

Aus der Klinik für Infektiologie und Mikrobiologie  
der Universität zu Lübeck  
Direktor: Prof. Dr. med. Jan Rupp

**Identifikation und Charakterisierung von Markergensequenzen**  
**zur Differenzierung von klonalen Linien**  
**von *Staphylococcus epidermidis***

Inauguraldissertation  
zur  
Erlangung der Doktorwürde  
der Universität zu Lübeck

-Aus der Sektion Medizin-

Vorgelegt von  
  
Philipp Maximilian Maurer

aus Haan

1. Berichterstatter: Prof. Dr. Knobloch

2. Berichterstatter: Prof. Dr. Manz

Tag der mündlichen Prüfung: 20.08.2019

Zum Druck genehmigt: Lübeck, den 20.08.2019

Promotionskommission der Sektion Medizin

# Inhaltsverzeichnis

|                                                                              |           |
|------------------------------------------------------------------------------|-----------|
| Inhaltsverzeichnis                                                           | I         |
| Tabellenverzeichnis                                                          | IV        |
| Abbildungsverzeichnis                                                        | VI        |
| Abkürzungsverzeichnis                                                        | VII       |
| <b>1. Einleitung</b>                                                         | <b>1</b>  |
| 1.1 Der Genus <i>Staphylococcus</i>                                          | 1         |
| 1.2 Das Mikrobiom der Haut                                                   | 3         |
| 1.3 Koagulase-negative Staphylokokken                                        | 4         |
| 1.4 Die Pathogenität der Koagulase-negativen Staphylokokken                  | 5         |
| 1.5 Der Prozess der Biofilmbildung durch <i>S. epidermidis</i>               | 7         |
| 1.6 Die Differenzierung von pathogenen und kommensalen <i>S. epidermidis</i> | 10        |
| 1.7 Die SCCmec-Typisierung                                                   | 10        |
| 1.8 <i>Multi-Locus-Sequence-Typing</i>                                       | 13        |
| 1.8.1 Das Prinzip des MLST-Verfahren                                         | 13        |
| 1.8.2 Das MLST-Verfahren für <i>Staphylococcus epidermidis</i>               | 14        |
| 1.9 Das erweiterte MLST <sup>+</sup> -Verfahren und das core-Genom           | 15        |
| 1.10 Fragestellung                                                           | 16        |
| <b>2. Material</b>                                                           | <b>18</b> |
| 2.1 Bakterienstämme                                                          | 18        |
| 2.2 Einwegprodukte                                                           | 18        |
| 2.3 Chemikalien und Nährmedien                                               | 18        |
| 2.4 Geräte                                                                   | 19        |
| 2.5 Kits                                                                     | 19        |
| 2.6 Programme und Datenbanken                                                | 20        |

|           |                                                                      |           |
|-----------|----------------------------------------------------------------------|-----------|
| <b>3.</b> | <b>Methoden</b>                                                      | <b>21</b> |
| 3.1       | Kultur und Anzuchtbedingungen                                        | 21        |
| 3.2       | Isolierung der chromosomalen DNA der<br>Mikroorganismen              | 21        |
| 3.3       | Qualitäts- und Quantitätsbestimmung                                  | 22        |
| 3.4       | Erstellung der DNA- <i>libraries</i>                                 | 23        |
| 3.5       | Sequenzierung der DNA mittels <i>Next-Generation<br/>Sequencing</i>  | 24        |
| 3.5.1     | Cluster-Bildung                                                      | 25        |
| 3.5.2.    | Sequenzierung                                                        | 26        |
| 3.6       | Sequenzanalysen                                                      | 27        |
| 3.7       | Schema der Identifizierung von Markersequenzen                       | 28        |
| 3.8       | Identifizierung von Homologien zu bereits<br>annotierten Genomen     | 29        |
| 3.9       | <i>Multi Locus Sequence Typing</i>                                   | 30        |
| 3.10      | Core-Genom und MLST+                                                 | 30        |
| 3.11      | Bestimmung der Varianz des markerpositiven<br><i>contigs</i>         | 32        |
| 3.12      | Statistische Auswertung                                              | 34        |
| <b>4.</b> | <b>Ergebnisse</b>                                                    | <b>35</b> |
| 4.1       | Erstellung eines <i>core</i> -Genoms                                 | 35        |
| 4.2       | Charakterisierung des Sequenztyps 5 von<br><i>S. epidermidis</i>     | 35        |
| 4.2.1     | Identifizierung der Markersequenz                                    | 35        |
| 4.2.2     | Klonale Verwandtschaft in Relation zum<br>Auftreten des Markers      | 38        |
| 4.2.3     | Aufbau des Markers                                                   | 40        |
| 4.2.4     | Identifizierung der Varianz des<br>markerpositiven Sequenzabschnitts | 41        |
| 4.3       | Charakterisierung des Sequenztyps 59 von<br><i>S. epidermidis</i>    | 44        |
| 4.3.1     | Identifizierung der Markersequenz                                    | 44        |

|            |                                                                              |             |
|------------|------------------------------------------------------------------------------|-------------|
| 4.3.2      | Klonale Verwandtschaft in Relation zum<br>Auftreten des Markers              | 46          |
| 4.3.3      | Aufbau des Markers                                                           | 47          |
| 4.3.4      | Identifizierung der Varianz des<br>markerpositiven Sequenzabschnitts         | 48          |
| 4.4        | Charakterisierung des Sequenztyps 2 von<br><i>S. epidermidis</i>             | 50          |
| 4.4.1      | Identifizierung der Markersequenzen                                          | 50          |
| 4.4.2      | Klonale Verwandtschaft in Relation zum<br>Auftreten der Marker               | 56          |
| 4.4.3      | Aufbau des ST2-Markers 1                                                     | 58          |
| 4.4.4      | Identifizierung der Varianz des ST2-Marker-1-<br>positiven Sequenzabschnitts | 60          |
| 4.4.5      | Aufbau des ST2-Markers 2                                                     | 64          |
| 4.4.6      | Identifizierung der Varianz des ST2-Marker-2-<br>positiven Sequenzabschnitts | 65          |
| 4.4.7      | Aufbau des ST2-Markers 3                                                     | 68          |
| 4.4.8      | Identifizierung der Varianz des ST2-Marker-3-<br>positiven Sequenzabschnitts | 70          |
| <b>5.</b>  | <b>Diskussion</b>                                                            | <b>73</b>   |
| 5.1        | Voraussetzungen für diese Arbeit                                             | 73          |
| 5.2        | Charakterisierung des ST5-Markers                                            | 74          |
| 5.3        | Charakterisierung des ST59-Markers                                           | 79          |
| 5.4        | Charakterisierung der ST2-Marker                                             | 82          |
| 5.4.1      | ST2-Marker 1                                                                 | 87          |
| 5.4.2      | ST2-Marker 2                                                                 | 93          |
| 5.4.3      | ST2-Marker 3                                                                 | 98          |
| 5.5        | Ausblick                                                                     | 101         |
| <b>6.</b>  | <b>Zusammenfassung</b>                                                       | <b>102</b>  |
| <b>7.</b>  | <b>Literaturverzeichnis</b>                                                  | <b>X</b>    |
| <b>8.</b>  | <b>Anhang</b>                                                                | <b>XX</b>   |
| <b>9.</b>  | <b>Danksagung</b>                                                            | <b>XXX</b>  |
| <b>10.</b> | <b>Lebenslauf</b>                                                            | <b>XXXI</b> |

## Tabellenverzeichnis

|                                                                                                                         |    |
|-------------------------------------------------------------------------------------------------------------------------|----|
| Tabelle 1.1: Faktoren der spezifischen, primären Adhäsion von <i>S. epidermidis</i>                                     | 8  |
| Tabelle 1.2 Durch Studien identifizierte, häufige SCCmec-Elemente in pathogenen <i>S. epidermidis</i>                   | 12 |
| Tabelle 1.3 Die <i>housekeeping genes</i> der MLST-Schemata unterschiedlicher Arbeitsgruppen                            | 14 |
| Tabelle 1.4 Übersicht über die MLST-Sequenztypen, die in Blutkulturen im Rahmen anderer Studien identifiziert wurden    | 15 |
| Tabelle 2.1: Verwendete Einwegprodukte sowie ihre Hersteller                                                            | 18 |
| Tabelle 2.2: Verwendete Chemikalien und Nährmedien sowie ihre Hersteller                                                | 18 |
| Tabelle 2.3: Verwendete Geräte sowie ihre Hersteller                                                                    | 19 |
| Tabelle 2.4: Verwendete Kits sowie ihre Hersteller                                                                      | 19 |
| Tabelle 2.5: Verwendete Programme und Datenbanken sowie ihre Provider                                                   | 20 |
| Tabelle 4.1: Anzahl der ST5-markerpositiven Stämme in Relation zur Gesamtzahl der Stämme der jeweiligen Kategorie       | 37 |
| Tabelle 4.2: Nachweis des ST5-Markers innerhalb der Gruppe aller verwendeten Isolate                                    | 38 |
| Tabelle 4.3: Nachweis des ST5-Markers in pathogenen und Krankenhausumwelt-Isolaten der Stammsammlung der Arbeitsgruppe  | 38 |
| Tabelle 4.4: Anzahl der ST59-markerpositiven Stämme in Relation zur Gesamtzahl der Stämme der jeweiligen Kategorie      | 45 |
| Tabelle 4.5: Nachweis des ST59-Markers innerhalb der Gruppe aller verwendeten Isolate                                   | 45 |
| Tabelle 4.6: Nachweis des ST59-Markers in pathogenen und Krankenhausumwelt-Isolaten der Stammsammlung der Arbeitsgruppe | 45 |
| Tabelle 4.7: Anzahl der ST2-markerpositiven Stämme in Relation zur Gesamtzahl der Stämme der jeweiligen Kategorie       | 52 |
| Tabelle 4.8: Nachweis des ST2-Markers 1 innerhalb der Gruppe aller verwendeten Isolate                                  | 54 |

|                                                                                                                           |        |
|---------------------------------------------------------------------------------------------------------------------------|--------|
| Tabelle 4.9: Nachweis des ST2-Markers 1 in pathogenen und Krankenhausumwelt-Isolaten der Stammsammlung der Arbeitsgruppe  | 54     |
| Tabelle 4.10: Nachweis des ST2-Markers 2 innerhalb der Gruppe aller verwendeten Isolate                                   | 55     |
| Tabelle 4.11: Nachweis des ST2-Markers 2 in pathogenen und Krankenhausumwelt-Isolaten der Stammsammlung der Arbeitsgruppe | 55     |
| Tabelle 4.12: Nachweis des ST2-Markers 3 innerhalb der Gruppe aller verwendeten Isolate                                   | 55     |
| Tabelle 4.13: Nachweis des ST2-Markers 3 in pathogenen und Krankenhausumwelt-Isolaten der Stammsammlung der Arbeitsgruppe | 55     |
| Tabelle 8.1: Verwendete Isolate, ihr MLST-Sequenztyp, Markernachweis und Herkunft                                         | XX     |
| Tabelle 8.2: Verwendete Isolate, ihre Herkunft, Isolatqualität und ihr N50-Wert                                           | XXIV   |
| Tabelle 8.3: ORFs der ST5-Markersuche und ihre Länge                                                                      | XXVII  |
| Tabelle 8.4: ORFs der ST59-Markersuche und ihre Länge                                                                     | XXVII  |
| Tabelle 8.5: Erste ORF-Abfolge der ST2-Markersuche und ihre Länge                                                         | XXVII  |
| Tabelle 8.6: Zweite ORF-Abfolge der ST2-Markersuche und ihre Länge                                                        | XXVIII |
| Tabelle 8.7: Dritte ORF-Abfolge der ST2-Markersuche und ihre Länge                                                        | XXVIII |
| Tabelle 8.8: Auftreten des Block 7 in Non-ST2-Stämmen in Relation zum ST2-Marker 1                                        | XXIX   |

## Abbildungsverzeichnis

|                                                                                                                                                   |    |
|---------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Abbildung 4.1 <i>Minimum spanning tree</i> der verwendeten ST5-Isolate                                                                            | 40 |
| Abbildung 4.2 Genkarte des ST5-Markers                                                                                                            | 41 |
| Abbildung 4.3 Das Ergebnis des <i>alignment</i> eines Ausschnitts der ST5-marker-tragenden Sequenz sowie sieben markernegativer Referenzsequenzen | 42 |
| Abbildung 4.4 <i>Minimum spanning tree</i> der verwendeten ST59-Isolate                                                                           | 47 |
| Abbildung 4.5 Das Ergebnis des <i>alignment</i> eines Ausschnitts der ST59-marker-tragenden Sequenz sowie acht markernegativer Referenzsequenzen  | 49 |
| Abbildung 4.6 <i>Minimum spanning tree</i> der verwendeten ST2-Isolate                                                                            | 56 |
| Abbildung 4.7 Genkarte des ST2-Markers 1                                                                                                          | 59 |
| Abbildung 4.8 Das Ergebnis des <i>alignment</i> der ST2-Marker-1-positiven Sequenz sowie 14 markernegativer Referenzsequenzen                     | 61 |
| Abbildung 4.9 Genkarte des ST2-Markers 2                                                                                                          | 64 |
| Abbildung 4.10 Das Ergebnis des <i>alignment</i> der ST2-Marker-2-positiven Sequenz sowie sieben markernegativer Referenzsequenzen                | 66 |
| Abbildung 4.11 Genkarte des ST2-Markers 3                                                                                                         | 69 |
| Abbildung 4.12 Das Ergebnis des <i>alignment</i> der ST2-Marker-3-positiven Sequenz sowie sechs markernegativer Referenzsequenzen                 | 71 |

## Abkürzungsverzeichnis

|             |                                                                                                                       |
|-------------|-----------------------------------------------------------------------------------------------------------------------|
| Aae / AtlE  | Autolysin                                                                                                             |
| Aap         | <i>accumulation-associated protein</i>                                                                                |
| <i>agr</i>  | <i>accessory gene regulator</i>                                                                                       |
| <i>arcC</i> | Gen der <i>carbamate kinase</i>                                                                                       |
| <i>aroE</i> | Gen der <i>shikimate 5-dehydrogenase</i>                                                                              |
| ATP         | Adenosintriphosphat                                                                                                   |
| Bap         | <i>biofilm associated protein</i>                                                                                     |
| Bhp         | Bap homologes Protein                                                                                                 |
| bp          | <i>base pair</i>                                                                                                      |
| C           | Cytosin                                                                                                               |
| °C          | Grad Celsius                                                                                                          |
| CCR         | <i>cassette chromosome recombinase</i>                                                                                |
| CDP         | Cytidindiphosphat                                                                                                     |
| cgMLST      | <i>core genome Multi Locus Sequence Typing</i>                                                                        |
| CGPTase     | CDP-Glycerol: Poly(glycerophosphat)<br>Glycerophosphotransferase                                                      |
| <i>dfp</i>  | Gen der <i>fused 4'-phosphopantothenoyleysteine decarboxylase/phosphopantothenoyleysteine synthetase, FMN-binding</i> |
| DNA         | <i>desoxyribonucleic acid</i> = Desoxyribonukleinsäure                                                                |
| dNTP        | Deoxynucleosid-Triphosphate                                                                                           |
| DR          | <i>direct repeats</i>                                                                                                 |
| dsDNA       | doppelsträngige DNA                                                                                                   |
| Embp        | <i>extracellular matrix-binding protein</i> = Extrazelluläre-Matrix-bindendes Protein                                 |
| Fbe         | Fibrinogen bindendes Protein, auch SdrG genannt                                                                       |
| G           | Guanin                                                                                                                |
| g           | Schwerebeschleunigung der Erde                                                                                        |
| GehD        | extrazelluläres Lipase-Protein                                                                                        |
| <i>glpF</i> | Gen einer <i>glycerol kinase</i>                                                                                      |
| <i>glpK</i> | Gen einer <i>glycerol kinase</i>                                                                                      |
| <i>gmK</i>  | Gen der <i>guanylate kinase</i>                                                                                       |
| <i>gtr</i>  | Gen des <i>ABC transporter</i>                                                                                        |

|                  |                                                                          |
|------------------|--------------------------------------------------------------------------|
| HCl              | Salzsäure                                                                |
| Hg <sup>0</sup>  | elementares Quecksilber                                                  |
| Hg <sup>2+</sup> | Quecksilberkationen                                                      |
| <i>hsp60</i>     | Gen des <i>heat-shock protein 60</i>                                     |
| <i>ica</i>       | <i>intercellular adhesion</i>                                            |
| IR               | <i>inverted repeats</i>                                                  |
| IS               | Insertionssequenz                                                        |
| ISS              | <i>integration site sequence</i>                                         |
| J                | <i>junkyard, joining</i>                                                 |
| kDa              | Kilo-Dalton                                                              |
| KNS              | Koagulase-negative Staphylokokken                                        |
| LTA              | Lipoteichonsäure                                                         |
| <i>mecA</i>      | Gen für das Penicillin-bindende Protein PBP2a                            |
| MLST             | <i>Multi Locus Sequence Typing</i>                                       |
| MRSE             | methicillinresistenter <i>Staphylococcus epidermidis</i> -Stamm          |
| MRSA             | methicillinresistenter <i>Staphylococcus aureus</i> -Stamm               |
| MSCRAMM          | <i>microbial surface component recognizing adhesive matrix molecules</i> |
| MSSA             | methicillinsensibler <i>Staphylococcus aureus</i> -Stamm                 |
| MST              | <i>minimum spanning tree</i>                                             |
| <i>mutS</i>      | Gen des <i>DNA mismatch repair protein</i>                               |
| µm               | Mikrometer                                                               |
| NCBI             | <i>National Center for Biotechnology Information</i>                     |
| NGS              | <i>next generation sequencing</i>                                        |
| nm               | Nanometer                                                                |
| N50              | gewichteter Median der Sequenzlängen eines <i>de-novo-assembly</i>       |
| ORF              | <i>open reading frame</i>                                                |
| <i>orfX</i>      | konservierter <i>open reading frame</i> unbekannter Funktion             |
| <i>oriC</i>      | Ursprung der Replikation                                                 |
| PBP              | Penicillin-bindendes Protein                                             |
| PCR              | <i>polymerase chain reaction</i>                                         |
| PFGE             | Pulsfeldgelelektrophorese                                                |

|             |                                                     |
|-------------|-----------------------------------------------------|
| PIA         | interzelluläres Polysaccharid-Adhäsion              |
| PSM         | <i>phenol soluble modulins</i>                      |
| <i>pta</i>  | Gen der <i>phosphate acetyltransferase</i>          |
| <i>pyrR</i> | Gen des <i>pyrimidine operon regulatory protein</i> |
| S.          | <i>Staphylococcus</i>                               |
| SCCmec      | <i>Staphylococcal Chromosome Cassette mec</i>       |
| Sdr         | <i>serine-aspartate dipeptid-repeats protein</i>    |
| SDS         | Natriumdodecylsulfat                                |
| SNP         | <i>small-nuclear polymorphism</i>                   |
| <i>spp</i>  | Subspezies                                          |
| SSSS        | <i>staphylococcal scalded skin syndrome</i>         |
| ST          | Sequenztyp                                          |
| <i>tpiA</i> | Gen der <i>triosephosphate isomerase</i>            |
| Tris        | Tris(hydroxymethyl)-aminomethan                     |
| TSB         | <i>tryptic soy broth</i>                            |
| TSS         | <i>toxic shock syndrome</i>                         |
| Tween       | Polyethylenglykolsorbitanmonolaurat                 |
| T4SS        | <i>type four secretion system</i>                   |
| UKSH        | Universitätsklinikum Schleswig-Holstein             |
| upm         | Umdrehung pro Minute                                |
| UV          | Ultraviolett                                        |
| WTA         | Zellwandteichonsäure                                |
| <i>xpt</i>  | Gen der <i>xanthine phosphoribosyltransferase</i>   |
| <i>yqiL</i> | Gen der <i>acetyl coenzyme A acetyltransferase</i>  |

## 1. Einleitung

### 1.1 Der Genus *Staphylococcus*

Staphylokokken gehören zu den Vertretern der grampositiven Kokken. Ein einzelnes Bakterium besitzt einen Durchmesser von circa 1 bis 1,5 µm (Sperber and Tatini, 1975). Sie sind nicht sporenbildend und können sowohl unter aeroben als auch unter anaeroben Bedingungen wachsen. Aufgrund ihres Zellteilungsmechanismus sowie ihrer Unbeweglichkeit bilden sich vermehrende Staphylokokken traubenförmige Gruppen. Bevorzugt siedeln sich Staphylokokken auf der Haut sowie der Nasenschleimhaut von Warmblütern an. Als grampositive Kokken mit einem geringen GC-Gehalt (33%) (Kuroda et al., 2001), zählen sie zu der Abteilung der Firmicuten. Innerhalb dieser Abteilung sind sie Teil der Klasse *Bacilli*. Sie gehören zusammen mit *Jeotgalicoccus*, *Macrococcus*, *Nosocomiicoccus* und *Salinicoccus* der Familie der *Staphylococcaceae* an (Becker et al., 2014). Staphylokokken wachsen bevorzugt auf Columbia oder Soja-Tripton-Agar mit 5% fibrinfreiem Schafsblut (oder Pferdeblut) unter aeroben Bedingungen (Becker et al., 2014; Ljungh and Wadström, 1995). Dort bilden sie in Abhängigkeit von dem artspezifischen Pigment weiße oder goldgelbe Kolonien aus. Die Art *S. aureus* zeigt auf Blutagar um die Kolonien jeweils eine blasse Zone der  $\beta$ -Hämolyse. Die  $\beta$ -Hämolyse stellt eine vollständige Zellyse der Erythrozyten dar (Kuroda et al., 2001; Mims, et al., 2006). Staphylokokken, insbesondere solche mit Antibiotikaresistenzen, besitzen häufig zusätzlich intermediäre Resistenzen gegen Antiseptika. Dies begünstigt ihr Überleben in der Krankenhaus-Umwelt (Taheri et al., 2016).

Staphylokokken lassen sich von Streptokokken anhand eines Katalase-Tests differenzieren. Katalase ist ein Enzym, das die Reaktion von, durch Stoffwechselprozesse in Staphylokokken anfallendem, Wasserstoffperoxid in Wasser und elementaren Sauerstoff katalysiert. Im Gegensatz zu Streptokokken besitzen Staphylokokken in der Regel dieses Enzym und sind daher Katalase-positiv (Foster, 1996; Park et al., 2008).

Die Staphylokokken-Arten werden im klinisch-diagnostischen Kontext anhand ihrer biochemischen Eigenschaften unterschieden. Dabei ist das bekannteste Unterscheidungskriterium der Koagulasetest. Das Prinzip des Koagulasetests beruht auf der Plasmaagglutination. Diese wird ausgelöst durch die von *S. aureus* sezernierten Proteine Staphylokoagulase sowie dem von-Willebrand-Faktor-Bindungsprotein. Diese beiden Proteine binden Prothrombin und bilden dadurch den Komplex des Staphylothrombins. Das Staphylothrombin wiederum katalysiert die Reaktion des Plasma-löslichen Fibrinogens in unlösliches Fibrin. Durch den so entstehenden Fibrinwall ist *S. aureus in vivo* vor äußeren Einflüssen geschützt beispielsweise dem Immunsystem des Wirtes oder Antibiotika. Im labormedizinischen Kontext wird diese Eigenschaft *in vitro* für den Koagulasetest verwendet, um eine Differenzierung der *S. aureus* gegenüber den klinisch wichtigen Staphylokokken, die das Enzym Plasmakoagulase nicht sezernieren, zu ermöglichen. Die Durchführung eines Koagulasetests dauert bis zu 24 h (Sperber and Tatini, 1975; Vanassche et al., 2013). Aufgrund des geringeren Zeitaufwands ist im Rahmen der Differenzierung zwischen *S. aureus* und koagulase-negativen Staphylokokken (KNS) der weniger zeitaufwendige Test des *clumping factor* verbreitet. Der *clumping factor* fungiert bei *S. aureus* als Fibrinogen-Rezeptor, der gebunden an der bakteriellen Zelloberfläche vorliegt und ebenfalls durch Plasmaagglutination nachgewiesen werden kann (Personne et al., 1997).

Von den mehr als 16 Staphylokokkenarten besitzen *S. aureus*, *S. epidermidis* sowie *S. saprophyticus* die größte klinische Relevanz (Mims, et al., 2006). *S. aureus* ist der bekannteste der drei genannten Keime. Bevorzugt siedeln diese Bakterien in der Nasenschleimhaut. 20% der Bevölkerung sind von *S. aureus* als Kommensal persistent nasal besiedelt und weitere 30% sind intermittierende Träger dieses Bakteriums. Nasale Träger von *S. aureus* haben ein erhöhtes Risiko eine Infektion mit diesem Bakterium zu erleiden, weshalb die Eradikation von infektionsgefährdeten Populationen, zum Beispiel chirurgischen Patienten, empfohlen wird (Wertheim et al., 2005). *S. aureus* verfügt über eine große Varianz an Pathogenitätsfaktoren. Zu diesen Faktoren von *S. aureus* zählen, neben der schon beschriebenen Koagulase und dem Fibrinogen-bindenden *clumping factor*, Fibronectin-bindende Proteine, Protein A, Mukopeptide

und eine Kapselstruktur. Zusätzlich produziert *S. aureus* diverse Toxine, die den Wirtsorganismus beeinflussen (Hof et al., 2014). Ein bekannter Vertreter ist das Panton-Valentine Leukozidin, das durch Porenbildung die polymorphnuklearen Neutrophilen des Immunsystems schädigt. Im Zusammenhang mit diesem Leukozidin wurde bereits die nekrotisierende Pneumonie beschrieben (Labandeira-Rey et al., 2007). Weitere Toxine sind membranschädigende Stoffe wie die Hämolysine, Lebensmittelvergiftung auslösende Enterotoxine, epidermolytische Toxine, toxisches Schocksyndrom-Toxin und Staphylokinase (Hof et al., 2014). Aufgrund dieser Vielzahl an unterschiedlichen Pathogenitätsfaktoren können *S. aureus* eine Bandbreite unterschiedlicher Erkrankungen auslösen. Zu den typische Erkrankungen zählen Abszesse im Rahmen der Impetigo follicularis oder Mastitis puerperalis, Furunkel, Karbunkel, Osteomyelitis, Biofilm-assoziierte Katheterinfektionen, *Staphylococcal scalded Skin Syndrome* (SSSS), *toxic shock syndrome* (TSS), Sepsis und Enterotoxin-assoziierte Enteritis oder Enterokolitis.

## 1.2 Das Mikrobiom der Haut

Der Begriff "Mikrobiom" ist definiert als das gesamte genetische Material einer mikrobiellen Gemeinschaft (Grice and Segre, 2011). Zwischen dem Wirt und den Mikrobiota der Haut besteht ein symbiontisches Verhältnis. Die Mikroorganismen schützen den Wirt vor einer Fehlbesiedlung mit pathogenen Erregern und spielen eine wichtige Rolle in der Ausbildung der Immuntoleranz. Es wird zwischen residenten und transienten Mikroorganismen unterschieden. Die residente, kommensale Flora regeneriert sich selbstständig nach einer Störung und kann dauerhaft auf der Haut nachgewiesen werden. Die transiente Flora hingegen persistiert für einen kürzeren Zeitraum, bevor sie nicht mehr nachgewiesen werden kann (Dréno et al., 2016). Die Haut bietet ein abwechslungsreiches Habitat für die Mikroorganismen aufgrund ihres dreischichtigen Aufbaus aus Epidermis, Dermis und Subkutis mit vielen unterschiedlichen Kompartimenten. Die dermalen Hauptkompartimente können in feucht, trocken und talgig unterschieden werden, die jeweils eine Nische für eine angepasste Gruppe von Mik-

roorganismen bilden. Beispielsweise Staphylokokken siedeln bevorzugt auf feuchter Haut, wie der Ellenbeuge sowie auf talgiger Haut. Die lipophilen Propionibakterien hingegen präferieren als Habitat talgige oder trockene Kompartimente (Grice et al., 2009). Die Hautanhangsgebilde und Hautfalten besitzen ebenfalls ein, für sie typisches, mikrobielles Profil (Grice and Segre, 2011). Bestimmten Organismen konnte bereits auch die Manipulation der Hautflora nachgewiesen werden. *Propionibacterium acnes* kann eine Reduktion des Wachstums von methicillinresistenten *S. aureus*-Stämmen bewirken. Ähnliche sowie weitere Effekte wurden für *S. epidermidis* beschrieben. Diese Staphylokokkenart kann durch die Induktion von  $\beta$ -Defensinen, die zu den menschlichen, antimikrobiellen Peptiden zählen, das Immunsystem zur Abwehr von *S. aureus* anregen. Zusätzlich sind *S. epidermidis* in der Lage, antivirale Mastzellen zu aktivieren, eine unkontrollierte Immunreaktion im Rahmen der Wundheilung zu unterdrücken sowie die Reifung von T-Zellen zu stimulieren (Dréno et al., 2016).

Die Haut des ungeborenen Kindes ist bis zur Geburt steril, ihre Besiedlung beginnt unmittelbar mit dem Geburtsvorgang im Fall der vaginalen Entbindung beziehungsweise kurz nach der Geburt per Schnittentbindung. Die Zusammensetzung der Besiedlung mit Mikroorganismen ändert sich im weiteren Verlauf des Lebens durch unterschiedliche Faktoren wie beispielsweise dem Hormonstatus oder der Schweißbildung. (Grice et al., 2009).

### 1.3 Koagulase-negative Staphylokokken

Die Gruppe der klinisch relevanten, Koagulase-negativen Staphylokokken (KNS) wird aufgeteilt in die *S. epidermidis*-ähnliche Gruppe, in die *S. saprophyticus*-Gruppe und in *S. lugdunensis*. Die diagnostische Differenzierung zwischen *S. saprophyticus* und den anderen humanassoziierten KNS wird durch einen Plättchentest auf Novobiocin-Empfindlichkeit ermöglicht. Die *S. saprophyticus*-Gruppe ist resistent gegen Novobiocin. Zu dieser Gruppe zählen unter anderem die klinisch relevanten *S. saprophyticus*, *S. xylosus* und *S. cohnii*. Im Ge-

gensatz dazu ist die Mehrzahl der weiteren humanpathogenen KNS Novobiocin-sensibel (Hébert et al., 1988). Am häufigsten auf der menschlichen Haut nachgewiesen wurden die KNS-Arten *S. epidermidis* und *S. haemolyticus*. Zu den weiteren KNS des dermalen Mikrobioms zählen unter anderem *S. saprophyticus*, *S. hominis*, *S. cohnii*, *S. warneri*, *S. xylosus* und *S. simulans*. Hierbei werden für einige der unterschiedlichen KNS jeweils typische Habitat identifiziert. Einer der beiden häufigsten KNS, *S. haemolyticus*, lässt sich häufig von der Haut der Beine und der Arme isolieren. *S. epidermidis* als das zweite, häufigste KNS konnte insbesondere von humiden Regionen wie den Nasenvorhöfen oder den Achselhöhlen isoliert werden (Kloos and Musselwhite, 1975). *S. auricularis* siedelt bevorzugt im äußeren Gehörgang. *S. capitis* dominiert auf der behaarten Kopfhaut und den Armen (Becker et al., 2014).

#### **1.4 Die Pathogenität der Koagulase-negativen Staphylokokken**

KNS zählen zu den typischen Hautkommensalen. Trotz diesem Umstand wurde den KNS ebenfalls ein großes Pathogenitätspotential nachgewiesen. *S. saprophyticus* beispielsweise ist aufgrund seines spezifischen Adhäsins der häufigste, grampositive Erreger des Harnwegsinfekts in jungen Frauen (Kline et al., 2010). Die gesamte Gruppe der KNS konnte als häufigster Erreger aus Blutkulturen isoliert werden (Karlowsky et al., 2004). Zu den häufigsten, pathogenen Vertretern dieser Gruppe zählen *S. epidermidis*, gefolgt von *S. haemolyticus* und *S. hominis* (Spanu et al., 2003). *S. epidermidis* stellt auch den häufigsten Endokarditis-Erreger der Gruppe der KNS dar, gefolgt von *S. lugdunensis* (Petti et al., 2008). In Abhängigkeit von der Erkrankung gibt es demnach Unterschiede in der epidemiologischen Verteilung der pathogenen KNS. Die Infektionen verlaufen je nach Art der KNS unterschiedlich schwer. Für *S. lugdunensis*-verursachte Endokarditiden wurden im Vergleich zu den *S. epidermidis*-Endokarditiden schwerere Krankheitsverläufe beobachtet (Anguera et al., 2005). Es konnte die Verwendung von medizinischen Fremdkörpern als begünstigender Faktor von KNS-Infektionen identifiziert werden. Zu den betroffenen medizinischen Fremdkörper zählen intravaskuläre sowie intrathekale Kathetersysteme,

Schrittmacherelektroden, Harnwegskatheter, prothetischer Gelenkersatz und weitere dauerhafte oder transiente Implantate aus Metall beziehungsweise Kunststoff. *S. epidermidis* gilt in diesem Zusammenhang als häufigster Erreger (Ortega-Peña et al., 2015; Ziebuhr et al., 2006). Für die KNS bedingte Fremdkörperinfektion konnten prädisponierenden Faktoren in Form von onkologischen Erkrankungen, Chemotherapie sowie Immunschwäche nachgewiesen werden (Pektaş et al., 2015). Weitere Risikofaktoren sind Thrombozytopenie, für Neonaten ein Gewicht niedriger als 1000 g, Leukozytopenie, Malnutrition, Einführen eines Katheters kurz vor Beginn einer Chemotherapie, Dauer der Katheterisierung, Ort des Katheters, mangelnde hygienische Maßnahmen bei Kathetereinlage sowie die mangelnde Pflege der Kathetereinlage (Gominet et al., 2017; Mahieu et al., 2001; Pektaş et al., 2015; Viana Taveira et al., 2017). Im Fall von einliegenden Kathetern galt die Manipulation durch medizinisches Personal als Risiko für eine KNS-Infektion, z.B. die Gabe von Lipidemulsionen. Die Manipulationen können Keimverschleppung von anderen Körperregionen des Patienten oder von anderen, besiedelten Patienten zur Folge haben (Mahieu et al., 2001).

Ein Virulenzfaktor für Fremdkörper-assoziierte Infektionen stellt die Fähigkeit der KNS zur Biofilmbildung dar. Der Biofilm entzieht sie dem Zugriff durch das angeborene Immunsystem. Ein weiterer Effekt ist ein Biofilm-bedingter Schutz vor Antibiotika (Ziebuhr et al., 2006). Die Ursache für die erschwerte, antibiotische Therapie von katheterbedingten Infektionen durch biofilmbildende KNS sind die persistierenden Bakterien in den tiefen Biofilmschichten. Die persistierenden Zellen können dem antibiotischen Effekt mittels ihres verlangsamten Stoffwechsel ausweichen. Die übrigen planktonisch-vorliegenden Bakterien werden durch das Immunsystem nach Antibiotikagabe beseitigt. Da die persistierenden Bakterien allerdings im Biofilm vor den Zellen des Immunsystems geschützt sind, bilden sie ein Reservoir für ein erneutes Aufflammen der Infektion (Lewis, 2007). Zu den weiteren Schutzmechanismen des Biofilms vor Antibiotika zählen eine verminderte Penetration in den Biofilm durch die erhöhte Bakteriendichte sowie die erhöhte Masse an extrazellulärer Matrix, dem Ausbilden von Effluxsystemen und der Aktivierung von stressbedingten Regulons (Gominet et al., 2017). Im Fall des Vorliegens eines infizierten Fremdkörpers,

sollte dieser entfernt und eine systemische Antibiotikatherapie begonnen werden (Mermel et al., 2009). Für die antibiotische Therapie hat sich eine Kombinationstherapie bewährt, da der im klinischen Kontext häufige, Biofilm-bildenden Keim *S. epidermidis* eine Multiresistenz gegenüber diversen Antibiotika besitzt. In Abhängigkeit von der jeweiligen Studie konnten in unterschiedlichem Ausmaß Resistenz gegen Fluorchinolone und Glykopeptide (Raad et al., 1998) beziehungsweise Makrolide, Fluorchinolone, Mupirocin, Rifampicin, Fusidinsäure, Aminoglykoside, Trimethoprim/Sulfamethoxazol sowie Oxacilin (Deplano et al., 2016) beobachtet werden. Aus diesem Grund besitzt die Kombinationstherapie bestehend aus zwei Antibiotika von Biofilm-bildenden *S. epidermidis* eine höhere Effektivität (Saginur et al., 2006).

### **1.5 Der Prozess der Biofilmbildung durch *S. epidermidis***

Der Biofilm stellt einen wichtigen Faktor für die Pathogenität der KNS dar und ist bereits für *S. epidermidis* gut untersucht worden. In einem Biofilm liegen die Bakterien in mehreren Schichten vor, die in eine amorphe Substanz eingebettet sind. Maßgeblicher Bestandteil dieser amorphen Substanz ist das bakterielle, interzelluläre Polysaccharid-Adhäsion (PIA). PIA setzt sich aus Polymeren von N-Acetylglucosaminen mit diversen Seitengruppen zusammen. Durch Gelfiltration konnten zwei unterschiedliche Bestandteile des PIA identifiziert werden. Der Hauptanteil (mindestens 80%) bildet Polysaccharid I als ein unverzweigtes Homoglykan mit einer Länge von 130 Einheiten. Es besteht aus bis zu 85%  $\beta$ -1,6 verknüpften N-Acetylglucosamin-Einheiten und mindestens 15% deacetyliertem Glucosaminen. Dies bedingt eine positive Ladung des Moleküls. Bis zu 20% von PIA bildet Polysaccharid II, das strukturell vergleichbar ist mit dem Polysaccharid I. Es besteht ausschließlich aus Glucosaminen, die ebenfalls in Teilen acetyliert vorliegen. Der Anteil von nicht-acetylierten Glucosaminen ist im Polysaccharid II kleiner. Polysaccharid II besitzt zusätzlich einen geringen Anteil an Phosphaten und durch Esterbindung verknüpfte Succinate. Dadurch erhält das Molekül einen anionischen Charakter (Mack et al., 1996). Die Kombination aus PIA, Zellwandteichonsäure und weiteren Proteinen bilden zusammen die amorphe Substanz, die die Grundlage des Biofilms darstellt (Otto, 2008).

Die Biofilmbildung wurde in drei Schritte unterteilt. Im ersten Schritt der primären Adhäsion binden die *S. epidermidis* an eine unbelebte Oberfläche. Zu diesen unbelebten Oberflächen zählen sowohl synthetische Polymere von beispielsweise Kathetern als auch humane Matrixproteine. Es wird die unspezifische von der spezifischen Adhäsion unterschieden. Unspezifische Adhäsion erfolgt durch hydrophobe und elektrostatische Wechselwirkungen. Spezifische Adhäsion ist durch die Oberflächenmoleküle der *S. epidermidis* möglich (Tab 1.1) (Kloos and Bannerman, 1994). Über die sogenannten MSCRAMMs (*microbial surface components recognizing adhesive matrix molecules*) sind *S. epidermidis* in der Lage menschliche Matrixproteine zu binden, zum Beispiel Fibronektin. Die kovalente Bindung dieser MSCRAMMs an das bakterielle Peptidoglykan wird katalysiert durch die Enzymgruppe der Sortasen, insbesondere durch die Sortase A (Otto, 2008).

Tabelle 1.1: Faktoren der spezifischen, primären Adhäsion von *S. epidermidis*

| Adhäsionsmolekül                                      | Bindungspartner                                            | Referenz                                                                |
|-------------------------------------------------------|------------------------------------------------------------|-------------------------------------------------------------------------|
| Aae (Autolysin)                                       | Vitronectin, Fibrinogen<br>Fibronektin                     | Heilmann et al., 2003                                                   |
| AtlE (Autolysin)                                      | Vitronectin, Polystyrol                                    | Heilmann et al., 1997                                                   |
| Bhp (Bap homologes Protein)                           | Polystyrol                                                 | Cucarella et al., 2001;<br>Gill et al., 2005                            |
| SdrF                                                  | Kollagen, Keratin,<br>Polytetrafluorethylen,<br>Polystyrol | Arrecubieta et al., 2007;<br>Toba et al., 2013;<br>Trivedi et al., 2017 |
| SdrH                                                  |                                                            | Bowden et al., 2005                                                     |
| Teichonsäure                                          | Fibronektin,                                               | Hussain et al., 2001                                                    |
| Embp (Extrazelluläre-Matrix-bindendes Protein)        | Fibronektin,<br>Plasminogen, Heparin,<br>Hyaluronsäure     | Williams et al., 2002                                                   |
| Fbe (Fibrinogen bindendes Protein, auch SdrG genannt) | Fibrinogen                                                 | Hartford et al., 2001;<br>Nilsson et al., 1998                          |
| GehD (extrazelluläre Lipase)                          | Kollagen                                                   | Bowden et al., 2002                                                     |

Im zweiten Schritt der Akkumulation kommt es zu einer mehrlagigen Anordnung der bakteriellen Zellen. Nur ein Teil der Bakterien steht in direktem Kontakt mit der besiedelten Oberfläche. Die durch PIA vermittelten, interzellulären Verknüpfungen bilden die Grundlage der weiteren Akkumulation von Bakterien

ohne Oberflächen-Kontakt. Die Synthese von PIA wird durch die Proteine der Strukturgene des Operon *icaADBC* ermöglicht (Gerke et al., 1998). Die Proteinprodukte von *icaA* und *icaD* besitzen N-Acetylglucosaminyltransferase-Aktivität. Das Strukturgen *icaB* kodiert für eine PIA-Deacetylase. Dem Gen *icaC* konnte ein vermuteter PIA Exporter zugeordnet werden. Das regulatorische Gen stellt *icaR* dar (Otto, 2008). Das Proteinprodukt IcaR wirkt als Repressor der Transkription des *ica*-Operons (Conlon et al., 2002). Neben diesem PIA-abhängigen Mechanismus konnte auch eine PIA-unabhängige Biofilmbildung nachgewiesen werden. Für die Synthese eines PIA-unabhängigen Biofilms spielt das 220 kDa große *accumulation-associated protein* (Aap) eine entscheidende Rolle. Es substituiert hierbei die Funktion von PIA im Zuge der interzellulären Adhäsion. Aap muss allerdings zunächst durch Spaltung mittels einer Protease in seine 140 kDa schwere, aktive Isoform überführt werden (Rohde et al., 2005). Nach Synthese eines Biofilms können einzelne Bakterien sich ablösen und disseminiert streuen. Dieser Vorgang beruht auf mechanischen Kräften wie dem Blutfluss oder auf Faktoren, die befallene Matrixproteine abbauen (Otto, 2008). Neben den mechanischen Kräften wurde ebenfalls eine bakteriell-kontrollierte Biofilmablösung beschrieben. Der Prozess der Ablösung wird in *S. epidermidis* vermutlich gesteuert durch den Regulator *agr* (*accessory gene regulator*) des *quorum sensing system*, allerdings ist dieser Prozess weniger gut erforscht (Vuong et al., 2004). Zwei potentielle Mechanismen wurden bereits formuliert: Einerseits wird eine enzymatische Ablösung bedingt durch Proteasen und Nukleasen vermutet, allerdings ist die Studienlage dahingehend noch nicht eindeutig. Ein anderer möglicher Mechanismus stellt die Ablösung durch grenzflächenaktive, Surfactant-ähnliche Peptide dar, den sogenannten *phenol soluble modulins* (PSM). Die Synthese von PSM unterliegt der Kontrolle durch *agr* (Otto, 2013). Insbesondere das PSM $\beta$  konnte als zentrales Peptid dieses Mechanismus identifiziert werden (Wang et al., 2011). Durch die kontrollierte Ablösung sind die Biofilm-bildende Bakterien in der Lage in andere Körperkompartimenten zu streuen und auf diesem Weg neue Foki zu erschließen (Otto, 2013).

## 1.6 Die Differenzierung von pathogenen und kommensalen *S. epidermidis*

Im labormedizinischen Kontext stellt *S. epidermidis* eine Herausforderung dar: Wurde das Bakterium in Blutkulturen nachgewiesen, handelt es sich entweder um einen pathogenen Stamm oder um eine Kontamination durch ein Kommensal (Tokars, 2004). Im Rahmen der Hospitalisierung (Widerström et al., 2016) sowie von antibiotischer Prophylaxe (Archer, 1991) wurde eine Veränderung der Population der *S. epidermidis* beschrieben. Unter den genannten Bedingungen besitzen die Biofilm bildenden, methicillinresistenten Stämme einen Überlebensvorteil. Folglich sind die Unterschiede zwischen diesen selektierten Kommensalen und pathogenen Stämmen von Interesse. In beiden Gruppen wurden mehrere Genloci der primären Adhäsion als potentielle Differenzierungsmerkmale untersucht. Es konnte kein Unterschied in der Verteilung der Gene *aap*, *bhp*, *fbe*, *embp* und *atlE* zwischen kommensalen und infektiösen Stämmen festgestellt werden, so dass diese Loci nicht als Marker herangezogen werden können. Zusätzlich wurde das Auftreten des PIA-kodierenden Operons *icaADBC*, der Biofilm-assoziierten Insertionssequenzen IS256 sowie IS257 und des Methicillinresistenz kodierenden *mecA*-Gens zwischen beiden *S. epidermidis*-Gruppen untersucht. Es konnte kein Unterschied zwischen den Gruppen nachgewiesen werden (Bradford et al., 2006; Rohde et al., 2004).

## 1.7 Die SCCmec-Typisierung

Die *mecA*-bedingte Methicillinresistenz beruht auf der Kodierung des Penicillin-bindenden Proteins PBP2a (auch PBP2' oder MecA genannt), dessen Affinität zu Penicillinen deutlich vermindert ist. Aus diesem Grund kann das PBP2a ebenfalls in Anwesenheit von Penicillinen als Transpeptidase das Peptidoglykangerüst der bakteriellen Zellwand synthetisieren (Figueiredo and Ferreira, 2014). Das *mecA*-Gen liegt innerhalb des *mec*-Komplex eingebettet vor. Zu dem *mec*-Komplex zählen zusätzlich die Regulationsgene *mecI* und *mecR*. Dieser Komplex ist auf einem mobilen genetischen Element lokalisiert, dem SCCmec-Element genannten Kassetten-Chromosom. Auf Sequenzebene

wird es eingerahmt von direkten Wiederholungen (*direct repeats*, DR), die die jeweilige Integrationsseitensequenz (*integration site sequence*, ISS) enthalten. Die ISS kann von der Kassetten-Chromosom-Rekombinase (*cassette chromosome recombinase*, CCR) erkannt werden. An das SCC*mec*-Element schließen sich auf Sequenzebene invertierte Wiederholungen (*inverted repeats*, IR) an. Die entscheidenden Genkomplexe des Elements sind *ccr* sowie *mec*. Diese beiden Komplexe sind in variable Regionen eingebettet, den *junkyard* oder auch *joining* genannte Regionen (J). Es liegen drei Regionen vor, J1 bis J3. Sie bestehen aus scheinbar nicht-funktionalen Pseudogenen sowie Fragmenten von Transposons und Integrons. Die Kombination aus der jeweiligen Art von Rekombinase *ccr* (*ccrAB* oder *ccrC*) sowie *mec*-Komplex bestimmt den Typ des SCC*mec*-Elements. Der Subtyp des entsprechenden Elements wird durch den Aufbau der drei J-Regionen festgelegt. Die typische Insertionsstelle des SCC*mec*-Elements ist das 3' Ende des *orfX*. Die Gensequenz dieses *orfX* kodiert ein Protein unbekannter Funktion und ändert sich durch die Insertion des SCC*mec*-Elements trotz eines Basenaustauschs ihres 3' Endes nicht (Hiramatsu et al., 2013). Neben dem *orfX* wurden mindestens sechs weitere ISS für SCC*mec* in der *oriC*-Region beschrieben (Takeuchi et al., 2005). Die Typen von SCC*mec*-Elementen sind phylogenetisch unterschiedlich alt. Die Elemente mit Typ I bis III sind älter, treten gehäuft in Krankenhaus-assoziierten MRSA-Stämmen auf und tragen häufig weitere Antibiotikaresistenzen. Die Typen IV bis V sind jünger, wurden in MRSA-Stämmen außerhalb von Kliniken nachgewiesen und beinhalten häufig nur den *mec*-Komplex. Dafür sind die SCC*mec*-Typen IV- sowie V-tragenden *S. aureus* virulenter und schneller wachsend als solche mit den Typen I-III. Dies könnte an der geringeren Größe des Elements der Typen IV-V liegen, als auch an ihrer geringeren Interaktion mit dem bakteriellen Stoffwechsel aufgrund des Fehlens weiterer Resistenzgene. Statt dem *mec*-Komplex kann das SCC-Element auch andere Resistenzgene enthalten, beispielsweise für Quecksilberresistenz. Diese SCC-Elemente können zusätzlich neben dem SCC*mec* an die gleiche ISS eingefügt werden. Die Insertion von diversen SCC-Elementen ist eine weitere Erklärung für die hohe genetische Varianz der Staphylokokken sowie ihre Anpasstheit an lebenswidrige Bedingungen (Hiramatsu et al., 2013). Eine weitere Besonderheit der SCC*mec*-Elemente ist ihr Gentransfer über Art-Grenzen

hinweg. Studien lieferten Hinweise darauf, dass methicillinsensible *S. aureus* (MSSA) durch den Transfer der *mecA*-Gen-enthaltenden SCC*mec*-Kassette von methicillinresistenten *S. epidermidis* (MRSE) ebenfalls methicillinresistent wurden (Bloemendaal et al., 2010; Wielders et al., 2001). Im ersten Fall (Bloemendaal et al., 2010) konnte ausschließlich ein Unterschied in Form einer Punktmutation zwischen dem SCC*mec* Element des MRSE-Isolats und dem MRSA-Stamm nachgewiesen werden. Die Punktmutation war mit einem Basenpaaraustausch (Valin gegen Alanin) verbunden. In dieser Studie wurde aufgrund des geringen Unterschieds ein horizontaler Gentransfer zwischen dem MRSE und dem MSSA angenommen. Im zweiten Fall gab es keinen Unterschied zwischen dem Restriktionsenzym-Schnittmuster des *mecA*-Gens des methicillinresistenten *S. epidermidis* sowie dem Schnittmuster des isolierten MRSA. Dies weist auf einen identischen Aufbau des *mecA*-Gens hin, so dass der *S. epidermidis* als potentieller Überträger des *mecA*-Gens beschrieben wurde (Wielders et al., 2001)

Es konnten bereits 11 unterschiedliche Typen von SCC*mec* identifiziert werden, deren Aufbau auf der Internetplattform <http://www.sccmec.org/> öffentlich verfügbar sind (Ito, 2009).

In *S. epidermidis* konnten die fünf großen SCC*mec*-Element-Typen I bis V nachgewiesen werden. Verschiedene Arbeitsgruppen haben unterschiedliche SCC*mec*-Element-Typen gehäuft mit pathogenen *S. epidermidis* in Verbindung gebracht (Tab. 1.2). Je nach Region konnte ein lokal vorherrschender Typ identifiziert werden. Folglich kann aufgrund des Typs der SCC*mec*-Kassette nicht auf die Pathogenität eines Stammes geschlossen werden.

Tabelle 1.2 Durch Studien identifizierte, häufige SCC*mec*-Elemente in pathogenen *S. epidermidis*

| Referenz               | Die drei häufigsten SCC <i>mec</i> -Typen bezogen auf pathogene <i>S. epidermidis</i> |
|------------------------|---------------------------------------------------------------------------------------|
| Saffari et al., 2016   | IV (54%), I (15%), II (10%)                                                           |
| Oliveira et al., 2016  | III (38%), IVd (12%), III+IVa (8%)                                                    |
| Salgueiro et al., 2017 | IV (69%), III (15%), V (15%)                                                          |
| Soroush et al., 2016   | V (24%), III (8%), IV (6%)                                                            |

## 1.8 Multi-Locus-Sequence-Typing

Das *Multi-Locus-Sequence-Typing*-(MLST-)Verfahren stellt eine schnell durchführbare Alternative zu weiteren, zeitaufwendigeren Methoden der Klärung von bakteriellen Verwandtschaftsverhältnissen dar. Zu diesen weiteren Methoden zählen unter anderem die Pulsfeldgelelektrophorese (PFGE) oder die DNA-DNA-Hybridisierung, die beide bereits zur Charakterisierung von *S. epidermidis* angewandt wurden (Villari et al., 2000). Mit der molekularen Typisierung sollen zwei wichtige epidemiologische Faktoren geklärt werden. Einerseits soll die verwandtschaftliche Einordnung in einen weltweiten Kontext ermöglicht werden. Andererseits soll geklärt werden, ob pathogene Isolate eines lokalisierten Ausbruchs auf einen Stamm zurückzuführen sind (Maiden et al., 1998). Für den Begriff Ausbruch gibt es zwei mögliche Definitionen: Die erste Definition lautet, dass ein Ausbruch definiert wird als das Auftreten von mehr Fällen einer bestimmten Erkrankung als erwartet an einem bestimmten Ort oder in einer spezifischen Bevölkerungsgruppe innerhalb eines Zeitraums. Ein Ausbruch kann ebenfalls definiert sein als der häufigere Nachweis eines Pathogens wider Erwarten bei ansonsten gleichen Bedingungen (Quainoo et al., 2017). Im Rahmen dieser Arbeit wird auf die zweite Definition zurückgegriffen.

### 1.8.1 Das Prinzip des MLST-Verfahren

Im MLST-Verfahren werden sieben Genloci der sogenannten *housekeeping genes* von unterschiedlichen Stämmen miteinander verglichen. *Housekeeping genes* liegen in einer Spezies als differentielle Polymorphismen vor, die betreffenden evolutionäre Veränderungen finden jedoch sehr langsam statt. Jedem Allelmuster dieser Genloci wird ein sogenannter Sequenztyp zugeordnet (Maiden et al., 1998; Wang et al., 2003). Es wurden für mehrere Spezies bereits entsprechende MLST-Schemata entwickelt. Hierzu zählen beispielsweise *S. aureus* (Enright et al., 2000), *Neisseria meningitidis* (Maiden et al., 1998), *Campylobacter jejuni* (Dingle et al., 2001) oder auch Pilze wie *Candida albicans* (Bougnoux et al., 2004). Die ermittelten MLST-Schemata liegen auf zwei

Internetplattformen gesammelt vor: <http://www.mlst.net> (Aanensen and Spratt, 2005) und <http://pubmlst.org> (Jolley et al., 2004). Dies ermöglicht den globalen Austausch und eine einheitliche, laborunabhängige Typisierung zu epidemiologischen Zwecken (Urwin and Maiden, 2003).

### 1.8.2 Das MLST-Verfahren für *Staphylococcus epidermidis*

Zur Analyse von Verwandtschaftsverhältnissen von *S. epidermidis* konnte sich das verbesserte MLST-Schema von Thomas *et. al.* durchsetzen (Tab. 1.2). Es beinhaltet die Genloci *arcC*, *aroE*, *gtr*, *mutS*, *pyrR*, *tpiA* und *yqiL* (Thomas et al., 2007). Diese Kombination aus Genloci setzt sich zusammen aus Loci von bereits bestehenden MLST-Schemata für *S. epidermidis*. Die beiden ersten Schemata stammen von Wisplinghoff *et. al.* und Wang *et. al.* und sind äquivalent zu dem MLST-Schema für *S. aureus* (Wang et al., 2003; Wisplinghoff et al., 2003). Aus diesen Arbeiten stammen die Genloci *arcC*, *aroE*, *tpiA* und *yqiL* des verbesserten Schemas. Aus dem unpublizierten MLST-Schema von Peacock *et. al.* entstammen die Loci *gtr*, *mutS* sowie *pyrR* für das verbesserte Schema.

Tabelle 1.3 Die *housekeeping genes* der MLST-Schemata unterschiedlicher Arbeitsgruppen

| <i>housekeeping gene</i>                                                                       | Referenz                              |
|------------------------------------------------------------------------------------------------|---------------------------------------|
| <i>arcC</i> , <i>aroE</i> , <i>gtr</i> , <i>mutS</i> , <i>pyrR</i> , <i>tpiA</i> , <i>yqiL</i> | Thomas <i>et. al.</i> , 2007          |
| <i>dfp</i> , <i>gmk</i> , <i>gtr</i> , <i>mutS</i> , <i>pta</i> , <i>pyrR</i> , <i>xpt</i>     | Peacock <i>et. al.</i> , unpubliziert |
| <i>aroE</i> , <i>glpF</i> , <i>gmk</i> , <i>hsp60</i> , <i>pta</i> , <i>tpiA</i> , <i>yqiL</i> | Wang <i>et. al.</i> , 2003            |
| <i>arcC</i> , <i>aroE</i> , <i>glpK</i> , <i>gmk</i> , <i>pta</i> , <i>tpiA</i> , <i>yqiL</i>  | Wisplinghoff <i>et. al.</i> , 2003    |

Es konnten im Rahmen mehrerer Studien ein Zusammenhang zwischen bestimmten Sequenztypen und pathogenen *S. epidermidis* hergestellt werden. Jedoch schwankt in Abhängigkeit von der betrachteten Studie der Anteil der einzelnen MLST-Sequenztypen in der gesamten Stichprobe pathogener *S. epidermidis*. Bei einem Vergleich der aktuellen Studienlage (Tab. 1.3) sind die Sequenztypen ST2, ST5, ST59 und ST83 die vier häufigsten Sequenztypen, die in Blutkulturen nachgewiesen wurden. Aufgrund dieser Studienlage und

dem gehäuftem Nachweis der drei MLST-Sequenztypen ST2, ST5 sowie ST59 im Rahmen von Infektionen im Universitätsklinikum Schleswig-Holstein Lübeck, liegt der Schwerpunkt dieser Arbeit auf diesen drei klonalen Linien von *S. epidermidis*. Diese drei Sequenztypen konnten jedoch ebenfalls als Teil der kommensalen Flora von Patienten nachgewiesen werden (Miragaia et al., 2007). Folglich ermöglicht der MLST-Sequenztyp ebenfalls nicht die Differenzierung von pathogenen und kommensalen *S. epidermidis*.

Tabelle 1.4 Übersicht über die MLST-Sequenztypen, die in Blutkulturen im Rahmen anderer Studien identifiziert wurden

| Referenz                                                         | Zahl der untersuchten Blutkulturen | Die drei häufigsten MLST-Sequenztypen der Blutkulturen                                |
|------------------------------------------------------------------|------------------------------------|---------------------------------------------------------------------------------------|
| Oliveira et al., 2016                                            | n = 62                             | ST2 (27%), ST6 (11%), ST59 (6%), ST23 (6%)                                            |
| Bender et al., 2015                                              | n = 13                             | ST22 (38%), ST2 (38%), ST5 (23%)                                                      |
| Sharma et al., 2014                                              | n = 75                             | ST2 (27%), ST5 (17%), ST59 (6%), ST83 (6%)                                            |
| Mendes et al., 2012                                              | n = 49                             | ST5 (20%), ST2 (16%), ST83 (10%)                                                      |
| Iorio et al., 2012                                               | n = 13                             | ST2 (46%), ST23 (30%), ST53 (15%) ST59 (15%)                                          |
| Summe der MLST-typisierten Blutkulturen der untersuchten Studien | n = 212                            | ST2 (26%), ST5 (12%), ST59 (5%), ST83 (5%), ST23 (4%), ST6 (3%), ST22 (2%), ST53 (1%) |

Aufgeführt sind publizierte Studien mit dem Untersuchungszeitraum 2017 bis einschließlich 2010, in denen mehr als 10 Blutkulturisolat dem MLST-Verfahren nach Thomas unterzogen wurden

### 1.9 Das erweiterte MLST<sup>+</sup>-Verfahren und das core-Genom

Es zeigte sich im Verlauf der Verwendung von MLST, dass sich dieses Verfahren nur auf nah verwandte Bakterien anwenden lässt. Innerhalb einer Gattung werden teilweise mehrere MLST-Schemata benötigt. Mit Hilfe des MLST-Schema konnte beispielsweise nicht zwischen weiteren Subgruppen eines Ausbruchs von *Neisseria meningitidis* unterscheiden werden, die durch erweiterte MLST-Verfahren entdeckt wurden (Jolley et al., 2012). Zu diesen erweiterten MLST-Verfahren zählt auch das *whole-genome* MLST. Hierbei wird ein

Gen-zu-Gen-Vergleich ausgeführt um alle äquivalenten Genloci von unterschiedlichen Stämmen miteinander zu vergleichen. Dieses Verfahren eignet sich allerdings nur für nah verwandte Stämme oder *single-clone* pathogene Isolate mit geschlossenem Genom. Im Gegensatz dazu wird im Rahmen des *core*-Genom MLST (cgMLST) ein erstelltes *core*-Genom mit unterschiedlichen Stämmen verglichen. Das *core*-Genom ist ein Zusammenschluss von Genloci einer Bakterienart, die in allen zur Erstellung verwendeten Referenzgenomen auftreten. Demgegenüber steht das *accessory*-Genom aus den Genloci, die variabel in den Referenzgenomen auftreten (Maiden et al., 2013). Nach Erstellung des *core*-Genom werden die *small-nuclear* Polymorphismen (SNP) von beispielsweise einem Ausbruchsstamm zu den Genloci des *core*-Genom identifiziert. Die SNP stellen die Allelvariation der Genloci dar und können zwischen unterschiedlichen Stämmen verglichen werden. Nah verwandte Stämme teilen sich die gleichen Allelvariationen (Köser et al., 2012). Es wurden bereits cgMLST-Schemata unter anderem für *Salmonella enterica* (Toro et al., 2016), *Enterococcus faecium* (de Been et al., 2015), *Mycobacterium tuberculosis* (Kohl et al., 2014) und *Acinetobacter baumannii* (Higgins et al., 2017) erstellt. Die bereits annotierten *core*-Genome liegen auf der Internetplattform <http://www.cgmlst.org/ncs> der Ridom GmbH vor. Bei der Bewertung des cgMLST konnte eine hohe Genauigkeit und Reproduzierbarkeit zwischen fünf klinischen Laboren festgestellt werden (Mellmann et al., 2017). Aufgrund der Identifizierung der klonalen Verwandtschaft können auch Transmissionereignisse erkannt werden, da die verwandten Stämme sich in nur einer geringen Anzahl von SNP unterscheiden (Mellmann et al., 2016).

### 1.10 Fragestellung

Die bisherigen Studien konnten keinen Marker für pathogene *S. epidermidis* identifizieren. Im Rahmen dieser Studien zeigte sich, dass die klassischen Resistenz- und Biofilm-assoziierten Gene für die Differenzierung von pathogenen und kommensalen Stämmen von Patienten ungeeignet sind. Diese mangelnde Differenzierung ist vereinbar mit der Arbeit von Hercun (Hercun, 2011) sowie

weiteren Studien (Widerström et al., 2016), die eine Kolonisierung des Patienten mit potentiell pathogenen *S. epidermidis* aus der Klinikumwelt vermuteten. Aus Materialien der klinischen Diagnostik (Blutkulturen) konnte nachgewiesen werden, dass die pathogenen Isolate weniger divers in ihrer genetischen Struktur sind. Mit Pathogenität wurden insbesondere drei MLST-Sequenztypen (ST2, ST5, ST59) in Verbindung gebracht. Die Fragestellung dieser Arbeit war, ob sich für diese Sequenztypen Markersequenzen extrahieren lassen, die den jeweiligen Sequenztyp verlässlich zuordnen. Im Fall des Vorliegens geeigneter Sequenzen sollte anhand der kodierten Proteinprodukte die erhöhte Pathogenität des entsprechenden Sequenztyps erklärt werden. Ebenfalls sollte die Eignung der potentiellen Marker untersucht werden, pathogene von apathogenen *S. epidermidis* zu differenzieren.

## 2. Material

### 2.1 Bakterienstämme

Die Bakterienstämme entstammen der Stammsammlung der Arbeitsgruppe Knobloch, die hauptsächlich Patienten-Proben des UKSH repräsentieren. Die *Staphylococcus epidermidis*-Stämme mit Linezolidresistenz wurden im Verlauf der Entstehung dieser Arbeit der Sammlung hinzugefügt. Ausnahme bilden die von unbelebten Oberflächen gesammelten Stämme mit den Bezeichnungen ENVH, ENVP und ENVL sowie die bovinen Mastitis-Stämme. Zusätzlich wurde auf eine Auswahl von bereits auf der Plattform NCBI annotierten *S. epidermidis* zurückgegriffen (Tab. 8.1).

### 2.2 Einwegprodukte

Tabelle 2.1: Verwendete Einwegprodukte sowie ihre Hersteller

| Produkt                        | Firma                  |
|--------------------------------|------------------------|
| High Sensitivity DNA Chip      | Agilent Technologies   |
| Kryoröhrchen Cryobank - blau   | Mast Diagnostica GmbH  |
| Qubit® Assay Tubes 500µl       | Invitrogen             |
| Reagenzreservoirs 25ml, steril | VWR International GmbH |

### 2.3 Chemikalien und Nährmedien

Tabelle 2.2: Verwendete Chemikalien und Nährmedien sowie ihre Hersteller

| Produkt                                          | Hersteller                |
|--------------------------------------------------|---------------------------|
| Bacto Tryptic Soy Broth (TSB)- Medium 5 ml       | Becton Dickinson          |
| Blut-Agarplatten (COS-Platten)                   | bioMérieux SA             |
| Ethanol                                          | Carl Roth GmbH+Co KG      |
| Glycin                                           | AppliChem GmbH            |
| Lysostaphin                                      | Genmedics                 |
| Proteinase K, recombinant, PCR Grade             | Roche Diagnostics GmbH    |
| Tris-Puffer (10 mmol Tris HCl+0,1%Tween; pH 8,5) | Sigma-Aldrich Chemie GmbH |

## 2.4 Geräte

Tabelle 2.3: Verwendete Geräte sowie ihre Hersteller

| Name                                          | Hersteller           |
|-----------------------------------------------|----------------------|
| 2100 Bioanalyzer                              | Agilent Technologies |
| - 80 °C Gefrierschrank Sanyo                  | Sanyo                |
| Inkubationsschüttler SM 30 control TH30       | Edmund Bühler        |
| Inkubator HERACell 240 mit 5% CO <sub>2</sub> | Heraeus              |
| NanoPhotometers™ P330                         | Implen               |
| Qubit® 2.0 Fluorometer                        | Invitrogen           |
| Thermocycler FlexCycler                       | Analytik Jena AG     |
| Thermomixer comfort                           | Eppendorf AG         |
| Vortex Genius 3                               | Ika                  |
| Wasserbad                                     | GFL                  |
| Zentrifuge 5430                               | Eppendorf            |

## 2.5 Kits

Tabelle 2.4: Verwendete Kits sowie ihre Hersteller

| Name                                    | Hersteller           | Verwendung                                            |
|-----------------------------------------|----------------------|-------------------------------------------------------|
| Agencourt AMPure XP 5ml Kit             | Beckman Coulter GmbH | DNA-Sequenzierung                                     |
| Agilent High Sensitivity DNA Kit        | Agilent Technologies | Bioanalyzer-Messung                                   |
| MiSeq Reagent Kit v3                    | Illumina Inc.        | DNA-Sequenzierung                                     |
| Nextera® XT DNA Library Preparation Kit | Illumina             | DNA- <i>library</i> -Erstellung                       |
| Nextera® XT Index                       | Illumina Inc.        | DNA- <i>library</i> -Erstellung                       |
| PhiX Control v3                         | Illumina Inc.        | Kontrolle für DNA-Sequenzierung                       |
| Qubit® dsDNA HS Assay Kit               | Invitrogen           | DNA-Isolierung                                        |
| UltraClean® Microbial DNA Isolation Kit | Mo Bio               | DNA-Isolierung                                        |
| WATER, DNase-, RNase-free               | Illumina Inc.        | DNA- <i>library</i> -Erstellung und DNA-Sequenzierung |

## 2.6 Programme und Datenbanken

Tabelle 2.5: Verwendete Programme und Datenbanken sowie ihre Provider

| Name                                                                        | Hersteller/Provider                                  |
|-----------------------------------------------------------------------------|------------------------------------------------------|
| NCBI GenBank                                                                | <i>National Center for Biotechnology Information</i> |
| CLC Genomics Workbench<br>8.0.1                                             | Qiagen                                               |
| MLST-Datenbank<br>( <a href="http://www.mlst.net">http://www.mlst.net</a> ) | Imperial College London                              |
| Microsoft Word 2007                                                         | Microsoft Corporation                                |
| Microsoft Excel 2007                                                        | Microsoft Corporation                                |
| Ridom SeqSphere + 2.3.1                                                     | ridom bioinformatics                                 |
| Mauve - multiple genome<br>alignment                                        | Darling lab der University of Technology<br>Sydney   |
| RStudio 0.98.1091                                                           | RStudio Inc.                                         |

### 3. Methoden

#### 3.1 Kultur und Anzuchtbedingungen

Für die Anzucht der Bakterien zur Gewinnung der chromosomalen DNA für die Genomanalyse wurden die Stämme auf Blut-Agarplatten (COS-Platten) bei 37°C über Nacht aus den Kryokulturen reaktiviert. Von den Blut-Agarplatten wurde eine Kolonie in 5 ml TSB + 1 % Glycin als Flüssigkultur überführt und bei 150 upm und 37°C über Nacht im Schüttelinkubator kultiviert. Die Zugabe der 1 % Glycin zu dem Medium sollte die Ausbildung der Pentaglycinketten begünstigen (Petit et al., 1968).

#### 3.2 Isolierung der chromosomalen DNA der Mikroorganismen

Zur Isolierung der DNA aus den Staphylokokken-Kulturen wurde das Mo Bio UltraClean® Microbial DNA Isolation Kit verwendet. Aus einer Übernachtskultur in TSB + 1 % Glycin wurde 2 ml in ein Eppendorfgefäß übernommen und bei 16.500 g für 2 Minuten zentrifugiert. Der Überstand wurde verworfen und der Schritt wurde erneut mit 2 ml Übernachtskultur wiederholt, so dass ein Pellet aus insgesamt 4 ml Übernachtskultur sich bildete. Dieses Pellet wurde in 300 µl *MicroBead*-Lösung resuspendiert. Diese enthält Puffersysteme, die die Zellen in der Lösung homogenisieren und die Lyse vereinfachen. Abweichend vom Protokoll wurde für den Aufschluss der Zellen zusätzlich Lysostaphin, eine Glycyl-Glycin-Endopeptidase zugesetzt. Lysostaphin spaltet spezifisch die Pentaglycinketten des Peptidoglykan-Netzwerks von Staphylokokken. Es wurden 5 µl einer Lysostaphin-Lösung mit einer Konzentration von 10 mg/ml zu den resuspendierten Zellen gegeben und für 30 Minuten bei 37°C inkubiert. Anschließend wurde die Suspensionslösung zur weiteren Lyse mit 25 µl Proteinase K (20 mg/ml) versetzt, bevor sie für weitere 90 Minuten bei 65°C inkubiert wurde. Diese Lösung wurde in *MicroBead* Tubes überführt und anschließend mit 50 µl MD1-Lösung, die neben SDS auch weitere Chemikalien zur Zellyse enthält, versetzt. Die Tubes wurden für 10 Minuten bei maximaler Geschwindigkeit gevor-

text, um durch die Kombination aus chemischer und mechanischer Lyse die DNA aus den Bakterien freizusetzen. Die sich anschließende Zentrifugation bei 10.000 g für 30 Sekunden diente dem Absenken der verbliebenen Zellbestandteile. Der Überstand wurde in ein neues Tube überführt sowie mit 100 µl MD2-Lösung versetzt, die die Nicht-DNA-Bestandteile präzipitieren lässt. Nach einer Inkubation bei 4°C für 5 Minuten wurden die Tubes mit 10.000 g zentrifugiert. Der Überstand wurde abgenommen und mit 900 µl MD3-Lösung versetzt, die aufgrund des hohen Salzgehalts die DNA-Bindung an einen Silica-Filter ermöglicht. 700 µl dieser Lösung wurden auf einen *Spin Filter* aus Silica-Membran gegeben, um bei 10.000 g für 30 Sekunden zentrifugiert zu werden. Die abgefilterte Lösung wurde verworfen, da sie aus nicht an den Filter gebundenen Nicht-DNA-haltigen Stoffen besteht. Der Prozess wurde mit dem verbleibenden Überstand wiederholt. Zu dem nun verbliebenen Überstand im Filter wurden 300 µl MD4-Lösung, einer Waschlösung auf Ethanolbasis, die die am Filter gebundene DNA weiter aufreinigt, hinzugefügt, bevor der Filter erneut bei 10.000 g für 30 Sekunden zentrifugiert wurde. Die abgefilterte Lösung wurde verworfen. Nach einer erneuten Zentrifugation bei 10.000 g für 1 Minute, wurde der Filter in ein neues Tube überführt. Auf den Filter wurden 50 µl MD5-Lösung gegeben, die Tris-Puffer zur Auswaschung der DNA aus dem Filter enthält. Die Tube wurde bei 10.000 g für 30 Sekunden zentrifugiert. Da die DNA nur in salzhaltiger Lösung an den Silica-Filter binden kann, aber nicht in salzfreiem Tris-Puffer, lag sie nun gelöst im Überstand vor.

### 3.3 Qualitäts- und Quantitätsbestimmung

Es wurde mittels NanoPhotometers™P330 eine photometrische Messung der Extinktion bei 260 nm und 280 nm mit dem Ziel durchgeführt, die DNA-Konzentration sowie -Reinheit zu bestimmen. Der Quotient aus beiden Absorptionswerten erlaubt eine Aussage für das Maß an Verunreinigung durch Proteine. Ein 260 nm/280 nm Absorptionswert-Quotient von 1,8 bis 2 dient als ein Zeichen für eine qualitativ gute Isolierung. Das Prinzip bei der Quotientenbildung ist die unterschiedlich starke Absorption von Proteinen beziehungsweise DNA bei die-

sen beiden Wellenlängen. Eine Extinktion von 1 bei 260 nm bedeutet, dass in der Lösung eine Konzentration einer doppelsträngigen DNA von 50 µg/ml gegeben ist. Dieses Verfahren dient der Bestimmung der Qualität der Probe.

Vor Erstellung der *DNA-libraries* wurde die DNA-Konzentration mit dem fluorometrischen Verfahren des Qubit® 2.0 Fluorometer und Qubit® dsDNA HS Assay Kit nach Herstellerangaben bestimmt. Dieses Verfahren ermöglichte eine genauere Bestimmung der Konzentration als die Extinktionsmessung. Das Prinzip beruht darauf, dass nur die doppelsträngigen DNA-Moleküle (dsDNA) mit einem spezifischen Fluoreszenzfarbstoff angefärbt und gemessen werden. Mögliche Kontaminationen mit RNA, Proteinen oder Salzen können mit dieser Methode nicht nachgewiesen werden.

### 3.4 Erstellung der *DNA-libraries*

Zur Erstellung der *DNA-libraries* wurde das Nextera® XT DNA Library Preparation Kit nach Herstellerangaben verwendet. Zunächst wurde im Prozess der Tagmentierung die DNA der jeweiligen Probe mit Hilfe von Transposomen in DNA-Fragmente zerschnitten und gleichzeitig mit Adaptersequenzen versehen. Im zweiten Schritt wurden die so gewonnenen DNA-Fragmente mittels Polymerase-Ketten-Reaktion vervielfältigt. Daraufhin erfolgte die Aufreinigung der *library* DNA an magnetischen *beads*, die mit Hilfe ihrer physikalischen Wechselwirkung die Negativ-geladene DNA binden. Im magnetischen Ständer akkumulierten die *beads* an einer Gefäßwand, so dass der DNA-freie Überstand abgenommen werden kann. Verbliebene Pufferlösung wurde in einem zweifach ausgeführten Schritt mit Ethanol ausgewaschen. Die DNA-bindenden *beads*, wurden zunächst getrocknet, bevor sie mit Resuspensionspuffer versetzt werden. Nach erneutem Schütteln mittels Vortex und 2 Minuten Inkubationszeit war die DNA aus den *beads* in den Resuspensionspuffer überführt. Nach erneuter Inkubation im magnetischen Ständer wurde die DNA-haltigen Pufferlösung von den nun DNA-freien *beads* getrennt.

Die Qualität der DNA-*libraries* wurde im nächsten Schritt durch die Bestimmung der Fragment-Größe mittels des Agilent Technology 2100 Bioanalyzer in 1 µl unverdünnter Probe bestimmt. Dafür wurden zunächst die High Sensitivity DNA Chips mit Hilfe des Agilent High Sensitivity DNA Kits mit einem siebenden Polymer sowie einem Fluoreszenzfarbstoff präpariert. Die 16 *wells* eines Chips sind mittels Mikro-Kanäle untereinander verbunden, sodass sie nach der Gel-Präparation ein in sich geschlossenen, elektrischen Kreis bilden. Jedes *well* verfügte über eine eigene Elektrode, sodass in jedem *well* durch einen elektrischen Gradienten eine kontrollierte, elektrophoretische Auftrennung stattfinden kann. Die Polymerstruktur in Kombination mit einem konstantem Masse-zu-Ladungs-Verhältnis bewirken im elektrischen Feld eine Auftrennung der dsDNA nach Größe. Je kleiner das Fragment, desto schneller durchwandert es die Matrix. Das dem Gel zugefügte Färbemittel interkalierte mit den DNA-Fragmenten, sodass Komplexe entstanden, deren Fluoreszenz vom BioAnalyzer durch Laser-Induktion nachgewiesen werden konnte. Der Analyzer erstellte aus den so gewonnenen Daten einerseits ein Bandenmuster und andererseits ein Elektropherogramm. Durch die Zugabe verschiedener Standards mit bekannter Größe, konnte das Programm die Fragment-Größe erfassen und in einen Kontext zu den Fragment-Größen der Proben setzen.

Nach dieser Qualitätsprüfung folgte der Arbeitsschritt der abschließenden Alkalisierung der einzelnen Proben mit NaOH. Durch das NaOH denaturierte die Doppelstrang- in Einzelstrang-DNA, die in annähernd gleicher Menge gepoolt wurde. Anhand ihrer Adapter- und Index-Sequenzen lassen sich die gepoolten DNA-Fragmente nach der Sequenzierung zum ursprünglichen Bakterienstamm zurückverfolgen.

### **3.5 Sequenzierung der DNA mittels *Next-Generation Sequencing***

Diese Arbeitsschritte wurde mit dem Illumina MiSeq® System sowie dem MiSeq Reagent Kit v3 nach Herstellerangaben durchgeführt.

Vor Beginn der eigentlichen Sequenzierung folgte der Arbeitsschritt der Verdünnung und Denaturierung nach dem Hersteller-Protokoll für *beads*-basierte *library*-Herstellung. Hierbei wurden 24 µl der *library* in einem Reaktionsgefäß mit 576 µl vorgekühltem Hybridisierungspuffer 1 versetzt. Nach einer einminütigen Zentrifugation wurde das Reaktionsgefäß in einem auf 98°C vorgeheiztem Inkubator für 2 Minuten inkubiert, bevor es auf Eis für 5 Minuten gekühlt worden ist. Anschließend erfolgte die Beladung der *cartridge* des MiSeq Systems mit der so gewonnenen Probe.

Das Prinzip der Sequenzierung mittels Illumina MiSeq® beruht auf mehreren Schritten:

### 3.5.1 Cluster-Bildung

Für die Cluster-Bildung wurden *flow cells* vorgegeben, die an ihrer Oberfläche zu den *library*-Adaptoren komplementäre Oligonukleotid-Sequenzen gebunden hatten. Auf diese *flow cells* wurden die *DNA-libraries* aufgetragen. Die *DNA*-Fragmente mit der komplementären Adaptersequenz wurden an die *flow cell* Oligonukleotid-Sequenz in bestimmten Clustern gebunden und mittels *bridge amplification* auf dem *flow cell* vervielfältigt. Es kam zur komplementären Bindung des *DNA*-Fragments und der lokalisierten *flow cell*-Sequenz über den Adapter 1. Nach der ersten Polymerisation sowie dem Auswaschen des ursprünglichen Fragments blieb die an der *flow cell* gebundene, komplementäre Sequenz des Fragments zurück. Das Prinzip der *bridge amplification* verläuft ähnlich dem PCR-Prozess. Der Adapter 2 der *flow cell* gebundenen Sequenz bindet komplementär an die passende Oligonukleotidsequenz der *flow cell*. Durch die darauf folgenden Schritte der Polymerisation sowie der Denaturierung entstand im gleichen Cluster eine weitere, gebundene Kopie des ursprünglichen Fragments. Der Arbeitsschritt der *bridge amplification* generierte so lokal eine große Menge an klonalen Kopien des ursprünglichen Fragments. Dies bedeutet eine Signalverstärkung für die folgende Sequenzierung.

### 3.5.2. Sequenzierung

Das Illumina MiSeq®System verwendet die *sequencing by synthesis* Technologie. Hierbei werden Fluoreszenz-markierte Deoxynucleosid-Triphosphate (dNTP) verwendet. Diese wirken als reversible Terminatoren der DNA-Synthese, da ihr modifizierter Nukleotid-Abschnitt bei Einbau in den komplementären DNA-Strang durch die DNA-Polymerase einen Kettenabbruch induziert. Jedes dNTP mit der gleichen Base hat ein bestimmtes Fluoreszenzsignal. Zunächst werden auf die *flow cell* mit den entsprechenden Clustern die Primer, die DNA-Polymerasen sowie die markierten dNTP gegeben. Durch Einbau der ersten Base am komplementären DNA-Strang kommt es zum Kettenabbruch. Das Fluoreszenzsignal wird durch einen Laser angeregt und für jeden Cluster gemessen. Das ermöglicht die Zuordnung von Fluoreszenzsignal zu Cluster und damit auch zur ersten Base des entsprechenden Clusters. Im nächsten Schritt wird mittels einer enzymatischen Reaktion der Nucleotidteil gespalten, der zuvor den weiteren Einbau von dNTP in den komplementären DNA-Strang unterbunden hat. Die nächste markierte Base wird eingebaut. Dies führt zum Kettenabbruch. Das Fluoreszenzsignal wird bestimmt und den Clustern kann die zweite Base zugeordnet werden. Mit jedem weiteren Zyklus kann eine weitere Base des ursprünglichen Fragments bestimmt werden. Diese so gewonnene Basenabfolge wird als *read* bezeichnet. Ein Vorteil der vorangegangenen Cluster-Bildung ist, dass das Fluoreszenzsignal verstärkt wird. Da auf dem *flow cell* die Kopien des ursprünglichen Fragment an beiden Adaptoren gebunden vorliegen, werden diese von beiden Enden ausgehend sequenziert. Es erfolgt eine sogenannte *paired-end*-Sequenzierung. Dabei werden sowohl *forward reads*, in der Leserichtung des Fragments, als auch *reverse reads*, entgegen der Leserichtung des eigentlichen Fragments, gebildet. Mit Hilfe von bioinformatischen Programmen wird aus den komplementären *forward* und *reverse reads* *read pairs* gebildet. Der Vorteil hierbei ist, dass repetitive Abschnitte der Sequenz besser abgebildet werden können als mit *single-read*-Verfahren, da der Abstand zwischen den einzelnen *paired-end reads* auf die Referenzsequenz vorgegeben ist.

Von der Arbeitsgruppe neu sequenzierte Isolate werden in der weiteren Arbeit auch als eigen sequenzierte Isolate bezeichnet. Entstammen Genome der Sequenzdatenbank NCBI werden sie im Folgenden bereits annotierte Stämme genannt.

### 3.6 Sequenzanalysen

Die durch NGS gewonnen Rohdaten wurden im *pairing* und *trimming* nach den Standard-Voreinstellungen der CLC Genomics Workbench prozessiert. Im Prozess des *pairing* wurden die Einzelstrang-*read*-Sequenzen ihrem komplementären Partner zugeordnet. Durch das *trimming* wurden nicht eindeutig zugeordnete Basen, sowie die bestehenden Adaptersequenzen aus den Rohsequenzen entfernt. Es wurden sogenannte getrimmte *paired reads* gebildet.

Das *mapping* erfolgte unter den Standardeinstellungen der CLC Genomics Workbench gegen das auf NCBI vollständig annotierte Genom des *Staphylococcus epidermidis*-Stammes RP62A (NCBI Accession number: CP000029.1). Hierbei wurden die *paired reads* mit dem Grundgerüst des RP62A-Genoms abgeglichen. Jene *paired reads*, die nicht homolog zu Sequenzen dieses Genoms waren, wurden als sogenannte *unmapped reads* gespeichert.

Durch den Schritt des *de-novo-assembly* entsprechend den Workbench vorgegebenen Einstellungen werden die Sequenzen anhand von Sequenzüberlappungen zu längeren Sequenzabschnitten verbunden. Diese längeren Sequenzabschnitte werden *contigs* genannt. Zur Qualitätsbewertung eines *de-novo-assembly* werden die mittlere *contig*-Länge und der N50-Wert herangezogen. Je größer die beiden Werte eines *de-novo-assembly* sind, desto besser ist dessen Qualität. Der N50-Wert ist definiert als gewichteter Median der Sequenzlängen eines *de-novo-assembly* (Earl et al., 2011) und wird berechnet mit Hilfe der Addition der Länge der längsten *contigs* bis 50% der totalen *contig*-Länge erreicht sind. Ein *de-novo-assembly* der *paired reads* der Rohdaten als auch ein *de-novo-assembly* der *unmapped reads* des *mapping* wurden durch-

geführt. Das *de-novo-assembly* der *paired reads* entspricht einem Entwurf des Genoms für den entsprechenden Stamm. Neben diesen so gewonnen Genomen wurde in die Analysen eine Auswahl von *S. epidermidis*-Genomen einbezogen, die bereits von anderen Arbeitsgruppen sequenziert wurden. Diese Auswahl aus den, zu dem damaligen Zeitpunkt verfügbaren, *S. epidermidis*-Genomen wurden der NCBI-Datenbank entnommen (Tab. 8.1).

Anhand eines *gene-prediction*-Algorithmus der CLC Genomics Workbench konnte auf den Sequenzabschnitten *open reading frames* (ORFs) identifiziert werden. Die ORFs entsprechen den potentiellen proteinkodierenden Sequenzen eines *contigs*, da der Algorithmus die Sequenz zwischen potentiellen Start- und Stop-Codonen identifiziert. Es wurde sowohl der *forward*- als auch der *reverse*-Strang auf ORFs untersucht. Die Länge eines ORF musste mindestens 300 bp umfassen und über das Start-Codon AUG oder GUG verfügen.

Durch die Basic-Local-Alignment-Search-Tool (BLAST)-Funktion der CLC Genomics Workbench wurden nach den Voreinstellungen Homologien zwischen einzelnen Sequenzen ermittelt.

### 3.7 Schema der Identifizierung von Markersequenzen

Die Suche nach den sequenztypspezifischen Markersequenzen erfolgte mit Hilfe der Stämme, die im Rahmen dieser Arbeit sequenziert wurden. Es wurde ein *mapping* gegen das ST10-Referenzgenom RP62A als subtrahierende Sequenzanalyse durchgeführt. Sollten sequenztypspezifische Sequenzen vorliegen, waren sie Teil der *unmapped reads*. Aus diesen *unmapped reads* konnte mittels des *de-novo-assembly*-Verfahrens entsprechend den Einstellungen der CLC Genomics Workbench *contigs* gebildet werden. Für den folgenden Schritt wurden ausschließlich *contigs* mit einer Länge von mindestens 5.000 bp verwendet. Mit dem *gene-prediction*-Algorithmus der CLC Genomics Workbench wurde entsprechend den Voreinstellungen auf jenen differentiellen *contigs* ORFs gesucht. Im folgenden Schritt wurde diese Gruppe aus allen identifizierten ORFs einem BLAST nach den Standard-Einstellungen der Workbench gegen sich

selbst durchgeführt. Es wurden jene ORFs weiter untersucht, die mindestens drei Übereinstimmungen mit ORFs anderer Isolate des gleichen Sequenztyps hatten. Ein zweiter BLAST dieser gefilterten ORFs gegen die *contigs* der *de-novo-assemblies* der ursprünglichen *reads* der eigenen Isolatsammlung wurde durchgeführt. Auf diesem Weg wurde überprüft, ob die ORFs sequenztypspezifisch auftreten. Erfüllte ein ORF oder eine ORF-Abfolge diese Kriterien, wurden sie auf einem *contig* eines qualitativ hochwertig sequenzierten Stammes mittels BLAST identifiziert. Ein BLAST dieses *contig* wurde gegen die *contigs* der Stammsammlung durchgeführt. Im Folgenden wurde die finale Markersequenz extrahiert: Hierfür wurde der Abschnitt des entsprechenden *contigs*, der den zuvor identifizierten ORF beziehungsweise ORF-Abfolge enthält, sowie der unmittelbar benachbarte Sequenzteil herausgeschnitten. Diese benachbarten Abschnitte sollten ausschließlich den Sequenzteil umfassen, dem durch das BLAST nur Homologien in anderen Stämmen des entsprechenden Sequenztyps zugeordnet werden konnte. Von den so extrahierten Sequenzen wurden ausschließlich solche mit einer Länge von mindestens 500 bp als Marker berücksichtigt. Die Homologie zwischen den Markern und sowohl den eigen sequenzierten als auch den NCBI-entnommenen Genomen konnten mittels BLAST der Workbench ermittelt werden. Ein Stamm wurde als markerpositiv definiert, falls er einen Marker-homologen Sequenzanteil von 75% oder mehr besitzt. Für die Stämme, die Markeranteile auf unterschiedlichen Sequenzabschnitten (*contigs*) vorweisen, wurde die Summe der Längen der homologen Anteile gebildet. Entsprechend den gewählten Kriterien wurden diese Summen geprüft und die Stämme als markerpositiv oder -negativ bewertet. Zusätzlich wurde mit Hilfe des *gene-prediction*-Algorithmus der CLC Genomics Workbench entsprechend den Voreinstellungen die ORFs der finalen Markersequenz ermittelt.

### 3.8 Identifizierung von Homologien zu bereits annotierten Genomen

Zur Identifizierung von Homologien zwischen Sequenzen und bereits annotierten Genomen wurde das *Basic Local Alignment Search Tool* (BLAST) der Se-

quenzdatenbank der U.S. National Library of Medicine verwendet. Mehrere, unterschiedliche BLAST-Anwendungen stehen für diesen Zweck zur Verfügung. Der dominierende Anteil der Sequenzen wurde der Anwendung *megablast* unterzogen. Der *megablast* ist geeignet, Homologien zwischen eng verwandten Stämmen zu identifizieren. Eine Sequenz wurde dem sogenannten *discontiguous megablast* unterworfen. Durch den Algorithmus des *discontiguous megablast* können Homologien zu weniger verwandten Sequenzen mit einem höheren Maß an Diskrepanz zugeordnet werden. Die Anwendung des *discontiguous megablast* wird im Rahmen des entsprechenden Stammes gesondert erwähnt.

### 3.9 Multi Locus Sequence Typing

Das *Multi Locus Sequencing Typing* (MLST) System ermöglicht die Identifizierung einzelner Stämme in einer Art bakterieller Organismen. Hierfür wurden bei einer Population *Staphylococcus epidermidis* die Allelvariationen der sieben konservierten, sogenannten *housekeeping*, Genen bestimmt. Ein BLAST mit den Standard-Voreinstellungen der CLC Genomics Workbench wurde mit den bekannten Allelmutationen der *housekeeping* Genen *arcC*, *aroE*, *gtr*, *mutS*, *pyrR*, *tpiA* und *yqiL* (Thomas et al., 2007) durchgeführt. Anhand der Kombination der einzelnen Mutationen konnte den Stämmen dieser Arbeit ein Sequenztyp zugeordnet werden, der auf der Plattform <http://www.mlst.net> definiert wurde.

### 3.10 Core-Genom und MLST+

Mit Hilfe des Programms Ridom SeqSphere wurde ein *core genome* erstellt. Als Referenzgenom diente das Genom des *Staphylococcus epidermidis*-Stammes RP62A (NCBI Accession Number: CP000029.1). Voraussetzungen für die Referenzsequenz sind die drei Kriterien der vollständigen Sequenzierung, An-

notierung sowie allgemeine Zugänglichkeit. Mittels der MLST+ *target-definer*-Funktion wurde ein Gen-für-Gen-Vergleich durchgeführt. Im Rahmen dieses Verfahrens wurden entsprechend den Voreinstellungen *targets* genannte Genloci des Referenzgenoms herausgefiltert, die für das *core genome*-Grundgerüst geeignet waren (Leopold et al., 2014). Ausgeschlossen wurden beispielsweise Genloci ohne Start-oder Stop-Codon. Mit diesem Grundgerüst wurde ein BLAST des Programms Ridom SeqSphere gegen die beiden Vergleichsgenome der ebenfalls auf NCBI zugänglichen *S. epidermidis* Stämme ATCC 12228 (NCBI Accession Number: AE015929.1) und PM221 (NCBI Accession Number: HG813242.1) durchgeführt. Die Genloci des *core genome*-Grundgerüsts bildeten das endgültige *core genome*, die in beiden Genomen mit einer *sequence identity* von 90% oder mehr sowie 100% Übereinstimmung vorkamen. Neben dem *core genome* wurde ein *accessory genome* sowie das *discarded genome* erstellt. Das *accessory genome* beinhaltet jene Genloci, die sich gegenseitig im Referenzgenom überlappten, nicht in allen Vergleichsgenomen zu finden waren, oder mindestens 80% nicht funktionale Stop-Codons in den beiden Vergleichsgenomen besaßen. Das *discarded genome* besteht aus Genloci des Referenzgenoms, die zueinander homolog waren oder nicht funktionierende Start/Stop-Codons enthielten. Die Gene des *accessory genome* sowie des *discarded genome* wurden nicht weiter für die Analyse der klonalen Verwandtschaft verwendet.

Die Sequenzabschnitte der *de-novo*-assemblierten Rohdaten sind auf ihre Homologie zu den Loci des *core genome* untersucht worden. Entsprechend den Herstellerangaben liegt ein erfolgreich validiertes *core genome* vor, wenn mindestens 98,5% der Zielloci des *core genome* Homologie zu dem erneut sequenzierten Referenzgenom besitzen. Aus diesem Grund wurde ebenfalls der eigen sequenzierte Referenzstamm RP62A auf Homologie zu dem *core genome* untersucht. Die Ergebnisse des *built-in* BLAST für jedes Genom der untersuchten Isolate wurde differenziert in *good targets*, *failed targets* und *not found targets*. Die *good targets* sind definiert als die Gene des jeweiligen Stammes, die extrahiert wurden sowie die vom Programm vorgegebenen Einstellungen erfüllten. *Failed targets* waren solche Genloci, die extrahiert wurden, allerdings nicht die Einstellungen erfüllten. *Not found targets* konnten nicht aus den Stämmen ex-

trahiert werden. Zur weiteren phylogenetischen Untersuchung wurde durch SeqSphere eine Vergleichstabelle (*comparison table*) aus den so gewonnenen Daten gebildet. In dieser Vergleichstabelle wurden die Allelvariationen der Genloci des *core*-Genom der einzelnen Stämme gelistet. Mit Hilfe der Angaben der Vergleichstabelle wurde ein *minimum spanning tree* (MST) jeweils für die Sequenztypen 2, 5 sowie 59 erzeugt. Nur die Stämme mit weniger als 80 fehlenden Genloci des *core genome* wurden zur Bildung eines MST verwendet. Hierbei wurde die Einstellung "paarweises Ignorieren fehlender Werte" ("*pairwise ignore missing values*") gewählt. Durch diese Einstellung wurden für die Erstellung des MST jeweils die Genloci des *core genome* des jeweiligen Stammes ignoriert, die nicht in diesem Stamm identifiziert werden konnten. Ridom SeqSphere verknüpfte im MST die Stämme miteinander, die die geringsten Unterschiede in den Allelvariationen der Genloci des *core*-Genoms besaßen. Alle Genome von Stämmen mit weniger als 100 Genloci Unterschied werden in der weiteren Arbeit als ein zusammengehöriges Cluster bezeichnet. Zusätzlich identifizierte Ridom SeqSphere nah verwandte Stämme, die sich in maximal 50 Allelen unterschieden und bildete aus ihnen sogenannte Komplexe. Im MST wurden die Markerkombinationen der entsprechenden Sequenztypen unterschiedlich angefärbt.

### 3.11 Bestimmung der Varianz des markerpositiven *contigs*

Mit Hilfe des *progressive Mauve Alignment* sollte der Grad an genetischer Varianz der markertragenden *contigs* identifiziert werden. Hierfür wurde jeweils ein Abschnitt eines markerpositiven *contigs* sowie homologe Abschnitte von *contigs* markernegativer Stämme einem *alignment* unterzogen. Mittels der CLC Genomics Workbench wurden die benötigten Sequenzstücke für das Alignment extrahiert: Im BLAST-Ergebnis der Markersequenz gegen die Sequenzabschnitte der de-novo-assemblierten Rohdaten (*contigs*) wurde jener Sequenzabschnitt des entsprechenden Sequenztyps identifiziert, in das der Marker mitig inseriert. Aus diesem Sequenzabschnitt (*contig*) wurde eine Sequenz extrahiert, die den Markerbereich sowie die jeweils einrahmenden 1500 bp umfasst.

Diese gekürzte Sequenz des markerpositiven Sequenzabschnitts wurde in der CLC Genomics Workbench einem BLAST gegen die Sequenzabschnitte (*contigs*) der markernegativen Stämme unterzogen. Aus dem Ergebnis wurde eine Auswahl von *contigs* extrahiert, die jeweils zwei homologe Abschnitte besitzen. Aus dieser Auswahl *contigs* der markernegativen Stämme wurde jeweils der Abschnitt herausgeschnitten, der von den beiden homologen Sequenzstücke umfasst wird. Die erste Base des Sequenzausschnitts ist die erste Base der ersten Homologie beziehungsweise die letzte Base des Ausschnitts ist die letzte Base der zweiten Homologie. Diese Sequenzausschnitte als Referenzsequenzen sowie das gekürzte markertragende *contig* wurden aus der CLC Genomics Workbench als fasta-Dateien exportiert und dem *progressive Mauve Alignment* entsprechend den Standardeinstellungen unterzogen

*Mauve* ist ein Programm, das zum *alignment* von orthologen und xenologen Regionen von zwei oder mehr Genomen dient, die sowohl lokalen als auch größeren Sequenzänderungen im Verlauf ihrer Entwicklung unterworfen worden sind. Bakterielle Rekombination hat als Nebeneffekt Sequenzänderungen, wobei eine Neuordnung oder Inversion von Sequenzfragmenten stattfinden kann, die ortholog zu Sequenzen anderer Stämme sind. *Mauve* erkannte konservierte Segmente, die in sich frei von dieser Neuordnung sind. Diese wurden als *locally collinear blocks* (LCBs) in einer je Block spezifischen Farbe markiert. Dadurch ermöglichte *Mauve* die Identifizierung von *rearrangement*-Mustern. Das Ergebnis des *progressive Mauve Alignment* war in zwei unterschiedlichen Formen verfügbar. Einerseits gab es das LCB-Color-Schema, in dem die von *Mauve* identifizierten LCBs, die zueinander homolog und frei von *rearrangement* waren, als Block mit eigener Farbe markiert waren. Das *Backbone-Color*-Schema war die andere Darstellungsmöglichkeit der Ergebnisse des *progressive Mauve Alignment* und wurde im Rahmen dieser Arbeit ausgewertet. Die zentrale, schwarze Linie in den Abbildungen stellte die Sequenz dar. Das *backbone* entspricht den Regionen, die in allen einbezogenen Sequenzen konserviert vorliegen. Die entsprechenden Regionen wurden in einem Malven-Farbtönen gefärbt. Im *Backbone-Color*-Schema wurde jeder weitere Sequenzblock, der konserviert in mehr als einer der verwendeten Sequenzen vorlag, in einem für diesen Sequenzblock spezifischem Farbtönen markiert. Lagen die Blöcke oberhalb der Se-

quenz, waren sie in *forward*-Richtung kodiert. Lagen sie unterhalb der Sequenz, lagen sie in inverser Form auf der Sequenz vor. Regionen, die außerhalb der Blöcke lagen, waren frei von Homologie zu den verwendeten Sequenzen. Die Höhe der farbigen Ausfüllung gibt das durchschnittliche Level der Konservierung des entsprechenden Sequenzblocks wieder. Regionen innerhalb eines Blocks mit komplett fehlender farblicher Ausfüllung stehen für Abschnitte, die spezifisch für die entsprechende Sequenz sind. Die Randbereiche der farbigen Blöcke stehen für Bruchstellen des genomischen *rearrangements* (Darling et al., 2004).

### 3.12 Statistische Auswertung

Die statistische Auswertung erfolgte anhand des Statistikprogramms RStudio (*RStudio, Inc.*) und dem Tabellenkalkulationsprogramm Microsoft Excel (*Microsoft Corporation*). Im Rahmen der deskriptiven Auswertung wurde die statistische Signifikanz mittels eines exakten Tests nach Fisher berechnet. Als statistisches Signifikanzniveau wurde  $\alpha \leq 0,05$  gewählt.

## 4. Ergebnisse

### 4.1 Erstellung eines *core*-Genoms

Für diese Arbeit wurde auf insgesamt 95 eigen sequenzierte Genome zurückgegriffen (Tab. 8.1). Von den sequenzierten Isolaten dieser Arbeit wurden 70 Isolate als potentiell pathogen und 25 Isolate als Umweltstämme eingestuft. Zusätzlich wurden für die weiteren Sequenzanalysen 84 öffentlich zugängliche Genome (*public draft genomes*) verwendet. Diese 84 Gensätze stellen eine Auswahl von *S. epidermidis*-Genomen dar, die zu diesem Zeitpunkt auf NCBI verfügbar waren.

Im Rahmen dieser Arbeit wurde mit Hilfe von SeqSphere durch Verwendung der Genome RP62A, ATCC 12228 sowie PM221 ein *core*-Genom aus 1946 Zielsequenzen (*targets*) und ein akzessorisches Genom aus 473 Genloci gebildet. Die Identifizierung der Genloci des zuvor definierten *core*-Genoms in den Sequenzen der assemblierten Rohdaten erfolgte nach den Standardvoreinstellungen von SeqSphere. Zu den abgeglichenen Sequenzen gehörten auch die des durch die Arbeitsgruppe neu sequenzierten und assemblierten RP62A-Stammes. Es konnten 99.79% der Zielgenloci des *core*-Genoms (1942 Zielgene) als homolog zu den Sequenzen dieses Referenzgenoms zugeordnet werden. Damit gilt das erstellte *core*-Genom als valide, da entsprechend der Herstellervorgaben ein valides *core*-Genom in mindestens 98,5% der Zielloci Homologie zu dem erneut sequenzierten Referenzgenom aufweisen muss.

### 4.2 Charakterisierung des Sequenztyps 5 von *S. epidermidis*

#### 4.2.1 Identifizierung der Markersequenz

Mit Hilfe einer subtrahierenden Sequenzanalyse (*mapping*) gegen das ST10-Referenzgenom RP62A sollten ST5 spezifische Sequenzen identifiziert werden.

Die nach der Subtraktion verbleibenden Sequenzteile (*paired unmapped reads*) der ST5-Stämme wurde neu assembliert (*de-novo-assembly*). Zum Zeitpunkt der Analyse erfüllten 64 Sequenzen die getroffene Definition für den folgenden Schritt. Mit der Suchfunktion für ORFs konnte aus einer Auswahl dieser Sequenzen eine Abfolge aus zwei ORFs (Tab. 8.3) identifiziert werden, die ausschließlich Homologien zu den Sequenzen von ST5-Isolaten besitzen.

Auf dem hochqualitativ sequenzierten ST5-Stamm Bt1p3 (*average coverage*: 44.827 kb, N50: 95.772 kb) konnte diese ORF-Abfolge auf einem *de-novo-assemblierten* Sequenzabschnitt identifiziert werden (*contig 11*). Dieser Sequenzabschnitt (*contig 11*) wurde auf Homologien zu den Sequenzen aller verfügbaren, eigen sequenzierten Stämme untersucht. Eine finale Markersequenz konnte entsprechend den definierten Bedingungen mit einer Länge von 3.161 bp extrahiert werden. Die Homologie zwischen dem Marker und sowohl den eigenen als auch den öffentlich annotierten Genomen (*public draft genomes*) wurde mit Hilfe der Workbench bestimmt (BLAST). Die Bedingung für die Definition "markerpositiv" erfüllten 63% der ST5-Stämme (Tab. 4.1). Ebenfalls markerpositiv waren 0,6% der Isolate mit abweichendem Sequenztyp. In der Gruppe aller verwendeten Stämme konnte der Marker signifikant häufiger ( $p \leq 0,01$ ) in den ST5-Stämmen als in den Stämmen mit anderem Sequenztyp identifiziert werden. In der Gruppe der im Rahmen dieser Arbeit sequenzierten Stämme konnten 71% der pathogenen ST5-Isolate als markerpositiv ermittelt werden. Keines der 63 pathogenen Isolate mit anderem Sequenztyp ist nach der getroffenen Definition markerpositiv. Folglich tritt der Marker signifikant häufiger ( $p \leq 0,01$ ) in ST5-Stämmen auf als in den anderen Stämmen innerhalb der Gruppe der pathogenen Isolate. In der Gruppe der Umweltstämme waren 50% der ST5-Stämme dieser Arbeitsgruppe markerpositiv. Es konnte in keinem der Umwelt-Isolate mit abweichendem ST der Marker nachgewiesen werden. Der ST5-Marker tritt nicht signifikant häufiger in Umwelt-Stämmen mit ST5 als in denen mit abweichendem Sequenztyp auf. Von den bereits annotierten ST5-Genomen wurden 67% als markerpositiv identifiziert. Einem Stamm der öffentlich zugänglichen Genome mit anderem Sequenztyp als ST5 konnte der Marker zugeordnet werden. Der Marker tritt statistisch signifikant häufiger ( $p \leq 0,01$ ) in be-

reits annotierten Genomen mit ST5 auf als in jenen mit einem abweichenden Sequenztyp.

Tabelle 4.1: Anzahl der ST5-markerpositiven Stämme in Relation zur Gesamtzahl der Stämme der jeweiligen Kategorie

|         | Eigene klinische Isolate | Eigene Umwelt-Isolate | Eigene Mastitis-Isolate | Öffentliche vorläufige Genome ( <i>public draft genomes</i> ) | Gesamtergebnis |
|---------|--------------------------|-----------------------|-------------------------|---------------------------------------------------------------|----------------|
| ST5     | 5/7**                    | 3/6                   |                         | 4/6**                                                         | 12/19**        |
| Non-ST5 | 0/63**                   | 0/7                   | 0/12                    | 1/78**                                                        | 1/160**        |

\*\* statistisch hochsignifikant ( $p \leq 0,01$ )

Die Korrelation von Pathogenität und Markernachweis wurde in unterschiedlichen Gruppen unabhängig vom Sequenztyp untersucht, die sich wie folgt zusammensetzen: Die Gruppe aller pathogenen *S. epidermidis* wurde gebildet aus den 70 eigen sequenzierten und elf bereits annotierten Isolaten. Für die bereits annotierten Isolate gilt die Annahme, dass Stämme aus Blutkulturen oder von Katheterspitzen als pathogen einzustufen sind. Die im Rahmen dieser Arbeit sequenzierten 13 Umwelt-Stämme, zwölf Kuhmastitis-Isolate und 41 Umwelt-Isolate anderer Arbeitsgruppen bildeten zusammen die Gruppe aller verwendeten *S. epidermidis* der Umwelt. Von den bereits annotierten Genomen mussten 32 der 84 Stämme für die Korrelation zwischen Pathogenität und Markernachweis ausgeschlossen werden. Die mögliche Pathogenität dieser Isolate ließ sich nicht eindeutig anhand der Informationen von NCBI klären. Die Gruppe aus eigen sequenzierten Stämmen besteht aus 70 pathogenen *S. epidermidis* und elf Stämmen der klinischen Umwelt. Die beschriebene Zusammensetzung der einzelnen Gruppen ist identisch für die folgenden Korrelationen von Pathogenität und Markernachweis der anderen Sequenztypen.

Innerhalb der Gruppe aller verwendeten Isolate konnte der ST5-Marker nicht signifikant häufiger in den pathogenen Stämmen als in den Umwelt-Stämmen nachgewiesen werden (Tab. 4.2). In der Gruppe der eigen sequenzierten, pathogenen und Krankenhausumwelt-Isolate konnte unabhängig vom Sequenztyp

keine signifikante Verteilung der markerpositiven Stämme festgestellt werden (Tab. 4.3).

Tabelle 4.2: Nachweis des ST5-Markers innerhalb der Gruppe aller verwendeten Isolate

|               | Pathogene Isolate | Umwelt-Isolate |
|---------------|-------------------|----------------|
| Markerpositiv | 8                 | 3              |
| Markernegativ | 73                | 63             |

Tabelle 4.3: Nachweis des ST5-Markers in pathogenen und Krankenhausumwelt-Isolaten der Stammsammlung der Arbeitsgruppe

|               | Pathogene Isolate | Klinikumwelt-Isolate |
|---------------|-------------------|----------------------|
| Markerpositiv | 5                 | 2                    |
| Markernegativ | 65                | 9                    |

#### 4.2.2 Klonale Verwandtschaft in Relation zum Auftreten des Markers

Die Analyse der klonalen Verwandtschaft ergab, dass die verwendeten Isolate mehrheitlich einem gemeinsamen Cluster ( $n=17$ ) zugehörig sind. Zwei Stämme liegen mit einer großen genetischen Varianz der Genloci des *core*-Genoms außerhalb dieses Clusters. Beide Stämme sind von der Arbeitsgruppe sequenzierte markernegative Umwelt-Isolate. Alle weiteren Umwelt-Stämme fügen sich in das Cluster ein. Ein markerpositives Umwelt-Isolat bildet das Zentrum des MST. Dieses Isolat ist Teil des Komplex 2 ( $n=3$ , rötlich), das eines der drei identifizierten Komplexe darstellt. Eine minimale genetische Distanz von 68 Genloci trennt den zentralen Stamm des Komplex 2 von Komplex 1 ( $n=2$ , bläulich). Die beiden Stämme des Komplex 1 trennt ein Abstand von einem Locus. Ein weiteres markerpositives Isolat liegt als Bindeglied zwischen dem zentralen Stamm von Komplex 2 und dem Komplex 3 ( $n=3$ , türkis). Dieses Bindeglied trennt eine Distanz von 79 Genloci vom Komplex 2 und eine Distanz von 54 Genloci vom Komplex 3. Dieser dritte Komplex befindet sich weiter in der Peripherie des Clusters. Ohne Berücksichtigung der drei Komplexe streut die minimale genetische Distanz innerhalb des Clusters zwischen 54 und 100 Genloci. Im Zen-

trum des MST finden sich mehrheitlich markerpositive Stämme. Die markernegativen Isolate liegen mit einer minimalen genetischen Distanz von mindestens 54 Loci zu einem markerpositiven Stamm in der Peripherie. Im Komplex 3 findet sich interessanterweise ein markerpositiver Stamm als Bindeglied zwischen zwei markernegativen Stämmen. Ebenso lässt sich in der Peripherie des Clusters ein markernegativer Stamm finden, der die Verknüpfung zwischen zwei markerpositiven Stämmen darstellt. Diese beiden auffälligen Stämme mit abweichendem Markermuster sind Isolate, die von anderen Arbeitsgruppen sequenziert wurden. Zum Ausschluss einer schlechten Sequenzierqualität wurden die veröffentlichten N50-Werte recherchiert. Der auffällige, markertragende Stamm (NIH06004) des Komplex 3 besitzt einen N50-Wert von 88.654 kb. Das auffällige, markernegative Isolat in der Clusterperipherie (M23864\_W2\_grey), besitzt den N50-Wert 92.830 kb. Eine schlechte Qualität der Sequenzierung kann aufgrund dieser Werte ausgeschlossen werden.

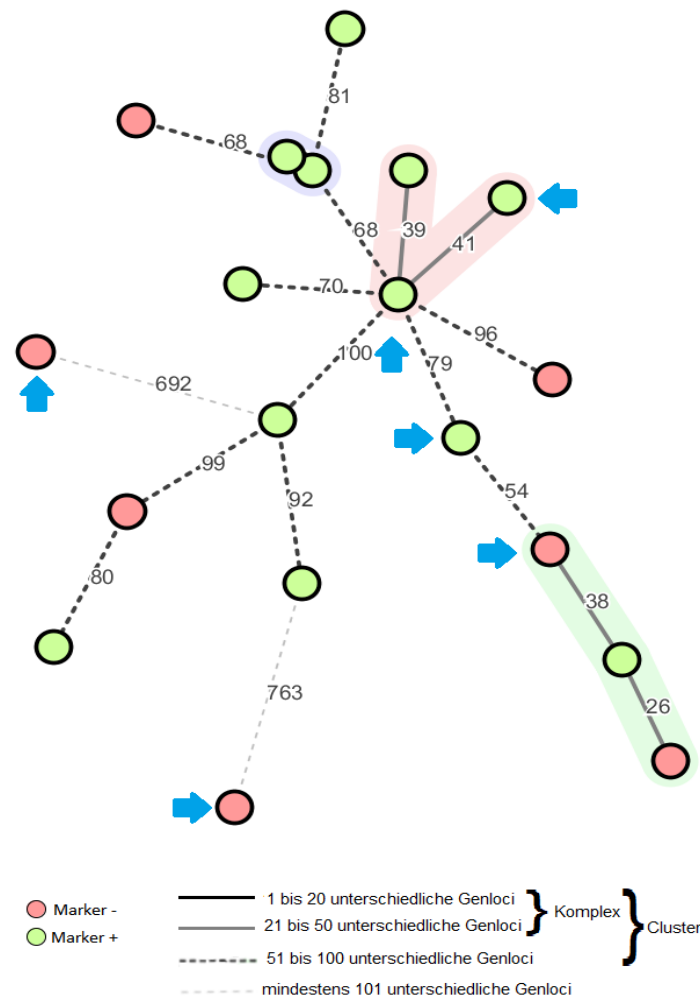


Abbildung 4.1 *Minimum spanning tree* der verwendeten ST5-Isolate

Die Stämme sind entsprechend ihres Markerauftritts markiert. Grün markierte Punkte stehen für die markerpositiven, rot markierte für die marker negativen Stämme. Linien verbinden die Stämme mit den geringsten Unterschieden zueinander, bezogen auf die Genloci des *core*-Genoms. Die Zahl neben der Verknüpfung gibt die differierenden Genorte wieder. Ist keine Zahl erkennbar, differieren die Isolate höchstens in zwei Genloci. Stämme mit Unterschieden in 100 oder weniger Genloci bilden ein Cluster. Ein Cluster wurde identifiziert. Isolate mit Differenzen in 50 oder weniger Genloci wurden als farblich hervorgehobener Komplex zusammengefasst. Der Komplex 1 (bläulich), Komplex 2 (rötlich) und Komplex 3 (türkis) wurden identifiziert. Eigen sequenzierten Umwelt-Isolate wurden mit einem blauen Pfeil markiert.

#### 4.2.3 Aufbau des Markers

Der ST5-Marker besitzt eine Länge von 3.161 bp und verfügt über zwei ORFs. Die beiden ORFs wurden durch die Anwendung *megablast* der Datenbank des NCBI auf homologe, bereits annotierte Proteine untersucht. Der erste ORF hat

eine Länge von 738 bp. Es besteht eine Homologie der letzten 719 bp zu einem 719 bp langen potentiellen Protein unbekannter Funktion des Stammes JCSC1435 von *S. haemolyticus* (Accession number: AP006716.1). Der zweite, 857 bp lange ORF besitzt eine Teilhomologie zu dem 896 bp umfassenden, potentiellen Proteinprodukt mit ungeklärter Funktion des *S. haemolyticus* JCSC1435. Die Homologie beider ORFs besitzt eine Identität (*identity*) von 99% und eine Abdeckung der Referenzsequenzen (*query cover*) von 100%. Interessanterweise ist die Leserichtung beider ORFs auf dem Marker gegensätzlich. Im Genom des *S. haemolyticus* JCSC1435 liegen beide ORFs hingegen mit gleicher Leserichtung und benachbart vor.

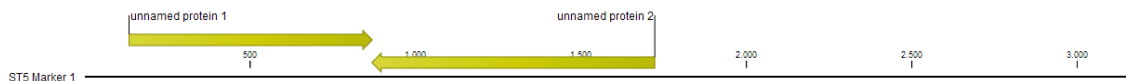


Abbildung 4.2 Genkarte des ST5-Markers

Dargestellt sind die beiden ORFs (grüne Pfeile), die jeweils homolog zu einem potentiellen Protein unbekannter Funktion (*unnamed protein*) des annotierten Stammes JCSC1435 von *S. haemolyticus* sind.

Weiterhin wurde die *megablast*-Funktion auf die komplette Markersequenz angewandt. Dadurch konnten 60% (*query cover*) dieser Sequenz mit einer Identität (*identity*) von 99% als homolog zu dem Genom des Stammes JCSC1435 von *S. haemolyticus* JCSC1435 zugeordnet werden.

#### 4.2.4 Identifizierung der Varianz des markerpositiven Sequenzabschnitts

In dem Sequenzabschnitt des assemblierten Stammes Bt1p3 (*contig* 11) liegt die Markersequenz mittig vor. Entsprechend der zuvor angegebenen Basenpaarlängen wurde die markertragende Sequenz für das *alignment* extrahiert.

Nach einem Vergleich dieser Sequenz mit denen weiterer ST5-Isolaten (BLAST) wurde eine Auswahl von sieben Referenzsequenzen aus den Gensätzen markernegativer Isolate für das *alignment* entnommen. Im Schema des *alignment* inseriert der Marker auf dem markerpositiven Sequenzabschnitt zwischen Basenpaar 1.500 und 4.661. Auffällig ist, dass *upstream* des Markers eine 200 bp lange Sequenz ohne Homologie zu den Referenzsequenzen ermittelt wurde.



Abbildung 4.3 Das Ergebnis des *alignment* eines Ausschnitts der ST5-marker-tragenden Sequenz sowie sieben markernegativer Referenzsequenzen. Der markerpositive Abschnitt ist die Bezugssequenz über den Referenzen. Der Marker liegt auf dem markerpositiven Sequenzabschnitt zwischen Basenpaar 1.500 und 4.661. Die markernegativen Referenzsequenzen A bis G folgen. Homologe Abschnitte zwischen den einzelnen Sequenzen sind als Block mit einem spezifischen Farbton markiert. Je höher der Grad an Konservierung des jeweiligen Blocks, desto größer ist der Block. Zum besseren Verständnis wurden diese Sequenzblöcke nummeriert. Block 2, der in allen Sequenzen konserviert vorliegt, ist malvenfarben. Abschnitte, die nur in einer Sequenz vorkommen, besitzen keinen farblichen Block.

Alle verwendeten Sequenzen teilen sich als *backbone* den kurzen, konservierten Abschnitt 2. Die markernegativen Sequenzen besitzen ein ähnliches Muster. Auffällig ist, dass sich bestimmte Blockkombinationen in gleicher Orientierung als längere Fragmente zwischen den einzelnen Sequenzen verschieben. Dies ist bedingt durch Translokationen. Die Translokationsbruchstellen liegen zwischen den ermittelten Fragmenten. Das erste Fragment besteht aus den Blöcken 1 bis 3. Das zweite Fragment aus den Blöcken 9 bis 11. Das dritte Fragment aus den Blöcken 8 und 12. Die Sequenzblöcke 4 bis 7 bilden das vierte Fragment. Die Blöcke 4 bis 7 treten in Abhängigkeit von der Referenz variabel auf. Sie behalten dabei ihre chronologische Abfolge: Falls ein Sequenzblock fehlt, sind die weiteren Blöcke mit einer niedrigeren Nummer ebenfalls nicht vorhanden. Die markerpositive Sequenz besteht aus der Abfolge des ersten, des vierten und dem Block 8 des dritten Fragments. Anhand der Fragmentabfolgen lassen sich die Referenzsequenzen zusammenfassen. Die Referenzen A, B und E bilden gemeinsam eine Gruppe. Sie besitzen die Abfolge aus dem zweiten, dem ersten, dem vierten und dem dritten Fragment. Referenzsequenz B und E besitzen vom vierten Fragment nur Block 7. Referenz A beinhaltet die Blöcke 6 bis 7 des vierten Fragments. Sequenz C sowie D können zu einer zweiten Gruppe zusammengefasst werden. Sie besitzen die Abfolge aus dem vierten, dem dritten, dem zweiten und dem ersten Fragment in der genannten Reihenfolge. Referenzsequenz D besitzt Sequenzblock 7 des vierten Fragments. Sequenz C hingegen die Blöcke 5 bis 7. Referenz G kann keiner der anderen Gruppe zugeordnet werden. Sie setzt sich zusammen aus dem dritten, dem zweiten und dem ersten Fragment. Es ist die einzige Referenz ohne Anteile des vierten Fragments. Innerhalb des Blocks 12 liegt auf dieser Sequenz eine stammspezifische Insertion vor. In den Sequenzen A, B, C, D, E sowie G lassen sich demnach die Blöcke 1 und 3 mit einem hohen Konservierungsgrad in gleicher Ausrichtung ermitteln. Referenzsequenz F lässt sich ebenfalls keiner Gruppe zuteilen. Diese Sequenz besitzt die Blöcke 4 bis 7 des vierten Fragments am Sequenzanfang. Im Gegensatz zu den anderen Sequenzen liegen auf dieser Referenz die Blöcke 10 und 2 eingebettet in einem Abschnitt vor, der keine Homologie zu den anderen Sequenzen besitzt. Aus diesem Grund wurde die Sequenzierqualität des Referenzsequenz F zugehörigen Stammes (M06) überprüft. Der N50-Wert des Isolats beträgt 81.780 kb. Eine

schlechte Qualität der Sequenzierung kann demnach ausgeschlossen werden. Durch das *alignment* konnte gezeigt werden, dass der ST5-Marker in eine Sequenz mit einer hohen genetischen Variabilität inseriert.

### 4.3 Charakterisierung des Sequenztyps 59 von *S. epidermidis*

#### 4.3.1 Identifizierung der Markersequenz

Nach der subtrahierenden Sequenzanalyse (*mapping*) gegen das Referenzgenom RP62A wurden die verbliebenen, kurzen Sequenzen (*paired unmapped reads*) der ST59-Stämme anhand von Sequenzüberlappung zu längeren Fragmenten zusammengefasst (*de-novo-assembly*). Es konnten zum Zeitpunkt der Untersuchung 24 Sequenzen identifiziert werden, die die Vorgaben für die weitere Suche erfüllten. Auf einer dieser Sequenzen konnten zwei einzelne ORFs sowie eine Abfolge aus zwei ORFs (Tab. 8.4) identifiziert werden, die ausschließlich homolog zu ST59-Isolaten waren.

Die ORFs wurden auf den Sequenzabschnitten der *de-novo*-assemblierten Rohdaten (*contigs*) des Stammes AEEG06, der über einen hohen Qualitätsgrad der Sequenzierung verfügt (N50-Wert: 129.954 kb; *average coverage*: 41.687 kb), auf einem Sequenzabschnitt (*contig* 5) identifiziert. Es wurden potentielle Markersequenzen entsprechend dem Vergleich der Homologien zwischen diesem Sequenzabschnitt (*contig* 5) und dem Sequenzdatensatz der eigenen Stammsammlung extrahiert. Nur eine Sequenz erfüllte die geforderten Bedingungen. Dieser ST59-Marker besitzt eine Länge von 858 bp. Für die Definition "markerpositiv" ausreichende Homologien konnten in 23% aller verfügbaren ST59-Genome identifiziert werden (Tab. 4.4). Ungeachtet der einzelnen untersuchten Subgruppen konnten keine Homologien zu Stämmen mit abweichendem Sequenztyp ermittelt werden. Hieraus ergibt sich, dass der Marker in allen verwendeten ST59-Isolaten signifikant häufiger ( $p \leq 0,01$ ) auftritt als in den anderen verwendeten Stämme mit abweichendem Sequenztyp. Die markerpositiven Stämme entsprechen der Hälfte der pathogenen ST59-Isolate, die im Rah-

men dieser Arbeit sequenziert wurden. Der Marker fehlt in den pathogenen *S. epidermidis* mit einem anderen Sequenztyp als ST59. In der Gruppe der pathogenen Stämme tritt der ST59-Marker signifikant häufiger ( $p < 0,01$ ) in den ST59-Stämmen auf als in Isolaten mit abweichendem Sequenztyp. Der Marker konnte weder in Umwelt-, noch in bovinen Mastitis-, und ebenfalls nicht in bereits annotierten Genomen unabhängig vom Sequenztyp identifiziert werden.

Tabelle 4.4: Anzahl der ST59-markerpositiven Stämme in Relation zur Gesamtzahl der Stämme der jeweiligen Kategorie

|          | Eigene klinische Isolate | Eigene Umwelt-Isolate | Eigene Mastitis-Isolate | <i>public draft genomes</i> | Gesamtergebnis |
|----------|--------------------------|-----------------------|-------------------------|-----------------------------|----------------|
| ST59     | 3/6**                    |                       | 0/4                     | 0/3                         | 3/13**         |
| Non-ST59 | 0/64**                   | 0/13                  | 0/8                     | 0/81                        | 0/166**        |

\*\* statistisch hochsignifikant ( $p \leq 0,01$ )

Im Rahmen der Korrelation von Markernachweis mit Pathogenität konnte in der Gruppe aller verwendeter Genome der ST59-Marker nicht signifikant häufiger in den pathogenen Stämmen als in den Umwelt-Stämmen nachgewiesen werden (Tab. 4.5). In der Gruppe der eigen sequenzierten Isolate tritt der ST59-Marker nicht signifikant häufiger in den pathogenen als in den Klinikumwelt-Stämmen auf (Tab. 4.6).

Tabelle 4.5: Nachweis des ST59-Markers innerhalb der Gruppe aller verwendeten Isolate

|               | Pathogene Isolate | Umwelt-Isolate |
|---------------|-------------------|----------------|
| Markerpositiv | 3                 | 0              |
| Markernegativ | 78                | 66             |

Tabelle 4.6: Nachweis des ST59-Markers in pathogenen und Krankenhausumwelt-Isolaten der Stammsammlung der Arbeitsgruppe

|               | Pathogene Isolate | Klinikumwelt-Isolate |
|---------------|-------------------|----------------------|
| Markerpositiv | 3                 | 0                    |
| Markernegativ | 67                | 11                   |

#### 4.3.2 Klonale Verwandtschaft in Relation zum Auftreten des Markers

Die Analyse der klonalen Verwandtschaft zeigte, dass die drei markerpositiven ST59-Isolate hochgradig verwandt sind. Die markernegativen Stämme streuen hingegen vermehrt in der Zahl differierender Genloci. Der MST besteht aus drei Clustern, die durch eine minimale genetische Distanz in 121 beziehungsweise 488 Loci getrennt sind. Die markerpositiven Stämme sind Teil des ersten Clusters (n=5). Sie bilden zusammen den Komplex 1 (n=3, rötlich). Eine genetische Distanz von einem beziehungsweise von sieben Loci differenziert die drei Stämme voneinander. Ebenfalls Teil des ersten Clusters sind zwei markernegative Stämme. Eine genetische Distanz von 94 Loci liegt zwischen dem Komplex 1 und dem nächsten markernegativen Isolat. Dieser Stamm unterscheidet sich in 92 Loci von dem anderen markernegativen Isolat des gleichen Clusters sowie in 121 Genloci von einem markernegativen Stamm des zweiten Clusters (n=6). Dieser Stamm des zweiten Clusters bildet mit zwei weiteren Isolaten den Komplex 2 (n=3, bläulich). Innerhalb des Komplex 2 ist die genetische Distanz von 47 beziehungsweise 50 Loci deutlich größer als im Komplex 1. Ausgehend von dem Komplex 2 verzweigt sich das zweite Cluster zu den drei weiteren Isolaten. Der erste Zweig enthält zwei weitere Stämme und endet ohne weitere Verknüpfungen. Der peripher gelegene Stamm dieses Zweigs ist ein pathogener ST59-Stamm. Der zweite Zweig zieht zu einem Stamm des zweiten Clusters. Dieser Stamm stellt mit einer minimalen genetischen Distanz von 488 Loci das Bindeglied zu dem dritten Cluster (n=2) dar. Das dritte Cluster besteht aus zwei markernegativen, pathogenen Stämmen, die eine genetische Distanz von 64 Loci trennt. Für die drei markernegativen, pathogenen Isolate (B8190, B4249\_E1, SKC09) wurde zum Vergleich der Qualität der Sequenzierung der jeweilige N50-Wert bestimmt. Die N50-Werte betragen 84.111 kb (B8190), 57.624 kb (B4249\_E1) sowie 65.050 kb (SKC09). Diese Werte zeigen eine hohe Qualität der Sequenzierung an.

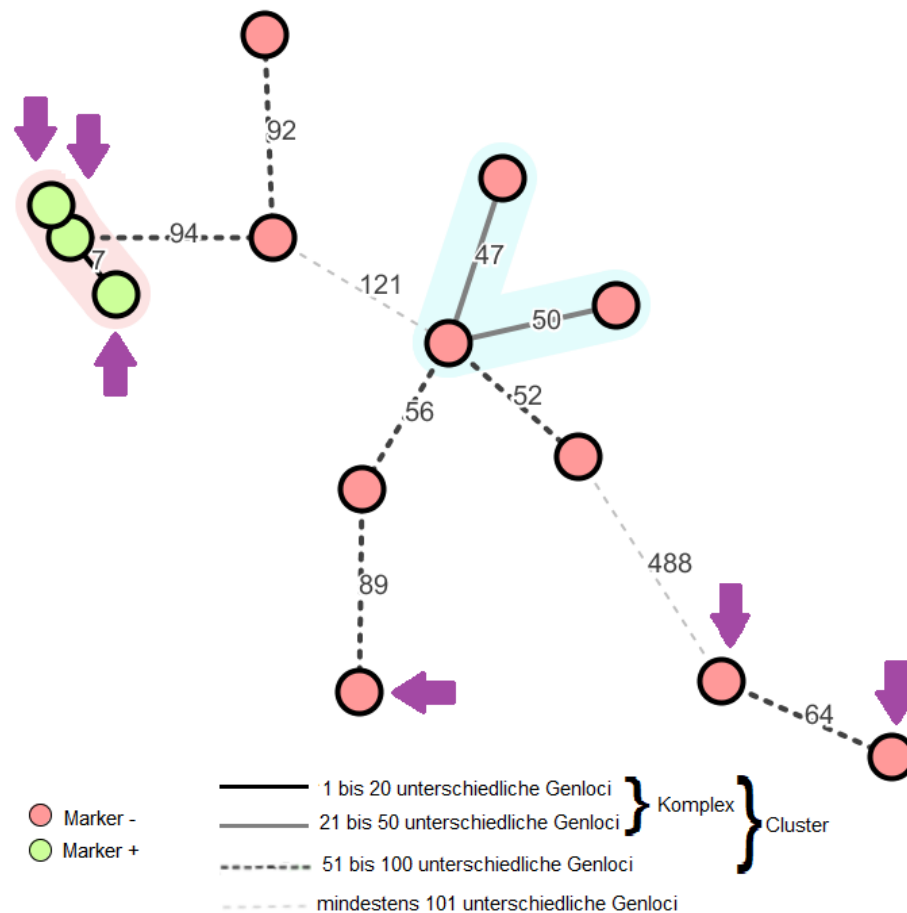


Abbildung 4.4 *Minimum spanning tree* der verwendeten ST59-Isolate

Die Stämme sind entsprechend ihres Markerauftritts markiert. Grün markierte Punkte stehen für die markerpositiven, rot markierte die marker negativen Stämme. Linien verbinden die Stämme mit den geringsten Unterschieden zueinander, bezogen auf die Genloci des *core*-Genoms. Die Zahl neben der Verknüpfung gibt die differierenden Genorte wieder. Ist keine Zahl erkennbar, differieren die Isolate höchstens in zwei Genloci. Stämme mit Unterschieden in 100 oder weniger Genloci bilden ein Cluster. Drei Cluster wurden identifiziert. Isolate mit Differenzen in 50 oder weniger Genloci wurden als farblich hervorgehobener Komplex zusammengefasst. Der Komplex 1 (rötlich) und der Komplex 2 (bläulich) wurden identifiziert. Die pathogenen Stämme wurden mit einem violetten Pfeil markiert.

#### 4.3.3 Aufbau des Markers

Dem 858 bp langen ST59-Marker konnte durch den *gene prediction*-Algorithmus der CLC Genomics Workbench keine ORFs zugeordnet werden. Ein *megablast* des gesamten Markers zur Identifizierung von Homologien zu annotierten Proteinen der Datenbank NCBI war nicht möglich. Aus diesem Grund wurde der *discontiguous megablast* von NCBI angewendet. Es konnte eine

Homologie mit einer Identität von 77% sowie einer Sequenzüberlappung von 99% zu dem Genom des Stammes *S. haemolyticus* JCSC1435 (*Accession number*: AP006716.1) identifiziert werden. Auf diesem Genom umfasst die Homologie zu dem Marker zwei proteinkodierende Sequenzen. Es konnte eine vollständige Homologie zu einem potentiellen Protein unbekannter Funktion sowie eine Teilhomologie von 448 bp zu einem weiteren, 455 bp langen, potentiellen Proteinprodukt ungeklärter Funktion zugeordnet werden.

#### **4.3.4 Identifizierung der Varianz des markerpositiven Sequenzabschnitts**

Der markerpositive Sequenzabschnitt (*contig* 5) des Isolats AEEG06 trägt den ST59-Marker mittig in der Sequenz. Aus diesem Sequenzabschnitt wurden der Marker sowie die flankierenden Sequenzstücke entsprechend der beschriebenen Methoden herausgearbeitet. Eine Auswahl aus acht homologen Sequenzen wurde getroffen, die die geforderten Bedingungen erfüllten. Ihr homologer Anteil wurde entsprechend dem Schema extrahiert. Die Markerregion befindet sich auf der gekürzten, markertragenden Sequenz zwischen dem Basenpaar 1.500 und dem Basenpaar 2.358. Die flankierende Sequenz besteht aus dem *backbone*, das sich aus den Sequenzblöcken 1, 2 sowie 4 zusammensetzt. *Upstream* in direkter Nachbarschaft zu dem Marker findet sich der kurze Sequenzblock 3 mit niedrigem Konservierungsgrad, der nicht in allen Sequenzen enthalten ist.

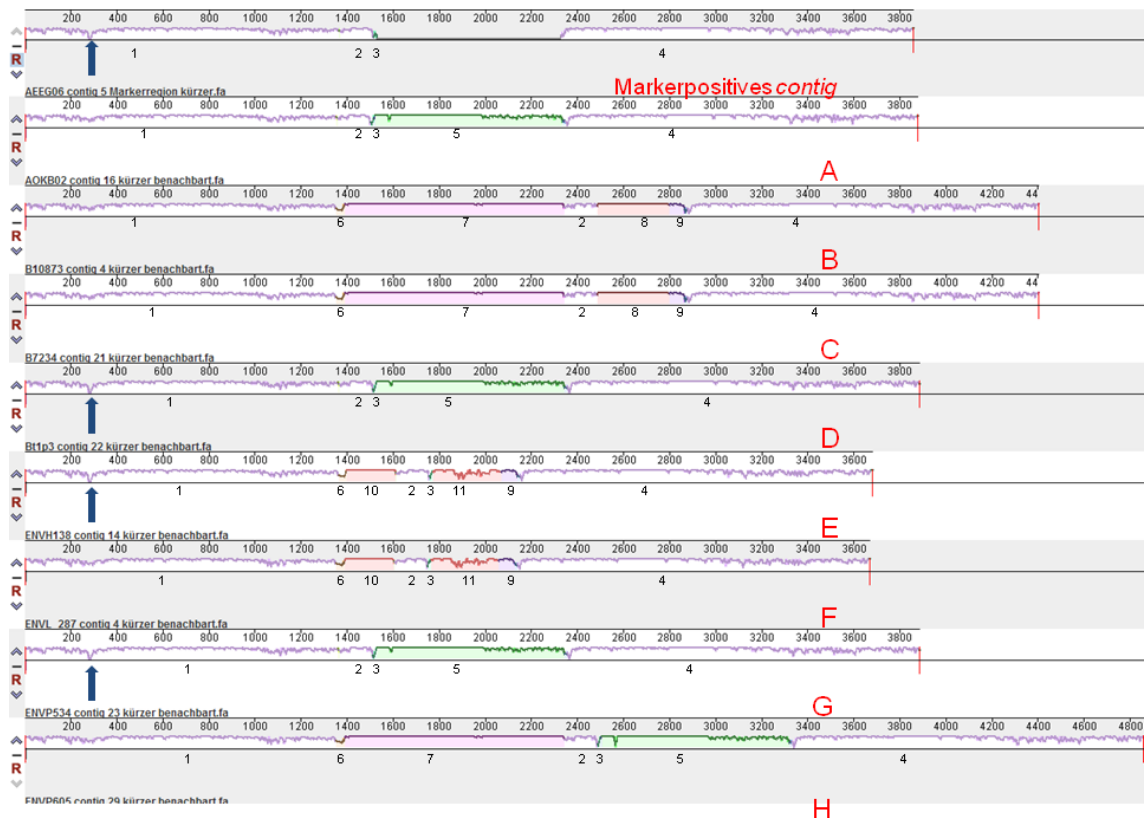


Abbildung 4.5 Das Ergebnis des *alignment* eines Ausschnitts der ST59-markertragenden Sequenz sowie acht markernegativer Referenzsequenzen

Der markerpositive Abschnitt ist die Bezugssequenz über den Referenzen. Der Marker liegt auf dem markerpositiven Abschnitt zwischen Basenpaar 1.500 und 2.358. Die markernegativen Referenzsequenzen A bis H folgen. Homologe Abschnitte zwischen den einzelnen Sequenzen sind als Block mit einem spezifischen Farbton markiert. Je höher der Grad an Konservierung des jeweiligen Blocks, desto größer ist der Block. Zum besseren Verständnis wurden diese Sequenzblöcke nummeriert. Block 1, 2 und 4, die in allen Sequenzen konserviert vorliegen, sind malvenfarben. Abschnitte, die nur in einer Sequenz vorkommen, besitzen keinen farblichen Block. Die Sequenzen D, E, G sowie die markertragende Sequenz besitzen jeweils einen Abschnitt ohne Homologie zueinander an Position des Basenpaares 278 (blauer Pfeil).

Die Sequenzblöcke 3 und 5 lassen sich mit einem hohen Konservierungsgrad in den Referenzsequenzen A, D sowie G finden. In diesen Sequenzen schließt sich der 829 bp lange Sequenzblock 5 *downstream* von Block 3 an. Der Sequenzblock 5 befindet sich an der Position, die auf der markerpositiven Sequenz durch den Marker eingenommen wird. In der Referenzsequenz H konnten neben dem *backbone* ebenfalls die konservierten Blöcke 3 und 5 identifiziert werden. Allerdings besitzt diese Referenzsequenz im *backbone* zusätzlich noch an der Position von Basenpaar 1350 eine Abfolge bestehend aus dem Block 6 sowie 7. Getrennt werden die Sequenzblöcke 7 und 3 durch den Block 2 des

*backbone*. Ein zu Referenzsequenz H ähnlichen Aufbau findet man in den Sequenzen B sowie C. Der Unterschied ist, dass beide Sequenzen *downstream* von Block 2 eine 392 bp lange Abfolge aus dem Sequenzblock 8 sowie 9 besitzen. Referenzsequenz E und F besitzen ebenfalls eine ähnliche Struktur zu der Sequenz H. Ein Unterschied ist, dass sich auf den Sequenzen E sowie F *downstream* von Block 6 der Sequenzblock 10 anschließt. An gleicher Position findet sich auf Referenz H der Block 7. Ein weiterer Unterschied der Sequenzen E sowie F besteht in dem Vorliegen der Blöcke 11 sowie 9 *downstream* von Block 3. Referenz H verfügt an dieser Position über den Block 5. In der markerpositiven Sequenz sowie den Referenzsequenzen D, E und G lässt sich an der Position von Basenpaar 276 innerhalb von Block 1 jeweils eine 9 bp lange Insertion ohne Homologie zueinander ermitteln.

Im Vergleich zum *alignment* von ST2-Marker 1 liegen größere Schwankungen des Konservierungsgrads des *backbone* vor. Die untersuchten Sequenzen konnten anhand ihrer Sequenzblöcke nicht in eine Reihenfolge gebracht werden, die eine möglicherweise stattgehabte Evolution darstellen.

## 4.4 Charakterisierung des Sequenztyps 2 von *S. epidermidis*

### 4.4.1 Identifizierung der Markersequenzen

Durch die subtrahierende Sequenzanalyse (*mapping*) der ST2-Stämme gegen das Referenzgenom RP62A verblieben Sequenzfragmente (*paired unmapped reads*). Diese Fragmente wurden zu längeren Sequenzen anhand von Überlappungen zusammengefasst (*de-novo-assembly*). Von den *de-novo*-assemblierten Sequenzfragmenten erfüllten 145 Sequenzen die vorgegebenen Definitionen für die weitere Markersuche. In dem Satz der 145 Sequenzen konnten drei ORF-Abfolgen auf drei unterschiedlichen Sequenzfragmenten entsprechend der beschriebenen Methoden identifiziert werden. Diese ORF-Abfolgen waren ausschließlich homolog zu den zu diesem Zeitpunkt sequenzierten ST2-Stämmen. Die erste Abfolge bestand aus 21 ORFs (Tab. 8.5). Diese Aneinanderreihung

von ORFs konnte einem Sequenzabschnitt (*contig* 68) des hochqualitativ sequenzierten Stamm B12142 (N50-Wert: 81.755 kb; *average coverage*: 22.350 kb) als homolog zugeordnet werden. Die zweite Abfolge bestand aus sieben ORFs (Tab. 8.6) und wurde im Gegensatz des qualitativ hochwertig sequenzierten Stammes B10873 (N50-Wert: 64.352 kb; *average coverage*: 27.424 kb) auf einem Sequenzabschnitt (*contig* 46) identifiziert. Die dritte Abfolge bestand aus zwei ORFs (Tab. 8.7) und konnte einem Sequenzabschnitt (*contig* 34) des Stammes B10873 zugeordnet werden. Nach dem beschriebenen Vergleich der Homologien zwischen dem jeweiligen Sequenzabschnitt (*contig*) und den Sequenzabschnitten der eigenen Stammsammlung, wurden die drei Markersequenzen entsprechend den Einschlusskriterien extrahiert.

Nicht alle ST2-Stämme tragen die Marker (Tab. 4.7). In der Gruppe aller verwendeten Isolate konnten 91% aller ST2-Isolate als positiv für die Markerkombination identifiziert werden. Als "positiv" für die Markerkombination wurden die Stämme gewertet, die mindestens einen der drei ST2-Marker besaßen. Bei den 61 markerpositiven ST2-Stämmen handelt es sich um vier bereits im Rahmen anderer Arbeiten sequenzierte Stämme, 51 klinische sowie sechs Umwelt-Stämme, die im Rahmen dieser Arbeit sequenziert wurden. Von den verwendeten Isolaten mit abweichendem Sequenztyp wurden 9% als positiv für die Markerkombination ermittelt. Die Kombination konnte signifikant häufiger in Stämmen mit ST2 als in solchen mit anderem Sequenztyp nachgewiesen werden.

In keinem der Kuhmastitis-assoziierten Isolate konnte einer der ST2-Marker nachgewiesen werden.

Tabelle 4.7: Anzahl der ST2-markerpositiven Stämme in Relation zur Gesamtzahl der Stämme der jeweiligen Kategorie

|         |          | Eigene klinische Isolate | Eigene Umwelt-Isolate | Eigene Mastitis-Isolate | <i>public draft genomes</i> | Gesamtergebnis | Markerkombination |
|---------|----------|--------------------------|-----------------------|-------------------------|-----------------------------|----------------|-------------------|
| ST2     | Marker 1 | 46/51**                  | 2/6                   |                         | 2/5                         | 50/62**        | 61/62**           |
|         | Marker 2 | 42/51**                  | 4/6                   |                         | 3/5*                        | 49/62**        |                   |
|         | Marker 3 | 45/51**                  | 2/6                   |                         | 1/5                         | 48/62**        |                   |
| Non-ST2 | Marker 1 | 1/19**                   | 1/7                   | 0/12                    | 6/79                        | 8/117**        | 11/117**          |
|         | Marker 2 | 1/19**                   | 1/7                   | 0/12                    | 7/79*                       | 9/117**        |                   |
|         | Marker 3 | 0/19**                   | 0/7                   | 0/12                    | 2/79                        | 2/117**        |                   |

\* statistisch signifikant ( $p \leq 0,05$ ), \*\* statistisch hochsignifikant ( $p \leq 0,01$ )

Wenn mindestens einer der drei Marker nachgewiesen wurde, galt ein Isolat als positiv für die Markerkombination

Der ST2-Marker 1 besitzt eine Länge von 34.798 bp. In der Gruppe aller verwendeten Isolate tritt der Marker 1 in 81% der Stämme mit ST2 und in 7% der Isolate mit abweichendem Sequenztyp auf. Innerhalb dieser Gruppe tritt der Marker 1 signifikant häufiger ( $p \leq 0,01$ ) in ST2-Isolaten auf als in Stämmen mit anderem Sequenztyp. In der Subgruppe der eigen sequenzierten, pathogenen Isolate sind 90% der ST2-Stämme positiv für den Marker 1. Ein Stamm (5%) der klinisch relevanten Stämme mit abweichendem ST (Stamm ACEF09) wurde als markerpositiv ermittelt. In der Gruppe der pathogenen Stämme konnte der ST2-Marker 1 signifikant häufiger ( $p \leq 0,01$ ) in ST2-Stämmen als in Isolaten mit abweichendem Sequenztyp nachgewiesen werden. In 33% der eigen sequenzierten ST2-Umwelt-Stämme sowie 40% der bereits annotierten ST2-Stämme konnte der Marker 1 nachgewiesen werden. In den Stämmen mit anderem Sequenztyp konnten 17% der sequenzierten Umwelt-Stämme sowie 8% der annotierten Genome als markerpositiv ermittelt werden. Weder innerhalb der Gruppe der Umwelt-Isolate noch in dem Satz von Genomen anderer Ar-

beitsgruppen konnte der Marker 1 signifikant häufiger in ST2-Isolaten als in Stämmen mit abweichendem Sequenztyp ermittelt werden.

Der ST2-Marker 2 weist 7.714 bp auf. In der Gruppe aller verwendeten Isolate konnte der Marker 2 in 79% der Stämme mit ST2 und in 8% der Isolate mit anderem Sequenztyp nachgewiesen werden. Innerhalb dieser Gruppe ist der Nachweis von Marker 2 signifikant häufiger in ST2-Isolaten als in Stämmen mit anderem Sequenztyp. In der Gruppe der eigen sequenzierten Stämme waren nach der getroffenen Definition 82% der klinisch relevanten ST2-Stämme positiv für den Marker 2. In 5% der pathogenen Stämme mit einem anderen Sequenztyp konnte der Marker ebenfalls nachgewiesen werden. Innerhalb der Gruppe der pathogenen Isolate tritt der Marker 2 signifikant häufiger ( $p \leq 0,01$ ) in ST2-Stämmen als in Isolaten mit abweichendem Sequenztyp auf. In der Gruppe eigen sequenzierter Umwelt-Stämme konnten 67% der Isolate mit ST2 sowie 16% der Stämme mit einem anderen Sequenztyp als markerpositiv ermittelt werden. Innerhalb der Gruppe der Umwelt-Stämme ist der Nachweis von Marker 2 nicht signifikant häufiger in ST2-Stämmen als in solchen mit anderem Sequenztyp möglich. Von den öffentlich zugänglichen ST2-Genomen erfüllten 60% die Definitionen für "markerpositiv". 9% dieser bereits annotierten Genome mit abweichendem Sequenztyp waren ebenfalls markerpositiv. Der Marker 2 tritt signifikant häufiger ( $p \leq 0,05$ ) in den bereits annotierten Isolaten mit ST2 auf als in solchen mit einem anderen Sequenztyp.

Der ST2-Marker 3 besteht aus 7.568 bp. Innerhalb der Gruppe aller verwendeten Stämme ist der Marker 3 in 77% der Stämme mit ST2 sowie in 2% der Stämme mit abweichendem Sequenztyp identifiziert worden. In dieser Gruppe konnte der Marker 3 signifikant häufiger ( $p \leq 0,01$ ) in ST2-Isolaten als in Stämmen mit abweichendem Sequenztyp ermittelt werden. Es konnten 90% der pathogenen ST2-Stämme als markerpositiv identifiziert werden. Der Marker 3 konnte nicht in einem pathogenen Stamm mit abweichendem Sequenztyp nachgewiesen werden. Innerhalb der Gruppe pathogener Stämme tritt der Marker signifikant häufiger ( $p \leq 0,01$ ) in ST2-Isolaten als in Stämmen mit anderem Sequenztyp auf. In der Gruppe der Umwelt-Isolate konnte in 33% der ST2-Stämme der Marker 3 nachgewiesen werden. Der Marker wurde nicht in einem

Umwelt-zugehörigen *S. epidermidis* mit anderem Sequenztyp nachgewiesen. Innerhalb der Gruppe der Umweltstämmen konnte der Marker 3 nicht signifikant häufiger in ST2-Isolaten als in Stämmen mit abweichendem Sequenztyp ermittelt werden. Der Marker 3 konnte in 20% der bereits annotierten ST2-Stämme identifiziert werden. 3% der bereits annotierten Genome mit anderem Sequenztyp besaßen ebenfalls den ST2-Marker 3. In der Gruppe der bereits annotierten Genome konnte der Marker 3 nicht signifikant häufiger in ST2-Stämmen als in solchen mit abweichendem Sequenztyp nachgewiesen werden.

Von den sequenzierten *S. epidermidis* mit dem ST2 (n=62) entstammen 34 Isolate dem gleichen Ausbruch von linezolidresistenten *S. epidermidis*.

Der ST2-Marker 1 konnte signifikant häufiger ( $p \leq 0,01$ ) in den pathogenen Stämmen aller verwendeten Genome als in den Umwelt-Stämmen nachgewiesen werden (Tab. 4.8). Innerhalb der Stammsammlung der Arbeitsgruppe tritt der Marker 1 signifikant ( $p \leq 0,05$ ) häufiger in den pathogenen *S. epidermidis* auf als in den Klinikumwelt-Isolaten (Tab. 4.9).

Tabelle 4.8: Nachweis des ST2-Markers 1 innerhalb der Gruppe aller verwendeten Isolate

|               | Pathogene Isolate | Umwelt-Isolate |
|---------------|-------------------|----------------|
| Markerpositiv | 49**              | 4**            |
| Markernegativ | 32**              | 62**           |

\*\* statistisch hochsignifikanter Unterschied ( $p \leq 0,01$ )

Tabelle 4.9: Nachweis des ST2-Markers 1 in pathogenen und Krankenhausumwelt-Isolaten der Stammsammlung der Arbeitsgruppe

|               | Pathogene Isolate | Klinikumwelt-Isolate |
|---------------|-------------------|----------------------|
| Markerpositiv | 47*               | 3*                   |
| Markernegativ | 23*               | 8*                   |

\* statistisch signifikant ( $p \leq 0,05$ )

Der ST2-Marker 2 wurde signifikant häufiger ( $p \leq 0,01$ ) in den pathogenen Stämmen aller verwendeten Genome als in den Umwelt-Stämmen nachgewiesen (Tab. 4.10). Der Marker 2 konnte nicht signifikant häufiger in den eigen se-

quenzierten, pathogenen Isolaten identifiziert werden als in den Krankenhausumwelt-Isolaten der Arbeitsgruppe (Tab. 4.11).

Tabelle 4.10: Nachweis des ST2-Markers 2 innerhalb der Gruppe aller verwendeten Isolate

|               | Pathogene Isolate | Umwelt-Isolate |
|---------------|-------------------|----------------|
| Markerpositiv | 47**              | 7**            |
| Markernegativ | 34**              | 59**           |

\*\* statistisch hochsignifikant ( $p \leq 0,01$ )

Tabelle 4.11: Nachweis des ST2-Markers 2 in pathogenen und Krankenhausumwelt-Isolaten der Stammsammlung der Arbeitsgruppe

|               | Pathogene Isolate | Klinikumwelt-Isolate |
|---------------|-------------------|----------------------|
| Markerpositiv | 43                | 4                    |
| Markernegativ | 27                | 7                    |

Der ST2-Marker 3 wurde signifikant häufiger ( $p \leq 0,01$ ) in den pathogenen als in den Umwelt-Stämmen nachgewiesen (Tab. 4.12). Der ST2-Marker 3 tritt signifikant ( $p \leq 0,01$ ) häufiger in den pathogenen *S. epidermidis* auf als in den Klinikumwelt-Stämmen (Tab. 4.13).

Tabelle 4.12: Nachweis des ST2-Markers 3 innerhalb der Gruppe aller verwendeten Isolate

|               | Pathogene Isolate | Umwelt-Isolate |
|---------------|-------------------|----------------|
| Markerpositiv | 45**              | 2**            |
| Markernegativ | 36**              | 64**           |

\*\* statistisch hochsignifikant ( $p \leq 0,01$ )

Tabelle 4.13: Nachweis des ST2-Markers 3 in pathogenen und Krankenhausumwelt-Isolaten der Stammsammlung der Arbeitsgruppe

|               | Pathogene Isolate | Klinikumwelt-Isolate |
|---------------|-------------------|----------------------|
| Markerpositiv | 45**              | 2**                  |
| Markernegativ | 25**              | 9**                  |

\*\* statistisch hochsignifikant ( $p \leq 0,01$ )

#### 4.4.2 Klonale Verwandtschaft in Relation zum Auftreten der Marker

Im Rahmen der klonalen Analyse zeigten sich im MST drei Cluster (Abb. 4.6). Jedem der im MST aufgeführten Stämme wurde die jeweilige Markerkombination zugeordnet. In zwei der drei Cluster tritt jeweils der ST2-Marker 2 als Markersequenz prädominant auf.

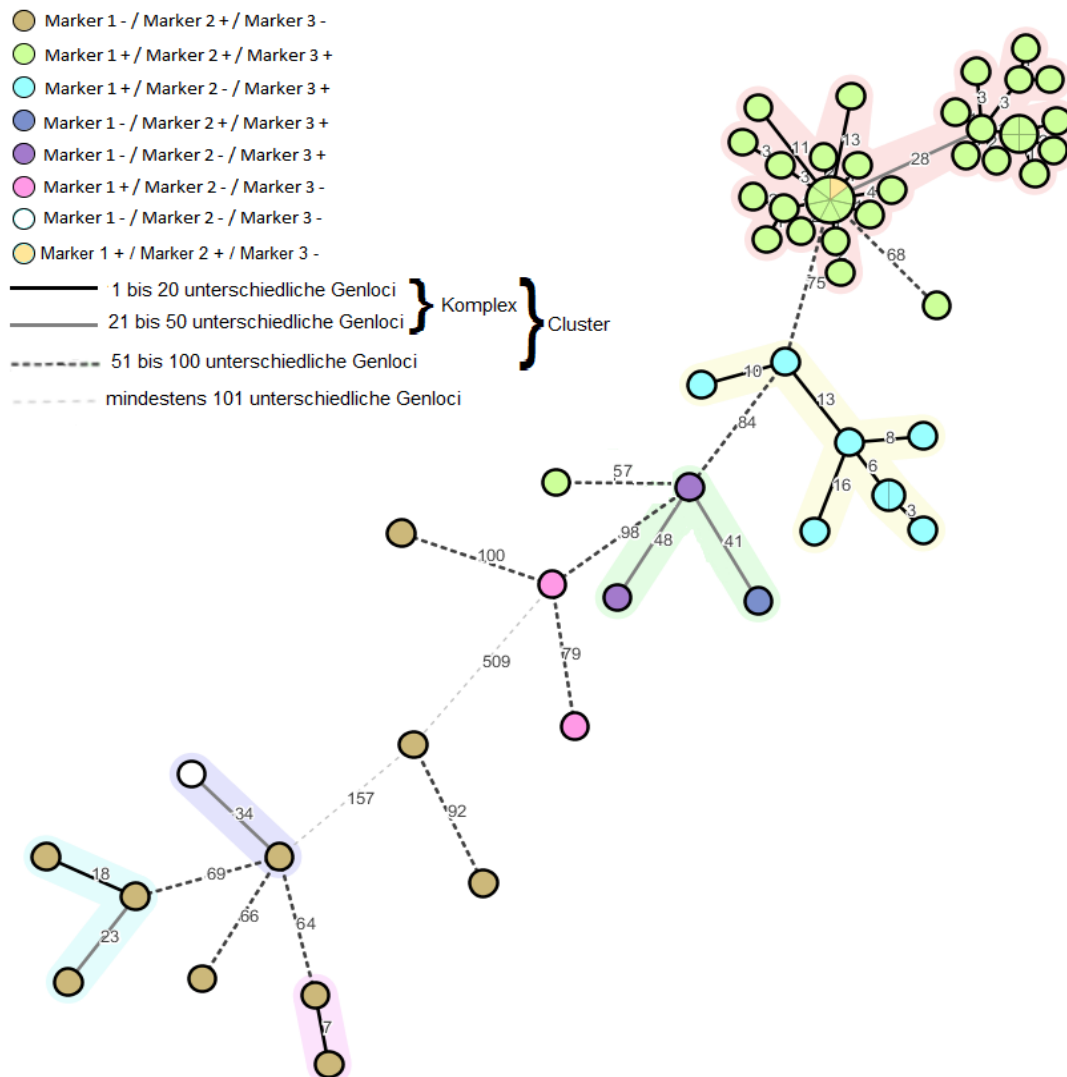


Abbildung 4.6 *Minimum spanning tree* der verwendeten ST2-Isolate

Die Stämme sind entsprechend ihres Markerauftretens markiert.

Unterscheidet sich eine Anzahl von Isolaten nicht in ihren Genloci, ist der entsprechende Punkt größer sowie unterteilt. Einzelnen Abschnitte des Punktes entsprechen je einem Stamm. Linien verbinden die Stämme mit den geringsten Unterschieden zueinander, bezogen auf die Genloci des core-Genoms. Die Zahl neben der Verknüpfung gibt die differierenden Genorte wieder. Ist keine Zahl erkennbar, differieren die Isolate höchstens in zwei Genloci. Stämme mit Unterschieden in 100 oder weniger Genloci bilden ein Cluster. Drei Cluster wurden identifiziert. Isolate mit Differenzen in 50 oder weniger Genloci bilden zusammen einen farblich hervorgehobenen Komplex. Komplex 1 (rötlich), Komplex 2 (gelblich), Komplex 3 (grünlich), Komplex 4 (violett), Komplex 5 (türkis) und Komplex 6 (magenta) wurden identifiziert.

Der größte Cluster ( $n = 52$ ) teilt sich auf in den Komplex 1 (rötlich), den Komplex 2 und den Komplex 3 sowie mehrere Isolate ohne Komplexzugehörigkeit. Mit Ausnahme von zwei Stämmen besitzen die drei Komplexe jeweils eine bestimmte Markerkombination. Der Komplex 1 ( $n=36$ ) besteht mit einer Ausnahme aus Stämmen, die über alle drei ST2-Marker verfügen. Im Komplex 1 streut die genetische Distanz der Stämme zueinander zwischen null und 28 Genloci. Zu den 36 Stämmen dieses Komplexes zählen die 34 *S. epidermidis*, die im Rahmen eines Ausbruchs von linezolidresistenten, koagulasenegativen Staphylokokken sequenziert wurden. Interessanterweise liegt im Zentrum des Komplexes ein Stamm, der nur über ST2-Marker 1 und 2 verfügt. Dieser Stamm (LRKNS009) zählt ebenfalls zu den oben genannten 34 linezolidresistenten Isolaten. Er unterscheidet sich in den Genloci des gebildeten *core*-Genoms nicht von sechs weiteren Isolaten, die die drei Marker tragen. Zum Ausschluss einer schlechten Sequenzierqualität wurde sein N50-Wert herangezogen. Dieser liegt bei 87.411 kb und schließt eine schlechte Qualität der Sequenzierung aus. Unterschiede in 68 Genloci trennen ein Isolat ohne Komplexzugehörigkeit, das positiv für die drei Marker ist, von dem Komplex 1. Eine Distanz von 75 Genloci liegt zwischen dem Komplex 1 und dem Komplex 2 (gelblich). Der Komplex 2 ( $n=8$ ) besteht aus Stämmen, die ST2-Marker 1- sowie 3-positiv sind. Die genetische Distanz innerhalb des Komplex 2 streut zwischen null bis 16 Loci. Zwischen dem Komplex 2 und dem Komplex 3 (grünlich) liegt eine genetische Distanz von 84. Der Komplex 3 ( $n=3$ ) setzt sich aus zwei Marker 3-positiven Isolaten und einem Marker 2- sowie 3-positiven Stamm zusammen. Zentral in diesem Komplex befindet sich einer der beiden Marker 3-positiven Stämme. Innerhalb des Komplex 3 wird das Isolat durch eine genetische Distanz von 48 Genloci von einem weiteren Stamm getrennt, der ebenfalls positiv für Marker 3 ist. Den dritten Stamm des Komplexes trennt eine Distanz von 41 Genloci von dem zentralen Isolat. Dieser dritte Stamm trägt die ST2-Marker 2 und 3. Den zentralen Stamm des Komplex 3 trennt eine genetische Distanz von 57 Loci von einem solitären vorliegenden Stamm. Er ist positiv für die drei ST2-Marker. Ein weiterer Zweig des zentralen Isolats des Komplex 3 bildet mit einer Distanz von 98 Genloci ein Stamm, der nur den Marker 1 besitzt. Von diesem Isolat ziehen zwei Verknüpfungen zu solitären Stämmen: Den Marker 1-positiven Stamm trennen Unterschiede in 100 Genloci von einem Marker 2-positiven

und Differenzen in 79 Loci von einem weiteren Marker 1-positiven Isolat. Eine genetische Distanz von 509 Genloci trennt den zentralen Marker 1-positiven Stamm der MST-Aufzweigung von einem Isolat des zweiten Cluster.

Das zweite Cluster (n=2) besteht aus Stämmen, die über den Marker 2 verfügen und die eine genetische Distanz von 92 Loci trennt. Unterschiede in 157 Genloci trennen den zweiten Cluster von einem Marker 2-positiven Stamm des Komplex 4 (violett) des dritten Clusters. Dieser Stamm ist der zentrale Ausgangspunkt für die Aufzweigung innerhalb des dritten Clusters (n=8). Eine genetische Distanz von 34 Genloci trennt diesen Stamm von dem zweiten, markerlosen Isolat (NIH05005) des Komplex 4 (n=2). Um eine schlechte Sequenzierqualität dieses Stammes auszuschließen, wurde auf den N50-Wert der Plattform NCBI zurückgegriffen. Der N50-Wert des markerlosen Stammes beträgt 112.534 kb. Die Sequenzierung ist damit als qualitativ hochwertig zu werten. Differenzen in 69 Loci unterscheiden den zentralen Stamm des dritten Clusters von einem Stamm des Komplex 5 (türkis). Es finden sich Unterschiede in 18 beziehungsweise 23 Genloci zwischen diesem Stamm und den anderen beiden Isolaten des Komplex 5 (n=3). Die drei Stämme des Komplex 5 tragen ebenfalls ausschließlich Marker 2. Der zentrale Stamm des dritten Clusters differiert in 66 Genloci von einem solitären, Marker 2-positiven Stamm und in 64 Loci von dem Komplex 6 (magenta). Dieser Komplex (n=2) besteht aus Marker 2-positiven Stämmen. Sie differieren in sieben Genloci voneinander.

#### 4.4.3 Aufbau des ST2-Markers 1

Der ST2-Marker 1 ist 34.798 bp lang und durch den *gene prediction*-Algorithmus der CLC Genomics Workbench konnten 19 ORFs identifiziert werden. Mit Ausnahme des ersten sowie dritten ORFs liegen alle weiteren ORFs in gleicher Orientierung auf der Markersequenz vor. Sie wurden aufgrund ihrer Nachbarschaft in Gruppen unterteilt. Den ORFs einer Gruppe konnten mittels des *megablast* der Datenbank von NCBI vollständige Homologien mit mindestens 99% Identität zu den bereits annotierten Proteinen des *S. epidermidis*-Stammes

ATCC 12228 (*Accession number*: AE015929.1) zugeordnet werden. Die erste Gruppe besteht aus zwei ORFs. ORF 1 ist eine Integrase. Dem zweiten ORF konnte Homologie zu einem hypothetischen Protein zugeordnet werden. Die Orientierung der beiden ORFs innerhalb der ersten Gruppe ist gegensätzlich. Die zweite Gruppe besteht aus den ORFs 3 bis 5, die jeweils ein hypothetisches Protein kodieren. Die Orientierung des ORF 3 ist gegensätzlich zu der Ausrichtung der anderen beiden ORFs. Die dritte Gruppe setzt sich aus ORF 6 bis 9 zusammen, deren Leserichtung identisch ist und die jeweils Homologie zu einem hypothetischen Protein besitzen. Die vierte Gruppe besteht aus ORF 10 bis 14, die ebenfalls benachbart mit der gleichen Orientierung auf der Sequenz vorliegen. ORF 10 sowie 14 sind homolog zu jeweils einem hypothetischen Protein. Der ORF 11 der vierten Gruppe ist eine putativen ATPase TraE. Der ORF 12 wurde als homolog zu einem *secretory antigen SsaA-like* Protein identifiziert. Der ORF 13 kodiert für ein Replikation-assoziiertes Protein. Die fünfte Gruppe besitzt ORF 15 bis 19. Der ORF 15 und ORF 17 zeigen Homologie zu je einem hypothetischen Protein. Der ORF 16 ist eine DNA-Topoisomerase III topB. Der ORF 18 kodiert für eine Sortase. Der ORF 19 zeigte Homologie zu einem LtrC-like Protein.

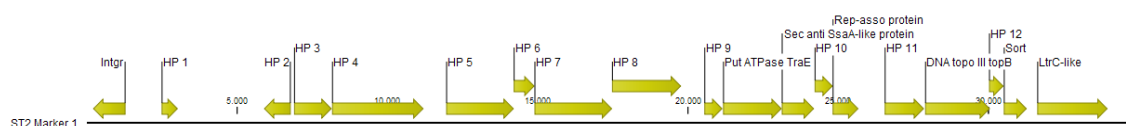


Abbildung 4.7 Genkarte des ST2-Markers 1

Dargestellt sind die 19 ORFs (grüne Pfeile). Die Abkürzungen am jeweiligen ORF geben das Protein der homologen Sequenz des *S. epidermidis*-Stammes ATCC 12228 an. Es konnten Homologie zu zwölf hypothetischen Proteinen (HP), einer Integrase (Intgr), einer putativen ATPase TraE (Put ATPaseTraE), einem *secretory antigen SsaA-like* Protein (Sec anti SsaA-like protein), einem *replication-associated* Protein (Rep-asso protein), einer DNA-Topoisomerase III topB (DNA topo III topB), einer Sortase (Sort) sowie einem LtrC-like Protein (LtrC-like) identifiziert werden.

Für die gesamte Markersequenz konnte mittels *megablast* der NCBI Datenbank eine Homologie mit einer Identität von 99% zu dem vollständig annotierten Genom von *S. epidermidis* ATCC12228 identifiziert werden.

#### **4.4.4 Identifizierung der Varianz des ST2-Marker-1-positiven Sequenzabschnitts**

Das Schema für die Vorbereitungen des *alignment* wich von dem bereits beschriebenen Schema ab. Bessere Ergebnisse wurden erzielt mit der Verwendung des ungekürzten markerpositive Sequenzabschnitt (*contig 40*) des Stammes B647, in den der Marker mittig inseriert. Zusätzlich wurden die 14 ungekürzten, homologen Sequenzabschnitte der markernegativen Stämme für das *alignment* extrahiert. Die vergleichende Analyse zeigte, dass der Marker in eine Sequenz mit einem hohen Grad an Konservierung inseriert.

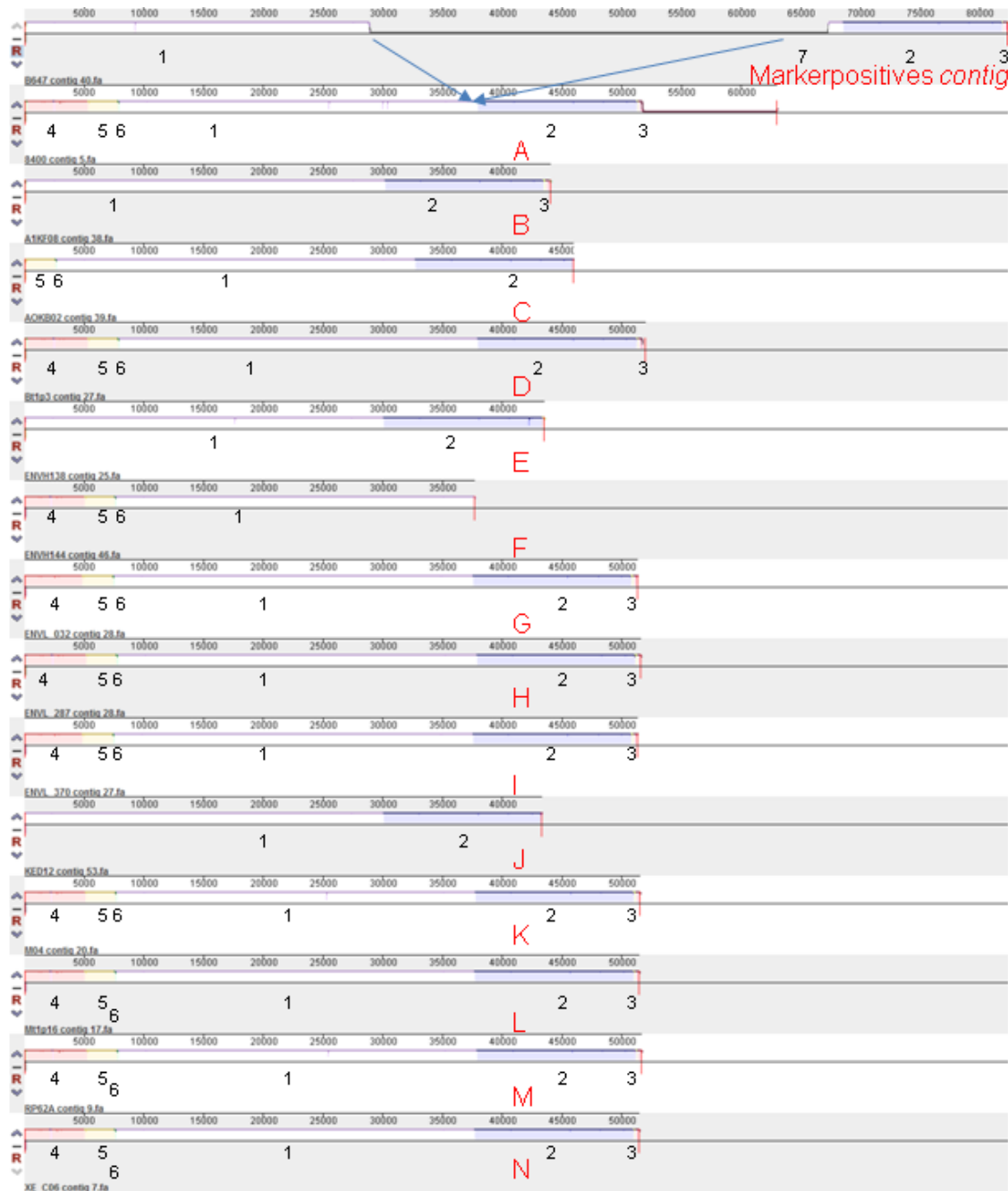


Abbildung 4.8 Das Ergebnis des *alignment* der ST2-Marker-1-positiven Sequenz sowie 14 markernegativer Referenzsequenzen

Der markerpositive Abschnitt ist die Bezugssequenz über den Referenzen. Der Marker liegt auf dem markerpositiven Sequenzabschnitt (*contig*) zwischen Basenpaar 28.877 und 63.674. Die markernegativen Referenzsequenzen A bis N folgen. Homologe Abschnitte zwischen den einzelnen Sequenzen sind als Block mit einem spezifischen Farbton markiert. Je höher der Grad an Konservierung des jeweiligen Blocks, desto größer ist der Block. Zum besseren Verständnis wurden diese Sequenzblöcke nummeriert. Block 1, der in allen Sequenz konserviert vorliegt, ist malvenfarben. Abschnitte, die nur in einer Sequenz vorkommen, besitzen keinen farblichen Block. Die Markerinsertionsstelle im Block 1 wurde anhand der markernegativen Referenz A durch zwei hellblaue Pfeile verdeutlicht.

In der markerpositiven Sequenz befindet sich der ST2-Marker 1 zwischen dem Basenpaar 28.877 und 63.674. Der Marker liegt als Insertion im Sequenzblock 1, der das *backbone* der eingespeisten Sequenzen darstellt. Interessanterweise befindet sich direkt *downstream* des Markers der 3.582 bp lange Block 7, der keine Homologie zu den Referenzen aufweist. Dieser Block 7 wurde auf Homologien zu den verwendeten Stämmen mit anderem Sequenztyp als ST2 untersucht (BLAST). Hierbei konnten Homologien zu 13 Isolaten mit abweichendem Sequenztyp identifiziert werden (Tab. 8.8). Dies sind Stämme mit Sequenzabschnitten (*contigs*), die mindestens zu 41% teilhomolog zu dem ST2-Marker 1 sind. Von diesen 13 identifizierten Stämmen zählen zehn zu den markerpositiven Isolaten. Die vollständige Sequenz des Marker 1 ist mehrheitlich nicht gekoppelt an den Sequenzblock 7. Beispielsweise im Genom des markerpositiven Stammes VCU041 liegen Marker sowie der Block 7 auf unterschiedlichen Sequenzabschnitten vor. Auffällig ist auch, dass die markernegativen Stämme (Sc119, CIM40, APO35) Sequenzen mit 17% oder weniger Teilhomologie zu Block 7 besitzen. Da der Block 7 kein Bestandteil des ursprünglich extrahierten Markers ist, wurde seine Lage in dem Gensatz des Stammes der Markerextraktion identifiziert. Hierfür wurden die Homologien des Block 7 zu dem Stamm der Markerextraktion (B12142) ermittelt (BLAST). Die Insertionsstelle des Block 7 befindet sich auf einem anderen Sequenzabschnitt (*contig* 3) als die des vollständigen Marker 1 (*contig* 68). *Upstream* des Block 7 befindet sich ein 156 bp langes Fragment des Markers 1.

*Downstream* des Sequenzblocks 1 befinden sich auf der markerpositiven Sequenz die Blöcke 2 sowie 3, die auch in einigen homologen Sequenzen der markernegativen Stämmen ermittelt wurden.

Die markernegativen Sequenzen lassen sich anhand des Musters ihrer Sequenzblöcke unterteilen. Die Referenzsequenzen E sowie J bilden eine Gruppe. Sie bestehen aus jeweils Sequenzblock 1 und 2. Referenzsequenz F kann keiner Gruppe zugeordnet werden, da sie nur über die Sequenzblöcke 4, 5, 6 sowie 1 verfügt. Sie ist die einzige, zu der markerpositiven Sequenz homologe, Referenz, die nicht über den Sequenzblock 2 verfügt. Sequenz B kann ebenfalls keiner Gruppe zugeordnet werden, da auf dieser Sequenz sich an die Se-

quenzblöcke 1 und 2 noch Block 3 anschließt. Sequenz C ist auch ohne Gruppenzugehörigkeit. Sie setzt sich aus den Blöcken 5, 6, 1 sowie 2 zusammen. Die größte Gruppe bilden die Referenzsequenzen A, D, G, H, I, K, L, M sowie N. Sie besitzen die Sequenzblöcke 4, 5, 6, 1, 2 sowie 3 in der genannten Reihenfolge. Sequenz A beinhaltet zusätzlich eine nicht farblich hervorgehobene Nukleinsäureabfolge ohne Homologie zu den anderen Referenzen, die sich *downstream* von Block 3 befindet.

Für die Sequenz *up-* sowie *downstream* des Markers konnte Homologie mit einer Identität von 99% und einer Sequenzüberlappung von 99% zu bereits annotierten Proteinen des *S. epidermidis*-Stammes ATCC 12228 auf der Plattform NCBI ermittelt werden (*megablast*). Die *upstream* liegende Sequenz ist homolog zu 14 konservierten hypothetischen Proteinen, einer Protoporphyrinogenoxidas, einer Ferrochelatase, einer Uroporphyrinogendecarboxylase, einem Signaltransduktionsprotein TRAP, einem mutmaßlichen Cadmiumresistenztransporter, einem ABC Transporter ecsB, einem ABC-Transporter-ecsA-ähnlichen Protein, einem Hit-ähnlichen Protein, dem Protein PrsA, einem hypothetischen Protein, einem CMP-bindenden Faktor 1, einem Zwei-Komponenten-Reaktions-Regulationssystem, einer Fumarathydratase der Klasse II, einer mutmaßlichen rRNA Methylase sowie einem Fragment einer Transposase. Der Block 7 *downstream* der Markerinsertionsstelle kodiert für zwei hypothetische Proteine, ein unvollkommenes Phagenresistenzgen sowie eine 180 bp lange Teilhomologie zu einem weiteren, 827 bp langen hypothetischen Protein. Für die hochkonservierte, *downstream* liegende Sequenz konnte Homologie zu einem Pflanzen-Stoffwechsel-Dehydrogenase-ähnlichem Protein, zehn hypothetischen Proteinen, einer S-Adenosylmethioninsynthetase, einer Phosphoenolpyruvatcarboxykinase, einer O-Succinylbenzolsäure(OSB)-Synthetase, einer O-Succinylbenzolsäure-CoA-Ligase, zwei Integrasen und einem Phage-Phi-PVL-Amidase-ähnlichen Protein festgestellt werden.

#### 4.4.5 Aufbau des ST2-Markers 2

Auf dem 7.714 bp langen ST2-Marker 2 konnten mit Hilfe der Workbench fünf ORFs identifiziert werden. Im Vergleich zu dem ST2-Marker 1 ist eine Einteilung in Gruppen schwieriger, da der Abstand zwischen den einzelnen ORFs deutlich größer ist. Durch die BLAST-Funktion der NCBI-Datenbank konnten für die einzelnen ORFs Homologien zu Schwermetallresistenz-assoziierten Proteinprodukten des *S. epidermidis*-Stammes NTUH-7684 (Accession number: LC085180.1) mit einer Referenzabdeckung (*query cover*) sowie einer Identität von 100% identifiziert werden. Der ORF 1 kodiert für ein regulatorisches Protein des *merR*-Gen. Dem ORF 2 konnte durch die BLAST-Funktion einem hypothetischen Protein des annotierten Gensatzes zugeordnet werden. Der ORF 3 und ORF 4 lassen sich, aufgrund der Nachbarschaft auf der Sequenz und ihrer gleichen Orientierung, zu einer Gruppe zusammenfassen. Der ORF 3 ist eine Quecksilberreduktase des *merA*-Gens. ORF 4 konnte der Organoquecksilberlyase des *merB*-Gens zugeordnet werden.



Abbildung 4.9 Genkarte des ST2-Markers 2

Dargestellt sind die fünf ORFs (grüne Pfeile). Die Abkürzungen am jeweiligen ORF geben das Protein der jeweiligen homologen Sequenz des *S. epidermidis*-Stammes NTUH-7684 an. Es konnten Homologien zu zwei hypothetischen Proteinen (HP), einem regulatorischen Protein (*merR*), einer Quecksilberreduktase (*merA*) sowie einer Organoquecksilberlyase (*merB*) identifiziert werden.

Der kompletten Sequenz des Markers 2 konnten Homologien zu mehreren Genomen zugeordnet werden. Zu einer Auswahl der, mit einer Identität von 100%, identifizierten Gensätze gehören das Genom von *S. aureus* M121 (Accession number: CP007670.1), das Chromosom 1 des *S. epidermidis*-Stammes BPH0662 (Accession number: LT571449.1) und die *composite island* SCCpbp4 *region* *S. epidermidis* ATCC 12228 (Accession number: BK001539.1).

#### 4.4.6 Identifizierung der Varianz des ST2-Marker-2-positiven Sequenzabschnitts

Der ST2-Marker 2 inseriert mittig in einem Sequenzabschnitt (*contig* 10) des Stammes ENVH350. Gemäß dem beschriebenen Schema wurde der Marker mit der einrahmenden Sequenz extrahiert sowie seine Homologien zu den markernegativen Stämmen der Stammsammlung (BLAST) ermittelt. Entsprechend dem Schema wurden sieben homologe, gekürzte Referenzsequenzen exzidiert und gemeinsam mit der markertragenden Sequenz einem *alignment* unterzogen. Der markerpositive Sequenzabschnitt weist einen hohen Sequenzanteil ohne weitere Homologien auf, der durchbrochen ist von schmalen Sequenzstücken mit Teilhomologie zu den Referenzen. Auf dem markertragenden Sequenzabschnitt befindet sich der Marker zwischen dem Basenpaar 1.500 und 9.213. Es ließ sich kein *backbone* identifizieren, das sich alle Sequenzen teilen.

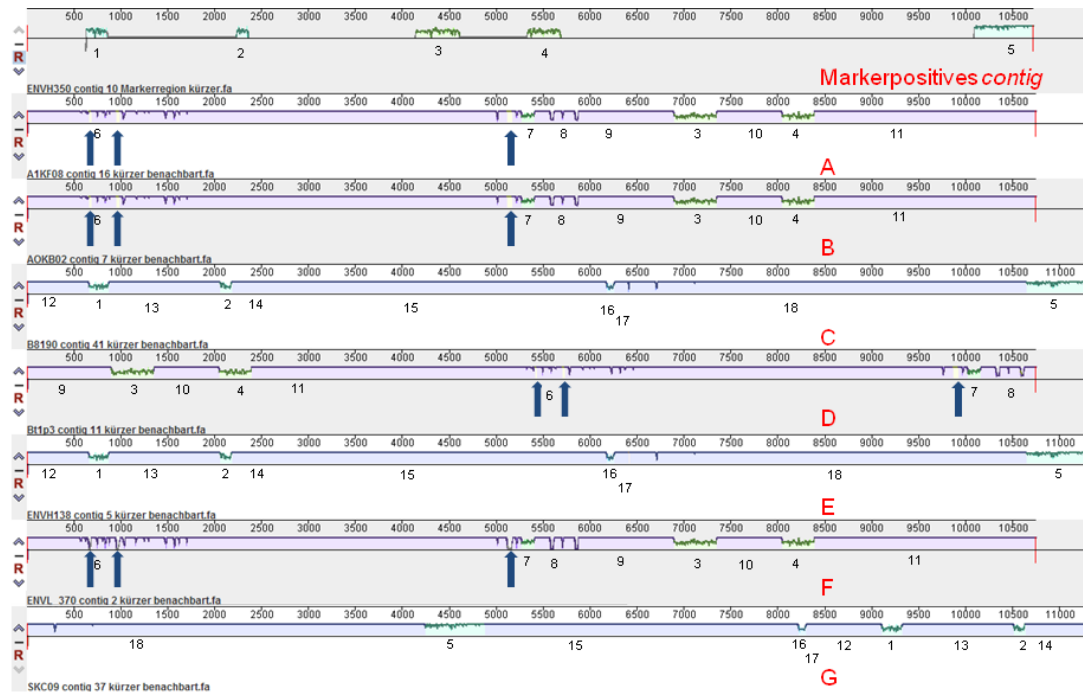


Abbildung 4.10 Das Ergebnis des *alignment* der ST2-Marker-2-positiven Sequenz sowie sieben markernegativer Referenzsequenzen

Der markerpositive Abschnitt ist die Bezugssequenz über den Referenzen. Der Marker liegt auf dem markerpositiven Sequenzabschnitt zwischen Basenpaar 1.500 und 9.213. Die markernegativen Referenzsequenzen A bis G folgen. Homologe Abschnitte zwischen den einzelnen Sequenzen sind als Block mit einem spezifischen Farbton markiert. Je höher der Grad an Konservierung des jeweiligen Blocks, desto größer ist der Block. Zum besseren Verständnis wurden diese Sequenzblöcke nummeriert. Es gibt keinen Block, der konserviert in allen Abschnitten vorliegt. Abschnitte, die nur in einer Sequenz vorkommen, besitzen keinen farblichen Block. Die dunkelblauen Pfeile markieren kurze Insertionsabschnitte in den Referenzen A, B, D und F.

Auf der markerpositiven Sequenz findet sich der Sequenzblock 1 vor dem Marker. *Downstream* zu dem Marker liegt der Sequenzblock 5. Auffällig ist, dass keiner der beiden homologen Blöcke die Markerinsertionsstelle unmittelbar flankiert. Der Marker wird stattdessen eingerahmt durch Abschnitte, die nicht homolog zu den Referenzen sind. Der Block 2 liegt im Bereich der Markerinsertionsstelle zwischen Basenpaar 2232 und 2350. Die Blöcke 3 sowie 4 liegen mittig im Bereich des Markers.

Die Referenzsequenzen der markernegativen Stämme lassen sich anhand ihrer homologen Sequenzblöcke in zwei Gruppen unterteilen. Die erste Gruppe aus den Referenzen A, B, D und F besteht aus den Sequenzblöcken 6, 7, 8, 9, 3,

10, 4 und 11. Die schwach ausgeprägte Homologie der ersten Referenzgruppe zu der markerpositiven Sequenz besteht demnach aus dem Sequenzblock 3 sowie 4. Auf der markerpositiven Sequenz trennt die beiden Blöcke eine Sequenz ohne Homologie zu den Referenzen, während die Blöcke auf den homologen Referenzen der ersten Gruppe durch den Block 10 getrennt werden. Die Stämme der ersten Gruppe lassen sich weiter unterteilen. Die Referenzen A, B und F besitzen die Sequenzblöcke in der zuvor genannten Reihenfolge. Die Sequenz D hingegen verfügt über die Sequenzabfolge aus 9, 3, 10, 4, 11, 6, 7 sowie 8. Aufgrund dieses Unterschieds ließen sich zwei Fragmente mit einer jeweils festen Sequenzblockabfolge identifizieren. Das erste Fragment besteht aus den Blöcken 6, 7 und 8. Das zweite Fragment beinhaltet die Blöcke 9, 3, 10, 4 sowie 11. Eine Translokation ist die Ursache für die unterschiedliche Anordnung der Fragmente in den Stämmen der ersten Gruppe. Die genannten Referenzen unterscheiden sich weiterhin durch drei Insertionen von wenigen Basenpaaren in Block 6. Die Insertionen der Sequenzen A, B und D sind homolog zueinander. Im Gegensatz dazu besitzen die Insertionen der Sequenz F keine Homologie zu den anderen verwendeten Sequenzen.

Die zweite Gruppe besteht aus den Referenzsequenzen C, E und G. Sie besitzen die Abfolge aus den Sequenzblöcken 12, 1, 13, 2, 14, 15, 16, 17, 18 sowie 5. Die schwache Homologie von Sequenzblock 1 und 2 und die ausgeprägtere Homologie von Block 5 verbindet die zweite Gruppe von Referenzen mit der Sequenz des markerpositiven Stammes. Die zusammengehörigen Blöcke 1 und 2 werden auf den markernegativen Referenzen durch den Block 13 getrennt. Auf der markerpositiven Sequenz befindet sich zwischen den beiden Blöcken ein Sequenzstück ohne Homologie zu den verwendeten Sequenzen. Diese Anordnung ist ähnlich der Anordnung der Sequenzblöcke 3 und 4. Interessant in diesem Zusammenhang ist der geringere Konservierungsgrad der Blöcke 1 bis 4 im Vergleich zu Block 5. Es entsteht für die markerpositive Sequenz das Bild einer Mosaikstruktur.

Die zweite Referenzgruppe lässt sich ebenfalls weiter unterteilen. Die Referenzen C und E besitzen die Sequenzblöcke in der genannten Abfolge. Auf Sequenz G hingegen liegen die Blöcke in der Reihenfolge 18, 5, 15, 16, 17, 12, 1,

13, 2 und 14 vor. Es lassen sich in diesen drei Referenzen drei Fragmente mit einer festen Abfolge von Sequenzblöcken identifizieren. Das erste Fragment besteht aus den Blöcken 18 sowie 5. Die Blöcke 15, 16 und 17 bilden das zweite Fragment. Die Abfolge aus Block 12, 1, 13, 2 sowie 14 stellt das dritte Fragment dar. Zwei Translokationen sind die Ursache für die veränderte Abfolge der beschriebenen drei Fragmente innerhalb der zweiten Gruppe.

Der Sequenz *up-* sowie *downstream* des Markers konnten mit Hilfe der Plattform NCBI (*megablast*) Homologien zu bereits annotierten Proteinen des *S. epidermidis*-Stammes NTUH-7684 zugeordnet werden. Für die *upstream* liegende Sequenz konnte eine 763 bp lange Teilhomologie zu der 792 bp langen Insertionssequenz IS431 identifiziert werden. Es besteht Homologie zu der entsprechenden Transposase und Teilhomologie zu invertierten Basenpaar-Wiederholungen (*inverted repeats*) der IS431. Die *downstream* liegende Sequenz ist vollständig homolog zu derselben 792 bp langen IS431. Die Homologie umfasst ebenfalls die Transposase von IS431 und die invertierten Basenpaar-Wiederholungen (*inverted repeats*).

#### 4.4.7 Aufbau des ST2-Markers 3

Auf dem 7.568 bp langen ST2-Marker 3 konnten vier ORFs identifiziert werden. Die ersten beiden ORFs überlappen sich auf der Markersequenz in drei Basenpaaren und besitzen die gleiche Orientierung. Beiden ORFs konnte durch die Funktion *megablast* der Plattform NCBI Homologien mit einer Identität von mindestens 98% zu Genen des *S. epidermidis*-Stammes NTUH-7684 (*Accession number*: LC085180.1) zugeordnet werden, die mit Schwermetallresistenzen assoziiert sind. Der 477 bp lange ORF 1 kodiert für eine Teilhomologie mit einer Identität von 98% zu der 497 bp langen Sequenz eines hypothetischen Proteins. Der zweite ORF ist vollständig homolog zu einem weiteren hypothetischen Protein. Der ORF 3 besitzt eine Homologie mit 100% Identität zu einem hypothetischen Protein des Kassettenchromosoms des *S. haemolyticus*-Stamm SH480 (*Accession number*: AB477967.1). Das 1.551 bp lange ORF 4

ist teilhomolog zu der insgesamt 2.267bp langen Sequenz einer CDP-Glycerol: Poly(glycerophosphat) Glycerophosphotransferase (CGPTase) des Kassettenchromosoms von *S. haemolyticus* SH480. Die Teilhomologie besitzt eine Identität von 99%. Weitere Teilhomologien mit einer geringeren Identität konnten dem ORF 4 zugeordnet werden. Hierzu zählt die Teilhomologie mit einer Länge von 1.551 bp zu dem Gensatz SCC4610 und Typ V Staphylokokken Kasette C des *S. aureus*-Stammes JCSC4610 (*Accession number*: AB773816.1). Auf diesem Gensatz besteht die Teilhomologie zu der 2.328 bp langen Sequenz eines Teichonsäurebiosyntheseprotein F mit einer Identität von 77%.



Abbildung 4.11 Genkarte des ST2-Markers 3

Dargestellt sind die ORFs (grüne Pfeile) über der schwarzen Sequenz. ORF 1 und ORF 2 konnte je ein Protein der jeweiligen, homologen Sequenz des *S. epidermidis*-Stammes NTUH-7684 zugeordnet werden. ORF 3 sowie ORF 4 besitzen Homologie zu dem *S. haemolyticus*-Stammes SH480. Es konnten Homologien zu drei hypothetischen Proteinen (HP) sowie eine Teilhomologie zu einer CDP-Glycerol: Poly(glycerophosphat) Glycerophosphotransferase (CDP-glyc glycerophosphotransferase) identifiziert werden.

Durch die Durchführung der BLAST-Funktion von NCBI mit der vollständigen Sequenz des Markers 3 (7.568 bp Länge) konnte diesem mit einer Referenzabdeckung (*query cover*) von 100% sowie einer Identität von 99% das Kassettenchromosom von *S. haemolyticus*-Stammes SH480 zugeordnet werden. Teilhomologien mit hoher Identität jedoch niedriger Referenzabdeckung zu weiteren annotierten Gensätzen, wie dem des *S. epidermidis*-Stammes NTUH-7684 (1.483 bp Referenzabdeckung) oder der *composite island SCCpbp4 region* des *S. epidermidis*-Stammes ATCC12228 (*Accession number*: BK001539.1) (2.287 bp Referenzabdeckung) wurden ebenfalls identifiziert.

#### 4.4.8 Identifizierung der Varianz des ST2-Marker-3-positiven Sequenzabschnitts

Der Marker 3 befindet sich mittig in einem Sequenzabschnitt (*contig* 5) des Stammes ENVH150. Der markerpositive Teil des Sequenzabschnitts (*contig* 5) wurde extrahiert, sowie dessen Homologien in den verbliebenen markernegativen Stämmen mit abweichendem Sequenztyp identifiziert. Eine Auswahl aus sechs Referenzsequenzen wurde dem Methodenteil entsprechend in gekürzter Form extrahiert und gemeinsam mit dem gekürzten markerpositiven Sequenzabschnitt (*contig* 5) einem *alignment* unterzogen. In der schematischen Abbildung liegt der Marker auf dem markertragenden Sequenzabschnitt zwischen dem Basenpaar 1.500 und 9.067. Die Referenzen besitzen Homologien zu den markerflankierenden Sequenzstücken. Diese Sequenzstücke liegen jedoch getrennt voneinander auf unterschiedlichen Referenzen vor. Folglich konnte die eigentliche Markerinsertionsstelle mit dem gewählten Schema der Sequenzauswahl nicht identifiziert werden.

Der markerpositive Abschnitt ist die Bezugssequenz über den Referenzen. Der Marker liegt auf dem markerpositiven Sequenzabschnitt zwischen Basenpaar 1.500 und 9.067. Die markernegativen Referenzsequenzen A bis F folgen. Homologe Abschnitte zwischen den einzelnen Sequenzen sind als Block mit einem spezifischen Farbton markiert. Je höher der Grad an Konservierung des jeweiligen Blocks, desto größer ist der Block. Zum besseren Verständnis wurden diese Sequenzblöcke nummeriert. Es gibt keinen Block, der konserviert in allen Abschnitten vorliegt. Abschnitte, die nur in einer Sequenz vorkommen, besitzen keinen farblichen Block. Die dunkelblauen Pfeile markieren kurze Insertionsabschnitte in den Referenzen A, B, D, E und F.

Der Marker wird auf dem markerpositiven Sequenzabschnitt flankiert von dem Sequenzblock 1, 3 und 4. *Upstream* des Blocks 1 befindet sich ein Sequenzstück ohne Homologie zu den Referenzen. Im Bereich des Markers liegt der Sequenzblocks 2 mit einem geringen Grad an Homologie. Dieser Sequenzblock 2 wurde ebenfalls in den markernegativen Referenzsequenzen A, B, D, E sowie F jeweils zwischen Sequenzblock 11 und 12 gefunden. Sowohl die Anordnung der Sequenzblöcke 1, 2 sowie 4, als auch der Abstand der Blöcke zueinander, unterscheiden sich von der markertragenden Sequenz. Aufgrund des Auftretens der unterschiedlichen Sequenzblöcke kann zwischen den

markernegativen Referenzen unterschieden werden. Die Referenzsequenz C besteht aus dem Block 3. Eine Sequenz mit vergleichbarem Aufbau liegt nicht vor. Die Referenzsequenzen A, B, D, E sowie F setzen sich aus den weiteren elf identifizierten Sequenzblöcken zusammen und bilden eine gemeinsame Gruppe. Aufgrund ihrer Anordnung kann diese Gruppe weiter unterteilt werden. Die erste Subgruppe aus den Sequenzen A, E und F besteht aus der Abfolge von Sequenzblock 5, 1, 6, 7, 8, 9, 4, 10, 11, 2 und 12. Die Referenzsequenzen B sowie D bilden die zweite Subgruppe. Sie besitzen die Reihenfolge aus Sequenzblock 11, 2, 12, 6, 7, 8, 9, 4, 10, 5 und 1. Es konnten für die Referenzsequenzen drei Fragmente mit fester Abfolge von Sequenzblöcken identifiziert werden. Das erste Fragment besteht aus den Blöcken 5 und 1. Das zweite Fragment setzt sich zusammen aus den Sequenzblöcken 6, 7, 8, 9, 4 sowie 10. Die Sequenzblöcke 11, 2 und 12 bilden das dritte Fragment. Die unterschiedliche Anordnung der Fragmente zwischen den Subgruppen lässt sich auf zwei Translokation zurückführen. Ein weiterer Unterschied zwischen den einzelnen Referenzsequenzen sind die kleineren Insertionen, die eine Länge von wenigen Basenpaaren besitzen und an den gleichen Positionen in den Referenzsequenzen auftreten. Diese bevorzugten Punkte befinden sich *downstream* von Block 6 sowie 7 und innerhalb von Block 12.

## 5. Diskussion

### 5.1 Voraussetzungen für diese Arbeit

*S. epidermidis* ist eine physiologische Komponente der Hautmikrobiota des Menschen (Dréno et al., 2016; Grice and Segre, 2011; Grice et al., 2009; Kloos and Musselwhite, 1975). *S. epidermidis*-Infektionen sind häufig als nosokomial erworben beschrieben worden und im besonderen Maße assoziiert mit der Verwendung von medizinischen Kunststoff-Fremdkörpern (Ortega-Peña et al., 2015; Ziebuhr et al., 2006). Diese besondere Affinität zu Fremdkörpern beruht auf der Fähigkeit der Biofilmbildung auf Polymer-Oberflächen (Gominet et al., 2017; Lewis, 2007; Ziebuhr et al., 2006). Die Biofilmbildung wird ermöglicht durch Virulenzfaktoren der primären Adhäsion und der Akkumulation (Gerke et al., 1998; Kloos and Bannerman, 1994; Otto, 2008; Rohde et al., 2005). Die aktuelle Studienlage zeigte, dass die Differenzierung zwischen kommensalen und pathogenen *S. epidermidis*-Gruppen des Patienten durch den Nachweis von den Biofilm-Faktoren kodierenden Genen *icaA*, *aap*, *bhp*, *fbe*, *embp*, *atlE* und den Biofilm-assoziierten Insertionselementen IS256 sowie IS257 nicht möglich ist. Darüber hinaus konnte auch keine Korrelation zwischen dem Methicillinresistenz-kodierenden Locus *mecA* und Pathogenität in kommensalen *S. epidermidis* der Klinik ermittelt werden (Bradford et al., 2006; Rohde et al., 2004). Der Grund hierfür ist die Selektion durch die Krankenhausumwelt von *icaADBC*-, *mecA*- und IS256-positiven, kommensalen *S. epidermidis* (Rohde et al., 2004). Im Rahmen weiterer Studien konnten Stämme mit Pathogenität assoziiert werden, die bestimmten klonalen Linien angehören, den MLST-Sequenztypen 2, 5 und 59 (Tab. 1.4). Diese Sequenztypen konnten allerdings auch in human-kolonisierenden Isolaten nachgewiesen werden, so dass auch sie keine eindeutige Differenzierung von pathogenen und apathogenen *S. epidermidis* ermöglichen (Miragaia et al., 2007). Der Schwerpunkt der vorliegenden Arbeit war die Identifizierung von Gensequenzen, die jeweils signifikant häufiger in einer dieser drei klonalen Linien auftreten. Im positiven Fall der Identifizierung derartiger Markersequenzen sollten diese näher charakterisiert werden, um die Adaption dieser Stämme an die Klinikumwelt zu erklären. Zu dieser Charakterisierung zählten die Identifikation von homologen Gense-

quenzen, die Analyse klonaler Verwandtschaft und die Untersuchung der Varianz des markerpositiven Sequenzabschnitts. Zur Analyse der klonalen Verwandtschaft wurde ein *core*-Genom für *S. epidermidis* erstellt. Von den Genloci des ermittelten *core*-Genoms konnten 99,79% als homolog zu dem eigen sequenzierten Referenzgenom zugeordnet werden. Entsprechend den Angaben des Herstellers des verwendeten Programms wurde das *core*-Genom für die Untersuchung klonaler Verwandtschaft mit diesem Wert erfolgreich validiert. Für diese Arbeit wurde auf die Genome von 95 sequenzierten Isolaten der Stammsammlung der Arbeitsgruppe Knobloch sowie auf 84 von anderen Arbeitsgruppen sequenzierte *S. epidermidis*-Stämme zurückgegriffen.

## 5.2 Charakterisierung des ST5-Markers

Durch die Sequenzanalyse der eigen sequenzierten 13 ST5-Stämmen konnte eine 3.161 bp lange ST5 spezifische Markersequenz ermittelt werden. Die weitere Analyse der Markersequenz erfolgte anhand eigen sequenzierter Genome mit abweichendem Sequenztyp sowie anhand bereits annotierter Genome. Der ST5-Marker konnte signifikant häufiger den ST5-Stämmen als Isolaten mit abweichendem Sequenztyp zugeordnet werden in der Gruppe aller verwendeter Genome, der Subgruppe pathogener Stämme sowie der Subgruppe von bereits annotierten Isolaten. Eine Korrelation von Marker und MLST-Sequenztyp 5 in der Subgruppe der Umwelt-Isolate war nicht möglich (Tab. 4.1). Die mangelnde Korrelation in der Gruppe der Umwelt-Isolate ist vermutlich auf die kleine Stichprobengröße zurückzuführen. Durch eine erneute Untersuchung des Markernachweises in einer größeren Stichprobe kann eine mögliche Assoziation von Marker und Sequenztyp in der Subgruppe der Umwelt-Isolate erneut validiert werden. Dies konnte im zeitlichen Rahmen dieser Doktorarbeit nicht mehr durchgeführt werden. Die genannten Ergebnisse zeigen, dass der Markernachweis mit dem Sequenztyp 5 in den Gruppen korreliert, die im Rahmen der klinischen Diagnostik von besonderem Interesse sind. Durch den PCR-Nachweis des Markers kann somit ein *S. epidermidis* mit fraglichem Sequenz-

typ zeitnah sowie kostengünstig entsprechend charakterisiert werden. Dies macht eine zeitaufwändigere und fehleranfälligere Sequenzierung überflüssig.

Da der MLST-Sequenztyp 5 auch in kommensalen *S. epidermidis* nachgewiesen worden ist (Miragaia et al., 2007), wurde zusätzlich die Korrelation zwischen den Merkmalen "Pathogenität" und "Markernachweis" ermittelt. Es konnte keine Korrelation zwischen Auftreten des ST5-Markers und Pathogenität entsprechend der einleitenden Fragestellung in der Gruppe aller verwendeten Isolate festgestellt werden (Tab. 4.2). Ebenso konnte der Marker nicht signifikant häufiger in den eigen sequenzierten, pathogenen Isolaten als in den Klinikumwelt-Stämmen nachgewiesen werden (Tab. 4.3). Die Ursache für das Ausbleiben der Marker-Pathogenität-Korrelation beruht auf dem niedrigen Anteil der ST5-Stämme innerhalb der jeweils untersuchten Stichprobe pathogener *S. epidermidis* (9% und 10%). Den Hauptanteil der invasiven Stämme bildeten die sequenzierten ST2-Isolate (63% und 73%), die nicht über eine ST5-spezifische Sequenz verfügen.

Die Untersuchung der klonalen Verwandtschaft der einzelnen ST5-Stämme zeigte, dass sich der Hauptteil der untersuchten Stämme zu einem Cluster zusammenschließen ließ (Abb. 4.1). Bei den beiden Isolaten außerhalb des Clusters handelt es sich um markernegative Umwelt-Isolate. Dies zeigt, dass in der Umwelt auch andere ST5-Isolate mit stark abgeändertem Gensatz erfolgreich sind. Alle weiteren Umwelt-Stämme liegen jedoch innerhalb des Clusters. Eine Trennung zwischen den Umwelt-Stämmen und den pathogenen Isolaten innerhalb des MST ist nicht möglich. Innerhalb des Clusters ließen sich drei Komplexe mit einem variablen Abstand zueinander ermitteln (68 und 79+54 Genloci). Die beiden zentralen Komplexe enthalten ausschließlich markerpositive Isolate (n=2 und n=3) und im peripheren Komplex treten markernegative Stämme prädominant auf (n=3). Dies lässt sich in Einklang bringen mit dem Ergebnis, dass markerpositive Stämme zentral und markernegative peripher im Cluster vorliegen. Der ST5-Marker ist folglich besser geeignet eine Gruppe ST5-Stämme des zentralen Clusters als nah verwandt zu identifizieren als das MLST-Schema, das Thomas et. al. 2007 publizierten. Im MST konnte ein markerpositiver Stamm als Bindeglied zwischen

zwei markernegativen Stämmen sowie ein markernegatives Isolat als Verbindung von zwei markerpositiven Stämmen identifiziert werden. Der ST5-Marker scheint nach diesem Ergebnis in einer Subgruppe der Isolate vorzuliegen. Eine weitere mögliche Erklärung wäre, dass die publizierten Genome dieser beiden auffälligen Stämme das Produkt einer fehlerhaften Sequenzierung sind. Da beide Stämme von anderen Arbeitsgruppen sequenziert worden sind, wurde zum Vergleich der Sequenzierqualität auf die von NCBI veröffentlichten N50-Werte zurückgegriffen. Beide N50-Werte (88.654 kb, 92.830 kb) belegen eine gute Sequenzierqualität der annotierten Isolate. Dies widerlegt die Theorie der mangelhaften Sequenzierung. Es muss jedoch berücksichtigt werden, dass ohne Ringschluss nie das komplette Genom vorliegt und möglicherweise aus diesem Grund der Marker nicht in einem der auffälligen Stämme nachgewiesen werden kann.

Interessanterweise bilden die 19 verwendeten ST5-Isolate nur ein Cluster mit zwei solitär abgetrennten Stämmen. Im Vergleich konnten drei Cluster für die 13 Stämme des ST59 in der Analyse der klonalen Verwandtschaft identifiziert werden. Die Vermutung liegt nahe, dass die ST5-Stämme eine geringere Heterogenität als die ST59-Isolate besitzen. Um die vermutete geringere Diversität zu beschreiben, wird die Isolierung und Sequenzierung von *S. epidermidis* mit ST5 aus einer Krankenhaus-fernen Umwelt in einer weiterführenden Arbeit angeregt.

Um die möglicherweise kodierten Virulenzfaktoren zu identifizieren, wurde der ST5-Marker auf Homologien zu den bereits annotierten Sequenzen der Plattform NCBI untersucht. Von der ST5-Markersequenz weisen 60% der Sequenz Homologie zu dem Genom des *S. haemolyticus* JCSC1435 auf. Diese Tatsache lässt sich mit der Fähigkeit der Staphylokokken-Gattung erklären, genetische Informationen über die Art-Grenzen hinweg zu transferieren. Dieser Transfer ist beispielsweise bereits für das SCCmec-Element mit dem Methicillinresistenzkodierenden *mecA*-Gen zwischen methicillinresistenten *S. epidermidis* (MRSE) und methicillinsensiblen *S. aureus* (MSSA) beschrieben worden (Bloemendaal et al., 2010; Wiolders et al., 2001). Des weiteren zeigte sich, dass die Orientierung der ORFs des ST5-Markers gegensätzlich ist (Abb. 4.2). Dies lässt die

Annahme zu, dass beide ORFs über voneinander unabhängige Promotoren verfügen. Da beide ORFs im Genom des *S. haemolyticus* JCSC1435 in gleicher Leserichtung benachbart vorliegen, bilden sie in diesem Genom vermutlich eine Transkriptionseinheit, die sich den gleichen Promotor teilt. Den beiden ORFs des Markers konnten anhand der Informationen von NCBI nur zwei Proteine mit ungeklärter Funktion zugeordnet werden, so dass Rückschlüsse auf einen Einfluss auf die Pathogenität durch den ST5-Marker nicht möglich waren.

Das *alignment* der markerpositiven sowie der markernegativen Sequenzen (Abb. 4.3) verdeutlicht die hohe genetische Varianz des markerpositiven Sequenzabschnitts. Der konservierte Sequenzblock 2, den sich alle verwendeten Sequenzen teilen (*backbone*), umfasst nur wenige Basenpaare. Die Kuhmastitis-assoziierte Referenzsequenz F ist die einzige Sequenz ohne Block 1 sowie 3. Alle weiteren Sequenzen verfügen über diese beiden Sequenzblöcke in gleicher Ausrichtung. Die Blöcke 1 bis 3 zählen somit zum konservierten Block, den sich die human-assoziierten *S. epidermidis* teilen. Ein Erklärungsansatz für das Fehlen der Blöcke 1 sowie 3 in Referenz F könnte eine geringere Qualität der entsprechenden Sequenzierung sein. Dies konnte durch den N50-Wert (81.780 kb) ausgeschlossen werden. Die Sequenzblöcke konnten zu vier Fragmenten zusammengefasst werden. Dieser Zusammenschluss beruht auf den unterschiedlichen Anordnungen der Sequenzblöcke bei gleicher Orientierung auf den Sequenzen der humanpathogenen Sequenzen. Aus diesem Grund entsteht die Theorie, dass die verwendeten Sequenzen im Zuge ihrer Entwicklung Translokationen zu unterschiedlichen Zeitpunkten durchlaufen haben. Alle Sequenzblöcke liegen mit gleicher Orientierung auf den verwendeten Sequenzen vor. Dies schließt das Vorliegen einer Inversion eines Sequenzstücks aus. Die Bruchstellen der Translokationen liegen jeweils zwischen den beschriebenen vier Fragmenten. Das vierte Fragment stellt eine Besonderheit dar. Die Sequenzblöcke (Sequenzblock 4 bis 8) des vierten Fragments treten häufig bruchstückhaft mit einem bestimmten Muster auf: Ab dem fehlenden Block fehlen alle weiteren Sequenzblöcke mit niedrigerer Nummer. Dies lässt sich ebenfalls mit der Theorie der Translokation vereinbaren. Durch die Verschiebung des vierten Fragments gehen die randständigen Blöcke dieser Abfolge verloren. Die Anordnung der Blöcke zu längeren Fragmenten lässt sich

nur eingeschränkt auf die Kuhmastitis assoziierte Referenzsequenz F anwenden. Zusätzlich besitzt die bovinpathogene Referenz F einen größeren Bereich ohne Homologie zu den anderen humanpathogenen Sequenzen. Im Laufe der Entwicklung innerhalb des Gensatzes dieses bovinpathogenen Stammes haben weitere Mutationen stattgefunden, die sich von denen der humanassoziierten Isolate deutlich unterscheiden. Diese Mutationen erklären den relativ großen Sequenzanteil ohne Homologie zu den humanpathogenen Stämmen. Die markerpositive Sequenz stellt ebenfalls eine Besonderheit dar: Sie besteht aus dem ersten Fragment, dem vierten Fragment und dem Block 8 des dritten Fragments. Für sie gilt demnach auch nur eingeschränkt die beschriebene Einteilung der Sequenzblöcke in Fragmente. Dies ist auf das Schema der Extraktion von Sequenzen für das *alignment* zurückzuführen. Der Block 12 liegt vermutlich *downstream* von Block 8. Um dies zu prüfen ist ein weiteres *alignment* mit der entsprechend extrahierten Sequenz nötig. Eine Prüfung dieser Frage konnte im Rahmen dieser Arbeit nicht mehr durchgeführt werden.

*Upstream* des Markers konnte ein Sequenzblock von 200 bp ermittelt werden, der keine Homologie zu den anderen Sequenzen besitzt. Im Zuge der Translokation kann jener 200 bp lange Sequenzblock aus einem anderen Bereich des Genoms *upstream* eingefügt worden sein, der in der Auswahl verwendeter Referenzsequenzen nicht zu identifizieren war. In diesem Fall liegt der 200 bp lange Block auf anderen Sequenzen der markernegativen Referenzstämmen vor. Eine weitere Theorie ist, dass dieser Sequenzblock stattdessen auf anderen markernegativen Stämmen als den verwendeten Referenzen nachweisbar ist. Eine dritte mögliche Theorie ist, dass die sequenztypspezifische Sequenz länger ist als der extrahierte Marker. Eine Untersuchung auf Homologien des 200 bp langen Blocks zu den markernegativen Stämmen ist aufgrund seiner geringen Größe nicht durchgeführt worden. Unabhängig von der Mindestgröße der sequenztypspezifischen Sequenz ist der Marker mit seiner Länge von 3.161 bp ausreichend lang für einen Nachweis im Rahmen der Polymerase-Ketten-Reaktion (PCR). Im Rahmen der PCR werden in der Regel Sequenzabschnitte amplifiziert, die kleiner als 1.000 bp sind.

### 5.3 Charakterisierung des ST59-Markers

Durch die Analyse der zehn eigen sequenzierten ST59-Stämme konnte eine 858 bp lange Markersequenz identifiziert werden. Der Marker wurde mittels der weiteren 169 sequenzierten Isolate weiter untersucht. Der Marker tritt signifikant häufiger in ST59-Isolaten auf als in Stämmen mit anderem Sequenztyp. Dies gilt für die Gruppe aller verwendeter Isolate sowie für die Gruppe pathogener Isolate der Stammsammlung der AG Knobloch (Tab. 4.4). Nicht möglich war der Markernachweis in den eigen sequenzierten ST59-bovinpathogenen Stämmen, in den ST59-Isolaten anderer Arbeitsgruppen und in Isolaten eines anderen Sequenztyp. Auf dieser Beobachtung fußt die Theorie, dass der identifizierte ST59-Marker nur in einer Subgruppe von ST59-Isolaten der Stammsammlung der Arbeitsgruppe Knobloch auftritt. Die Ursache für die Identifizierung eines Markers einer ST59-Subgruppe ist die Stichprobengröße ( $n=10$ ), die für die Durchführung der Markersuche verwendet wurde. Aufgrund der geringen Probenzahl ist die Wahrscheinlichkeit größer eine Sequenz zu extrahieren, die nur in einem Teil der verwendeten Stämme auftritt. Dass 50% der eigen sequenzierten, pathogenen *S. epidermidis* mit ST59 den Marker nicht besitzen, ist vereinbar mit der Theorie des Markernachweises in einer Subgruppe der verwendeten ST59-Isolate. Diese Subgruppe scheint adaptiert an ihren Herkunftsort zu sein. Der ST59-Marker hat aus diesem Grund Relevanz für die klinische Diagnostik des Universitätsklinikums Lübeck. In einer zukünftigen Arbeit sollte untersucht werden, ob weitere ST59-spezifische Sequenzen extrahierbar sind, die in einer größeren Anzahl von ST59-Isolaten vorkommen. Dem Schema zur Identifizierung von Markersequenzen sollte in dieser Untersuchung mit einer größeren Stichprobenzahl von ST59-Isolaten nachgegangen werden.

Da der Sequenztyp alleine nicht ausreicht zur Identifikation von pathogenen *S. epidermidis* (Miragaia et al., 2007), wurde zusätzlich die Marker-Pathogenität-Korrelation untersucht. Sowohl in der Gruppe aller verwendeten Stämme als auch in der Gruppe der Stammsammlung der AG Knobloch konnte keine Korrelation zwischen Markernachweis und Pathogenität ermittelt werden (Tab. 4.5 und Tab. 4.6). Da der Marker nur in einer Subgruppe von eigen sequenzierten, pathogenen ST59-Stämmen auftritt, ist eine solche Korrelation in beiden Grup-

pen mit einem hohen Anteil pathogener Stämme anderen Sequenztyps nicht zu erwarten.

Durch die Analyse der klonalen Verwandtschaft wurde deutlich, dass die drei markerpositiven Isolate hochgradig verwandt sind und gemeinsam einen Komplex bilden (Abb. 4.4). Dies belegt die Theorie, dass die ST59-Marker-positiven Stämme eine Subgruppe von ST59-Isolaten mit einer geringen genetischen Diversität sind. Es bleibt fraglich, warum die anderen drei pathogenen ST59-Stämme, die im Rahmen dieser Arbeit sequenziert wurden, nicht den Marker tragen. Im Fall dieser drei Stämme kann spekuliert werden, dass eine fehlerhafte Sequenzierung vorliegt. Diese Theorie konnte anhand der entsprechend hohen N50-Werte (57.624, 84.111, 65.050) jedoch widerlegt werden. Auffälligerweise besitzen die markernegativen pathogenen Stämme eine deutliche genetische Distanz vom Komplex der markerpositiven Isolate. Die drei Stämme gehören anderen Clustern an als der Komplex der markerpositiven Isolate. Dies ist ein Hinweis auf das Vorliegen von unterschiedlichen Subgruppen von ST59-Isolaten, die an das Überleben im Universitätsklinikum Schleswig-Holstein adaptiert sind.

Um die Ursache für die festgestellte Adaption an das Überleben im Klinikum Lübeck zu klären, wurden die Homologien des ST59-Markers zu bereits annotierten Sequenzen der Datenbank von NCBI bestimmt. Da der ST59-Marker nur schwach homolog zu dem auf NCBI annotierten *S. haemolyticus* JCSC1435 (*discontiguous megablast* statt *megablast*, Identität 77%) ist, kann angenommen werden, dass der ST59-Marker nicht im Rahmen eines Gentransfers zwischen den markerpositiven Stämmen und dem *S. haemolyticus* ausgetauscht wurde. Der Marker ist teilhomolog zu jeweils einem Proteinprodukt unbekannter Funktion des *S. haemolyticus*-Isolats. Mit Hilfe der Algorithmen der CLC Genomics Workbench konnten keine ORFs auf der Markersequenz identifiziert werden, obwohl für die Markererstellung zwei ORFs verwendet wurden. Die Ursache beruht auf dem letzten Schritt der Markerextraktion aus dem Gensatz des qualitativ hochwertig sequenzierten Stammes. In diesem Schritt wurde die Markersequenz unter Ausschluss von Homologien zu Sequenzen von Stämmen mit abweichendem Sequenztyp extrahiert. Die Extraktion wird eine Verkürzung

der *up-* oder *downstream* liegenden Strukturmerkmale der beiden ORFs zur Folge gehabt haben, so dass eine Identifikation der ORFs nicht mehr möglich ist.

Im *alignment* wurde deutlich, dass der ST59-Marker-positive Sequenzabschnitt über einen großen konservierten Anteil verfügt, der ebenfalls in den markernegativen Sequenzen nachgewiesen werden konnte (*backbone*) (Abb. 4.5). Der konservierte Sequenzabschnitt besitzt drei Punkte (zwischen den Blöcken 1, 2 und 4 sowie an der Position von Basenpaar 276 des Blocks 1), an denen häufig genetische Veränderungen stattfinden. Der Marker ist eine Insertion an einer dieser Positionen. Interessanterweise besitzen vier der neun verwendeten Sequenzen eine jeweils andere Abfolge von neun Basenpaaren an gleicher Position (Basenpaar 276). Dies zeigt, dass diese Position ein bevorzugter Punkt für Mikroevolution ist. Die vier genannten Sequenzen unterscheiden sich in ihren *downstream* liegenden, weiteren Insertionen. Folglich ist die identifizierte Mikroevolution unabhängig von den weiteren genetischen Ereignissen.

Im Vergleich zu der Analyse der Varianz der ST 2-Marker 1-positiven Sequenz fällt ein geringerer Grad an Konservierung der Sequenz auf, die sich alle Stämme teilen (Block 1, 2, 4). Dieses Ergebnis lässt sich einordnen zu den bereits beschriebenen genetischen Ereignissen, die gehäuft an bestimmten Positionen der verwendeten Sequenzen stattfinden. Aufgrund des *rearrangement* von Sequenzfragmenten innerhalb der konservierten Sequenz sinkt die Homologie der konservierten drei Sequenzblöcke zueinander. Ein eindeutiges Schema für eine möglicherweise stattgefundenene Evolution der Sequenzen des *alignment* konnte nicht aufgestellt werden. Für ein solches Schema müssen demnach mehr Stämme in das *alignment* eingeschlossen werden. Dies war im zeitlichen Rahmen der Arbeit nicht mehr möglich.

## 5.4 Charakterisierung der ST2-Marker

Durch die Untersuchung der 57 Stämme konnten drei Markersequenzen mit unterschiedlicher Länge identifiziert werden. Mindestens einer der drei Marker konnte in der Gruppe aller verwendeten Isolate in 91% der ST2-Isolate nachgewiesen werden (Tab. 4.7). Die so definierte Markerkombination tritt signifikant häufiger in ST2-Stämmen auf als in solchen mit abweichendem Sequenztyp. Dies zeigt die hohe Sensitivität bei Verwendung aller drei Marker für den MLST-Sequenztyp 2. Zwei Genome, die durch eine andere Arbeitsgruppe annotiert wurden, konnten nicht eindeutig durch das MLST-Verfahren als ST2 identifiziert werden. Diese beiden Genome wurden zunächst als unklarer ST zu den Stämmen mit abweichendem Sequenztyp gezählt. Sollten diese beiden Stämme nach erneuter Sequenzierung dem ST2 zugeordnet werden, würde das einen Anstieg der Sensitivität auf 94% zur Folge haben.

Innerhalb der Gruppe aller verwendeten Isolate sowie in der Gruppe der pathogenen Isolate konnte jeder der drei ST2-Marker signifikant häufiger in Isolaten mit ST2 als in solchen mit anderem Sequenztyp nachgewiesen werden. Folglich sind die einzelnen Marker im Kontext der klinischen Diagnostik geeignet pathogene ST2-Isolate des Universitätsklinikums Lübeck sowie ST2-Isolate der gesamten Stichprobe als ST2 zu identifizieren. Der ST2-Marker 2 konnte signifikant häufiger in der Gruppe bereits annotierter Genome in ST2-Isolaten als in Stämmen mit anderem ST identifiziert werden. Der Marker 2 ist demnach ebenfalls im Rahmen eines PCR-Schnelltest-Verfahrens für ST2-Stämme unterschiedlicher Herkunft geeignet. Das Signifikanzniveau dieses Ergebnisses ist allerdings niedriger als das für den ST2-Nachweis innerhalb der anderen beiden Gruppen ( $p \leq 0.05$  gegen  $p \leq 0,01$ ). Diese schwächere Korrelation sowie die fehlende Korrelation für diese untersuchte Gruppe der anderen beiden ST2-Marker erklärt sich im Schema der Markersuche. Für die primäre Markersuche wurden ausschließlich Stämme der Stammsammlung der eigenen Arbeitsgruppe verwendet. Somit steigt die Wahrscheinlichkeit eine Sequenz zu isolieren, die sensitiver für die eigen sequenzierten Stämme ist. In der Gruppe der Umwelt-Isolate korrelieren die Marker nicht mit dem Merkmal ST2. Die Ursache liegt ebenfalls im Schema der Markersuche, da prädominant (89%) pathogene

Stämme für die primäre Markersuche verwendet wurden. Aus diesem Grund steigt die Wahrscheinlichkeit eine Sequenz zu extrahieren, die sensitiv für die Gruppe pathogener ST2-Isolate ist.

Wie auch für die beiden Marker anderer Sequenztypen zuvor wurde die Korrelation zwischen dem Markernachweis und Pathogenität analysiert. In der Gruppe aller verwendeten Isolate treten die drei Marker jeweils signifikant häufiger in den pathogenen Stämmen auf als in den Umwelt-Isolaten (Tab. 4.8, 4.10, 4.12). In der klinischen Diagnostik macht der Nachweis eines der drei Marker in einem *S. epidermidis* das Vorliegen eines pathogenen Stammes wahrscheinlicher. Dieses Ergebnis muss allerdings unter dem Umstand betrachtet werden, dass die Umwelt-Isolate nicht alle der Krankenhausumwelt entstammen. In diese Gruppe der Umwelt-Isolate wurden neben Krankenhausumwelt-Isolate unter anderem auch veterinäre *S. epidermidis* als Vergleichsgruppe einbezogen. Dies ist bei Betrachtung der aktuellen Studienlage sinnvoll: Rohde et. al. beschrieben 2004 einen Shift der kommensalen *S.-epidermidis*-Population in der Krankenhausumwelt. Die nosokomial besser adaptierten *S.-epidermidis*-Populationen werden durch die Klinikumwelt begünstigt und verdrängen andere Stämme. Anhand einer aktuellen Studie konnten erste Rückschlüsse auf den zeitlichen Verlauf des beschriebenen Shifts kommensaler *S. epidermidis* in der Klinikumwelt gezogen werden. In dieser Studie waren Patienten, die aus der krankenhausesfernen Umwelt auf die Intensivstation aufgenommen wurden, zunächst zu 22% von MRSE besiedelt. Nach drei Tagen konnte bei 86% und nach fünf Tagen bei 100% der Patienten MRSE nachgewiesen werden (Widerström et al., 2016). Da Methicillinresistenz ein Zeichen erhöhter Adaption in der Krankenhausumwelt darstellt, dauert der Kolonisierungsprozess in diesem beschriebenen Fall auf einer Intensivstation höchstens 120 Stunden. Folglich ist innerhalb dieses Zeitraums der Nachweis eines apathogenen Umweltisolats als Kontamination aus der Hautflora in einer Blutkultur möglich. Aus diesem Grund ist die Einbeziehung von Krankenhaus-fern gewonnenen *S. epidermidis* in die Beurteilung der Assoziation von Marker und Pathogenität gerechtfertigt. Die relevante Stichprobe der Studie für den zeitlichen Kolonisierungsverlauf bestand aus neun Patienten, die innerhalb der ersten 24 Stunden auf die Intensivstation gelegt wurden (Widerström et al., 2016).

Für eine genauere Untersuchung der zeitlichen Veränderungen der *S. epidermidis*-Besiedlung ist in Zukunft eine weitere Studie mit einer größeren Stichprobe sinnvoll.

In der Gruppe der eigen sequenzierten Isolate konnte der Marker 1 statistisch signifikant häufiger den pathogenen *S. epidermidis* als den Klinikumwelt-Stämmen zugeordnet werden (Tab. 4.9). Damit zeigt sich die Eignung des Markers 1 Sequenztyp-unabhängig zwischen den Klinikumwelt- und pathogenen Stämmen zu differenzieren. Die Korrelation könnte auf den höheren Anteil von Isolaten mit dem ST2 zurückzuführen sein. Je nach epidemiologischer Studie streut der Anteil der ST2-Stämme in Relation zu allen identifizierten, Bakteriämie-assoziierten Isolaten (Tab. 1.4). In einer schwedischen Langzeitstudie über 29 Jahre konnten allerdings 65% der pathogenen *S. epidermidis* dem ST2 zugeordnet werden (Ahlstrand et al., 2014). Die Relation dieser Studie ist ähnlich dem Anteil pathogener ST2-Isolate der Stammsammlung zu *S. epidermidis* mit abweichendem Sequenztyp (ca. 72%). Folglich ist der gewählte ST2-Anteil für die Korrelation von Markernachweis und Pathogenität berechtigt. Den größten Anteil (ca. 66,67%) der pathogenen Lübecker ST2-Stämme bilden 34 Stämme eines Ausbruchsgeschehens von linezolidresistenten *S. epidermidis* im Jahr 2012 bis zum Jahr 2014. Ausbruch ist in dieser Arbeit definiert als der häufigere Nachweis eines bestimmten Pathogens örtlich begrenzt innerhalb eines bestimmten Zeitraums. Die 34 Stämme entstammen dem gleichen Klon, der sich an die Verhältnisse des UKSH Lübeck angepasst hat. Diese Tatsache könnte als Verzerrung des Ergebnisses der Korrelation von Marker 1 und Pathogenität gedeutet werden, da alle Nachkommen dieses Klons markerpositiv sind. Diese Erklärung relativiert sich jedoch bei Betrachtung der klonalen Verwandtschaft der ST2-Stämme (Abb. 4.6): Die linezolidresistenten Stämme bilden einen gemeinsamen Komplex, unterscheiden sich jedoch voneinander in einer kleinen Anzahl von Genloci. Damit zeigt sich, dass der Marker 1 in dieser Gruppe von linezolidresistenten *S. epidermidis* über den Zeitraum von über 2 Jahren konstant vorlag und nicht durch Mutationen verändert wurde. Eine weitere Theorie bezogen auf eine mögliche ungerechtfertigte Marker-Pathogenität-Korrelation ist, dass die Klinikumwelt-Stämme eines bestimmten Krankenhauses die Markersequenz durch Deletion verloren haben. In diesem Fall wären alle

apathogenen Stämme anderer Kliniken markerpositiv und die Marker-Pathogenität-Korrelation würde sich auf jene Klinik beschränken. Dies kann widerlegt werden, da die markernegativen Krankenhausumwelt-Isolate aus mehreren unterschiedlichen Städten stammen (Hamburg, Lübeck, Jelenia Góra in Polen). Für das Merkmal Pathogenität eines klinischen *S. epidermidis*-Isolats ist der Marker 1 demnach aussagekräftig.

In der Gruppe eigen sequenzierter Stämme trat der ST2-Marker 2 nicht signifikant häufiger in den pathogenen als in den Krankenhausumwelt-Stämmen auf (Tab 4.11). Der Marker ist aus diesem Grund nur eingeschränkt für die Identifizierung von pathogenen *S. epidermidis*-Isolaten im Rahmen der klinischen Diagnostik geeignet. Die Ursache liegt im Prozess der Markersuche. Der Schwerpunkt lag hierbei auf der Identifikation von Markern, die ausschließlich in Stämmen eines Sequenztyps vorkommen. Der Marker 2 konnte entsprechend der Zielsetzung dieser Arbeit als eine sequenztypspezifische Sequenz identifiziert werden (Tab. 4.7). Eine wiederholte Suche nach Markern spezifisch für Pathogenität in *S. epidermidis* sollte sequenztypunabhängig mittels einer differentiellen Sequenzanalyse stattfinden. Die Subtraktion sollte mit Genomen von Umwelt-Isolaten stattfinden, die qualitativ hochwertig sequenziert wurden (*mapping*). Die verbliebenen Sequenzen sollten vergleichbar mit dem in dieser Arbeit verwendeten Schema auf ihr Vorliegen in apathogenen und pathogenen Stämmen untersucht werden.

In der Gruppe der eigen sequenzierten Genome war der Nachweis des ST2-Markers 3 in pathogenen *S. epidermidis* signifikant häufiger als in den Krankenhausumwelt-Stämmen (Tab. 4.13). Der Marker 3 ist aufgrund der höheren Signifikanz ( $p \leq 0,01$  statt  $p \leq 0,05$ ) besser als der Marker 1 für die Differenzierung zwischen pathogenen und Klinikumwelt-Isolaten geeignet. Wie für Marker 1 gilt, dass die markernegativen Klinikumwelt-Stämme aus unterschiedlichen Krankenhäusern stammen. Die Theorie der Deletion der Markersequenz in nur einer Subgruppe von Klinikumwelt-Stämmen eines bestimmten Krankenhauses schließt sich aus. Ebenso wie Marker 1 konnte die Markersequenz über einen Zeitraum von zwei Jahren mehrheitlich in den Stämmen eines Ausbruchs von linezolidresistenten KNS nachgewiesen werden (33 von 34 Stämmen). Diese

Tatsachen zeigen den hohen Grad der Eignung des ST2 Markers 3 zur Identifikation von pathogenen Isolaten.

Die Analyse der klonalen Verwandtschaft der ST2-Stämme zeigt, dass eine Zuordnung von Markerkombinationen zu einzelnen Komplexen sowie Clustern möglich ist (Abb. 4.6). In Komplex 1 finden sich überwiegend Stämme, die über drei Marker verfügen (35 von 36 Stämmen). Die Ursache für die enge Verwandtschaft mit Komplexbildung ist darin begründet, dass die Stämme von Komplex 1 einem Ausbruch von linezolidresistenten, koagulasenegativen *S. epidermidis* entstammen (34 von 36 Stämmen). Dieser Ausbruch ist vermutlich auf einen Klon zurückzuführen, der über einen Zeitraum von über zwei Jahren stetig in der betroffenen Klinik weitergetragen wurde. Eines der resistenten Isolate ist auffälligerweise nur positiv für Marker 1 und 2 getestet worden. Dies ist ein Stamm, der zeitlich zu Beginn des Ausbruchgeschehens isoliert wurde. Im zeitlichen Verlauf isolierte Stämme des Ausbruchs verfügen über Marker 3. Diese beiden Tatsachen lassen sich mit der Theorie des Überlebensvorteils auf dem Ökosystem Mensch für Marker 3-positive Stämme vereinbaren (Kap. 5.4.3). Eine weitere mögliche Erklärung für das Fehlen des Markers 3 in dem beschriebenen Isolat wäre eine verminderte Qualität der Sequenzierung. Zum Qualitätsvergleich wurde hierfür der N50-Wert verwendet. Dieser weist in diesem Fall auf eine qualitativ hochwertige Sequenzierung hin (N50-Wert: 87.411 kb). Sequenzierungen ohne Ringschluss erfassen allerdings nie das vollständige Genom. Die Möglichkeit besteht, dass das entsprechende Sequenzstück zufällig nicht in die Sequenzierung einbezogen worden ist. Eine erneute Sequenzierung des entsprechenden Stammes (LRKNS009) hätte Klärung schaffen können, wurde allerdings aus zeitlichen Gründen nicht mehr durchgeführt.

Im Komplex 2 finden sich ausschließlich Stämme mit Marker 1 sowie 3 (8 von 8 Stämmen). Im Komplex 3 finden sich mehrheitlich Marker 3-positive Stämme (2 von 3 Stämmen). Stämme mit Marker 2 bilden den zweiten Cluster (2 von 2 Stämmen) und hauptsächlich den dritten Cluster (7 von 8 Stämmen). Aufgrund des gleichen Markerauftretens könnten der zweite sowie dritte Cluster Bestandteil eines größeren, gemeinsamen Clusters sein. Diese Theorie ist ver-

einbar mit der Tatsache, dass nur eine sehr geringe genetische Distanz (157 Genloci) zwischen beiden Clustern besteht. Weniger wahrscheinlich ist der Zusammenschluss des ersten und zweiten Clusters zu einem größeren Cluster aufgrund des Vorliegens einer höheren genetischen Distanz (509 Genloci).

Die auffällige Abwesenheit aller drei Marker in einem Stamm des dritten Cluster konnte primär nicht auf eine schlechte Qualität der Sequenzierung zurückgeführt werden, da der N50-Wert (112.534 kb) dies widerlegte. Wie zuvor erläutert umfassen Sequenzierungen allerdings nie das komplette Genom. Aus diesem Grund wurde das möglicherweise vorhandene Sequenzstück mit Marker 2 nicht sequenziert. Eine erneute Sequenzierung dieses Stammes war nicht möglich, da er als öffentlich annotiertes Genom nicht zu der Stammsammlung der Arbeitsgruppe Knobloch zählte. Im Rahmen der Analyse der klonalen Verwandtschaft zeigte sich, dass die jeweils vorliegende ST2-Marker-Kombination zu Identifikationen von Clustern sowie Komplexen geeignet ist. Folglich ist das jeweilige Auftreten der ST2-Marker-Kombination besser als das einfache MLST-Verfahren geeignet, nah verwandte ST2-Isolate einander zuzuordnen.

#### **5.4.1 ST2-Marker 1**

Um die Faktoren der Adaption an die Krankenhausumwelt durch den 34.798 bp langen ST2-Marker 1 zu identifizieren, wurden die Homologien der einzelnen ORFs bestimmt. Den fünf ORF-Gruppen konnten hierbei mit der größten Übereinstimmung Sequenzen des bereits annotierten Genoms des ST8-Stammes ATCC 12228 zugeordnet werden (Abb. 4.7). Die Homologie zu diesem Isolat ist auf horizontalen Gentransfer zurückzuführen, der zwischen Staphylokokken-Arten bereits für das *SCCmec*-Element beschrieben wurde (Bloemendaal et al., 2010; Wielders et al., 2001). Die erste Gruppe besteht aus ORF 1 und 2. Aufgrund der gegensätzlichen Orientierung beider ORFs ist davon auszugehen, dass sie keine gemeinsame Transkriptionseinheit bilden. Der ORF 1 kodiert eine Integrase. Dies ist vereinbar mit der Theorie, dass es sich bei dem Marker um ein mobiles, genetisches Element handelt. Fraglich bleibt, ob der Marker

noch seine Position innerhalb des Genoms ändern oder durch horizontalen Gentransfer an andere Bakterien weitergegeben werden kann. Das Vorliegen einer intakten Integrase kann im Rahmen zukünftiger Sequenzanalysen geprüft werden. ORF 2 und die ORFs der zweiten sowie der dritten ORF-Gruppe sind jeweils homolog zu Proteinprodukten mit unbekannter Funktion. Die ORFs 6 bis 9 der dritten Gruppe liegen benachbart in gleicher Orientierung vor. Demnach ist anzunehmen, dass sie eine Transkriptionseinheit im Sinne eines Operons bilden. Die Klärung dieser Theorie bedarf weiterer Sequenzanalysen. Die ORFs 10 bis 14 der vierten Gruppe liegen ebenfalls benachbart mit gleicher Orientierung vor. Die Theorie des Vorliegens eines Operons lässt sich aus diesem Grund ebenfalls für die vierte ORF-Gruppe aufstellen. Die Promotorregion wäre vor dem ORF 10 zu lokalisieren, das ein hypothetisches Protein kodiert. Der ORF 11 dieser Gruppe ist homolog zu der Sequenz einer vermuteten ATPase TraE. Die ATPase TraE stellt durch ATP-Spaltung Energie bereit und ist im Zusammenhang mit dem *type four secretion system* (T4SS) beschrieben worden. Das T4SS spielt sowohl bei grampositiven als auch gramnegativen Bakterien eine wichtige Rolle für den interzellularen Stoffaustausch oder den Transport von Effektorproteinen in Wirtszellen (Rangrez et al., 2010; Zhang et al., 2012). Der ORF 12 kodiert für ein *secretory-antigen-SsaA-like*-Protein. Das Protein SSaA besitzt ein hohes Antigenpotential. In Blutproben von Patienten mit *S. epidermidis*-bedingter Endokarditis konnten erhöhte Titer der anti-SSaA-IgG nachgewiesen werden. Demnach stellt das Protein SSaA einen Faktor dar, den eine Subgruppe invasiver *S. epidermidis* besitzt. Für SSaA konnte Homologie zu einem Biosynthese-Protein des Antioxidans Staphyloxanthin von *S. aureus* nachgewiesen werden (Lang et al., 2000). Folglich ist zu vermuten, dass es sich um einen Faktor handelt, der einen Einfluss auf die Pathogenität der entsprechenden Marker 1-positiven *S. epidermidis* besitzt. ORF 13 kodiert für ein hypothetisches Protein. ORF 14 der vierten ORF-Gruppe kodiert ein nicht näher beschriebenes, replikationsassoziiertes Protein. Aufgrund ihrer gleichen Orientierung wie auch wegen ihrer direkten Nachbarschaft kann für die ORFs der fünften Gruppe von Marker 1 ebenfalls eine Operonstruktur angenommen werden. Vor dem ORF 15, das ein hypothetisches Protein kodiert, muss die entsprechende Promotorregion vermutet werden. ORF 16 ist homolog zu der DNA-Topoisomerase III topB. Die Topoisomerase dient der Entspannung

der positiven Superspiralisierung im Zuge der Replikation. Der Mechanismus der DNA-Topoisomerase besteht aus einem Dreischritt. Zunächst kommt es zu einem vorübergehenden DNA-Strangbruch. Anschließend wird einer der DNA-Stränge durch den Bruch geführt. Im letzten Schritt werden die entspiralisierten Strangfragmente erneut verknüpft. Für die DNA-Topoisomerase III *topB* konnte nachgewiesen werden, dass sie nicht hauptsächlich an der Entspiralisierung der DNA-Stränge beteiligt ist, sondern in Abwesenheit einer Gyrase oder der Topoisomerase IV unterstützend auf den Fortschritt der Replikation wirkt (Perez-Cheeks et al., 2012). Die Sequenz der DNA-Topoisomerase III *topB* ist bereits als Insertionspunkt beschrieben worden. In dem beschriebenen Fall wurde neben einem neuen Pathogenitätsfaktor zusätzlich eine weitere, intakte DNA-Topoisomerase III *topB* in die Sequenz inseriert (Carter et al., 2013). Mit Hilfe der BLAST-Anwendung der Sequenzdatenbank von NCBI konnte allerdings für die markerflankierenden Sequenzstücke des markerpositiven Sequenzabschnitts (*contig* 40 des Stammes B647) keine Homologie zu einem Fragment der DNA-Topoisomerase III *topB* identifiziert werden. Die Theorie einer Insertion in die Gensequenz der DNA-Topoisomerase III *topB* vergleichbar mit dem beschriebenen Fall ist demnach widerlegt. Interessanterweise wurden jedoch für die Regionen *up-* sowie *downstream* des Markers 1 Homologien zu Strukturen von mobilen genetischen Elementen ermittelt (Transposasen, Integrasen, Phagen-Bestandteile). Dies lässt sich vereinbaren mit der zuvor aufgestellten Theorie, dass es sich bei der Markersequenz um ein mobiles genetisches Element handelt. Das Vorliegen von Phagen-Protein-kodierenden Sequenzen ist als möglicher Hinweis auf die Einbettung des Markers in einen Bakteriophagen zu werten.

Eine Unterscheidung erfolgt zwischen temperenten und lytischen Bakteriophagen. Temperente Phagen werden im Rahmen des lysogenen Vermehrungszyklus in die bakterielle Sequenz eingebaut und liegen dort zunächst ruhend als Prophage vor. Im Rahmen der bakteriellen Zellteilung werden die temperenten Phagen an die Tochterzellen weitergegeben. Durch äußere Einflüsse, beispielsweise DNA-schädigende UV-Strahlung oder Antibiotika, sind sie in der Lage in den lytischen Vermehrungszyklus zu wechseln. Im Verlauf dieses Zyklus werden intrazellulär Phagenbestandteilen synthetisiert sowie zu neuen

Virionen zusammengesetzt. Die Virionen werden im letzten Schritt des Zyklus durch Lyse des Bakteriums freigegeben. Bestimmte Prophagen können jedoch auch stabil in ihrem Wirtsbakterium vorliegen. Durch den Einbau des Prophagen in das bakterielle Genom können temperente Phagen aus mehreren Gründen zu der bakteriellen Fitness beitragen. Infizierte Bakterien können nicht durch weitere temperente oder lytische Bakteriophagen der gleichen Prophagen-Gruppe infiziert werden. Temperente Phagen können als Vektoren Pathogenitätsfaktoren anderer Bakterienstämme transferieren, zum Beispiel Exotoxine, Adhäsionsproteine oder immunmodulierende Proteine. Der Prozess des Gentransfers durch Phagen wird Transduktion genannt. In dem Bakterium ruhende Prophagen können im Lauf der Evolution eine zunehmend wichtige Rolle in der Stoffwechselregulation des Wirts spielen. Bestimmte Phagen enthalten eine Kopie des Gens, das ihr Zielgen für die Integration in die bakterielle Sequenz ist. Dadurch entsteht für das infizierte Bakterium zunächst kein Genverlust durch die Integration des Phagens. Lytische Bakteriophagen hingegen bauen ihr Erbgut nicht in die bakterielle Sequenz ein, sondern durchlaufen ausschließlich den lytischen Vermehrungszyklus (Fortier and Sekulovic, 2013).

Ob Marker 1 Bestandteil einer Phagensequenz ist, muss mit einer alternativen Suchstrategie geklärt werden. Falls der Marker Teil eines Phagens ist, muss geklärt werden, ob dieser ein stabiler oder ein potentiell lytischer Prophage ist. ORF 18 kodiert eine Sortase. Sortasen lassen sich anhand ihrer Primärsequenz wie auch ihrer Funktion in sechs unterschiedliche Klassen, A bis F, unterteilen. Ihre Funktion als Cystein-Transpeptidasen besteht aus der Verknüpfung von Proteinen an das Peptidoglykangerüst der Zellwand oder dem Zusammenschluss von Proteinen zur Bildung von Pili. Es ist davon auszugehen, dass der ORF homolog zu der Sequenz einer Sortase der Klasse A ist. Grund für diese Annahme ist die Homologie des ORFs zu der Sortase-kodierenden Sequenz eines *S. epidermidis*-Isolats (ATCC 12228). *S. epidermidis* gehören zu der Firmicuten-Unterfamilie von grampositiven Bakterien, die über Sortasen der Klasse A verfügen. Aufgrund der Eigenschaft eine Vielzahl von Proteinen mit unterschiedlicher Funktion mit der Zellwand zu verknüpfen, stellt die Sortase A einen entscheidenden Faktor in der bakteriellen Adhäsion, Ernährung, Invasion und Manipulation des Wirtsimmunsystems dar. Die Abwesenheit einer funktio-

nierenden Sortase senkt nachweislich die Pathogenität in *S. aureus* (Spirig et al., 2011). Der Zusammenhang aus Deletion der Sortase A und einer abgeschwächten Pathogenität wurde ebenfalls für *S. epidermidis* beschrieben (Tan et al., 2017). ORF 19 ist ein LtrC-like Protein. Das LtrC-like Protein ist im Zusammenhang mit der Genanalyse des Plasmids pLgLA39 des *Lactobacillus gasserii* Stammes LA39 (*Accession number*: NC\_011839.1) beschrieben worden. Auf der Plattform NCBI waren keine Informationen zu der genauen Funktion des Proteins verfügbar. Auf dem Plasmid pLgLA39 finden sich Gene für die Plasmid-Grundfunktionen (Replikation, Erhalt, Konjugation), für die Produktion von dem Bakteriocin Gassericin A sowie Resistenz gegen diesen Wirkstoff. Die antibakteriellen Bakteriocine eines Bakteriums richten sich in der Regel gegen nah verwandten Stämme, die den Wirkstoff nicht synthetisieren. Die Stämme mit entsprechender Bakteriocinsynthese erlangen einen Wachstumsvorteil, da sie neben der Wirkstoff-Synthese auch die entsprechende Resistenz entwickeln (Ito et al., 2009). Da Plasmide als mobile genetische Elemente durch Konjugation zwischen unterschiedlichen Bakterien ausgetauscht werden können, erklärt dies die vorliegende Homologie zwischen der Sequenz von Staphylokokken und einer *Lactobacillus* zugehörigen Sequenz. Zusammenfassend lässt sich aufgrund der identifizierten Homologien der ORFs die Aussage treffen, dass die Markersequenz Faktoren beinhaltet, die einen Einfluss auf die Pathogenität von markerpositiven Stämmen haben können.

Das *alignment* der markerpositiven Sequenz und der markernegativen Referenzen zeigt, dass der Marker 1 in den konservierten Sequenzblock (*backbone*) inseriert (Abb. 4.8). *Downstream* des Markers befindet sich der Block 7 mit einer Länge von 3.582 bp, der nur auf der markerpositiven Sequenz (B647) identifiziert wurde. Dies lässt sich mit der Theorie vereinbaren, dass es sich um eine weitere sequenztypspezifische Sequenz handelt. Der Block 7 ist vollständig homolog zu zwei hypothetischen Proteinen, teilhomolog zu einem hypothetischen Protein sowie vollständig homolog zu einem Phagenresistenzgenfragment. Dies kann ein Hinweis darauf sein, dass Block 7 ein Prophagenfragment ist. Da ein Phagenresistenzgenfragment vorliegt, entsteht die Theorie, dass der Prophage durch Mutationen verändert wurde. Aufgrund dieser Mutationen kann der Prophage seine Fähigkeit verloren haben, in den lytischen Zyklus zu wech-

seln. Um diese Theorien zu prüfen, müssen zukünftig Sequenzanalysen durchgeführt werden. In dem Stamm der Markerextraktion (B12142) konnte der Block 7 auf einem anderen Sequenzabschnitt identifiziert werden als der vollständige ST2-Marker 1. Entsprechend dem verwendeten Markerschema wurde die größtmögliche sequenztypspezifische Sequenz aus diesem Stamm mit der Dissoziation von Block 7 extrahiert, so dass der Marker nicht den Block 7 umfasst. Der extrahierte Marker besitzt eine Länge von 34.798 bp und ist deswegen für den ST2 Nachweis im Rahmen einer PCR geeignet, obwohl eine möglicherweise längere sequenztypspezifische Sequenz vorliegt. Der Block 7 ist in den untersuchten Stämmen mehrheitlich nicht gekoppelt an den Marker 1 (Tab. 8.8), tritt aber in den gleichen Stämmen an anderer Position im Genom auf. Aufgrund der gewählten Definition für das Merkmal "markerpositiv" (Markeranteil in Summe auf einem Stamm  $\geq 75\%$  der Basen des gesamten Markers) können auch markernegative Stämme teilhomolog für den Marker sein. Die entsprechenden markernegativen Stämme mit einer geringen Teilhomologie zu dem Marker 1 besitzen ebenfalls eine geringe Teilhomologie zu dem Block 7. Der Block 7 kann somit ein weiterer Marker sein. Die Homologie des Blocks 7 kann in weiteren Untersuchungen in Stämmen mit ST2 und abweichendem Sequenztyp untersucht werden, um dies zu bestätigen. Diese Analysen konnten im Rahmen dieser Arbeit aus zeitlichen Gründen nicht mehr durchgeführt werden.

Der konservierte Sequenzblock rahmt den Marker 1 sowie den Block 7 ein und tritt in allen Referenzsequenzen des *alignment* auf. Er wird auf den Sequenzen *upstream* sowie *downstream* flankiert von weiteren Sequenzblöcken. Die Reihenfolge sowie die Orientierung der Blöcke ist gleichbleibend. In Abhängigkeit von der jeweiligen Sequenzlänge fehlen die entsprechenden Blöcke bei erhaltener Reihenfolge. Für das *alignment* des Markers 1 wurde im Gegensatz zu den anderen Markersequenzen vollständige Sequenzabschnitte (*contigs*) verwendet. Aus diesem Grund ist das Fehlen von Sequenzblöcken auf einen Sequenzabbruch im Rahmen des Assemblierungsprozesses der Workbench (*de-novo-assembly*) zurückzuführen. Konnten Fragmente der Rohdaten (*reads*) nicht mehr einem Sequenzabschnitt (*contig*) zugeordnet werden, fehlen diesem Sequenzabschnitt im *alignment* die entsprechenden Sequenzblöcke. Passend

zu dieser Theorie lässt sich das Ergebnis einordnen, dass keine weiteren Insertionen oder Verschiebungen von Sequenzfragmenten zwischen den Referenzen nachgewiesen wurden. Abgesehen von der Insertion des Markers sowie Block 7 ist der entsprechende markerpositive Sequenzabschnitt frei von genetischem *rearrangement* und somit konserviert bis auf einen bevorzugten Punkt für Insertionen innerhalb des Block 1.

#### 5.4.2 ST2-Marker 2

Für den 7.714 bp langen ST2-Marker 2 konnten Homologien zu Faktoren identifiziert werden, die eine Adaption an die Klinikumwelt darstellen (Abb. 4.9). Die ORFs des Markers sind homolog zu Abschnitten des *mer*-Operons eines *S. epidermidis*. Das *mer*-Operon kodiert für die Proteine zur Detoxifikation von Quecksilber. Die Quecksilberresistenz war bis circa 1977 ein Faktor der Anpassung für Bakterienstämme an die Klinikumwelt, da bis zu diesem Zeitpunkt Quecksilberverbindungen als Bestandteil von Desinfektionsmitteln weit verbreitet waren. Nach Abschaffung der Verwendung von Quecksilber zur Flächendesinfektion wurde ein Rückgang resistenter Bakterienpopulationen beobachtet (Porter et al., 1982). Entsprechend der aktuellen Studienlage konnte Staphylokokken eine Neigung zur Bildung von Desinfektionsmittel-Resistenzen nachgewiesen werden (Taheri et al., 2016).

Quecksilberionen sind aus mehreren Gründen für die Bakterienzelle toxisch. Sowohl in ionisierter Form als auch in alkylgebundener Form können sie biologische Membranen passieren. Im Zellinnern schädigen sie in gebundener oder ionisierter Form irreversibel die Nukleinsäuren. Eine weitere Ursache für die Toxizität von Quecksilber ist seine Reaktionsfreudigkeit mit Amino-, Carboxy-, Phosphat- und Imidazolgruppen. Dies bewirkt die Inaktivierung von zentralen Proteinen, wie zum Beispiel der Laktatdehydrogenase oder der Glutathionperoxidase. Zusätzlich besitzen sie eine hohe Affinität zu Cysteinresten der Thiolgruppen von Proteinen. Durch die Bindung des ionisierten Quecksilbers an die Cysteinreste verlieren die betroffenen Proteine ihre Funktion. Entsprechend

resistente Bakterienstämme sind damit in einer Umwelt mit höheren Quecksilberkonzentrationen überlebensfähig und haben einen Wachstumsvorteil gegenüber solchen, die sensibel sind. Aus diesem Grund haben sich das *mer*-Operon sowie weitere Resistenzmechanismen entwickelt, die durch horizontalen Gentransfer in diversen Bakterienarten verbreitet sind. Als Vektoren für die *mer*-Gene konnten Plasmide, Chromosome, Transposons und Integrons nachgewiesen werden (Mathema et al., 2011).

Von dem Gensatz des *mer*-Operons konnten drei der fünf ORFs des Markers entsprechend als homolog zugeordnet werden. ORF 1 wurde Homologie zum *merR*-Gen nachgewiesen. Das Protein MerR reguliert durch Bindung an die Promotor-Operator-Region des *mer*-Operon die Expression sowohl der weiteren Strukturgene als auch die eigene Transkription. Liegt ionisiertes Quecksilber ( $\text{Hg}^{2+}$ ) in der Bakterienzelle vor, kommt es zur Bindung an MerR. Daraufhin ändert sich die Sekundär-, Tertiär- sowie Quartärstruktur des Proteins gefolgt von einer veränderten Konformation der Sequenz der Promotor-Operator-Region. Dies hat eine vermehrte Expression der Strukturgene des *mer*-Operon zur Folge. In Abwesenheit von ionisiertem Quecksilber ( $\text{Hg}^{2+}$ ) wirkt das MerR als Repressor (Chang et al., 2015). ORF 2 ist homolog zu einem hypothetischen Protein des gleichen Gensatzes wie ORF 1. Aus diesem Grund kann angenommen werden, dass ORF 2 ebenfalls für ein Schwermetallresistenz-assoziiertes Protein kodiert. ORF 3 ist homolog zum *merA*-Gen, das für eine Quecksilberreduktase kodiert. Die NAD(P)H-abhängige Quecksilberreduktase MerA gehört zu der Gruppe der Flavinoxidoreduktasen. Sie reduziert die toxischen Quecksilberionen ( $\text{Hg}^{2+}$ ) zu elementarem Quecksilber ( $\text{Hg}^0$ ), das als flüchtiges, weniger toxisches Gas in das äußere Periplasma transportiert wird (Mathema et al., 2011). ORF 4 ist eine Organoquecksilberlyase des *merB*-Gens. Die Organoquecksilberlyase MerB ist ein weiteres entscheidendes Enzym im Entgiftungsprozess von Quecksilber. MerB katalysiert die Protonierung von Kohlenstoff-Quecksilber-Bindungen. Dabei wird das zuvor organisch gebundene Quecksilber in seine ionisierte, weniger toxische, immobilere Form ( $\text{Hg}^{2+}$ ) überführt. Die Kombination aus *merA* und *merB* ist das entscheidende Charakteristikum einer Breitspektrum-Resistenz vermittelt durch das *mer*-Operon. Im Fall einer Breitspektrum-Resistenz liegen die Strukturgene *merRTPAGBDE*

vor. Im Rahmen einer solchen Resistenz können sowohl Quecksilberionen als auch Alkylquecksilberverbindungen entgiftet werden. Im ersten Schritt löst MerB in Alkylverbindungen vorliegendes Quecksilber. Die auf diesem Weg freigesetzten Ionen werden wiederum durch die Quecksilberreduktase MerA zum weniger schädlichen elementaren Quecksilber ( $\text{Hg}^0$ ) reduziert. Die Schmalspektrum-Resistenz ermöglicht ausschließlich die Entgiftung von Quecksilberionen. Sie wird kodiert durch die Strukturgene *merRTPADE* (Mathema et al., 2011). Die sequenzierten, Marker 2-positiven Stämme besitzen demnach Proteine, die homolog sind zu den beiden Schlüsselenzymen sowie einem Regulator der *mer*-Operon kodierten Breitspektrum-Resistenz. Durch die BLAST-Anwendung der Datenbank NCBI konnten der markerflankierenden Sequenz keine weiteren Homologien zu Genen des *mer*-Operon zugeordnet werden. Stattdessen wurden *up*- und *downstream* der Markerinsertion die Transposase sowie invertierte Basenpaar-Wiederholungen (*inverted repeats*) der Insertionssequenz IS431 ermittelt. Für die vollständige Quecksilber-Schmalspektrum-Resistenz fehlen somit die Strukturgene *merT*, *merP*, *merD* sowie *merE*. Neben diesen Strukturgenen fehlt das Gen *merG* für die Breitspektrum-Resistenz (Mathema et al., 2011). Die Markersequenz ist somit nicht Bestandteil eines *mer*-Operons. Die Frage bleibt, ob phänotypisch eine Schmal- oder Breitspektrum-Resistenz trotz Abwesenheit weiterer *mer*-Operonbestandteile ausgeprägt wird. Dies muss durch Anzucht der entsprechenden Marker 2-positiven Stämme mit anschließender Resistenztestung geprüft werden. Sollte eine Resistenz gegen ionisches Quecksilber oder gegen organische Quecksilberverbindungen ermittelt werden, sind zwei mögliche Erklärungen denkbar. Einerseits könnten *mer*-Operonbestandteilen außerhalb der Markersequenz vorliegen. Diese Strukturgene würden die Gene umfassen, die für die Ausbildung einer Schmal- oder Breitspektrum-Resistenz fehlen. Andererseits könnten andere Proteine die Funktion der fehlenden Strukturgene substituieren und eine Quecksilberresistenz mit verändertem Mechanismus vorliegen.

Im *alignment* zeigen vier der fünf Sequenzblöcke (Block 1 bis 4) des markerpositiven Stammes eine mäßige Homologie, jedoch konnte allen fünf Blöcken eine hohe Varianz im Vergleich zu den Sequenzen markernegativer Stämme nachgewiesen werden (Abb. 4.10). Interessanterweise konnten keine

markerflankierenden Sequenzblöcke konserviert in den Gensequenzen marker-negativer Stämme identifiziert werden. Dies ist ein Hinweis darauf, dass die sequenztypspezifische Sequenz größer ist als der ursprünglich extrahierte Marker. Die sequenztypspezifische Sequenz ist in diesem Fall unterbrochen durch die kurzen Blöcke (Block 1 bis 4) mit geringer Homologie zu den Referenzen. Die markerpositive Sequenz zeigt damit das Bild einer Mosaikstruktur. Entsprechend dem Schema der Markersuche ist diese Mosaikstruktur mit den schwachen Homologie verantwortlich für die gewählte Größe des extrahierten Markers sowie dem Ausschluss der weiteren, sequenztypspezifischen Sequenz. Aufgrund des beschriebenen Mosaiks der markerpositiven Sequenz konnte die potentielle Markerinsertionsstelle auf den Referenzen im *alignment* nicht identifiziert werden. Das gewählte Schema der Sequenzauswahl für das *alignment* ist demnach nicht anwendbar für Sequenzen mit einer hohen genetischen Varianz, wie dem Marker 2.

Die sequenztypspezifische Sequenz umfasst damit ebenfalls die Transposase sowie die invertierten Basenpaare der Insertionssequenz IS431. Aufgrund des Vorliegens der flankierenden Insertionssequenzen und der Mosaikstruktur der Sequenz des markerpositiven Stammes entsteht die Theorie, dass der Marker Bestandteil eines mobilen genetischen Elements ist. Insertionssequenzen bestehen aus Basenpaar-Wiederholungen, die eine Transposase einrahmen. Sie gehören zu den kleinen mobilen genetischen Einheiten, die allein aufgrund ihrer Transposase an einer anderen Position der Sequenz inserieren können (Mahillon and Chandler, 1998). Die Markersequenz wird durch zwei Insertionssequenzen IS431 eingerahmt. Dies lässt die Theorie zu, dass der Marker 2 ein zusammengesetztes Transposon (*composite transposon*) genanntes mobiles genetische Element ist. Der Aufbau eines solchen Transposons besteht aus einer Gensequenz, die durch zwei Insertionssequenzen eingerahmt und als solche Einheit im Genom verschoben wird. Der Mechanismus wurde für *S. epidermidis* bereits im Zusammenhang der Insertion eines IS257-tragenden, zusammengesetzten Transposons in einem Plasmid beschrieben: Im ersten Schritt wurde eine eigenständige Kopie der Insertionssequenz IS257 im Rahmen der Transposition in das Plasmid inseriert. Im zweiten Schritt kam es zur Rekombination zwischen der IS257 des Plasmids und der IS257 einer

zirkulär vorliegenden Gensequenz. Durch die Rekombination wurde die zirkuläre Gensequenz in das Plasmid eingebaut. Dort lag die Sequenz eingerahmt von den beiden IS257 vor (Brenciani et al., 2016). Die Insertionssequenz IS431 wurde bereits im Zusammenhang mit dem *rearrangement* sowie dem horizontalen Gentransfer der Regulatorregion des Methicillinresistenz-kodierenden *mecA*-Gens zwischen Staphylokokken beschrieben (Kobayashi et al., 2001). Dies kann mit der Theorie vereinbart werden, dass der Marker 2 als zusammengesetztes Transposon durch einen IS431-vermittelten Transfer in die Gensätze der Marker 2-positiven Stämme gelangt ist. In zukünftigen Sequenzanalysen muss geprüft werden, ob dieses Transposon weiterhin seine Position im Genom ändern kann.

Die verwendeten Referenzsequenzen können anhand ihrer konservierten Sequenzblöcke in zwei Gruppen unterteilt werden. Innerhalb dieser beiden Gruppen ist eine weitere Differenzierung anhand der unterschiedlichen Anordnung der Sequenzblöcke möglich. Die unterschiedliche Anordnung lässt sich in beiden Gruppen auf Translokationen zurückführen. Eine Inversion ließ sich aufgrund der unveränderten Orientierung der beschriebenen Fragmente ausschließen. Neben diesen Translokationen fanden in der ersten Gruppe von Referenzen (A, B, D, F) an den gleichen Positionen innerhalb des Sequenzblocks 6 drei Insertionen mit einer Länge von wenigen Basenpaaren statt. Die Insertionen in Block 6 von drei der Referenzen (A, B, D) sind homolog zueinander, obwohl eine der drei Referenzen (D) aufgrund der Translokation eine veränderte Abfolge von Sequenzblöcken aufweist. Folglich haben die beschriebenen Insertionen unabhängig von der Translokation stattgefunden. Zwischen den Referenzsequenzen der zweiten Gruppe (C, E, G) wurden drei Fragmente in zwei zeitlich versetzten Schritten transloziert. Die verwendeten Referenzen beider Gruppen sind demnach Sequenzen, in denen gehäuft Translokationen auftreten.

### 5.4.3 ST2-Marker 3

Es konnte ein dritter ST2-Marker mit einer Länge von 7.568 bp identifiziert werden. ORF 1 und 2 besitzen Homologien zu dem gleichen Schwermetallresistenz-kodierenden Gensatz des *S. epidermidis*-Stamm NTUH-7684 wie schon Marker 2 (Abb. 4.11). Den beiden ORFs konnten Sequenzen von hypothetischen Proteinen zugeordnet werden. Die Proteine beider ORFs können demnach mit einer Schwermetallresistenz assoziiert sein. Ebenfalls möglich ist, dass beide ORFs Antibiotikaresistenz-assoziierte Proteine kodieren. Antibiotikaresistenz-kodierende Gene werden häufig gemeinsam mit Schwermetallresistenz-assoziierten Genen über horizontalen Gentransfer innerhalb von Bakterienpopulationen weitergegeben (Mathema et al., 2011). Beide ORFs besitzen auf dem Marker die gleiche Orientierung. Dies ist ein Hinweis auf das Vorliegen einer genetischen Einheit, die gekoppelt vererbt wird. ORF 3 ist homolog zu der Sequenz des Kassettenchromosoms eines *S. haemolyticus*. Genkassetten wie die *SCCmec*-Kassette ermöglichen das Verbreiten von Genen in Bakterienpopulationen (Bloemendaal et al., 2010; Wiolders et al., 2001). ORF 3 kodiert für ein Protein mit unklarer Funktion. Weitere Informationen zu der Funktion waren auf der Plattform NCBI nicht gegeben. Eine zukünftige Analyse der Proteinprodukte der ORF 1 bis 3 kann ihre jeweilige Funktion klären. ORF 4 des Markers besitzt eine Teilhomologie zu einer CDP-Glycerol: Poly(glycerophosphat) Glycerophosphotransferase (CGPTase). Für *S. epidermidis* wurde bereits 1999 die Struktur einer CGPTase beschrieben, die durch das Gen *tagF* kodiert wird und für die Synthese von Zellwandteichonsäure (WTA) eine entscheidende Rolle spielt. Das *tagF*-Gen ist in Organismen von *S. epidermidis* nicht gekoppelt an eine Operon-Struktur (Fitzgerald and Foster, 2000). Dies erklärt, weshalb der Marker 3 keine weiteren Homologien zu Genen von Enzymen der WTA-Synthese aufweist.

Die CGPTase des *tagF*-Gens ist ein zytoplasmatisches, membrangebundenes Enzym, das die Bildung der WTA-Polymerstruktur katalysiert. Im beschriebenen Mechanismus wird das Substrat CDP-Glycerol durch Spaltung der Pyrophosphatbindung über eine Phosphodiesterbrücke kovalent an ein Lipidgrundgerüst gebunden. Durch diesen Reaktionsschritt erfolgt eine Erweite-

rung der Polymerkette der WTA. Eine Modifikation des so gewonnenen Polymers kann im weiteren Verlauf erfolgen, zum Beispiel durch Verknüpfung mit Alanin. Am Ende der Prozessierung wird es exportiert, um kovalent durch eine Phosphodiesterbindung an die Mureinsäure des Peptidoglykangerüsts gebunden zu werden. WTA sind Bestandteil der Peptidoglykanschicht aller grampositiver Bakterien. Gemeinsam mit den membrangebundenen Lipoteichonsäuren (LTA) dienen sie der Adhäsion sowie Kolonisierung von Oberflächen und werden für die Ausbildung von Biofilmen benötigt. WTA und LTA haben einen regulatorischen Einfluss auf Autolysinvorgänge. Beide Stoffgruppen stellen einen Schutz vor äußeren Einflüssen dar. Zu diesen äußeren Einflüssen zählen unter anderem Lyse durch Proteine der unspezifischen Immunabwehr. Der WTA von *S. aureus* wurde in diesem Zusammenhang ein Schutz vor antibiotisch wirksamen Fettsäuren der menschlichen Talgdrüsen zugeschrieben. Aufgrund ihrer Position an der Peptidoglykanschicht beeinflussen beide Arten der Teichonsäure Rigidität und Durchlässigkeit des bakteriellen Zelläußeren. Zusätzlich bieten sie Schutz vor Bakteriophagen. Dies wird durch bestimmte Phagen deutlich, die Enzyme besitzen, die ihnen das Eindringen durch die WTA in das entsprechende Bakterium erleichtern (Lovering et al., 2010).

Des weiteren konnte für *S. aureus* ein Zusammenhang zwischen einer WTA-reicheren, dickeren Zellwand und einer abnehmenden Empfindlichkeit für Daptomycin beschrieben werden. Aufgrund der höheren WTA-Konzentration besaßen die resistenten Stämme eine höhere, positive Ladung der Zellwand. Die Daptomycin-Resistenz wurde von Bertsche et. al. auf zwei mögliche Mechanismen zurückgeführt. Der erste Mechanismus ist der ladungsabhängige Efflux von Calciumkomplex-gebundenem Daptomycin. Der zweite Mechanismus stellt die verminderte Penetration des Antibiotikums der Bakterienzelle aufgrund von sterischen Wechselwirkungen dar (Bertsche et al., 2013). Eine Zellwand mit einem größeren WTA-Anteil ist demnach ein Pathogenitätsfaktor. Auf der Markersequenz wurde allerdings nur eine Teilhomologie zu einer CPGTase identifiziert. Die Frage bleibt, ob diese Teilhomologie für die phänotypische Ausprägung einer WTA-reicheren Zellwand ausreicht. Um dies zu prüfen, muss der WTA-Anteil der Zellwand zwischen Marker 3-positiven und -negativen Stämmen in einer zukünftigen Studie verglichen werden.

Im durchgeführten *alignment* (Abb. 4.12) konnte nicht die potentielle Markerinsertionsstelle in den Referenzsequenzen identifiziert werden, da die beiden markerflankierenden Sequenzblöcke auf jeweils unterschiedlichen Referenzen getrennt vorliegen. Folglich ist der markerpositive Sequenzabschnitt ein sehr variabler Bereich, in dem ein hohes Maß an genetischem *rearrangement* stattfindet. Ähnlich wie für ST2-Marker 2 wird die möglicherweise längere, sequenztypspezifische Sequenz *upstream* durchbrochen von einem kurzen homologen Sequenzblock (Block 1). Zur Überprüfung dieser Theorie kann die Sequenz *upstream* dieses Blocks auf Homologien zu ST2-Genomen und solchen mit abweichendem Sequenztyp untersucht werden. Diese Analyse war im zeitlichen Rahmen dieser Arbeit nicht mehr möglich. Für die Durchführung des PCR-Verfahrens zum Nachweis des Sequenztyps 2 ist die extrahierte Markersequenz mit der Länge von 7.568 bp jedoch ausreichend.

Ein kurzer Sequenzblock (Block 2) wurde sowohl innerhalb des Markers als auch in der zweiten Kategorie markernegativer Referenzen (A, B, D, E, F) identifiziert. Dieser Block kann aufgrund der Länge weniger Basen (144 bp) sowie des geringen Grads an Homologie als Artefakt gewertet werden.

Anhand der Abfolge ihrer Fragmente aus den gleichen Sequenzblöcken lassen sich die fünf Referenzen der identifizierten Gruppe (A, B, D, E, F) in zwei Subgruppen dividieren. Die Verschiebung der Fragmente ist auf zwei Translokationen zurückzuführen. Eine Inversion ist aufgrund der gleich bleibenden Orientierung der Fragmente unwahrscheinlich. Innerhalb der Sequenzblöcke der Gruppe markernegativer Referenzen (A, B, D, E, F) liegen je drei weitere Insertionen mit einer Länge weniger Basenpaare an jeweils gleicher Position vor. Diese Insertionen differieren teilweise zwischen den fünf Referenzsequenzen. Die unterschiedlichen Insertionen an gleicher Position liefern einen Hinweis auf mikroevolutionäre Prozesse, die die Referenzsequenzen voneinander unterscheiden.

## 5.5 Ausblick

Im Verlauf der Erstellung dieser Arbeit traten Fragen auf, die weitere Abklärung im Rahmen zukünftiger Untersuchungen benötigen. Der Einfluss des ST5-Markers auf die Pathogenität konnte anhand der identifizierten Homologien nicht eingegrenzt werden. Da nur ein Marker für eine schleswig-holsteinische Subgruppe von klinischen ST59-Stämmen identifiziert werden konnte, bleibt die Frage bestehen, ob ein überregionaler ST59-Marker identifizierbar ist. Eine Analyse der Proteinprodukte aller identifizierten Marker verspricht mehr Informationen über die kodierten Proteine und ihren gegebenenfalls vorliegenden Einfluss auf die Adaptation in der Klinikumwelt. Insbesondere der ST2 Marker 2 mit seinen Strukturgenen des *mer*-Operon bietet einen interessanten Ansatzpunkt für weiterführende Forschung. Weitere Sequenzanalysen sind nötig, um die Promotorstrukturen der vermuteten Operon-Transkriptionseinheiten der Markersequenzen näher zu beleuchten. Weiterhin konnten durch die Varianz-Analysen der markerpositiven Sequenzen des ST59-Markers und des ST2-Marker 3 mehrere bevorzugte Insertionspunkte identifiziert werden. Weitere Untersuchungen können Hinweise liefern auf eventuell vorliegende mikroevolutionäre Prozesse.

## 6. Zusammenfassung

In der klinischen Diagnostik besteht die Schwierigkeit, zwischen kommensalen und pathogenen *S. epidermidis* zu unterscheiden. Die klassischen Virulenzfaktoren der primären Adhäsion sowie Akkumulation im Rahmen der Biofilmbildung haben sich als nicht ausreichend für die Differenzierung herausgestellt. Die MLST-Sequenztypen 2, 5 sowie 59 konnten gehäuft in Blutkulturen nachgewiesen werden. *S. epidermidis* mit jenen Sequenztypen liegen jedoch ebenfalls in der Umwelt vor, weswegen der Sequenztyp nicht für die Identifikation pathogener *S. epidermidis* ausreicht. Um eine sequenztypspezifische Markergensequenz zu ermitteln, wurde eine subtrahierende Sequenzanalyse durchgeführt. Je ein Marker für die MLST-Sequenztypen 5 sowie 59 und drei Marker für den MLST-Sequenztyp 2 konnten identifiziert werden. Diese fünf Marker treten innerhalb der Stichprobe aller verwendeter Isolate sowie innerhalb der analysierten Gruppe von pathogenen Isolaten jeweils signifikant häufiger in den Stämmen mit entsprechendem Sequenztyp auf. Folglich identifizieren die Markersequenzen zuverlässig das Merkmal Sequenztyp. Jedoch konnten nur die drei ST2-Marker mit dem Merkmal Pathogenität in der Stichprobe aller verwendeten Isolate korreliert werden. Ausschließlich ST2-Marker 1 und 3 sind für die Differenzierung zwischen kommensalen Patienten-Isolaten und pathogenen Stämmen geeignet. Für den Sequenztyp 5 und 59 konnten jeweils auf den Markern nur Homologien zu Sequenzen hypothetischer Proteine identifiziert werden. Eine Aussage zur möglichen Funktion der beiden Marker war aus diesem Grund nicht möglich. Für die ST2-Marker konnte durch die Analyse der kodierten Proteine Folgendes festgestellt werden: Der ST2-Marker 1 besteht aus mehreren Operoneinheiten, deren Strukturgene unter anderem für Proteine der Zellwandzusammensetzung und von Sekretionssystemen kodieren. Die Strukturgene des ST2-Marker 2 wurden dem Quecksilberresistenz-kodierenden *mer*-Operon zugeordnet. Der ST2-Marker 3 kodiert für ein wichtiges Enzym der Zellwandteichonsäure-Synthese. In Studien wurde eine Korrelation zwischen Zellwandanteil von Zellwandteichonsäure mit erhöhter Antibiotikaresistenz festgestellt. Diese kodierten Faktoren betreffen somit Oberflächenstrukturen und Resistenzen der markerpositiven Stämme, die Einfluss auf die Adaption an die Klinikumwelt besitzen.

## 7. Literaturverzeichnis

Aanensen, D.M., and Spratt, B.G. (2005). The multilocus sequence typing network: mlst.net. *Nucleic Acids Res.* 33, W728–W733.

Ahlstrand, E., Hellmark, B., Svensson, K., and Söderquist, B. (2014). Long-Term Molecular Epidemiology of *Staphylococcus epidermidis* Blood Culture Isolates from Patients with Hematological Malignancies. *PLoS ONE* 9.

Anguera, I., Del Río, A., Miró, J.M., Matínez-Lacasa, X., Marco, F., Gumá, J.R., Quaglio, G., Claramonte, X., Moreno, A., Mestres, C.A., et al. (2005). *Staphylococcus lugdunensis* infective endocarditis: description of 10 cases and analysis of native valve, prosthetic valve, and pacemaker lead endocarditis clinical profiles. *Heart Br. Card. Soc.* 91, e10.

Archer, G.L. (1991). Alteration of cutaneous staphylococcal flora as a consequence of antimicrobial prophylaxis. *Rev. Infect. Dis.* 13 Suppl 10, S805-809.

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

Becker, K., Heilmann, C., and Peters, G. (2014). Coagulase-Negative *Staphylococci*. *Clin. Microbiol. Rev.* 27, 870–926.

de Been, M., Pinholt, M., Top, J., Bletz, S., Mellmann, A., van Schaik, W., Brouwer, E., Rogers, M., Kraat, Y., Bonten, M., et al. (2015). Core Genome Multilocus Sequence Typing Scheme for High- Resolution Typing of *Enterococcus faecium*. *J. Clin. Microbiol.* 53, 3788–3797.

Bender, J., Strommenger, B., Steglich, M., Zimmermann, O., Fenner, I., Lensing, C., Dagwadordsch, U., Kekulé, A.S., Werner, G., and Layer, F. (2015). Linezolid resistance in clinical isolates of *Staphylococcus epidermidis* from German hospitals and characterization of two cfr-carrying plasmids. *J. Antimicrob. Chemother.* 70, 1630–1638.

Bertsche, U., Yang, S.-J., Kuehner, D., Wanner, S., Mishra, N.N., Roth, T., Nega, M., Schneider, A., Mayer, C., Grau, T., et al. (2013). Increased Cell Wall Teichoic Acid Production and D-alanylation Are Common Phenotypes among Daptomycin-Resistant Methicillin-Resistant *Staphylococcus aureus* (MRSA) Clinical Isolates. *PLoS ONE* 8.

Bloemendaal, A.L.A., Brouwer, E.C., and Fluit, A.C. (2010). Methicillin Resistance Transfer from *Staphylococcus epidermidis* to Methicillin-Susceptible *Staphylococcus aureus* in a Patient during Antibiotic Therapy. *PLoS ONE* 5.

Bougnoux, M.-E., Aanensen, D.M., Morand, S., Théraud, M., Spratt, B.G., and d'Enfert, C. (2004). Multilocus sequence typing of *Candida albicans*: strategies, data exchange and applications. *Infect. Genet. Evol. J. Mol. Epidemiol. Evol. Genet. Infect. Dis.* 4, 243–252.

Bowden, M.G., Visai, L., Longshaw, C.M., Holland, K.T., Speziale, P., and Hook, M. (2002). Is the GehD lipase from *Staphylococcus epidermidis* a collagen binding adhesin? *J. Biol. Chem.* 277, 43017–43023.

- Bowden, M.G., Chen, W., Singvall, J., Xu, Y., Peacock, S.J., Valtulina, V., Speziale, P., and Höök, M. (2005). Identification and preliminary characterization of cell-wall-anchored proteins of *Staphylococcus epidermidis*. *Microbiol. Read. Engl.* *151*, 1453–1464.
- Bradford, R., Abdul Manan, R., Daley, A.J., Pearce, C., Ramalingam, A., D'Mello, D., Mueller, Y., Uahwatanasakul, W., Qu, Y., Grando, D., et al. (2006). Coagulase-negative staphylococci in very-low-birth-weight infants: inability of genetic markers to distinguish invasive strains from blood culture contaminants. *Eur. J. Clin. Microbiol. Infect. Dis. Off. Publ. Eur. Soc. Clin. Microbiol.* *25*, 283–290.
- Brenciani, A., Morroni, G., Mingoia, M., Varaldo, P.E., and Giovanetti, E. (2016). Stability of the cargo regions of the cfr-carrying, multiresistance plasmid pSP01 from *Staphylococcus epidermidis*. *Int. J. Med. Microbiol. IJMM* *306*, 717–721.
- Carter, A.T., Stringer, S.C., Webb, M.D., and Peck, M.W. (2013). The Type F6 Neurotoxin Gene Cluster Locus of Group II *Clostridium botulinum* Has Evolved by Successive Disruption of Two Different Ancestral Precursors. *Genome Biol. Evol.* *5*, 1032–1037.
- Chang, C.-C., Lin, L.-Y., Zou, X.-W., Huang, C.-C., and Chan, N.-L. (2015). Structural basis of the mercury(II)-mediated conformational switching of the dual-function transcriptional regulator MerR. *Nucleic Acids Res.* *43*, 7612–7623.
- Conlon, K.M., Humphreys, H., and O'Gara, J.P. (2002). *icaR* Encodes a Transcriptional Repressor Involved in Environmental Regulation of *ica* Operon Expression and Biofilm Formation in *Staphylococcus epidermidis*. *J. Bacteriol.* *184*, 4400–4408.
- Cucarella, C., Solano, C., Valle, J., Amorena, B., Lasa, Í., and Penadés, J.R. (2001). Bap, a *Staphylococcus aureus* Surface Protein Involved in Biofilm Formation. *J. Bacteriol.* *183*, 2888–2896.
- Darling, A.C.E., Mau, B., Blattner, F.R., and Perna, N.T. (2004). Mauve: Multiple Alignment of Conserved Genomic Sequence With Rearrangements. *Genome Res.* *14*, 1394–1403.
- Deplano, A., Vandendriessche, S., Nonhoff, C., Dodémont, M., Roisin, S., and Denis, O. (2016). National surveillance of *Staphylococcus epidermidis* recovered from bloodstream infections in Belgian hospitals. *J. Antimicrob. Chemother.* *71*, 1815–1819.
- Dingle, K.E., Colles, F.M., Wareing, D.R., Ure, R., Fox, A.J., Bolton, F.E., Bootsma, H.J., Willems, R.J., Urwin, R., and Maiden, M.C. (2001). Multilocus sequence typing system for *Campylobacter jejuni*. *J. Clin. Microbiol.* *39*, 14–23.
- Dréno, B., Araviiskaia, E., Berardesca, E., Gontijo, G., Sanchez Viera, M., Xiang, L.F., Martin, R., and Bieber, T. (2016). Microbiome in healthy skin, update for dermatologists. *J. Eur. Acad. Dermatol. Venereol. JEADV* *30*, 2038–2047.
- Earl, D., Bradnam, K., St. John, J., Darling, A., Lin, D., Fass, J., Yu, H.O.K., Buffalo, V., Zerbino, D.R., Diekhans, M., et al. (2011). Assemblathon 1: A competitive assessment of de novo short read assembly methods. *Genome Res.* *21*, 2224–2241.
- Enright, M.C., Day, N.P.J., Davies, C.E., Peacock, S.J., and Spratt, B.G. (2000). Multilocus Sequence Typing for Characterization of Methicillin-Resistant and Methicillin-Susceptible Clones of *Staphylococcus aureus*. *J. Clin. Microbiol.* *38*, 1008–1015.

Figueiredo, A.M.S., and Ferreira, F.A. (2014). The multifaceted resources and microevolution of the successful human and animal pathogen methicillin-resistant *Staphylococcus aureus*. *Mem. Inst. Oswaldo Cruz* 109, 265–278.

Fitzgerald, S.N., and Foster, T.J. (2000). Molecular Analysis of the tagF Gene, Encoding CDP-Glycerol:Poly(glycerophosphate) Glycerophosphotransferase of *Staphylococcus epidermidis* ATCC 14990. *J. Bacteriol.* 182, 1046–1052.

Fortier, L.-C., and Sekulovic, O. (2013). Importance of prophages to evolution and virulence of bacterial pathogens. *Virulence* 4, 354–365.

Foster, T. (1996). *Staphylococcus*. In *Medical Microbiology*, S. Baron, ed. (Galveston (TX): University of Texas Medical Branch at Galveston), p.

Gerke, C., Kraft, A., Süßmuth, R., Schweitzer, O., and Götz, F. (1998). Characterization of the N-acetylglucosaminyltransferase activity involved in the biosynthesis of the *Staphylococcus epidermidis* polysaccharide intercellular adhesin. *J. Biol. Chem.* 273, 18586–18593.

Gill, S.R., Fouts, D.E., Archer, G.L., Mongodin, E.F., DeBoy, R.T., Ravel, J., Paulsen, I.T., Kolonay, J.F., Brinkac, L., Beanan, M., et al. (2005). Insights on Evolution of Virulence and Resistance from the Complete Genome Analysis of an Early Methicillin-Resistant *Staphylococcus aureus* Strain and a Biofilm-Producing Methicillin-Resistant *Staphylococcus epidermidis* Strain. *J. Bacteriol.* 187, 2426–2438.

Gominet, M., Compain, F., Beloin, C., and Lebeaux, D. (2017). Central venous catheters and biofilms: where do we stand in 2017? *APMIS Acta Pathol. Microbiol. Immunol. Scand.* 125, 365–375.

Grice, E.A., and Segre, J.A. (2011). The skin microbiome. *Nat. Rev. Microbiol.* 9, 244–253.

Grice, E.A., Kong, H.H., Conlan, S., Deming, C.B., Davis, J., Young, A.C., Bouffard, G.G., Blakesley, R.W., Murray, P.R., Green, E.D., et al. (2009). Topographical and Temporal Diversity of the Human Skin Microbiome. *Science* 324, 1190–1192.

Hartford, O., O'Brien, L., Schofield, K., Wells, J., and Foster, T.J. (2001). The Fbe (SdrG) protein of *Staphylococcus epidermidis* HB promotes bacterial adherence to fibrinogen. *Microbiol. Read. Engl.* 147, 2545–2552.

Hébert, G.A., Crowder, C.G., Hancock, G.A., Jarvis, W.R., and Thornsberry, C. (1988). Characteristics of coagulase-negative staphylococci that help differentiate these species and other members of the family Micrococcaceae. *J. Clin. Microbiol.* 26, 1939–1949.

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

Heilmann, C., Thumm, G., Chhatwal, G.S., Hartleib, J., Uekötter, A., and Peters, G. (2003). Identification and characterization of a novel autolysin (Aae) with adhesive properties from *Staphylococcus epidermidis*. *Microbiol. Read. Engl.* 149, 2769–2778.

Hercun, M. (2011). Epidemiologie, Prävalenz Virulenz-assoziiierter Gene und klonale Verwandtschaften von *S. epidermidis*. Lübeck.

- Higgins, P.G., Prior, K., Harmsen, D., and Seifert, H. (2017). Development and evaluation of a core genome multilocus typing scheme for whole-genome sequence-based typing of *Acinetobacter baumannii*. *PLoS ONE* 12.
- Hiramatsu, K., Ito, T., Tsubakishita, S., Sasaki, T., Takeuchi, F., Morimoto, Y., Katayama, Y., Matsuo, M., Kuwahara-Arai, K., Hishinuma, T., et al. (2013). Genomic Basis for Methicillin Resistance in *Staphylococcus aureus*. *Infect. Chemother.* 45, 117–136.
- Hof, H., Dörries, R., Geginat, G., Schlüter, D., and Wendl, C. (2014). *Duale Reihe - Medizinische Mikrobiologie* (Stuttgart: Georg Thieme).
- Hussain, M., Heilmann, C., Peters, G., and Herrmann, M. (2001). Teichoic acid enhances adhesion of *Staphylococcus epidermidis* to immobilized fibronectin. *Microb. Pathog.* 31, 261–270.
- Iorio, N.L.P., Caboclo, R.F., Azevedo, M.B., Barcellos, A.G., Neves, F.P.G., Domingues, R.M.C.P., and dos Santos, K.R.N. (2012). Characteristics related to antimicrobial resistance and biofilm formation of widespread methicillin-resistant *Staphylococcus epidermidis* ST2 and ST23 lineages in Rio de Janeiro hospitals, Brazil. *Diagn. Microbiol. Infect. Dis.* 72, 32–40.
- Ito, Y., Kawai, Y., Arakawa, K., Honme, Y., Sasaki, T., and Saito, T. (2009). Conjugative Plasmid from *Lactobacillus gasseri* LA39 That Carries Genes for Production of and Immunity to the Circular Bacteriocin Gassericin A. *Appl. Environ. Microbiol.* 75, 6340–6351.
- Jolley, K.A., Chan, M.-S., and Maiden, M.C. (2004). mlstdbNet – distributed multi-locus sequence typing (MLST) databases. *BMC Bioinformatics* 5, 86.
- Jolley, K.A., Hill, D.M.C., Bratcher, H.B., Harrison, O.B., Feavers, I.M., Parkhill, J., and Maiden, M.C.J. (2012). Resolution of a Meningococcal Disease Outbreak from Whole-Genome Sequence Data with Rapid Web-Based Analysis Methods. *J. Clin. Microbiol.* 50, 3046–3053.
- Karlowsky, J.A., Jones, M.E., Draghi, D.C., Thornsberry, C., Sahm, D.F., and Volturo, G.A. (2004). Prevalence and antimicrobial susceptibilities of bacteria isolated from blood cultures of hospitalized patients in the United States in 2002. *Ann. Clin. Microbiol. Antimicrob.* 3, 7.
- Kline, K.A., Ingersoll, M.A., Nielsen, H.V., Sakinc, T., Henriques-Normark, B., Gattermann, S., Caparon, M.G., and Hultgren, S.J. (2010). Characterization of a Novel Murine Model of *Staphylococcus saprophyticus* Urinary Tract Infection Reveals Roles for Ssp and Sdrl in Virulence. *Infect. Immun.* 78, 1943–1951.
- Kloos, W.E., and Bannerman, T.L. (1994). Update on clinical significance of coagulase-negative staphylococci. *Clin. Microbiol. Rev.* 7, 117–140.
- Kloos, W.E., and Musselwhite, M.S. (1975). Distribution and Persistence of *Staphylococcus* and *Micrococcus* Species and Other Aerobic Bacteria on Human Skin. *Appl. Microbiol.* 30, 381–395.
- Kobayashi, N., Alam, M.M., and Urasawa, S. (2001). Genomic Rearrangement of the *mec* Regulator Region Mediated by Insertion of IS431 in Methicillin-Resistant *Staphylococci*. *Antimicrob. Agents Chemother.* 45, 335–338.

- Kohl, T.A., Diel, R., Harmsen, D., Rothgänger, J., Walter, K.M., Merker, M., Weniger, T., and Niemann, S. (2014). Whole-genome-based *Mycobacterium tuberculosis* surveillance: a standardized, portable, and expandable approach. *J. Clin. Microbiol.* 52, 2479–2486.
- Köser, C.U., Holden, M.T.G., Ellington, M.J., Cartwright, E.J.P., Brown, N.M., Ogilvy-Stuart, A.L., Hsu, L.Y., Chewapreecha, C., Croucher, N.J., Harris, S.R., et al. (2012). Rapid Whole-Genome Sequencing for Investigation of a Neonatal MRSA Outbreak. *N. Engl. J. Med.* 366, 2267–2275.
- Kuroda, M., Ohta, T., Uchiyama, I., Baba, T., Yuzawa, H., Kobayashi, I., Cui, L., Oguchi, A., Aoki, K., Nagai, Y., et al. (2001). Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*. *Lancet Lond. Engl.* 357, 1225–1240.
- Labandeira-Rey, M., Couzon, F., Boisset, S., Brown, E.L., Bes, M., Benito, Y., Barbu, E.M., Vazquez, V., Höök, M., Etienne, J., et al. (2007). *Staphylococcus aureus* Panton-Valentine leukocidin causes necrotizing pneumonia. *Science* 315, 1130–1133.
- Lang, S., Livesley, M.A., Lambert, P.A., Littler, W.A., and Elliott, T.S.J. (2000). Identification of a novel antigen from *Staphylococcus epidermidis*. *FEMS Immunol. Med. Microbiol.* 29, 213–220.
- Leopold, S.R., Goering, R.V., Witten, A., Harmsen, D., and Mellmann, A. (2014). Bacterial Whole-Genome Sequencing Revisited: Portable, Scalable, and Standardized Analysis for Typing and Detection of Virulence and Antibiotic Resistance Genes. *J. Clin. Microbiol.* 52, 2365–2370.
- Lewis, K. (2007). Persister cells, dormancy and infectious disease. *Nat. Rev. Microbiol.* 5, 48–56.
- Ljungh, A., and Wadström, T. (1995). Growth conditions influence expression of cell surface hydrophobicity of staphylococci and other wound infection pathogens. *Microbiol. Immunol.* 39, 753–757.
- Lovering, A.L., Lin, L.Y.-C., Sewell, E.W., Spreter, T., Brown, E.D., and Strynadka, N.C.J. (2010). Structure of the bacterial teichoic acid polymerase TagF provides insights into membrane association and catalysis. *Nat. Struct. Mol. Biol.* 17, 582–589.
- Mack, D., Fischer, W., Krokotsch, A., Leopold, K., Hartmann, R., Egge, H., and Laufs, R. (1996). The intercellular adhesin involved in biofilm accumulation of *Staphylococcus epidermidis* is a linear beta-1,6-linked glucosaminoglycan: purification and structural analysis. *J. Bacteriol.* 178, 175–183.
- Mahieu, L.M., De Dooy, J.J., Lenaerts, A.E., Ieven, M.M., and De Muynck, A.O. (2001). Catheter manipulations and the risk of catheter-associated bloodstream infection in neonatal intensive care unit patients. *J. Hosp. Infect.* 48, 20–26.
- Mahillon, J., and Chandler, M. (1998). Insertion sequences. *Microbiol. Mol. Biol. Rev. MMBR* 62, 725–774.
- Maiden, M.C.J., Bygraves, J.A., Feil, E., Morelli, G., Russell, J.E., Urwin, R., Zhang, Q., Zhou, J., Zurth, K., Caugant, D.A., et al. (1998). Multilocus sequence typing: A portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc. Natl. Acad. Sci. U. S. A.* 95, 3140–3145.

Maiden, M.C.J., Jansen van Rensburg, M.J., Bray, J.E., Earle, S.G., Ford, S.A., Jolley, K.A., and McCarthy, N.D. (2013). MLST revisited: the gene-by-gene approach to bacterial genomics. *Nat. Rev. Microbiol.* **11**, 728–736.

Mathema, V.B., Thakuri, B.C., and Sillanpää, M. (2011). Bacterial mer operon-mediated detoxification of mercurial compounds: a short review. *Arch. Microbiol.* **193**, 837–844.

Mellmann, A., Bletz, S., Böking, T., Kipp, F., Becker, K., Schultes, A., Prior, K., and Harmsen, D. (2016). Real-Time Genome Sequencing of Resistant Bacteria Provides Precision Infection Control in an Institutional Setting. *J. Clin. Microbiol.* **54**, 2874–2881.

Mellmann, A., Andersen, P.S., Bletz, S., Friedrich, A.W., Kohl, T.A., Lilje, B., Niemann, S., Prior, K., Rossen, J.W., and Harmsen, D. (2017). High Interlaboratory Reproducibility and Accuracy of Next-Generation-Sequencing-Based Bacterial Genotyping in a Ring Trial. *J. Clin. Microbiol.* **55**, 908–913.

Mendes, R.E., Deshpande, L.M., Costello, A.J., and Farrell, D.J. (2012). Molecular Epidemiology of *Staphylococcus epidermidis* Clinical Isolates from U.S. Hospitals. *Antimicrob. Agents Chemother.* **56**, 4656–4661.

Mermel, L.A., Allon, M., Bouza, E., Craven, D.E., Flynn, P., O'Grady, N.P., Raad, I.I., Rijnders, B.J.A., Sherertz, R.J., and Warren, D.K. (2009). Clinical Practice Guidelines for the Diagnosis and Management of Intravascular Catheter-Related Infection: 2009 Update by the Infectious Diseases Society of America. *Clin. Infect. Dis.* **49**, 1–45.

Mims, C., Dockrell, H.M., Goering, R.V., Roitt, I., Wakelin, D., and Zuckerman, M. (2006). *Medizinische Mikrobiologie - Infektiologie mit Virologie und Immunologie* (München: Elsevier).

Miragaia, M., Thomas, J.C., Couto, I., Enright, M.C., and de Lencastre, H. (2007). Inferring a population structure for *Staphylococcus epidermidis* from multilocus sequence typing data. *J. Bacteriol.* **189**, 2540–2552.

Nilsson, M., Frykberg, L., Flock, J.-I., Pei, L., Lindberg, M., and Guss, B. (1998). A Fibrinogen-Binding Protein of *Staphylococcus epidermidis*. *Infect. Immun.* **66**, 2666–2673.

Oliveira, C.F. de, Cavanagh, J.P., Fredheim, E.G.A., Reiter, K.C., Rieger, A., Klingenberg, C., d'Azevedo, P.A., and Sollid, J.E. (2016). Coagulase-negative staphylococci in Southern Brazil: looking toward its high diversity. *Rev. Soc. Bras. Med. Trop.* **49**, 292–299.

Ortega-Peña, S., Colín-Castro, C., Hernández-Duran, M., López-Jácome, E., and Franco-Cendejas, R. (2015). [Microbiological characteristics and patterns of resistance in prosthetic joint infections in a referral hospital]. *Cir. Cir.* **83**, 371–377.

Otto, M. (2008). Staphylococcal biofilms. *Curr. Top. Microbiol. Immunol.* **322**, 207–228.

Otto, M. (2013). Staphylococcal infections: mechanisms of biofilm maturation and detachment as critical determinants of pathogenicity. *Annu. Rev. Med.* **64**, 175–188.

Park, B., Nizet, V., and Liu, G.Y. (2008). Role of *Staphylococcus aureus* Catalase in Niche Competition against *Streptococcus pneumoniae*. *J. Bacteriol.* **190**, 2275–2278.

- Pektaş, A., Kara, A., and Gurgey, A. (2015a). Cohort Study: Central Venous Catheter-Related Complications in Children with Hematologic Diseases at a Single Center. *Turk. J. Hematol.* 32, 144–151.
- Pektaş, A., Kara, A., and Gurgey, A. (2015b). Cohort Study: Central Venous Catheter-Related Complications in Children with Hematologic Diseases at a Single Center. *Turk. J. Hematol.* 32, 144–151.
- Perez-Cheeks, B.A., Lee, C., Hayama, R., and Mariani, K.J. (2012). A Role for Topoisomerase III in *Escherichia coli* Chromosome Segregation. *Mol. Microbiol.* 86, 1007–1022.
- Personne, P., Bes, M., Lina, G., Vandenesch, F., Brun, Y., and Etienne, J. (1997). Comparative performances of six agglutination kits assessed by using typical and atypical strains of *Staphylococcus aureus*. *J. Clin. Microbiol.* 35, 1138–1140.
- Petit, J.F., Strominger, J.L., and Söll, D. (1968). Biosynthesis of the peptidoglycan of bacterial cell walls. VII. Incorporation of serine and glycine into interpeptide bridges in *Staphylococcus epidermidis*. *J. Biol. Chem.* 243, 757–767.
- Petti, C.A., Simmon, K.E., Miro, J.M., Hoen, B., Marco, F., Chu, V.H., Athan, E., Bukovski, S., Bouza, E., Bradley, S., et al. (2008). Genotypic Diversity of Coagulase-Negative *Staphylococci* Causing Endocarditis: a Global Perspective. *J. Clin. Microbiol.* 46, 1780–1784.
- Porter, F.D., Silver, S., Ong, C., and Nakahara, H. (1982). Selection for mercurial resistance in hospital settings. *Antimicrob. Agents Chemother.* 22, 852–858.
- Quainoo, S., Coolen, J.P.M., van Hijum, S.A.F.T., Huynen, M.A., Melchers, W.J.G., van Schaik, W., and Wertheim, H.F.L. (2017). Whole-Genome Sequencing of Bacterial Pathogens: the Future of Nosocomial Outbreak Analysis. *Clin. Microbiol. Rev.* 30, 1015–1063.
- Raad, I., Alrahwani, A., and Rolston, K. (1998). *Staphylococcus epidermidis*: emerging resistance and need for alternative agents. *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.* 26, 1182–1187.
- Rangrez, A.Y., Abajy, M.Y., Keller, W., Shouche, Y., and Grohmann, E. (2010). Biochemical characterization of three putative ATPases from a new type IV secretion system of *Aeromonas veronii* plasmid pAC3249A. *BMC Biochem.* 11, 10.
- Rohde, H., Kalitzky, M., Kröger, N., Scherpe, S., Horstkotte, M.A., Knobloch, J.K.-M., Zander, A.R., and Mack, D. (2004). Detection of virulence-associated genes not useful for discriminating between invasive and commensal *Staphylococcus epidermidis* strains from a bone marrow transplant unit. *J. Clin. Microbiol.* 42, 5614–5619.
- Rohde, H., Burdelski, C., Bartscht, K., Hussain, M., Buck, F., Horstkotte, M.A., Knobloch, J.K.-M., Heilmann, C., Herrmann, M., and Mack, D. (2005). Induction of *Staphylococcus epidermidis* biofilm formation via proteolytic processing of the accumulation-associated protein by staphylococcal and host proteases. *Mol. Microbiol.* 55, 1883–1895.
- Saffari, F., Widerström, M., Gurram, B.K., Edebro, H., Hojabri, Z., and Monsen, T. (2016). Molecular and Phenotypic Characterization of Multidrug-Resistant Clones of *Staphylococcus epidermidis* in Iranian Hospitals: Clonal Relatedness to Healthcare-Associated Methicillin-Resistant Isolates in Northern Europe. *Microb. Drug Resist.* 22, 570–577.

- Saginur, R., StDenis, M., Ferris, W., Aaron, S.D., Chan, F., Lee, C., and Ramotar, K. (2006). Multiple Combination Bactericidal Testing of Staphylococcal Biofilms from Implant-Associated Infections. *Antimicrob. Agents Chemother.* 50, 55–61.
- Salgueiro, V.C., Iorio, N.L.P., Ferreira, M.C., Chamon, R.C., and dos Santos, K.R.N. (2017). Methicillin resistance and virulence genes in invasive and nasal *Staphylococcus epidermidis* isolates from neonates. *BMC Microbiol.* 17.
- Sharma, P., Satorius, A.E., Raff, M.R., Rivera, A., Newton, D.W., and Younger, J.G. (2014). Multilocus Sequence Typing for Interpreting Blood Isolates of *Staphylococcus epidermidis*. *Interdiscip. Perspect. Infect. Dis.* 2014.
- Soroush, S., Jabalameli, F., Taherikalani, M., Amirmozafari, N., Fooladi, A.A.I., Asadollahi, K., Beigverdi, R., and Emaneini, M. (2016). Investigation of biofilm formation ability, antimicrobial resistance and the staphylococcal cassette chromosome mec patterns of methicillin resistant *Staphylococcus epidermidis* with different sequence types isolated from children. *Microb. Pathog.* 93, 126–130.
- Spanu, T., Sanguinetti, M., Ciccaglione, D., D'Inzeo, T., Romano, L., Leone, F., and Fadda, G. (2003). Use of the VITEK 2 System for Rapid Identification of Clinical Isolates of Staphylococci from Bloodstream Infections. *J. Clin. Microbiol.* 41, 4259–4263.
- Sperber, W.H., and Tatini, S.R. (1975). Interpretation of the Tube Coagulase Test for Identification of *Staphylococcus aureus*. *Appl. Microbiol.* 29, 502–505.
- Spirig, T., Weiner, E.M., and Clubb, R.T. (2011). Sortase enzymes in Gram-positive bacteria. *Mol. Microbiol.* 82, 1044–1059.
- Taheri, N., Ardebili, A., Amouzandeh-Nobaveh, A., and Ghaznavi-Rad, E. (2016). Frequency of Antiseptic Resistance Among *Staphylococcus aureus* and Coagulase-Negative Staphylococci Isolated From a University Hospital in Central Iran. *Oman Med. J.* 31, 426–432.
- Takeuchi, F., Watanabe, S., Baba, T., Yuzawa, H., Ito, T., Morimoto, Y., Kuroda, M., Cui, L., Takahashi, M., Ankai, A., et al. (2005). Whole-Genome Sequencing of *Staphylococcus haemolyticus* Uncovers the Extreme Plasticity of Its Genome and the Evolution of Human-Colonizing Staphylococcal Species. *J. Bacteriol.* 187, 7292–7308.
- Tan, C., Wang, J., Hu, Y., Wang, P., and Zou, L. (2017). *Staphylococcus epidermidis*  $\Delta$ Sortase A strain elicits protective immunity against *Staphylococcus aureus* infection. *Antonie Van Leeuwenhoek* 110, 133–143.
- Thomas, J.C., Vargas, M.R., Miragaia, M., Peacock, S.J., Archer, G.L., and Enright, M.C. (2007). Improved Multilocus Sequence Typing Scheme for *Staphylococcus epidermidis*. *J. Clin. Microbiol.* 45, 616–619.
- Toba, F.A., Visai, L., Trivedi, S., and Lowy, F.D. (2013). The Role of Ionic Interactions in the Adherence of the *S. epidermidis* Adhesin SdrF to Prosthetic Material. *FEMS Microbiol. Lett.* 338, 24–30.
- Tokars, J.I. (2004). Predictive value of blood cultures positive for coagulase-negative staphylococci: implications for patient care and health care quality assurance. *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.* 39, 333–341.
- Toro, M., Retamal, P., Ayers, S., Barreto, M., Allard, M., Brown, E.W., and Gonzalez-Escalona, N. (2016). Whole-Genome Sequencing Analysis of *Salmonella enterica*

Serovar Enteritidis Isolates in Chile Provides Insights into Possible Transmission between Gulls, Poultry, and Humans. *Appl. Environ. Microbiol.* 82, 6223–6232.

Trivedi, S., Uhlemann, A.-C., Herman-Bausier, P., Sullivan, S.B., Sowash, M.G., Flores, E., Khan, S., Dufrêne, Y.F., and Lowy, F.D. (2017). The Surface Protein SdrF Mediates *Staphylococcus epidermidis* Adherence to Keratin. *J. Infect. Dis.*

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

Vanassche, T., Peetermans, M., Van Aelst, L.N.L., Peetermans, W.E., Verhaegen, J., Missiakas, D.M., Schneewind, O., Hoylaerts, M.F., and Verhamme, P. (2013). The role of staphylothrombin-mediated fibrin deposition in catheter-related *Staphylococcus aureus* infections. *J. Infect. Dis.* 208, 92–100.

Viana Taveira, M.R., Lima, L.S., de Araújo, C.C., and de Mello, M.J.G. (2017). Risk factors for central line-associated bloodstream infection in pediatric oncology patients with a totally implantable venous access port: A cohort study. *Pediatr. Blood Cancer* 64, 336–342.

Villari, P., Sarnataro, C., and Iacuzio, L. (2000). Molecular Epidemiology of *Staphylococcus epidermidis* in a Neonatal Intensive Care Unit over a Three-Year Period. *J. Clin. Microbiol.* 38, 1740–1746.

Vuong, C., Kocianova, S., Yao, Y., Carmody, A.B., and Otto, M. (2004). Increased colonization of indwelling medical devices by quorum-sensing mutants of *Staphylococcus epidermidis* in vivo. *J. Infect. Dis.* 190, 1498–1505.

Wang, R., Khan, B.A., Cheung, G.Y.C., Bach, T.-H.L., Jameson-Lee, M., Kong, K.-F., Queck, S.Y., and Otto, M. (2011). *Staphylococcus epidermidis* surfactant peptides promote biofilm maturation and dissemination of biofilm-associated infection in mice. *J. Clin. Invest.* 121, 238–248.

Wang, X.-M., Noble, L., Kreiswirth, B.N., Eisner, W., McClements, W., Jansen, K.U., and Anderson, A.S. (2003). Evaluation of a multilocus sequence typing system for *Staphylococcus epidermidis*. *J. Med. Microbiol.* 52, 989–998.

Wertheim, H.F.L., Melles, D.C., Vos, M.C., van Leeuwen, W., van Belkum, A., Verbrugh, H.A., and Nouwen, J.L. (2005). The role of nasal carriage in *Staphylococcus aureus* infections. *Lancet Infect. Dis.* 5, 751–762.

Widerström, M., Wiström, J., Edebro, H., Marklund, E., Backman, M., Lindqvist, P., and Monsen, T. (2016). Colonization of patients, healthcare workers, and the environment with healthcare-associated *Staphylococcus epidermidis* genotypes in an intensive care unit: a prospective observational cohort study. *BMC Infect. Dis.* 16.

Wielders, C.L., Vriens, M.R., Brisse, S., de Graaf-Miltenburg, L.A., Troelstra, A., Fleer, A., Schmitz, F.J., Verhoef, J., and Fluit, A.C. (2001). In-vivo transfer of *mecA* DNA to *Staphylococcus aureus* [corrected]. *Lancet Lond. Engl.* 357, 1674–1675.

Williams, R.J., Henderson, B., Sharp, L.J., and Nair, S.P. (2002). Identification of a Fibronectin-Binding Protein from *Staphylococcus epidermidis*. *Infect. Immun.* 70, 6805–6810.

Wisplinghoff, H., Rosato, A.E., Enright, M.C., Noto, M., Craig, W., and Archer, G.L. (2003). Related Clones Containing SCCmec Type IV Predominate among Clinically

Significant *Staphylococcus epidermidis* Isolates. *Antimicrob. Agents Chemother.* 47, 3574–3579.

Zhang, W., Rong, C., Chen, C., and Gao, G.F. (2012). Type-IVC Secretion System: A Novel Subclass of Type IV Secretion System (T4SS) Common Existing in Gram-Positive Genus *Streptococcus*. *PLoS ONE* 7.

Ziebuhr, W., Hennig, S., Eckart, M., Kränzler, H., Batzilla, C., and Kozitskaya, S. (2006). Nosocomial infections by *Staphylococcus epidermidis*: how a commensal bacterium turns into a pathogen. *Int. J. Antimicrob. Agents* 28 *Suppl 1*, S14-20.

(2009). Classification of Staphylococcal Cassette Chromosome *mec* (SCC*mec*): Guidelines for Reporting Novel SCC*mec* Elements. *Antimicrob. Agents Chemother.* 53, 4961–4967.

## 8. Anhang

Tabelle 8.1: Verwendete Isolate, ihr MLST-Sequenztyp, Markernachweis und Herkunft

| Stamm    | ST | <i>housekeeping genes (arcC /aroE / gtr / mutS / pyrR / tpiA / yqiL)</i> | ST2-Marker 1 | ST2-Marker 2 | ST2-Marker 3 | ST5-Marker | ST59-Marker | Herkunft |
|----------|----|--------------------------------------------------------------------------|--------------|--------------|--------------|------------|-------------|----------|
| VCU028   | 1  | 1/2/2/2/1/1/10                                                           | -            | -            | -            | -          | -           | PDG      |
| VCU081   | 1  | 1/2/2/2/1/1/10                                                           | -            | -            | -            | -          | -           | PDG      |
| ABDK11   | 2  | 7/1/2/2/4/1/1                                                            | -            | +            | +            | -          | -           | HL       |
| Ae-e06   | 2  | 7/1/2/2/4/1/1                                                            | -            | +            | -            | -          | -           | HL       |
| ALKB06   | 2  | 7/1/2/2/4/1/1                                                            | -            | +            | -            | -          | -           | HL       |
| B10873   | 2  | 7/1/2/2/4/1/1                                                            | +            | +            | +            | -          | -           | HL       |
| B12142   | 2  | 7/1/2/2/4/1/1                                                            | +            | -            | +            | -          | -           | HL       |
| B12916   | 2  | 7/1/2/2/4/1/1                                                            | +            | -            | +            | -          | -           | HL       |
| B13258   | 2  | 7/1/2/2/4/1/1                                                            | -            | -            | +            | -          | -           | HL       |
| B13587   | 2  | 7/1/2/2/4/1/1                                                            | +            | -            | +            | -          | -           | HL       |
| B16261   | 2  | 7/1/2/2/4/1/1                                                            | +            | -            | +            | -          | -           | HL       |
| B2461    | 2  | 7/1/2/2/4/1/1                                                            | +            | -            | +            | -          | -           | HL       |
| B2731    | 2  | 7/1/2/2/4/1/1                                                            | +            | -            | +            | -          | -           | HL       |
| B59      | 2  | 7/1/2/2/4/1/1                                                            | +            | -            | +            | -          | -           | HL       |
| B647     | 2  | 7/1/2/2/4/1/1                                                            | +            | -            | +            | -          | -           | HL       |
| B7234    | 2  | 7/1/2/2/4/1/1                                                            | +            | +            | +            | -          | -           | HL       |
| ENVH113  | 2  | 7/1/2/2/4/1/1                                                            | +            | -            | -            | -          | -           | HH       |
| ENVH131  | 2  | 7/1/2/2/4/1/1                                                            | +            | +            | +            | -          | -           | HH       |
| ENVH150  | 2  | 7/1/2/2/4/1/1                                                            | -            | -            | +            | -          | -           | HH       |
| ENVH350  | 2  | 7/1/2/2/4/1/1                                                            | -            | +            | -            | -          | -           | HH       |
| ENVL300  | 2  | 7/1/2/2/4/1/1                                                            | -            | +            | -            | -          | -           | HL       |
| ENVP605  | 2  | 7/1/2/2/4/1/1                                                            | -            | +            | -            | -          | -           | JG       |
| Kt1p58   | 2  | 7/1/2/2/4/1/1                                                            | -            | +            | -            | -          | -           | HL       |
| LRKNS009 | 2  | 7/1/2/2/4/1/1                                                            | +            | +            | -            | -          | -           | HL       |
| LRKNS010 | 2  | 7/1/2/2/4/1/1                                                            | +            | +            | +            | -          | -           | HL       |
| LRKNS014 | 2  | 7/1/2/2/4/1/1                                                            | +            | +            | +            | -          | -           | HL       |
| LRKNS016 | 2  | 7/1/2/2/4/1/1                                                            | +            | +            | +            | -          | -           | HL       |
| LRKNS017 | 2  | 7/1/2/2/4/1/1                                                            | +            | +            | +            | -          | -           | HL       |
| LRKNS018 | 2  | 7/1/2/2/4/1/1                                                            | +            | +            | +            | -          | -           | HL       |
| LRKNS020 | 2  | 7/1/2/2/4/1/1                                                            | +            | +            | +            | -          | -           | HL       |
| LRKNS026 | 2  | 7/1/2/2/4/1/1                                                            | +            | +            | +            | -          | -           | HL       |
| LRKNS029 | 2  | 7/1/2/2/4/1/1                                                            | +            | +            | +            | -          | -           | HL       |
| LRKNS032 | 2  | 7/1/2/2/4/1/1                                                            | +            | +            | +            | -          | -           | HL       |
| LRKNS037 | 2  | 7/1/2/2/4/1/1                                                            | +            | +            | +            | -          | -           | HL       |
| LRKNS038 | 2  | 7/1/2/2/4/1/1                                                            | +            | +            | +            | -          | -           | HL       |
| LRKNS039 | 2  | 7/1/2/2/4/1/1                                                            | +            | +            | +            | -          | -           | HL       |
| LRKNS042 | 2  | 7/1/2/2/4/1/1                                                            | +            | +            | +            | -          | -           | HL       |
| LRKNS046 | 2  | 7/1/2/2/4/1/1                                                            | +            | +            | +            | -          | -           | HL       |
| LRKNS048 | 2  | 7/1/2/2/4/1/1                                                            | +            | +            | +            | -          | -           | HL       |
| LRKNS050 | 2  | 7/1/2/2/4/1/1                                                            | +            | +            | +            | -          | -           | HL       |
| LRKNS054 | 2  | 7/1/2/2/4/1/1                                                            | +            | +            | +            | -          | -           | HL       |
| LRKNS056 | 2  | 7/1/2/2/4/1/1                                                            | +            | +            | +            | -          | -           | HL       |
| LRKNS057 | 2  | 7/1/2/2/4/1/1                                                            | +            | +            | +            | -          | -           | HL       |
| LRKNS059 | 2  | 7/1/2/2/4/1/1                                                            | +            | +            | +            | -          | -           | HL       |
| LRKNS062 | 2  | 7/1/2/2/4/1/1                                                            | +            | +            | +            | -          | -           | HL       |

|                |    |                 |   |   |   |   |   |     |
|----------------|----|-----------------|---|---|---|---|---|-----|
| LRKNS064       | 2  | 7/1/2/2/4/1/1   | + | + | + | - | - | HL  |
| LRKNS068       | 2  | 7/1/2/2/4/1/1   | + | + | + | - | - | HL  |
| LRKNS072       | 2  | 7/1/2/2/4/1/1   | + | + | + | - | - | HL  |
| LRKNS074       | 2  | 7/1/2/2/4/1/1   | + | + | + | - | - | HL  |
| LRKNS075       | 2  | 7/1/2/2/4/1/1   | + | + | + | - | - | HL  |
| LRKNS076       | 2  | 7/1/2/2/4/1/1   | + | + | + | - | - | HL  |
| LRKNS077       | 2  | 7/1/2/2/4/1/1   | + | + | + | - | - | HL  |
| LRKNS079       | 2  | 7/1/2/2/4/1/1   | + | + | + | - | - | HL  |
| LRKNS080       | 2  | 7/1/2/2/4/1/1   | + | + | + | - | - | HL  |
| LRKNS081       | 2  | 7/1/2/2/4/1/1   | + | + | + | - | - | HL  |
| LRKNS082       | 2  | 7/1/2/2/4/1/1   | + | + | + | - | - | HL  |
| LRKNS084       | 2  | 7/1/2/2/4/1/1   | + | + | + | - | - | HL  |
| Lt3p20         | 2  | 7/1/2/2/4/1/1   | - | + | - | - | - | HL  |
| pK_B01         | 2  | 7/1/2/2/4/1/1   | - | + | - | - | - | HL  |
| VCU041         | 2  | 7/1/2/2/4/1/1   | + | - | - | - | - | PDG |
| NIH05005       | 2  | 7/1/2/2/4/1/1   | - | - | - | - | - | PDG |
| NIH05001       | 2  | 7/1/2/2/4/1/1   | - | + | - | - | - | PDG |
| NIH04008       | 2  | 7/1/2/2/4/1/1   | - | + | - | - | - | PDG |
| 41tr           | 2  | 7/1/2/2/4/1/1   | + | + | + | - | - | PDG |
| A1KF08         | 5  | 1/1/1/2/2/1/1   | - | - | - | + | - | HL  |
| AOKB02         | 5  | 1/1/1/2/2/1/1   | - | - | - | + | - | HL  |
| Bt1p3          | 5  | 1/1/1/2/2/1/1   | - | - | - | + | - | HL  |
| ENVH138        | 5  | 1/1/1/2/2/1/1   | - | - | - | - | - | HH  |
| ENVH144        | 5  | 1/1/1/2/2/1/1   | - | - | - | + | - | HH  |
| ENVL_032       | 5  | 1/1/1/2/2/1/1   | - | - | - | - | - | HL  |
| ENVL_287       | 5  | 1/1/1/2/2/1/1   | - | - | - | + | - | HL  |
| ENVL_370       | 5  | 1/1/1/2/2/1/1   | - | - | - | + | - | HL  |
| ENVP534        | 5  | 1/1/1/2/2/1/1   | - | - | - | - | - | JG  |
| KED12          | 5  | 1/1/1/2/2/1/1   | - | - | - | - | - | HL  |
| Mt1p16         | 5  | 1/1/1/2/2/1/1   | - | - | - | + | - | HL  |
| Qt1p44         | 5  | 1/1/1/2/2/1/1   | - | - | - | - | - | HL  |
| XE_C06         | 5  | 1/1/1/2/2/1/1   | - | - | - | + | - | HL  |
| M23864_W2_grey | 5  | 1/1/1/2/2/1/1   | - | - | - | - | - | PDG |
| VCU045         | 5  | 1/1/1/2/2/1/1   | - | - | - | + | - | PDG |
| NIH08001       | 5  | 1/1/1/2/2/1/1   | - | - | - | - | - | PDG |
| NIH06004       | 5  | 1/1/1/2/2/1/1   | - | - | - | + | - | PDG |
| M0881          | 5  | 1/1/1/2/2/1/1   | - | - | - | + | - | PDG |
| VCU037         | 5  | 1/1/1/2/2/1/1   | - | - | - | + | - | PDG |
| NIH05003       | 7  | 1/1/1/2/4/1/1   | - | - | - | - | - | PDG |
| NIHLM020       | 7  | 1/1/1/2/4/1/1   | - | - | - | - | - | PDG |
| ATCC 12228     | 8  | 2/1/7/1/1/1/1   | + | + | - | - | - | PDG |
| VCU057         | 8  | 2/1/7/1/1/1/1   | + | - | - | - | - | PDG |
| M04            | 10 | 1/1/1/1/3/1/1   | - | - | - | - | - | Kob |
| M22            | 10 | 1/1/1/1/3/1/1   | - | - | - | - | - | Kob |
| RP62A          | 10 | 1/1/1/1/3/1/1   | - | - | - | - | - | PDG |
| VCU071         | 19 | 8/7/12/4/12/2/2 | - | - | - | - | - | PDG |
| NIHLM018       | 20 | 1/1/2/2/1/1/3   | - | - | - | - | - | PDG |
| T679           | 22 | 7/1/2/2/4/7/1   | - | - | - | - | - | PDG |
| VCU120         | 22 | 7/1/2/2/4/7/1   | - | - | - | - | - | PDG |
| VCU117         | 23 | 7/1/2/1/3/3/1   | - | - | - | - | - | PDG |
| VCU126         | 23 | 7/1/2/1/3/3/1   | - | - | - | - | - | PDG |
| T423           | 48 | 7/1/2/2/4/1/4   | - | - | - | - | - | PDG |

|             |     |                    |   |   |   |   |   |     |
|-------------|-----|--------------------|---|---|---|---|---|-----|
| SK135       | 57  | 1/1/1/1/2/1/1      | - | - | - | - | - | PDG |
| AEEG06      | 59  | 2/1/1/1/2/1/1      | - | - | - | - | + | HL  |
| AHKF08      | 59  | 2/1/1/1/2/1/1      | - | - | - | - | + | HL  |
| B4249_E1    | 59  | 2/1/1/1/2/1/1      | - | - | - | - | - | HL  |
| B8190       | 59  | 2/1/1/1/2/1/1      | - | - | - | - | - | HL  |
| KEB09       | 59  | 2/1/1/1/2/1/1      | - | - | - | - | + | HL  |
| M01         | 59  | 2/1/1/1/2/1/1      | - | - | - | - | - | Kob |
| M25         | 59  | 2/1/1/1/2/1/1      | - | - | - | - | - | Kob |
| M26         | 59  | 2/1/1/1/2/1/1      | - | - | - | - | - | Kob |
| M29         | 59  | 2/1/1/1/2/1/1      | - | - | - | - | - | Ber |
| SKC09       | 59  | 2/1/1/1/2/1/1      | - | - | - | - | - | HL  |
| BCM-HMP0060 | 59  | 2/1/1/1/2/1/1      | - | - | - | - | - | PDG |
| VCU109      | 59  | 2/1/1/1/2/1/1      | - | - | - | - | - | PDG |
| AU12-03     | 59  | 2/1/1/1/2/1/1      | - | - | - | - | - | PDG |
| NIHLM087    | 72  | 8/2/2/4/9/6/9      | - | - | - | - | - | PDG |
| NIHLM095    | 72  | 8/2/2/4/9/6/9      | - | - | - | - | - | PDG |
| NIHLM039    | 73  | 1/5/2/6/2/1/6      | - | - | - | - | - | PDG |
| VCU123      | 73  | 1/5/2/6/2/1/6      | - | - | - | - | - | PDG |
| VCU105      | 73  | 1/5/2/6/2/1/6      | - | - | - | - | - | PDG |
| 1457        | 86  | 1/2/2/1/1/1/1      | - | - | - | - | - | HH  |
| ACEF09      | 89  | 1/1/2/1/2/1/1      | + | - | - | - | - | HL  |
| NIH051475   | 89  | 1/1/2/1/2/1/1      | + | + | - | - | - | PDG |
| M05         | 93  | 1/13/2/2/21/1/7    | - | - | - | - | - | Kob |
| M17         | 102 | 1/1/2/6/2/11/10    | - | - | - | - | - | Mü  |
| M06         | 109 | 1/1/20/6/2/16/10   | - | - | - | - | - | Kob |
| M19         | 109 | 1/1/20/6/2/16/10   | - | - | - | - | - | Mü  |
| M35         | 112 | 1/1/21/2/2/16/7    | - | - | - | - | - | Ber |
| 8400        | 114 | 1/1/2/6/2/1/7      | - | - | - | - | - | HH  |
| IS-K        | 130 | 1/1/1/2/1/1/1      | - | - | - | - | - | PDG |
| B14018      | 130 | 1/1/1/2/1/1/1      | - | - | - | - | - | HL  |
| VCU127      | 136 | 1/1/7/2/2/1/1      | - | - | - | - | - | PDG |
| Sci25       | 153 | 2/1/6/2/2/1/1      | - | - | - | - | - | PDG |
| Sci31       | 153 | 2/1/6/2/2/1/1      | - | - | - | - | - | PDG |
| ENVH115     | 168 | 7/1/2/22/4/1/1     | + | + | - | - | - | HH  |
| NIHLM040    | 170 | 25/19/17/4/23/10/2 | - | - | - | - | - | PDG |
| NIHLM088    | 184 | 16/1/2/1/2/1/1     | - | - | - | - | - | PDG |
| NIH051668   | 185 | 27/1/2/2/4/1/1     | + | + | - | - | - | PDG |
| NIHLM003    | 218 | 1/1/2/6/2/16/1     | - | - | - | - | - | PDG |
| W23144      | 290 | 12/29/5/5/11/4/4   | - | - | - | - | - | PDG |
| NIHLM001    | 323 | 1/1/7/6/2/1/41     | - | - | - | - | - | PDG |
| NIHLM008    | 324 | 1/13/6/6/2/20/1    | - | - | - | - | - | PDG |
| NIHLM015    | 325 | 12/29/9/8/6/5/27   | - | - | - | - | - | PDG |
| NIHLM031    | 326 | 19/16/5/6/3/19/10  | - | - | - | - | - | PDG |
| NIHLM021    | 327 | 1/1/2/1/4/1/1      | - | - | - | - | - | PDG |
| NIHLM023    | 328 | 12/29/5/5/7/5/4    | - | - | - | - | - | PDG |
| NIHLM037    | 329 | 12/29/9/8/6/5/8    | - | - | - | - | - | PDG |
| VCU129      | 329 | 12/29/9/8/6/5/8    | - | - | - | - | - | PDG |
| NIHLM049    | 330 | 2/1/2/2/34/1/1     | - | - | - | - | - | PDG |
| NIHLM053    | 331 | 12/35/5/5/10/16/21 | - | - | - | - | - | PDG |
|             |     |                    |   |   |   |   |   |     |

|                    |      |                        |   |   |   |   |   |     |
|--------------------|------|------------------------|---|---|---|---|---|-----|
| NIHLM057           | 331  | 12/35/5/5/10/16/<br>21 | - | - | - | - | - | PDG |
| NIHLM061           | 332  | 28/29/5/5/11/4/4       | - | - | - | - | - | PDG |
| NIHLM067           | 333  | 2/1/2/1/2/1/7          | - | - | - | - | - | PDG |
| NIHLM070           | 334  | 1/1/33/6/2/16          | - | - | - | - | - | PDG |
| VCU125             | 384  | 1/1/22/2/2/1/25        | - | - | - | - | - | PDG |
| NIH04003           | 430  | 1/1/1/34/2/1/1         | - | - | - | + | - | PDG |
| UC7032             | 595  | 49/3/9/5/3/4/4         | - | + | - | - | - | PDG |
| M21                | Neu  | 1/1/2/2/2/1/10         | - | - | - | - | - | Gie |
| 1057               | Neu  | 14/1/2/2/4/1/1         | - | + | - | - | - | HH  |
| FRI909             | Neu  | 28/16/5/5/3/22/<br>11  | - | - | - | - | - | PDG |
| VCU144             | Neu  | 1/13/2/1/2/1/29        | - | - | - | - | - | PDG |
| BVS058A4           | Neu  | 24/22/6/4/29/10/<br>13 | - | - | - | - | - | PDG |
| VCU065             | Neu  | 2/6/2/1/1/1/1          | - | + | - | - | - | PDG |
| VCU118             | Neu  | 28/29/15/7/3/34/<br>11 | - | - | - | - | - | PDG |
| VCU128             | Neu  | 28/44/5/5/7/5/11       | - | - | - | - | - | PDG |
| 12142587           | Neu  | 8/24/12/23/12/2/<br>2  | - | - | - | - | - | PDG |
| MC19               | Neu  | 28/16/5/5/8/22/<br>12  | - | - | - | - | - | PDG |
| EA13A              | Neu  | 1/1/1/2/2/1/4          | - | - | - | - | - | PDG |
| 14.1.R1.SE         | Neu  | 49/25/9/30/7/5/<br>11  | - | - | - | - | - | PDG |
| APO27              | Neu  | 28/16/5/5/8/22/<br>12  | - | - | - | - | - | PDG |
| APO35              | Neu  | 28/16/5/5/8/22/<br>12  | - | - | - | - | - | PDG |
| CIM37_<br>Sentile1 | Neu  | 28/16/5/5/8/22/<br>12  | - | - | - | - | - | PDG |
| CIM40              | Neu  | 28/16/5/5/8/22/<br>12  | - | - | - | - | - | PDG |
| MC16               | Neu  | 28/16/5/5/8/22/<br>12  | - | - | - | - | - | PDG |
| MC28               | Neu  | 28/16/5/5/8/22/<br>12  | - | - | - | - | - | PDG |
| Sci19              | Neu  | 28/16/5/5/8/22/<br>12  | - | - | - | - | - | PDG |
| Sci22              | Neu  | 28/16/5/5/8/22/<br>12  | - | - | - | - | - | PDG |
| Sentile2           | Neu  | 28/16/5/47/8/22/<br>38 | - | - | - | - | - | PDG |
| WI05               | Neu  | 28/16/5/5/8/22/<br>12  | - | - | - | - | - | PDG |
| WI09               | Neu  | 28/16/5/5/8/22/<br>12  | - | - | - | - | - | PDG |
| IS-250             | 130? | 1/1/1/x/1/1/1          | - | - | - | - | - | PDG |
| 36-1               | 2?   | 7/1/2/2/4/1/x          | + | + | + | - | - | PDG |
| 528m               | 2?   | 7/1/x/2/4/1/1          | + | + | + | - | - | PDG |

ST Sequenztyp, SE *Staphylococcus epidermidis*, PDG *public draft genome*, HL Lübeck, HH Hamburg, JG Jelenia Gora, Kob Koblenz, Mü München, Gie Giessen, B Berlin, n.a. nicht angegeben, x nicht zuzuordnen

Tabelle 8.2: Verwendete Isolate, ihre Herkunft, Isolatqualität und ihr N50-Wert

| Stamm    | Herkunft | Isolat-<br>qualität | N50-Wert |
|----------|----------|---------------------|----------|
| VCU028   | PDG      | n.a.                | 94.488   |
| VCU081   | PDG      | n.a.                | 65.590   |
| ABDK11   | HL       | P                   | 33.156   |
| Ae-e06   | HL       | P                   | 82.965   |
| ALKB06   | HL       | P                   | 46.754   |
| B10873   | HL       | P                   | 64.352   |
| B12142   | HL       | P                   | 81.755   |
| B12916   | HL       | P                   | 81.519   |
| B13258   | HL       | P                   | 69.211   |
| B13587   | HL       | P                   | 70.893   |
| B16261   | HL       | P                   | 83.041   |
| B2461    | HL       | P                   | 61.130   |
| B2731    | HL       | P                   | 83.037   |
| B59      | HL       | P                   | 82410    |
| B647     | HL       | P                   | 88.100   |
| B7234    | HL       | P                   | 53.162   |
| ENVH113  | HH       | KU                  | 96.475   |
| ENVH131  | HH       | KU                  | 81.413   |
| ENVH150  | HH       | KU                  | 84.738   |
| ENVH350  | HH       | U                   | 96.213   |
| ENVL300  | HL       | KU                  | 52.802   |
| ENVP605  | JG       | KU                  | 42.320   |
| Kt1p58   | HL       | P                   | 71.810   |
| LRKNS009 | HL       | P                   | 87.441   |
| LRKNS010 | HL       | P                   | 87.366   |
| LRKNS014 | HL       | P                   | 57.706   |
| LRKNS016 | HL       | P                   | 65.672   |
| LRKNS017 | HL       | P                   | 50.005   |
| LRKNS018 | HL       | P                   | 10.745   |
| LRKNS020 | HL       | P                   | 82.847   |
| LRKNS026 | HL       | P                   | 57.793   |
| LRKNS029 | HL       | P                   | 76.150   |
| LRKNS032 | HL       | P                   | 52.894   |
| LRKNS037 | HL       | P                   | 69.342   |
| LRKNS038 | HL       | P                   | 23.122   |
| LRKNS039 | HL       | P                   | 62.006   |
| LRKNS042 | HL       | P                   | 82.847   |
| LRKNS046 | HL       | P                   | 67.415   |
| LRKNS048 | HL       | P                   | 43.861   |
| LRKNS050 | HL       | P                   | 96.634   |
| LRKNS054 | HL       | P                   | 96.231   |
| LRKNS056 | HL       | P                   | 95.269   |
| LRKNS057 | HL       | P                   | 53.490   |
| LRKNS059 | HL       | P                   | 47.741   |
| LRKNS062 | HL       | P                   | 46.433   |
| LRKNS064 | HL       | P                   | 81.933   |
| LRKNS068 | HL       | P                   | 83.070   |
| LRKNS072 | HL       | P                   | 82.914   |
| LRKNS074 | HL       | P                   | 36.559   |
| LRKNS075 | HL       | P                   | 25.070   |
| LRKNS076 | HL       | P                   | 53.493   |

|                |      |      |         |
|----------------|------|------|---------|
| LRKNS077       | HL   | P    | 16.953  |
| LRKNS079       | HL   | P    | 62.942  |
| LRKNS080       | HL   | P    | 71.115  |
| LRKNS081       | HL   | P    | 38.439  |
| LRKNS082       | HL   | P    | 43.676  |
| LRKNS084       | HL   | P    | 70.948  |
| Lt3p20         | HL   | P    | 72.146  |
| pK_B01         | HL   | P    | 66.998  |
| VCU041         | PDG  | n.a. | 69.330  |
| NIH05005       | PDG  | B    | 112.534 |
| NIH05001       | PDG  | B    | 41.387  |
| NIH04008       | PDG  | B    | 55.960  |
| 41tr           | PDG  | n.a. | 63.976  |
| A1KF08         | HL   | P    | 79.213  |
| AOKB02         | HL   | P    | 87.473  |
| Bt1p3          | HL   | P    | 95.772  |
| ENVH138        | HH   | KU   | 80.451  |
| ENVH144        | HH   | KU   | 72.273  |
| ENVL_032       | HL   | KU   | 96.026  |
| ENVL_287       | HL   | KU   | 89.326  |
| ENVL_370       | HL   | U    | 82.922  |
| ENVP534        | JG   | KU   | 62.493  |
| KED12          | HL   | P    | 60.596  |
| Mt1p16         | HL   | P    | 51.476  |
| Qt1p44         | HL   | P    | 96.264  |
| XE_C06         | HL   | P    | 83.009  |
| M23864_W2_Grey | PDG  | U    | 92.830  |
| VCU045         | PDG  | n.a. | 76.713  |
| NIH08001       | PDG  | B    | 88.613  |
| NIH06004       | PDG  | B    | 88.654  |
| M0881          | PDG  | B    | 102.303 |
| VCU037         | PDG  | n.a. | 116.325 |
| NIH05003       | PDG  | B    | 112.666 |
| NIHLM020       | PDG  | H    | 111.908 |
| ATCC12228      | PDG  | U    | n.a.    |
| VCU057         | PDG  | n.a. | 51.219  |
| M04            | Kob  | Bov  | 110.070 |
| M22            | Kob  | Bov  | 82.786  |
| RP62A          | n.a. | P    | 82.701  |
| VCU071         | PDG  | n.a. | 82.765  |
| NIHLM018       | PDG  | H    | 100.250 |
| T679           | PDG  | n.a. | 21.967  |
| VCU120         | PDG  | n.a. | 109.995 |
| VCU117         | PDG  | n.a. | 66.010  |
| VCU126         | PDG  | n.a. | 68.885  |
| T423           | PDG  | n.a. | 34.570  |
| SK135          | PDG  | U    | 276.877 |
| AEEG06         | HL   | P    | 129.954 |
| AHKF08         | HL   | P    | 129.772 |
| B4249_E1       | HL   | P    | 57.624  |
| B8190          | HL   | P    | 84.111  |
| KEB09          | HL   | P    | 22.716  |
| M01            | Kob  | Bov  | 210.996 |

|             |     |      |         |
|-------------|-----|------|---------|
| M25         | Kob | Bov  | 110.103 |
| M26         | Kob | Bov  | 81.301  |
| M29         | Ber | Bov  | 48.888  |
| SKC09       | HL  | P    | 65.050  |
| BCM-HMP0060 | PDG | H    | 60.334  |
| VCU109      | PDG | n.a. | 92.250  |
| AU12-03     | PDG | K    | 81.766  |
| NIHLM087    | PDG | H    | 231.717 |
| NIHLM095    | PDG | H    | 210.651 |
| NIHLM039    | PDG | H    | 141.870 |
| VCU123      | PDG | n.a. | 65.375  |
| VCU105      | PDG | n.a. | 144.444 |
| 1457        | HH  | P    | 81.508  |
| ACEF09      | HL  | P    | 87.896  |
| NIH051475   | PDG | B    | 82.215  |
| M05         | Kob | Bov  | 97.613  |
| M17         | Mü  | Bov  | 70.532  |
| M06         | Kob | Bov  | 81.780  |
| M19         | Mü  | Bov  | 78.425  |
| M35         | Ber | Bov  | 71.026  |
| 8400        | HH  | P    | 109.639 |
| IS-K        | PDG | n.a. | 95.197  |
| B14018      | HL  | P    | 72.299  |
| VCU127      | PDG | n.a. | 80.044  |
| Sci25       | PDG | M    | 47.479  |
| Sci31       | PDG | M    | 24.307  |
| ENVH115     | HH  | KU   | 79.867  |
| NIHLM040    | PDG | H    | 129.468 |
| NIHLM088    | PDG | H    | 137.963 |
| NIH051668   | PDG | B    | 94.672  |
| NIHLM003    | PDG | H    | 120.024 |
| W23144      | PDG | U    | 45.524  |
| NIHLM001    | PDG | H    | 96.622  |
| NIHLM008    | PDG | H    | 180.423 |
| NIHLM015    | PDG | H    | 188.412 |
| NIHLM031    | PDG | H    | 116.494 |
| NIHLM021    | PDG | H    | 77.530  |
| NIHLM023    | PDG | H    | 166.800 |
| NIHLM037    | PDG | H    | 108.277 |
| VCU129      | PDG | n.a. | 152.156 |
| NIHLM049    | PDG | H    | 125.653 |
| NIHLM053    | PDG | H    | 75.775  |
| NIHLM057    | PDG | H    | 210.416 |
| NIHLM061    | PDG | H    | 127.052 |
| NIHLM067    | PDG | H    | 122.425 |
| NIHLM070    | PDG | H    | 108.860 |
| VCU125      | PDG | n.a. | 69.546  |
| NIH04003    | PDG | B    | 95.333  |
| UC7032      | PDG | U    | 83.821  |
| M21         | Gie | Bov  | 78.536  |
| 1057        | HH  | P    | 113.091 |
| FRI909      | PDG | n.a. | 129.563 |
| VCU144      | PDG | n.a. | 77.528  |
| BVS058A4    | PDG | n.a. | 842.138 |

|                |     |      |         |
|----------------|-----|------|---------|
| VCU065         | PDG | n.a. | 82.730  |
| VCU118         | PDG | n.a. | 70.187  |
| VCU128         | PDG | n.a. | 108.019 |
| 12142587       | PDG | n.a. | 93.807  |
| MC19           | PDG | M    | 84.904  |
| EA13A          | PDG | n.a. | n.a.    |
| 14.1.R1.SE     | PDG | U    | 38.797  |
| APO27          | PDG | M    | 42.531  |
| APO35          | PDG | M    | 47.820  |
| CIM37_Sentile1 | PDG | M    | 42.105  |
| CIM40          | PDG | M    | 97.905  |
| MC16           | PDG | M    | 54.777  |
| MC28           | PDG | M    | 21.099  |
| ScI19          | PDG | M    | 94.372  |
| ScI22          | PDG | M    | 10.280  |
| Sentile2       | PDG | n.a. | 13.549  |
| WI05           | PDG | M    | 78.843  |
| WI09           | PDG | M    | 96.764  |
| IS-250         | PDG | n.a. | 90.927  |
| 36-1           | PDG | n.a. | 10.922  |
| 528m           | PDG | n.a. | 45.618  |

PDG *public draft genome*, HL Lübeck, HH Hamburg, JG Jelenia Gora, Kob Koblenz, Mü München, Gie Giessen, Ber Berlin, P Pathogen, H Hautabstrich, U Umwelt, KU Klinikumwelt, B Blutkultur, Bov Bovin, K Katheter, M Maus, n.a. nicht angegeben

Tabelle 8.3: ORFs der ST5-Markersuche und ihre Länge

| <i>open reading frame</i> | Länge  |
|---------------------------|--------|
| ORF 1                     | 738 bp |
| ORF 2                     | 858 bp |

Tabelle 8.4: ORFs der ST59-Markersuche und ihre Länge

| <i>open reading frame</i> | Länge  |
|---------------------------|--------|
| ORF 1                     | 312 bp |
| ORF 2                     | 339 bp |
| ORF 3                     | 435 bp |
| ORF 4                     | 408 bp |

Tabelle 8.5: Erste ORF-Abfolge der ST2-Markersuche und ihre Länge

| <i>open reading frame</i> | Länge    |
|---------------------------|----------|
| ORF 1                     | 1.248 bp |
| ORF 2                     | 3.039 bp |
| ORF 3                     | 303 bp   |
| ORF 4                     | 2.247 bp |

|        |          |
|--------|----------|
| ORF 5  | 693 bp   |
| ORF 6  | 2.580 bp |
| ORF 7  | 2.298 bp |
| ORF 8  | 402 bp   |
| ORF 9  | 327 bp   |
| ORF 10 | 597 bp   |
| ORF 11 | 1.950 bp |
| ORF 12 | 1.071 bp |
| ORF 13 | 603 bp   |
| ORF 14 | 858 bp   |
| ORF 15 | 321 bp   |
| ORF 16 | 351 bp   |
| ORF 17 | 1.305 bp |
| ORF 18 | 2.157 bp |
| ORF 19 | 480 bp   |
| ORF 20 | 762 bp   |
| ORF 21 | 2.337 bp |

Tabelle 8.6: Zweite ORF-Abfolge der ST2-Markersuche und ihre Länge

| <i>open reading frame</i> | Länge    |
|---------------------------|----------|
| ORF 1                     | 408 bp   |
| ORF 2                     | 354 bp   |
| ORF 3                     | 681 bp   |
| ORF 4                     | 348 bp   |
| ORF 5                     | 1.644 bp |
| ORF 6                     | 651 bp   |
| ORF 7                     | 549 bp   |

Tabelle 8.7: Dritte ORF-Abfolge der ST2-Markersuche und ihre Länge

| <i>open reading frame</i> | Länge  |
|---------------------------|--------|
| ORF 1                     | 681 bp |
| ORF 2                     | 477 bp |

Tabelle 8.8: Auftreten des Block 7 in Non-ST2-Stämmen in Relation ST2-Marker 1

| Stamm     | Anteil des Block 7 in Basen pro 3.582 bp | Lage des Sequenzblocks 7 in Relation zum ST2-Marker 1 | Anteil des ST2-Marker 1 in Basen pro 34.798 bp | Marker-nachweis |
|-----------|------------------------------------------|-------------------------------------------------------|------------------------------------------------|-----------------|
| ENVH115   | 95% (3.414 bp)                           | Dissoziiert                                           | 100% (34.798 bp)                               | P               |
| ACEF09    | 100% (3.582 bp)                          | Dissoziiert                                           | 100% (34.798 bp)                               | P               |
| ATCC12228 | 100% (3.582 bp)                          | <i>upstream</i> des vollständigen Markers             | 100% (34.798 bp)                               | P               |
| VCU057    | 73% (2.624 bp)                           | Dissoziiert                                           | 97% (33.792 bp)                                | P               |
| Scl19     | 11% (389 bp)                             | Dissoziiert                                           | 71% (24.662 bp)                                | P               |
| CIM40     | 7% (254 bp)                              | Dissoziiert                                           | 42% (14.620 bp)                                | N               |
| APO35     | 17% (617 bp)                             | Dissoziiert                                           | 56% (20.381 bp)                                | N               |
| NIH051475 | 100% (3.582 bp)                          | Dissoziiert                                           | 100% (34.798 bp)                               | P               |
| NIH051668 | 100% (3.582 bp)                          | <i>upstream</i> des vollständigen Markers             | 100% (34.798 bp)                               | P               |
| 528m      | 100% (3.582 bp)                          | <i>upstream</i> eines trunkierten Markers             | 99% (34.746 bp)                                | P               |
| 41tr      | 100% (3.582 bp)                          | <i>upstream</i> eines trunkierten Markers             | 99% (34.739 bp)                                | P               |
| VCU041    | 92% (3.311 bp)                           | Dissoziiert                                           | 99% (34.624 bp)                                | P               |
| 36-1      | 99% (3.570 bp)                           | Dissoziiert                                           | 96% (33.391 bp)                                | P               |

P positiv, N negativ

## 9. Danksagung

Ich möchte mich zunächst bei meinem Doktorvater Prof. Dr. Johannes Knobloch für die Bereitstellung des Themas bedanken. Er hatte ein offenes Ohr für meine diversen Fragen, Probleme sowie Vorschläge. Ebenso hat er den letzten Dissertationsentwurf kritisch gelesen.

Des weiteren danke ich dem ehemaligen Institutsleiter, Prof. Dr. Werner Solbach sowie dem aktuellen Institutsleiter, Prof. Dr. Jan Rupp, für die Möglichkeit der Promotion in der Klinik für Infektiologie und Mikrobiologie. Ich bedanke mich bei den weiteren Mitarbeiterinnen und Mitarbeitern des Instituts für die freundliche Aufnahme. Insbesondere danke ich den Mitgliedern der Arbeitsgruppe Knobloch für das entspannte Arbeitsklima und ihre Unterstützung. Besonders erwähnt werden soll in diesem Zusammenhang Justina Bimschas für ihre freundliche Art sowie ihre Hilfe bei den einzelnen Sequenzierungen. Kristina Bäsler gebührt ebenfalls ein großer Dank, ohne sie wäre diese Arbeit nicht möglich gewesen. Ihre Arbeit bei den Sequenzierungen war eine wertvolle Hilfe für die Erstellung der Dissertation. Zusätzlich hat sie mich sehr unterstützt bei den diversen Problemen, die im Prozess der Erstellung dieser Dissertation entstanden sind.

Meinen Freunden, insbesondere Frau Hannah Wedman und Frau Theresa Fernandes danke ich für die konstruktive Kritik an einzelnen Abschnitten dieser Dissertation. Ihre Unterstützung ist seit annähernd dem Beginn meines Studiums eine verlässliche Stütze. Auch Nora Kiobassa möchte ich für die Kritik einzelner Abschnitte meiner Arbeit danken.

Ebenfalls danke ich meinen Eltern, Martina und Karl-Heinz Maurer, sowie meiner Schwester, Lisa Charlotte Maurer, für die vielen aufbauenden Worte und die Ermöglichung all meiner bisherigen Vorhaben. Ich bin meinen Eltern dankbar für die hilfreiche Kritik an meinem Dissertationsentwurf.

Ebenso möchte ich mich bei jenen Menschen bedanken, die mir in meinem bisherigen Leben hilfreich zur Seite standen oder mich prägten, aber im Rahmen dieser Danksagung nicht mehr genannt werden konnten.

Philipp Maximilian Maurer

## 10. Lebenslauf

Name: Philipp Maximilian Maurer

Geburtsdatum: 3.12.1991

Geburtsort: Haan

Familienstand: ledig



Ausbildung:

1998-2002            Grundschule Millrath

2002-2011           Gymnasium Hochdahl

Mai 2011            Abschluss: Abitur

seit Oktober 2011   Medizinstudium an der Universität zu Lübeck

seit April 2014:    Doktorarbeit in der Klinik für Infektiologie und  
Mikrobiologie Lübeck bei Prof. Dr. Knobloch