

Aus dem Institut für Mikrobiologie und Hygiene

Klinische Medizin

der Medizinischen Fakultät

der Universität des Saarlandes, Homburg/Saar

**Next Generation Sequencing und  
DNA-Microarray-Analyse von invasiven und  
kolonisierenden Methicillin-resistenten  
*Staphylococcus aureus*-Stämmen aus dem  
klonalen Komplex 5**

***Dissertation zur Erlangung des Grades eines Doktors der Medizin  
der Medizinischen Fakultät***

der UNIVERSITÄT DES SAARLANDES

2015

vorgelegt von: Jan-Hendrik Brink

geb. am: 15.12.1987 in Braunschweig

## Zusammenfassung

Die klinische Bedeutung des Methicillin-resistenten Bakteriums *Staphylococcus aureus* (MRSA) ist insbesondere als Erreger nosokomialer Krankheiten seit Jahren hoch und weiterhin wachsend. Für MRSA besteht eine Vielzahl an unterschiedlichen Typisierungsmethoden. In jüngster Zeit sind die beiden molekularen Methoden DNA-Microarray sowie Vollgenomsequenzierung (engl. whole genome sequencing oder kurz WGS) hinzugekommen, welche ein immer höheres Diskriminierungspotential aufweisen und unterschiedliche individuelle Vorzüge besitzen. In dieser Arbeit wurden diese beiden Methoden genutzt, um eine definierte Kohorte von 12 kolonisierenden und 15 invasiven MRSA Isolaten, welche an der Universität des Saarlands gesammelt wurden, retrospektiv zu untersuchen. In die Studie eingeschlossen wurden Isolate aus dem klonalen Komplex 5 mit dem spa-Typ t504 und t003. Eine bioinformatische Datenanalyse wurde genutzt, um genetische Unterschiede zwischen den Gruppen zu identifizieren und Fragen bezüglich der Phylogenie und der Virulenz zu beantworten.

Als Ergebnis der Abstammungsanalyse anhand von Kerngenom-SNP-Daten ergibt sich, dass alle Stämme mit dem spa-Typ t504 sowie eine weitere Gruppe von miteinander eng verwandten t003-Stämmen jeweils eigenständige Kladen lokalen Ursprungs bilden, welche sich von den bekannten deutschlandweit verbreiteten übrigen t003-Stämmen abgrenzen.

Die WGS-Analyse hat darüber hinaus das Gen *ebhB* (zellwandassoziiertes Oberflächenprotein) mit Virulenz in Verbindung gebracht. Innerhalb von *ebhB* konnten bei 7 von 12 invasiven Isolaten 7 verschiedene genetische Variationen gefunden werden, während in den kolonisierenden Isolaten keine Mutationen auftraten.

Die Microarray-Genprofile identifizierten eine Untergruppe von 5 invasiven, für das Gen *hdsxCC15* positiven, Isolaten. Darüber hinaus konnten die Microarray-Profile aber nicht mit der Phylogenie der untersuchten Kohorte assoziiert werden.

Es zeigt sich, dass die Analyse des Kerngenoms mittels WGS den herkömmlichen Techniken (spa-Typisierung, DNA-Microarray) überlegen ist, wenn es darum geht, die Phylogenie und Virulenz von MRSA-Isolaten zu erfassen. Für die Zukunft ist es daher wichtig, die Datenauswertung des WGS weiter zu entwickeln und zu standardisieren.

## Summary

The clinical importance of the methicillin resistant bacterium *Staphylococcus aureus* (MRSA) as a pathogen responsible for nosocomial diseases has been increasing enormously for years and is still growing. There is a large variety of different typing methods available. Recently, molecular methods such as DNA microarray and whole genome sequencing (WGS) have been developed, which show an increased discrimination potential and possess different individual advantages. In this study, these two methods were applied in order to retrospectively investigate a defined cohort of 12 colonizing and 15 invasive MRSA isolates, collected at the University of Saarland. Included in the study were isolates of the clonal complex 5 with spa type t504 and t003. Bioinformatic data analysis was employed to identify genetic differences between the groups and to answer questions on phylogeny and virulence.

The phylogenetic analysis based on core genome SNP data leads to the result that all strains with spa type t504, as well as a further group of closely related t003 strains, form independent clades of local origin, which can be clearly distinguished from the well-known remaining, widely-spread German t003 strains. Furthermore, the whole genome analysis could associate gene *ebhB* (cell wall associated surface protein) with virulence. Within *ebhB*, different genetic variations could be detected for 7 out of 12 invasive isolates, while no such mutations occurred in the colonizing isolates.

The microarray gene profiles identified a sub-group of 5 invasive isolates positive for gene *hsdSxCC15*. However, the microarray profiles could not be associated with the phylogeny of the cohort under investigation.

It has been demonstrated that the analysis of the core genome via WGS is superior to the established methods (protein A-typing, DNA-microarray) for investigating the phylogeny and virulence of closely related MRSA isolates. The task ahead will be to further develop and standardize the WGS data-analysis.

# Inhaltsverzeichnis

1	Einleitung .....	1
	1.1.1 Methicillin-Resistenz.....	2
	1.1.2 Klinische Bedeutung von <i>Staphylococcus aureus</i> .....	3
1.1	<i>Staphylococcus aureus</i> : Mikrobiologie und Taxonomie.....	1
	1.2.1 Pulsfeld-Gelelektrophorese .....	5
1.2	Typisierungsmethoden für <i>Staphylococcus aureus</i> .....	4
	1.2.2 Multilocus Sequenz-Typisierung .....	6
	1.2.3 Protein A–Typisierung.....	7
	1.2.4 DNA-Microarray .....	8
	1.2.5 Vollgenomsequenzierung.....	9
1.3	Das Genom von <i>Staphylococcus aureus</i> .....	10
	1.3.1 Virulenzfaktoren von <i>Staphylococcus aureus</i> .....	12
1.4	Zielsetzung .....	14
2	<b>Material und Methoden .....</b>	<b>15</b>
2.1	Material.....	15
2.2	Studienisolate .....	15
2.3	Protein A-Typisierung .....	16
2.4	Genotypisierung mittels Microarray .....	17
2.5	Vollgenomsequenzierung .....	18
2.6	Bioinformatische Auswertung .....	18
3.1	Vollgenomanalyse .....	24
3.2	Ergebnisse der DNA-Microarray Analyse .....	29
3	<b>Ergebnisse .....</b>	<b>24</b>
4	Diskussion.....	30
5	Literaturverzeichnis .....	36

6	Publikationen / Danksagung.....	39
---	---------------------------------	----

7	Anhang .....	40
---	--------------	----

6.1	Publikation im Rahmen der Dissertation.....	39
-----	---	----

6.2	Danksagung .....	39
-----	------------------	----

7.1	Protokoll der spa-Typisierung.....	40
-----	------------------------------------	----

7.2	Protokoll der Microarray-Messung.....	41
-----	---------------------------------------	----

7.3	Protokoll der DNA Aufreinigung für WGS.....	43
-----	---	----

7.4	Tabelle 2: Potentielle Virulenzgene .....	44
-----	---	----

Lebenslauf ..... Fehler! Textmarke nicht definiert.

# 1 Einleitung

## ***Staphylococcus aureus*: Mikrobiologie und Taxonomie**

*Staphylococcus aureus* ist aus der Gruppe der Staphylokokken das bedeutendste und virulenteste Bakterium. Das Wort Staphylokokkus kommt von den beiden 1.1 altgriechischen Wörtern für „Weintraube“ und „Kern“. Nachdem durch den bekannten deutschen Chirurgen Billroth 1874 die Streptokokken ihren Namen erhielten, war es der schottische Chirurg Ogston, der das Wort Staphylokokkus einführte [1]. 1884 hat der Deutsche Anton J. Rosenbach, wiederum ein Chirurg, für die in ihrem Wachstum goldfarben erscheinenden Stämme den Namen *Staphylococcus aureus* (aus dem Lateinischen „aurum“ für Gold) eingeführt [2]. Die über 30 pathogenen Staphylokokkenspezies sind unbeweglich, aerob - aber auch fakultativ anaerob - und können extra- wie auch intrazellulär leben. Mit Hilfe ihrer polysaccharid Kapsel sind Staphylokokken widerstandsfähig und auch bei nicht idealen Umweltbedingungen lange fähig zu überleben.

Auf Blutagar Nährmedien wächst *Staphylococcus aureus* in kleinen traubenförmigen Kolonien mit einer goldenen Färbung und einer  $\beta$ -Hämolyse. Dadurch grenzt es sich bereits makroskopisch von seinen Artverwandten, den Koagulase-negativen Staphylokokken (CoNS) ab. Diese formen weiße sowie kleinere Kolonien, weil sie dieses Enzym nicht besitzen, was auch mit einer geringeren Virulenz einhergeht [3].

Eine einfache Möglichkeit, die klinisch relevanten Gram-positiven Kokken zu unterscheiden, besteht in dem Katalase- und dem Koagulase-Test: Während Streptokokken hier negativ reagieren, besitzen Staphylokokken dieses Enzym. *Staphylococcus aureus* zeichnet sich singular außerdem dadurch aus, dass er, wie oben aufgezeigt, das Oberflächenenzym Koagulase besitzt, welches Fibrinogen spalten kann. Darüber hinaus gibt es auch noch einen Latexagglutinationstest, der das für *Staphylococcus aureus* charakteristische Protein A nachweist [3].

### 1.1.1 Methicillin-Resistenz

MRSA steht für *Methicillin-Resistenten Staphylococcus aureus*. Bei Methicillin handelt es sich um das erste penicillinasefeste Penicillin, welches 1959 eingeführt wurde, nachdem die ersten *Staphylococcus aureus*-Stämme bereits kurze Zeit nach dem ersten Einsatz von Penicillin-Resistenzen gegen dieses entwickelt hatten. Bereits im Jahr 1961 tauchten Berichte in Großbritannien auf, die über eine Methicillin-Resistenz berichteten – dies gilt als der Anfang der MRSA-Entwicklung [4]. Die beschriebene Resistenzentwicklung beruht auf einer Veränderung des penicillin-bindenden Proteins, eines für Gram-positive Bakterien notwendigen Proteins, welches für die Zellwandsynthese zuständig ist. Für gewöhnlich binden  $\beta$ -lactam-Antibiotika, zu denen die Penicilline zählen, an diese Proteine und verhindern somit die Quervernetzung der Bakterienzellwand. Durch die resultierende mangelnde Stabilität entfaltet sich eine bakterizide Wirkung.

*Staphylococcus aureus* trägt diese Fähigkeit, alternative Penicillinbindeproteine zu produzieren, auf dem sogenannten *mecA* Gen, welches wiederum auf einem mobilen genetischen Element, dem „Staphylococcal cassette chromosome *mec*“ (SCC*mec*), liegt [5]. Vom Gen *mecA* gibt es vier verschiedene Untertypen und es wird angenommen, dass dieser Virulenzfaktor auf unbekanntem Wege erst in Koagulase-negative Staphylokokken und dann weiter über horizontalen Gentransfer in *Staphylococcus aureus* gelangt ist. Enright *et al.* [5] beschreiben, dass es mehrere Fälle gibt, in denen epidemisch erfolgreiche Methicillin-sensible *Staphylococcus aureus* (MSSA)-Klone mehrfach der Sprung zu MRSA gelungen ist und dass der horizontale Gentransfer des *mecA*-Gens mehrmalig in der Entwicklungsgeschichte der Staphylokokken vorgekommen sein muss. Ein Hinweis hierauf ist die Tatsache, dass es eng verwandte *Staphylococcus aureus*-Klone gibt, die unterschiedliche SCC*mec*-Typen haben. Eine Folge dieser Entwicklung ist, dass weniger als 5% der klinischen Isolate in schwerwiegenden Staphylokokken-Infektionen sensibel für Penicillin sind [3] und eine Therapie mit Reserveantibiotika wie zum Beispiel Vancomycin eingeleitet werden muss.

### 1.1.2 Klinische Bedeutung von *Staphylococcus aureus*

*Staphylococcus aureus* ist einer der wichtigsten Krankenhauskeime überhaupt und einer der wichtigsten Erreger nosokomialer Erkrankungen, von denen es in Deutschland etwa 400.000 bis 600.000 gibt [6]. Insbesondere auf Intensivstationen ist er ein gefürchteter Erreger. In Deutschland wurde gezeigt, dass *Staphylococcus aureus* mit einem Anteil von 16,5% der häufigste nosokomiale Erreger ist und der Anteil der Methicillin-resistenten Stämme kontinuierlich zugenommen hat [7].

Für die USA gibt es Daten, die zeigen, dass sich die Problematik der MRSA-Keime weiterhin verschärft. So gilt für den Zeitraum von 1995-2000, dass sich die Zahl der Krankenhausaufenthalte, die mit MRSA assoziiert waren, um bis zu 62% gesteigert hat [8].

Dass MRSA weltweit an Krankenhäusern und auch zunehmend außerhalb dieser Einrichtungen verbreitet ist und sich auf dem Vormarsch befindet, liegt daran, dass *Staphylococcus aureus* ein sogenannter Kommensale ist. Ein Kommensale ist ein Erreger, der regelmäßig die Haut und auch die Schleimhäute von Menschen besiedelt. Das Hauptreservoir von *Staphylococcus aureus* sind die menschlichen Nasenvorhöfe, die er bei 20% der Menschen dauerhaft sowie bei 30% zeitweise besiedelt [9]. An anderen Körperstellen, wie zum Beispiel unter den Achseln sowie in der Analregion, kann ebenfalls MRSA nachgewiesen werden. Dabei ist interessant, dass für einen *permanenten* Keimträger eine Besiedelung durch ein und denselben Keim typisch ist, während *temporäre* MRSA-Keimträger unterschiedliche Keime für unterschiedlich lange Zeiträume beherbergen. Letzteres findet sich zum Beispiel häufig im Rahmen eines Krankenhausaufenthaltes [10]. Gerade durch die Fähigkeit von *Staphylococcus aureus*, seinen Wirt, den Menschen, zeitweise zu besiedeln, steigt das Gefährdungspotential der Übertragung durch medizinisches Personal.

*Staphylococcus aureus* ist ein überaus vielseitiger Krankheitserreger, der sich vieler unterschiedlicher Pathogenitätsfaktoren bedient, die entweder toxinvermittelt sind oder auf einer direkten Gewebsinvasivität beruhen. Haut- und Weichteilinfektionen, wie die Follikulitis oder Prozesse, die mit der Bildung eines Abszesses einhergehen, also Furunkel oder Karbunkel, sind überaus typische *Staphylococcus aureus*-Erkrankungen. Sind die tieferen Weichteilgewebe betroffen, kann es zu Entzündungen des Knochenmarks (Osteomyelitis) oder zu tiefen Muskelabszessen

kommen. Darüber hinaus kann eine Infektion auch in die Blutbahn streuen und dann septisch verlaufen und die inneren Organe und hier insbesondere das Herz und die Herzklappen schädigen [3]. Des Weiteren können Staphylokokken auch sogenannte Biofilme auf medizinischen Implantaten bilden, wo sie unter einem Netz quer- vernetzten körpereigenen Fibrins vor Antibiotika geschützte Kolonien formen. [3]

In letzter Zeit wurden hoch virulente, extrem invasive MRSA-Stämme, sogenannte community acquired MRSA beobachtet, deren Prävalenz regional sehr unterschiedlich auf um die 0.2% geschätzt wird. Diese Stämme haben die Fähigkeit, nekrotisierende Weichteilinfektionen (Nekrotisierende Faszitis) sowie nekrotisierende Lungenentzündungen mit einer Sterblichkeit von bis zu 70% zu verursachen [11].

## Typisierungsmethoden für *Staphylococcus aureus*

### 1.2

Typisierungsverfahren haben einen hohen Stellenwert in der Infektionsüberwachung und insbesondere im Rahmen der Diagnostik und Prävention von Ausbrüchen nosokomialer Krankheiten mit schwerwiegenden Verläufen auf Intensivstationen. Bei bereits stattgefundenen Ausbrüchen dienen Typisierungsverfahren dazu, Übertragungswege aufzudecken und die weitere Ausbreitung zu verhindern [12]. Prinzipiell kann man die phänotypischen von den genotypischen Verfahren zur Typisierung von Bakterien unterscheiden. Früher standen die rein phänotypischen Methoden im Vordergrund, die zum Beispiel darauf beruhten, ein Antibiotikaresistenzprofil experimentell zu bestimmen. Auch die Einteilung in unterschiedliche Serotypen, wie sie zum Beispiel zum Unterscheiden von Salmonellen angewandt wird, fällt in diese Kategorie. Bei der Einteilung in Serotypen werden mit Hilfe von Antikörpern gegen bestimmte ausgewählte antigene Determinanten beispielsweise Oberflächenproteine nachgewiesen.

Heutzutage stehen vermehrt die genotypischen Typisierungsverfahren im Vordergrund, da diese ein viel höheres Diskriminierungspotential besitzen. Dies bedeutet, dass genotypische Methoden in der Lage sind, epidemiologisch nicht verwandte Stämme voneinander zu trennen und somit Infektionsketten entweder zu bestätigen oder auszuschließen. Dabei sind einige dieser Typisierungsmethoden so

exakt, dass sie Einzelnukleotid-Polymorphismen (SNPs, single nucleotide polymorphism) detektieren können. Ein SNP ist ein genetischer Polymorphismus, der von einer Population geteilt wird. Bestehen Einzelnukleotidvariationen zwischen nur zwei Organismen oder treten sie nur vereinzelt innerhalb einer Population auf, werden sie SNV (single nucleotide variation) genannt.

Weiterhin sind die hochauflösenden genotypischen Typisierungsverfahren so detailliert, dass sie den Austausch einer einzelnen Base im DNA-Molekül bestimmen können. Dadurch sind sie nicht nur in der Lage, lokale endemische Klone eines MRSA-Stammes in einem Krankenhaus oder auch in einer Region zu definieren, sondern können diese auch von allen anderen dort auftretenden Erregern der gleichen Spezies abgrenzen [13].

Im Folgenden sollen nun die wichtigsten genotypischen Verfahren zur Typisierung von *Staphylococcus aureus*-Isolaten dargestellt werden. Hierbei soll insbesondere der Fokus auf die Protein A-Typisierung (*spa*-Typisierung) und die DNA-Microarrays fallen, da diese beiden Methoden im Rahmen der hier vorgelegten Arbeit angewandt wurden.

### **1.2.1 Pulsfeld-Gelelektrophorese**

Bei der Pulsfeld-Gelelektrophorese (PFGE) handelt es sich um ein auch heute noch weit verbreitetes Verfahren, das im Review von Sabat *et al.* [14] noch als Goldstandard der epidemiologischen Typisierung bezeichnet wird. Beim PFGE-Verfahren werden durch eine DNA-Endonuklease, also durch ein Enzym, das DNA-Moleküle spalten kann, 30kb bis zu 10Mb große DNA-Fragmente erstellt, die aus hochaufgereinigter, genomischer DNA des zu untersuchenden Bakteriums bestehen. Diese Fragmente, die abhängig von der Spaltstelle eine unterschiedliche Größe besitzen, werden im nächsten Schritt mithilfe eines Agarose-Gels getrennt. Ein besonderes Merkmal des Elektrophorese-Verfahrens ist die angelegte Spannung, die nicht kontinuierlich in eine Richtung, sondern abwechselnd gepulst in einem Winkel links und rechts zur Laufrichtung angelegt wird. So gelingt es, abgrenzbare Banden auch von großen Molekülen zu erhalten. Bei einer kontinuierlichen Spannung würden diese auf einer Höhe landen und wären nicht mehr voneinander abgrenzbar. Die auf

diese Weise voneinander getrennten DNA-Fragmente sind als sogenanntes PFGE-Muster sichtbar und können interpretiert werden.

Das PFGE-Verfahren zeichnet sich durch ein hohes Diskriminierungspotential aus und ist relativ kostengünstig, sodass es hervorragend für umfangreiche epidemiologische Untersuchungen genutzt werden kann. Von Nachteil sind der große technische Aufwand sowie der benötigte Zeit- und Arbeitsaufwand. Als weitere Kritikpunkte sind die Subjektivität der Interpretation des PFGE-Bandenmusters und eine geringere Möglichkeit zum Vergleichen der Ergebnisse sowie das Veröffentlichende der erhobenen Daten zu nennen [15].

### 1.2.2 Multilocus Sequenz-Typisierung

Bei der Multilocus Sequenz-Typisierung (MLST) handelt es sich um eine standardisierte Methode zur Bestimmung der Phylogenie (Verwandtschaft) von MRSA-Stämmen. Bei diesem Verfahren werden für *Staphylococcus aureus* sieben Haushaltsgene mittels PCR vermehrt und anschließend sequenziert. Als Haushaltsgene bezeichnet man Gene, die unabhängig von Zelltyp, Zellstadium und äußeren Einflüssen exprimiert werden [16]. Das Charakteristische des MLST-Verfahrens besteht darin, dass unterschiedlichen Sequenzen der verschiedenen Gene Identifikationsnummern zugeordnet werden. Aus der Zusammensetzung aller sequenzierten Gene und den dazugehörigen jeweils eindeutigen Nummern ergibt sich dann der sogenannte „Sequence Type“ oder kurz ST-Typ eines *Staphylococcus aureus*-Isolates [14].

Die großen Vorteile dieser Methode sind die international einheitliche, standardisierte Nomenklatur sowie die Verfügbarkeit großer zentraler Datenbanken (<http://www.mlst.net>). So werden üblicherweise *Staphylococcus aureus*-Isolate nach ihrem ST-Typ und nach ihren vier unterschiedlichen *mecA*-Genen (siehe 1.1.1) benannt. Eine vollständige Bezeichnung eines endemischen Stammes sieht dann z.B. wie folgt aus: ST 247-MRSA – I [14].

Die MLST-Methode dient darüber hinaus auch der Einteilung in sogenannte Klonale Komplexe (clonal complex, CC), und kann mit Hilfe der Software „BURST“ (<http://www.mlst.net>) durchgeführt werden. MRSA-Isolate werden allgemein üblich

innerhalb eines gleichen klonalen Komplexes gruppiert, wenn fünf der sieben sequenzierten Haushaltsgene identisch sind [17].

Nachteile der MLST-Methode sind die recht hohen Kosten, die zurzeit bei rund 50 Euro pro Isolat liegen [14], und das zeitintensive und arbeitsaufwendige Verfahren. Um ein Ergebnis zu erhalten, müssen sieben unterschiedliche Gene untersucht werden und eine weitere Unterscheidung innerhalb von ST MLST-Typen ist ebenfalls nicht möglich.

### 1.2.3 Protein A-Typisierung

Bei der Protein A-Typisierung handelt es sich im Gegensatz zur MLST-Typisierung um ein genomisches Typisierungsverfahren, bei dem anstelle von mehreren Genen nur ein einziges Gen per PCR amplifiziert und anschließend die Sequenz sequenziert wird. Demzufolge nennt man dieses Verfahren auch Single Locus Sequenz-Typisierung (SLST).

Die *spa*-Typisierung ist das wohl am meisten genutzte Typisierungsverfahren aus der SLST-Gruppe. Bei dieser Methodik macht man sich zu Nutze, dass es innerhalb der für das Protein A codierenden Sequenz eine Region gibt, die sich durch eine Aneinanderreihung von kurzen, sich wiederholenden Nukleotidsequenzen auszeichnet. Dort wo diese Sequenzen direkt beieinander liegen und sie sich zwischen verschiedenen *Staphylococcus aureus*-Klonen nur in ihrer Anzahl unterscheiden, nennt man diese Region „variable number tandem repeats“ oder „VNTR“ [14]. Innerhalb der VNTR-Region des *spa* Gens gibt es eine große genetische Vielfalt bestehend aus einer variablen Anzahl kurzer, sich wiederholender Basenpaarabfolgen oder auch Punktmutationen in diesen Bereichen. Auch bei diesem Verfahren wird jeder dieser eindeutig zugeordneten sequenzierten Sequenzen ein spezifischer Code zugeordnet. Hierdurch zeichnet sich auch dieses Verfahren durch seine standardisierte Nomenklatur sowie durch eine im Internet verfügbare Datenbank (<http://spaserver.ridom.de>) und die zugehörige benutzerfreundliche Software aus.

Die *spa*-Typisierung ist ein kostengünstiges, schnelles und einfach zu handhabendes Verfahren, das im Vergleich zu PFGE oder MLST ein etwas geringeres

Diskriminierungspotential aufweist. Dennoch ist es das am einfachsten zu handhabende, nützlichste und effizienteste Werkzeug für lokale, nationale und auch internationale Typisierungsprogramme [14].

#### **1.2.4 DNA-Microarray**

Bei der DNA-Microarray handelt es sich um ein Verfahren, mit dessen Hilfe mehrere hundert Gene anhand charakteristischer Nukleotidsequenzen nachgewiesen werden können. Der Nachweis beruht auf dem Prinzip der DNA-Hybridisierung. Auf einer festen Unterlage, dem sogenannten Chip, werden auf kleinstem Raum maschinell PCR-Amplifikate mit einer Länge von ungefähr 200 Basenpaaren befestigt (die sogenannten Sonden). Danach wird die zu untersuchende, aufgereinigte DNA markiert und auf den Chip hinzugegeben. Komplementäre Sequenzen werden fest gebunden und die überschüssige nicht bindende DNA wird in mehreren Waschschrritten entfernt. Die am Ende übrigbleibenden Signale werden von einem automatisierten Lesegerät mit angeschlossenen Computer und zugehöriger Software ausgewertet.

Je nach verwendeten Proben auf dem Chip kann man eine Vielzahl unterschiedlicher Ergebnisse erzielen. So ist es möglich, beispielsweise das Resistenzprofil von MRSA-Isolaten zu erfassen und anhand der nachgewiesenen Gene ihr Virulenzpotential zu betrachten. Auf diese Weise können auch hochvirulente cMRSA-Stämme, die das seltene PVL-Gen [11] tragen, detektiert werden. Darüber hinaus ist unter epidemiologischen Gesichtspunkten interessant, dass man mit Hilfe der Array-Technologie MRSA-Isolate auch in die schon erwähnten klonalen Komplexe einteilen kann [18].

Die Vorteile der Microarray-Technik liegen in der großen Zahl der nachweisbaren Gene sowie darin, dass auch extragenomische Elemente wie beispielsweise Plasmide miterfasst werden. Nachteilig sind die hohen Kosten, insbesondere für die hoचाussagekräftigen DNA-Microarrays mit mehreren 1000 Sonden, die sogenannten high density-Arrays. Darüber hinaus besteht eine gewisse Limitierung alleine dadurch, dass nur Gene nachgewiesen werden können, die auf dem starren Array als Sonden vertreten sind.

### 1.2.5 Vollgenomsequenzierung

Nachdem in den 50er Jahren die DNA-Struktur von Watson und Crick veröffentlicht wurde, bestand ein großes Interesse, dieses Molekül zu entschlüsseln und zu lesen. Der erste große Schritt in diese Richtung wurde 1977 getan, als Sanger das Kettenabbruchverfahren entwickelte. Dieses Verfahren der ersten Generation beruhte darauf, dass jeweils markierte und strukturell modifizierte Didesoxynukleotidtriphosphate zu einem Kettenabbruch während der DNA-Replikation führten. Die so entstandenen Fragmente unterschiedlicher Länge, aber mit immer der jeweils gleichen letzten Base, konnten dann mittels Gelelektrophorese entschlüsselt werden [19]. In den späten 90er Jahren wurde dieses Verfahren weiter verfeinert, und die nötigen Reaktionsschritte wurden durch den Einsatz von fluoreszenzmarkierten Basen als Reagenzien und den Einsatz von automatisierten Kapillarelektrophoreseverfahren minimiert [20].

Erst in den letzten Jahren wurde das „Next Generation Sequencing“ (NGS) entwickelt, das sich grundlegend von den Ansätzen der ersten Generation, basierend auf der Sangermethode, unterscheidet. Der Unterschied besteht darin, dass ein einzelner Lauf millionenfache kurze Einzelsequenzen, sogenannte „reads“, ergibt. Dies gelingt dadurch, dass die zu untersuchende Gesamt-DNA in kleine Fragmente zerteilt wird, diese dann per Adapter an einer spezialisierten Oberfläche („flow cell“) verankert und dort ungefähr tausendfach vermehrt wird. Zum Sequenzieren wird nun eine Technik angewandt, die sich SBS nennt, was für „sequencing by synthesis“ steht. Dies bedeutet, dass parallel jeweils immer ein fluoreszenz-markiertes Nukleotid eingebaut und dieses Signal per Laser gemessen wird. Im nächsten Schritt wird die Fluoreszenzfarbe enzymatisch abgespalten, wodurch der Einbau des nächsten Nukleotids ermöglicht wird. Dieses wird gemessen und am Ende entstehen Millionen von *reads* mit einer Länge von 35-700 Basenbaren.

Das Prinzip des NGS ermöglicht auch eine Anwendung in der Typisierung von MRSA-Stämmen, Kosten und Zeitaufwand hierfür sind soweit gesunken, dass es auf absehbare Zeit das Verfahren der Wahl werden dürfte. Schon im Jahr 2012 lagen die Gesamtkosten bei 75-100 Euro für ein einziges Isolat [14].

Die Vorteile des NGS liegen vor allem in der höchsten erreichbaren Auflösung, da die Genauigkeit bis zum Unterscheiden von SNPs reicht.

Die Nachteile ergeben sich aus der riesigen Datenmenge und deren aufwendiger Verarbeitung, denn die millionenfachen kurzen redundanten Einzelsequenzen haben für sich allein betrachtet zunächst keine Aussagekraft. Grundsätzlich gibt es für die bio-informatische Auswertung zwei Möglichkeiten: Die erste und einfachere besteht darin, die entstandenen *reads* mit einem Referenzgenom übereinander zu legen. Die Einzelsequenzen werden zusammengesetzt, um Informationen über die enthaltenen Gene und Variationen des untersuchten Genoms zu erhalten. Je nach verwendetem Referenzgenom lassen sich hiermit aber nur bis zu 90 Prozent der Daten in Übereinstimmung bringen. Die zweite Möglichkeit ist, die entstandenen *reads* mit bioinformatischen Verfahren anhand ihrer Überlappungen so zusammensetzen, dass die Gesamtsequenz entsteht. Dieser Ansatz, welcher „*de novo*“ („von Neuem“) genannt wird, ist deutlich arbeitsintensiver und gelingt je nach Länge der verwendeten *reads* nicht immer fehlerfrei.

### 1.3 **Das Genom von *Staphylococcus aureus***

Das Genom von *Staphylococcus aureus* besteht, wie bei Bakterien üblich, aus einem ringförmig angelegten DNA-Molekül und zusätzlich einem oder auch mehreren Plasmiden. Das Gesamtgenom hat üblicherweise eine Länge von ungefähr 2,8 Megabasenpaaren und die Plasmide ungefähr 25.000. Dies kann jedoch von Klon zu Klon variieren, denn das Erbgut von *Staphylococcus aureus* ist keineswegs statisch, sondern zeichnet sich durch eine stetige Anpassung aus. Man spricht hier von klonalen Organismen, da Bakterien allein durch ihre Art der Vermehrung zunächst einmal weitestgehend von einer Generation zur nächsten genetisch identisch bleiben.

Von den um die 2,8 Millionen Basenpaaren sind ungefähr 85% codierende Regionen, und wenn man die offenen Leserahmen (engl. „open reading frames“) betrachtet, also die Regionen, die dem Bereich zwischen einem Start- und einem Stopcodon entsprechen, kommt man auf eine Anzahl von rund 2600 [21]. Alle diese

Bereiche codieren potentiell für eine Aminosäuresequenz und letztendlich auch für ein Protein.

Die Diversität verschiedener Populationen wird durch eine Kombination aus horizontalem Gentransfer, Rekombination und Mutation erreicht.

Der horizontale Gentransfer zeichnet sich dadurch aus, dass er unter Bakterien auch über Artgrenzen hinaus stattfindet und zu großen Entwicklungssprüngen führen kann. Auf diese Weise ist, wie oben erwähnt, die Resistenz gegenüber Methicillin in *Staphylococcus aureus* gelangt (siehe 1.1.1.).

Bei der Rekombination handelt es sich um eine Neuordnung von genetischem Material innerhalb eines Organismus. Bei Bakterien geschieht dies auch durch parasexuelle Prozesse wie die Konjugation, bei der Bakterien Erbgut untereinander austauschen, oder die Transduktion, bei der neues Erbmaterial mittels Viren in die Bakterien gelangt. Schließlich gibt es noch die Mutation, zum Beispiel in Form von Punktmutationen, die zu einem Basenaustausch führen können. Dieser Prozess tritt spontan auf und spielt in der Evolution eine wichtige Rolle, da er über einen großen Zeitraum verläuft und zu einer immer besseren Anpassung des Organismus an seine Umwelt führt. Klinisch relevant ist, dass die Evolution von *Staphylococcus aureus*-Keimen sowohl während einer asymptomatischen Kolonisation als auch im Verlaufe einer invasiven Erkrankung innerhalb der Blutbahn des infizierten Wirtes geschehen kann.

Das Genom von *Staphylococcus aureus* kann in ein Kerngenom (core genome) und ein Zusatzgenom (accessory genome) unterteilt werden. Das Kerngenom macht ungefähr 75% des Gesamtgenoms aus und enthält alle Gene, die für Wachstum und Überleben essentiell sind [22]. Darüber hinaus sind auch Gene enthalten, die für staphylokokkentypische Virulenzfaktoren codieren, wie zum Beispiel Proteine, die für die Adhäsion an Zellen zuständig sind, und auch Toxine und Exoenzyme, die das invasive Wachstum begünstigen.

Wichtigster Bestandteil des Zusatzgenoms sind die mobilen genetischen Elemente, die ungefähr 15-20% des Gesamtgenoms ausmachen [23]. Zu den mobilen genetischen Elementen gehören Plasmide, Transposons, Bakteriophagen und auch die Bereiche, die für einen Großteil der Virulenz verantwortlich sind, die sogenannten

Pathogenitätsinseln. Diese mobilen Elemente unterliegen einer hohen Evolutionsrate und werden insbesondere durch horizontalen Gentransfer innerhalb der Staphylokokken ausgetauscht. Sie beinhalten einen Großteil der Erbinformationen, die die Fähigkeit zur Resistenz, Toxizität, Virulenz und Umgehung des Immunsystems beinhalten. Durch diese Bereiche des Genoms ist *Staphylococcus aureus* in der Lage, sich an Lebensräume und Wirte anzupassen.

Während sich das Kerngenom durch eine recht langsame Mutationsrate von ca. 2,6 Mutationen/Megabasen und Jahr (Young *et al.*) auszeichnet und hochkonserviert ist, haben die mobilen genetischen Elemente eine viel höhere Mutationsrate und zeigen eine ausgeprägte Vielfalt in unterschiedlichen Isolaten [23].

Darüber hinaus gibt es im Genom von *Staphylococcus aureus* mehrere Regionen, in denen repetitive Sequenzen auftreten, die meist in den nicht codierenden Abschnitten liegen und von denen es ungefähr 70 gibt [21]. Repetitive Abschnitte treten allerdings auch in codierenden Abschnitten auf, etwa in der polymorphen X-Region des Protein A. Die Gemeinsamkeit zwischen diesen repetitiven Regionen und den mobilen genetischen Elementen besteht darin, dass sie eine hohe Mutationsrate besitzen und die Analyse von evolutionsassoziierten Mutationen der Vollgenomsequenzen erschweren.

### **1.3.1 Virulenzfaktoren von *Staphylococcus aureus***

*Staphylococcus aureus* besitzt eine ausgesprochen große Anzahl z.T. redundanter Virulenzfaktoren, die auf vielfältige Weise dem Bakterium dazu dienen, seinen Wirt zu infizieren und in ihm zu persistieren. Unterschieden werden kann hierbei zwischen strukturgebundenen und sekretierten Proteinen. Zu den strukturgebundenen Proteinen gehört das sich in der Kapsel befindende Protein A, das Antikörper am Fc-Fragment binden kann und somit die Aufnahme in Fresszellen verhindert. Des Weiteren bindet *Staphylococcus aureus* an menschliche Endothelzellen. Dieser Prozess wird unter anderem durch mehrere Fibronectin und Kollagen bindende Proteine vermittelt. Alle Proteine, die der Bindung an Wirtsgewebe dienen, werden unter dem Begriff MSCRAMMS („microbial surface components recognizing adhesive matrix molecules“) zusammengefasst. Mit Hilfe dieser Gruppe von

Proteinen gelingt es, *Staphylococcus aureus* auch an nicht biologischem Material wie Prothesen festzuhalten. Zu den weiteren Pathogenitätsfaktoren zählt die Fähigkeit zur Bildung von Biofilmen, mit deren Hilfe die Bakterien an anorganischem Material unter Ausbildung einer Matrix aus Polysacchariden anheften. Während einer stattgefundenen Infektion bekämpft *Staphylococcus aureus* aktiv das Immunsystem, beispielsweise mit Hämolytinen, die in der Lage sind, Erythrozyten und andere Zellen zu zerstören. Darunter fallen auch die Leukozidine, die Leukozyten lysieren, indem sie deren Zellmembran mit Poren durchsetzen. Das invasive Wachstum während einer Infektion wird durch zahlreiche sekretierte Enzyme vermittelt, die in der Lage sind, das Wirtsgewebe aufzulösen. Darunter fallen zum Beispiel: Alpha-Hämolysin, Beta-Hämolysin und Gamma-Hämolysin sowie weitere Metalloexoproteasen, Elastasen oder Lipasen [24]. Durch diese Enzymausstattung lässt sich die Neigung zum tiefen Vordringen ins Gewebe mit der Ausbildung von Abszessen erklären. In diese Klasse fällt auch das Pantone-Valentine-Leukozidin (PVL), das Phagen-kodiert und für schwere nekrotisierende Pneumonien verantwortlich ist [25].

Auch die Ausstattung an Toxinen ist umfangreich: Es gibt eine Gruppe von hitzebeständigen Enterotoxinen, die als Superantigene fungieren und Lebensmittelvergiftungen auslösen. Darüber hinaus gibt es die Exfoliatine A und B sowie die Epidermiolysine, die zur Blasenbildung in der Haut führen. *Staphylococcus aureus* bildet darüber hinaus den Clumping Factor A und B, mit dessen Hilfe unter der Mitwirkung von Fibronektin das Anheften an Fibrin vermittelt wird.

Darüber hinaus zählen alle Gene, die Antibiotikaresistenzen vermitteln, zu den Virulenzfaktoren, wie etwa das modifizierte Penicillinbindepotein PBP2 oder Transportproteine, die aktiv Antibiotika aus dem Bakterium herauspumpen.

## Zielsetzung

1.4 Aufgabenstellung dieser Studie ist zum einen die genomische Analyse von MRSA-Proben von 12 invasiven MRSA-Isolaten aus Blutkulturen und 15 kolonisierenden MRSA-Isolaten aus Nasenabstrichen, die am Universitätsklinikum des Saarlandes (UKS) gewonnen wurden. In einem zweiten Schritt sollen diese Proben miteinander verglichen werden. Das in dieser Studie untersuchte Vergleichskollektiv ist fokussiert auf den klonalen Komplex 5 und, innerhalb dieses Komplexes, auf die lokalen im Saarland vorzufindenden t504-Stämme sowie auf die in Deutschland weit verbreiteten t003-Stämme.

Für die Typisierung werden verschiedene Methoden eingesetzt, darunter die Singlelocus Typisierungsmethode Protein A-Typisierung sowie Genotypisierung mittels DNA-Microarray. Im Mittelpunkt dieser Arbeit steht die Vollgenomanalyse mit der Zielsetzung, einen präzisen und umfassenden genomischen Vergleich anzustellen.

Zielsetzung ist es, die genetische Ausstattung von invasiven und kolonisierenden Isolaten zu vergleichen und die Frage zu beantworten, ob sich die Isolate der verschiedenen Gruppen bezüglich ihrer Virulenz unterscheiden. Die Vollgenomsequenzierung bietet des Weiteren auch die Möglichkeit, die Phylogenie der untersuchten Isolate abzubilden und einen möglichen Zusammenhang zwischen Phylogenie und Virulenz aufzudecken. Darüber hinaus sollen die Möglichkeiten der Vollgenomanalyse genutzt werden, um potentielle neue Virulenzfaktoren in MRSA zu finden.

## 2 Material und Methoden

### Material

Als Nährmedien dienen:

Blutplatten: 5% Schafsblut in TSA der Firma Becton Dickinson

2.1

TSB Medium: 30,0 g Bacto Tryptic Soy Broth (TSB) (Difco), ad 1 l aqua dest

Des Weiteren wurden verwendet die Kits:

1. „DNeasy Blood and Tissue Kit“ der Firma Qiagen
2. „*S.aureus* Genotyping Kit 2.0“ der Firma Alere Technologies
3. “Agencourt AMPure XP PCR Purification Kit” der Firma Beckman Coulter

sowie die Primer:

1. *spa1113F* (5'-TAAAGACGATCCTTCGGTGAGC-3') und
2. *spa1496R* (5'-TTTGCTTTTGCAATGTCATTTACTG-3') der Firma Roche

2.2

### Studienisolate

Die 15 kolonisierenden MRSA-Isolate stammen von einem Patientenkollektiv, das im Rahmen einer Screening-Studie im Jahre 2012 am Universitätsklinikum des Saarlandes (UKS) erfasst wurde. Diese MRSA-Isolate wurden schon im Rahmen einer Studie aus der Medizinischen Mikrobiologie Homburg von Ruffing *et al.* untersucht, bei der es um Genotypisierung von MRSA- und-MSSA Stämmen mittels Microarray und spa-Typisierung geht [26]. Bei den Isolaten handelt es sich um kolonisierende Bakterien nicht kritisch erkrankter Patienten, deren spa-Typ und Microarray-Genprofil bereits bekannt sind.

Dem gegenüber steht die Gruppe der invasiven Isolate, die aus Blutkulturen schwer erkrankter oder septischer Patienten im Rahmen eines stationären Klinikaufenthalts am UKS isoliert wurden.

Die Auswahl dieser 12 invasiven Isolate wurde wie folgt durchgeführt: Mit der Laborsoftware [M/Lab] wurde in der Datenbank für Blutkulturen am Institut für Mikrobiologie und Hygiene des UKS eine Suchanfrage für den Zeitraum vom

1.1.2007 bis 1.1.2013 nach im Blut von Patienten nachgewiesenen und anschließend eingefrorenen MRSA-Isolaten gestellt. Die gefundenen Proben wurden mittels Protein A-Typisierung untersucht und diejenigen Isolate, die den spa-Typ t504 oder t003 aufwiesen, wurden in die Studie aufgenommen. So entstanden vier Gruppen mit mindestens je vier Isolaten: invasive t003-Isolate, invasive t504-Isolate, kolonisierende t003-Isolate und kolonisierende t504-Isolate.

## Protein A-Typisierung

2.3 Von den am Vortag auf TSA-Blutagar angeimpften Isolaten wurde eine Einzelkolonie in destilliertes nukleasefreies Wasser eingerührt. Die Lyse der Bakterien geschah durch 10 Minuten Kochen bei 95°C. Nach kurzem Zentrifugieren wurde eine PCR angesetzt, um die polymorphe X-Region des *spa*-Gens zu vervielfältigen. Benutzt wurden hierzu mittels Hochleistungsflüssigkeitschromatographie gereinigte Vorwärts- und Rückwärtsprimer: spa1113f und spa1514r ([www.ridom.de](http://www.ridom.de)). Nach der Anfertigung eines Kontrollgels mittels Elektrophorese, das sicherstellen sollte, dass das Amplifikat rein ist, wurde das PCR-Amplifikat aufgereinigt. Hierzu wurde das „Agencourt AMPure XP“ Kit der Firma Beckman Coulter benutzt. Im letzten Schritt wurde mit dem „GenomeLab™ DTCS Quick StartKit“ (Beckman Coulter) eine Sequenzierungs-PCR angesetzt, die einen Mastermix beinhaltet, der alle Reagenzien enthält, die für die anschließende Sequenzierung nach dem Sanger-Prinzip notwendig sind. Darunter fallen die Reaktionspuffer sowie die essentiellen zum Kettenabbruch führenden farb-markierten Desoxynukleotidtriphosphate. Es wurden jeweils zwei Ansätze pro Probe angefertigt, einmal mit Vorwärtsprimer und einmal mit Rückwärtsprimer.

Nach Ablauf der PCR wurden nach einem letzten Aufreinigungsschritt die Proben im Sequencer von Beckman Coulter des Typs „Genome Lab GeXP“ sequenziert. Zuletzt wurde mit Hilfe der Software Ridom Staph Type V.2.1.1 (Ridom GmbH) aus zwei zusammengehörigen Vorwärts- und Rückwärtssequenzen der spa-Typ der Isolate ermittelt und in eine Datenbank eingetragen [27] (Vollständiges Protokoll siehe Anhang).

## Genotypisierung mittels Microarray

2.4 Die für die Hybridisierungsversuche mit dem Microarray benötigte DNA muss viel höher aufgereinigt und aufkonzentriert sein als für die spa-Typisierung erforderlich, deshalb wurde hier folgendes Protokoll angewandt: Die über Nacht bei 37°C auf Schafblutagar inkubierten Isolate wurden einer chemischen Lyse, bestehend unter anderem aus Reaktionspuffer, Lysosthaphin, Lysozym und Ribonuklease A, unterzogen. Im weiteren Verlauf wurde noch Proteinase K hinzugegeben, um die unerwünschten Proteine im Lysat zu beseitigen. Das weitere Aufreinigen der DNA wurde unter Verwendung des Blood and Tissue Kits der Firma Qiagen vorgenommen. Das komplette Lysat wurde auf die im Kit enthaltenen Qiagen-Säulen aufgegeben, die eine Membran besitzen, in der die DNA während mehrerer Zentrifugationsschritte mit Waschpuffern hängen bleibt. Im finalen Schritt wurde mit Hilfe eines Elutionspuffers die gereinigte, konzentrierte DNA aus dieser Membran herausgelöst und in ein Eppendorfgefäß abzentrifugiert.

Vor der Durchführung der aufwendigen Array-Versuche wurde die DNA noch auf ihre Qualität hinsichtlich Konzentration und Reinheit geprüft. Dies geschah durch eine photometrische Messung mit dem Gerät Nano Drop 2000 (PEQLAB Biotechnologie GmbH). Die Qualitätsgrenze der für die Microarray zu verwendenden DNA lag bei einer Mindestkonzentration von 100ng/µl und einem A260/A280-Quotienten von mindestens 1,8. Dieser Quotient sagt etwas über die Reinheit der gemessenen DNA aus und wäre im Falle einer Verunreinigung durch Proteine deutlich kleiner.

Auf den verwendeten Microarray-Strips „IdentiBAC MA“ der Firma Alere Technologies aus Jena befinden sich 334 einzelsträngige modifizierte Oligonukleotidsonden, die im Falle einer erfolgreichen Hybridisierung eine besonders feste Bindung mit der Streptavidin-markierten Erreger-DNA eingehen. Nach stattgefundenener Hybridisierung und mehreren gründlichen Waschvorgängen mit unterschiedlichen Waschpuffern wurden die Gen-Spots auf dem Array mit Hilfe des Enzyms Meerrettichperoxidase sichtbar gemacht, das mit dem Streptavidin reagiert und zur Ausbildung von grauen Spots führt. Anschließend wurden diese Farbsignale mit dem hierfür entwickelten automatisierten Reader „Array Mate“ und der zugehörigen Software „Iconoclust“ (Alere Technologies, Jena) ausgewertet. Als Resultat der 334 in unterschiedlichen Graustufen vorhandenen Spots ergab sich ein Genprofil,

bestehend aus 174 einzelnen Genen, die in Speziesfaktoren, Virulenz- und Resistenzgene sowie MSCRAMMS unterteilt sind. Des Weiteren nimmt die Software eine Einteilung in die vier SCCmec-Typen vor und gibt den ST-Typ sowie die Zugehörigkeit zu den klonalen Komplexen an [26] (vollständiges Protokoll siehe Anhang 7.3).

## **Vollgenomsequenzierung**

2.5 Für die Präparation der genomischen DNA wurde ein leicht modifiziertes Protokoll von dem der Protein A–Typisierung angewandt, da für eine Vollgenomsequenzierung eine höhere Konzentration und Reinheit benötigt wird als für eine Singlelocus Sequenz Typisierung. Für die DNA-Aufreinigung wurde das „DNeasy Blood and Tissue Kit“ der Firma „Qiagen“ genutzt. Um eine ausreichende Menge an Bakterien zu erhalten, wurden die Proben über Nacht in BHI Medium inkubiert. Nach dem Zentrifugieren fand eine mechanische Lyse der Zellen im Zellhomogenisierer statt. Unerwünschte Proteine wurden mithilfe der im Kit mitgelieferten Proteinase K und unerwünschte RNA mit einer RNase der Firma „Appllichem“ entfernt. Daraufhin wurde die DNA in mehreren Zentrifugationsläufen hochkonzentriert (siehe 3.2.). Im letzten Schritt wurde ein 1% Agarosegel zur Kontrolle laufengelassen und die Konzentration der DNA mit dem NanoDrop Gerät fotometrisch gemessen (vollständiges Protokoll siehe Anhang). Schließlich wurde die qualitativ hohe DNA mit einer Konzentration von mindestens 80ng/µl an die Sequenzierungsabteilung des Helmholtz-Zentrums für Infektionsforschung in Braunschweig übergeben und dort auf einem Illumina MiSeq-System sequenziert.

2.6

## **Bioinformatische Auswertung**

Die Auswertung der Daten aus der Vollgenomanalyse wurde zusammen mit dem Institut für Bioinformatik in Saarbrücken und der Außenstelle des Robert Koch-Instituts für Infektionsforschung in Wernigerode durchgeführt. Das Prinzip der Auswertung besteht darin, verschiedene spezialisierte Computerprogramme hintereinanderschalten und zur einfacheren und schnelleren Bearbeitung der

Daten in einer sogenannten Pipeline zusammenzufassen. Eine Pipeline ist in einer einfachen Programmiersprache wie zum Beispiel Java geschrieben und basiert auf einer Eingabezeile, in die die entsprechenden Befehle hineingeschrieben werden können. Das Prinzip der im Rahmen dieser Arbeit angewandten Pipeline soll nun im Folgenden in den Grundzügen vorgestellt werden. Die Pipeline besteht aus einem Programmiercode, genannt Skript. Das Skript ruft die Programme nacheinander mit den vorgegebenen Einstellungen auf. Das vollständige Durchlaufen der kompletten Pipeline mit einer Datenmenge von 30 Ganzgenomsequenzen kann bis zu 12 Stunden dauern.

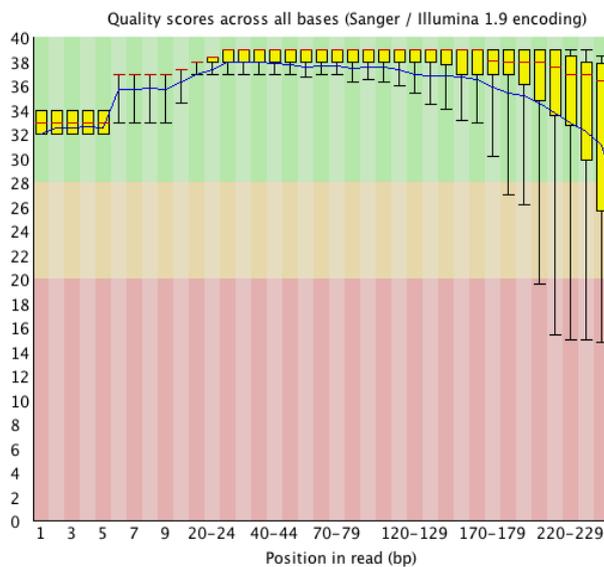
Im ersten Schritt wurde eine Qualitätskontrolle der Sequenzierungsdaten mit Hilfe des Programms FastQC durchgeführt. Die Qualität der einzelnen Reads wurde überprüft, indem die jeweiligen Qualitätswerte für jede einzelne Base ermittelt und grafisch dargestellt wurden. Der Qualitätswert ist als der Wert definiert, der logarithmisch abhängig ist von der Wahrscheinlichkeit  $P$ , mit der eine Base innerhalb der Reads falsch bestimmt wurde. Die Formel lautet:

$$Q = -\log_{10} P$$

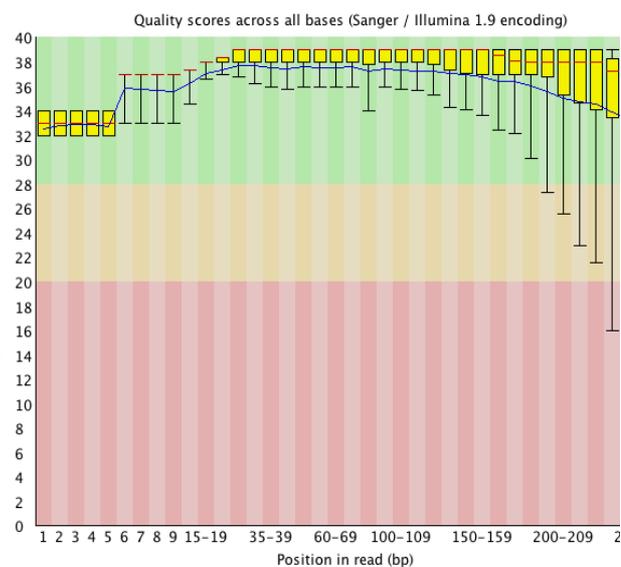
Somit bedeutet zum Beispiel ein Wert von  $Q=30$  für eine Base, dass eine Wahrscheinlichkeit von 1 zu 1000 besteht, dass diese Base falsch identifiziert wurde. Ein  $Q30$  Wert gilt als Standard der Analyse des Next Generation Sequencing und garantiert mit 99,9 prozentiger Wahrscheinlichkeit eine korrekte Sequenzierung.

Zu niedrige  $Q$ -Werte würden außerdem zu mehr falschen positiven Single Nukleotid Variationen (SNP) führen und somit die Aussagekraft der Daten herabsetzen. Die folgenden Abbildungen aus dem Programm FastQC zeigen grafisch einen Überblick über alle Qualitätswerte für alle Basen aller Reads. Aus technischen Gründen nimmt die Qualität am Ende eines Reads mit zunehmender Länge ab, was aus den Boxplot-Grafiken der Abbildung 1 gut ersichtlich wird.

### Qualitätskontrolle der kolonisierenden Isolate



### Qualitätskontrolle der invasiven Isolate



**Abbildung 1:** Beispiel Qualitätskontrollen der sequenzierten Isolate. Diese Analyse wurde für jede einzelne Probe unabhängig durchgeführt. Für jede Position ist der Qualitätswert als Boxplot dargestellt, wobei die rote Linie den Mittelwert und die blaue Linie die durchschnittliche Qualität darstellt.

Aus der obigen Abbildung 1 wird ersichtlich, dass die Sequenzierungsqualität für alle Isolate mit durchschnittlichen Qualitätswerten über 30 in einem tolerablen Bereich liegt. Im Falle einer nicht ausreichenden Qualität der Daten in diesem ersten Schritt der Pipeline, also mit Werten weit unter 30, würden diese von der Analyse ausgeschlossen.

Der nächste Schritt der Auswertung von Ganzgenomdaten nach der Beurteilung ihrer Qualität besteht darin, die einzelnen Reads im Mapping zusammzusetzen. In dieser Pipeline wurde nicht der *de novo*-Weg gewählt, sondern die Assemblierung mit Hilfe eines Referenzgenoms. Als Referenzgenom wurde das *Staphylococcus aureus*-Genom NC\_017340.1 [28] benutzt. Es eignet sich deshalb besonders gut, weil es sich durch eine enge Verwandtschaft mit den zu untersuchenden Isolaten auszeichnet. Beim Referenzgenom NC\_017340 handelt es sich um einen in Europa endemischen MRSA-Stamm mit dem Sequencetyp ST225, der außerdem von einem vermuteten Vorfahren des CC5-Komplexes abstammt [29].

Für das Mapping in dieser Pipeline wurde das Programm BWA und Samstats zur Qualitätskontrolle gewählt [30, 31]. BWA ist ein Programm, das dafür programmiert

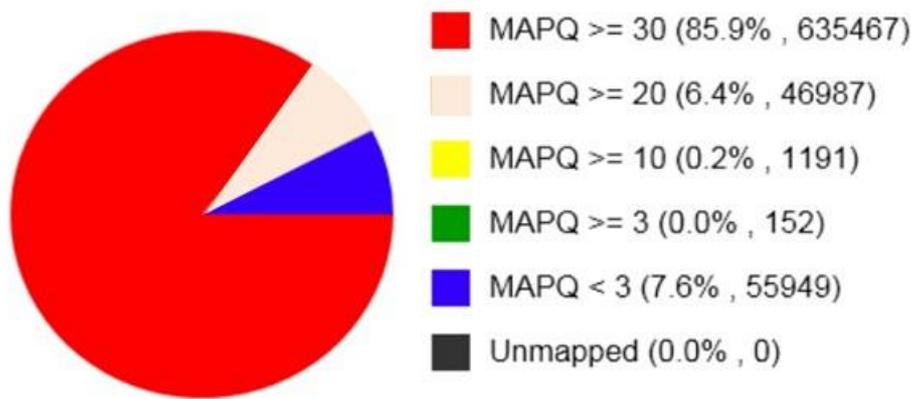
wurde, kurze Reads gegen ein Referenzgenom abzugleichen und somit zu einem Contig zusammenzusetzen.

Bowtie gibt Dateien im \*.sam Format (von Sequence Alignment/Map) aus. Die entstehenden Dateien sind kompakt in der Größe und können auch von anderen Programmen wie zum Beispiel Samstats verarbeitet werden.

Der nächste Schritt, der auf das Mapping folgt, ist eine erneute Qualitätskontrolle mithilfe des Programms Samstats. Im Folgenden soll auf die Bedeutung der Mapping-Qualität eingegangen werden: Die Berechnung erfolgt analog zur Berechnung der Qualitätskontrolle der Sequenzierungsdaten. Ein Qualitätswert von 30 für die Mapping-Qualität eines Reads bedeutet somit auch, dass eine Wahrscheinlichkeit von 1 zu 1000 besteht, dass dieser Read falsch im Vergleich zum Referenzgenom angeordnet wurde. Faktoren, die das Mapping und damit dessen Qualität beeinflussen, sind die Qualität und die Länge der Reads sowie die Anzahl und Menge der Repeatstrukturen des Referenzgenoms sowie die Sensitivität des benutzten Mapping-Programms.

Um die Analyse der in dieser Arbeit verwendeten Daten präzise zu gestalten, wurde auch in diesem Schritt ein Qualitätswert von 30 für das Mapping verwendet. Dies geht auf Kosten der Datenmenge, da alle anderen Reads und auch schon entstandene Contigs gefiltert werden.

Diese Aufgabe übernimmt das Programm Samstats innerhalb der Pipeline. Es bewertet die Mapping-Qualität und stellt sie grafisch dar, wie hier am Beispiel eines Datensatzes von einem Isolat, das aus einer Blutkultur gewonnen wurde, exemplarisch dargestellt werden soll.



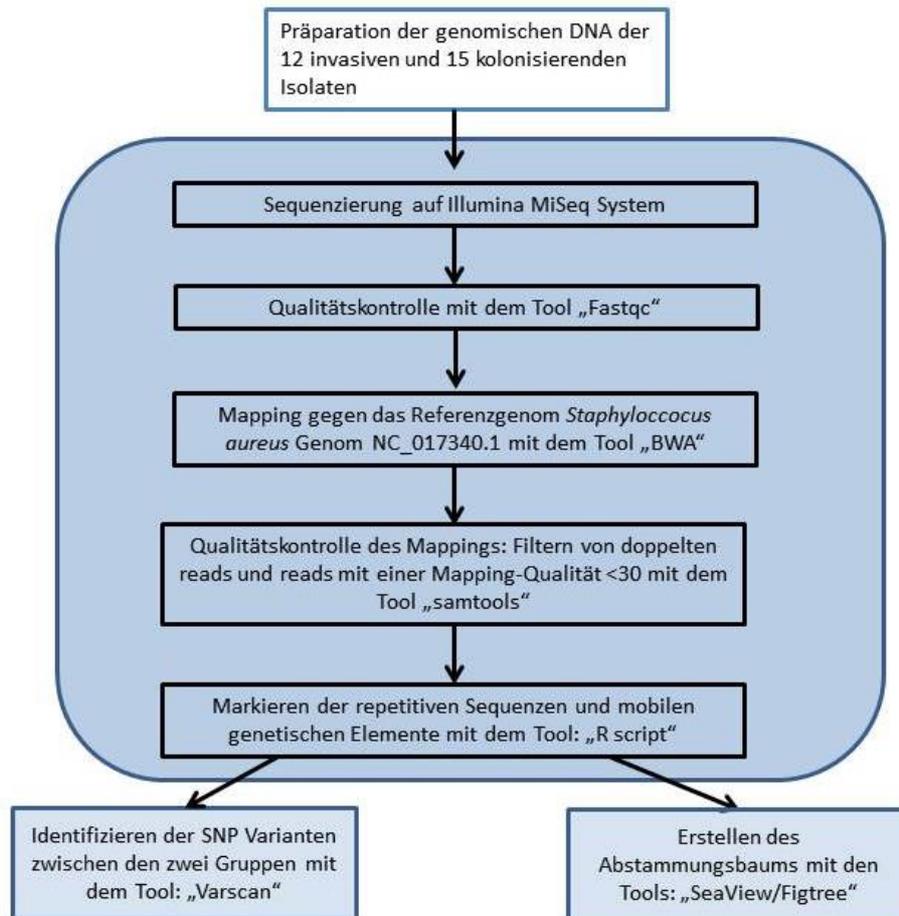
**Abbildung 2:** Qualitätskontrolle der Mapping-Qualität am Beispiel des Blutkultur Isolats S1. Das Kreisdiagramm zeigt die Anteile der Reads mit ihrer jeweiligen Mapping Qualität. In diesem Fall konnte für 85.9% der Reads ein Mapping mit einem Qualitätswert von 30 erreicht werden. Es gab darüber hinaus keine Reads, die nicht gemappt werden konnten. Diese Qualitätskontrolle wurde für jede einzelne Probe durchgeführt.

Analog zu den Qualitätskriterien für die Sequenzierung wurde auch für das Mapping ein Qualitätswert von 30 als Grenze angenommen, um die weitere Analyse der SNPs so aussagekräftig wie möglich zu gestalten.

Im nächsten Schritt der Pipeline wurden mithilfe des Programms Samtools alle gemappten Reads gefiltert, Duplikate entfernt und die Daten in einem anderen Format gespeichert. Dieses Format nennt sich \*.bam, ist eine vereinfachte komprimierte Version des \*.sam Formats und besitzt darüber hinaus einen Index, der die Position der gemappten Reads beziehungsweise auch größerer überlappender DNA-Fragmente in Bezug auf das Referenzgenom festlegt. Diese Indexerstellung erfolgt auch mit dem Programm Samtools und ist deshalb so wichtig, weil man später nach der folgenden SNP-Analyse weiß, an welcher Position eine Mutation auftritt. Dies ist möglich, weil das Referenzgenom durchgehend annotiert ist und somit auch das zugehörige Gen bekannt ist, in dem sich die Mutation befindet.

Im finalen Schritt der Datenanalyse wurden die Genabschnitte maskiert, die in repetitiven Regionen sowie in den mobilen genetischen Elementen auftreten, um falsch positive Ergebnisse bei der Suche nach Mutationen möglichst zu vermeiden.

Abbildung 3 fasst noch einmal den Aufbau sowie die Bestandteile und deren Anordnung innerhalb der verwendeten Pipeline zusammen.



**Abbildung 3:** Aufbau der Pipeline: die einzelnen Arbeitsschritte und die angewandten Programme.

Die Pipeline ermöglicht es, Mutationen zu finden, die nur innerhalb der Gruppe der invasiven MRSA-Isolate vorhanden sind. Dies ermöglicht die Identifikation von möglichen Virulenzgenen und Mutationen, die für eine höhere Virulenz von invasiven Stämmen verantwortlich sind. Aus diesem Grund wurde eine Analyse auf Mutationen mit dem Programm VarScan2 [32] durchgeführt, insbesondere auf Einzelnukleotidvariationen sowie Deletionen und Insertionen. Die genetischen Variationen zwischen den zwei verschiedenen Gruppen, den invasiven MRSA-Isolaten und den kolonisierenden Isolaten, wurden berechnet und den jeweils zugehörigen Genen zugeordnet. Daraufhin wurde das Auftreten dieser Variationen nach der auftretenden Position geordnet und die Anzahl dieser Mutationen in den beiden verschiedenen Gruppen gezählt.

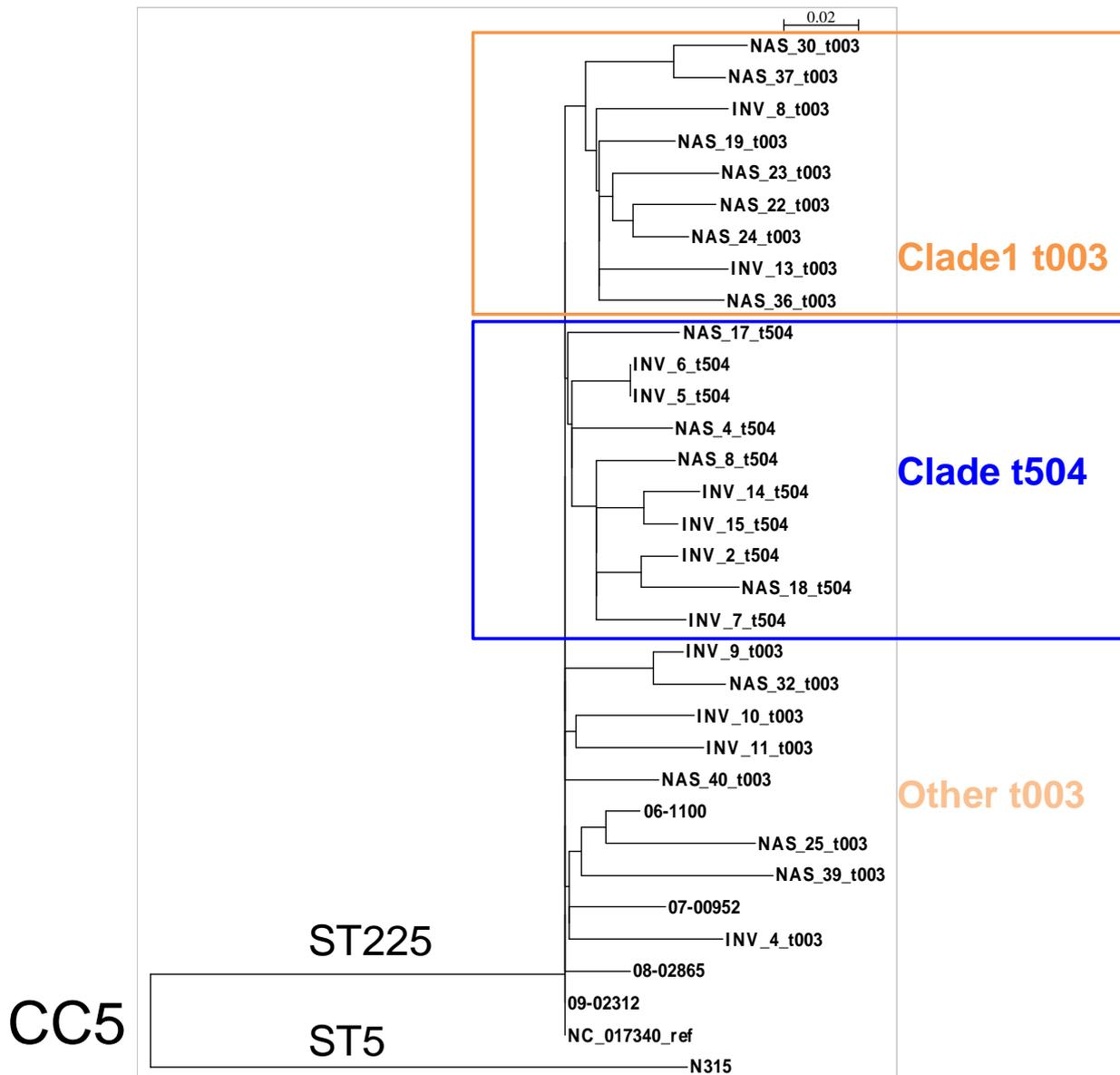
Die zweite Funktion der Pipeline besteht darin, die Phylogenie darzustellen. Der phylogenetische Abstammungsbaum wurde mit den Programmen *Seaview* [33] und *Figtree* [34] erstellt. Um eine höhere Aussagekraft zu erhalten, wurden alle dargestellten Genome mit dem Referenzgenom in Bezug gesetzt. Das Referenzgenom stellt somit den Startpunkt dar, während die Anzahl der SNPs die Länge der Äste der im Diagramm dargestellten entwicklungsgeschichtlichen Beziehung der untersuchten Genome bestimmt. Um die Aussagekraft zu erhöhen, wurden außerdem vier für Deutschland repräsentative CC5-Stämme hinzugefügt, um die Relation der im Rahmen dieser Arbeit untersuchten lokalen saarländischen Stämme zu diesen zu erkennen.

## 3 Ergebnisse

### 3.1 Vollgenomanalyse

Die Vollgenomsequenzierung der Bakterien-DNA mit dem Illumina MiSeq System hat, bei einer 110-fachen Abdeckung, Reads mit einer Länge von 251 Basenpaaren ergeben. Eine anschließende Analyse hat gezeigt, dass von den Reads 81-87% auf das Referenzgenom NC\_017340.1 [28] gemappt werden konnten.

Die Auswertung unter phylogenetischen Aspekten der Vollgenomanalyse ergibt, dass sich die untersuchten Isolate in drei geschlossenen Clustern wiederfinden. Diese werden jeweils auch als eine Klade bezeichnet, was in der Abstammungsbiologie besagt, dass diese definierte Gruppe sich dadurch auszeichnet, dass sie einen gemeinsamen Vorfahren und alle seine Nachfahren enthält. Diese phylogenetischen Verwandtschaftsverhältnisse zeigt nachfolgend Abbildung 4:



**Abbildung 4:** Phylogenetischer Baum der untersuchten Isolate sowie vier weitere repräsentative CC5 Stämme. Die Wurzel bildet das Referenzgenom NC\_017340.1

Es zeigt sich, dass sich alle saarländischen Isolate mit dem spa-Typ t504 klar von allen anderen Isolaten mit dem spa-Typ t003 abgrenzen lassen und ihre eigene Klade formen (*Clade t504*). In der Gruppe der t003-Stämme zeigt sich hingegen eine größere Vielfalt: Es entsteht eine Gruppe von 9 eng verwandten Stämmen, die eine individuelle Klade bilden und sich somit von den anderen deutschen t003-Stämmen abgrenzen (*Clade1 t003*). Die dritte Klade (*Other t003*) ist vielfältiger und besteht aus 9 t003-Isolaten sowie den anderen vier deutschen als Referenz hinzugefügten t003-Isolaten. Abbildung 4 verdeutlicht die große Vielfalt innerhalb der Gruppe der t003-Isolate und im Gegensatz dazu die nähere Verwandtschaft innerhalb der Gruppe der

t504-Isolate. Darüber hinaus zeigt sich, dass es keine direkte Assoziation zwischen der Phylogenie und der Herkunft der Proben (Blut oder Nasenabstriche) gibt. Bei der Klade 1 t003 handelt es sich um einen lokal endemischen Stamm ohne Assoziation zu den weit verbreiteten anderen t003.

Die Suche nach SNPs außerhalb mobiler genetischer Elemente und repetitiver Regionen ergab 535 verschiedene Treffer zwischen allen 12 invasiven und den 15 nicht invasiven Isolaten. Unter diesen 535 SNPs befanden sich 56 INDELS, also Mutationen, die mit einer potentiell größeren Veränderung der Proteinstruktur einhergehen. Die Klade *other t003* zeigte hierbei die meisten ( $56 \pm 11$ ) und die der saarländischen die wenigsten ( $36 \pm 7$ ) Mutationen. Es ist hervorzuheben, dass 40% der gefundenen Variationen in annotierten Genabschnitten liegen und 176 Gene betreffen, von denen wiederum 18 Gene laut der VFDB-Datenbank mit Virulenz assoziiert sind [35]. Von diesen 18 bekannten und mit Virulenz assoziierten Genen fanden sich Mutationen nur in der Gruppe der invasiven Isolate, während sich in der Vergleichsgruppe der kolonisierenden Isolate keine Mutationen fanden.

Darüber hinaus wurden anhand von SNPs, die sich in mindestens zwei oder mehr invasiven und in keinem der kolonisierenden Isolate befinden, 36 Gene identifiziert, von denen zwei bereits bekannte (*atl*, *hly*) Virulenzgene sind. Die weiteren Gene sind als neue potentielle Virulenzgene anzusehen. Einen Ausschnitt zeigt Tabelle 1 (komplette Tabelle siehe Anhang).

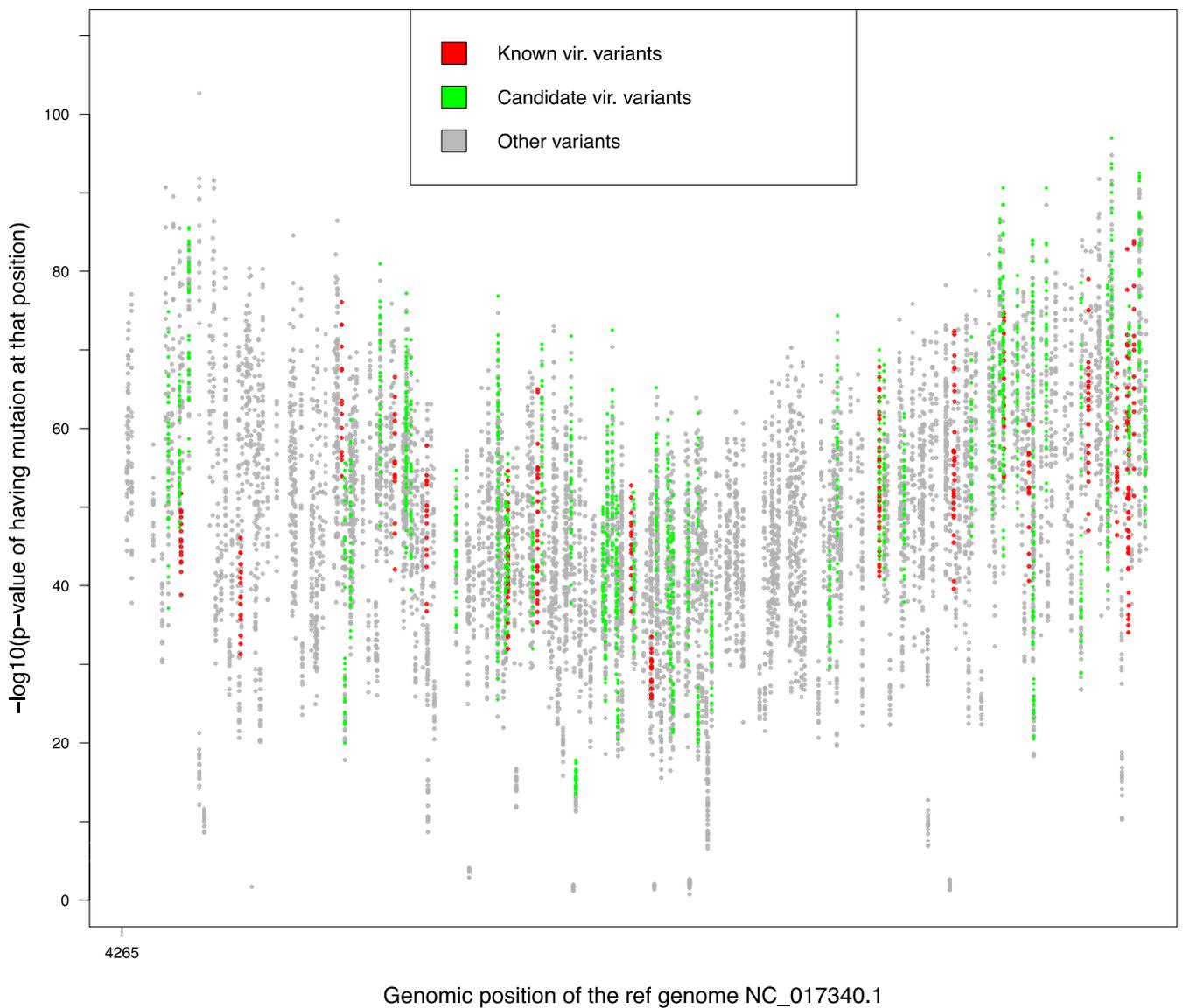
**Tabelle 1:** Liste der neuen potentiellen Virulenzgene, die in mindestens zwei invasiven Isolaten mit mehrere Variationen aufwiesen

Gene Name	Gene Symbol	Description	No. of invasive isolates with variants	Mutations	Amino acid changes	Ka/Ks
SA2981_1390	<b>ebhB</b>	Putative Staphylococcal surface anchored protein; adhesin emb	7	C to T	S to N	4.92
SA2981_1815	<b>pfoS/R</b>	Regulatory protein	3	C to CAAT	add R	0.74
SA2981_1256	<b>glpF</b>	Glycerol uptake facilitator protein	2	C to CT	ORF shifted	0.79
SA2981_2486	<b>feoB</b>	Ferrous iron transport protein B	2	C to T	V to M	50.00
SA2981_2564	<b>yvcP</b>	Two-component response regulator	2	G to T	P to T	50.00

Um die Bedeutung dieser so identifizierten Gene weiter zu untersuchen, stellt sich die Frage, ob es sich um stumme Mutationen handelt oder aber um solche, die die Aminosäuresequenz auch wirklich beeinflussen. Ein Maß, um dies zu quantifizieren, ist das  $k_a/k_s$ -Verhältnis. Dieses ist definiert als das Verhältnis aus Anzahl von nicht synonymen Substitutionen an einer Position des Genoms zu den synonymen Substitutionen an dieser Stelle. Ein hohes  $k_a/k_s$ -Verhältnis spricht für einen höheren selektiven Druck auf den Organismus, der mit einer Änderung der Aminosäurestruktur einhergeht.

Interessanterweise haben 12 der potentiellen Virulenzgene einen  $k_a/k_s$ -Wert größer als 1. Um mehr Informationen über diese potentiellen Virulenzgene zu erhalten und auch ihre Bedeutung besser einschätzen zu können, wurde die Stelle im Genom, an der diese Mutationen stattfinden, mit der Position von bereits bekannten Virulenzgenen innerhalb des Genoms verglichen. Hierbei zeigte sich, dass SNPs signifikant häufiger an Genlokalisationen auftreten, die bereits mit Pathogenität assoziiert sind als im Vergleich zu zufälligen SNP-Positionen ( $p$ -Wert= 0,035). Grafisch ist dies in der folgenden Abbildung 5 in Form eines Manhattan Plot dargestellt.

Abschließend lässt sich zu den Ergebnissen der WGS-Analyse sagen, dass das Gen *ebhB* hervorzuheben ist, da in 7 invasiven Isolaten innerhalb dieses Gens Mutationen gefunden wurden, während in diesem Gen in allen anderen kolonisierenden Isolaten keine Mutationen auftraten. Darüber hinaus wurde für *ebhB* eine  $K_a/K_s$ - Ratio von 4,92 errechnet, was auf eine hohe Signifikanz hinweist.



**Abbildung 5:** Manhattan Plot, der die Verteilung aller gefundener SNPs in Bezug auf das ganze Referenzgenom zeigt. Dabei sind grau markiert alle 535 gefundenen SNPs sowie grün alle SNPs, die sich in mindestens zwei oder mehr Isolaten fanden. Rot markiert sind bekannte Virulenzgene. Die Abbildung verdeutlicht die Nähe der grün markierten potentiellen Virulenzgene zu den bereits bekannten Virulenzgenen.

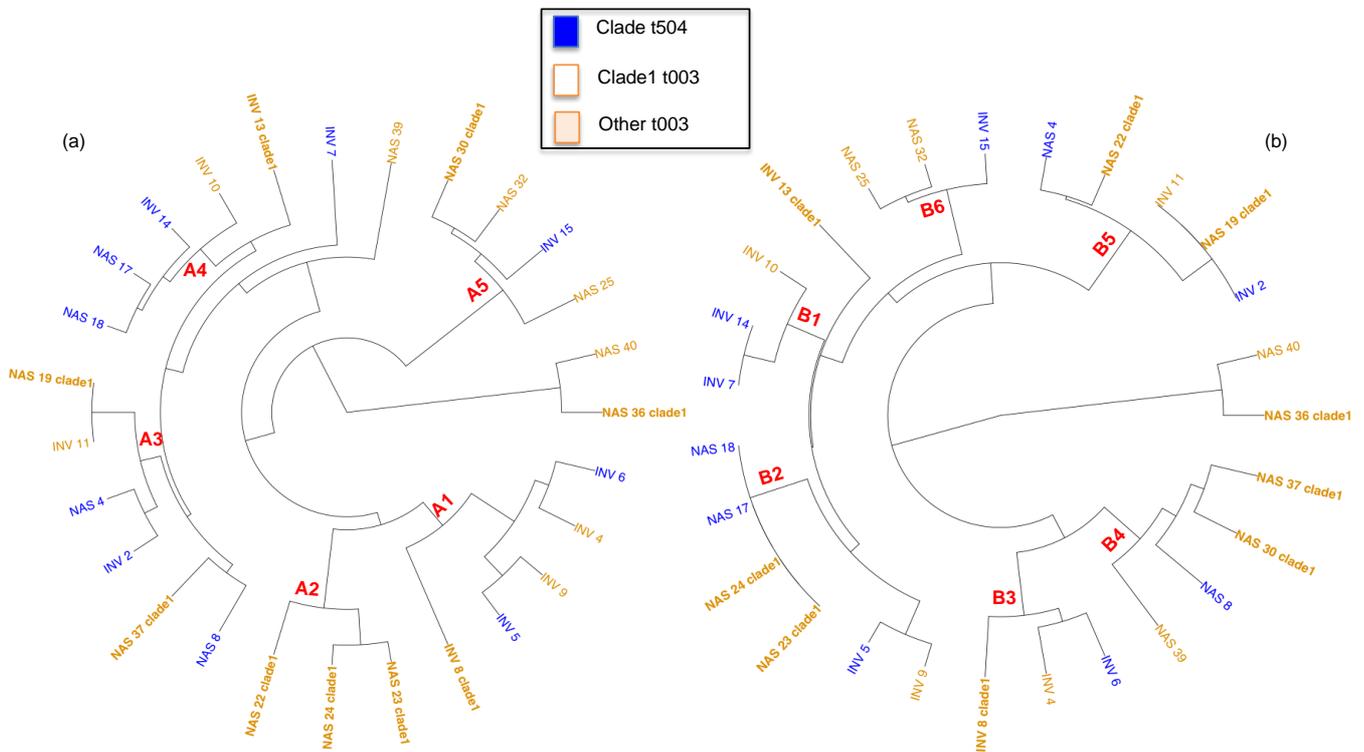
## Ergebnisse der DNA-Microarray Analyse

3.2 Die hierarchische Cluster-Analyse aller 330 auf dem Array vertretenen Gene [18] ergibt eine Einteilung der untersuchten MRSA-Isolate in 5 Cluster, in denen mindestens je drei Isolate vertreten sind (siehe Abbildung 6 a und 6b). In der Cluster-Gruppe A1 finden sich ausschließlich fünf invasive Isolate (P-Wert 0,01 im Fisher-Signifikanztest), die sich dadurch auszeichnen, dass sie für das Gen *hsdSxCC15* (das für eine Deoxyribonuklease-Untereinheit codiert) und das Gen *Q2YUB* (das für ein Effluxprotein codiert) positiv getestet wurden.

Dem gegenüber steht Gruppe A2, die nur Isolate enthält, die aus Nasenabstrichen stammen und positiv auf das Gen *ccrB.4* getestet wurden (P-Wert 0,23 im Fisher Signifikanztest). Die Gruppen A3 bis A5 sind gemischt mit keiner klaren Neigung zu invasiver oder kolonisierender Herkunft.

Eine Einteilung nur auf Basis der 174 annotierten Virulenzgene ergab eine Einteilung in 6 Cluster (Abbildung 6 b). In der Gruppe B1 und B3 finden sich invasive Isolate, die sich durch das Vorhandensein bestimmter Varianten des für Staphylokokken typischen superantigen-ähnlichen Proteins auszeichnen. In den Gruppen B2 und B4 fanden sich wiederum nur kolonisierende Isolate, während die anderen Gruppen gemischt zusammengesetzt waren. Ein Anwenden des Fisher-Signifikanztests ergab hier keine signifikanten P-Werte.

Abschließend wurde zusammenfassend versucht, einen Zusammenhang aus den Ergebnissen der SNP-Analyse, der Phylogenie und der Mikroarray-Gruppierung herzustellen. Hierbei ergab sich, dass die Microarray-Profile nicht mit der Phylogenie der untersuchten Microarray-Profile assoziiert sind. Es konnte aber eine statistisch signifikante Untergruppe von invasiven Isolaten (Cluster A1) ausgemacht werden.



**Abbildung 6 a/b:** Gruppierung anhand der Microarray-Analyse auf Basis aller 330 auf dem Array vertretenen Gene (a) und nur auf Basis der 174 Virulenzgene (b).

## 4 Diskussion

Das Konzept, ausgehend von dem mit WGS ermittelten Genotyp Rückschlüsse auf den Phänotyp zu ziehen, ist von hohem wissenschaftlichen Interesse. Hierzu bestehen auch bereits einige Arbeiten anderer Forschungsgruppen. Laabei *et al.* haben anhand von 90 MRSA-Isolaten ein Vorhersagemodell entwickeln können, das auf Basis von signifikanten SNP-Profilen an bestimmten Stellen des Genoms Aussagen über die Toxizität von *Staphylococcus aureus* treffen kann [36]. Dies war möglich mithilfe eines lernenden Algorithmus' und einer angeschlossenen Bestätigung im Tiermodell. Dies ist ein Ansatz, allein von der genetischen Sequenz eines Bakteriums ausgehend eine Aussage über das Potential zur Verursachung von schwerwiegenden Infektionen treffen zu können.

In dem Review von Priest *et al.* [37], das sich mit der gleichen Grundthematik beschäftigt, wird außerdem die These aufgestellt, dass es, auch wenn die

Genomstruktur bekannt ist, noch eine Vielzahl von weiteren Faktoren gibt, die die tatsächliche Virulenz *in vivo* beeinflussen, wie zum Beispiel die regulatorischen Prozesse und die Epistase, d.h. die Interaktion von Genen. Deshalb wird in der Arbeit von Priest *et al.* die Notwendigkeit bioinformatischer Methoden aufgezeigt, die komplexe genetische Zusammenhänge simulieren können.

Auch in dieser Arbeit wurde WGS benutzt, um SNPs in Genen zwischen der Gruppe der invasiven und der Gruppe der kolonisierenden Isolate aufzudecken mit dem Ziel, herauszufinden, ob ein Zusammenhang zwischen Genotyp und Virulenz besteht und insbesondere in welchen Genen dieser zu finden ist. Als Ergebnis konnten 36 Gene als potentiell neue Virulenzfaktoren ermittelt werden. Insbesondere das Gen *ebhB* kann mit einer Virulenz in *Staphylococcus aureus* in Verbindung gebracht werden. Das Auftreten von SNPs in in den invasiven Isolaten ist statistisch hochsignifikant. Außerdem gehen diese Mutationen mit einer Veränderung der Aminosäurestruktur einher: An Position 1503065 des Genoms befindet sich ein Cytosin anstelle eines Thymins, was zu einem Wechsel der Aminosäuren von Serin zu Asparagin führt. Dies stellt eine bedeutende Mutation dar, da die Proteinstruktur verändert wurde. *EbhB* ist in der Literatur als ein zellwandassoziiertes Oberflächenprotein beschrieben, welches menschliches Fibronectin binden kann und somit der Zelladhäsion dient, was bei Gram-positiven Bakterien ein gängiger Mechanismus ist und zur Pathogenität von *Staphylococcus aureus* beiträgt [38].

Die Expression von *EbhB* wird von dem Regulationssystem „accessory gene regulatur“ (*agr*), einem Gen, das die Virulenz in Infektionssituationen hochregeln kann, gesteuert [38]. *Agr* ist das wichtigste Regulationssystem, das verantwortlich dafür ist, ob der adhärenz oder der toxische Phänotyp von *Staphylococcus aureus* exprimiert wird, und beeinflusst viele Gene des akzessorischen Genoms.

Auch den anderen Genen, in denen sich SNP-Mutationen fanden (siehe Tabelle 2 im Anhang) und die Mutationen in nur zwei oder drei der invasiven Isolate aufwiesen (Fisher-Signifikanztest mit einem P-Wert von 0,2 - 0,7), sollte man Beachtung schenken und diese weiter auf eine Assoziation mit Virulenz hin untersuchen. Aufgrund ihrer Lage in der Nähe zu bekannten Virulenzgenen im Genom (siehe Abbildung 5) lässt sich vermuten, dass diese 20 Gene direkt oder mindestens indirekt mit Virulenz assoziiert sein könnten. Diese Tatsache unterstützt die

Hypothese, dass es sich bei den neuen potentiellen Virulenzgenen tatsächlich um Virulenzfaktoren handelt, da in *Staphylococcus aureus* diese häufig genetisch eng beieinander liegen und auf Pathogenitätsinseln organisiert sind. Darüber hinaus muss festgestellt werden, dass sich nur 40% der hier dargestellten Mutationen in annotierten Genabschnitten befinden, die sich in Datenbanken wie der VFDB-Datenbank [35] wiederfinden. Über die anderen 60% kann man im Moment noch wenig Aussagen treffen. Diese Tatsache wird sich aber aufgrund des Wissenszuwachses und der Komplettierung der Gendatenbanken in Zukunft ändern.

In dieser Arbeit wurde unter anderem die Phylogenie von MRSA-Isolaten aus dem CC5 anhand einer SNP-Analyse des Kerngenoms untersucht. Dies ist eine etablierte Technik für phylogenetische und epidemiologische Fragestellungen [13]. In einer großen über sechs Jahre laufenden und im Jahr 2011 veröffentlichten Studie von Schaumburg *et al.* zeigte sich, dass innerhalb des klonalen Komplexes 5 Stämme mit dem spa-Typ t003 die dominierenden Erreger in Deutschland sind [39]. In einer zur gleichen Zeit durchgeführten MRSA-Screening-Studie von Ruffing *et al.* an Patienten des Universitätsklinikums des Saarlands fand sich eine hohe Prävalenz eines neuen spa-Typs t504, dessen Herkunft und Entwicklungsgeschichte unbekannt war [26] und der großes Interesse bezüglich seiner Herkunft geweckt hat. Mit der in der vorgelegten Arbeit angewandten Technik der Vollgenomanalyse lässt sich nun zeigen, dass dieser saarländische t504-Stamm tatsächlich Vorfahren aus der großen Kohorte der in Deutschland weit verbreiteten t003-Stämme hat. Es ist somit nachvollziehbar, dass sich im Laufe der klonalen Verbreitung der Bakterien in jüngerer Vergangenheit ein MRSA-Klon abgespalten und sich dann im Saarland ein lokal erfolgreicher MRSA-Stamm etabliert hat. Diese Hypothese wird neben dem Fakt, dass alle Isolate mit dem Typ t504 eine eigene Klade bilden, auch dadurch unterstützt, dass innerhalb der Klade der t504-Isolate die Verwandtschaft untereinander enger ist. Des Weiteren hat WGS in dieser Arbeit zu dem Ergebnis geführt, dass die Isolate der Klade1 t003 als lokale endemische Klone von MRSA angesehen werden können, die sich von den anderen t003-Isolaten aufgrund des SNP-Profiles abgrenzen. Die Bedeutung liegt qualitativ darin, dass hier ein neuer, vorher wissenschaftlich noch nicht beschriebener MRSA Klon beschrieben wurde. Dies bestätigt die Bedeutung von lokal endemischen MRSA-Klonen und ihrem bekannten Verbreitungsmuster in Form von regionalen Klonen in Krankenhäusern

und ähnlichen Einrichtungen des Gesundheitswesens [13]. Da sich in der gleichen Klade auch Isolate von kolonisierenden MRSA von Aufnahmepatienten aus dem größeren Umkreis der Uniklinik wiederfinden, liegt die Vermutung nahe, dass es sich bei der Klade1 t003 tatsächlich um einen regionalen Klon handelt, der in der Region des Saarlandes endemisch ist. Hiermit reproduzieren wir die Ergebnisse von Nübel *et al.* [13] und zeigen abermals, dass WGS gut geeignet ist, lokal endemische MRSA-Klone zu erfassen. Zusammenfassend zeigt sich, dass anhand der WGS-Analyse keine Korrelation zwischen Virulenz und Phylogenie in der hier untersuchten Studienkohorte besteht. Das bedeutet, dass WGS, so wie es auch in dieser Arbeit angewandt wurde, zwar dafür geeignet ist, MRSA phylogenetisch zu gruppieren, dass es aber nicht möglich ist, einen direkten Zusammenhang zwischen phylogenetischer Zugehörigkeit und Virulenz herzustellen. Als Grund hierfür ist anzunehmen, dass der horizontale Genaustausch unabhängig von der Phylogenie stattfindet [40].

Ergänzend zu der WGS-Analyse wurde die Microarray-Analyse durchgeführt und hierbei ergab sich in der Cluster-Analyse folgendes Resultat: Anhand der genetischen Profile der DNA-Microarray-Profile konnte nur innerhalb der Gruppe der invasiven Isolate ein eigenständiges Cluster ausfindig gemacht werden, das sich durch bekannte Virulenzfaktoren von den anderen Isolaten der Gruppe A1 abgrenzt (Abbildung 6 a). Im Gegensatz zu der vorausgegangenen SNP-Analyse des Kerngenoms wurden in der DNA-Microarray-Analyse Isolate unterschiedlicher Abstammung nicht getrennt, das heißt, es finden sich sowohl t003-Isolate als auch t504-Isolate. Diese Art der Typisierung ist also dafür geeignet, eine Untergruppe von invasiven Isolaten zu charakterisieren, die sich durch ein charakteristisches Hybridisierungsprofil im DNA-Microarray auszeichnen.

Zusammenfassend zeigt sich, dass sich im Rahmen dieser Arbeit die Genomanalyse mittels DNA-Microarray einfacher und praktikabler gezeigt hat, um Aussagen über die Virulenz zu treffen, während die WGS-Analyse interessante Antworten bezüglich der Entstehung des im Saarland endemischen t504 spa-Typs geben konnte. Die WGS-Analyse ist darüber hinaus als einzige Methode in der Lage, die Phylogenie darzustellen. Darüber hinaus hat sich WGS als sehr gutes Verfahren erwiesen, um potentielle neue Virulenzfaktoren zu entdecken und um unser Wissen über *Staphylococcus aureus* als humanpathologisches Bakterium zu erweitern. Aufgrund der

Möglichkeit, alle Typisierungen und Genomanalysen in Zukunft alleine auf Basis einer WGS-Analyse durchzuführen, zeigt sich WGS den anderen Typisierungsmethoden überlegen. So wurde bereits in anderen Arbeiten, z.B. von Bartels *et al.*, gezeigt, dass WGS mit einer hohen Zuverlässigkeit auch dafür geeignet ist, den spa-Typ zu erfassen [41]. Darüber hinaus wäre es prinzipiell auch möglich, ein virulenzbasiertes Clustering, wie es die Microarray-Analyse erlaubt, anhand der WGS-Daten zu erheben. Um diesen Weg zu beschreiten, ist es jedoch notwendig, die zur Verfügung stehende komplette Datenmenge des Genoms optimal zu verarbeiten und mit vollständigen Referenzgenomsequenzen abzugleichen, nach Möglichkeit inklusive des akzessorischen Genoms. An dieser Stelle fehlt es aber noch an standardisierten Verfahrensweisen, insbesondere im bioinformatischen Teilbereich, um reproduzierbare Ergebnisse zu erzielen. Aufgrund der weiterhin stattfindenden Kostensenkung und der immer breiteren Anwendung dieser Methode und ihrer vielseitigen Möglichkeiten kann man WGS als neuen Goldstandard zum Typisieren von MRSA bezeichnen.

Diese Studie kann als Pilotstudie angesehen werden, die die Durchführbarkeit und die Praktikabilität der Vollgenomanalyse für MRSA zeigt. In Zukunft sollten sich weitere, überregionale Studien mit einer höheren Fallzahl, sowie möglicherweise weiteren Protein A-Typen anschließen. Für die im Rahmen dieser Studie durchgeführten Analysen ist kritisch anzumerken, dass die Anzahl der invasiven Isolate hätte höher sein sollen, um den Daten mehr Aussagekraft zu verleihen.

Es ist weiter anzumerken, dass die im Rahmen dieser Arbeit erhobenen Ergebnisse der WGS-Analyse nur auf dem einen Referenzgenom NC\_017340 [28] basieren. Hiermit wurde den mobilen genetischen Elementen, die bis zu 20% des Genoms ausmachen, nicht genügend Beachtung geschenkt. Andere Arbeiten, wie zum Beispiel die Studie von Leopold *et al.* [42], arbeiten mit *de novo*-Assemblierung, was jedoch einen viel größeren technischen und zeitlichen Aufwand darstellen würde, und im Rahmen dieser Pilotstudie nicht durchführbar war. Eine andere einfachere Alternative wäre, mehrere unterschiedliche Referenzgenome zu benutzen. An dieser Stelle soll auch noch einmal der Unterschied zwischen dem Kerngenom und den restlichen mobilen genetischen Elementen hervorgehoben werden. Das Kerngenom ist hochkonserviert, *d.h.* Isolate, die derselben Abstammungslinie angehören, unterscheiden sich trotz großer lokaler und temporärer Unterschiede innerhalb ihrer

codierenden Sequenz nur durch wenige hundert SNPs. Im Gegensatz dazu können die Stämme unterschiedliche mobile genetische Elemente aufweisen und ein unterschiedliches Virulenzverhalten zeigen. Um diesem Sachverhalt gerecht zu werden, besteht ein Bedarf an neuen bioinformatischen Konzepten.

Dass das Gen *EbhB* mit Virulenz assoziiert ist, konnte in dieser Arbeit gezeigt werden. Für weitere Studien wäre es interessant zu ergründen, ob die in *EbhB* gefundenen Mutationen von vornherein auf eine gesteigerte Virulenz schließen lassen oder diese als Resultat einer Anpassung der Bakterien unter den Bedingungen einer Infektion innerhalb des infizierten Wirts auftreten. Hier wäre es notwendig, weitere Untersuchungen anzustellen, idealerweise anhand eines Genomvergleichs ein und desselben MRSA-Klons vor und nach einer invasiven Infektion (möglicherweise an einem Tiermodell). Ergänzend könnten Funktionsanalysen des *ebhB*-Genprodukts angeschlossen werden.

Insbesondere der Zusammenhang von Genotyp und daraus resultierendem Phänotyp erfordert neue innovative bioinformatische Modelle. Es wäre empfehlenswert, die Methode des WGS weiter zu entwickeln, um zu einer detaillierteren Vorstellung und einem besseren Verständnis von *Staphylococcus aureus* und auch den im Rahmen dieser Arbeit identifizierten potentiellen neuen Virulenzfaktoren, vor allem *ebhB*, zu gelangen. Es ist zu konstatieren, dass es heutzutage keine großen finanziellen und technischen Probleme mehr darstellt, das gesamte Erbgut eines Bakteriums wie *Staphylococcus aureus* bis auf jedes einzelne Basenpaar auszulesen, aber dass wir mit dem Verarbeiten der darin enthaltenen Informationen noch viele Herausforderungen vor uns liegen haben. So lässt sich in Zukunft vielleicht noch besser verstehen, bei welchen MRSA-Stämmen aus harmlosen Kommensalen für den Menschen gefährliche invasive Krankheitserreger werden. Auch für die klinische Medizin stellt dies einen vielversprechenden Ausblick dar, um die Diagnostik möglichst effizient durchzuführen. Insbesondere Infektionsketten sind mit WGS sehr gut nachvollziehbar und auf Ausbruchssituationen kann schneller reagiert werden. Es ist denkbar, dass die antimikrobielle Therapie in Zukunft vielleicht einmal gezielt gegen den Genotyp des vorgefunden Krankheitserregers gerichtet werden kann.

## 5 Literaturverzeichnis

1. Ogston A. Report upon Micro-Organisms in Surgical Diseases. *British medical journal*. 1881 Mar 12;1(1054):369 b2-75.
2. Rosenbach A. Mikro - Organismen bei den Wund-Infektions-Krankheiten des Menschen. Wiesbaden, J.F. Bergmann. 1884.
3. Lowy FD. Staphylococcus aureus infections. *The New England journal of medicine*. 1998 Aug 20;339(8):520-32.
4. Jevons MP. *Brit med J* 1961;1:124-5.
5. Enright MC, Robinson DA, Randle G, Feil EJ, Grundmann H, Spratt BG. The evolutionary history of methicillin-resistant Staphylococcus aureus (MRSA). *Proceedings of the National Academy of Sciences of the United States of America*. 2002 May 28;99(11):7687-92.
6. Gastmeier P, Sohr D, Geffers C, Behnke M, Daschner F, Ruden H. Mortality risk factors with nosocomial Staphylococcus aureus infections in intensive care units: results from the German Nosocomial Infection Surveillance System (KISS). *Infection*. 2005 Apr;33(2):50-5.
7. Geffers C, Zuschneid I, Sohr D, Ruden H, Gastmeier P. [Microbiological isolates associated with nosocomial infections in intensive care units: data of 274 intensive care units participating in the German Nosocomial Infections Surveillance System (KISS)]. *Anesthesiologie, Intensivmedizin, Notfallmedizin, Schmerztherapie : AINS*. 2004 Jan;39(1):15-9.
8. Klein E, Smith DL, Laxminarayan R. Hospitalizations and deaths caused by methicillin-resistant Staphylococcus aureus, United States, 1999-2005. *Emerging infectious diseases*. 2007 Dec;13(12):1840-6.
9. Gordon RJ, Lowy FD. Pathogenesis of methicillin-resistant Staphylococcus aureus infection. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*. 2008 Jun 1;46 Suppl 5:S350-9.
10. Edwards AM, Massey RC, Clarke SR. Molecular mechanisms of Staphylococcus aureus nasopharyngeal colonization. *Molecular oral microbiology*. 2012 Feb;27(1):1-10.
11. Zetola N, Francis JS, Nuermberger EL, Bishai WR. Community-acquired methicillin-resistant Staphylococcus aureus: an emerging threat. *The Lancet infectious diseases*. 2005 May;5(5):275-86.
12. Wichelhaus A, Volker Schäfer und Volker Brade. Typisierungsverfahren in der Infektionsepidemiologie. *Chemotherapie Journal*. 2000(2).
13. Nubel U, Nitsche A, Layer F, Strommenger B, Witte W. Single-nucleotide polymorphism genotyping identifies a locally endemic clone of methicillin-resistant Staphylococcus aureus. *PloS one*. 2012;7(3):e32698.
14. Sabat AJ, Budimir A, Nashev D, Sa-Leao R, van Dijl J, Laurent F, et al. Overview of molecular typing methods for outbreak detection and epidemiological surveillance. *Euro surveillance : bulletin European sur les maladies transmissibles = European communicable disease bulletin*. 2013;18(4):20380.
15. Trindade PA, McCulloch JA, Oliveira GA, Mamizuka EM. Molecular techniques for MRSA typing: current issues and perspectives. *The Brazilian journal of infectious diseases : an official publication of the Brazilian Society of Infectious Diseases*. 2003 Feb;7(1):32-43.

16. Koolman K-HR. Taschenatlas Biochemie des Menschen. 2009.
17. Deurenberg RH, Vink C, Kalenic S, Friedrich AW, Bruggeman CA, Stobberingh EE. The molecular evolution of methicillin-resistant *Staphylococcus aureus*. *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases*. 2007 Mar;13(3):222-35.
18. Monecke S, Slickers P, Ehricht R. Assignment of *Staphylococcus aureus* isolates to clonal complexes based on microarray analysis and pattern recognition. *FEMS immunology and medical microbiology*. 2008 Jul;53(2):237-51.
19. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*. 1977 Dec;74(12):5463-7.
20. Pavlopoulos GA, Oulas A, Iacucci E, Sifrim A, Moreau Y, Schneider R, et al. Unraveling genomic variation from next generation sequencing data. *BioData mining*. 2013;6(1):13.
21. Kuroda M, Ohta T, Uchiyama I, Baba T, Yuzawa H, Kobayashi I, et al. Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*. *Lancet*. 2001 Apr 21;357(9264):1225-40.
22. Lindsay JA, Holden MT. Understanding the rise of the superbug: investigation of the evolution and genomic variation of *Staphylococcus aureus*. *Functional & integrative genomics*. 2006 Jul;6(3):186-201.
23. Lindsay JA, Knight GM, Budd EL, McCarthy AJ. Shuffling of mobile genetic elements (MGEs) in successful healthcare-associated MRSA (HA-MRSA). *Mobile genetic elements*. 2012 Sep 1;2(5):239-43.
24. Plata K, Rosato AE, Wegrzyn G. *Staphylococcus aureus* as an infectious agent: overview of biochemistry and molecular genetics of its pathogenicity. *Acta biochimica Polonica*. 2009;56(4):597-612.
25. Francis JS, Doherty MC, Lopatin U, Johnston CP, Sinha G, Ross T, et al. Severe community-onset pneumonia in healthy adults caused by methicillin-resistant *Staphylococcus aureus* carrying the Panton-Valentine leukocidin genes. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*. 2005 Jan 1;40(1):100-7.
26. Ruffing U, Akulenko R, Bischoff M, Helms V, Herrmann M, von Muller L. Matched-cohort DNA microarray diversity analysis of methicillin sensitive and methicillin resistant *Staphylococcus aureus* isolates from hospital admission patients. *PLoS one*. 2012;7(12):e52487.
27. Harmsen D, Claus H, Witte W, Rothganger J, Claus H, Turnwald D, et al. Typing of methicillin-resistant *Staphylococcus aureus* in a university hospital setting by using novel software for spa repeat determination and database management. *Journal of clinical microbiology*. 2003 Dec;41(12):5442-8.
28. *Staphylococcus aureus* 04-02981 chromosome, complete genome NC\_017340. at [http://www.ncbi.nlm.nih.gov/nuccore/NC\\_017340](http://www.ncbi.nlm.nih.gov/nuccore/NC_017340).
29. Nubel U, Dordel J, Kurt K, Strommenger B, Westh H, Shukla SK, et al. A timescale for evolution, population expansion, and spatial spread of an emerging clone of methicillin-resistant *Staphylococcus aureus*. *PLoS pathogens*. 2010 Apr;6(4):e1000855.
30. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009 Aug 15;25(16):2078-9.

31. Langmead B. Aligning short sequencing reads with Bowtie. *Current protocols in bioinformatics / editorial board, Andreas D Baxevanis [et al]*. 2010 Dec;Chapter 11:Unit 11 7.
32. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome research*. 2012 Mar;22(3):568-76.
33. Gouy M, Guindon S, Gascuel O. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular biology and evolution*. 2010 Feb;27(2):221-4.
34. FigTree. (2009). at <<http://tree.bio.ed.ac.uk/software/figtree/>>.
35. Chen L, Yang J, Yu J, Yao Z, Sun L, Shen Y, et al. VFDB: a reference database for bacterial virulence factors. *Nucleic acids research*. 2005 Jan 1;33(Database issue):D325-8.
36. Laabei M, Recker M, Rudkin JK, Aldeljawi M, Gulay Z, Sloan TJ, et al. Predicting the virulence of MRSA from its genome sequence. *Genome research*. 2014 May;24(5):839-49.
37. Priest NK, Rudkin JK, Feil EJ, van den Elsen JM, Cheung A, Peacock SJ, et al. From genotype to phenotype: can systems biology be used to predict *Staphylococcus aureus* virulence? *Nature reviews Microbiology*. 2012 Nov;10(11):791-7.
38. Clarke SR, Harris LG, Richards RG, Foster SJ. Analysis of Ebh, a 1.1-megadalton cell wall-associated fibronectin-binding protein of *Staphylococcus aureus*. *Infection and immunity*. 2002 Dec;70(12):6680-7.
39. Schaumburg F, Kock R, Mellmann A, Richter L, Hasenberg F, Kriegeskorte A, et al. Population dynamics among methicillin-resistant *Staphylococcus aureus* isolates in Germany during a 6-year period. *Journal of clinical microbiology*. 2012 Oct;50(10):3186-92.
40. Juhas M. Horizontal gene transfer in human pathogens. *Critical reviews in microbiology*. 2015 Feb;41(1):101-8.
41. Bartels MD, Petersen A, Worning P, Nielsen JB, Larner-Svensson H, Johansen HK, et al. Comparing whole-genome sequencing with Sanger sequencing for spa typing of methicillin-resistant *Staphylococcus aureus*. *Journal of clinical microbiology*. 2014 Dec;52(12):4305-8.
42. Leopold SR, Goering RV, Witten A, Harmsen D, Mellmann A. Bacterial whole-genome sequencing revisited: portable, scalable, and standardized analysis for typing and detection of virulence and antibiotic resistance genes. *Journal of clinical microbiology*. 2014 Jul;52(7):2365-70.

## 6 Publikationen / Danksagung

### Publikation im Rahmen der Dissertation

Eingereichte Publikation (In Revision)

6.1 Journal: Infection, Genetics and Evolution  
Corresponding Author: Lutz von Müller

Co-Authors: Mohamed Hamed; Daniel P Nitzsche, PhD; Ulla Ruffing; Mathias Steglich; Janina Dordel, PhD; Duy Nguyen; Jan–Hendrik Brink; Gursharan S Chhatwal; Mathias Herrmann *et al.* Whole Genome Sequence Typing and Microarray Profiling of Commensal and Invasive Methicillin-Resistant *Staphylococcus aureus* Isolates: Clues to Phylogeny and Invasiveness

### 6.2 Danksagung

Mein Dank gilt Herrn Professor Dr. med. Lutz von Müller für die Betreuung der vorgelegten Arbeit sowie für wertvolle Hinweise und Anregungen. Des Weiteren gilt mein Dank der Arbeitsgruppe von Prof. Dr. Singh Chhatwal, insbesondere Dr. Patric Nitsche-Schmitz vom Helmholtz-Zentrum für Infektionsforschung in Braunschweig, die die Vollgenomsequenzierung ermöglicht hat. Bedanken möchte ich mich darüber hinaus bei der Arbeitsgruppe von Herrn Professor Dr. Volkhard Helms vom Institut für Bioinformatik in Saarbrücken, insbesondere bei Mohamed Hamed für die Auswertung der bioinformatischen Daten und das Erstellen der Grafiken, die in dieser Arbeit verwendet wurden, sowie für das ausführliche Erläutern der bioinformatischen Abläufe, die für den theoretischen Hintergrund der hier vorgelegten Arbeit von großer Relevanz waren.

Ich danke schließlich meinen Eltern für die Realisierung und fortlaufende Unterstützung meines Medizinstudiums.

## 7 Anhang

### Protokoll der spa-Typisierung

Die Versuchsanordnung wurde wie nachfolgend beschrieben durchgeführt. Die verwendeten Reagenzien stammen von den Firmen Qiagen (Dann-Aufreinigung) und

#### 7.1 Beckman Coulter (Sequenzierung).

- Eine Bakterienkolonie gründlich in 200 µl Aqua dest. einrühren
- 10 min bei 95°C kochen (Lyse der Zellen)
- 3 min zentrifugieren bei 130 rpm, um Bakterienbestandteile zum Sedimentieren zu bringen
- PCR wie folgt: 10 µl Aqua, 12,5 µl PCR Mastermix 0,25 µl Vorwärtsprimer spa1113f und 0,25 µl Rückwärtsprimer spa1514r ([www.ridom.de](http://www.ridom.de)) mit dem Programm 7 min bei 95°C, dann 30 Zyklen: 95°C für 30 s, 55°C für 1 min, 72°C für 3 min, 72°C für 10 min. Danach 72°C für 10 min, dann Aufbewahrung bei 10°C
- Kontrollgel: Gellauf bei 150V für 35 min
- Verdau des PCR Produktes mit 4 µl Exonuclease (20 min inkubieren, um Enzym zu aktivieren und danach 20 min inkubieren, um Enzym wieder zu deaktivieren)
- Sequenzierungs PCR mit je einem Vorwärts- und einem Rückwärtsansatz
- PCR Aufreinigung mit Agencourt AMPure XP PCR Purification Kit: 10 µl Amplifikat mit 18 µl AMPure XP vermischen. Die dort enthaltenden Magnetkugeln binden die DNA.
- 2 min auf die im Kit enthaltenden Magnetplatte stellen, um die Magnetkugeln mit gebundener DNA zu isolieren
- Überstand verwerfen, ohne die Magnetkügelchen zu berühren
- 2fach mit Ethanol waschen
- Platte von der Magnetplatte nehmen und mit 30 µl Elutionspuffer die DNA herauslösen
- Platte wieder auf die Magnetplatte stellen und die gereinigte DNA abpipettieren. Die Magnetkügelchen bleiben zurück.

Im letzten Schritt kann das so aufgereinigte Amplifikat auf die Platten des Sequencers aufgetragen und sequenziert werden.

## Protokoll der Microarray-Messung

Die Versuchsanordnung wurde wie nachfolgend beschrieben durchgeführt. Die 7.2 verwendeten Reagenzien stammen von der Firma Alere Technologies und sind alle im "S. aureus Genotyping Kit 2.0" enthalten. Im ersten Schritt wird die DNA in folgenden Schritten aufgereinigt:

- Eppendorfgefäß aus dem Clondiag-Kit 1 min bei 130.000 Umdrehungen herunterzentrifugieren, um Lysepellets auf den Boden zu bringen
- 200 µl A1 Lysepuffer in Eppendorfgefäß hinzugeben und vortexen
- 5 min zum Lösen stehen lassen
- Bakterienmaterial mit Hilfe einer Impföse hinzufügen
- Bei 37°C auf Thermorüttler 55rpm zur Enzymaktivierung inkubieren
- 200 µl ALL Lysepuffer sowie 40 µl Proteinkinase K zugeben und vortexen
- 45 min bei 56°C auf Thermorüttler bei 550rpm zur Enzymdeaktivierung inkubieren
- Anzentrifugieren und 200 µl Ethanol hinzugeben, sodass DNA ausfällt
- Übergabe auf Qiagen-Kit-Säule
- 1 min bei 6000 g zentrifugieren, Durchfluss verwerfen und Säule auf ein neues Tube aus dem Kit stecken
- 500 µl Waschpuffer AW1 auf die Säule geben
- 1 min bei 6000g zentrifugieren
- Durchfluss verwerfen und wieder auf eine neue Säule übertragen
- 500 µl Waschpuffer2 AW2
- 3 min auf 13000g zentrifugieren, dann Durchfluss verwerfen und nochmal 1 min zentrifugieren

- Säule auf ein frisches Eppendorfgefäß stellen, 40 µl Elutionspuffer hinzufügen und 4 min bei Raumtemperatur inkubieren
- Im letzten Schritt 1 min bei 6000 g abzentrifugieren, um die konzentrierte DNA zu erhalten

Danach erfolgt eine Bestimmung der DNA-Konzentration mittels Nanodrop 2000-Messung und anschließend die Messung am Microarray-Gerät ArrayMate (Alere Technologies) nach folgendem Protokoll:

- Vorlegen von 4,9 µl B1 (Labelling Buffer/Mastermix) + 0,1 µl B2 (Labelling-Enzym)
- Hinzugeben der zuvor aufgereinigten DNA (Menge angepasst an die jeweilige Konzentration)
- Mischen und zentrifugieren
- PCR zum Einbau der Farbe nach folgendem Schema: Deckel vorheizen auf 105°C, dann 300 sec bei 96°C und danach 20 sec bei 50°C. Anschließend 45 Zyklen mit je 30 sec bei 72°C und 20 sec bei 96°C
- Nach PCR-Lauf 90 µl C1-Hybridisierungspuffer in jedes Eppendorfgefäß geben und mischen
- Die Cups des Clondiag-Kits spülen und mit 200 µl C1 Puffer bei 55°C inkubieren
- Den PCR-Mix hinzugeben und bei 55°C 1 Stunde lang inkubieren
- 1 µl C3 (Meerrettich-Peroxidase) + 99 µl C4 (Konjugatpuffer) pro Cup vorbereiten
- Reagenzien in den Cups verwerfen und 3 mal mit Waschpuffer C2 spülen
- Je 100 µl des C3/C4 Mastermixes hinzugeben und bei 30°C 10 min inkubieren
- Nach den 10min C3/C4 entfernen und verwerfen und 2mal mit C5 Waschpuffer2 waschen
- 100 µl D1 (Meerrettich-Peroxidase) in jedes einzelne Cup pipettieren und 5 min bei Raumtemperatur stehenlassen

Nun kann die Messung am Gerät durchgeführt werden.

## Protokoll der DNA Aufreinigung für WGS

Die Versuchsanordnung wurde wie nachfolgend beschrieben durchgeführt. Die verwendeten Reagenzien stammen von der Firma Qiagen und sind in dem DNeasy Blood and Tissue Kit enthalten.

7.3

- Zellen aus dem BHI-Medium durch 7 min Zentrifugieren bei 3700 g ernten
- Mit PBS-Lösung waschen
- Bakterienpellets in 180 µl TE Puffer und 20 µl RNase der Firma Applichem geben
- Die Suspension auf Zirkon Kugeln geben und im Zellhomogenisierer 30 sec bearbeiten
- 200 µl AL Puffer hinzugeben und Kugeln herunterzentrifugieren
- Das Bakterienlysat in ein neues Eppendorfgefäß geben, 25 µl Proteinase K hinzugeben und bei 56 C inkubieren
- 200 µl Ethanol hinzufügen, um die DNA auszufällen
- 1 min bei 6000 g zentrifugieren, Durchfluss verwerfen und Säule auf ein neues Tube aus dem Kit stecken
- 500 µl Waschpuffer AW1 auf die Säule geben
- 1 min bei 6000g zentrifugieren
- Durchfluss verwerfen und wieder auf eine neue Säule übertragen
- 500 µl Waschpuffer2 AW2
- 3 min auf 13000 g zentrifugieren, dann Durchfluss verwerfen und noch einmal 1 min zentrifugieren
- Säule auf ein frisches Eppendorfgefäß stellen, 40 µl Elutionspuffer hinzufügen und 4 min bei Raumtemperatur inkubieren
- Im letzten Schritt 1 min bei 6000 g abzentrifugieren, um die konzentrierte DNA zu erhalten

**Tabelle 2: Potentielle Virulenzgene**

7.4

LOCUS_TAG	GENE NAME	Description	SNP_POS	REF	ALT	Amino acid change	P-value of the variant
SA2981_0120	sbnD	Siderophore staphylobactin biosynthesis protein SbnD	131858	G	T	W to C	4.05E-51
SA2981_0148	-		162408	T	C	none (D)	2.21E-50
SA2981_0542	-		616048	A	G	none (G)	2.83E-52
SA2981_0561	-		631881	G	A	R to K	3.54E-53
SA2981_0710	-		783925	T	C	none (I)	1.36E-48
SA2981_0711	-		784904	A	G	Q to R	2.83E-61
SA2981_0724	-		797606	G	GA	ORF shifted	1.81E-47
SA2981_0874	-		921757	G	A	G to D	2.97E-53
SA2981_0978	-		1035879, 1036742	A, G	G, A	I to V; M to I	4.15E-40, 1.86E-70
SA2981_1074	-		1131989	C	T	T to I	3.51E-44
SA2981_1100	mutS2	Recombination inhibitory protein MutS2	1157221	C	T	none (S)	9.61E-67
SA2981_1178	prkC	Serine/threonine protein kinase PrkC, regulator of stationary phase	1237441	G	T	A to S	2.81E-52
SA2981_1251	-		1324647	C	A	T to N	2.45E-37
SA2981_1256	glpF	Glycerol uptake facilitator protein	1331327	C	CT	ORF shifted	2.41E-29
SA2981_1260	miaA	tRNA delta(2)-isopentenylpyrophosphate transferase	1336837	A	G	D to G	8.05E-45
SA2981_1284	thrC	Threonine synthase	1361487	C	T	S to F	1.61E-30
SA2981_1288	-		1365826	T	C	T to A	1.59E-26
SA2981_1323	trpD	Anthranilate phosphoribosyltransferase	1409763	T	C	R to R	2.78E-38
SA2981_1390	ebhB	Putative Staphylococcal surface anchored protein; adhesin emb	1503065	C	T	S to N	7.67E-41
SA2981_1468	gnd	6-phosphogluconate dehydrogenase, decarboxylating	1585512	G	A	none (F)	2.40E-23
SA2981_1511	sodA	Manganese superoxide dismutase; Superoxide dismutase (Fe)	1622766	C	T	G to D	2.46E-33
SA2981_1815	pfoS/R	Regulatory protein	1946895	C	CAAT	add R	1.93E-31
SA2981_1826	tagG	Teichoic acid translocation permease protein TagG	1967972	A	T	F to Y	2.88E-51
SA2981_2019	kdpD	Osmosensitive K <sup>+</sup> channel histidine kinase KdpD	2151346	G	GA	ORF shifted	1.37E-53
SA2981_2265	metT	Methionine transporter MetT	2394447	A	T	none (G)	5.88E-62

LOCUS_TAG	GENE_NAME	Description	SNP_POS	REF	ALT	Amino acid change	P-value of the variant
SA2981_2284	-		2414503	G	A	A to T	7.91E-77
SA2981_2294	tcaB	Teicoplanin resistance associated membrane protein TcaB	2424068	A	G	I to T	1.70E-62
SA2981_2329	-		2462332	C	T	none (L)	1.14E-59
SA2981_2366	-		2504026	G	A	P to S	8.70E-67
SA2981_2367	-		2504949	A	G	I to V	3.05E-58
SA2981_2370	-		2507139	C	CA	ORF shifted	1.69E-24
SA2981_2400	opp-1F	Oligopeptide transporter putative ATPase domain protein	2541624	C	T	A to T	4.11E-76
SA2981_2486	feoB	Ferrous iron transport protein B	2637050	C	T	V to M	3.31E-61
SA2981_2556	-		2710563	T	C	none (K)	2.77E-50
SA2981_2564	yvcP	Two-component response regulator YvcP	2721091	G	T	P to T	2.38E-71
SA2981_2642	-		2812721	G	T	R to L	3.21E-48