

Aus dem Universitätsklinikum Münster
Poliklinik für Parodontologie
- Kommissarischer Leiter: Univ.-Prof. Dr. med. D. Harmsen -

**Evaluation des BURP-Algorithmus
für die phylogenetische Analyse von
*Staphylococcus aureus***

INAUGURAL - DISSERTATION
zur
Erlangung des doctor medicinae

der Medizinischen Fakultät
der Westfälischen Wilhelms-Universität Münster

vorgelegt von

Berssenbrügge, Christoph
aus Georgsmarienhütte

2008

Gedruckt mit Genehmigung der Medizinischen Fakultät der Westfälischen
Wilhelms-Universität Münster

Dekan: Univ.-Prof. Dr. med. V. Arolt

1. Berichterstatter: Univ.-Prof. Dr. med. D. Harmsen

2. Berichterstatter: Univ.-Prof. Dr. rer. nat. H. Karch

Tag der mündlichen Prüfung: 24. Januar 2008

Aus dem Universitätsklinikum Münster

Poliklinik für Parodontologie

- Kommissarischer Leiter: Univ.-Prof. Dr. med. D. Harmsen -

Referent: Univ.-Prof. Dr. med. D. Harmsen

Koreferent: Univ.-Prof. Dr. rer. nat. H. Karch

ZUSAMMENFASSUNG

Evaluation des BURP-Algorithmus für die phylogenetische Analyse von

Staphylococcus aureus

Berssenbrügge, Christoph

Für epidemiologische Ausbruchsuntersuchungen von *Staphylococcus aureus* hat sich in den vergangenen Jahren die Sequenzierung der VNTR-Region des *spa* Gens etabliert. Zudem wird seit einiger Zeit vermutet, dass diese *spa* Typisierung evolutionsbiologisch auch ältere Prozesse darstellt. Als ein komplexes Mutationsereignis in genetischen Repeatregionen wird das sog. „slipped-strand mispairing“ angenommen, das durch das sog. EDSI-Modell formal beschrieben wird.

Der darauf aufbauende neue Algorithmus BURP (Based Upon Repeat Patterns) gruppiert *S. aureus* Isolate basierend auf den *spa* Repeatprofilen. Zur Evaluation des Algorithmus für die phylogenetische Analyse von *S. aureus* wurden drei in der Literatur beschriebene Datensätze untersucht. Außerdem wurden 447 Isolate mittels MLST charakterisiert, die als Referenzverfahren für Langzeitstudien gilt.

Die BURP-Gruppierung von *S. aureus* Isolaten sowie ihre Übereinstimmung mit den Gruppierungen durch herkömmliche Methoden war abhängig von den gewählten BURP-Parametern. Es ließ sich in allen Datensätzen eine Konkordanz von mindestens 87% zu Referenzverfahren erzielen. In einem populationsbasierten Datensatz betrug die Konkordanz zu einer auf MLST-Daten beruhenden Gruppierung bis zu 97%.

Zusammenfassend ermöglicht BURP auf der Basis eines einzelnen Genlocus automatische Gruppierungen von *S. aureus* Isolaten, die mit denjenigen klassischer Gruppierungsmethoden weitgehend korrelieren. Diese Arbeit stellt zudem die Grundlage für weitere phylogenetische Studien an *S. aureus* und anderen bakteriellen Spezies sowie für mögliche krankenhaushygienische Anwendungen.

Tag der mündlichen Prüfung: 24. Januar 2008

Meinen Eltern

Inhaltsverzeichnis

1.	Einleitung	1
1.1	<i>Staphylococcus aureus</i>	1
1.2	Typisierungsmethoden.....	2
1.3	Fragestellung.....	10
2.	Material und Methoden	11
2.1	Labormaterial.....	11
2.1.1	Laborgeräte.....	11
2.1.2	Verbrauchsmaterialien.....	11
2.1.3	Chemikalien.....	11
2.1.4	Gebrauchslösungen.....	11
2.2	Labormethoden.....	12
2.2.1	Extraktion der DNA.....	12
2.2.2	Amplifikation der DNA.....	12
2.2.3	Aufreinigung der Amplifikationsprodukte.....	14
2.2.4	Sequenzierung der DNA.....	14
2.3	Auswertungsmethoden.....	15
2.3.1	Auswertung der Daten der MLST und eBURST.....	15
2.3.2	Auswertung der Daten der <i>spa</i> Typisierung und BURP.....	17
2.3.3	Statistische Auswertungen.....	18
2.4	Datensätze.....	19
2.4.1	Datensatz Münster.....	20
2.4.2	Datensatz Korea.....	21
2.4.3	Datensatz Robinson.....	21
2.4.4	Datensatz Nottingham.....	22
3.	Ergebnisse	23
3.1	Datensatz Münster.....	23
3.2	Datensatz Korea.....	29
3.3	Datensatz Robinson und Datensatz Nottingham.....	31
4.	Diskussion	33
4.1	Datensatz Münster.....	33

4.1.1	eBURST im Datensatz Münster	33
4.1.2	Einfluss des Ausschluss-Parameters auf die BURP-Einteilung ..	34
4.1.3	Einfluss des Kosten-Parameters auf die BURP-Einteilung	34
4.1.4	Einfluss des Ausschluss-Parameters auf die Konkordanz	35
4.1.5	Einfluss des Kosten-Parameters auf die Konkordanz.....	36
4.1.6	Beispiele für die Konkordanz.....	37
4.1.7	Selektionsbias im Datensatz Münster	38
4.2	Datensatz Koreen	39
4.3	Datensatz Robinson.....	40
4.4	Datensatz Nottingham.....	41
4.5	Einschränkungen	43
4.5.1	Problem der biologischen Repräsentativität	43
4.5.2	Problem der Typisierbarkeit von Isolaten	44
4.5.3	Problem der Wahl der BURP-Parameter.....	45
4.5.4	Problem der Angabe von Gründungsvätern	46
4.5.5	Problem der chromosomalen homologen Rekombinationen.....	47
4.6	Ausblick	50
5.	Literaturverzeichnis	53
6.	Verzeichnis der Abkürzungen	59
7.	Verzeichnis der Abbildungen, Tabellen und Diagramme.....	61
7.1	Verzeichnis der Abbildungen.....	61
7.2	Verzeichnis der Tabellen.....	61
7.3	Verzeichnis der Diagramme	61
8.	Danksagung.....	62
9.	Lebenslauf	63
Anhang	I
Anhang A:	Datensatz Münster	I
Anhang B:	Datensatz Koreen	VII
Anhang C:	Dendrogramm, Datensatz Koreen.....	VIII
Anhang D:	Datensatz Robinson.....	IX
Anhang E:	Datensatz Nottingham.....	XI

1. Einleitung

1.1 *Staphylococcus aureus*

Der Wundkeim *Staphylococcus (S.) aureus* ist weltweit der häufigste Erreger nosokomialer (d.h. im Krankenhaus erworbener) Infektionen [3]. Er wird u.a. bei Wundinfektion, Pneumonie, Endokarditis, Sepsis, Toxic-Shock-Syndrom, septischer Arthritis, hämatogener Osteomyelitis, Abszessbildung oder Katheter-assoziierten Infektionen nachgewiesen.

Parallel zu der steigenden Bedeutung von *S. aureus* als Infektionserreger hat sich die Resistenzsituation gegenüber einer Reihe von Antibiotika deutlich und zunehmend verschlechtert. Besonders kritisch sind dabei Infektionen durch Methicillin-resistente *S. aureus* (MRSA). Nach Einführung des Methicillins für die Behandlung von Penicillin-resistenten *S. aureus* im Jahr 1959 traten MRSA erstmals im Jahr 1961 in Großbritannien auf und wurden innerhalb kürzester Zeit weltweit nachgewiesen, wo sie neben sporadischen Infektionen zunehmend Infektionsausbrüche auslösten [19, 42]. Für MRSA gibt es nur noch wenige Möglichkeiten der antibiotischen Therapie, denn es wirken nur noch Reserveantibiotika wie Glykopeptide (z.B. Vancomycin), die allerdings meist sehr teuer und nebenwirkungsreicher sind. Zudem konnte gezeigt werden, dass systemische Infektionen mit MRSA im Vergleich zu solchen mit Methicillinsensiblen *S. aureus* (MSSA) eine erhöhte Mortalität aufweisen [2]. Neben längeren und schwereren Krankheitsverläufen für die betroffenen Patienten bedeutet das Auftreten von MRSA äußerst arbeitsaufwändige und für Krankenhäuser sehr kostenintensive Konsequenzen, im Ausbruchsfall kann die Schließung ganzer Stationen nötig sein.

Weltweit stellen MRSA-Infektionen ein eskalierendes Problem in stationären Einrichtungen dar. Neben Ländern mit kaum noch beherrschbarer MRSA-Situation (u.a. Japan, USA, Spanien, Italien, Frankreich, England), die einen Anteil von 20-60% MRSA an allen *S. aureus* Infektionen aufweisen, sind

diejenigen Länder hervorzuheben, die infolge strikter Kontroll- und Präventionsstrategien ihre MRSA-Inzidenzen auf wenige Prozent beschränken konnten (Niederlande, skandinavische Länder). In Deutschland stieg der MRSA-Anteil an allen Infektionen mit *S. aureus* von 1,7 Prozent (1990) über 15,2 Prozent (1998) und 20,7 Prozent (2001) auf 22,6% (2004) [24]. Deutschland liegt damit im europäischen Mittelfeld, in den USA gibt es regional stark schwankende MRSA-Raten bis über 30 Prozent [36].

Außerhalb des Krankenhauses verursachen sog. community-acquired MRSA (cMRSA) zunehmend medizinische Probleme [22]. Doch über die Problematik der Methicillin-Resistenz hinaus besteht das Risiko einer Ausbreitung der gegenwärtig insbesondere in Japan und in den USA beobachteten MRSA-Stämme mit zusätzlich verminderter Empfindlichkeit gegenüber Glykopeptiden (GISA oder VISA, erstmals 1997 beschrieben [17]) sowie des erstmals 2002 nachgewiesenen VRSA (Vancomycin-resistente MRSA [1]).

S. aureus gehört zu der physiologischen Hautflora, wobei er vorrangig im Vestibulum nasi angetroffen wird. In einer britischen Studie war in dieser Lokalisation etwa ein Viertel der älteren Bevölkerung mit dem Keim kolonisiert [15]. Doch auch auf unbelebten Gegenständen kann der Erreger aufgrund seiner hohen Umweltresistenz lange persistieren und von dort wieder auf Menschen übertragen werden. Zur Vermeidung nosokomialer Infektionen werden Hygienemaßnahmen empfohlen, unter denen auch die hygienische Händedesinfektion des medizinischen Personals als sehr wirksam gilt [32].

1.2 Typisierungsmethoden

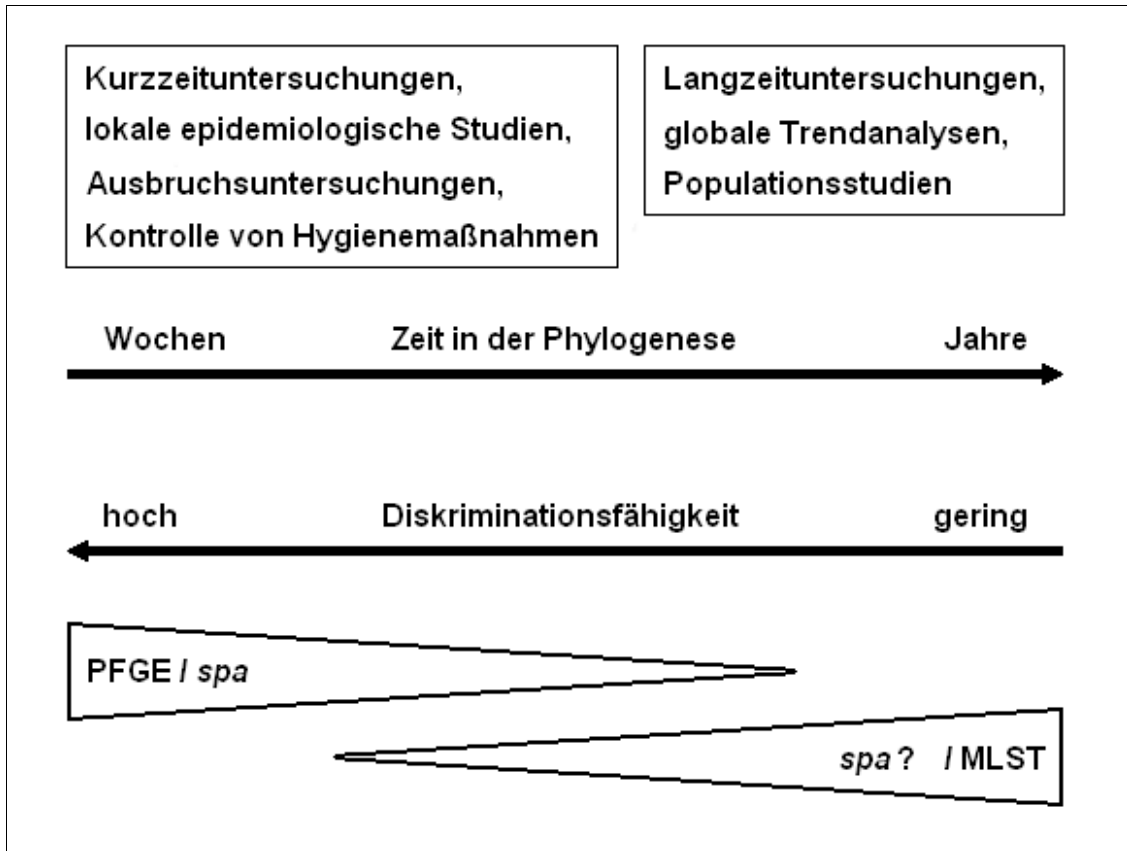
Für die Untersuchung von Krankheitserregern stehen verschiedene epidemiologische Typisierungsmethoden zur Verfügung. Generell dienen sie der Untersuchung von Populationsdynamiken und Ausbreitungstudien von Bakterien oder anderen Mikroorganismen mit klonaler, nicht-sexueller Reproduktion. Spezielle Fragestellungen sind dabei Populationsstudien von

Bakterien, Untersuchungen über Infektionswege und Ausbrüche sowie die Überwachung von Infektionskrankheiten [44]. Um Infektionsquellen und klonale Beziehungen zwischen Isolaten von *S. aureus* zu untersuchen, ist die Typisierung von Isolaten nicht nur in klinischen Proben, sondern auch in Populationsstudien in zunehmendem Maße ein wichtiges Werkzeug geworden. So ist es möglich, die Übertragungswege und die Epidemiologie von Isolaten genauer aufzuklären und Hinweise auf einen möglichen Ausbruch eines einzelnen Stammes besser bewerten zu können. Auf diese Weise können die Indikationen zu geeigneten hygienischen oder therapeutischen Maßnahmen gegebenenfalls zügiger gestellt werden.

Wegen der unterschiedlichen genetischen Variationsgeschwindigkeiten („clock speed“), die den Zielstrukturen von Typisierungsmethoden zugrunde liegen können, wird zwischen Langzeituntersuchungen und Kurzzeituntersuchungen unterschieden. Bisher konnte keine Methode etabliert werden, die für beide Zwecke gleichermaßen geeignet ist [23]. Somit muss derzeit in Abhängigkeit von der jeweiligen Fragestellung individuell eine geeignete Typisierungsmethode ausgewählt werden (vgl. Abbildung 1-1).

Kriterien für die Charakterisierung oder Evaluation von Typisierungsmethoden werden nach Struelens et al. in die beiden Kategorien Effektivität („performance“) und Effizienz („convenience“) eingeteilt [44]. Zu der ersten Gruppe werden Typisierbarkeit, Reproduzierbarkeit, Stabilität und Diskriminationsfähigkeit gezählt, die relativ einfach quantifizierbar sind. Beispielsweise wurde für die *spa* Typisierung von *S. aureus* die Stabilität des *spa* Locus in vitro und in vivo von verschiedenen Gruppen untersucht [11, 12, 21, 37, 39]. Die Kriterien der zweiten Gruppe, Anwendungs- und Interpretationsaufwand, sind rechnerisch nur schwierig zu bestimmen, aber mindestens ebenso bedeutsam [18, 39, 45]. So hat jede bekannte Technik ihre eigenen Vor- und Nachteile, die abgewogen werden sollten, bevor ein Verfahren zur Anwendung kommt [46].

Abbildung 1-1: Variationsgeschwindigkeit von Typisierungsmethoden.



Für epidemiologische Untersuchungen von Bakterien werden bei Kurzzeituntersuchungen Typisierungsmethoden mit möglichst hoher Diskriminationsfähigkeit verwendet [39]. Genau diese hohe Diskriminationsfähigkeit fehlt üblicherweise Methoden, die globale Dynamiken abbilden können und dadurch Langzeitauswertungen und Populationsstudien ermöglichen [40]. Inwieweit im *spa* Locus nicht nur genetische Kurzzeitinformationen enthalten sind, sondern auch phylogenetisch lange Zeiträume widerspiegelt werden, ist eine Frage dieser Arbeit.

Für Ausbruchsuntersuchungen und für Vorhersagen zukünftiger Trends galt lange Zeit die bandenbasierte Pulsfeldgelelektrophorese (PFGE, *Sma*I Makrorestriktionsanalyse) als Goldstandard [31]. Die PFGE bietet eine sehr hohe Diskriminationsfähigkeit, die für Kurzzeituntersuchungen gewünscht ist, für Langzeitauswertungen aber zu hoch ist [45]. In Hinblick auf den Anwendungsaufwand ist die PFGE mit vier bis sechs Arbeitstagen personal-, zeit- und damit kostenintensiv und daher nicht zufrieden stellend. Sie kann keine zeitnahe Untersuchung von Ausbrüchen gewährleisten, sondern liefert meist nur retrospektive Ergebnisse [29, 45], was zudem durch einen hohen Interpretationsaufwand erschwert wird. Die PFGE lässt sich nur mit großem Aufwand so standardisieren, dass ein Ergebnisvergleich zwischen verschiedenen Untersuchungsansätzen oder gar verschiedenen Laboratorien möglich ist [31, 48]. Aufgrund dieser Probleme wird sie im klinischen Alltag üblicherweise nicht

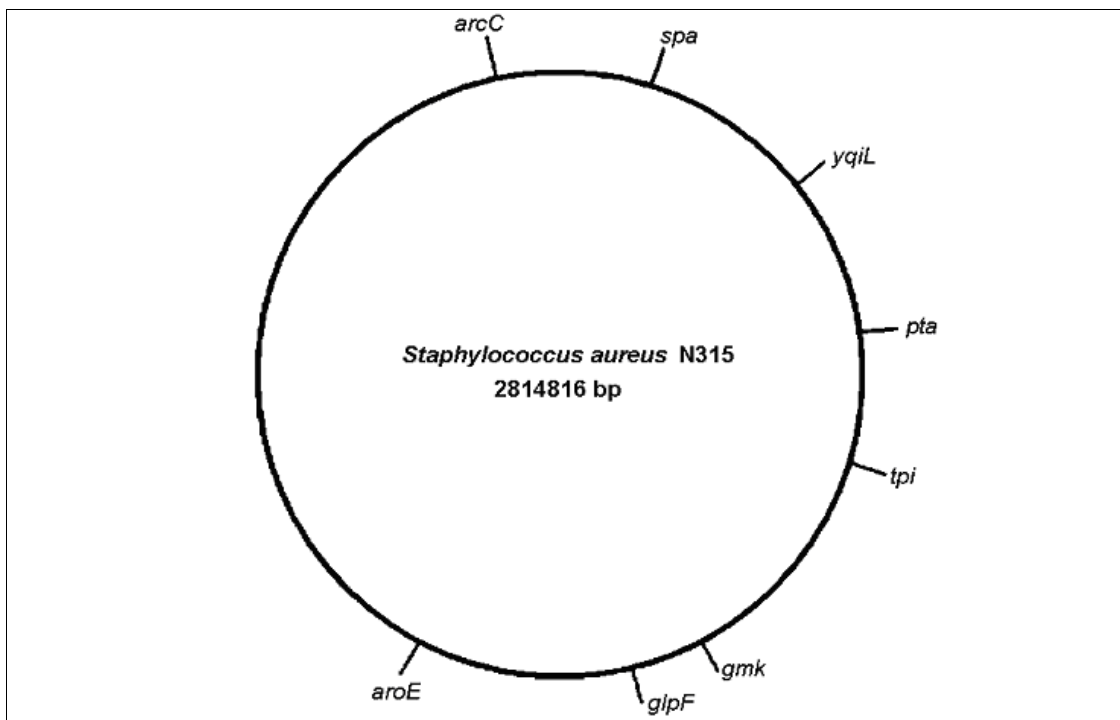
genutzt, ihr Einsatz ist auf wenige Zentren und Referenzlaboratorien beschränkt.

Viele der für die PFGE geschilderten Probleme existieren bei sequenzbasierten Methoden nicht, weshalb diese in den letzten Jahren an Bedeutung gewonnen haben. Diese Entwicklung wurde auch durch kostensenkende technische Fortschritte in diesem Bereich begünstigt. Die Verfahren liefern innerhalb von ein bis zwei Tagen mit vergleichsweise geringem Arbeits- und Interpretationsaufwand eindeutige Ergebnisse, was eine schnelle Diagnostik und damit eine zeitnahe Reaktion in Form von Hygienemaßnahmen ermöglicht [39]. Die Daten lassen sich relativ leicht elektronisch verwalten, und das Internet ermöglicht einen einfachen Datenaustausch zwischen verschiedenen Laboratorien [16, 27].

Die MLST (Multi Locus Sequence Typing) charakterisiert einen Bakterienstamm durch Polymorphismen seiner Nukleotidsequenzen interner Fragmente von sieben Haushaltsgenen (vgl. Abbildung 1-2). Die Sequenzen haben jeweils eine Länge von etwa 450 bp. Für jedes Genfragment an jedem einzelnen Locus werden verschiedene Sequenzen als verschiedene Allele angesehen, und es werden ihnen verschiedene Nummern zugeteilt. So ist jeder Stamm durch die Allele an den sieben Loci, sein allelisches Profil, definiert. Jedem allelischen Profil wird ein Sequenztyp (ST) zugeordnet, der eine eindeutige Beschreibung für den Bakterienstamm ist. Isolate mit demselben Sequenztyp werden als Angehörige eines Klons angesehen. Isolate mit einer einzelnen Abweichung im allelischen Profil (SLV, single locus variants) oder zwei Abweichungen (DLV, double locus variants) können je nach Definition als verwandt bezeichnet werden. Das Maß der allelischen Übereinstimmung kann damit als Gradmesser für die Verwandtschaft gelten. Die MLST wurde nach der Entwicklung am Modellorganismus *Neisseria meningitidis* durch Maiden et al. [27] durch Enright et al. an *S. aureus* adaptiert [4, 5]. Die diskriminatorische Fähigkeit der MLST ist höher als die der Phagentypisierung und erreicht je nach untersuchter Stammsammlung beinahe die der PFGE [13]. Dennoch kann die MLST insbesondere in

Kombination mit dem Gruppierungsalgorithmus eBURST für evolutionäre Globaluntersuchungen von *S. aureus* genutzt werden, sie gilt oft sogar als Referenzmethode [4, 5, 7]. Sie ist mit der notwendigen Sequenzierung von sieben Loci jedoch vergleichsweise teuer und daher ebenfalls kein routine-taugliches Typisierungsverfahren für *S. aureus* Isolate.

Abbildung 1-2: Positionen der MLST-Loci und des *spa* Locus.

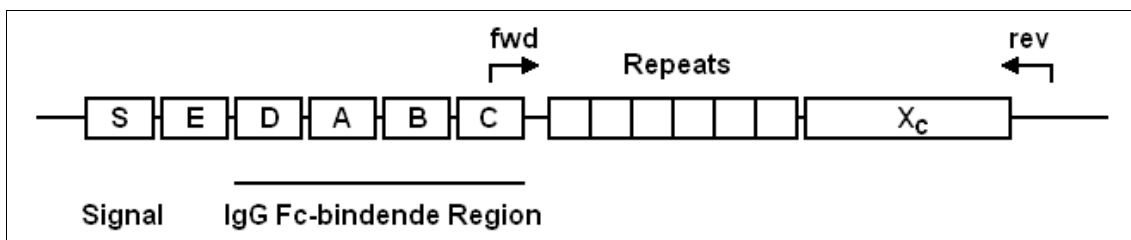


Die chromosomalen Positionen der sieben Loci der MLST sowie des *spa* Locus in der Kreisrepräsentation des Referenzstamms *S. aureus* N315 (GenBank acc. no. BA000018, <http://www.ncbi.nih.gov/Genbank>, Datum 07.12.2004), den Kuroda et al. im Jahr 2001 für die erste Genomsequenzierung eines *S. aureus* nutzten [26]. Erstellt mit MacPlasmap Version 2.1 (CGC Scientific, Inc.). Der *spa* Locus befindet sich relativ nahe an dem MLST-Locus *arcC*.

Im Vergleich zu der MLST existiert mit der *spa* Typisierung eine Methode, die bei ebenso einfachem Datenaustausch durch die Bestimmung nur eines Locus (Single Locus Sequence Typing, SLST) wesentlich kostengünstiger ist. Frénay et al. zeigten als erste, dass mit der Sequenz eines Genabschnitts des Protein A Gens eine Typisierung von *S. aureus* (*spa*) möglich ist [11, 12]. Protein A ist ein Bestandteil der Zellwand von *S. aureus*. Das entsprechende Gen hat eine Länge von etwa 2150 bp und besteht aus drei verschiedenen Regionen, der Fc-Region, der hochvariablen X-Region und dem C-Terminus. Die molekulare Basis der *spa* Typisierung ist die DNA-Sequenzanalyse der X-Region. Diese

besteht aus einer variablen Anzahl von Repeats (Variable Number of Tandem Repeats, VNTR [47]), kurze wiederkehrende Sequenzabschnitte mit einer Länge von 21-27 Basenpaaren (vgl. Abbildung 1-3). Aus der Art und der Abfolge der Repeats resultiert der *spa* Typ und damit das Typisierungsergebnis. Dieser Ansatz wurde intensiv verfolgt [39, 45]. Im November 2004 wurde das europäische Forschernetzwerk „SeqNet.org“ gegründet [10].

Abbildung 1-3: Der *spa* Locus von *S. aureus*.



Repräsentation des *spa* Gens, das für Protein A codiert. Die Kästchen zeigen die Segmente für die Signalsequenz (S), die IgG-bindende Region (A-D), eine zu A-D homologe Region (E) und den COOH-Terminus, der aus den Repeats und der Zellwandankerssequenz Xc besteht. In diesem Beispiel liegen sechs Repeats zwischen der Fc-Region und der Xc-Region. Die Positionen des Forward- und des Reverse-Primers werden durch Pfeile gezeigt. Aus der Art und der Abfolge der Repeats resultiert der *spa* Typ. Modifiziert nach Frénay et al. [11] und Shopsin et al. [39].

Wegen eines Mangels an geeigneter Software für die Identifikation und eines zunächst fehlenden Konsens bezüglich der Nomenklatur (fortlaufende alphabetische oder fortlaufende numerische Bezeichnung) der *spa* Repeats verzögerte sich die Verbreitung der Methode bis zu der Einführung der Ridom StaphType Software durch Harmsen et al. [16]. Diese Software ermöglicht eine effektive Auswertung der Sequenzdaten. Sie liest automatisch Sequenzen ein, erkennt Repeatmuster und verwaltet in einer hinterlegten Datenbank neben den Typisierungsergebnissen epidemiologisch relevante Daten. Zudem ist sie so angelegt, dass über einen Internet-basierten Server die Typisierungsergebnisse zwischen einzelnen Laboratorien synchronisiert werden können, was eine für alle Nutzer geltende einheitliche (fortlaufende numerische) Nomenklatur für neu auftretende Repeats oder *spa* Typen ermöglicht. Somit ist es möglich, Vergleiche durchzuführen und überregionale Dynamiken zu untersuchen.

Bei der Betrachtung von Datenbanken über *S. aureus* gewinnt man den Eindruck, dass es zwischen bestimmten *spa* Repeats und Repeatprofilen sowie bestimmten MLST-Sequenztypen Assoziationen gibt. Es wurde daher die Hypo-

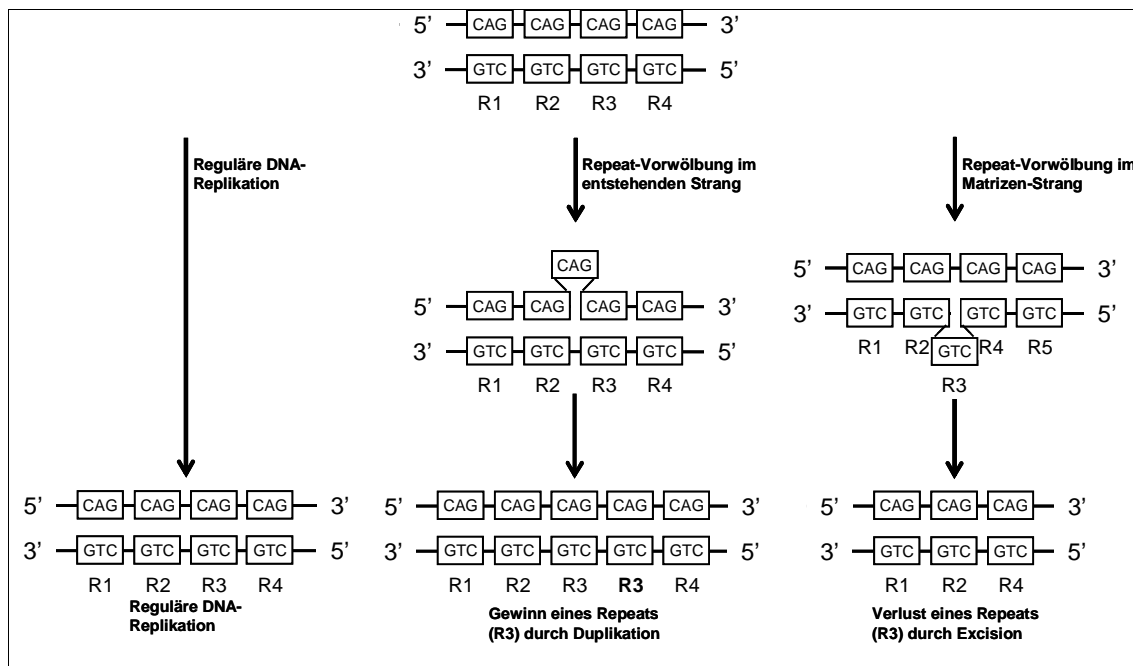
these aufgestellt, dass die *spa* Region nicht nur die Kurzzeit-, sondern auch die Langzeitepidemiologie von *S. aureus* reflektiert [21, 39, 47]. Basierend auf den Repeatprofilen wurden bereits manuelle Gruppierungen von *spa* Typen vorgenommen [23]. Bislang stand aber kein automatischer oder objektiver Algorithmus zur Verfügung, um aus den *spa* Typen auf die klonale Verwandtschaft verschiedener *spa* Typen zu schließen.

In einem herkömmlichen Modell für Mutationsereignisse wurde angenommen, dass die fortschreitende Veränderung von DNA-Sequenzen ausschließlich auf Punktmutationen beruhe, also auf Basensubstitutionen, -insertionen oder -deletionen, wobei die beiden zuletzt genannten Ereignisse unter dem Begriff „Indels“ zusammengefasst wurden. In der Literatur ist dieses Modell als SI-Modell bezeichnet worden [38]. Algorithmen zum Vergleich von Sequenzen, dem sogenannten Alignment, beruhten lange auf diesem Modell. Sequenzen wurden virtuell ineinander überführt, indem schrittweise Punktmutationen an ihnen simuliert wurden. Um den Aufwand zu quantifizieren, der für diese Umwandlungsprozesse nötig ist, kann die Zahl der Ereignisse zugrunde gelegt werden und in Kosten angegeben werden. Wenn man als ein Grundprinzip der Evolution die Vermeidung von Aufwand oder Umwegen annimmt (größtmögliche Sparsamkeit, „maximum parsimony“), minimieren gute Alignments die Zahl der Veränderungen in der evolutionären Entwicklung und damit die Kosten.

Es wird angenommen, dass die in VNTR-Loci beobachtete Variabilität durch komplexe Mutationsereignisse verursacht wird, die nicht nur einzelne Nukleotidpositionen, sondern ganze Abschnitte von Genen betreffen. Während der DNA-Replikation können Replikationsfehler durch Vorwölbungen oder Schleifenbildungen in der Repeatregion zur Deletion oder Insertion ganzer Repeats führen, die dann als Exzisionen und Duplikationen bezeichnet werden können. Die englischsprachige Bezeichnung für diese Mutationsereignisse lautet „slipped-strand mispairing“, was übersetzt werden kann mit „Fehlanlagerung der DNA-Stränge durch Auslassen einer Schlaufe“ (vgl. Abbildung 1-4). Die

betreffenden Genabschnitte sind dabei nicht nur länger als bei Einzelnukleotidmutationen, sondern sie enthalten das Muster des Tandem Repeats in der korrespondierenden Sequenz. Herkömmliche Punktmutationen, d.h. Substitution oder Indels (Insertion oder Deletion) einzelner Nukleotide, können weiterhin zusätzlich zu diesen Mutationsprozessen stattfinden [47].

Abbildung 1-4: Mechanismus des „slipped-strand mispairing“.



Schematische Repräsentation des SSM-Mechanismus während der DNA-Replikation eines VNTR (Variable Number of Tandem Repeats)-Locus. Einzelne Repeats (R1, R2, R3, R4, R5) werden durch Kästchen angezeigt. Vorwölbungen oder Schleifenbildungen, also ungepaarte Basen, die den regulären DNA-Doppelstrang unterbrechen, resultieren in einer Verkürzung oder Verlängerung. Eine Vorwölbung im Matrizen-Strang führt zu einer geringeren Zahl an Repeats im entstehenden Strang und umgekehrt. Während der Replikation kann die Länge des Locus um mehrere Repeats zu- oder abnehmen, wenn in einem Strang mehrfache Vorwölbungen auftreten. Die Vorwölbungen können in beiden Strängen auftreten, und der Effekt von Duplikation oder Exzision kann durch das Auftreten des gegensätzlichen Ereignisses neutralisiert werden. Modifiziert nach van Belkum et al. [47]. Dieser Mechanismus ist Grundlage des EDSI-Modells, auf dem der BURP-Algorithmus aufbaut.

Um diese Abläufe zu beschreiben, entwickelten Sammeth et al. das bisherige SI-Modell weiter zu dem EDSI-Modell (Exzision, Duplikation, Substitution, Indels) [38]. Auf der Grundlage des EDSI-Modells klassifiziert der Algorithmus BURP (Based Upon Repeat Patterns) *spa* Typen von *S. aureus* ausgehend von den Repeatprofilen in klonale Komplexe, sog. BURP-Cluster oder *spa*-CC. Der Algorithmus findet theoretisch mögliche Mutationsprozesse und schreibt ihnen Kosten entsprechend dem Aufwand ihrer Veränderungen zu. Am Ende wird derjenige Veränderungsprozess ausgewählt, der die geringsten Kosten

verursacht. Wenn man für die Natur das oben erwähnte Prinzip der Vermeidung von Aufwand oder Umwegen annimmt, kann man diese Kosten als ein Maß für die genetische Distanz der verglichenen Sequenzen ansehen.

1.3 Fragestellung

Zur Evaluation des BURP-Algorithmus für die phylogenetische Analyse von *S. aureus* wurden in der vorliegenden Arbeit vier Datensätze erstmals mit BURP untersucht. Dadurch werden Antworten auf die folgenden Fragen gegeben:

Inwieweit sind im *spa* Locus genetische Informationen enthalten, die phylogenetisch weit in die Vergangenheit reichende Zeiträume widerspiegeln?

Wie wirken sich Veränderungen an den beiden BURP-Parametern, die von dem Anwender einzustellen sind, auf die Gruppierungsergebnisse aus?

Kann BURP zuverlässig Abstammungslinien reproduzieren, die durch etablierte Typisierungs- und Gruppierungsmethoden angezeigt werden? Wie ist dabei insbesondere die Beziehung zu der für Langzeituntersuchungen als Referenzmethode geltenden MLST?

2. Material und Methoden

2.1 Labormaterial

2.1.1 Laborgeräte

Es wurden Standardlaboreinrichtungen der Firmen Bender und Hobein (Zürich, CH), Eppendorf (Hamburg), Heidolph (Schwabach), Kirsch (Offenburg), von Keutz (Reiskirchen) und Merck Eurolab (Leuven, B) verwendet. Als Thermocycler wurden die Modelle TGradient und T1Thermocycler der Firma Biometra (Göttingen) genutzt.

2.1.2 Verbrauchsmaterialien

Standardplastik- und Glaswaren wurden von den Firmen ABgene House (Surrey, UK), Corning Incorporated (Corning, NY, USA) und Eppendorf bezogen. Zusätzlich wurden von der Firma Millipore (Schwallbach) 96-well Mikrotiterplatten, MultiScreen Column Loader 45 µl und MultiScreen 96-well Filtration Plates verwendet.

2.1.3 Chemikalien

Es wurden Standardchemikalien von den Firmen Bio-Rad (München sowie Herkules, CA, USA), Biozym (Hessisch Oldendorf), Invitrogen life technologies (Karlsruhe), Merck (Darmstadt), Roth (Karlsruhe) und Sigma Aldrich (Deisenhofen) verwendet. Aqua bidestillata entstammte der Eigenherstellung des Instituts für Hygiene (Münster), das HPLC Wasser LiChrosolv wurde von Merck (Darmstadt) bezogen.

2.1.4 Gebrauchslösungen

Für die Herstellung des TE-Puffers wurden 121,44 mg Tris, 3,72 mg EDTA und 100 ml HPLC Wasser verwendet, mit 32% Salzsäure (ca. 80 µl) wurde ein pH von 7,5 erreicht. Für den 5x-ABI-Sequenzierungspuffer wurden 400 µl der Stammlösung A (12,1 g Tris, 80 µl HPLC Wasser, 32% HCl (ca. 1,5 ml) auf pH 9, ad 100 ml HPLC Wasser), 10 µl der Stammlösung B (20,3 mg MgCl₂ in 100 ml HPLC Wasser) und 590 µl HPLC Wasser verwendet.

2.2 Labormethoden

Die MLST wurde mit Modifikationen nach dem Schema von Enright et al. durchgeführt [4].

2.2.1 Extraktion der DNA

Um DNA zu erhalten, wurden mit einer Öse Zellen einer Kolonie von *S. aureus* in destilliertem Wasser gelöst und mit 100 µl 6% InstaGene™ Matrix Solution (Bio Rad, München) für 20 min bei 56 °C inkubiert. Die Suspension wurde gevortext und für 8 min bei 100 °C erhitzt. Dann wurde für 2-3 min bei 8000 g zentrifugiert. 2 µl des Überstandes wurden für die Amplifikation verwendet. Die Lösungen wurden bei -20 °C aufbewahrt, bevor die PCR durchgeführt wurde.

2.2.2 Amplifikation der DNA

Zur Amplifikation der jeweiligen Genabschnitte der bakteriellen DNA diente die PCR. Sämtliche Oligonukleotide (Primer) stammten von der Firma Sigma Genosys (Steinheim) und wurden in TE-Puffer gelöst (Konzentration: 5 µmol/l). Es wurden HPLC-aufgereinigte Primer (vgl. Tabelle 2-1) verwendet. Detaillierte Informationen zu den untersuchten Genen, den Amplikons der PCR und den für die MLST genutzten Sequenzen finden sich in Tabelle 2-2. Es wurden PuRe Taq Ready-To-Go PCR Beads (Amersham Biosciences, Freiburg) mit einem Gesamtvolumen von 25 µl, das 10 pmol jedes Primers enthielt, genutzt. Entsprechend den Längen der Primer und ihrer Basenzusammensetzung galten

die folgenden PCR-Bedingungen: Die thermischen Reaktionen begannen mit einer initialen Denaturierung (10 min bei 80 °C), danach folgten 35 Zyklen von Denaturierung (45 s bei 94 °C), Anlagerung (60 s bei 55 °C) und Elongation (90 s bei 72 °C) mit einer einzelnen abschließenden Elongation (10 min bei 72 °C). Die erfolgreiche Amplifikation wurde mittels Gelelektrophorese eines Teils des PCR-Produkts dokumentiert.

Tabelle 2-1: Verwendete synthetische Oligonukleotide (Primer).

Locus	Primer	Sequenz
<i>arcC</i>	<i>arcC</i> -165f	5'- TTG ATT CAC CAG CGC GTA TTG TC -3'
	<i>arcC</i> -692r	5'- AGG TAT CTG CTT CAA TCA GCG -3'
<i>aroE</i>	<i>aroE</i> -38f	5'- ATC GGA AAT CCT ATT TCA CAT TC -3'
	<i>aroE</i> -529r	5'- GGT GTT GTA TTA ATA ACG ATA TC -3'
<i>glpF</i>	<i>glpF</i> -46f	5'- CTA GGA ACT GCA ATC TTA ATC C -3'
	<i>glpF</i> -578r ^a	5'- TGG TAA AAT CGC ATG TGC AAT TC -3'
<i>gmk</i>	<i>gmk</i> -46f ^a	5'- ATC GTT TTA TCA GGA CCA TCT G -3'
	<i>gmk</i> -490r	5'- TCA TTA ACT ACA ACG TAA TCG TA -3'
<i>pta</i>	<i>pta</i> -71f	5'- GTT AAA ATC GTA TTA CCT GAA GG -3'
	<i>pta</i> -601r	5'- GAC CCT TTT GTT GAA AAG CTT AA -3'
<i>tpi</i>	<i>tpi</i> -306f	5'- TCG TTC ATT CTG AAC GTC GTG AA -3'
	<i>tpi</i> -736r	5'- TTT GCA CCT TCT AAC AAT TGT AC -3'
<i>yqiL</i>	<i>yqiL</i> -45f	5'- CAG CAT ACA GGA CAC CTA TTG GC -3'
	<i>yqiL</i> -598r	5'- CGT TGA GGA ATC GAT ACT GGA AC -3'

Die Primer-Nomenklatur bezeichnet die 3'-Nukleotidposition des Primers innerhalb des Gens von *S. aureus* (GenBank acc. no. BA000018, Datum 07.12.2004). ^a leicht geänderte Primer im Vergleich zu der Erstbeschreibung [4].

Tabelle 2-2: Daten der MLST für *S. aureus*.

Locus		Gen		Amplikon		MLST-Sequenz	
Name	ORF ^a	Beginn ^b	Länge (bp)	Intervall ^c	Länge (bp)	Intervall	Länge (bp)
<i>arcC</i>	SA2425 (-)	2723049	969	143 / 712	570	199 / 654	456
<i>aroE</i>	SA1424 (-)	1629141	807	16 / 551	536	64 / 519	456
<i>glpF</i>	SA1140 (+)	1296691	819	25 / 600	576	100 / 564	465
<i>gmk</i>	SA1052 (+)	1191032	624	25 / 512	488	73 / 501	429
<i>pta</i>	SA0545 (+)	638340	987	49 / 623	575	97 / 570	474
<i>tpi</i>	SA0729 (+)	835146	762	284 / 758	475	313 / 714	402
<i>yqiL</i>	SA0342 (+)	399116	1182	23 / 620	598	58 / 573	516

Daten nach der Genomsequenz *S. aureus subsp. aureus* N315 (GenBank acc. no. BA000018, Datum 07.12.2004). ^a ORF und Transkriptionsrichtung nach GenBank und Kuroda et al. [26]. ^b Position innerhalb des Genoms (erstes Nukleotid). ^c Positionen innerhalb des Gens (erstes Nukleotid / letztes Nukleotid). Zum Vergleich *spa* Locus: ORF SA0107 (-), Position des ersten Nukleotids innerhalb des Genoms: 124001.

2.2.3 Aufreinigung der Amplifikationsprodukte

Das Amplifikat wurde mit zwei hydrolytischen Enzymen, Exonuklease I (New England Biolabs, Frankfurt) und Shrimp alkalischer Phosphatase (USB Amersham, Cleveland, OH, USA) aufgereinigt, um überschüssige Einzelstrang-DNA zu entfernen. Dazu wurden 5 µl des Amplifikats mit jeweils 1 µl der beiden Enzyme (1 U/µl) für 30 min bei 37 °C inkubiert. Danach wurden die Enzyme für 15 min bei 80 °C hitzeinaktiviert und die Lösungen anschließend auf 4 °C gekühlt.

2.2.4 Sequenzierung der DNA

Die Amplifikate wurden mit dem ABI Prism BigDye Terminator v3.1 Ready Reaction Cycle Sequencing Kit (Applied Biosystems, Darmstadt) sequenziert. Für die Sequenzierung wurden dieselben Primer verwendet wie für die Amplifikation (vgl. Tabelle 2-1). Für die Sequenzierungsreaktion wurden 0,5 µl des Premix, 1,8 µl ABI-Sequenzierungspuffer, 10 pmol des Primers, und 2 µl des gereinigten PCR Produkts in einem Gesamtvolumen von 10 µl genutzt. Alle Sequenzierungsreaktionen wurden mit 25 Zyklen Denaturierung (10 s bei 96 °C), Anlagerung (5 s bei 55 °C) und Elongation (4 min bei 60 °C) bei einem „thermal ramping“ von 1 °C/s durchgeführt.

Das Produkt der Sequenzierungsreaktion wurde für die abschließende Aufreinigung über Kunstharzsäulen geleitet. Es wurden 45 µl Sephadex G 50 Superfine Kunstharz Pulver (AmershamPharmaciaBiotech, Uppsala, S) mit 300 µl HPLC Wasser in eine MultiScreen HV Platte gegeben, für drei Stunden bei Raumtemperatur quellen gelassen und das restliche Wasser abzentrifugiert (5 min bei 2300 U/min). Nach einer Wasserspülung mit 150 µl HPLC und anschließender Zentrifugation (5 min bei 1065 g) wurde die gesamte Sequenzierungsreaktion (10 µl) auf die Säule aufgetragen und durch Zentrifugation (5 min bei 1065 g) in eine Mikrotiterplatte eluiert. Diese aufgereinigte DNA wurde in einer Vakuumzentrifuge („Speedvac“ Concentrator 5301 von Eppendorf, Hamburg)

für 20 min bei 45 °C getrocknet. Danach wurde sie entsprechend den Angaben des Herstellers mit HiDi Formamid (Applied Biosystems) vorbereitet, um in den automatischen Kapillarsequenzierer ABI Prism 3100 Avant Genetic Analyzer (Applied Biosystems) geladen zu werden, der mit der Sequenzierungssoftware 3100 Avant Data Collection Software (Applied Biosystems) arbeitete.

2.3 Auswertungsmethoden

2.3.1 Auswertung der Daten der MLST und eBURST

Die Nukleotidsequenzen wurden mit der Software Sequence Navigator Version 1.0.1 (Applied Biosystems) analysiert, die Zuordnung von Allelen und Sequenztypen erfolgte mit der Online-MLST-Datenbank (<http://saureus.mlst.net>). Provisorische Sequenztypen über ST1000, die bereits in der Literatur verwendet worden waren, wurden beibehalten [13]. Für diejenigen allelischen Profile, für die in der Online-MLST-Datenbank zum Zeitpunkt der Beendigung der Datensammlung (Datum 09.09.2005) kein Sequenztyp angegeben war, wurden weitere provisorische Sequenztypen über ST2000 bzw. ST3000 definiert (vgl. Tabelle 2-3).

Tabelle 2-3: Allelische Profile der provisorischen Sequenztypen.

ST	allelisches Profil ^a	ST	allelisches Profil	ST	allelisches Profil
1001	1-1-25-1-1-1-1	1014	3-51-1-1-4-4-3	1027	7-6-28-27-11-40-43
1003	3-1-1-8-4-1-1	1017	23-1-4-1-5-5-4	1029	3-1-31-1-29-5-3
1004	10-1-1-1-12-1-1	1018	7-6-30-5-8-5-6	1030	3-4-1-4-1-6-3
1005	2-2-24-2-2-2-2	1021	5-13-1-1-12-11-13	1031	3-2-1-2-13-3-11
1008	10-11-23-6-10-3-2	1024	6-5-6-6-7-17-19	2001	3-3-6-1-4-4-16
1009	10-14-23-6-14-3-2	1025	13-13-1-1-12-11-33	3001	2-3-1-1-4-47-3
1012	10-47-8-26-26-33-2	1026	3-1-1-1-1-5-38	3002	6-5-6-2-7-11-5

^a Reihenfolge: *arcC*, *aroE*, *glpF*, *gmk*, *pta*, *tpi*, *yqiL*. Allelische Profile mit Sequenztypen über ST1000 gehören zum Datensatz Nottingham, über ST2000 zum Datensatz Robinson. Für Sequenztypen über ST3000 wurde die MLST erstmals im Rahmen dieser Arbeit durchgeführt.

Die Häufigkeiten der im Datensatz vorhandenen Sequenztypen wurden mit der Software START Version 1.0.8 (K. Jolley, Oxford, UK, <http://outbreak.ceid.ox.ac.uk/software.htm> [20]) bestimmt.

Mit der Software BLAND Version 1.0 (persönliche Korrespondenz D. A. Robinson, Bath, UK) wurde die Beziehung von Nukleotiddiversität zu allelischer Diversität wie von Feil et al. beschrieben analysiert [6]. Dabei wird die Zahl der allelischen Unterschiede als Parameter für die Zeit seit der Divergenz zweier Sequenztypen gewertet. Es wird ein paarweiser Vergleich aller im Datensatz enthaltenen Sequenztypen durchgeführt, wobei jeder Sequenztyp nur einmal berücksichtigt wird, so dass die unterschiedliche Häufigkeit von Sequenztypen im Datensatz nicht in das Ergebnis eingeht. Als Ergebnis wird eine Kurve angegeben, in der die Zahl der unterschiedlichen Allele (Alleldivergenz, y-Achse) aufgetragen wird gegen den Quotienten der durchschnittlichen Zahl der Nukleotidunterschiede pro nicht-identisches Allel (Nukleotiddivergenz, x-Achse). Für die Tendenz der Kurve macht es dabei keinen Unterschied, nicht den Quotienten auf die y-Achse aufzutragen, sondern nur die durchschnittliche Zahl der Nukleotidunterschiede, wohl aber für die absolute Höhe der Werte. Im Rahmen dieser Arbeit wurde die zuvor verwendete Nomenklatur beibehalten.

Für Gruppierungen der MLST-Daten wurde das von Feil et al. beschriebene Programm eBURST v2 (B. Li, <http://eburst.mlst.net>) genutzt, das im Anschluss an den früheren Algorithmus BURST entwickelt worden war [5, 7, 8]. MLST-Datensätze werden durch eBURST in klonale Komplexe eingeteilt, das sind Gruppen von verwandten Sequenztypen, die sich nicht überschneiden. Die Gruppeneinteilung basiert auf der Zahl der Unterschiede zwischen den allelischen Profilen der einzelnen Sequenztypen. Dabei kann von dem Anwender eingestellt werden, welches Ausmaß an allelischen Unterschieden erlaubt wird, um zwei verschiedene Sequenztypen noch in eine gemeinsame Gruppe zu sortieren. Es besteht die Wahl zwischen einer voreingestellten strengen Gruppeneinteilung (jeder Sequenztyp hat mindestens einen SLV mit sechs gemeinsamen Allelen in der Gruppe), die in dieser Arbeit als eBURST 6/7 bezeichnet wird, und einer erweiterten Gruppeneinteilung, bei der lediglich fünf von sieben Allelen übereinstimmen müssen (SLV oder DLV), hier als eBURST 5/7 bezeichnet. Darüber hinaus erkennt das Programm die am wenigsten aufwendigen Abstammungsmuster von dem Gründungsvater

innerhalb jedes klonalen Komplexes. eBURST nutzt die Zahl von Verwandten, um Gründungsväter zu definieren. Derjenige Sequenztyp mit den meisten SLV wird als Gründungsvater definiert. Andere Sequenztypen können zu Subgründungsvätern definiert werden, basierend auf ihrer Zahl von SLV. So kann eBURST die Diversifikation von Bakterienklonen modellhaft beschreiben. Ebenfalls Teil des Programms ist die Funktion der Populationsübersicht („population snapshot“). Dabei wird jeder Sequenztyp durch einen Punkt repräsentiert, dessen Größe mit der Zahl der Isolate des Sequenztyps korreliert. SLV werden durch Linien miteinander verbunden.

2.3.2 Auswertung der Daten der *spa* Typisierung und BURP

Mit BURP, das auf dem EDSI-Modell [38] basiert und neu in der Software Ridom StaphType Version 1.3 (Ridom GmbH, Würzburg, <http://www.ridom.de> [16]) enthalten ist, steht ein Algorithmus zur Verfügung, der die Isolate eines Datensatzes automatisch nach ihren *spa* Typen in BURP-Cluster (*spa*-CC) klassifiziert. BURP bewertet den Grad von Verwandtschaft zwischen verschiedenen *spa* Typen basierend auf den Kosten für ihre Umwandlung ineinander. Damit kann eine Bildung von Clustern durchgeführt werden, die auch visualisiert werden kann.

Der Anwender hat die Möglichkeit, zwei BURP-Parameter zu variieren: Der Ausschluss-Parameter bietet die Möglichkeit, *spa* Typen, die nicht eine gewisse Mindestzahl an Repeats haben, von der weiteren Analyse auszuschließen („exclude *spa* types that are shorter than ... repeats“). Der Kosten-Parameter beschreibt die maximale Höhe der Kosten, die erlaubt ist, um zwei *spa* Typen als miteinander verwandt zu bezeichnen („cluster *spa* types together if cost is less or equal ...“).

Durch Kalkulation eines „founder-score“ kann BURP Gründungsväter und Subgründungsväter definieren. Der *spa* Typ mit dem höchsten „founder-score“ ist Gründungsvater eines Clusters (blaue Farbe in der graphischen

Repräsentation). Der spa-CC wird nach diesem *spa* Typ benannt. Falls mehrere *spa* Typen einen „founder-score“ in identischer Höhe haben, gehen sie gleichwertig in den Namen des spa-CC ein. Ähnlich wie in eBURST, aber nicht basierend auf der Anzahl der Single Locus Variants, sondern ebenfalls auf dem „founder-score“, werden als Subgründungsväter diejenigen *spa* Typen definiert, die den zweithöchsten Score haben (gelbe Farbe).

2.3.3 Statistische Auswertungen

Typisierbarkeit bezeichnet die Fähigkeit einer Methode, für jedes untersuchte Isolat ein eindeutiges Ergebnis zu erbringen. Als nicht-typisierbar werden diejenigen Isolate bezeichnet, für die kein oder lediglich ein mehrdeutiges Resultat zu erzielen ist. Reproduzierbarkeit ist die Fähigkeit eines Tests, bei wiederholter Durchführung des Tests eine identische Typisierung zu erbringen. Der Begriff der Stabilität bezeichnet die Fähigkeit einer Typisierungsmethode, trotz phänotypischer oder genomischer Variation die klonale Verwandtschaft von Stämmen zu erkennen, die *in vitro* oder *in vivo* von einem gemeinsamen Vorgänger abstammen. Die Quantifizierung von Typisierbarkeit, Reproduzierbarkeit und Stabilität ist relativ einfach durch Angabe von Prozentwerten, die idealerweise möglichst nahe bei 100% liegen sollten [44].

Einige Typisierungsmethoden neigen dazu, Organismen in wenige große Gruppen einzuteilen, während andere dieselben Organismen in viele kleine Cluster gruppieren. Der Begriff der Diskriminationsfähigkeit bezieht sich auf die Fähigkeit, eindeutig zwischen zwei nicht-verwandten Isolaten zu unterscheiden. Für die Quantifizierung der Diskriminationsfähigkeit hat sich der Diskriminatorische Index (DI) als geeignet erwiesen. In der englischsprachigen Literatur wird der DI als „Index of diversity“ [13, 41] oder auch als „Simpson’s index of diversity“ bezeichnet [18, 23, 33]. Er ist definiert als die durchschnittliche Wahrscheinlichkeit, dass eine Typisierungsmethode zwei nicht-verwandte Isolate mit unterschiedlichen Typen bezeichnet. Der DI ist abhängig von der Zahl der Isolate, von der Zahl der Typen und von der Häufigkeits-

verteilung der Isolate in den Typen. Konfidenzintervalle (KI) des DI wurden errechnet wie von Grundmann et al. beschrieben [14].

Die Konkordanz als Maßzahl für die Übereinstimmung zweier Gruppierungsmethoden wurde durch Paarvergleiche errechnet wie von Robinson et al. beschrieben [35]. Ein Paarvergleich der Klassifizierung von n Isolaten in spa-CC durch BURP, in CC durch eBURST oder in jegliche Art von Gruppe durch andere mögliche Methoden bedeutet, dass alle möglichen Paarungen von Isolaten betrachtet werden, das sind $[\frac{1}{2} (n^2-n)]$ Paarvergleiche (vgl. Tabelle 2-4).

Tabelle 2-4: Schema für die Berechnung der Konkordanz im Paarvergleich.

		Methode 1	
		Match	Mismatch
Methode 2	Match	a	b
	Mismatch	c	d

Für jedes mögliche Paar von zwei Isolaten wird bestimmt, ob ihre Einklassifizierung durch die jeweilige Gruppierungsmethode übereinstimmt (Match) oder nicht übereinstimmt (Mismatch). Dann wird betrachtet, ob eine Match-Match-, eine Match-Mismatch- oder eine Mismatch-Mismatch-Kombination vorliegt. Die Konkordanz K zweier Gruppierungsmethoden ergibt sich in diesem Paarvergleich aus der Summe der Match-Match-Kombinationen und der Mismatch-Mismatch-Kombinationen, dividiert durch die Zahl aller möglichen Vergleiche. $K = (2(a+d)) / (n^2-n)$. Sie ist umso höher, je mehr Match-Match- oder Mismatch-Mismatch-Kombinationen auftreten und umso niedriger, je mehr Match-Mismatch-Kombinationen auftreten.

In der Ridom StaphType Software Version 1.3 ist eine Berechnungsfunktion sowohl für die Konkordanz als auch für den diskriminatorischen Index mit Konfidenzintervall enthalten.

2.4 Datensätze

Der vorliegenden Arbeit liegen vier Datensätze zugrunde.

2.4.1 Datensatz Münster

Die diesem Datensatz zugrunde liegende Idee war es, für möglichst jeden an einem Stichtag bekannten *spa* Typ (Datum 29.07.2004, bis t430) mindestens ein Isolat durch die MLST zu typisieren. Durch Kooperation mit verschiedenen wissenschaftlichen Einrichtungen gelang es, für einen Großteil der bekannten *spa* Typen ein Isolat für die MLST-Sequenzierung zu finden oder die entsprechenden Daten zu übermitteln (vgl. Tabelle 2-5), weitere Isolate mit *spa* Typ und MLST Sequenztyp konnten aus der Literatur ergänzt werden [33]. Keine Kombination eines ST mit einem *spa* Typ sollte aber mehr als einmal vorkommen. Auf diese Weise ergaben sich für den Datensatz Münster (vgl. Anhang A) 447 Isolate, von denen für 207 Isolate der Sequenztyp bereits bekannt war. Für 240 Isolate wurden die MLST-Sequenzen, allelischen Profile und Sequenztypen im Rahmen der Laborarbeit erstmals bestimmt. Der Datensatz ist nach seinem Entstehungsort Münster benannt.

Tabelle 2-5: Herkunft der Isolate, Datensatz Münster.

Kooperation	Zahl der Isolate	ST bereits bekannt	ST neu bestimmt
Becker, K ^a	36	0	36
Enright, MC ^b	132	132	0
Grundmann, H ^c	48	48	0
Harmsen, D ^d	59	0	59
Kahl, BC ^a	39	0	39
Kearns, AM ^e	12	12	0
Peters, G ^a	1	0	1
Petersson, AC ^f	22	1	21
Sabat, A ^g	5	5	0
Sinha, B ^a	12	0	12
Vogel, U ^h	17	0	17
Westh, H ⁱ	50	1	49
Witte, W ^k	14	8	6
gesamt	447	207	240

Wissenschaftler, die Isolate oder Daten zur Verfügung stellten. ^a Institut für Medizinische Mikrobiologie, Münster. ^b Department of Biology and Biochemistry, Bath, UK. ^c University Hospital Nottingham, UK. ^d Institut für Hygiene, Münster. ^e Health Protection Agency, London, UK. ^f Lund University Hospital, Schweden. ^g National Institute of Public Health, Warschau, Polen. ^h Institut für Hygiene und Mikrobiologie, Würzburg. ⁱ Hvidovre Hospital, Dänemark. ^k Robert Koch Institut, Werningerode.

2.4.2 Datensatz Koreen

Die Zusammensetzung dieses Datensatzes (vgl. Anhang B) wurde in der Literatur eingehend beschrieben [9]. Er enthält den Referenz-Stamm (COL) sowie 36 *S. aureus* Isolate, die von 2077 mittels Multilocus Enzym Elektrophorese (MLEE) typisierten Isolaten aus weltweit zehn Ländern auf vier verschiedenen Kontinenten über einen Zeitraum von vier Jahrzehnten ausgewählt worden waren, darunter auch zwölf MRSA. Diese auch mit einem Gesamtgenom-Microarray untersuchten Isolate waren durch Koreen et al. ergänzend mittels *spa* Typisierung, *coa* Typisierung und Pulsfeldgelelektrophorese (PFGE) charakterisiert worden [23]. Nach diesem Autor wurde der Datensatz für die vorliegende Arbeit benannt. Die Einteilung in Microarray-Gruppen wurde mittels manueller Auslesung des Microarray-Dendrogramms bei drei Pearson-Korrelationskoeffizienten vorgenommen. Die Repeatprofile wurden mittels eines Internet-basierten Servers (<http://spaserver.ridom.de>) in die Ridom-Nomenklatur übersetzt [16]. Das Dendrogramm (vgl. Anhang C) wurde mit der Software MEGA (<http://www.megasoftware.net> [25]) erstellt.

2.4.3 Datensatz Robinson

Robinson et al. hatten 147 *S. aureus* Isolate aus achtzehn Ländern mit einer hochauflösenden Multilocus-Methode untersucht und für 107 dieser Isolate fünf Stammbäume (CC5, CC8, CC22, CC30 und CC45) angelegt, die diejenigen Linien beschreiben, die die fünf weltweit dominierenden MRSA-Linien enthalten oder solche MSSA, aus denen sich phylogenetisch später MRSA entwickelten [33]. Aus diesen 107 Isolaten wurde der Datensatz Robinson (vgl. Anhang D) erstellt. Die Nomenklatur folgte dem Vorschlag Enright's nach Sequenztyp und SCC*mec*-Typ [5], in Klammern folgte das Landeskürzel und gegebenenfalls eine Nummerierung. Dabei ergab eine Kontrolle der allelischen Profile (<http://www.mlst.net>, Datum 12.09.2005) Änderungen der im Stammbaum angegebenen Sequenztypen ST305, ST308 bzw. ST313 zu den Sequenztypen ST157, ST554 bzw. ST158. Das allelische Profil des angegebenen ST257

existierte in der Datenbank nicht, so dass hierfür der provisorische ST2001 definiert wurde. Die im Stammbaum angegebenen Repeatprofile wurden online (<http://spaserver.ridom.de>) in *spa* Typen übersetzt. Die Information über das Repeatprofil des Isolates ST45-MRSA-IV (Bel) wurde in persönlicher Korrespondenz mit dem Autor eingeholt. Wegen der von Robinson et al. in einer späteren Publikation beschriebenen Rekombination [34] wurde nicht nur der Gesamtdatensatz untersucht, sondern auch eine „Auswahl“ von 90 Isolaten aus diesem Datensatz ohne diejenigen siebzehn Isolate, die in der Originalpublikation den ST239 hatten oder von ihm abgeleitet worden waren [33]. Sechs Isolate aus dem Datensatz Robinson wurden auch in den Datensatz Münster übernommen.

2.4.4 Datensatz Nottingham

Der Datensatz Nottingham (vgl. Anhang E) enthält 111 Isolate aus einer britischen Kolonisationsstudie, die vor wenigen Jahren im Großraum Nottingham durchgeführt worden war [15]. Die MLST-Daten von 117 Isolaten dieser Stammsammlung waren bereits veröffentlicht [13]. Die *spa* Typisierung war im Institut für Hygiene, Münster, durchgeführt worden, wobei drei Isolate (Nummer 268, 337 und 398) nicht typisierbar gewesen waren. Ein Abgleich der von dem Autor in persönlicher Korrespondenz zur Verfügung gestellten Labor-daten führte wegen Unstimmigkeiten zum Ausschluss dreier weiterer Isolate (Nummer 116, 254 und 399). Von den 111 Isolaten sind 22 auch im Datensatz Münster enthalten.

3. Ergebnisse

Zur Evaluation des BURP-Algorithmus für die phylogenetische Analyse von *S. aureus* wurden vier Datensätze herangezogen. Es wurde untersucht, wie sich die BURP-Parameter auf die Gruppierungsergebnisse auswirken und inwieweit die gebildeten Cluster mit den Gruppierungen durch etablierte Methoden übereinstimmen.

3.1 Datensatz Münster

Der Datensatz Münster enthielt 447 Isolate, für die die *spa* Typen bereits bekannt waren. Für 207 Isolate war der Sequenztyp bereits bekannt, für 240 Isolate wurden die MLST-Daten hingegen neu bestimmt. Fünf weitere getestete Isolate waren auch unter Variation der PCR-Bedingungen mit der MLST nicht typisierbar, so dass sie nicht in den Datensatz aufgenommen wurden (Daten nicht gezeigt).

Tabelle 3-1: Häufigkeit von Sequenztypen, Datensatz Münster.

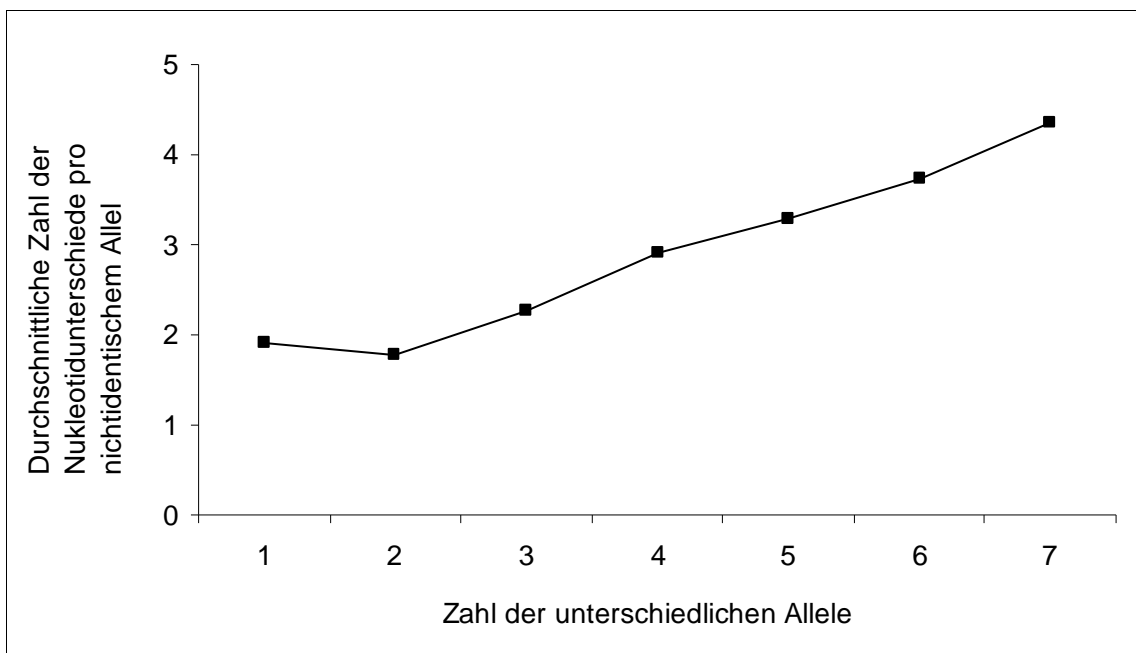
Häufigkeit des ST ^a	Anzahl der ST ^b	Sequenztyp
48 (10,74)	1 (10,74)	ST45
25 (5,59)	1 (5,59)	ST30
22 (4,92)	2 (9,84)	ST5, ST15
20 (4,47)	2 (8,95)	ST22, ST25
14 (3,13)	1 (3,13)	ST8
12 (2,68)	2 (5,37)	ST1, ST228
11 (2,46)	1 (2,46)	ST121
10 (2,24)	1 (2,24)	ST34
9 (2,01)	2 (4,03)	ST39, ST225
8 (1,79)	2 (3,58)	ST9, ST254
7 (1,57)	2 (3,13)	ST36, ST247
6 (1,34)	3 (4,03)	ST12, ST239, ST241
5 (1,12)	1 (1,12)	ST80
4 (0,89)	2 (1,79)	ST72, ST97
3 (0,67)	5 (3,36)	ST7, ST20, ST26, ST27, ST59
2 (0,45)	17 (7,61)	ST3, ST10, ST46, ST88, ST101, ST111, ST182, ST188, ST217, ST250, ST395, ST398, ST509, ST597, ST617, ST 627, ST1017
1 (0,22)	103 (23,04)	übrige 103 ST

^a Zahl der Isolate mit dem angegebenen Sequenztyp (in Klammern: Anteil an den Isolaten des Datensatzes in Prozent).

^b Zahl der Sequenztypen mit der angegebenen Häufigkeit (in Klammern: gemeinsamer Anteil der Isolate dieser Sequenztypen an den Isolaten des Datensatzes in Prozent).

Mit der MLST wurden 148 verschiedene Sequenztypen identifiziert. Mehrfache Aufnahme eines bestimmten Sequenztyps war nur bei Vorkommen verschiedener *spa* Typen möglich. Die Häufigkeitsverteilung der Sequenztypen ist aus Tabelle 3-1 ersichtlich. Der DI der MLST betrug 97,1% (95% KI: 96,5%-97,7%). Der mit BLAND errechnete Graph von allelischer Diversität und Nukleotiddiversität wird in Diagramm 3-1 gezeigt.

Diagramm 3-1: BLAND, Alleldivergenz und Nukleotiddivergenz.



Die 148 Sequenztypen des Datensatzes Münster wurde paarweise verglichen in Hinblick auf die durchschnittliche Zahl der unterschiedlichen Nukleotidpositionen pro Allel (y-Achse, dabei wurde nur die Zahl der nichtidentischen Allele zugrunde gelegt). Diese Werte wurden aufgetragen gegen die Zahl der unterschiedlichen Allele (x-Achse). Es besteht eine positive Beziehung zwischen den beiden Zahlen innerhalb des Datensatzes.

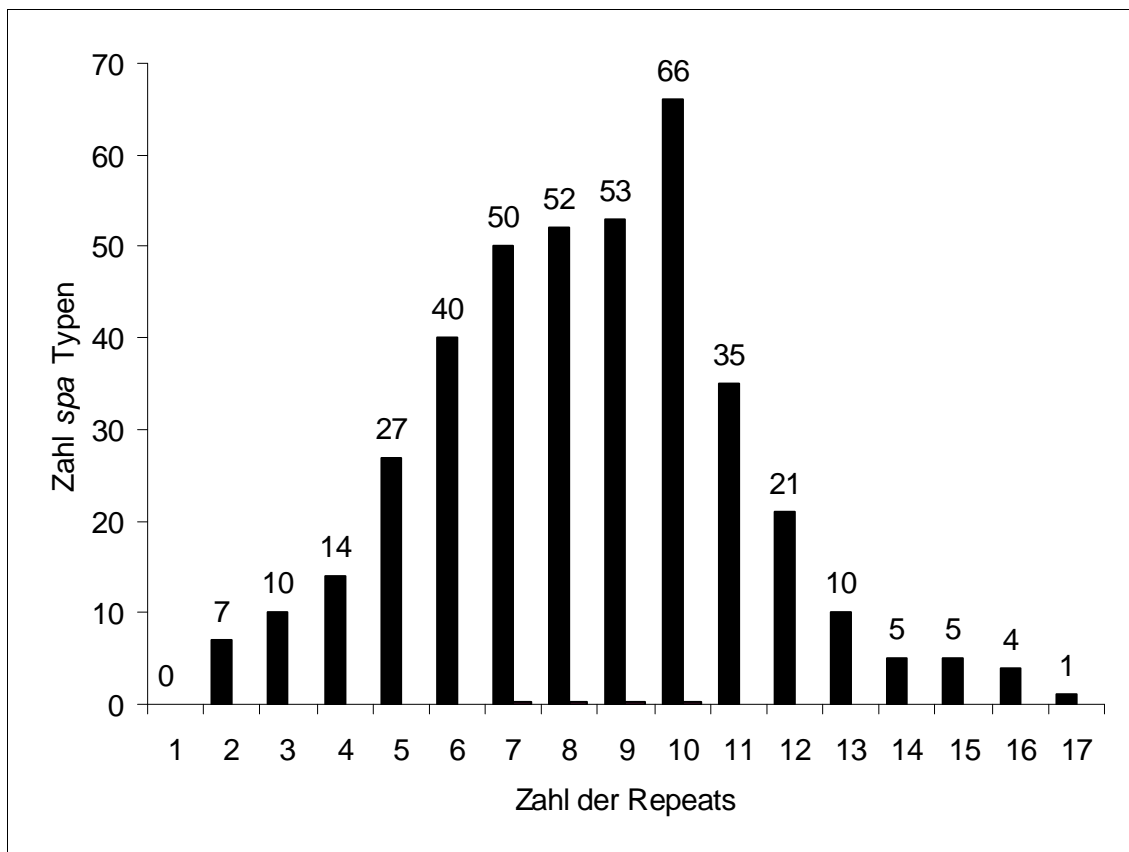
Mit der *spa* Typisierung wurden 400 *spa* Typen identifiziert. Die Häufigkeitsverteilung der *spa* Typen findet sich in Tabelle 3-2.

Die Länge der *spa* Typen reichte von zwei Repeats (sieben *spa* Typen) bis 17 Repeats (t294) (vgl. Diagramm 3-2). Der DI der *spa* Typisierung betrug 99,9% (95% KI: 99,9%-100%).

Tabelle 3-2: Häufigkeit von *spa* Typen, Datensatz Münster.

Häufigkeit des <i>spa</i> Typs ^a	Anzahl der <i>spa</i> Typen ^b	<i>spa</i> Typen ^c
6 (1,34)	1 (1,34)	t008 (ST8, ST247, ST250, ST254, ST1014, ST2001)
5 (1,12)	1 (1,12)	t005 (ST22, ST23, ST60, ST281, ST217)
4 (0,89)	1 (0,89)	t037 (ST239, ST241, ST240, ST158)
3 (0,67)	7 (4,70)	t001, t015, t018, t021, t030, t045, t084
2 (0,45)	21 (9,40)	t002, t003, t026, t041, t050, t052, t065, t078, t094, t100, t160, t166, t189, t192, t233, t236, t287, t298, t346, t362, t364
1 (0,22)	369 (82,55)	übrige 369 <i>spa</i> Typen

^a Zahl der Isolate mit dem angegebenen *spa* Typ (in Klammern: Anteil an den Isolaten des Datensatzes in Prozent). ^b Zahl der *spa* Typen mit der angegebenen Häufigkeit (in Klammern: gemeinsamer Anteil der Isolate dieser *spa* Typen an den Isolaten des Datensatzes in Prozent). ^c Für die drei häufigsten *spa* Typen in Klammern die zugehörigen Sequenztypen.

Diagramm 3-2: Zahl der Repeats von *spa* Typen, Datensatz Münster.

Die Zahl der Repeats der *spa* Typen im Datensatz Münster lag zwischen zwei und 17, *spa* Typen mit zehn Repeats machten die größte Gruppe aus (66 *spa* Typen).

Der Algorithmus eBURST teilte die 447 Isolate je nach der Einstellung der Parameter in unterschiedliche Gruppen oder Einzeltypen ein. Mit der voreingestellten Gruppendifinition (eBURST 6/7) wurden 23 Gruppen und 38 Einzeltypen definiert. Für 14 der Gruppen wurde ein Gründungsvater angegeben. Die Einteilung nach der erweiterten Gruppendifinition (eBURST 5/7)

ergab 22 Gruppen und 15 Einzeltypen, für 13 der Gruppen wurde ein Gründungsvater angegeben. Die Einteilung der einzelnen Isolate zu Gruppen oder Einzeltypen ist im Anhang angegeben (vgl. Anhang A). Die Populationsübersicht findet sich in Abbildung 3-1.

Abhängig von den gewählten Parametern gruppierte der BURP-Algorithmus unter Ausschluss einer bestimmten Zahl von *spa* Typen mit wenigen Repeats die Isolate in Gruppen und Einzeltypen. Ebenso wie die BURP-Gruppeneinteilung richtete sich auch die Konkordanz zwischen BURP und eBURST nach den gewählten Parametern (vgl. Tabelle 3-3). Die höchste Konkordanz mit 95,07% bzw. 95,27% zu den Einteilungen nach eBURST 6/7 bzw. eBURST 5/7 wurde bei einem Ausschluss-Parameter von fünf und einem Kosten-Parameter von vier gefunden. Die entsprechenden Informationen über die Gruppenzugehörigkeiten der einzelnen Isolate bei dieser Einteilung in 24 Gruppen und 40 Einzeltypen bei 31 ausgeschlossenen *spa* Typen findet sich im Anhang (vgl. Anhang A), eine graphische Repräsentation des *spa*-CC 008 wird in Abbildung 3-2 gezeigt.

Ein Beispiel hoher Konkordanz sind die Isolate des klonalen Komplexes CC12. Sämtliche Isolate dieses CC wurden von BURP dem *spa*-CC 160 zugeordnet, und es gab im gesamten Datensatz kein anderes Isolat, das diesem *spa*-CC 160 zugeteilt wurde. Dabei war es für die Zusammensetzung des CC12 gleichgültig, ob er nach eBURST 6/7 oder nach eBURST 5/7 definiert worden war. Lediglich ein Isolat (t336) dieses CC wurde von BURP wegen einer zu geringen Zahl an Repeats ausgeschlossen.

Tabelle 3-3: BURP und Konkordanz zu eBURST, Datensatz Münster.

BURP		Konkordanz zu eBURST (in %)	
Parameter	Statistik	eBURST 6/7 (23 gr, 38 sg)	eBURST CC 5/7 (22 gr, 15 sg)
3 ^a , 0 ^b	7 ^c , 0 ^d , 393 ^e	92,64	91,32
3, 1	7, 48, 210	92,99	91,71
3, 2	7, 54, 144	93,32	92,10
3, 3	7, 33, 74	94,12	93,59
3, 4	7, 21, 39	73,90	74,09
3, 5	7, 13, 20	45,69	46,43
3, 6	7, 5, 10	19,02	20,00
3, 7	7, 2, 3	10,12	11,45
3, 8	7, 1, 3	8,76	10,10
3, 9	7, 1, 3	8,76	10,10
4, 0	17, 0, 383	92,63	91,27
4, 1	17, 46, 204	92,98	91,68
4, 2	17, 52, 140	93,33	92,09
4, 3	17, 32, 74	94,16	93,66
4, 4	17, 23, 39	88,69	88,88
4, 5	17, 15, 20	58,07	58,58
4, 6	17, 5, 11	19,77	20,76
4, 7	17, 2, 3	10,22	11,58
4, 8	17, 2, 3	10,22	11,58
4, 9	17, 1, 3	8,83	10,19
5, 0	31, 0, 369	92,64	91,36
5, 1	31, 45, 193	93,03	91,80
5, 2	31, 52, 130	93,39	92,23
5, 3	31, 31, 69	94,23	93,85
5, 4	31, 24, 40	95,07	95,27
5, 5	31, 16, 21	79,83	80,36
5, 6	31, 6, 13	34,27	35,15
5, 7	31, 3, 5	24,50	25,60
5, 8	31, 3, 5	24,50	25,60
5, 9	31, 1, 4	9,33	10,59
6, 0	58, 0, 342	92,52	91,15
6, 1	58, 41, 175	92,96	91,65
6, 2	58, 47, 115	93,38	92,14
6, 3	58, 28, 59	94,14	93,73
6, 4	58, 22, 35	94,99	95,25
6, 5	58, 17, 19	90,95	91,53
6, 6	58, 9, 12	54,37	55,32
6, 7	58, 7, 5	38,36	39,33
6, 8	58, 6, 5	36,65	37,66
6, 9	58, 2, 4	23,40	24,60

Konkordanzen, Datensatz Münster ($n = 447$; 148 ST; 400 *spa* Typen). ^a minimal nötige Zahl von Repeats, um einen *spa* Typ in die Analyse einzuschließen. ^b maximal mögliche Kosten, von *spa* Typen innerhalb einer Gruppe. ^c Zahl der ausgeschlossenen *spa* Typen. ^d Zahl der gebildeten Gruppen. ^e Zahl der Einzeltypen. **Konkordanzmaxima.**

3.2 Datensatz Koreen

Für den Datensatz Koreen war bereits in der primären Veröffentlichung nur für 36 von 37 Isolaten ein *spa* Typ angegeben, weil das Isolat MSA2345 für die Studie nicht verfügbar gewesen war [23]. Für einen Vergleich mit den bereits von Koreen et al. untersuchten Typisierungsmethoden wurden in der vorliegenden Arbeit verschiedene Parametereinstellungen getestet, die jeweils zu unterschiedlichen Gruppierungen und Konkordanzen führten. Auch die Konkordanzmaxima zu den verschiedenen Typisierungsmethoden wurden dabei bei unterschiedlichen Parametereinstellungen gefunden (Daten nicht gezeigt). Am Ende wurde sowohl für den Ausschluss-Parameter als auch für den Kosten-Parameter die Höhe fünf gewählt, weil sich bei dieser Einstellung die höchste Konkordanz (87,2%) zu der 0.31 Microarray-Gruppierung ergab, die eine phylogenetisch weit in die Vergangenheit reichende und gleichzeitig objektive Gruppierung ist. Für diesen Datensatz lagen keine MLST-Daten vor.

Tabelle 3-4: Verschiedene Typisierungsmethoden, Datensatz Koreen.

Typisierungsmethode		Zahl der Isolate	Zahl der Gruppen	häufigste Gruppe in %	Diskriminatorischer Index (95% KI ^a)
Micro array	0.81	37	36	5,4	99,8 (99,4-100,3)
	0.5	37	16	29,7	89,2 (81,9-96,4)
	0.31	37	8	40,5	78,7 (69,0-88,3)
MLEE	MLEE Typ	36	15	41,7	81,6 (69,6-93,5)
	MLEE Lin	36	8	41,7	76,3 (66,2-86,5)
PFGE		36	28	18,9	96,3 (92,0-100,7)
coa	coa Typ	36	19	22,2	92,7 (87,7-97,7)
	coa Lin I	36	11	44,4	78,4 (66,0-90,8)
	coa Lin II	36	10	44,4	77,5 (65,5-89,4)
<i>spa</i>	<i>spa</i> Typ	36	29	16,7	97,3 (93,9-100,7)
	BURP ^b	34	13	38,2	82,9 (72,5-93,2)

Analyse des Datensatzes ($n = 37$; 29 *spa* Typen) aus der Publikation von Koreen et al. [23] mit verschiedenen Typisierungsmethoden. ^a Konfidenzintervall. ^b BURP-Parameter: Ausschluss von *spa* Typen kürzer als fünf Repeats, Kosten kleiner oder gleich fünf, dadurch Ausschluss von zwei *spa* Typen sowie Bildung von fünf Gruppen und neun Einzeltypen.

Die Isolate RF122 (zwei *spa* Repeats) und MSA2885 (drei *spa* Repeats) wurden wegen ihrer geringen Zahl an Repeats ausgeschlossen. Die detaillierte Einteilung in Microarray-Gruppen sowie die Übersetzung der *spa* Profile in *spa* Typen und nachfolgend die Einteilung in *spa*-CC ist im Anhang B enthalten. Eine Analyse der Einteilung mit Angabe des DI ist in Tabelle 3-4 wiedergegeben. Die Konkordanz der BURP-Einteilung zu den bereits von Koreen et al. untersuchten Klassifizierungsmethoden findet sich in Tabelle 3-5. Ein Dendrogramm, das die BURP-Einteilung mit derjenigen durch den Microarray bei einem Korrelationskoeffizienten von 0.31 vergleicht, ist im Anhang aufgeführt (vgl. Anhang C).

Tabelle 3-5: Konkordanzen, Datensatz Koreen.

Typisierungsmethode		% Konkordanz zwischen										
		Microarray			MLEE		PFGE	coa			spa	
		0.81	0.5	0.31	Typ	Lin		Typ	Lin I	Lin II	Typ	BURP
Microarray	0.81	-										
	0.5	89,3	-									
	0.31	78,8	89,5	-								
MLEE	Typ	81,7	91,1	94,1	-							
	Lin	76,5	87,1	91,7	94,8	-						
PFGE		96,3	93,8	83,3	86,4	81,2	-					
coa	Typ	92,7	91,7	84,4	86,6	82,0	94,1	-				
	Lin I	78,4	88,3	90,8	93,8	89,6	82,1	85,7	-			
	Lin II	77,5	87,3	91,7	92,8	88,6	81,1	84,8	99,0	-		
spa	Typ	97,3	92,5	82,4	85,5	80,0	98,4	95,4	81,1	80,2	-	
	BURP ^a	83,6	86,3	87,2	90,2	84,3	87,3	86,1	93,8	92,7	86,6	-

Modifiziert nach Koreen et al. [23]. ^a BURP-Parameter: Ausschluss von *spa* Typen kürzer als fünf Repeats, Kosten kleiner oder gleich fünf, dadurch Ausschluss von zwei *spa* Typen und Bildung von fünf Gruppen und neun Einzeltypen. Die Konkordanz zwischen dieser BURP-Gruppierung und der von Koreen et al. manuell durchgeführten *spa* Gruppierung beträgt 96,8%.

3.3 Datensatz Robinson und Datensatz Nottingham

Für die Datensätze Robinson (107 Isolate) und Nottingham (111 Isolate) lagen die einzelnen Daten der *spa* Typisierung und der Typisierung mittels MLST bereits vor.

Der DI im Datensatz Robinson betrug für die MLST 90,8% (95% KI: 88,7%-93,0%), für die *spa* Typisierung 93,9% (92,4%-95,3%). Die *spa* Typen mit der geringsten Zahl an Repeats waren im Datensatz Robinson die *spa* Typen t030 und t062 mit jeweils sechs Repeats. Daher schloss BURP bei einem Ausschluss-Parameter von sechs keinen *spa* Typ aus (Daten nicht gezeigt). Bei einem Kosten-Parameter von elf und höher, der zu einer Bildung von vier Gruppen führte, wurde das Konkordanzmaximum von 86,58% zu eBURST 6/7 gefunden. Weil für den Datensatz eine Einteilung in fünf Gruppen angestrebt wurde, wie es in der ursprünglichen Publikation der Fall gewesen war, wurde die detaillierte Einteilung (vgl. Anhang D) bei einem Kosten-Parameter von sieben erstellt, wobei die Konkordanz mit 86,16% leicht niedriger lag. Für die Auswahl ohne ST239 und die von ihm abgeleiteten Sequenztypen lag das Konkordanzmaximum genau bei diesem Kosten-Parameter (vgl. Tabelle 3-6).

In der Stammsammlung, die dem Datensatz Nottingham zugrunde liegt, betrug die Typisierbarkeit der *spa* Methode 97,4% (114 von 117 Isolaten). Der DI betrug in diesem Datensatz für die MLST 95,5% (95% KI: 94,0%-96,9%), für die *spa* Typisierung 98,0% (97,0%-98,9%). Weil ein Ausschluss-Parameter von fünf in den drei anderen Datensätzen die höchsten Konkordanzwerte ergeben hatte, wurde er für diesen Datensatz übernommen. Dabei fanden sich die Konkordanzmaxima zu den Einteilungen nach eBURST 5/7 und eBURST 6/7 bei unterschiedlichen Kosten-Parametern (vgl. Tabelle 3-6). Die detaillierte Einteilung bei einem Ausschluss-Parameter von fünf und einem Kosten-Parameter von sieben findet sich im Anhang (vgl. Anhang E).

Tabelle 3-6: BURP und Konkordanz zu eBURST, weitere Datensätze.

BURP Parameter	Robinson ^a						Nottingham ^b		
	Gesamtdatensatz ^c			Auswahl ^d			%Konkordanz zu eBURST		
	BURP Statistik	%Konkordanz zu eBURST		BURP Statistik	%Konkordanz zu eBURST		BURP Statistik	%Konkordanz zu eBURST	
		6/7 (3 gr, ^e 7 sg)	5/7 (3 gr, 3 sg)		6/7 (3 gr, 6 sg)	5/7 (3 gr, 3 sg)		6/7 (11 gr, 15 sg)	5/7 (10 gr, 11 sg)
5 ^g , 0 ^h	0 ⁱ , 0 ^k , 27 ^m	81,75	80,30	0, 0, 25	86,24	85,07	6, 0, 61	94,44	90,91
5, 1	0, 5, 16	81,64	80,23	0, 4, 16	86,99	85,87	6, 7, 41	95,81	92,29
5, 2	0, 6, 12	82,63	81,22	0, 5, 12	88,39	87,27	6, 9, 32	96,34	92,93
5, 3	0, 6, 7	84,20	85,01	0, 5, 7	90,61	92,63	6, 9, 26	97,01	93,64
5, 4	0, 6, 4	82,37	83,60	0, 6, 4	91,74	93,76	6, 9, 20	95,62	96,06
5, 5	0, 5, 1	85,31	86,69	0, 5, 1	95,91	98,13	6, 8, 18	95,44	96,17
5, 6	0, 5, 1	85,31	86,69	0, 5, 1	95,91	98,13	6, 10, 13	95,48	96,50
5, 7	0, 5, 0	86,16	87,53	0, 5, 0	97,10	99,33	6, 9, 9	94,38	96,36
5, 8	0, 5, 0	86,16	87,53	0, 5, 0	97,10	99,33	6, 8, 9	93,28	95,26
5, 9	0, 5, 0	86,16	87,53	0, 5, 0	97,10	99,33	6, 5, 7	66,89	68,86
5, 10	0, 5, 0	86,16	87,53	0, 5, 0	97,10	99,33	n.b.	n.b.	n.b.
5, 11	0, 4, 0	86,58	88,91	0, 5, 0	97,10	99,33	n.b.	n.b.	n.b.
5, 12	0, 4, 0	86,58	88,91	0, 5, 0	97,10	99,33	n.b.	n.b.	n.b.
5, 13	0, 4, 0	86,58	88,91	0, 5, 0	97,10	99,33	n.b.	n.b.	n.b.
5, 14	0, 2, 0	47,84	50,17	0, 2, 0	47,59	49,81	n.b.	n.b.	n.b.

^a Datensatz ($n = 107$) von Robinson et al. [33]. ^b Stammsammlung ($n = 111$; 44 ST; 67 *spa* Typen) von Grundmann et al. [13]. ^c Gesamtdatensatz ($n = 107$; 23 ST; 27 *spa* Typen). ^d Auswahl ($n = 90$; 17 ST; 25 *spa* Typen) ohne Isolate mit ST239 und die von ihm abgeleiteten Sequenztypen (ST157, ST158, ST240, ST241, ST246). ^e Zahl der gebildeten Gruppen. ^f Zahl der Einzeltypen. ^g minimal nötige Zahl von Repeats, um einen *spa* Typ in die Analyse einzuschließen. ^h maximal mögliche Kosten zwischen zwei *spa* Typen innerhalb einer Gruppe. ⁱ Zahl der ausgeschlossenen *spa* Typen. ^k Zahl der gebildeten Gruppen. ^m Zahl der Einzeltypen. Konkordanzmaxima.

Eine spätere umfangreichere Analyse der Parameter im Datensatz Nottingham ergab, dass bei einem Ausschluss-Parameter von sechs lediglich drei zusätzliche *spa* Typen ausgeschlossen wurden (Daten nicht gezeigt). Das höchste Konkordanzmaximum zu eBURST 5/7 betrug dann bei einem Kosten-Parameter von drei, der zu einer Bildung von neun Gruppen und 23 Einzeltypen führte, 97,09%. Zu eBURST 6/7 lag das Maximum bei einem Kosten-Parameter von sieben, der acht Gruppen und zehn Einzeltypen zur Folge hatte, bei 96,78%.

4. Diskussion

Für die Typisierung von *S. aureus* Isolaten steht eine Vielzahl von Methoden zur Verfügung. Gegenüber der bandenbasierten PFGE haben dabei sequenzbasierte Techniken, wie die MLST oder die *spa* Typisierung, mehrere Vorteile. Dazu zählt neben einer höheren Reproduzierbarkeit und einem geringeren Anwendungs- und Interpretationsaufwand nicht zuletzt der einfachere Datenaustausch. Darüber hinaus ist die PFGE zwar als Goldstandard für Ausbruchsuntersuchungen von *S. aureus* anerkannt, aber ihre Anwendbarkeit für phylogenetische Analysen steht in Frage [45]. Im Gegensatz zu der MLST, die für evolutionäre Untersuchungen von *S. aureus* verbreitet ist [4, 5], erwies sich die *spa* Typisierung als nützliche Routinemethode für Ausbruchsuntersuchungen. Darüber hinaus wurde für die *spa* Typisierung in früheren Studien eine hohe Diskriminationsfähigkeit beschrieben [45].

Allerdings existierte bislang kein Algorithmus, um für epidemiologische Studien Verwandtschaftsbeziehungen zwischen unterschiedlichen *spa* Typen festzustellen. Die Einführung von BURP in die Ridom StaphType Software ermöglicht die Gruppierung von *spa* Typen auf der Basis des EDSI-Modells [16, 38]. Ein Ziel dieser Arbeit war es, Gruppierungsergebnisse der *spa* Typisierung unter Anwendung von BURP mit denjenigen zu vergleichen, die durch Methoden erhalten werden, die für langzeitepidemiologische Studien anerkannt sind, insbesondere die MLST unter Anwendung von eBURST. Es konnte eine weitgehende Übereinstimmung der verschiedenen Gruppierungsmethoden gezeigt werden.

4.1 Datensatz Münster

4.1.1 eBURST im Datensatz Münster

Feil et al. veröffentlichten im Jahre 2004 eine Populationsübersicht auf der Grundlage der *S. aureus* MLST-Online-Datenbank, die zu dem damaligen Zeit-

punkt aus 1072 Isolaten und 191 Sequenztypen bestand [7]. Daher ist in der hier vorliegenden Populationsübersicht sowohl die Zahl als auch in vielen Fällen die Größe der Punkte geringer. Eine zusätzliche Analyse des Datensatzes Münster mit BURST ergab keine reproduzierbaren Ergebnisse. Die Größe des Datensatzes scheint die Rechenkapazität von BURST zu übersteigen. Dieses Problem war bereits früher beschrieben worden [7].

4.1.2 Einfluss des Ausschluss-Parameters auf die BURP-Einteilung

Unter Berücksichtigung des sog. „slipped-strand mispairing“, das durch das EDSI-Modell formal beschrieben wird, bietet die Zuordnung von *spa* Typen mit einer geringen Repeatzahl zu einem bestimmten *spa*-CC Schwierigkeiten auf. Daher war die Einführung eines Ausschluss-Parameters, durch den *spa* Typen mit einer geringen Repeatzahl von der Analyse ausgeschlossen werden, sinnvoll. Bei der Festlegung des Parameters ist auf der einen Seite der übertriebene Ausschluss von *spa* Typen zu vermeiden. Auf der anderen Seite besteht bei zu niedrig gewähltem Ausschluss-Parameter die Gefahr, leichtfertig die Definition von Verwandtschaftsbeziehungen zuzulassen, weil sich ein bestimmter *spa* Typ mit wenigen Repeats rechnerisch relativ leicht in einen *spa* Typ mit einer größeren Zahl von Repeats umwandeln lässt.

Im Datensatz Münster wird eine Auswirkung des Ausschluss-Parameters sichtbar. Bei einem höheren Ausschluss-Parameter werden mehr Isolate ausgeschlossen als bei einem niedrigeren Ausschluss-Parameter. Wird der Parameter zum Beispiel auf drei gesetzt, werden sieben *spa* Typen ausgeschlossen (diejenigen mit weniger als drei Repeats), wird er auf vier angehoben, werden insgesamt 17 *spa* Typen ausgeschlossen.

4.1.3 Einfluss des Kosten-Parameters auf die BURP-Einteilung

Auch die Variation des Kosten-Parameters wirkt sich auf die Größe der BURP-Cluster aus. Es können Verwandtschaftsketten von *spa* Typen gebildet werden.

Die *spa* Typen an den Endpunkten solcher Ketten können durchaus durch mehr Kosten voneinander entfernt sein als primär durch den Parameter erlaubt wird. In diesem Fall ist es bedeutsam, dass zwischen ihnen andere *spa* Typen liegen, die nach dem EDSI-Modell einander ähnlicher sind.

Diese Zusammenhänge lassen sich auch im Datensatz Münster betrachten. Wird die Grenze der erlaubten Kosten niedrig angesetzt, werden viele kleine Gruppen gebildet. Viele *spa* Typen gehen als Einzeltypen aus der Analyse hervor, weil das Programm für sie keine Verwandtschaft definiert. Bei einem Kosten-Parameter von null liegt die Exzessvariante dieses Zustands vor, bei der jeder *spa* Typ als Einzeltyp definiert wird. So kommt es, dass dann die Summe der Zahl der ausgeschlossenen *spa* Typen und der Zahl der Einzeltypen jeweils 400 ist. Mit steigender Höhe des Kosten-Parameters sinkt die Zahl der Einzeltypen, weil vermehrt *spa* Typen zu Gruppen zusammengefasst werden oder bereits bestehenden Gruppen zusortiert werden. Gleichzeitig kann die Zahl der Gruppen auch dadurch sinken, dass zuvor separate Gruppen zu einer gemeinsamen Gruppe zusammengefasst werden.

Der Kosten-Parameter zeigt Analogien zu der Einteilung nach eBURST, für die zwischen der voreingestellten (eBURST 6/7) und der erweiterten Gruppendefinition (eBURST 5/7) gewählt werden kann [7]. Dabei unterscheidet sich im Datensatz Münster die Zahl der Gruppen zwischen der voreingestellten Gruppeneffinition eBURST 6/7 (23 Gruppen) und der erweiterten Gruppeneffinition eBURST 5/7 (22 Gruppen) kaum. Die Zahl der Einzeltypen sinkt allerdings erheblich von 38 auf 15. Offenbar existiert eine große Zahl von DLV, die bei der voreingestellten Gruppeneffinition nicht erfasst werden, bei Erweiterung dann aber verbunden werden können.

4.1.4 Einfluss des Ausschluss-Parameters auf die Konkordanz

Innerhalb des in dieser Arbeit untersuchten Parameterspektrums hat der Ausschluss-Parameter auf die Konkordanz zu eBURST im Datensatz Münster

einen geringeren Einfluss als auf die oben analysierte Gruppeneinteilung. So liegt bei einem festen Kosten-Parameter von vier und dabei einem Anstieg des Ausschluss-Parameters von drei über vier und fünf auf schließlich sechs die Konkordanz zwischen BURP und eBURST (6/7) zunächst bei 73,90%, dann bei 88,69% und 95,07% und schließlich bei 94,99%. Dies leuchtet ein, weil von einer Heraufsetzung des Parameters nur diejenigen Isolate mit wenigen Repeats an der Grenze zu der Parameterdefinition betroffen sind, die meisten Isolate aber unbeeinflusst bleiben. Allerdings könnte durchaus auch ein *spa* Typ mit wenigen Repeats eine Bindefunktion innerhalb einer Verwandtschaftskette erfüllen, die durch Ausschluss dieses *spa* Typs verloren ginge. Dieser Fall könnte für den *spa* Typ t104 mit seinem an dieser Grenze liegenden Repeatprofil (r11r10r34r22r25, fünf Repeats) zutreffen, so dass ein über ihn mit dem Rest des *spa*-CC verwandter längerer *spa* Typ zum Einzeltyp würde, obwohl der Kosten-Parameter unverändert bliebe.

Für die definitive Gruppeneinteilung wurde der Ausschluss-Parameter letztlich auf fünf festgesetzt. Ein Grund dafür ist, dass eine weitere Erhöhung des Ausschluss-Parameters auf sechs den Ausschluss von 58 *spa* Typen (14,5% von 400 *spa* Typen) bedeutete. Bei einem Parameter von fünf hingegen werden lediglich 31 *spa* Typen ausgeschlossen (7,75% aller *spa* Typen), was tolerabel erscheint. Zum anderen bedeutete eine Absenkung des Parameters auf vier eine Verminderung der Konkordanz auf 88,69%. Bei dem gewählten Parameter ergibt sich dagegen die höchste Konkordanz zu den Einteilungen nach eBURST 6/7 und eBURST 5/7.

4.1.5 Einfluss des Kosten-Parameters auf die Konkordanz

Innerhalb des in dieser Arbeit untersuchten Parameterspektrums liegt die Konkordanz zwischen eBURST (sowohl in der voreingestellten als auch in der erweiterten Gruppendifinition) und BURP (bei festem Ausschluss-Parameter) für Kosten-Parameter zwischen null und vier bzw. fünf in einem weitgehend homogenen, hohen Bereich (meist über 90%). Das Konkordanzmaximum liegt

dabei jeweils bei einem Kosten-Parameter von drei oder vier. Dabei nimmt die Zahl der Gruppen deutlich zu und die der Einzeltypen deutlich ab. Bei darüber hinaus weiter ansteigendem Kosten-Parameter sinkt die Konkordanz bald auf niedrige Werte bis unter 10%. Diese geringe Konkordanz erklärt sich dadurch, dass in diesem Bereich teilweise nur noch ein oder zwei Gruppen und gleichzeitig nur wenige Einzeltypen definiert werden. Dann findet man im Paarvergleich in der BURP-Gruppierung fast ausschließlich Match-Match-Kombinationen, in der eBURST-Gruppierung aber bleibt es bei der bestehenden Verteilung von Matches und Mismatches.

Bei einem Kosten-Parameter von null werden keine Gruppen gebildet, sondern ausschließlich Einzeltypen, was eine Konkordanz von 100% zwischen *spa* Typen und BURP-Clustern bedeutet. Dieser Effekt ist unabhängig von dem Ausschluss-Parameter. Im Bereich von eBURST fände dies seine Analogie bei einer theoretischen Gruppendifinition eBURST 7/7, bei der alle sieben Allele übereinstimmen müssten.

4.1.6 Beispiele für die Konkordanz

Die beschriebenen hohen Konkordanzwerte zwischen der BURP-Einteilung von *spa* Daten und der für Langzeitstudien etablierten eBURST-Einteilung von MLST-Daten [4, 5, 7] legen nahe, dass in der genetischen Information des *spa* Locus tatsächlich nicht nur ein Kurzzeitsignal vorhanden ist, sondern auch ein Langzeitsignal, wie dies bereits seit einiger Zeit vermutet wurde [21, 23, 39, 47]. Dieses kann offenbar mit BURP erschlossen werden. Die enge Beziehung zwischen der Einteilung nach BURP und der Einteilung nach eBURST wird nun an Beispielen im Datensatz Münster für die oben genannte Parameter-einstellung erläutert.

Fast sämtliche Isolate des CC45 werden von BURP dem *spa*-CC 004 zugeordnet. Ausnahmen sind dabei lediglich die beiden Isolate id1912003 und id4092004 (beide ST617 und *spa*-CC 008) sowie diejenigen Isolate, die von

BURP wegen geringer Zahl von Repeats ausgeschlossen werden. Letztere gehen in die Berechnung der Konkordanz nicht ein, weil für sie kein Paarvergleich durchgeführt wird. Im Paarvergleich werden für den größten Teil der Isolate des CC45 Match-Match-Kombinationen gefunden, denn es ergibt sich in fast allen Fällen, dass zwei jeweils betrachtete Isolate von eBURST dem CC45 und von BURP dem spa-CC 004 zugeordnet werden. Match-Mismatch-Situationen treten bei denjenigen Paarvergleichen auf, bei denen eines der beiden erwähnten Isolate beteiligt ist. Wenn hingegen die beiden erwähnten Isolate als Paar verglichen werden, liegt wiederum eine Match-Match-Kombination vor.

Es gibt allerdings auch Gruppen nach eBURST, die von BURP keineswegs so eindeutig unterstützt werden. Die drei Isolate H19, D139 und 016m118F werden sowohl nach der voreingestellten Gruppendifinition eBURST 6/7 (dann als „gr 8“) als auch nach der erweiterten Gruppendifinition eBURST 5/7 (dann als „gr 9“) als jeweils ein klonaler Komplex sortiert. BURP hingegen findet bei der hier gewählten Parametereinstellung keine Verwandtschaftsbeziehung zwischen den drei Isolaten, sondern bezeichnet zwei von ihnen als Einzeltypen, das dritte wird dem spa-CC 166/240 zugeordnet. Der Paarvergleich dieser Isolate ergibt demnach ausschließlich Match-Mismatch-Kombinationen, was sich auf die Konkordanz mindernd auswirkt.

4.1.7 Selektionsbias im Datensatz Münster

Die sieben häufigsten Sequenztypen im Datensatz Münster werden sowohl bei eBURST 6/7 als auch bei eBURST 5/7 in separate CC klassifiziert. Fünf dieser Sequenztypen stellen in einer Publikation von Robinson et al. die Gründerväter der fünf großen klonalen Komplexe CC5, CC8, CC22, CC30 und CC45 dar [33]. Damit könnten sie nach dem Modell der Populationsstruktur von Bakterien, das von Maynard Smith et al. beschrieben wurde, als die derzeit dominanten Klone von *S. aureus* angesehen werden [28]. Zwar könnte man zunächst annehmen, dass dies sowie die Tatsache, dass die vier häufigsten

Sequenztypen auch in einer repräsentativen Kolonisationsstudie etwa ein Viertel (27,35%) der Isolate ausmachten [13], für eine gewisse Repräsentativität spricht, jedoch ist dies keine hinreichende Argumentation, denn die Art der Isolatauswahl führt zu dem Phänomen, dass Sequenztypen, von denen es viele *spa* Typen gibt, überproportional stark vertreten sind, und umgekehrt *spa* Typen, von denen es mehrere Sequenztypen gibt, überproportional stark vertreten sind. Dieser Selektionsbias verbietet eine Interpretation des errechneten DI und ist die Ursache für die deutliche Diskrepanz der Häufigkeiten einzelner *spa* Typen im Datensatz Münster und den in der Literatur oder im Ridom SpaServer angegebenen Häufigkeiten [16]. Aussagen über die epidemiologische Häufigkeit einzelner *spa* Typen sollten somit aus dem Datensatz Münster nicht getroffen werden.

4.2 Datensatz Koreen

Die erwähnten Parameter wurden gewählt, weil sich bei dieser Einstellung die höchste Konkordanz zu der 0.31 Microarray-Gruppierung ergab (87,2%), die als phylogenetisch am weitesten in die Vergangenheit des Bakteriums reichende Gruppierung angesehen wird. Die Tatsache, dass BURP bei geeigneter Parametereinstellung in der Lage ist, eine Gruppierung vorzunehmen, die die höchste Konkordanz zu eben dieser 0.31-Einteilung erbringt, ist als ein wichtiges Indiz dafür zu werten, dass im *spa* Locus tatsächlich ein phylogenetisch weit in die Vergangenheit reichendes Signal enthalten ist, und dass dieses durch BURP aufgedeckt werden kann. Das Dendrogramm stellt diesen Zusammenhang visuell dar. Weiter ergibt sich der interessante Befund, dass die Konkordanz zu den beiden weiteren Microarray-Gruppierungen mit ansteigendem Korrelationskoeffizienten sinkt (bei 0.5: 86,3%, bei 0.81: 83,6%). Diese Korrelationskoeffizienten bedeuten eine phylogenetisch weniger weit zurückreichende Analyse, und tatsächlich ergaben sich die Konkordanzmaxima für sie bei geringeren Kosten-Parametern.

Koreen et al. hatten in ihrer Publikation die auf *spa* Typen basierenden Verwandtschaftszuordnungen manuell vorgenommen und anhand eines globalen Sequenz-Alignments validiert [23]. Dabei wurde kein Isolat aufgrund zu geringer Zahl von Repeats ausgeschlossen, und die beiden Isolate MSA2099 und MSA2389 wurden aufgrund des identischen Beginns ihrer Repeatsequenzen einer gemeinsamen „lineage 8“ zugeteilt. Die Typisierung der beiden Isolate durch PFGE und *coa* Typisierung lautete sogar identisch, und nach MLEE und nach dem Microarray bei einem Korrelationskoeffizienten von 0.5 wurden sie einer gemeinsamen Gruppe zugeordnet. Es ist nicht auszuschließen, dass die Untersucher bei der manuellen Gruppierung in die „lineage 8“ durch die Ergebnisse der anderen Methoden beeinflusst waren, denn eine sichere Verblindung der Daten ist in der Publikation nicht dokumentiert. BURP hingegen klassifizierte die beiden Isolate bei dem gewählten Kosten-Parameter als Einzeltypen, weil der Abstand der beiden Repeatprofile größer als fünf war. Dennoch ist die Konkordanz zwischen der manuellen Klassifizierung und der BURP-Einteilung mit 96,8% hoch. Für Datensätze mit größerer Datenmenge aber, wie den Datensatz Münster, stellt das automatisch arbeitende Programm BURP eine große Erleichterung dar und ermöglicht eine objektive Gruppierung.

Für den Datensatz Koreen werden im Rahmen dieser Arbeit für die DI der verschiedenen Typisierungsmethoden Konfidenzintervalle angegeben, was in der ursprünglichen Publikation noch nicht der Fall war. Dabei erkennt man, dass der DI der *spa* Typisierung (97,3%) gegenüber demjenigen der PFGE (96,3%) nur insignifikant höher liegt, denn die 95% KI überschneiden sich. Die Ursache für die weiten Konfidenzintervalle und damit die geringe Signifikanz der Werte ist die geringe Größe des Datensatzes von nur 37 Isolaten.

4.3 Datensatz Robinson

Bei dem gewählten Ausschluss-Parameter von fünf schließt BURP kein Isolat aus, weil alle *spa* Typen eine höhere Zahl von Repeats enthalten. Auffällig sind

die abschnittsweise konstanten Gruppeneinteilungen und damit auch konstanten Konkordanzen bei ansteigendem Kosten-Parameter, insbesondere in der Auswahl ohne ST239 und die von ihm abgeleiteten Sequenztypen. Ab einem Kosten-Parameter von sieben tritt kein Einzeltyp mehr auf, und die dabei definierte Gruppeneinteilung in fünf Gruppen wird bis zu einem Kosten-Parameter von dreizehn beibehalten. Erst die Anwendung eines Kosten-Parameters von vierzehn führt zu einem weiteren Zusammenschluss von Gruppen. Für den Gesamtdatensatz wird in diesem Bereich lediglich eine Änderung zu vier Gruppen ab einem Kosten-Parameter von elf gefunden. Die Ursache für diesen konstanten Bereich ist die weitgehend homogene Zusammensetzung des Datensatzes. Viele Isolate haben denselben *spa* Typ, denn die ursprüngliche Isolatauswahl orientierte sich stark an der Eigenschaft der Methicillin-Resistenz und berücksichtigte vor allem die fünf weltweit dominierenden MRSA-Linien [33]. Damit liegt ein Selektionsbias vor, der auch die Ursache für die auffälligen Werte des DI ist, die mit 90,8% für die MLST und mit 93,9% für *spa* nicht nur niedriger liegen als in der betreffenden Publikation für alle *S. aureus* Isolate angegeben [33], sondern auch als in den anderen Datensätzen, die in dieser Arbeit untersucht wurden.

Das Phänomen des um etwa zehn Prozentpunkte höheren Konkordanzmaximums für die Auswahl ohne ST239 und die von ihm abgeleiteten Sequenztypen als für den Gesamtdatensatz entsteht durch die im Gesamtdatensatz häufigeren Match-Mismatch-Kombinationen u.a. durch Isolate des ST239, weil dieser Sequenztyp aus einer Rekombination entstanden ist [34]. Dies wird im Detail weiter unten diskutiert.

4.4 Datensatz Nottingham

Bei der BURP-Einteilung des Datensatzes Nottingham fällt bei einem festen Ausschluss-Parameter von fünf und dabei variiertem Kosten-Parameter auf, dass sich die Zahl der gebildeten BURP-Gruppen in einem Bereich von eins bis acht für den Kosten-Parameter nur unwesentlich zwischen sieben und zehn

bewegt. Dies ist ein anderes Ergebnis als im Datensatz Münster, wo die Zahl der Gruppen mit steigendem Kosten-Parameter kontinuierlich sinkt. Es könnte sein, dass im Datensatz Nottingham weniger *spa* Typen vorhanden sind, die eine Funktion als verbindende *spa* Typen einnehmen können, so dass die bestehenden Gruppen bei Erhöhung des Kosten-Parameters lediglich durch *spa* Typen ergänzt werden, die zuvor Einzeltypen waren. Tatsächlich sinkt die Zahl der Einzeltypen in dem angegebenen Bereich von 41 auf neun. Andererseits könnte das Ergebnis auch dadurch bedingt sein, dass zuvor separate Gruppen sehr wohl zusammengeschlossen werden, gleichzeitig aber *spa* Typen, die zuvor als Einzeltypen klassifiziert waren, zu neuen Gruppen zusammengefasst werden, so dass die Zahl der Gruppen in der Summe weitgehend konstant bleibt. Außerdem könnten auch beide Mechanismen nebeneinander vorliegen. Eine genauere Untersuchung dieser Zusammenhänge wurde im Rahmen dieser Arbeit nicht durchgeführt, könnte aber Gegenstand zukünftiger Studien sein.

Im Gegensatz zu der unterschiedlichen Entwicklung der Zahl der Gruppen verhält sich die Konkordanz zwischen eBURST und BURP in dem angegebenen Bereich ähnlich wie in den Datensätzen Münster und Robinson. Sie liegt relativ homogen über 90% mit einem Maximum von 97,01% (für eBURST 6/7) bzw. 96,50% (für eBURST 5/7). Interessant ist dabei, dass das Konkordanzmaximum bei der voreingestellten Gruppendifinition eBURST 6/7 bei einem mit drei ebenfalls vergleichsweise streng angesetzten Kosten-Parameter für BURP korreliert, während das Konkordanzmaximum für die erweiterte Gruppendifinition eBURST 5/7 bei dem ebenfalls großzügigeren Kosten-Parameter von sechs erreicht wird. Dies zeigt, dass der BURP-Algorithmus die phylogenetisch unterschiedlich weit in die Vergangenheit reichenden Einteilungen, wie sie durch eBURST 6/7 und eBURST 5/7 angezeigt werden, ebenfalls differenzieren kann. Sobald schließlich der Kosten-Parameter auf neun oder höhere Werte gesetzt wird, sinkt die Zahl der gebildeten Gruppen auf fünf und weniger, und die Konkordanzwerte fallen auf

Werte unter 70%. Ein solch deutlicher Abfall der Konkordanzwerte findet im Datensatz Münster bereits bei geringeren Kosten-Parametern statt.

Dies und die bei späteren Untersuchungen gefundenen noch leicht höheren Konkordanzmaxima bei anderen Parametern zeigen, dass die Auswirkungen der Parameter für verschiedene Datensätze unterschiedlich sind und die Angabe allgemeiner Empfehlungen für die Einstellung der BURP-Parameter mit Problemen behaftet ist.

Im Datensatz Nottingham, der unabhängig von einem Selektionsbias ist, weicht der DI für die MLST leicht ab von dem in der ursprünglichen Publikation [13] angegebenen Wert (95,7%), weil sechs Isolate nicht übernommen wurden. Hier wird belegt, dass die Diskriminationsfähigkeit von *spa* (98,0%) signifikant höher ist als diejenige der MLST (95,5%), denn die 95% KI der Indizes überschneiden sich nicht. Damit wird ein Ergebnis von Robinson et al. bestätigt, bei denen der DI der MLST (92,7%) signifikant niedriger war als derjenige der *spa* Typisierung (95,9%) [33].

4.5 Einschränkungen

Zunächst wird eine Einschränkung dieser Arbeit diskutiert, die in der Verschiedenheit der untersuchten Datensätze begründet ist. Die übrigen Abschnitte dieses Kapitels behandeln Probleme in der Anwendung des BURP-Algorithmus.

4.5.1 Problem der biologischen Repräsentativität

Von den vier in dieser Arbeit untersuchten Datensätzen stellen nur der Datensatz Koreen [23] und der Datensatz Nottingham [13, 15] repräsentative Abbilder natürlich auftretender Populationen dar. Daher entsprechen nur diese beiden Datensätze der Forderung nach einer adäquaten Auswahl der Testpopulation, wie sie Struelens et al. insbesondere für die Angabe der Typisierbarkeit und des DI formulierten [44]. Danach soll eine Testpopulation die für die

Natur oder für eine bestimmte Region oder Umgebung erwartete Diversität so repräsentativ wie möglich abbilden. In den beiden anderen Datensätzen liegt dagegen ein Selektionsbias vor, der sich teilweise auch auf die Diskriminationsfähigkeit (DI) der Typisierungsmethoden auswirkt, wie bereits oben diskutiert worden ist. Es könnte also die Evaluation des BURP-Algorithmus anhand dieser beiden Datensätzen in Frage gestellt werden. Jedoch ist festzustellen, dass in dem populationsbasierten Datensatz Nottingham die Konkordanz zwischen der BURP-Gruppierung und der Gruppierung nach eBURST sogar noch höher lag als in den beiden erwähnten Datensätzen. Zur Validierung der hier gefundenen Ergebnisse sollten Studien an weiteren repräsentativen Datenkollektiven unternommen werden.

4.5.2 Problem der Typisierbarkeit von Isolaten

Der BURP-Algorithmus ist auf eine hohe Typisierbarkeit durch die *spa* Methode angewiesen, weil ein nicht typisierbares Isolat auch die Anwendung von BURP für dieses Isolat hinfällig macht. Die geringe Typisierbarkeit der Phagentypisierung von *S. aureus*, die weit unter 100% lag, war einer der Hauptgründe für den Niedergang dieser Methode [37].

Zwar beträgt die rechnerische Typisierbarkeit in den Datensätzen Münster, Robinson und Nottingham sowohl für die *spa* Typisierung als auch für die MLST 100%. Weil für diese Datensätze aber ausschließlich typisierbare Isolate ausgewählt wurden, ist die Angabe dieser Zahl hier nicht sinnvoll. In einer älteren Arbeit an 320 Isolaten wurde für die *spa* Typisierung eine Typisierbarkeit von 100% festgestellt [39]. In der Stammsammlung, die dem Datensatz Nottingham zugrunde lag, betrug die Typisierbarkeit der *spa* Methode 97,4%. In einer Studie an 557 MRSA, die zwischen 1998 und 2003 an einem deutschen Universitätsklinikum isoliert worden waren, lag die Typisierbarkeit der *spa* Methode bei 98,6% und damit ebenfalls sehr hoch [30].

Die nicht vollständige Typisierbarkeit könnte durch *spa* Typen mit so vielen Repeats bedingt sein, dass die Bedingungen der PCR überschritten oder die technischen Möglichkeiten des Kapillarsequenzierers überfordert werden. Eine verminderte Affinität eines PCR-Primers zu seiner Annealingregion ist eine andere mögliche Ursache, kann aber prinzipiell bei jedem sequenzbasierten Verfahren durch geringe Polymorphismen in der Annealingregion auftreten und ist damit nicht spezifisch für die *spa* Typisierung [37]. Diese Problematik ist möglicherweise bei der MLST sogar von größerer Relevanz als bei der *spa* Typisierung, weil dort sieben Primerpaare und nicht nur zwei angewendet werden. Bereits ein einziges nicht typisierbares Allel verhindert dann die Bestimmung des Sequenztyps. Daher ist es umso erstaunlicher, dass in der Literatur keine Informationen über die Höhe der Typisierbarkeit der MLST gefunden werden konnten. Es fanden sich keinerlei Hinweise dafür, dass Isolate nicht typisierbar gewesen wären [4, 5, 13]. Im Rahmen der Laborversuche für diese Arbeit traf dies jedoch auf fünf Isolate zu, die daraufhin nicht in den Datensatz Münster eingeschlossen wurden. Es wäre wünschenswert, die genaue Typisierbarkeit der MLST durch Studien zu ermitteln, retrospektive Untersuchungen könnten dazu zunächst ausreichend sein.

4.5.3 Problem der Wahl der BURP-Parameter

Diese Arbeit dient der Evaluation des BURP-Algorithmus anhand etablierter Referenzmethoden. Er bedarf der Festlegung der beiden BURP-Parameter durch den Anwender. Zweck des Algorithmus ist es nicht, eine möglichst hohe Konkordanz zu bestimmten Referenzmethoden herbeizuführen, sondern eine angemessene Einteilung in Gruppen und Einzeltypen unter Berücksichtigung möglichst vieler *spa* Typen.

Die Definition von Verwandtschaftsbeziehungen von *spa* Typen mit wenigen Repeats bereitet dem BURP-Algorithmus Probleme, so dass solche Isolate aus der Analyse ausgeschlossen werden und möglicherweise basierend auf anderen Methoden wie PFGE oder MLST gruppiert werden müssen. In

Analogie zu der Definition der Typisierbarkeit auf Isolatebene nach Struelens et al. [44] könnte in Zukunft der Ausschluss von *spa* Typen durch BURP abhängig von der Höhe des Ausschluss-Parameters durch einen eigenen Wert beschrieben werden, beispielsweise als Quotient der Zahl der verbleibenden Isolate und der Gesamtzahl der Isolate im Datensatz. Eine mögliche Forderung könnte dann sein, dass dieser Wert über 90% liegen soll. Eine entsprechende Rechenfunktion ließe sich vielleicht sogar in den BURP-Algorithmus integrieren. Dann könnte der Ausschluss-Parameter bei Datensätzen, die einen niedrigen Anteil von *spa* Typen mit wenigen Repeats haben, eventuell automatisch höher gewählt werden als bei Datensätzen, die viele solcher Isolate enthalten. Trotz dieser Möglichkeit wird aber auch in Zukunft die Analyse von *spa* Typen mit wenigen Repeats eine Einschränkung darstellen, der sich der Anwender bewusst sein muss.

Für die Datensätze Münster, Robinson und Nottingham wird in dieser Arbeit ein Bereich des Kosten-Parameters beschrieben, in dem sich die Konkordanzwerte nur wenig verändern. Eine Konsequenz daraus könnte sein, diesen Parameter so hoch wie möglich zu wählen, um wenige Einzeltypen zu erhalten, aber so niedrig wie nötig, um die Konkordanz auf einem hohen Niveau zu erhalten. Es ist anzunehmen, dass der Kosten-Parameter bei Datensätzen mit wenigen Isolaten eher hoch gewählt werden muss, um überhaupt Verwandtschaftsbeziehungen zuzulassen. Möglicherweise jedoch muss die geeignete Parameterhöhe für jeden Datensatz individuell herausgefunden werden. Vielleicht ist auch für den Kosten-Parameter eine Programmierung dahingehend möglich, dass eine bestimmte Zahl von Gruppen oder eine bestimmte prozentuale Gruppengröße gebildet wird, oder dass nur ein bestimmter Anteil der *spa* Typen als Einzeltyp zugelassen wird.

4.5.4 Problem der Angabe von Gründungsvätern

Einzeltypen wurden bei der Anwendung von BURP mit dem Kürzel „sg“ und fortlaufenden Nummern bezeichnet. Die Benennung von Gründungsvätern

eines *spa*-CC ist eine wünschenswerte und informative Funktion des BURP-Algorithmus. Häufig kann BURP jedoch keinen Gründungsvater angeben, betroffene Gruppen wurden in dieser Arbeit mit dem Kürzel „gr“ und fortlaufenden Nummern versehen. Diese beiden Kompromisse in der Behandlung von Einzeltypen und von Gruppen sind von eBURST bekannt. Zusätzlich besteht aber bei dem BURP-Algorithmus das Problem, dass er sich oftmals nicht auf einen einzigen *spa* Typen als Gründungsvater festlegen kann. Dann werden gemeinsame Gründungsvaterschaften angegeben, wie im Falle des *spa*-CC 064/008 im Datensatz Koreen.

4.5.5 Problem der chromosomalen homologen Rekombinationen

Untersuchungen von Isolaten verschiedener Spezies, darunter auch *S. aureus*, haben bestätigt, dass ein relativ großer Anteil einer Bakterienpopulation zu einer vergleichsweise beschränkten Zahl von Clustern eng verwandter Genotypen gehört, die als klonale Komplexe bezeichnet werden [4, 27]. Klonale Komplexe bestehen nach dem Populationsmodell von Maynard-Smith et al. typischerweise aus einem einzelnen dominierenden Genotyp und einigen weiteren, weniger häufigen, engen Verwandten dieses Genotyps [8, 28]. Bakterienpezies mit hoher Rekombinationsrate sind hingegen *Streptococcus pneumoniae*, *Helicobacter pylori* und *Neisseria meningitidis* [6]. Zwar kann die MLST in solchen Fällen zur Identifikation von hypervirulenten zirkulierenden Stämmen zum Beispiel von *Neisseria meningitidis* für einen beschränkten Zeitraum nützlich sein, für phylogenetische Studien aber versagt sie [27]. Es leuchtet ein, dass dann auch eine SLST oder ein BURP-ähnlicher Algorithmus keine Aussage treffen können und ihre Anwendung daher nicht sinnvoll ist. Für Spezies mit klonaler Populationsstruktur hingegen wie *S. aureus* können sowohl die MLST als auch die *spa* Typisierung in Kombination mit BURP genutzt werden, um die Verwandtschaft individueller Isolate zu bestimmen.

Dennoch wurde die Langzeitevolution von *S. aureus* durch einige ungewöhnlich große chromosomale homologe Rekombinationen beeinflusst [5, 33, 34].

Beispielsweise wurde durch Übertragung eines Genomabschnitts von mindestens 557 kb aus einem Isolat mit ST30 und einem Isolat mit ST8 der neue Genotyp ST239 gebildet, wobei der ST239 von ST30 nur das Allel *arcC* (Allel 2) erhielt, alle übrigen MLST-Allele aber von ST8. Dabei legen die *spa* Typen der neu entstandenen Isolate nahe, dass der *spa* Locus genau diejenige Herkunft hat, die auch das *arcC* Allel hat [34]. In der Kreisrepräsentation des Chromosoms von *S. aureus* wird die vergleichsweise enge Nachbarschaft der beiden Loci *spa* und *arcC* im Vergleich zu den anderen MLST-Loci deutlich. Somit erscheint die Möglichkeit, dass eine Rekombination, die den *arcC* Locus erfasste, durchaus auch den *spa* Locus mit einschließen konnte, nicht unwahrscheinlich.

Diese und andere Rekombinationen stellen nicht nur für die MLST in Verbindung mit eBURST eine Herausforderung dar, sondern sie werfen ebenso die Frage auf, inwieweit die Anwendbarkeit der *spa* Typisierung und damit des BURP-Algorithmus auf solche Stämme gegeben ist. BURP definiert die Verwandtschaftsbeziehungen zwischen Isolaten allein nach dem *spa* Locus. Rekombinierte Genotypen lassen sich bei alleiniger Betrachtung eines einzelnen Locus allerdings nur bedingt als verwandt oder als nicht verwandt bezeichnen. Informationen und Hintergründe über stattgefundenene Rekombinationen, wie die MLST sie mit ihren sieben Allelen bietet, können nicht aufgedeckt werden. Es entstehen dann nicht nur kleinere Abweichungen als zufällige Fehler, sondern ganze Gruppen von Isolaten werden durch BURP anders klassifiziert als durch eBURST, so dass dann systematische Fehler vorliegen. Als Einzelmarkersystem stoßen die *spa* Typisierung und damit auch der BURP-Algorithmus hier an eine Grenze.

Im Datensatz Robinson wirken sich diese Hintergründe deutlich auf die Konkordanz aus, weil hier die Zahl der rekombinierten Sequenztypen relativ höher ist als in den übrigen Datensätzen. Weil der durch die erwähnte Rekombination neu entstandene Genotyp mit dem ST239 sechs seiner sieben MLST-Allele, nämlich alle bis auf *arcC*, von demjenigen Herkunftsisolat erhalten

hat, das gerade nicht auch seinen *spa* Typ beeinflusst hat, wird der neue Genotyp von eBURST als SLV der Gruppe des einen Herkunftsisolates zugeordnet, von BURP der Gruppe des anderen Herkunftsisolates. Das bedeutet für den Paarvergleich eine hohe Anzahl an Match-Mismatch-Kombinationen, was die Konkordanz senkt. Wenn im Datensatz Robinson die Isolate mit ST239 und die von ihm abgeleiteten Sequenztypen aus dem Gesamtdatensatz herausgenommen werden, wie dies für die Auswahl durchgeführt wurde, steigt die Konkordanz deutlich.

Um die Rekombinationsrate im Datensatz Münster abzuschätzen, wurde er mit dem Programm BLAND untersucht, mit dem Feil et al. bereits früher die Beziehung zwischen Nukleotiddivergenz und Alleldivergenz bei *S. aureus* und *Streptococcus pneumoniae* verglichen hatten. Für *Streptococcus pneumoniae* ist eine hohe Rate an Rekombinationen bekannt, und es hatte sich tatsächlich eine von der Alleldivergenz weitgehend unabhängige Nukleotiddivergenz gezeigt. Dagegen war bei *S. aureus* mit 75 Sequenztypen ein klarer positiver Trend zwischen Alleldivergenz und Nukleotiddivergenz vorhanden, was ein seltenes Auftreten von Rekombinationen bei diesem Bakterium nahelegte [6]. Diese Ergebnisse werden am Datensatz Münster, der mit 148 Sequenztypen beinahe doppelt so viele Sequenztypen enthält, bestätigt. Dies unterstützt die Vermutung einer weitgehend klonalen Populationsstruktur für *S. aureus* und damit die Zulässigkeit der Untersuchung des Bakteriums mit der *spa* Typisierung und BURP.

Dennoch muss man sich bei der Anwendung des BURP-Algorithmus der Möglichkeit von Ergebnisverzerrungen durch die beschriebenen und gegebenenfalls weitere chromosomale homologe Rekombinationen bewusst sein. Besondere Aufmerksamkeit ist dann notwendig, wenn die erwähnten Sequenztypen in einem zu untersuchenden Datensatz enthalten sind [34].

4.6 Ausblick

Der neue Algorithmus BURP ermöglicht automatische Gruppierungen von *spa* Typen nach ihren Repeatprofilen auf der Grundlage des EDSI-Modells [38]. Dies ist die erste Arbeit, die eine Evaluation von BURP anhand mehrerer Datensätze vornimmt. Sie zeigt eine weitreichende Korrelation der Gruppierung durch BURP mit denjenigen, die durch eBURST (Datensatz Münster, Datensatz Robinson, Datensatz Nottingham) oder andere Methoden (Datensatz Koreen) erzielt werden, die für langzeitepidemiologische Studien als etabliert gelten. Dies legt nahe, dass BURP die Fähigkeit besitzt, in den *spa* Repeatprofilen ein phylogenetisches Signal zu erkennen. Damit existiert erstmals ein aussagekräftiges Cluster-Verfahren unter Nutzung einer einzigen Repeatregion. Es arbeitet automatisch und ist damit für die Anwendung auf große Datensätze geeignet, bei denen manuelle Gruppierungen kaum noch möglich und zudem möglicherweise nicht objektiv sind.

Es wäre zu diesem Zeitpunkt verfrüht, allgemeine Empfehlungen für die Wahl der Höhe der beiden BURP-Parameter zu formulieren. Sie müssen vorerst auf den jeweils vorliegenden Datensatz individuell abgestimmt werden. Vielleicht lassen sich in Zukunft entsprechende Funktionen programmieren, die eine für den einzelnen Datensatz jeweils angemessene Parametereinstellung automatisch vornehmen, wie dies oben diskutiert worden ist. Auch in Hinblick auf diesen Aspekt sollten weitere Studien an unterschiedlichen Datensätzen und Stammsammlungen durchgeführt werden. Die Daten, die seit November 2004 durch die europäische Forscherinitiative SeqNet.org gesammelt werden, könnten sich auch dafür als sehr hilfreich erweisen [10]. Eine kritische Prüfung, ob eine automatisch vorgenommene Einteilung sinnvoll ist, sollte aber auch in diesen Fällen zumindest stichprobenartig durch medizinisches Personal erfolgen.

Der BURP-Algorithmus könnte gegebenenfalls auch auf andere Repeatregionen von *S. aureus* angewendet werden, wie dies beispielsweise eine

Untersuchung von Sabat et al. an den VNTR-Loci *spa*, *sdr*, *clfA*, *clfB*, *ssp* und *coa* mittels simultaner DNA-Amplifikation und anschließender Elektrophorese anregt [37]. Eine Sequenzierung aller genannten Loci und die anschließende Gruppierung mit einer BURP-ähnlichen Software könnte auch die Problemfälle der Isolate mit *spa* Typen mit nur wenigen Repeats lösen. Bei fünf Loci ist es wesentlich unwahrscheinlicher, dass ausschließlich solche mit wenigen Repeats vorliegen. Dies wird anschaulich am Beispiel des Isolates RF122 im Datensatz Koreen, das sich durch lediglich zwei *spa* Repeats, aber fünf *coa* Repeats auszeichnet [23]. Allerdings wäre bei einer Sequenzierung von mehreren Loci das Prinzip der SLST bereits wieder verlassen.

Schließlich erscheint auch eine Anwendung auf VNTR anderer Bakterienarten denkbar, falls sie dem Mutationsmechanismus des „slipped-strand mispairing“ unterliegen [47]. Voraussetzung dafür scheint aber eine klonale Populationsstruktur der betreffenden Bakterienart zu sein [8].

Das im Jahr 2006 von Mellmann et al. beschriebene Frühwarnsystem für MRSA-Ausbrüche in Krankenhäusern [30] könnte in Zukunft durch die BURP-Funktion ausgeweitet werden. Derzeit finden Alarmierungen durch das System bei mehrfachem Auftreten desselben *spa* Typs in dichter räumlicher Nachbarschaft, etwa auf einer Station, statt. Mit BURP wäre es durch die Identifikation eng verwandter *spa* Typen zusätzlich möglich, eine auf den kalkulierten Kosten basierende Warnung auszugeben und auf diese Weise schnelle Mutationsereignisse zu berücksichtigen, wie sie in der Literatur in vivo beschrieben sind [21]. So wäre es schneller als heute möglich, angemessene Hygienemaßnahmen einzuleiten.

Für zunehmend relevante genetische Eigenschaften von *S. aureus* könnten Zusammenhänge mit bestimmten *spa* Typen und BURP-Clustern gefunden werden. Beispielsweise könnte dies für die *lukS-lukF* Determinante gelten, die für das Panton-Valentin Leukocidin (PVL) kodiert, das gehäuft in cMRSA nachgewiesen wird [22, 43]. Zusätzlich steht zu befürchten, dass künftig

Glykopeptid-resistente *S. aureus* (VISA, VRSA) weitere Verbreitung finden werden [1]. Dies könnte die Beherrschbarkeit von MRSA-Infektionen durch Wegfall der therapeutischen Glykopeptid-Option weiter limitieren. Rechtzeitiges Erkennen von genetischen Zusammenhängen und Ausbreitungswegen durch den BURP-Algorithmus könnte zu geeigneten präventiven Maßnahmen und damit zur Vermeidung von Infektionen führen. Somit könnte der BURP-Algorithmus in Zukunft wertvolle mikrobiologische und krankenhaushygienische Beiträge zu der Bekämpfung von *S. aureus* leisten.

5. Literaturverzeichnis

- [1] **Centers for Disease Control and Prevention (CDC)** (2002) *Staphylococcus aureus* resistant to vancomycin - United States, 2002. *Morb Mortal Wkly Rep* 51: 565-567
- [2] **Cosgrove SE, Carmeli Y** (2003) The impact of antimicrobial resistance on health and economic outcomes. *Clin Infect Dis* 36: 1433-1437
- [3] **Emori TG, Gaynes RP** (1993) An overview of nosocomial infections, including the role of the microbiology laboratory. *Clin Microbiol Rev* 6: 428-442
- [4] **Enright MC, Day NPJ, Davies CE, Peacock SJ, Spratt BG** (2000) Multilocus sequence typing for characterization of methicillin-resistant and methicillin-susceptible clones of *Staphylococcus aureus*. *J Clin Microbiol* 38: 1008-1015
- [5] **Enright MC, Robinson DA, Randle G, Feil EJ, Grundmann H, Spratt BG** (2002) The evolutionary history of methicillin-resistant *Staphylococcus aureus* (MRSA). *Proc Natl Acad Sci USA* 99: 7687-7692
- [6] **Feil EJ, Cooper JE, Grundmann H, Robinson DA, Enright MC, Berendt T, Peacock SJ, Maynard Smith J, Murphy M, Spratt BG, Moore CE, Day NPJ** (2003) How clonal is *Staphylococcus aureus*? *J Bacteriol* 185: 3307-3316
- [7] **Feil EJ, Li BC, Aanensen DM, Hanage WP, Spratt BG** (2004) eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. *J Bacteriol* 186: 1518-1530
- [8] **Feil EJ, Spratt BG** (2001) Recombination and the population structures of bacterial pathogens. *Annu Rev Microbiol* 55: 561-590
- [9] **Fitzgerald JR, Sturdevant DE, Mackie SM, Gill SR, Musser JM** (2001) Evolutionary genomics of *Staphylococcus aureus*: insights into the origin of methicillin-resistant strains and the toxic shock syndrome epidemic. *Proc Natl Acad Sci USA* 98: 8821-8826
- [10] **Friedrich AW, Witte W, Harmsen D, de Lencastre H, Hryniewicz W,**

- Scheres J, Westh H** (2006) SeqNet.org: a European laboratory network for sequence-based typing of microbial pathogens. [online] Euro Surveill 11(1): E060112.4
- [11] **Frénay HME, Bunschoten AE, Schouls LM, van Leeuwen WJ, Vandenbroucke-Grauls CMJE, Verhoef J, Mooi FR** (1996) Molecular typing of methicillin-resistant *Staphylococcus aureus* on the basis of protein A gene polymorphism. Eur J Clin Microbiol Infect Dis 15: 60-64
- [12] **Frénay HME, Theelen JPG, Schouls LM, Vandenbroucke-Grauls CMJE, Verhoef J, van Leeuwen WJ, Mooi FR** (1994) Discrimination of epidemic and nonepidemic methicillin-resistant *Staphylococcus aureus* strains on the basis of protein A gene polymorphism. J Clin Microbiol 32: 846-847
- [13] **Grundmann H, Hori S, Enright MC, Webster C, Tami A, Feil EJ, Pitt T** (2002) Determining the genetic structure of the natural population of *Staphylococcus aureus*: a comparison of multilocus sequence typing with pulsed-field gel electrophoresis, randomly amplified polymorphic DNA analysis, and phage typing. J Clin Microbiol 40: 4544-4546
- [14] **Grundmann H, Hori S, Tanner G** (2001) Determining confidence intervals when measuring genetic diversity and the discriminatory abilities of typing methods for microorganisms. J Clin Microbiol 39: 4190-4192
- [15] **Grundmann H, Tami A, Hori S, Halwani M, Slack R** (2002) Nottingham *Staphylococcus aureus* population study: prevalence of MRSA among elderly people in the community. Br Med J 324: 1365-1366
- [16] **Harmsen D, Claus H, Witte W, Rothgänger J, Claus H, Turnwald D, Vogel U** (2003) Typing of methicillin-resistant *Staphylococcus aureus* in a university hospital setting by using novel software for *spa* repeat determination and database management. J Clin Microbiol 41: 5442-5448
- [17] **Hiramatsu K, Hanaki H, Ino T, Yabuta K, Oguri T, Tenover FC** (1997) Methicillin-resistant *Staphylococcus aureus* clinical strain with reduced vancomycin susceptibility. J Antimicrob Chemother 40: 135-136

- [18] **Hunter PR, Gaston MA** (1988) Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. *J Clin Microbiol* 26: 2465-2466
- [19] **Jevons MP** (1961) 'Celbenin'-resistant *Staphylococci*. *Br Med J* 1: 124-125
- [20] **Jolley KA, Feil EJ, Chan MS, Maiden MCJ** (2001) Sequence type analysis and recombinational tests (START). *Bioinformatics* 17: 1230-1231
- [21] **Kahl BC, Mellmann A, Deiwick S, Peters G, Harmsen D** (2005) Variation of the polymorphic region X of the protein A gene during persistent airway infection of cystic fibrosis patients reflects two independent mechanisms of genetic change in *Staphylococcus aureus*. *J Clin Microbiol* 43: 502-505
- [22] **Kazakova SV, Hageman JC, Matava M, Srinivasan A, Phelan L, Garfinkel B, Boo T, McAllister S, Anderson J, Jensen B, Dodson D, Lonsway D, McDougal LK, Arduino M, Fraser VJ, Killgore G, Tenover FC, Cody S, Jernigan DB** (2005) A clone of methicillin-resistant *Staphylococcus aureus* among professional football players. *N Engl J Med* 352: 468-475
- [23] **Koreen L, Ramaswamy SV, Graviss EA, Naidich S, Musser JM, Kreiswirth BN** (2004) *spa* typing method for discriminating among *Staphylococcus aureus* isolates: implications for use of a single marker to detect genetic micro- and macrovariation. *J Clin Microbiol* 42: 792-799
- [24] **Kresken M, Hafner D, Schmitz FJ, Wichelhaus TA** (2004) Resistenzsituation bei klinisch wichtigen Infektionserregern gegenüber Antibiotika in Deutschland und im mitteleuropäischen Raum. [online] <http://www.p-e-g.de>
- [25] **Kumar S, Tamura K, Nei M** (2004) MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinform* 5: 150-163
- [26] **Kuroda M, Ohta T, Uchiyama I, Baba T, Yuzawa H, Kobayashi I, Cui L, Oguchi A, Aoki K, Nagai Y, Lian J, Ito T, Kanamori M, Matsumaru**

- H, Maruyama A, Murakami H, Hosoyama A, Mizutani-Ui Y, Takahashi NK, Sawano T, Inoue R, Kaito C, Sekimizu K, Hirakawa H, Kuhara S, Goto S, Yabuzaki J, Kanehisa M, Yamashita A, Oshima K, Furuya K, Yoshino C, Shiba T, Hattori M, Ogasawara N, Hayashi H, Hiramatsu K (2001) Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*. *Lancet* 357: 1225-1240
- [27] Maiden MCJ, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, Zhang Q, Zhou J, Zurth K, Caugant DA, Feavers IM, Achtman M, Spratt BG (1998) Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci USA* 95: 3140-3145
- [28] Maynard Smith J, Feil EJ, Smith NH (2000) Population structure and evolutionary dynamics of pathogenic bacteria. *Bioessays* 22: 1115-1122
- [29] McDougal LK, Steward CD, Killgore GE, Chaitram JM, McAllister SK, Tenover FC (2003) Pulsed-field gel electrophoresis typing of oxacillin-resistant *Staphylococcus aureus* isolates from the United States: establishing a national database. *J Clin Microbiol* 41: 5113-5120
- [30] Mellmann A, Friedrich AW, Rosenkötter N, Rothgänger J, Karch H, Reintjes R, Harmsen D (2006) Automated DNA sequence-based early warning system for the detection of methicillin-resistant *Staphylococcus aureus* outbreaks. [online] *PLoS Med* (3)3: e33
- [31] Murchan S, Kaufmann ME, Deplano A, de Ryck R, Struelens M, Zinn CE, Fussing V, Salmenlinna S, Vuopio-Varkila J, El Solh N, Cuny C, Witte W, Tassios PT, Legakis N, van Leeuwen W, van Belkum A, Vindel A, Laconcha I, Garaizar J, Haeggman S, Olsson-Liljequist B, Ransjö U, Coombes G, Cookson B (2003) Harmonization of pulsed-field gel electrophoresis protocols for epidemiological typing of strains of methicillin-resistant *Staphylococcus aureus*: a single approach developed by consensus in 10 European laboratories and its application for tracing the spread of related strains. *J Clin Microbiol* 41: 1574-1585
- [32] Muto CA, Jernigan JA, Ostrowsky BE, Richet HM, Jarvis WR, Boyce JM, Farr BM (2003) SHEA guideline for preventing nosocomial

- transmission of multidrug-resistant strains of *Staphylococcus aureus* and *Enterococcus*. *Infect Control Hosp Epidemiol* 24: 362-386
- [33] **Robinson DA, Enright MC** (2003) Evolutionary models of the emergence of methicillin-resistant *Staphylococcus aureus*. *Antimicrob Agents Chemother* 47: 3926-3934
- [34] **Robinson DA, Enright MC** (2004) Evolution of *Staphylococcus aureus* by large chromosomal replacements. *J Bacteriol* 186: 1060-1064
- [35] **Robinson DA, Hollingshead SK, Musser JM, Parkinson AJ, Briles DE, Crain MJ** (1998) The IS1167 insertion sequence is a phylogenetically informative marker among isolates of serotype 6B *Streptococcus pneumoniae*. *J Mol Evol* 47: 222-229
- [36] **Rubin RJ, Harrington CA, Poon A, Dietrich K, Greene JA, Moiduddin A** (1999) The economic impact of *Staphylococcus aureus* infection in New York City hospitals. *Emerg Infect Dis* 5: 9-17
- [37] **Sabat A, Krzyszton-Russjan J, Strzalka W, Filipek R, Kosowska K, Hryniewicz W, Travis J, Potempa J** (2003) New method for typing *Staphylococcus aureus* strains: multiple-locus variable-number tandem repeat analysis of polymorphism and genetic relationships of clinical isolates. *J Clin Microbiol* 41: 1801-1804
- [38] **Sammeth M, Stoye J** (2006) Comparing tandem repeats with duplications and excisions of variable degree. *IEEE/ACM Trans Comput Biol Bioinform* 3: 395-407
- [39] **Shopsin B, Gomez M, Montgomery SO, Smith DH, Waddington M, Dodge DE, Bost DA, Riehman M, Naidich S, Kreiswirth BN** (1999) Evaluation of protein A gene polymorphic region DNA sequencing for typing of *Staphylococcus aureus* strains. *J Clin Microbiol* 37: 3556-3563
- [40] **Shopsin B, Gomez M, Waddington M, Riehman M, Kreiswirth BN** (2000) Use of coagulase gene (*coa*) repeat region nucleotide sequences for typing of methicillin-resistant *Staphylococcus aureus* strains. *J Clin Microbiol* 38: 3453-3456
- [41] **Simpson EH** (1949) Measurement of diversity. *Nature* 163: 688
- [42] **Stewart GT, Holt RJ** (1963) Evolution of natural resistance to the newer

- penicillins. *Br Med J* 1: 308-311
- [43] **Strommenger B, Kettlitz C, Weniger T, Harmsen D, Friedrich AW, Witte W** (2006) Assignment of *Staphylococcus* isolates to groups by *spa* typing, *Sma*I macrorestriction analysis, and multilocus sequence typing. *J Clin Microbiol* 44: 2533-2540
- [44] **Struelens MJ** (1996) Consensus guidelines for appropriate use and evaluation of microbial epidemiologic typing systems. *Clin Microbiol Infect* 2: 2-11
- [45] **Tang YW, Waddington MG, Smith DH, Manahan JM, Kohner PC, Highsmith LM, Li H, Cockerill III FR, Thompson RL, Montgomery SO, Persing DH** (2000) Comparison of protein A gene sequencing with pulsed-field gel electrophoresis and epidemiologic data for molecular typing of methicillin-resistant *Staphylococcus aureus*. *J Clin Microbiol* 38: 1347-1351
- [46] **Tenover FC, Arbeit R, Archer G, Biddle J, Byrne S, Goering R, Hancock G, Hébert GA, Hill B, Hollis R, Jarvis WR, Kreiwirth B, Eisner W, Maslow J, McDougal LK, Miller JM, Mulligan M, Pfaller MA** (1994) Comparison of traditional and molecular methods of typing isolates of *Staphylococcus aureus*. *J Clin Microbiol* 32: 407-415
- [47] **van Belkum A, Scherer S, van Alphen L, Verbrugh H** (1998) Short-sequence DNA repeats in prokaryotic genomes. *Microbiol Mol Biol Rev* 62: 275-293
- [48] **van Belkum A, van Leeuwen W, Kaufmann ME, Cookson B, Forey F, Etienne J, Goering R, Tenover F, Steward C, O'Brien F, Grubb W, Tassios P, Legakis N, Morvan A, El Solh N, de Ryck R, Struelens M, Salmenlinna S, Vuopio-Varkila J, Kooistra M, Talens A, Witte W, Verbrugh H** (1998) Assessment of resolution and intercenter reproducibility of results of genotyping *Staphylococcus aureus* by pulsed-field gel electrophoresis of *Sma*I macrorestriction fragments: a multicenter study. *J Clin Microbiol* 36: 1653-1659

6. Verzeichnis der Abkürzungen

acc. no.	accession number (engl.)
<i>arcC</i>	MLST-Locus (carbamate kinase, engl.)
<i>aroE</i>	MLST-Locus (shikimate dehydrogenase, engl.)
BLAND	Based On Locus And Nucleotide Differences (engl.)
bp	Basenpaar
BURP	Based Upon Repeat Patterns (engl.)
BURST	Based Upon Related Sequence Types (engl.)
CC	clonal complex (engl.)
CDC	Centers for Disease Control and Prevention (engl.)
cMRSA	community-acquired MRSA (engl.)
DI	diskriminatorischer Index
DLV	Double Locus Variant (engl.)
DNA	Desoxyribonukleinsäure
dNTP	Desoxynukleotidtriphosphat
eBURST	(engl.), Weiterentwicklung von BURST
EDTA	Ethylendiamintetraacetat
ex	excluded (engl.), aus einer Analyse ausgeschlossen
fwd	forward (engl.), Vorwärts-Primer
g	Vielfaches der Erdbeschleunigung in der Zentrifuge
GISA	Glykopeptid-intermediär suszeptibler <i>S. aureus</i>
<i>glpF</i>	MLST-Locus (glycerol kinase, engl.)
<i>gmk</i>	MLST-Locus (guanylate kinase, engl.)
gr	group(s) (engl.), Gruppe(n)
HPLC	high pressure liquid chromatographie (engl.)
kb	Kilobasen
KI	Konfidenzintervall
Lin	Lineage (engl.), Gruppierung
<i>mecA</i>	Methicillin resistance gene (engl.)
min	Minuten
MLEE	Multi Locus Enzyme Electrophoresis (engl.)

MLST	Multi Locus Sequence Typing (engl.)
MRSA	Methicillin-resistenter <i>S. aureus</i>
MSSA	Methicillin-sensibler <i>S. aureus</i>
µl	Mikroliter (10 ⁻⁶ l)
n.b.	nicht bestimmt
ORF	Open reading frame (engl.)
PCR	Polymerase chain reaction (engl.)
PFGE	Pulsfeldgelelektrophorese
<i>pta</i>	MLST-Locus (phosphate acetyltransferase, engl.)
rev	reverse (engl.), Rückwärts-Primer
RIDOM	Ribosomal Differentiation of Microorganisms (engl.)
s	Sekunden
<i>S. aureus</i>	<i>Staphylococcus aureus</i>
SCC <i>mec</i>	Staphylococcal cassette chromosome <i>mec</i> (engl.)
sg	singleton(s) (engl.), Einzeltyp(en)
SLST	Single Locus Sequence Typing (engl.)
SLV	Single Locus Variant (engl.)
<i>smal</i>	PFGE-Restriktionsenzym für <i>S. aureus</i>
<i>spa</i>	<i>S. aureus</i> Protein A
SSM	slipped-strand mispairing (engl.)
subsp.	subspecies
ST	Sequenztyp in der MLST
Taq-Polymerase	<i>Thermus aquaticus</i> DNA-Polymerase
TBE	Tris-Borsäure-EDTA
TE	Tris-HCl, EDTA
<i>tpi</i>	MLST-Locus (triosephosphate isomerase, engl.)
Tris	Trishydroxyethylaminoethan
U/min	Umdrehungen pro Minute
VISA	Vancomycin-intermediär suszeptibler <i>S. aureus</i>
VNTR	Variable Number of Tandem Repeats (engl.)
VRSA	Vancomycin-resistenter <i>S. aureus</i>
<i>yqiL</i>	MLST-Locus (acetyl coenzyme A acetyltransferase, engl.)

7. Verzeichnis der Abbildungen, Tabellen und Diagramme

7.1 Verzeichnis der Abbildungen

Abbildung 1-1: Variationsgeschwindigkeit von Typisierungsmethoden.

Abbildung 1-2: Positionen der MLST-Loci und des *spa* Locus.

Abbildung 1-3: Der *spa* Locus von *S. aureus*.

Abbildung 1-4: Mechanismus des „slipped-strand mispairing“.

Abbildung 3-1: Populationsübersicht, Datensatz Münster.

Abbildung 3-2: Graphische Repräsentation eines *spa*-CC, Datensatz Münster.

7.2 Verzeichnis der Tabellen

Tabelle 2-1: Verwendete synthetische Oligonukleotide (Primer).

Tabelle 2-2: Daten der MLST für *S. aureus*.

Tabelle 2-3: Allelische Profile der provisorischen Sequenztypen.

Tabelle 2-4: Schema für die Berechnung der Konkordanz im Paarvergleich.

Tabelle 2-5: Herkunft der Isolate, Datensatz Münster.

Tabelle 3-1: Häufigkeit von Sequenztypen, Datensatz Münster

Tabelle 3-2: Häufigkeit von *spa* Typen, Datensatz Münster.

Tabelle 3-3: BURP und Konkordanz zu eBURST, Datensatz Münster.

Tabelle 3-4: Verschiedene Typisierungsmethoden, Datensatz Koreen.

Tabelle 3-5: Konkordanzen, Datensatz Koreen.

Tabelle 3-6: BURP und Konkordanz zu eBURST, weitere Datensätze.

7.3 Verzeichnis der Diagramme

Diagramm 3-1: BLAND, Alleldivergenz und Nukleotiddivergenz.

Diagramm 3-2: Zahl der Repeats von *spa* Typen, Datensatz Münster.

8. Danksagung

Mein Dank gilt Herrn Prof. Dr. med. D. Harmsen für die Überlassung des Themas dieser Dissertation und für die Einführung in das wissenschaftliche Denken.

Herrn Prof. Dr. rer. nat. H. Karch danke ich für die Erlaubnis, die Laborversuche in den Räumlichkeiten seines Instituts durchführen zu dürfen.

Bei Herrn Dr. med. A. Mellmann bedanke ich mich für seine hilfreichen Ratschläge bei der Planung, Durchführung und Abfassung der Arbeit.

Für die lehrreiche Einführung in die praktische Labortätigkeit bedanke ich mich bei Frau U. Keckevoet.

Auch den weiteren Mitarbeitern des Instituts für Hygiene am Universitätsklinikum Münster danke ich für die stets freundliche und hilfsbereite Atmosphäre.

Der Studienstiftung des deutschen Volkes und dort insbesondere meiner Vertrauensdozentin Frau Prof. Dr. R.-E. Mohrmann danke ich für die Unterstützung meines Studiums sowie anregende und bereichernde Diskussionen.

Zuletzt danke ich besonders meinen Eltern Dr. med. G. und Dr. med. O. Berssenbrügge, die auch mir das Studium der Medizin nahe gebracht und ermöglicht haben. Sie haben mich an ihren Lebenserfahrungen teilhaben lassen und mich immer liebevoll begleitet.

9. Lebenslauf

Anhang

Anhang A: Datensatz Münster

Isolat	ST ^a	eBURST ^b		spa Typ	BURP ^e
		6/7 ^c	5/7 ^d		
015m134	1	1	1	t098	127
008spastapaur17	1	1	1	t114	127
008spastapaur47	1	1	1	t125	sg 6
CC007	1001	1	1	t127	127
HT2001254	1	1	1	t128	127
RKK65MSSA	1	1	1	t174	127
HT2001747	1	1	1	t175	sg 10
H462	1	1	1	t176	sg 11
D473	69	1	1	t210	sg 17
016m108	1	1	1	t273	gr 2
id4262004	1	1	1	t321	127
spa9	1	1	1	t343	gr 2
BAP94-050	1	1	1	t386	ex
021spa787	1	1	1	t405	127
Btn2242	5	5	5	t001	002
spa#a10035#03	5	5	5	t002	002
3441/02	5	5	5	t003	002
spa#k03#102#03	225	5	5	t003	002
NJ992(CDC4)	5	5	5	t010	002
VA11385	225	5	5	t014	002
VA74852003	225	5	5	t027	ex
id3772004	225	5	5	t035	002
Belgium97S101	5	5	5	t045	002
CDC12	225	5	5	t045	002
ST554-MRSA-II	554	5	5	t045	002
PolandMR1	5	5	5	t053	002
963SMALL(CDC1)	5	5	5	t062	002
007spaa57192002	5	5	5	t067	002
1620037	597	5	5	t071	002
016m43R	597	5	5	t072	002
016m105	5	5	5	t088	002
E3001	5	5	5	t105	002
015m057	5	5	5	t106	sg 4
015m093	5	5	5	t111	ex
015m218	5	5	5	t134	002
015m299	5	5	5	t143	002
id0692003	5	5	5	t149	002
id2922004	225	5	5	t151	002
Slovenia30	5	5	5	t178	002
C433	5	5	5	t179	002
Cuba4028	5	5	5	t187	002
D316	11	5	5	t212	002
BK519	5	5	5	t214	002
016m132	5	5	5	t242	002
id2542003	225	5	5	t264	002
id3722004	225	5	5	t283	002
004spa152004	627	5	5	t306	002
BAP01#263B	627	5	5	t311	002
BAP02#049	628	5	5	t312	sg 29
UK-3-2407	5	5	5	t319	002
id4412004	225	5	5	t322	ex
N39	5	5	5	t389	002
016m091	27	9	9	t099	099
H169	9	9	9	t100	100
CC117	27	9	9	t100	100
016m124	9	9	9	t158	sg 8
D410	27	9	9	t191	100
H150	9	9	9	t193	100
D472	109	9	9	t209	004
D295	9	9	9	t215	099
H116	9	9	9	t221	099
Bov3080	9	9	9	t235	sg 22
spa25	9	9	9	t337	sg 31
021spa751	9	9	9	t411	099
C126	12	12	12	t156	160
Not151	12	12	12	t160	160
H402	13	12	12	t160	160
D224	12	12	12	t183	160

H117	12	12	12	t213	160
spa51	12	12	12	t327	160
spa20	604	12	12	t336	ex
spa140	12	12	12	t366	160
016m068	15	15	15	t083	084
BK3691	15	15	15	t084	084
H291	18	15	15	t084	084
CC245	1025	15	15	t084	084
D61	15	15	15	t085	084
016m088	15	15	15	t094	084
CC011	1021	15	15	t094	084
016m095	15	15	15	t097	084
015m170	15	15	15	t120	084
015m301	15	15	15	t144	084
BK4332	15	15	15	t165	sg 9
C154	14	15	15	t181	sg 12
D508	58	15	15	t228	084
VA3392004	15	15	15	t254	084
016m254	15	15	15	t256	084
spa147	15	15	15	t279	084
spa8	15	15	15	t326	sg 30
spa86	15	15	15	t328	084
spa79	15	15	15	t335	084
CC102	15	15	15	t346	084
spa84	620	15	15	t346	084
spa88	15	15	15	t348	084
spa141	15	15	15	t358	084
spa121	15	15	15	t360	084
spa146	15	15	15	t368	084
021spa719	15	15	15	t385	084
021spa727	15	15	15	t393	084
B16522#01	15	15	15	t423	084
Bin2445	22	22	22	t005	005
C49	23	22	22	t005	005
H40	60	22	22	t005	005
CC136	217	22	22	t005	005
D374	134	22	22	t006	005
015m108	22	22	22	t016	005
VA59722004	22	22	22	t020	005
UK98/14719	22	22	22	t022	005
VA60882003	22	22	22	t025	005
spa#w2358#03	22	22	22	t032	005
HK1963	22	22	22	t060	005
016m082	22	22	22	t092	sg 2
015m007	22	22	22	t101	sg 3
015m025	22	22	22	t103	ex
id1342003	22	22	22	t107	005
D210	21	22	22	t192	005
CC443	217	22	22	t192	005
C13	22	22	22	t223	005
BK4777	22	22	22	t249	005
id3782004	22	22	22	t290	005
spaws1c22	622	22	22	t294	005
1480032004	606	22	22	t307	gr 1
6092004	22	22	22	t309	005
6082004	22	22	22	t310	005
spaid4662004	22	22	22	t372	sg 35
021spa732	1018	22	22	t395	005
021spa815	22	22	22	t403	005
H034680406	22	22	22	t417	sg 40
HO40780032	22	22	22	t420	005
CC021	25	25	25	t078	078
H112	26	25	25	t078	078
016m056	25	25	25	t079	ex
015m319	26	25	25	t081	078
016m028RW	25	25	25	t087	078
015m273	26	25	25	t140	078
016m101	25	25	25	t152	078
Not320	25	25	25	t167	078
H118	28	25	25	t168	078
C437	25	25	25	t173	078
D279	25	25	25	t227	ex
016m283	25	25	25	t244	078
016m251	25	25	25	t258	078
016m269	25	25	25	t259	sg 24
016m267	25	25	25	t260	078
016m253	25	25	25	t262	078
016m277	25	25	25	t263	078
VA26732004	25	25	25	t280	078
id3732004	25	25	25	t287	ex
CC057	1017	25	25	t287	ex

spa52	25	25	25	t329	078
spa46	25	25	25	t341	078
spa73	25	25	25	t349	078
spa44	25	25	25	t353	078
021spa734	1017	25	25	t396	078
021spa817	25	25	25	t401	078
ATCC43300	39	30	30	t007	012
Btn1260	30	30	30	t012	012
VA2138	30	30	30	t017	012
WW2703/97	30	30	30	t018	012
UK96/32010	36	30	30	t018	012
H137MRSA	38	30	30	t018	012
008spastapaur58	30	30	30	t019	012
E1410	30	30	30	t021	012
H399	33	30	30	t021	012
008spastapaur01	30	30	30	t076	012
016m227	34	30	30	t089	166/240
016m075	30	30	30	t090	012
016m085	30	30	30	t093	012
CC328	1005	30	30	t096	012
008spastapaur42	30	30	30	t122	012
015m224	598	30	30	t135	012
id1052003	34	30	30	t136	166/240
016m106RW	34	30	30	t153	166/240
016m170	34	30	30	t154	166/240
016m089	34	30	30	t155	sg 7
spa85	34	30	30	t166	166/240
H140	32	30	30	t205	012
D318	57	30	30	t217	012
C253	40	30	30	t219	012
CC335	30	30	30	t233	ex
Btn1873	39	30	30	t233	ex
id1892003	30	30	30	t238	012
016m116	30	30	30	t245	012
2494032004	30	30	30	t268	012
C2SAU0077	605	30	30	t275	012
C2SAU0032	30	30	30	t276	012
id3542004	30	30	30	t281	012
id3922004	30	30	30	t291	ex
CC036	39	30	30	t298	012
BAS96#002	36	30	30	t300	012
BAP01#173	36	30	30	t313	012
UK-3-2275	30	30	30	t318	012
UK-2-4791	39	30	30	t320	012
BAP02#136	36	30	30	t323	012
spa16	30	30	30	t338	012
spa36	30	30	30	t342	012
spa12	34	30	30	t352	166/240
spa134	30	30	30	t363	012
spa119	34	30	30	t365	166/240
spa113	34	30	30	t369	166/240
BAP03#046	36	30	30	t374	012
spa144	39	30	30	t378	012
BAP04#183	30	30	30	t381	012
021spa712	39	30	30	t382	012
021spa723	39	30	30	t384	012
2956	30	30	30	t391	ex
021spa770	39	30	30	t399	012
021spa804	378	30	30	t402	012
021spa782	30	30	30	t404	012
021spa791	30	30	30	t406	012
021spa813	39	30	30	t407	012
021spa798	34	30	30	t410	166/240
021spa748	37	30	30	t413	012
H040220406	36	30	30	t418	012
H042340015	36	30	30	t419	012
021spa759	30	30	30	t429	012
spa#k02#3058#02	45	45	45	t004	004
Btn2306	45	45	45	t015	004
Btn1206	319	45	45	t015	004
Btn2299	322	45	45	t015	004
015m0922000	45	45	45	t026	ex
CC087	47	45	45	t026	ex
id1352002	45	45	45	t028	004
id4222004	45	45	45	t029	004
NL011919-4	45	45	45	t031	004
id2652003	45	45	45	t033	004
Belgium97S99 (1432)	45	45	45	t038	004
id3612004	45	45	45	t040	004
id0482002	45	45	45	t043	004
VA7758	45	45	45	t049	004

IV

SwedenCCUG41787	45	45	45	t050	004
1520038	45	45	45	t061	004
WWPS95	45	45	45	t065	004
WW298/95	46	45	45	t065	004
id0312002	45	45	45	t069	004
id2892004	45	45	45	t073	004
007spaK032409	45	45	45	t077	004
016m063	45	45	45	t080	004
016m093	45	45	45	t095	004
015m004	45	45	45	t102	004
015m128	45	45	45	t116	004
008spastapaur43	45	45	45	t123	004
008spastapaur44	45	45	45	t124	004
008spastapaur65	45	45	45	t130	004
spaid3442004	45	45	45	t132	ex
015m281	45	45	45	t141	004
015m287	45	45	45	t142	004
015m381	45	45	45	t147	ex
016m120	45	45	45	t157	004
id2942004	45	45	45	t161	004
D49	53	45	45	t180	004
CDC8	45	45	45	t204	004
D98	54	45	45	t220	004
C233	45	45	45	t230	004
id1912003	617	45	45	t243	008
id2042003	45	45	45	t247	004
id2762004	45	45	45	t266	004
id3162004	45	45	45	t277	004
id3562004	45	45	45	t282	ex
BAP00#125	278	45	45	t295	004
id4092004	617	45	45	t305	008
spa56	45	45	45	t330	004
spa38	45	45	45	t331	004
spa34	45	45	45	t333	004
spa70	45	45	45	t339	ex
spa87	45	45	45	t340	004
M17	45	45	45	t350	004
spa109	45	45	45	t361	004
spa151	45	45	45	t362	ex
CC299	46	45	45	t362	ex
spa135	45	45	45	t370	004
spa132	45	45	45	t371	004
021spa710	1008	45	45	t383	004
A005a	45	45	45	t390	ex
021spa753	1009	45	45	t397	004
B24123#01	45	45	45	t424	004
<hr/>					
015m146	121	121	121	t119	159
016m107	121	121	121	t159	159
id1482003	121	121	121	t162	159
C3	51	121	121	t169	159
H560	121	121	121	t170	159
1488002003	3002	121	121	t269	sg 25
1419972004	121	121	121	t270	159
1452#97	121	121	121	t272	159
1876002004	121	121	121	t284	159
2878022004	121	121	121	t285	159
1451972004	121	121	121	t308	159
UK-2-1901	121	121	121	t314	gr 4
UK-2-7911	121	121	121	t317	gr 4
<hr/>					
Fin62305	296	188	1	t172	316
Can6428-011	188	188	1	t184	ex
spa65	188	188	1	t189	044
CC380	1003	188	1	t189	044
<hr/>					
E228	8	239	239	t008	008
Germany1869/98	247	239	239	t008	008
E213	250	239	239	t008	008
KD6505	254	239	239	t008	008
CC070	1014	239	239	t008	008
ST2001-MRSA-I	2001	239	239	t008	008
Germany1966/97	254	239	239	t009	008
France97151	8	239	239	t024	008
Slovenia54	239	239	239	t030	012
SwedenON 408/99	246	239	239	t030	012
id2782004	254	239	239	t036	008
ST158-MRSA-III	158	239	239	t037	012
UKNCTC11939	239	239	239	t037	012
ST240-MRSA-III	240	239	239	t037	012
Finland98541	241	239	239	t037	012
id0522002	241	239	239	t047	012
HK1444	254	239	239	t048	008
Germany134/93	247	239	239	t051	008

V

SwedenHS 2	247	239	239	t052	008
82MRSA	247	239	239	t054	008
016m010	8	239	239	t059	ex
VA64162004	8	239	239	t063	008
T12003	8	239	239	t064	008
007spai582002	8	239	239	t068	008
id4402004	241	239	239	t074	012
008spastapaur09	247	239	239	t075	008
015m026	8	239	239	t104	008
VA12935	8	239	239	t112	008
015m111	254	239	239	t115	008
008spastapaur75	247	239	239	t121	008
008spastapaur62	3001	239	239	t129	ex
015m211	254	239	239	t133	004
015m241	241	239	239	t137	012
015m245	241	239	239	t138	012
015m258	254	239	239	t139	008
015m347	254	239	239	t146	008
EMRSA2	8	239	239	t190	008
EMRSA8	250	239	239	t194	008
Cuba4005	94	239	239	t197	008
Spain5	247	239	239	t200	008
NL010548-1	8	239	239	t206	008
NCTC8325	8	239	239	t211	008
Finland37481	239	239	239	t234	012
016m252	8	239	239	t257	gr 6
BAP00#140	239	239	239	t297	012
BAP99#142	623	239	239	t298	012
BAP99#139	639	239	239	t299	012
BAS94#020	624	239	239	t301	008
BAS94#051	626	239	239	t303	008
BD92#10560	631	239	239	t304	008
spa1502004	630	239	239	t377	gr 6
2690	239	239	239	t387	012
H390	239	239	239	t388	012
021spa737	8	239	239	t394	008
BAP04#209	241	239	239	t421	012
M101	8	239	239	t430	008
016m065	395	395	395	t082	gr 3
1486032004	426	395	395	t271	sg 26
spa104	395	395	395	t367	gr 3
021spa823	1012	395	395	t412	004
D470	207	509	509	t226	sg 20
016m110	509	509	509	t239	sg 23
id1502003	509	509	509	t248	ex
BD04#3325	89	509	509	t375	sg 37
Germany1155-1/98	228	gr 1	5	t001	002
S99	228	gr 1	5	t013	002
VA55122004	228	gr 1	5	t023	sg 1
007spaa178562002	228	gr 1	5	t039	002
007spaw5402002	111	gr 1	5	t041	002
3531/02	228	gr 1	5	t041	002
id0512003	111	gr 1	5	t055	002
007spaa12152002	228	gr 1	5	t066	002
015m071	228	gr 1	5	t109	002
015m086	228	gr 1	5	t110	002
Slovenia33	228	gr 1	5	t188	002
Slovenia25	228	gr 1	5	t201	002
id2712004	228	gr 1	5	t265	002
VA36752004	228	gr 1	5	t288	ex
id2772004	398	gr 2	gr 1	t011	012
id2872004	398	gr 2	gr 1	t034	012
015m069	621	gr 2	gr 1	t108	012
Not38	101	gr 3	gr 4	t056	078
015m468	101	gr 3	gr 4	t150	078
CC266	505	gr 3	gr 4	t232	ex
015m305	154	gr 4	gr 6	t145	ex
BAP00#083	96	gr 4	gr 6	t380	ex
1520040	59	gr 5	gr 7	t163	316
D551	59	gr 5	gr 7	t199	316
D535	59	gr 5	gr 7	t216	316
UK-2-4820	359	gr 5	gr 7	t316	316
D115	123	gr 6	121	t171	159
021spa816	1024	gr 6	121	t408	159
D22	182	gr 7	gr 8	t196	gr 5
CC400	182	gr 7	gr 8	t364	gr 5
spa110	611	gr 7	gr 8	t364	gr 5
H19	10	gr 8	gr 9	t198	sg 14
D139	145	gr 8	gr 9	t218	sg 18
016m118F	10	gr 8	gr 9	t240	166/240
D302	97	gr 9	gr 3	t224	044

VI

NL011044-2	97	gr 9	gr 3	t231	044
2102032003	97	gr 9	gr 3	t267	044
spa114	97	gr 9	gr 3	t359	044
CC133	1026	gr 9	gr 3	t225	sg 19
H148	3	sg 1	1	t177	127
021spa771	3	sg 1	1	t398	ex
C56	6	sg 2	5	t207	008
C2	7	sg 3	gr 5	t091	084
016m194	7	sg 3	gr 5	t250	084
id3812004	7	sg 3	gr 5	t289	084
H42	20	sg 4	sg 1	t164	164
D17	20	sg 4	sg 1	t195	164
spa99	20	sg 4	sg 1	t357	164
C427	42	sg 5	30	t166	166/240
C316	49	sg 6	sg 2	t208	sg 16
H417	50	sg 7	sg 3	t185	sg 13
D97	55	sg 8	sg 4	t021	012
008spastapaur51	72	sg 9	sg 5	t126	324
VA136182003	72	sg 9	sg 5	t148	324
BAP03#020	72	sg 9	sg 5	t324	324
spa83	72	sg 9	sg 5	t347	sg 32
id0182002	80	sg 10	sg 6	t042	044
SwedenN8-890/99	80	sg 10	sg 6	t044	044
id1942003	80	sg 10	sg 6	t131	044
M762004	80	sg 10	sg 6	t376	044
UK-2-6416	80	sg 10	sg 6	t416	ex
016m072	88	sg 11	gr 2	t086	186
BK1563	88	sg 11	gr 2	t186	186
HT2001634	93	sg 12	sg 7	t202	sg 15
BAS97#479R	130	sg 13	sg 8	t373	sg 36
M7	152	sg 14	sg 9	t355	sg 33
ST157-MRSA-III	157	sg 15	239	t030	012
Slovenia20	222	sg 16	5	t001	002
ST231-MRSA-II	231	sg 17	5	t002	002
99.3759V	235	sg 18	sg 10	t052	008
NL013100-5	281	sg 19	30	t005	005
CAN6820-0616	289	sg 20	sg 11	t050	004
Btn2164	312	sg 21	sg 12	t229	sg 21
UK-2-3662	361	sg 22	sg 13	t315	002
Not290	529	sg 23	5	t222	ex
spa94	580	sg 24	gr 1	t356	sg 34
id3322004	599	sg 25	121	t278	sg 27
id3912004	600	sg 26	239	t292	008
id3992004	601	sg 27	gr 1	t293	sg 28
015m132	603	sg 28	30	t117	sg 5
M66	616	sg 29	22	t354	gr 1
id1972003	618	sg 30	gr 2	t237	044
016m104	619	sg 31	239	t118	ex
BAS94#035	625	sg 32	45	t302	004
BAP03#189	629	sg 33	gr 2	t325	186
021spa768	1004	sg 34	1	t409	sg 39
021spa730	1027	sg 35	sg 14	t392	sg 38
CC161	1029	sg 36	gr 3	t203	044
CC271	1030	sg 37	gr 5	t236	084
CC262	1031	sg 38	sg 15	t236	084

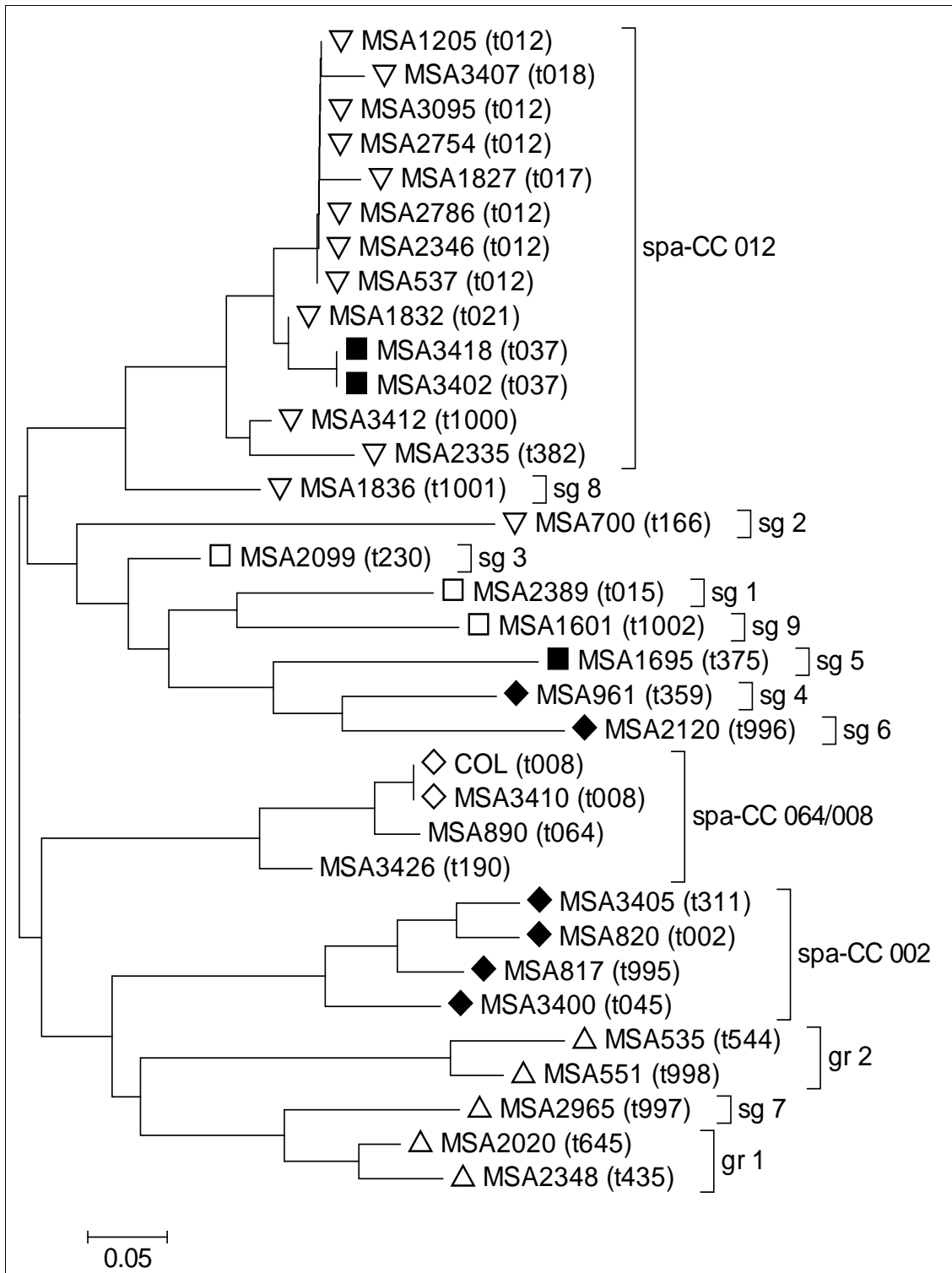
Datensatz der in Münster charakterisierten Isolate ($n = 447$; 148 ST; 400 *spa* Typen). Die Isolate sind primär nach eBURST 6/7 und sekundär nach *spa* Typen sortiert. gr: Gruppe ohne Angabe eines Gründungsvaters. sg: Einzeltyp. ^a nach MLST, ST über 1000 sind provisorische Sequenztypen für alleliche Profile, die zum Zeitpunkt der Untersuchung (Datum 03.10.2005) nicht Teil der MLST Datenbank waren. ^b Klassifikation durch eBURSTv2 separat für diesen Datensatz (online via <http://eburst.mlst.net/>, Datum 03.10.2005). ^c vor-eingestellte Gruppeneffinition. ^d erweiterte Gruppeneffinition. ^e BURP-Parameter: Ausschluss von *spa* Typen kürzer als fünf Repeats, Kosten kleiner oder gleich vier, dadurch Ausschluss von 31 *spa* Typen (ex) sowie Bildung von 24 Gruppen und 40 Einzeltypen.

Anhang B: Datensatz Koreen

Isolat	Microarray ^a			MLEE ^b		PFGE ^b	coa ^b			spa				
	0.81	0.5	0.31	Lin	Typ		Lin	Typ	Repeatorprofil	Lin ^b	Typ ^b	Repeatorprofil ^b	Typ ^c	BURP ^d
COL	sg 1	sg 1	gr 1	n.b.	n.b.	B1	1	3	MNOPQ	1	1	YHGFMBQBLO	t008	064/008
MSA3410	sg 2	sg 2	gr 1	F2	93	B2	1	3	MNOPQ	1	1	YHGFMBQBLO	t008	064/008
MSA890	sg 3	sg 3	sg 1	F2	93	1	1	20	MQ2OPQ	1	7	YHGCMBQBLO	t064	064/008
MSA3426	sg 4	sg 4	sg 2	F2	93	1	1	3	MNOPQ	1	4	YMBQBLO	t190	064/008
MSA817	sg 5	sg 5	gr 2	F4	114	1	2	8	USVDW	2	295	UJMBMDMEMK	t995	002
MSA961	sg 6	sg 6	gr 2	F4	147	1	3	19	UN2O2P2DNF	3	92	UJGFMBBPB	t359	sg 4
MSA820	sg 7	gr 1	gr 2	F2	91	G1	4	1	ABCDEF	2	2	T00000JMBMDMGMK	t002	002
MSA3400	sg 8	gr 1	gr 2	F2	91	G2	4	21	ARCDEF	2	47	TMDMGMK	t045	002
MSA3405	sg 9	gr 1	gr 2	F2	91	G1	4	6	ABSDEF	2	45	TJMBDMGMK	t311	002
MSA2120	sg 10	sg 7	gr 2	F4	146	1	5	29	S3ST3EF	4	294	UG2MFBBLB	t996	sg 6
RF122	sg 11	sg 8	gr 3	F10	195	1	6	26	K3L3M3N3O3	5	102	ZB	t529	ex
MSA2965	sg 12	gr 2	gr 3	F9	191	C1	7A	31	S2T2U2V2W2X2Y2D2	6	89	I2Z2EGMMJHM	t997	sg 7
MSA2348	sg 13	gr 2	gr 3	F9	189	C2	7A	27	S2T2U2V2W2X2D3Y2D2	6	293	I2Z2EMMMJH2M	t435	gr 1
MSA2020	sg 14	gr 2	gr 3	F9	189	C3	7A	31	S2T2U2V2W2X2Y2D2	6	79	I2Z2EGMJH2M	t645	gr 1
MSA535	sg 15	gr 3	gr 3	F1	66	E1	7B	30	R2T2U2Z2V3B3C3D2	7	292	D2KGFJMJJEMMMJQ	t544	gr 2
MSA551	sg 16	gr 3	gr 3	F1	70	E2	7B	30	R2T2U2Z2V3B3C3D2	7	291	D2GFMJJEMMMJQ	t998	gr 2
MSA2389	sg 17	gr 4	gr 4	D2	39	F1	8	10	YZA2B2C2D2	8	73	XKAKBEMKB	t015	sg 1
MSA1601	sg 18	gr 4	gr 4	D3	53	1	8	10	YZA2B2C2D2	8	299	XAKBBMBIB	t1002	sg 9
MSA2099	sg 19	gr 4	gr 4	D2	32	F1	8	10	YZA2B2C2D2	8	283	XKAKB	t230	sg 3
MSA3412	sg 20	gr 5	gr 5	H1	234	A3	9	25	GHIJKJ3	9	297	WGKKAKAOQQ	t1000	012
MSA3407	sg 21	gr 5	gr 5	H1	234	A4	9	22	GD3KL	9	16	WGKAKAOMQQQ	t018	012
MSA2885	sg 22	gr 5	gr 5	H1	234	A8	9	2	GHIJKL	9	298	WGM	t870	ex
MSA2335	sg 23	gr 5	gr 5	H1	234	A7	9	28	U3HIJKL	9	281	WGKKAMQQ	t382	012
MSA2754	gr 1	gr 6	gr 5	H1	234	A2	9	22	GD3KL	9	33	WGKAKAOMQQ	t012	012
MSA2345	gr 1	gr 6	gr 5	H1	234	n.b.	n.b.	n.b.	n.b.	n.b.	n.b.	n.b.	n.b.	n.b.
MSA1836	sg 24	gr 6	gr 5	H1	234	A6	9	2	GHIJKL	9	300	XKKAKAOM	t1001	sg 8
MSA1827	sg 25	gr 6	gr 5	H1	234	A1	9	23	GHE3JKL	9	296	WGKKAKAOMQQ	t017	012
MSA700	sg 26	gr 6	gr 5	H1	234	A5	9	22	GD3KL	10	301	ZZ2PNGKBKGBLB	t166	sg 2
MSA2786	sg 27	gr 6	gr 5	H1	234	A2	9	22	GD3KL	9	33	WGKAKAOMQQ	t012	012
MSA3095	sg 28	gr 6	gr 5	H1	234	A2	9	22	GD3KL	9	33	WGKAKAOMQQ	t012	012
MSA2346	sg 29	gr 6	gr 5	H1	234	A2	9	22	GD3KL	9	33	WGKAKAOMQQ	t012	012
MSA1205	sg 30	gr 6	gr 5	H1	234	A2	9	22	GD3KL	9	33	WGKAKAOMQQ	t012	012
MSA1832	sg 31	gr 6	gr 5	H1	234	A2	9	2	GHIJKL	9	43	WGKAKAOMQ	t021	012
MSA537	sg 32	gr 6	gr 5	H1	234	A2	9	22	GD3KL	9	33	WGKAKAOMQQ	t012	012
MSA3418	sg 33	gr 7	gr 6	F2	89	H1	9	2	GHIJKL	9	3	WGKAOMQ	t037	012
MSA3402	sg 34	gr 7	gr 6	F2	89	H2	9	2	GHIJKL	9	3	WGKAOMQ	t037	012
MSA1695	sg 35	sg 9	gr 6	F2	93	1	10	24	E2E3F3G3H3I3	11	302	Y2EJCMBPB	t375	sg 5

Datensatz modifiziert nach Koreen et al. [23]. Lin: Gruppierung („lineage“). gr: Gruppe. sg: Einzeltyp. ^a Gruppierung nach dem Dendrogramm des Microarrays bei den jeweiligen Pearson-Korrelationskoeffizienten. ^b übernommen aus der Originalpublikation. ^c nach Ridom SpaServer (<http://www.ridom.de>, Datum 07.10.2005). ^d spa-CC nach BURP (Ridom StaphType 1.2): Ausschluss von spa Typen kürzer als fünf Repeats, Kosten kleiner oder gleich fünf, dadurch dadurch Ausschluss von zwei spa Typen (ex) sowie Bildung von fünf Gruppen und neun Einzeltypen. Falls von BURP keine Gründungsväter definiert werden, wird die Zugehörigkeit zu einer Gruppe oder die Eigenschaft als Einzeltyp angegeben.

Anhang C: Dendrogramm, Datensatz Koreaen



Dendrogramm des Datensatzes Koreaen (nach Koreen et al. [23]) nach der *spa* Einteilung mit dem Programm MEGA. Eckige Klammern: Gruppierung nach BURP (Parameter: Ausschluss von *spa* Typen kürzer als fünf Repeats, Kosten kleiner oder gleich fünf, dadurch Ausschluss von zwei *spa* Typen sowie Bildung von fünf Gruppen und neun Einzeltypen). Runde Klammern: *spa* Typ. Symbole vor den Isolatnamen: Gruppierung nach dem Microarray bei einem Pearson-Korrelationskoeffizienten von 0.31 (gr 1 ◇, gr 2 ◆, gr 3 △, gr 4 □, gr 5 ▽; gr 6 ■; Einzeltypen nicht gekennzeichnet). Es zeigt sich eine weitgehende Korrelation der Einteilung durch den BURP-Algorithmus mit derjenigen durch den Microarray.

Anhang D: Datensatz Robinson

Isolat	eBURST ^a			spa		
	Profil ^b	6/7 ^c	5/7 ^d	Profil ^e	Typ ^f	BURP ^g
ST5-MRSA-I (Fra ^h)	1-4-1-4-12-1-10	5	5	TIMBMDMGMK	t001	010/002
ST5-MRSA-I (UK 1 ⁱ)	1-4-1-4-12-1-10	5	5	TIMBMDMGMK	t001	010/002
ST5-MRSA-I (UK 2)	1-4-1-4-12-1-10	5	5	TIMBMDMGMK	t001	010/002
ST5-MRSA-II (Fin)	1-4-1-4-12-1-10	5	5	TJMBMDMGMK	t002	010/002
ST5-MRSA-II (Jap)	1-4-1-4-12-1-10	5	5	TJMBMDMGMK	t002	010/002
ST5-MRSA-II (UK 1)	1-4-1-4-12-1-10	5	5	TJMBMDMGMK	t002	010/002
ST5-MRSA-II (UK 2)	1-4-1-4-12-1-10	5	5	TJMBMDMGMK	t002	010/002
ST5-MSSA (Den)	1-4-1-4-12-1-10	5	5	TJMBMDMGMK	t002	010/002
ST5-MSSA (Net)	1-4-1-4-12-1-10	5	5	TJMBMDMGMK	t002	010/002
ST5-MSSA (UK 1)	1-4-1-4-12-1-10	5	5	TJMBMDMGMK	t002	010/002
ST5-MSSA (UK 2)	1-4-1-4-12-1-10	5	5	TJMBMDMGMK	t002	010/002
ST5-MSSA (UK 3)	1-4-1-4-12-1-10	5	5	TJMBMDMGMK	t002	010/002
ST5-MSSA (UK 4)	1-4-1-4-12-1-10	5	5	TJMBMDMGMK	t002	010/002
ST5-MRSA-II (USA)	1-4-1-4-12-1-10	5	5	TMBMDMGMK	t010	010/002
ST5-MRSA-II (Ire)	1-4-1-4-12-1-10	5	5	TMDMGMK	t045	010/002
ST5-MRSA-III (Bel)	1-4-1-4-12-1-10	5	5	TMDMGMK	t045	010/002
ST225-MRSA-II (USA)	1-4-1-4-12-25-10	5	5	TMDMGMK	t045	010/002
ST554-MRSA-II (Ire)	1-2-1-4-12-1-10	5	5	TMDMGMK	t045	010/002
ST5-MRSA-II (USA)	1-4-1-4-12-1-10	5	5	TJMGMK	t062	010/002
ST5-MRSA-I (Slo)	1-4-1-4-12-1-10	5	5	TIMBMDMGMMMK	t178	010/002
ST5-MRSA-IV (USA)	1-4-1-4-12-1-10	5	5	TJMBMDMGMKK	t214	010/002
ST30-MRSA-IV (Ger 1)	2-2-2-2-6-3-2	36	36	WGKAKAOMQQQ	t018	037/021
ST30-MRSA-IV (Ger 2)	2-2-2-2-6-3-2	36	36	WGKAKAOMQQQ	t018	037/021
ST36-MRSA-II (Fin)	2-2-2-2-3-3-2	36	36	WGKAKAOMQQQ	t018	037/021
ST36-MRSA-II (UK 1)	2-2-2-2-3-3-2	36	36	WGKAKAOMQQQ	t018	037/021
ST36-MRSA-II (UK 2)	2-2-2-2-3-3-2	36	36	WGKAKAOMQQQ	t018	037/021
ST36-MRSA-II (UK 3)	2-2-2-2-3-3-2	36	36	WGKAKAOMQQQ	t018	037/021
ST36-MRSA-II (UK 4)	2-2-2-2-3-3-2	36	36	WGKAKAOMQQQ	t018	037/021
ST36-MRSA-II (UK 5)	2-2-2-2-3-3-2	36	36	WGKAKAOMQQQ	t018	037/021
ST36-MRSA-II (UK 6)	2-2-2-2-3-3-2	36	36	WGKAKAOMQQQ	t018	037/021
ST38-MRSA-II (UK)	2-2-10-2-3-3-2	36	36	WGKAKAOMQQQ	t018	037/021
ST30-MRSA-IV (Swe)	2-2-2-2-6-3-2	36	36	WGKAKAOMQ	t021	037/021
ST30-MSSA (Den)	2-2-2-2-6-3-2	36	36	WGKAKAOMQ	t021	037/021
ST30-MSSA (UK)	2-2-2-2-6-3-2	36	36	WGKAKAOMQ	t021	037/021
ST2001-MRSA-I (UK)	3-3-6-1-4-4-16	239	239	YHGFMBQBLO	t008	008
ST250-MRSA-I (Den)	3-3-1-1-4-4-16	239	239	YHGFMBQBLO	t008	008
ST250-MRSA-I (Swe)	3-3-1-1-4-4-16	239	239	YHGFMBQBLO	t008	008
ST250-MRSA-I (Uga)	3-3-1-1-4-4-16	239	239	YHGFMBQBLO	t008	008
ST250-MSSA (Den)	3-3-1-1-4-4-16	239	239	YHGFMBQBLO	t008	008
ST254-MRSA-I (UK 1)	3-32-1-1-4-4-3	239	239	YHGFMBQBLO	t008	008
ST254-MRSA-I (UK 2)	3-32-1-1-4-4-3	239	239	YHGFMBQBLO	t008	008
ST8-MRSA-II (USA)	3-3-1-1-4-4-3	239	239	YHGFMBQBLO	t008	008
ST8-MRSA-IV (Fra)	3-3-1-1-4-4-3	239	239	YHGFMBQBLO	t008	008
ST8-MSSA (Den 1)	3-3-1-1-4-4-3	239	239	YHGFMBQBLO	t008	008
ST8-MSSA (Den 2)	3-3-1-1-4-4-3	239	239	YHGFMBQBLO	t008	008
ST8-MSSA (UK 1)	3-3-1-1-4-4-3	239	239	YHGFMBQBLO	t008	008
ST8-MSSA (UK 2)	3-3-1-1-4-4-3	239	239	YHGFMBQBLO	t008	008
ST254-MRSA-IV (Ger)	3-32-1-1-4-4-3	239	239	YGFMBQBLQBLPO	t009	008
ST254-MRSA-IV (UK)	3-32-1-1-4-4-3	239	239	YGFMBQBLQBLPO	t009	008
ST239-MRSA-III (Fin)*	2-3-1-1-4-4-3	239	239	WGKAQQ	t030	037/021
ST239-MRSA-III (Ger)*	2-3-1-1-4-4-3	239	239	WGKAQQ	t030	037/021
ST239-MRSA-III (Net)*	2-3-1-1-4-4-3	239	239	WGKAQQ	t030	037/021
ST246-MRSA-III (Swe)*	2-3-1-12-4-4-3	239	239	WGKAQQ	t030	037/021
ST239-MRSA-III (Ger)*	2-3-1-1-4-4-3	239	239	WGKAOMQ	t037	037/021
ST239-MRSA-III (Pol 1)*	2-3-1-1-4-4-3	239	239	WGKAOMQ	t037	037/021
ST239-MRSA-III (Pol 2)*	2-3-1-1-4-4-3	239	239	WGKAOMQ	t037	037/021
ST239-MRSA-III (Por 1)*	2-3-1-1-4-4-3	239	239	WGKAOMQ	t037	037/021
ST239-MRSA-III (Por 2)*	2-3-1-1-4-4-3	239	239	WGKAOMQ	t037	037/021
ST239-MRSA-III (UK 1)*	2-3-1-1-4-4-3	239	239	WGKAOMQ	t037	037/021
ST239-MRSA-III (UK 2)*	2-3-1-1-4-4-3	239	239	WGKAOMQ	t037	037/021
ST239-MRSA-III (UK 3)*	2-3-1-1-4-4-3	239	239	WGKAOMQ	t037	037/021

ST239-MRSA-III (UK 4)*	2-3-1-1-4-4-3	239	239	WGKAOMQ	t037	037/021
ST240-MRSA-III (UK)*	2-3-1-1-21-4-3	239	239	WGKAOMQ	t037	037/021
ST241-MRSA-III (Fin)*	2-3-1-1-4-4-30	239	239	WGKAOMQ	t037	037/021
ST158-MRSA-III (Ire)*	2-3-1-1-4-4-2	239	239	WGKAOMQ	t037	037/021
ST247-MRSA-I (Por)	3-3-1-12-4-4-16	239	239	YHFGFMBQBLO	t051	008
ST247-MRSA-I (UK)	3-3-1-12-4-4-16	239	239	YHFGFMBQBLO	t051	008
ST247-MRSA-I (Bel)	3-3-1-12-4-4-16	239	239	YHFFMBQBLO	t054	008
ST8-MRSA-II (Ire)	3-3-1-1-4-4-3	239	239	YMBQBLO	t190	008
ST8-MRSA-II (UK 3)	3-3-1-1-4-4-3	239	239	YMBQBLO	t190	008
ST8-MRSA-II (UK 4)	3-3-1-1-4-4-3	239	239	YMBQBLO	t190	008
ST8-MRSA-IV (Aus)	3-3-1-1-4-4-3	239	239	YMBQBLO	t190	008
ST8-MRSA-IV (Ire 1)	3-3-1-1-4-4-3	239	239	YMBQBLO	t190	008
ST8-MRSA-IV (Ire 2)	3-3-1-1-4-4-3	239	239	YMBQBLO	t190	008
ST8-MRSA-IV (UK 1)	3-3-1-1-4-4-3	239	239	YMBQBLO	t190	008
ST8-MRSA-IV (UK 2)	3-3-1-1-4-4-3	239	239	YMBQBLO	t190	008
ST250-MRSA-I (UK)	3-3-1-1-4-4-16	239	239	YHGGGBBQBLO	t194	008
ST22-MRSA-IV (Ire)	7-6-1-5-8-8-6	sg 1	sg 1	TJEJNCMOMOKR	t005	022
ST22-MSSA (Net)	7-6-1-5-8-8-6	sg 1	sg 1	TJEJNCMOMOKR	t005	022
ST22-MSSA (UK 1)	7-6-1-5-8-8-6	sg 1	sg 1	TJEJNCMOMOKR	t005	022
ST22-MSSA (UK 2)	7-6-1-5-8-8-6	sg 1	sg 1	TJEJNCMOMOKR	t005	022
ST22-MSSA (UK 3)	7-6-1-5-8-8-6	sg 1	sg 1	TJEJNCMOMOKR	t005	022
ST22-MRSA-IV (UK 1)	7-6-1-5-8-8-6	sg 1	sg 1	TJEJNI2MNI2MOMOKR	t022	022
ST22-MRSA-IV (UK 2)	7-6-1-5-8-8-6	sg 1	sg 1	TJEJNI2MNI2MOMOKR	t022	022
ST22-MRSA-IV (UK 3)	7-6-1-5-8-8-6	sg 1	sg 1	TJEJNI2MNI2MOMOKR	t022	022
ST22-MRSA-IV (Ger)	7-6-1-5-8-8-6	sg 1	sg 1	TJJEJNI2MNI2MOMOKR	t032	022
ST22-MRSA-IV (Ire)	7-6-1-5-8-8-6	sg 1	sg 1	TJJEJNI2MNI2MOMOKR	t032	022
ST22-MRSA-IV (Swe)	7-6-1-5-8-8-6	sg 1	sg 1	TJJEJNI2MNI2MOMOKR	t032	022
ST22-MRSA-IV (UK 1)	7-6-1-5-8-8-6	sg 1	sg 1	TJJEJNI2MNI2MOMOKR	t032	022
ST22-MRSA-IV (UK 2)	7-6-1-5-8-8-6	sg 1	sg 1	TJJEJNI2MNI2MOMOKR	t032	022
ST22-MRSA-IV (UK 3)	7-6-1-5-8-8-6	sg 1	sg 1	TJJEJNI2MNI2MOMOKR	t032	022
ST22-MRSA-IV (UK 4)	7-6-1-5-8-8-6	sg 1	sg 1	TJJEJNI2MNI2MOMOKR	t032	022
ST45-MRSA-IV (Ger)	10-14-8-6-10-3-2	sg 2	sg 2	A2AKEEMBKB	t004	065
ST45-MRSA-IV (Fin)	10-14-8-6-10-3-2	sg 2	sg 2	XKAKBEMBKB	t015	065
ST45-MSSA (Cub)	10-14-8-6-10-3-2	sg 2	sg 2	XKAKBEMBKB	t015	065
ST45-MSSA (UK 1)	10-14-8-6-10-3-2	sg 2	sg 2	XKAKBEMBKB	t015	065
ST45-MSSA (UK 2)	10-14-8-6-10-3-2	sg 2	sg 2	XKAKBEMBKB	t015	065
ST45-MRSA-IV (Bel)	10-14-8-6-10-3-2	sg 2	sg 2	XE3 ^k BBEMBKB	t038	065
ST45-MRSA-IV (Swe)	10-14-8-6-10-3-2	sg 2	sg 2	XKAKBBMBKB	t050	065
ST45-MSSA (Ger)	10-14-8-6-10-3-2	sg 2	sg 2	A2AKBEMBKB	t065	065
ST45-MRSA-II (USA)	10-14-8-6-10-3-2	sg 2	sg 2	A2AMEEMBKB	t204	065
ST222-MRSA-I (Slo)	1-4-1-1-4-24-29	sg 3	5	TIMBMDMGMK	t001	010/002
ST228-MRSA-I (Ger 1)	1-4-1-4-12-24-29	sg 4	5	TIMBMDMGMK	t001	010/002
ST228-MRSA-I (Ger 2)	1-4-1-4-12-24-29	sg 4	5	TIMBMDMGMK	t001	010/002
ST231-MRSA-II (USA)	1-34-1-4-12-1-28	sg 5	5	TJMBMDMGMK	t002	010/002
ST235-MRSA-I (UK)	1-4-1-12-12-4-16	sg 6	sg 3	YFGFMBQBLO	t052	008
ST157-MRSA-III (Pol)*	2-3-26-1-4-39-3	sg 7	239	WGKAQQ	t030	037/021

Isolate ($n = 107$; 23 ST; 27 *spa* Typen) aus einer Publikation von Robinson et al. [33], der dortigen Nomenklatur folgend. Aus den allelischen Profilen resultieren entsprechend der MLST Datenbank (<http://www.mlst.net>, Datum 12.09.2005) vier Änderungen gegenüber den in der dortigen Grafik notierten Sequenztypen. Aus ST305, ST308 bzw. ST313 werden hier ST157, ST554 bzw. ST158, das allelische Profil für ST257 war in der MLST Datenbank nicht eingetragen, so dass hier der ST2001 definiert wird. Die Isolate sind primär nach eBURST 6/7 und sekundär nach *spa* Typen sortiert. gr: Gruppe ohne Angabe eines Gründungsvaters. sg: Einzeltyp. ^a Klassifikation durch eBURSTv2 separat für diesen Datensatz (online via MLST website, Datum 18.09.2005). ^b allelisches Profil in der Reihenfolge: *arcC*, *aroE*, *glpF*, *gmk*, *pta*, *tpi*, *yqiL*. ^c voreingestellte Gruppendifinition. ^d erweiterte Gruppendifinition. ^e Nomenklatur von Shopsin et al. [39]. ^f Übersetzung der Profile via Ridom SpaServer (<http://www.ridom.de>). ^g *spa* CC nach BURP: Ausschluss von *spa* Typen kürzer als fünf, Kosten kleiner oder gleich sieben, dadurch kein Ausschluss eines *spa* Typs sowie Bildung von fünf Gruppen ohne Einzeltypen. ^h Herkunftsland. ⁱ laufende Nummer. ^k persönliche Korrespondenz mit dem Autor der Publikation [33]. * Isolat gehört zu der Auswahl ohne ST239 und die von ihm abgeleiteten Sequenztypen.

Anhang E: Datensatz Nottingham

Isolat	ST ^a	eBURST ^b		spa Typ ^e	BURP ^f
		6/7 ^c	5/7 ^d		
001	15	15	15	t084	346
437	15	15	15	t084	346
628	15	15	15	t084	346
245	1025	15	15	t084	346
011	1021	15	15	t094	346
102	15	15	15	t346	346
042	15	15	15	t385	346
082	15	15	15	t393	346
126	30	30	30	t012	399
129	30	30	30	t012	399
371	30	30	30	t012	399
433	30	30	30	t012	399
614	30	30	30	t012	399
620	30	30	30	t012	399
147	30	30	30	t021	399
181	30	30	30	t021	399
426	30	30	30	t021	399
125	34	30	30	t166	gr 2
384	34	30	30	t166	gr 2
611	34	30	30	t166	gr 2
335	30	30	30	t233	ex
425	30	30	30	t318	399
436	378	30	30	t402	399
370	30	30	30	t404	399
396	30	30	30	t406	399
405	34	30	30	t410	gr 2
156	37	30	30	t413	399
260	30	30	30	t429	399
242	45	45	45	t015	383/230
394	45	45	45	t015	383/230
087	47	45	45	t026	ex
444	45	45	45	t230	383/230
299	46	45	45	t362	ex
009	1008	45	45	t383	383/230
183	1009	45	45	t397	sg 7
160	22	217	217	t005	005
390	22	217	217	t005	005
136	217	217	217	t005	005
401	22	217	217	t022	005
445	22	217	217	t032	005
342	22	217	217	t107	005
443	217	217	217	t192	005
428	22	217	217	t223	005
346	22	217	217	t310	005
103	1018	217	217	t395	005
465	22	217	217	t403	005
279	1	gr 1	gr 1	t127	127
356	1	gr 1	gr 1	t127	127
007	1001	gr 1	gr 1	t127	127
382	1	gr 1	gr 1	t405	127
024	8	gr 2	gr 3	t008	sg 1
085	8	gr 2	gr 3	t008	sg 1
110	8	gr 2	gr 3	t008	sg 1
315	8	gr 2	gr 3	t008	sg 1
316	8	gr 2	gr 3	t008	sg 1
375	8	gr 2	gr 3	t008	sg 1
442	8	gr 2	gr 3	t008	sg 1
070	1014	gr 2	gr 3	t008	sg 1
121	8	gr 2	gr 3	t394	399
117	27	gr 3	gr 6	t100	gr 1

XII

167	9	gr 3	gr 6	t411	gr 1
021	25	gr 4	gr 2	t078	401
023	25	gr 4	gr 2	t078	401
088	25	gr 4	gr 2	t078	401
145	25	gr 4	gr 2	t078	401
422	25	gr 4	gr 2	t078	401
441	25	gr 4	gr 2	t078	401
451	25	gr 4	gr 2	t078	401
603	25	gr 4	gr 2	t078	401
320	25	gr 4	gr 2	t167	401
057	1017	gr 4	gr 2	t287	ex
114	1017	gr 4	gr 2	t396	401
600	25	gr 4	gr 2	t401	401
328	1005	gr 5	30	t096	399
036	39	gr 5	30	t298	399
015	39	gr 5	30	t382	399
081	39	gr 5	30	t382	399
063	39	gr 5	30	t384	399
144	39	gr 5	30	t384	399
309	39	gr 5	30	t399	399
462	39	gr 5	30	t407	399
038	101	gr 6	gr 4	t056	401
266	505	gr 6	gr 4	t232	ex
139	188	gr 7	gr 1	t189	127
404	188	gr 7	gr 1	t189	127
440	188	gr 7	gr 1	t189	127
464	188	gr 7	gr 1	t189	127
380	1003	gr 7	gr 1	t189	127
403	3	sg 1	gr 1	t177	127
314	3	sg 1	gr 1	t398	ex
056	5	sg 2	sg 1	t002	gr 3
205	5	sg 2	sg 1	t002	gr 3
226	5	sg 2	sg 1	t002	gr 3
282	5	sg 2	sg 1	t002	gr 3
289	5	sg 2	sg 1	t002	gr 3
362	5	sg 2	sg 1	t002	gr 3
607	5	sg 2	sg 1	t002	gr 3
378	5	sg 2	sg 1	t088	gr 3
050	7	sg 3	gr 5	t091	346
151	12	sg 4	sg 2	t160	sg 2
029	20	sg 5	sg 3	t195	sg 3
272	59	sg 6	sg 4	t216	sg 4
400	182	sg 7	sg 5	t364	sg 5
290	1004	sg 8	gr 1	t409	sg 9
622	1012	sg 9	sg 7	t412	383/230
481	1024	sg 10	sg 8	t408	sg 8
133	1026	sg 11	sg 10	t225	127
093	1027	sg 12	sg 6	t392	sg 6
161	1029	sg 13	sg 9	t203	127
271	1030	sg 14	gr 5	t236	346
262	1031	sg 15	sg 11	t236	346

Isolate ($n = 111$; 44 ST; 67 *spa* Typen) aus einer Publikation von Grundmann et al. [13]. Die Isolate sind primär nach eBURST 6/7 und sekundär nach *spa* Typen sortiert. ^a MLST Sequenztypen wurden zugeordnet nach Korrespondenz mit dem Autor, ST über 1000 waren nicht Teil der MLST Datenbank (<http://www.mlst.net>, Datum 18.09.2005). Eine Kontrolle (<http://www.mlst.net>, Datum 18.09.2005) der in persönlicher Korrespondenz zur Verfügung gestellten allelischen Profile hatte die Änderung des Sequenztyps einiger Isolate zur Folge (Nummer 117 zu ST27, Nummer 136 zu ST217, Nummer 266 zu ST505, Nummer 272 zu ST59, Nummer 436 zu ST378, Nummer 443 zu ST217). ^b eBURST wurde separat für diesen Datensatz durchgeführt mit eBURSTv2 (online via MLST website, Datum 18.09.2005), Gründungsväter sind notiert, falls von dem Programm angegeben, ansonsten sind Gruppen (gr) oder Einzeltypen (sg) angegeben. ^c voreingestellte Gruppendifinition. ^d erweiterte Gruppendifinition. ^e die *spa* Typisierung war in Münster durchgeführt worden. ^f BURP-Parameter: Ausschluss von *spa* Typen kürzer als fünf Repeats, Kosten kleiner oder gleich sieben, dadurch Ausschluss von sechs *spa* Typen sowie Bildung von neun Gruppen und neun Einzeltypen. Gründungsväter sind notiert, falls von dem Programm angegeben, ansonsten sind Gruppen (gr) oder Einzeltypen (sg) angegeben.