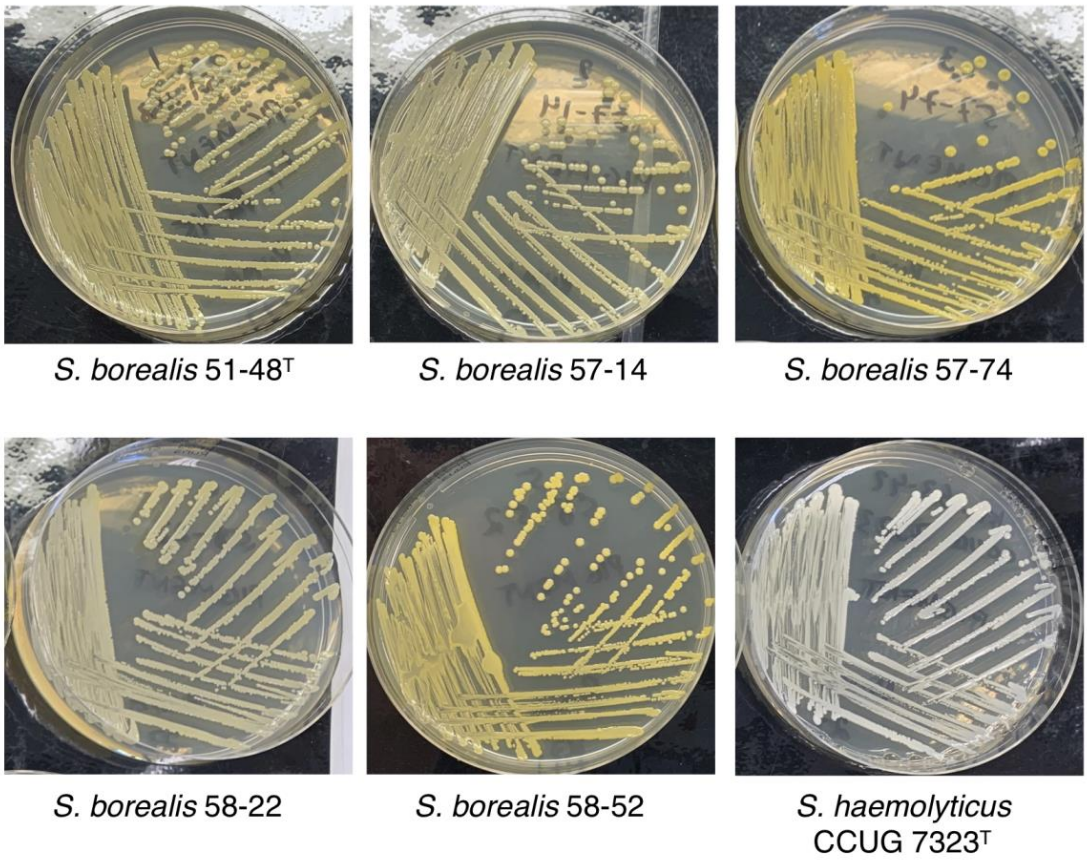


Figure 4



S. borealis 51-48^{T} | *S. borealis* 57-14 | *S. borealis* 57-74

S. borealis 58-22 | *S. borealis* 58-52 | *S. haemolyticus* CCUG 7323^{T}