Aus dem Bereich Theoretische Medizin und Biowissenschaften

Fachrichtung Medizinische Mikrobiologie und Hygiene

der Medizinischen Fakultät der

Universität des Saarlandes, Homburg/Saar

Genotypic and mass spectrometric comparative analysis of African and German *Staphylococcus aureus* isolates

DISSERTATION ZUR ERLANGUNG DES GRADES EINES DOKTORS DER NATURWISSENSCHAFTEN DER MEDIZINISCHEN FAKULTÄT DER UNIVERSITÄT DES SAARLANDES

2019

vorgelegt von

Ulla Ruffing

geboren am 13.09.1984 in Neunkirchen

Die vorliegende Arbeit wurde im Zeitraum von Februar 2011 bis Februar 2016 im Arbeitskreis von Herrn Prof. Dr. Mathias Herrmann im Institut für Medizinische Mikrobiologie und Hygiene der Universität des Saarlandes, Standort Universitätsklinikum des Saarlandes in Homburg/Saar, angefertigt.

Teile dieser Arbeit wurden bereits in Originalarbeiten, Vorträgen und Postern veröffentlicht, siehe Abschnitt 8 – Publikationen und Kongressbeiträge (S.177-180).

List of abbreviations

A. dest.	aqua destillatum
agr	accessory gene regulator
AFLP	amplified fragment length polymorphism
AIDS	acquired immune deficiency syndrome
APS	admission prevalence screening
B	blood culture
bp	base pairs
BURP	based upon repeat pattern
°C	degree celsius
CA	community-associated
CC	clonal complex
CHIPS	chemotaxis inhibitory protein
CRF	Case Report Form
DFG	Deutsche Forschungsgemeinschaft (German Society of Research)
DNA	desoxyribonucleid acid
d-UTP	2'-Deoxyuridin-5'-Triphosphat
edin	epidermal cell differentiation inhibitor
EDTA	ethylendiamintetraacetat
EEF	electronix excel file
Egc	enterotoxin gene cluster
ExoSAP	exonuclease I and Shrimp Alkaline Phosphatase
fig.	Figure
FR	Freiburg
g	gram
HA	healthcare-associated
HAC	hierarchical agglomerative clustering
H_2O	water
HIV	human immunodeficiency virus
HS	Homburg/Saar
IEC	immune evasion cluster
ISS	integration site sequence
IT	Ifakara/Tanzania
kDa	kilodalton
LA	livestock-associated
LG	Lambaréné/Gabon
LFM	latent factor model
MALDI-TOF	Matrix-Assisted Laser Desorption Ionization - Time of flight
MA	microarray
MLST	Multi-Locus-Sequence-Typing
MM	Manhiça/Mozambique
MMWR	morbidity and mortality weekly report

MRSA	Methicillin-resistant *Staphylococcus aureus*
MSCRAMMs	microbial surface component recognizing matrix molecules
MSSA	Methicillin-susceptible *Staphylococcus aureus*
MW	Münster/Westphalia
ng	nanogramm
N	nasal
O	wound infection
OD	optical density
PAK	Paketantrag (application package)
PBP	penicillin binding protein
PC	principal component
PCA	principal component analysis
PCR	polymerase-chain-reaction
PFGE	pulsed-field-gel-electrophoresis
PIA	intercellular polysaccharide adhesin
PNAG	poly-N-acetylglucosamin
PVL	Panton-Valentine leukocidin
RNase	ribonuclease
rpm	revolutions per minute
RT	room temperature
SA	*Staphylococcus aureus*
S. aureus	*Staphylococcus aureus*
sae	*S. aureus* exoprotein expression
sak	Staphylokinase
sar	staphylococcal accessory regulator
SCIN	staphylococcal complement inhibitor
SCC	staphylococcoal cassette chromosome
SERAM	secretable expanded repertoire adhesive molecules
SmaI	first restriction enzyme isolated from *Serratia marcescens*
spa	staphylococcus aureus protein A
SPSS	software package for statistical sciences
SSTI	skin and soft tissue infections
ST	Sequence Type
SVM	Support Vector Machine
tab	Table
TBE-Puffer	Tris-Borat-EDTA-Puffer
TSA	Trypticiase Soy Agar
tst1	encoding toxic shock syndrome toxin 1
UPGMA	unweighted pair group with arithmetic mean
V	volt
VISA	Vancomycin-intermediate resistant *Staphylococcus aureus*
VRSA	Vancomycin-resistant *Staphylococcus aureus*

WGS	whole genome sequencing
xg	x fold acceleration of gravity (9.82 m/s^2)
µl	microliter

Table of contents

Table of contents I
1 Summary 1
1.1 Summary 1
1.2 Zusammenfassung 3
2 Introduction 7
2.1 *Staphylococcus aureus* 7
2.1.1 Resistance of SA 8
2.1.2 Virulence of SA 10
2.2 SA epidemiology and the need of SA typing 14
2.2.1 The need of SA typing 14
2.2.2 SA typing methods 16
2.2.2.1 Phage typing 17
2.2.2.2 Amplified fragment length polymorphism (AFLP) 17
2.2.2.3 Pulsed-field gel electrophoresis (PFGE) 18
2.2.2.4 Staphylococcal cassette chromosome *mec* (SCC*mec*) - typing 19
2.2.2.5 SA protein A – typing (*spa*-typing) 19
2.2.2.6 Multi-locus-sequence-typing (MLST) 21
2.2.2.7 Whole genome sequencing (WGS) 22
2.2.2.8 DNA microarray (DNA MA) 23
2.2.2.9 Mass spectrometry using MALDI-TOF 24
2.2.3 SA epidemiology 26
2.2.3.1 Healthcare-, community- and livestock- associated SA 28
2.2.3.2 SA in Sub-Saharan Africa 30
2.3 Objectives of this thesis 33
3 Material and Methods 34
3.1 Material 34
3.1.1 Isolate collection 34
3.1.2 List of all analyzed SA isolates and DNA microarray data 36
3.1.3 Chemicals and consumables 37
3.1.3.1 Chemicals, media and consumables 37
3.1.3.2 Oligonucleotids 37
3.1.4 Instruments and Software 38
3.2 Methods 39
3.2.1 Culture, basic characterization and storage 39

3.2.2 DNA microarray 39
3.2.2.1 DNA-preparation 39
3.2.2.2 DNA concentration measurement 40
3.2.2.3 DNA microarray hybridization 40
3.2.2.4 DNA microarray data analysis by bioinformatic methods 42
3.2.2.4.1 Splits graph 42
3.2.2.4.2 Imputing ambiguous data points with Latent factor model (LFM) 46
3.2.2.4.3 Affinity propagation 48
3.2.2.4.4 Principal component analysis (PCA) 48
3.2.2.4.5 Hierarchical agglomerative clustering dendrogram (HAC) 49
3.2.2.4.6 Silhouette plot for clustered data 51
3.2.3 MALDI-TOF analysis 52
3.2.3.1 Culture and MALDI-TOF performance 52
3.2.3.2 Mass spectra analysis 52
3.2.3.2.1 Mass spectra preprocessing and PCA 52
3.2.3.2.2 Mass spectra identification projects with Support Vector Machine 53
3.2.3.2.3 Identification project or pairwise comparisons 54
3.2.3.2.4 Statistical analysis of peak intensities 54
3.2.4 *Spa*-typing 55
3.2.5 Multi locus sequence typing (MLST) 55
3.2.6 Whole genome sequencing for DNA MA concordance analysis and analysis of nucleotide sequences of three genes (*panD*, *rpmG,* and *rpmD*) 56
3.2.7 Statistical methods 56
3.2.7.1 Chi square test 56
3.2.7.2 Median of age 56
4 Results 57
4.1 Determination of clonal lineages and gene distribution of MRSA and matched MSSA by DNA MA 57
4.1.1 DNA MA and *spa*-typing data of stem collection 57
4.1.2 Cluster analysis of total stem collection 59
4.1.2.1 Splits graph of 46 MRSA and 46 MSSA based on *spa*-typing data ... 60
4.1.2.2 Splits graph of 46 MRSA and 46 MSSA based on DNA MA data 62
4.1.2.3 Comparison of the splits graph of 46 MRSA and 46 MSSA based on *spa*-typing and DNA MA data 63
4.1.3 Cluster analysis of 43 CC5 isolates by splits graphs based on *spa*-typing and DNA MA as well as by PCA and HAC based on DNA MA

..... 64

4.1.3.1 Splits graph of 43 CC5 isolates based on *spa*-typing data 64

4.1.3.2 Splits graphs of 43 CC5 isolates based on DNA MA data 65

4.1.3.3 Comparison of the splits graphs based on *spa*-typing and DNA MA data of 43 CC5 isolates 66

4.1.3.3.1 DNA MA data analysis of 43 CC5 MRSA/MSSA with PCA 66

4.1.3.3.2 DNA MA data analysis of 43 CC5 isolates with HAC and bioinformatic methods' data analysis comparison 67

4.1.3.3.3 Comparison of cluster analysis of 43 CC5 isolates with splits graph, PCA and HAC based on DNA MA data 69

4.1.4 Gene repertoire analysis 70

4.2 Characterization of Nigerian SA isolates by DNA MA analysis 74

4.3 Characterization and comparison of African and German SA isolates 77

4.3.1 Collection of African and German SA isolates 77

4.3.2 Clonal complexes and sequence types analyses 77

4.3.2.1 Analysis of clonal complexes with African or German SA origin 77

4.3.2.2 Association analysis of clonal complexes with clinical or nasal isolate origin 80

4.3.2.3 Antibiotic resistance and virulence-related gene presence as a function of geographical origin and clinical significance of SA isolates 83

4.3.3 Isolate cluster analysis 88

4.3.4 DNA MA and whole genome sequencing concordance analysis 90

4.4 Subtyping of SA isolates by MALDI-TOF 90

4.4.1 Evaluation of correct assignment of SA based on their total mass spectra to methicillin resistance, geographic origin, *spa*-type and strain lineage (CC/ST) 90

4.4.2 Analysis of potential CC121 specific total mass spectra peaks 95

5 Discussion 97

5.1 Differentiation of closely related SA strains of a hospital admission prevalence screening study in Saarland by DNA microarray 97

5.2 Characterization of Nigerian SA isolates by DNA MA 103

5.3 Genotypic characterization and comparison of Sub-Saharan African and German isolates 104

5.4 Potential of MALDI-TOF as strain differentiation method 110

5.5 Conclusion 112

6 Appendix 114

6.1 Manufacturer information of Alere Technologies to the DNA microarray targets 114

6.2 Case report forms 123
6.3 Characteristics of African and German healthy volunteers and patients 151
7 References 154
8 Publications / congress contributions 177
9 Acknowledgement 181

1 Summary

1.1 Summary

Staphylococcus aureus (SA) is a worldwide distributed, opportunistic pathogen colonizing 25-30% of the human population mostly without signs of infections. However, if infections (e.g. wound-, organ infections, sepsis) occurs, those are often associated with a significant morbidity and mortality. SA possesses a broad range of pathogenicity-, virulence- and resistance factors, whose presence is partially associated with specific phylogenetic developmental lines. Thus, the evaluation of transmission of infections is essentially based on phylogenetic analytics. This requires the development of reproducible, highly discriminatory typing methods for distinction of different SA strains. Thus, reproducible, polymerase chain reaction and sequencing based methods such as *spa* (SA protein A)-typing, multilocus sequencing typing (MLST) up to whole genome sequencing were developed. In contrast to *spa*-typing which is based on the sequencing of the single gene *spa*, MLST is based on the sequencing of seven housekeeping genes defining different sequence types (STs) which can be clustered to clonal complexes (CC). STs and CCs are today the grid for the international SA nomenclature in epidemiologic studies. DNA microarrays (MA) represents a special technique in which ST and CC are deduced from characteristic hybridization patterns of numerous genloci which are compared to a database of a large strain collection. In addition to the phylogenetic analysis the DNA MA hold a large number of locis for adhesion, virulence, toxin and resistance genes and thus provide a large amount of data. Comparison of multiple isolates investigated by DMA MA thus, requires supplementary bioinformatic methods for interpretation and presentation.

Overall, there are plenty of moleculargenetic based studies of infection epidemiology of SA. But, most of them referred to countries with a high level of microbiological diagnostic. For Africa there is only very limited epidemiological data on SA (inclusive methicillin-resistant SA [MRSA]) especially of non-nosocomial (this means not acquired in hospital) SA (e.g. community-associated (CA) methicillin-resistant or methicillin-susceptible SA [MRSA, MSSA]). In order to fill this knowledge gap, four studies were performed as basis for this doctoral thesis investigating a) different bioinformatic tools for analysis of the molecular epidemiology of SA, b) the epidemiology of African and German MSSA/MRSA originating outside healthcare

institutions and c) the applicability of mass spectrometry for analysis of SA epidemiology.

For characterization of the bioinformatic tools the subtyping potential of *spa*-typing versus DNA MA analysis were compared in a study with 46 nasal CA-MSSA and 46 CA-MRSA of infections, collected in the federal state of Saarland. Further, the results were evaluated by three bioinformatic methods (splits graph, hierarchical agglomerative clustering [HAC] and principal component analysis [PCA]). Finally, in order to test the applicability of the DNA MA analytics in the African cohort, a relatively small cohort of isolates of 52 Nigerian MSSA was analyzed by DNA MA and splits graph method.

In a big, multicentric, African-German, cross-sectional geographic comparative analysis, 1200 CA-SA isolates of healthy volunteers and patients of Sub-Saharan Africa and Germany were subsequently analyzed by DNA MA. In addition, association and comparative analyses of STs, CCs and gene composition of these isolates with respect to geographic (Germany or Africa) or clinical/nasal origin (nasal/asymptomatic or clinical/invasive) were performed. Finally, the potential applicability of identification of CCs or *spa*–types of CA-SA of CC5, CC8, CC15, CC30, CC45, CC121 and *spa*-types t002, t003, t504 based on their MALDI-TOF mass spectrometric profile, were also analyzed.

Expected results on the epidemiology of CA-MRSA/MSSA were firstly gained by the DNA MA study of the federal state of Saarland, showing a higher subtyping potential of the DNA MA compared to *spa*-typing. Evaluation of the DNA MA data by the three previously mentioned bioinformatic methods yielded a diverse assignment of CC5 isolates to 3 to 7 clusters with exception of 12 CC5 isolates that always clustered in two identical clusters. It does not make these methods a failure but the use of different mathematical algorithms analyzing hybridization pattern differences or similarities may lead to minor differences. Further, MSSA displayed a high degree of diversity in comparison to the CA-MRSA group, which was dominated by the CC5 Rhine-Hesse epidemic strain (89%).

The further analysis of African MSSA of the study with Nigerian MSSA identified a heterogeneous and divergent nature of the examined Nigerian MSSA and confirmed that the DNA MA is a suitable genotyping method of African SA. The results of the multicentric consortium study showed a total of 40 different CCs and STs, including 5 new STs. The majority, 15 of the 22 (68%) most common CC's, were either

predominant in Sub-Saharan Africa (Gabon, Mozambique, Tanzania) (CC80, CC88, CC121, CC152) or Germany (CC22, CC30, CC45, CC398). The rate of MRSA carriers in the African (3%) and German populations (1%) were very low, probably because the investigation was limited to CA-SA. Analysis of gene profiles identified a high prevalence of the Panton-Valentine leukocidin encoding genes lukF/S-PV in African isolates. Generally, continent-specific differences for the CCs and the CA-SA gene profiles could be identified. Furthermore, it became apparent that in addition to the good first-line characterization by DNA MA, standardized, complex analytical methods for association and epidemiological analyses are required.

The last investigation purpose, the evaluation of comparative mass spectra analysis showed that only SA of CC121 were correctly identified based on their mass spectrometric profile on the CC level, while no clear discrimination was achieved for other CCs or *spa*-types.

Taken together, this work identified country specific differences in the distribution of CCs, and specific genes of the examined regions, especially concerning only clinical isolates. The work provide first clues to the genetic based reasons which may explain the purported difference in clinical presentation and course of diseases caused by SA. Additionally, it was shown that DNA MA in contrast to mass spectrometry is a suitable tool for SA characterization e.g. in outbreak situations. However, it should be noted that as specific requirement of a challenging DNA MA based phylogenetic analyses, application of complex bioinformatic analysis methods are required for interpretation of multiple isolates, and that the characteristics and algorithms of the chosen bioinformatic method may result in slight differences in the data interpretation.

1.2 Zusammenfassung

Staphylococcus aureus (SA) ist ein weltweit verbreitetes, opportunistisches Pathogen, dass 25-30% der Bevölkerung häufig symptomfrei kolonisiert. Treten jedoch invasive Infektionen auf (z.B. Wund-, Organinfektion, Sepsis), so sind diese mit erheblicher Morbidität und Mortalität assoziiert. SA besitzt ein breites Spektrum an Pathogenitäts-, Virulenz- und Resistenzfaktoren, deren Vorhandensein teilweise mit besonderen phylogenetischen Entwicklungslinien assoziiert ist. So basiert die Bewertung von Infektionsübertragung und Ausbruchsgeschehen vielfach auch auf phylogenetischer Analytik. Dies erfordert die Entwicklung reproduzierbarer Typisierungsverfahren, die

die unterschiedlichen SA-Stämme auch ausreichend trennscharf charakterisieren können. Daher wurden reproduzierbare, einheitlich auswertbare Polymerasekettenreaktionen und Sequenzierungsmethoden wie *spa* (SA Protein A)-Typisierung, Multilokussequenztypisierung (MLST), bis hin zur Gesamtgenomsequenzierung entwickelt. Im Gegensatz zur *spa*-Typisierung, basierend auf der Sequenzierung des einzelnen Gens *spa*, basiert die MLST auf der Sequenzierung von insgesamt sieben Haushaltsgenen, die verschiedene Sequenztypen (ST) definieren, welche zu klonalen Komplexen (CC) gruppiert werden. STs und CCs bilden heutzutage das Gerüst der internationalen SA Nomenklatur in epidemiologischen Studien. Eine besondere Technik stellen DNA Microarrays (MA) dar, mit deren Hilfe ST und CC aus charakteristischen Hybridisierungsmustern zahlreicher Genloci durch Vergleich mit einer Datenbank einer großen Stammsammlung abgeleitet werden. Neben dieser phylogenetischen Zuordnung besitzt der DNA MA zahlreiche Loci für Adhäsions-, Virulenz- und Toxingene und liefert daher eine große Masse an Daten. Der Vergleich zahlreicher, mittels DNA MA untersuchter, Isolate erfordert daher zur Interpretation und Präsentation ergänzende bioinformatische Methoden.

Insgesamt gibt es für SA eine umfangreiche Studienlage zur molekulargenetisch typisierten Infektionsepidemiologie; überwiegend bezieht diese sich jedoch auf Länder mit einem hohen Standard mikrobieller Diagnostik. Für Afrika gibt es nur wenige molekular-epidemiologische Daten von SA (inklusive Methicillin-resistenter SA [MRSA]), insbesondere von nicht nosokomialen (d.h., nicht im Krankenhaus erworbenen) SA (wie z.B. ambulant erworbener (CA) Methicillin-resistenter oder -empfindlicher SA [MRSA, MSSA]). Um diese Wissenslücke zu füllen, wurden als Grundlage dieser Promotionsarbeit vier Studien durchgeführt zur Untersuchung a) verschiedener bioinformatischer Werkzeuge zur Analyse der molekularen Epidemiologie von SA, b) der Epidemiologie von afrikanischen und deutschen CA-MSSA/MRSA, die nicht in Gesundheitsinstitutionen erworben wurden, und c) der Anwendbarkeit der Massenspektrometrie zur Analyse der Epidemiologie von SA.

Zur Charakterisierung der bioinformatischen Werkzeuge wurde das Subtypisierungspotential der *spa*-Typisierung und der DNA MA Analyse in einer Studie mit 46 nasalen CA-MSSA und 46 von Infektionen stammenden CA-MRSA des Bundeslandes Saarland verglichen. Zudem wurden die Ergebnisse mit drei bioinformatischen Methoden („splits graph“, hierarchische agglomerative Gruppierung

[HAC] und Hauptkomponentenanalyse [PCA]) zur DNA Analyse evaluiert. Um die Anwendbarkeit der DNA MA Analytik auch in afrikanischen Kohorten zu überprüfen, wurde eine relativ kleine Isolatkohorte von 52 nigerianischen MSSA Isolaten mittels DNA MA und splits graph Analyse untersucht. In einer großen, multizentrischen, afrikanisch-deutschen, geographisch vergleichenden Querschnittsuntersuchung wurden 1200 CA-SA Isolate, gesunder Freiwilliger und von Patienten aus Sub-Sahara Afrika und Deutschland, mittels DNA MA analysiert. Zudem wurden Assoziations- und vergleichende Analysen der CC, ST und der Genausstattung der Isolate, hinsichtlich ihrer geographischen Herkunft (Deutschland oder Afrika) bzw. ihres klinischen/kommensalen Ursprung, durchgeführt. Schließlich wurde die mögliche Anwendbarkeit der Identifikation von CCs oder *spa*–Typen von CA-SA des CC5, CC8, CC15, CC30, CC45, CC121 und der *spa*- Typen t002, t003, t504 basierend auf MALDI-TOF Massenspektrometrieprofilen analysiert.

Die mit diesen Untersuchungen angestrebten Ergebnisse, zur Epidemiologie von CA-MSSA/MRSA, wurden zunächst mit der DNA MA Studie des Bundeslandes Saarland gewonnen, welche ein höheres Subtypisierungspotential für den DNA MA im Vergleich zur *spa*–Typisierung zeigte. Die Auswertung der DNA MA Daten mittels der zuvor genannten drei bioinformatischen Methoden zeigte eine unterschiedliche Zuordnung von CC5 Isolaten zu 3 bis 7 Isolatgruppen mit Ausnahme von zwölf CC5 Isolaten. Diese zwölf CC5 Isolate wurden immer zwei identischen Isolatgruppen zugeordnet. Dies heißt nicht, dass die Methoden fehlerhaft sind, sondern dass die Anwendung unterschiedlicher mathematischen Algorithmen, die entweder die Unterschiede oder die Ähnlichkeiten der Hybridisierungsmuster untersuchen, zu geringfügigen Unterschieden führen. Zudem zeigten die MSSA eine hohe Diversität im Vergleich zur CA-MRSA Gruppe, die vom CC5 Rhine-Hesse epidemischen Stamm dominiert wurde (89%).

Die weitere Analyse afrikanischer Stämme aus der Studie mit nigerianischen MSSA zeigte eine heterogene und divergente Natur der untersuchten nigerianischen MSSA und bestätigte, dass der DNA MA eine geeignete Methode zur Genotypisierung afrikanischer SA ist. Die Ergebnisse der multizentrischen deutsch-afrikanischen Konsortialstudie wies innerhalb der Isolatgruppe insgesamt 40 verschiedene CCs und STs nach, einschließlich 5 neuer STs. Die Mehrheit, 15 von 22 (68%) der CCs, die auch am häufigsten identifiziert wurden, waren entweder in Sub-Sahara Afrika (Gabon, Mozambique, Tanzania) (CC80, CC88, CC121, CC152) oder Deutschland (CC22,

CC30, CC45, CC398) vorherrschend. Die Rate an MRSA Trägern in der afrikanischen (3%) und deutschen Bevölkerung (1%) war sehr gering, vermutlich durch die auf CA-SA beschränkte Untersuchung. Die Analyse der Genprofile zeigte eine hohe Prävalenz der Panton-Valentin Leukozidin kodierenden Gene lukF/S-PV in afrikanischen Isolaten. Allgemein konnten kontinentspezifische Unterschiede für die CC und CA-SA Genprofile identifiziert werden. Zudem wurde ersichtlich, dass ergänzend zur guten Erstcharakterisierung durch den DNA MA, standardisierte, komplexe analytische Methoden für Assoziationsanalysen und epidemiologische Untersuchungen erforderlich sind.

Das letzte Untersuchungsziel, die Bewertung vergleichender Massenspektrometrieanalysen zeigte, dass - basierend auf ihrem Spektrometrieprofil - nur SA des CC121 korrekt identifiziert werden konnten, während keine korrekte Identifikation anderer CC oder *spa*-Typen möglich war.

Zusammenfassend identifizierte diese Arbeit länderspezifische Unterschiede in der Verteilung klonaler Komplexe und spezifischer Gene in den untersuchten Regionen, insbesondere bei alleiniger Betrachtung klinischer Isolate. Sie gibt wertvolle Hinweise auf die genetisch basierten Gründe, welche die angeblichen Unterschiede in der klinischen Präsentation und dem Verlauf von durch SA verursachten Infektionen in sich entwickelnden und industriellen Ländern erklären könnten. Zudem wurde gezeigt, dass der DNA MA im Gegensatz zur Massenspektrometrie eine geeignete Methode zur SA Charakterisierung z.B. in Ausbruchssituationen ist. Es darf jedoch nicht unerwähnt bleiben, dass als besondere Anforderung an eine anspruchsvolle, DNA MA-basierte phylogenetische Analytik mehrerer Isolate die Anwendung komplexer bioinformatischer Analysemethoden zur Interpretation erforderlich ist, und dass die Charakteristiken und Algorithmen der gewählten bioinformatischen Methode zu geringfügigen Unterschieden in der Dateninterpretation führen kann.

2 Introduction

2.1 *Staphylococcus aureus*

Staphylococcus aureus (SA) is a globally distributed major human pathogen and human commensal. Thus, it is a global serious public health threat and the causative agent of an economic burden for health care systems (Köck *et al.*, 2010). SA is the most important member of the staphylococcal family and responsible for a wide spectrum of infections ranging from harmless skin infections such as folliculitis up to life-threatening infections such as pneumonia, endokarditis and bacteremia (Gordon & Lowy, 2008). Staphylococci are gram- and catalase positive, facultative anaerobic, immobile and non-spore-forming bacteria with a diameter of 0.5 to 1.5 µm. Since 2010, staphylococci constitute their own family, the *Staphylococcaceae,* and are no longer part of the *Micrococacceae* (Lemon *et al.*, 2010). The genus *Staphylococcus* comprises at least 46 species, which can be divided into coagulase positive and negative staphylococci, 16 species of them can be found in humans. The name *Staphylococcus* was introduced by Ogston 1882 to describe micrococci responsible for inflammation and suppuration. The name derives from the Greek “staphyle” which translates as “bunch of grapes” and “kokkos”, meaning “grain” or “berry” (Ogston, 1882). The name describes its typical morphology, indicating that the round single bacteria cells build grape like clusters because of their manner of cell division in multiple levels. Two years later Rosenbach (1884) introduced the scientific name *Staphylococcus aureus* (SA), adding the Latin “aureus” for “golden”. The name describes the golden colony color caused by the carotenoid staphyloxantin of this staphylococcus species (Rosenbach, 1884).

SA colonizes the skin, groin, axilla, vagina, throat and mostly the vestibules of the nose as an opportunistic pathogen. Up to 50% (Kirkliauskienė *et al.*, 2010) of the human population are colonized transiently and 10 to 20% persistently (van Belkum *et al.*, 2009, Wertheim *et al.*, 2004). Colonization occurs without symptoms, but is associated with an increased infection risk for nasal carriers (Kluytmans *et al.*, 1997, Lowy, 1998). On the other hand, non-carriers in comparison to carriers have a fourfold higher mortality rate in cases of exogenous bacteremia (Wertheim *et al.*, 2004). Invasion into the host of SA is promoted by injuries (broken skin or mucosal surfaces), implanted

foreign material such as catheters or a compromised immune system. The bacterium is able to spread from its port of entry into adjacent tissues or the blood system.

The reason for the success and the high pathogenicity of SA as a human commensal and opportunistic pathogen are three characteristics: i) resistance genes, ii) virulence genes and iii) successful transmission. The first big challenge is its resistance to several antibiotic classes and the prevalence of multiresistant staphylococci, especially of methicillin-resistant SA (MRSA). The second challenge is its wide array of virulence determinants as part of the pathogenesis of various diseases (Archer, 1998, Feng *et al.*, 2008, Gill *et al.*, 2005, Takeuchi *et al.*, 2005). Third, they are able to spread successful in different environments: i) hospitals, health-care institutions and in the community as well as in ii) human and animals.

2.1.1 Resistance of SA

Betalactams has since the introduction of penicillin been the antibiotic class of choice for the treatment of SA infections. However, development of resistance has from the very beginning been a serious issue. Thus, only very shortly after penicillin was introduced in 1940 the first penicillin resistant SA was found in 1942, first in hospitals and afterwards also in the community (Lowy, 2003, Rammelkamp & Maxon, 1942). In the late 1960 the resistance was reported to be about 80% regardless of their origin (i.e. hospital- or community-associated). Today still 70-80% of methicillin-sensitive SA (MSSA) are penicillin resistant (Chambers & De Leo, 2009).

SA resistance against penicillin is mediated by the expression of the gene for β-lactamase *blaZ* and regulated by the adjacent antirepressor *blaI* and the repressor *blaR*. The β-lactamase confers resistance to penicillins by inactivation of penicillin through hydrolysis of its β-lactam ring.

The emergence of penicillin resistance in SA leads to the development of new antibiotics such as streptomycin, erythromycin and chloramphenicol in the 1950s. However, during clinical application SA developed resistances against these, too (Shanson, 1981, Udo *et al.*, 1999).

During the search of β-lactamase stable penicillin's as a new treatment option, methicillin was introduced as first of the semisynthetic penicillinase-resistant penicillin's in 1961. However, similarly to the development of the penicillin resistance, methicillin

resistance was discovered in the same year shortly after the introduction of methicillin (Jevons, 1961).

The by far most important resistance mechanism in MRSA is mediated by the *mec* genes. All MRSA strains contain a staphylococcal chromosomal cassette (SCC) carrying the *mec* operon. The SCC*mec* harbors the already well-known methicillin resistance genes *mecA*, *mecC* and the in SA just recently reported *mecB* (Ito *et al.*, 2012, Becker *et al.*, 2018). They all encode penicillin binding proteins (PBP).

The *mec* genes are located within the *mec* operon together with its regulatory genes (Berger-Bachi & Rohrer, 2002). The SCC*mec* is a mobile genetic element originally acquired from other *Staphylococcus* species. Thus, the SCC*mec* is a foreign genetic element from whom its mechanism of acquisition and spread in SA is not completely understood.

One assumption is that the SCC*mec* elements were acquired horizontally by SA from *Staphylococcus fleurettii* or *Staphylococcus sciuri* and that *mecA* positive, coagulase negative staphylococci as *Staphylococcus epidermidis*, may be a potential reservoir for the SCC*mec* element (Hiramatsu *et al.*, 2013, Rolo *et al.*, 2017, Tsubakishita *et al.*, 2010, Wu *et al.*, 1996). For example, SCC*mec* types II and III are too large to be transferred by bacteriophages (Lindsay & Holden, 2004).

Till yet 13 different SCC*mec* types (I to XIII) have been differentiated, five of them, SCC*mec* types I to V, are globally distributed (Baig *et al.*, 2018). The larger SCC*mec* allotypes I, II and III are normally found in healthcare-associated (HA)-MRSA and harbor multiple resistance determinants. The smaller, more mobile SCC*mec* allotypes IV and V are normally associated with community-associated (CA)-MRSA and don't carry multiple antibiotic-resistance genes (Baba *et al.*, 2002, Chambers & De Leo, 2009). However, SCC*mec* type IV could not be found only in CA-MRSA but also in HA-MRSA of sequence type (ST) 22 (ST22-IV) (Otter & French, 2012). Therefore, it is not possible to conclude from the SCC*mec* type if a SA isolate is a HA- or CA-MRSA isolate.

Vancomycin was increasingly used as last remaining antibiotic which MRSA were reliably susceptible, too (Hiramatsu *et al.*, 1997). However, after a long period the increasing burden of MRSA infections worldwide with MRSA rates up to 50% in bacteremia in European countries and 59% of MRSA in skin and soft tissue infections

in the USA (Köck *et al.*, 2010, Moran *et al.*, 2006) leads to the first Vancomycin intermediately resistant SA (VISA) strains detected in 1997 (Hiramatsu *et al.*, 1997, Smith *et al.*, 1999). VISA are strains showing a reduced susceptibility to vancomycin through adaptive mutation (Nelson *et al*., 2007, Howden *et al.*, 2014, Hanaki *et al.*, 1998). In Germany they were described for the first time in 1997 (Epidemiol Bulletin, 45/97). Later, complete vancomycin resistant SA (VRSA) were found. VRSA are caused by conjugal transfer of the *vanA* operon from *Enterococcus faecalis* (Hu *et al.*, 2016). So far, still less than 20 vancomycin resistant SA strains are described worldwide, with most of them reported from the USA and in one case from Portugal (Friães *et al.*, 2015, Limbago *et al*., 2014).

Until today SA developed resistances against all big antimicrobial classes including macrolides (e.g. erythromycin), lincosamides and streptogramin B, aminoglycosides [e.g. gentamicin], tetracyclines, linezolid, mupirocin, fusidic acid, oxazolidinone, glycopeptides (e.g. daptomycin), lipopeptides, gyrase like gyrase-inhibitors (e.g. ciprofloxacin) and many other unrelated antibiotic classes.

2.1.2 Virulence of SA

Pathogenesis of SA is driven by different virulence factors from adhesion mediating virulence factors, over biofilm promoting virulence factors to virulence factors like cytotoxins such as enterotoxins, hemolysins and leukocidins guaranteeing the SA nutrition, as well as virulence factors counteracting the host's immune system like the IEC (immune evasion cluster).

The start of pathogenesis is initiated by adhesion factors of SA. The first of the both predominant groups of adhesion factors are the cell wall associated proteins called microbial surface component recognizing matrix molecules (MSCRAMMs). MSCRAMMs share that they are covalently anchored to the cell wall by sortase A, a transpeptidase that recognizes the C-terminal cell-wall sorting signal LPXTG. The key representatives of the MSCRAMMs are the fibronectin binding proteins (FnBPA and FnBPB) (Foster, 2016), the fibrinogen binding proteins clumping factor A and B (ClfA and ClfB) (McDevitt *et al.*, 1994, Ni Eidhin *et al.*, 1998), the Ser-Asp dipeptide repeat surface proteins (SdrC and SdrD) (Foster & Hook, 1998), the bone sialoprotein binding protein (Bbp) (Vazquez *et al.*, 2011), the collagen binding protein (Cna) (Gillaspy *et al.*, 1998), and protein A (Spa) encoded by *spa*.

The second predominant group of adhesion factors are the secretable expanded repertoire adhesive molecules (SERAMS). These secreted adhesion proteins partially rebind to the bacterial cell surface (Chavakis *et al.*, 2007, Chavakis *et al.*, 2005, Foster *et al.*, 2014). Well known structurally unrelated SERAMs are the coagulase Coa, the fibrinogen binding protein FbpA, von-Willebrand-factor binding protein encoded by *vwb* and the extracellular adherence protein Eap, also referred to as Map (Chavakis *et al.*, 2005, McAdow *et al.*, 2012). All these adhesion genes enable SA to adhere and mediate the invasion of a range of different cell types (Chavakis *et al.*, 2007, Lacey *et al.*, 2016).

After adhesion, pathogenesis may be driven by biofilm formation which plays a major role in foreign body infections. A biofilm consists of a complex microbial flora within a self-synthesized matrix. One of the advantages of biofilm growth is the increased resistance to host defense mechanisms and antibiotics (Arciola *et al.*, 2012). A number of clinical SA isolates are able to synthesize the exopolysaccharide PIA (intercellular polysaccharide adhesin) also referred to as PNAG („poly-N-acetylglucosamin“). PIA synthesis is catalyzed by the products of the *ica*-operon components *icaA, icaB, icaC,* and *icaD*, whose expression is for example controlled by the negative regulator IcaR (Cue *et al.*, 2012, Otto, 2008). Other SA isolates can form a proteinaceous biofilm utilizing MSCRAMMs such as FnBPs, Spa, and the staphylococcal surface protein G (SasG). The SA surface component Bap (encoded by *bap*) is also involved in biofilm formation, which (in comparison to PIA) promotes not only the intercellular adhesion between SA cells but also the attachment to inert surfaces (Cucarella *et al.*, 2002, Taglialegna *et al.*, 2016).

Moreover, SA secretes several cytotoxins e.g. enterotoxins (Hu *et al.*, 2008, Hu *et al.*, 2016, Spaulding *et al.*, 2013) as the staphylococcal superantigen/ enterotoxin-like antigens (Argudin *et al.*, 2010), toxic shock syndrome-Toxin (TSST-1) (Becker *et al.*, 2003) and exfoliative toxins ETA and ETB (Grumann *et al.*, 2014, Nishifuji *et al.*, 2008). These toxins are known to have potent effects on cells of the host's immune system to inhibit the host's immune responses. However, many of them have other biological effects, too.

Cytotoxins such as hemolysins and leukocidins can form pores in diverse host cell types. The chromosomally encoded hemolysins α, β, γ and δ are responsible for the lysis of erythrocytes and may also harm other eukaryotic cell types (Vandenesch *et al.*,

2012). The function of the hemolysins and leukocidins is to ensure the supply of nutrients by lysis of host cells.

The best known leukocidin, Panton-Valentine leukocidin (PVL), consists of the subunits *lukF-PV* and *lukS-PV* (Shallcross *et al.*, 2013). Typical clinical manifestations of the well-known leukocidin PVL (Shallcross *et al.*, 2013) are chronic or recurrent skin and soft tissue infections (SSTI). In rare cases, PVL is also associated with severe life-threatening diseases such as necrotizing fasciitis and pneumonia (Gillet *et al.*, 2002, Lina *et al.*, 1999). The attention of PVL arised with the detection of an epidemiological association between the presence of the genes *lukF-PV* and *lukS-PV* and CA-MRSA (Vandenesch *et al.*, 2003). More than 75% of CA-MRSA carry the genes encoding for PVL (Daum *et al.*, 2002, Naimi *et al.*, 2003). Thus, PVL was assumed to be one of the criteria for definition of CA-MRSA (Otter & French, 2010). However, an increasing number of CA-MRSA clones do not contain the PVL encoding genes (He *et al.*, 2018).

Cytotoxins are causative for the invasiveness of SA and are responsible for typical clinical conditions. Their main function is to guarantee the nutrient supply and to counteract the host‘s immune system (Dinges *et al.*, 2000, Alonzo & Torres, 2014).

SA further secrets a number of lytic enzymes such as hyaluronate lyases, the zinc-dependent metalloprotease aureolysin (Aur), the enzyme proteases staphopain A, B and P (SspA, SspB, SspP) and the serine proteases SplA, SplB and SplE (Bailey *et al.*, 1997, Dubin, 2002, Redpath *et al.*, 1991). They are mainly involved in the destruction of the host tissue to enable the bacteria to spread from the initial infection site to new sites through lysis of specific cell structures. In this way its survival and nutrition will be guaranteed.

To counteract the host ‘s immune system SA possess the IEC. The IEC consists of three components, the immune modulator staphylococcal complement inhibitor (SCIN), the chemotaxis inhibitory protein of SA (CHIPS) and the plasminogen activator Staphylokinase (SAK) (Foster, 2005, McGuinness *et al.*, 2016, Collen, 1998).

Additionally, epidermal cell differentiation inhibitors (EDIN) belonging to a large group of virulence factors support the hematogenous spread of SA into deeper tissues (Lemichez *et al.*, 2010).

Most SA clinical isolates are also able to produce a polysaccharide capsule, which covers the outer layer of the bacteria and protect them from phagocytosis through host

immune cells. Eight capsule serotypes, CP1 to CP8, can be distinguished with CP5 and CP8 being predominant in clinical isolates (O'Riordan & Lee, 2004).

The virulence factor synthesis of SA is regulated by a complex regulatory network. These regulators are influenced by different conditions like the growth phase of the bacteria and environmental conditions like pH-value or available nutrients. Three well described regulation systems of this network are the quorum sensing system Agr (accessory gene regulator) (Novick, 2003, Somerville & Proctor, 2009), the two component systems SaeRS (*S. aureus* exoprotein expression RS) (Novick, 2003), and the Sar (staphylococcal accessory regulator) family (Liu *et al.*, 2016, Somerville & Proctor, 2009).

The Agr system plays an important role in the growth phase dependent regulation of various virulence genes. It suppresses the transcription of several cell wall associated proteins such as Protein A, coagulase and the fibronectin binding proteins. It also activates the transcription of several exoproteins such as α-toxin, β-haemolysin, TSST-1 and leukotoxins during the post-exponential phase. There are four distinguishable *agr* allelic groups: *agr*-type I, *agr*-type II, *agr*-type III and *agr*-type IV. They can be distinguished based on the cross-inhibition of the *agr*-response by their different autoinducing peptides (Bronner *et al.*, 2004, Vandenesch *et al.*, 2012).

A second important group are the Sar DNA binding proteins. The *sar* operon in SA comprises the three transcription units *sarA*, *sarB* and *sarC* encoding for the SarA protein. The expression of these transcripts is growth dependent (Bayer *et al.*, 1996). SarA can influence the transcription of different virulence genes such as *spa* encoding for the protein A and *hla* encoding for the alpha haemolysin (Novick, 2003, Somerville & Proctor, 2009). In addition it is an important activator of the *agr*-system (Cheung *et al.*, 1997).

The SaeRS two-component system plays an important role in the expression of over twenty virulence factors such as hemolysins, leukocidins, superantigens, surface proteins and proteases. The Sae system doesn’t have an influence on other regulation systems but it is known that the transcription of the *sae*-operon is influenced by other regulatory systems (Liu *et al.*, 2016).

2.2 SA epidemiology and the need of SA typing

2.2.1 The need of SA typing

Characterization including typing is an essential instrument to understand the epidemiology of pathogens. Reports of large SA outbreaks in the 1950's showed that a typing system for SA was required not only to understand individual, local outbreaks but also to understand transmission. Even today SA typing is the cornerstone in the understanding of the SA epidemiology and to investigate outbreaks especially in health care setting.

Hospitalized patients still have a higher risk of infection because they usually suffer from comorbidities associated with an increased infection risk (Herrmann *et al.*, 2013b, Cadena *et al.*, 2016). Furthermore, a thorough hygiene management can limit the pathogen transmission through doctors, nurses and other healthcare workers in hospitals. Studies showed that transmission between patients in settings without any infection control precautions as e.g. patient screenings and isolation at hospital-entry have a 16fold higher infection risk compared to patients with contact isolation (Tacconelli *et al.*, 2009). But also in the community the understanding of the mechanism of the successful spread SA could help to develop prevention strategies.

The first developed typing method was the phage typing (Blair & Williams, 1961). In the further decades, several different methods have been developed for MSSA and MRSA typing, surveillance, epidemiological investigations and outbreak detection. It started with traditional typing systems based on phenotype, such as phage-typing. Later, fingerprint-based methods such as amplified fragment length polymorphism (AFLP) and pulsed-field gel electrophoresis (PFGE) were developed.

Furthermore, the international spread from previous, locally successful clones showed that there was a need for a common language and typing methods standards for communication about SA infections between labs and countries with outbreaks of prevalent SA strains. Harmonization projects were necessary to define a common language (Murchan *et al.*, 2003). However, the attempt to create internationally, inter-lab comparable standards for fingerprint-based methods was not fully successful (Murchan *et al.*, 2003). This means that PFGE still is one of the methods of choice for the investigation for outbreaks but is not the method of choice for epidemiologic analyses.

For this purpose, polymerase chain reaction (PCR) based methods were developed and preferentially used such as *spa* (SA protein A)-typing, multilocus sequencing typing (MLST), DNA microarrays (DNA MA) and most recently whole genome sequencing (WGS). For MRSA SCC*mec*-typing also play a major role. Additionally, there were several attempts to prove the applicability of mass spectrometry as SA typing method.

Strictly, internationally standardized epidemiological SA typing conditions and strain type definitions allowed the introduction of freely accessible databases for example for *spa*-typing and MLST. However, it has to be mentioned that for *spa*-typing there are two nomenclatures from Ridom (Harmsen *et al.*, 2003) and Kreiswirth (Koreen *et al*., 2004), while there is only one nomenclature for MLST (Maiden *et al.*, 1998, Urwin & Maiden, 2003).

Actually, 17913 *spa*-types (www.spa.ridom.de/spatypes.shtml, 06.08.2018) and 4903 STs (www.pubmlst.org/data, retrieved till 06.08.2018, Figure 1) were described by the both established databases.

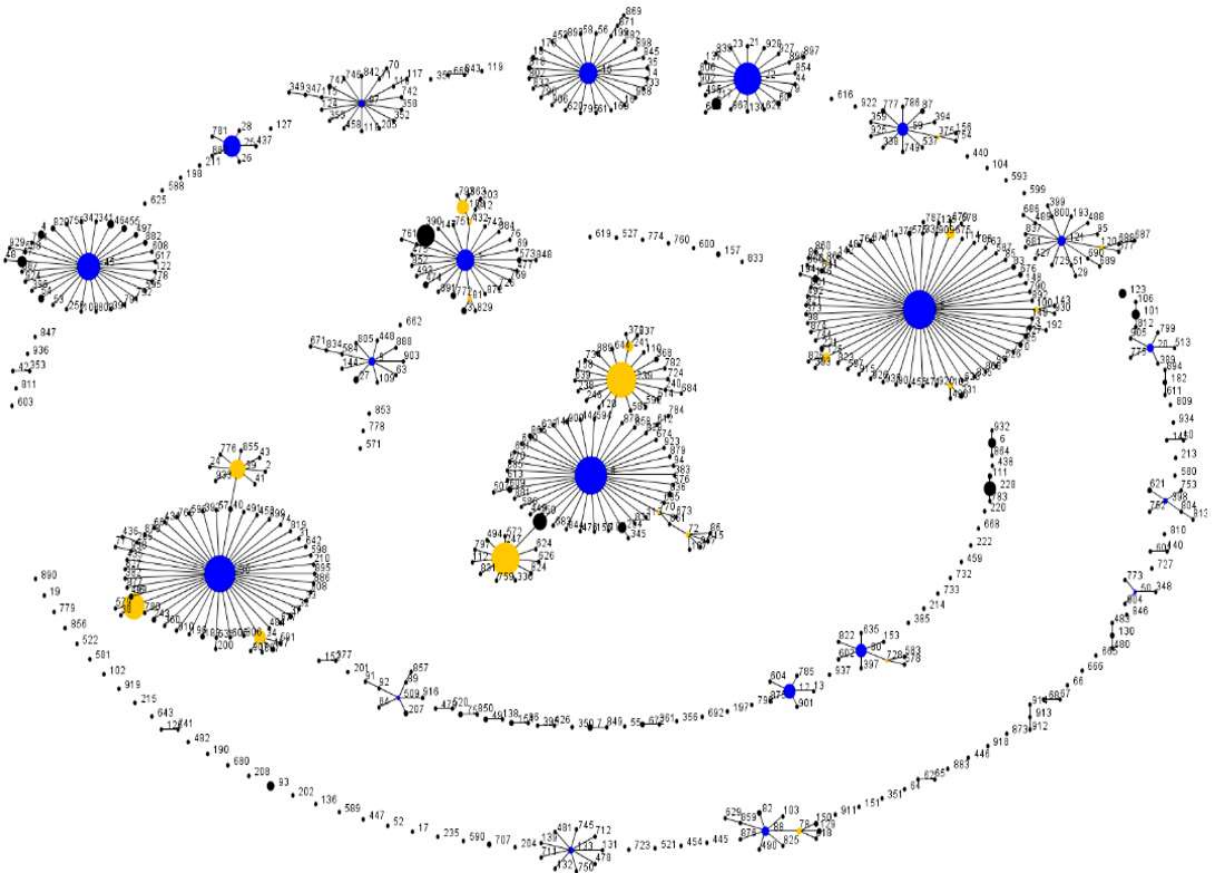


Figure 1: **Population structure of all known *Staphylococcus aureus* (SA) sequence types (ST).** The figure is created by the electronic based upon repeats sequence types (eBURST) of all SA STs from the multilocus sequence typing (MLST) database (http://saureus.mlst.net) as of July 2018. Each circle represents a single ST. Clonal complexes of STs (represented by dots) differing at only one out of seven alleles are joined by lines. Highlighted in blue are founding genotypes. Yellow dots represent subgroup founders. Size of dots indicates the frequency of a specific ST within the database.

These molecular typing methods are essential for the rapid identification of prevalent strains. They enhance the understanding of disease epidemiology and offers new insights into the evolution of SA. Molecular typing is important for their active surveillance to control and prevent the spread of SA in health-care settings and to act against their transmission, especially in outbreak situations.

2.2.2 SA typing methods

Today several different methods are available for differentiation of MSSA and MRSA, surveillance and outbreak detection. However, each of these different methods has its own advantages and disadvantages. They differ in their workload, accuracy, discriminatory power and reproducibility (van Belkum *et al.*, 2009, Weller, 2000). These characteristics must be considered when choosing the best appropriate method

regarding the time, geographical scale and the objectives of a study as well as the available resources.

2.2.2.1 Phage typing

Phage typing today is only in very limited use, however it was the cornerstone of SA typing for more than 30 years and thus is briefly described.

Phage typing is based on the classification of bacterial strains according to their susceptibility to a set of selected phages. This means that these strains are lysogenic, they carry bacteriophages. These phages are able to lyse specific but not all strains of a bacteria species. Applying a defined set of phages allowed to recognize a "susceptibility pattern" and by comparison of the "susceptibility patterns" the similarity or dissimilarity of the examined SA strains could be determined. For phage typing of human SA an internationally accepted set of 23 phages is used. It is important for interpretation and comparability of results that internationally recommended protocols for propagation and maintenance of phages are followed (Marples & Rosdahl, 1997). The phage typing results are easy to interpret and showed a fair amount of discriminatory power. In addition, it is a cost-effective method. On the other hand it is a time consuming and technically demanding method. However, the main disadvantage of this method is its limitation by poor reproducibility of the results and the high proportion SA strains which are not typeable by this method (Archer *et al.*, 1983, Coia *et al.*, 1990).

2.2.2.2 Amplified fragment length polymorphism (AFLP)

The amplified fragment length polymorphism (AFLP method) is similar to phage typing still in limited use. However, it was one of the first fingerprint-based methods and thus is briefly described exemplary.

AFLP method is based on the restriction of genomic DNA, which is cut with two restriction enzymes. In addition, double stranded adaptors are ligated to one of the sticky ends of the restriction fragments (Vos *et al.*, 1995). This enables a selective amplification by PCR of the restriction fragments ending with the ligated adaptor using complementary primers to the adaptor sequence, the restriction site sequence and several additional nucleotides. It is possible to specifically co-amplificate high numbers (between 50-100) of restriction fragments, often in combination with fluorescent dye-

labeled PCR primers. Computer-assisted analysis of high resolution banding patterns of AFLP analysis allowed the determination of genetic relatedness among the examined bacterial isolates (Mortimer *et al.*, 2001). AFLP are reported to be as discriminatory as pulsed-field gel electrophoresis (PFGE) (Zhao *et al.*, 2000). In comparison to PFGE it is a reproducible method which can be automated (Duim *et al.*, 1999). The disadvantages of this method are its working intensity and the associated costs.

2.2.2.3 Pulsed-field gel electrophoresis (PFGE)

As one of the fingerprint-based methods PFGE is still in the most settings the gold standard of bacterial typing in outbreak situations. The reasons are the high discriminatory power and the high in-house reproducibility. In 1991 PFGE was first applied for epidemiological surveillance of MRSA (Prevost *et al*., 1991). For PFGE bacterial DNA is digested by a restriction enzyme that only rarely cuts within the chromosome, thereby generating a small number of large chromosomal fragments. For the low G and C species SA (harboring a DNA chromosome with fewer [G] and cytosine [C] nucleotides than adenine [A] and thymine [T] nucleotides), the enzymes SmaI (used for MSSA) and ApaI (used for MRSA) are usually used, recognizing and cleaving CCCGGG and GGGCCC sites, respectively. Next, the DNA fragments are analyzed on an agarose gel. A specific feature of the PFGE is that the electric field constantly changes its direction, allowing the separation of the large DNA fragments and causing a high resolution.

A disadvantage of the PFGE is the low interlaboratory comparability because of its high susceptibility to even minor variations in the experimental settings and the interpretation of band marking (van Belkum *et al.,* 1998). Therefore, it was important to define international standards for an epidemiological SA typing language (Murchan *et al.*, 2003). However, it was not possible to create a common language or to build an interactive database for global analysis based on PFGE which already exist for other molecular typing methods such as *spa*-typing and MLST or will be conceivable for DNA MA and WGS data, respectively. Thus, PFGE is still applied for the investigation of outbreaks but not for epidemiological investigations. Moreover, it is a laborious and time-consuming method with low throughput (Hallin *et al*., 2007) why it is not suitable for long-term epidemiological investigations (Blanc *et al.*, 2002).

2.2.2.4 Staphylococcal cassette chromosome *mec* (SCC*mec*) - typing

For MRSA SCC*mec*-typing also play a major role. MRSA needs to be differentiated below the species level by rapid and comparable typing methods as SCC*mec*-typing.

The staphylococcal cassette chromosome *mec* is a mobile genomic island of SA. It is an unique staphylococcal vector for gene transfer with tremendous variation in size and structure as well as of tremendous variation of genes they carry. To date 13 types of SCC*mec* elements were identified (Baig *et al*., 2018). All SCC*mec* elements have in common that they carry the *mec* gene complex encoding methicillin resistance (Hiramatsu *et al.*, 2013). Thus, the acquisition of this foreign SCC*mec* element transforms MSSA to MRSA.
All SCC*mec* elements share the same characteristics: i) carriage of the *mec* gene operon (*mec*), ii) carriage of the *ccr* gene complex (*ccr*), which are iii) flanked by specific nucleotide sequences, direct and inverted repeats at both ends and iv) integration at the so-called integration site sequence (ISS) for SCC, located at the 3'-end of *orfX* or at the extremity of the SCC element v) joining regions (j-regions) other than *mec* and *ccr* within SCC*mec* (Hiramatsu *et al.*, 2013, Ito *et al.*, 2014, Tsubakisihita *et al.*, 2010). These 13 types were classified based on the combination of their different *mec*- and ccr-types. The *mec*-types were classified based on the presence or absence of specific insertion sequences. Furthermore, J-regions can be differentiated into three subgroups (J1-J3) (Hiramatsu *et al.*, 2013).

The SCC*mec* typing is based on several multiplex PCRs. Several multiplex PCRs are needed to identify all SCC*mec* types. Therefore, this method needs a lot of optimization, is time consuming and is cumbersome to introduce in the routine diagnostic, too (Ghaznavi-Rad *et al*., 2010).

2.2.2.5 SA protein A – typing (*spa*-typing)

spa-typing is a single locus sequencing method which is one of the methods applied in the investigation of outbreaks in hospitals and health care settings.

The staphylococcal protein A is a 42kDa protein of the MSCRAMMs (section 2.1.2). The open reading frame of *spa* is about 2150bp long and encompasses different regions with a varying degree of conservation: the well conserved Fc-region encoding the N-terminal end of the protein A, which interacts with the immunoglobulins, a hypervariable X-region used for typing, and the well conserved sequence encoding the

C-terminal end of the protein which is responsible for the cell wall anchorage (Guss *et al.*, 1984). The polymorphic X-region was sequenced for the first time in 1984 by Uhlen et al. (Uhlen *et al.*, 1984), while Frenay et al. examined the epidemiological potential of *spa*-typing for the first time in 1994 (Frenay *et al.*, 1996). The highly polymorphic X-region is a region consisting of a variable number of about 24bp long repeats (Uhlen *et al.*, 1984) that can be used to subdivide SA isolates into different *spa*-types. The *spa*-types are defined by varying repeat numbers as well as by numerous strain-specific repeat succession possibilities and changes through point mutations and deletions of the X_R region (Figure 2) (Frenay *et al.*, 1996). The distinctive polymorphism of this region is based on duplications, deletions, insertions and point mutations.

The *spa*-types are defined based on the repeat succession of the polymorphic X_R region (Figure 2). The repeats are defined by a unique base composition. Each new detected repeat is assigned a unique repeat code consisting of an “r” and a number for example r17.

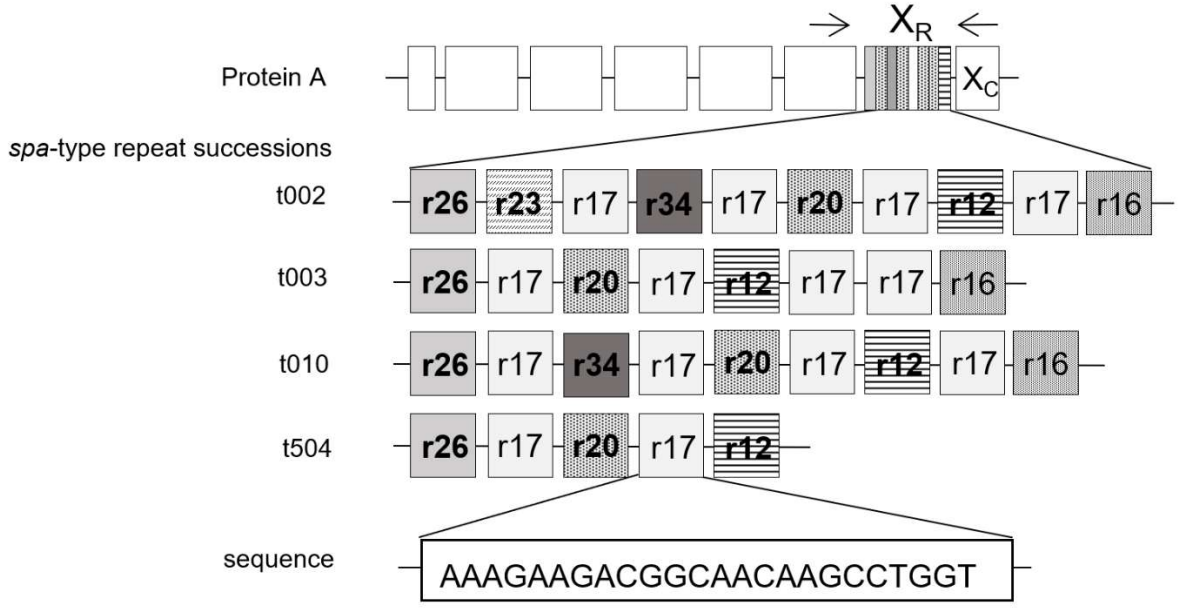


Figure 2: **Genetic organization of the *Staphylococcus aureus* (SA) protein A (*spa*)-gene and the principle of *spa*-typing.** Schematic representation of the protein A encoding *spa* open reading frame of SA. The boxes represent different gene segments encoding for the different parts of the protein A, with the hypervariable X-region and their repetitive elements (X_R), and the cell wall anchoring segment (X_c) at the end. Arrows symbolize the position of the primers used to amplify the X-region utilized for typing. The repeat succession of an isolate determines its *spa*-type. The repeats are defined by a unique base composition, and each repeat is assigned a unique repeat code consisting of an “r” and a number. Identical repeats are highlighted by identical colours, patterns and font.

Spa-typing as a single locus typing method of the hypervariable X-region of Protein A is nowadays a routinely applied typing method in diagnostics as well as in outbreak situations e.g. to check if a possible transmission of a specific SA strain from one to another patient occurred. The success of this method is based on its ease and speed, its interlaboratory reproducibility as well as its objective interpretation and access to a global, regularly updated database (Aires-de-Sousa *et al.*, 2006, Harmsen *et al.*, 2003, Shopsin *et al.*, 1999). This typing method is particularly suitable for the brief analysis of local outbreaks. However, a disadvantage of this method is that epidemiologically related isolates might be different due to frequent sequence changes in the X-region which may result in different *spa*-types (van Belkum *et al.*, 2009).

2.2.2.6 Multi-locus-sequence-typing (MLST)

The development of MLST started with the analysis of highly virulent *Neisseria meningitidis* isolates (Maiden *et al.*, 1998). MLST was developed as a fast, highly reproducible typing method for unambiguous strain characterization. In 2000, Enright and colleagues first published an article on SA characterization by MLST (Enright *et al.*, 2000). It is a highly discriminative method with sequencing of internal ~450bp long fragments of seven constitutively expressed genes, so-called housekeeping genes. The seven housekeeping genes used for SA MLST are *arcC* (carbamat kinase), *aroF* (shikimate dehydrogenase), *glpF* (glycerol kinase), *gmk* (guanylat kinase), *pta* (phosphate acetyltransferase), *tpi* (triosephosphat isomerase), and *ygiY* (acetyl coenzym A acetyltransferase). The sequences of each gene fragment of the different gene loci are compared and each allele is assigned a specific number. The assignment of the alleles of the seven loci and the combination of their numbers results in the allelic profile which defines the sequence type (ST) (Figure 3) (Aanensen & Spratt, 2005, Enright *et al.*, 2000). If the allelic profiles of six out of the seven housekeeping genes of two individual isolates are identical, then they have the same ST. If only five of the seven alleles are identical, these strains have different STs but belong to the same clonal complex (CC).

The sequence fragments of the housekeeping genes used in MLST are characterized by a slow accumulation of genetic variations. This makes it on the one hand a suitable method for long term and global epidemiological studies (Feil *et al.*, 2003), but on the other this could lead to an insufficient discriminatory potential in epidemiologically unrelated strains (Cooper & Feil, 2004). A low mutation rate of the sequence fragments

of seven housekeeping genes makes MLST most suitable for long-term and global epidemiological studies (Feil *et al.*, 2003). Therefore, it is necessary to examine seven gene fragments to achieve a sufficient discriminatory power, which (in combination with the high MLST stability) makes it a particularly suitable tool for the analysis of strain clonality in global, long-term studies (Feil & Enright, 2004, Hallin *et al.*, 2008).

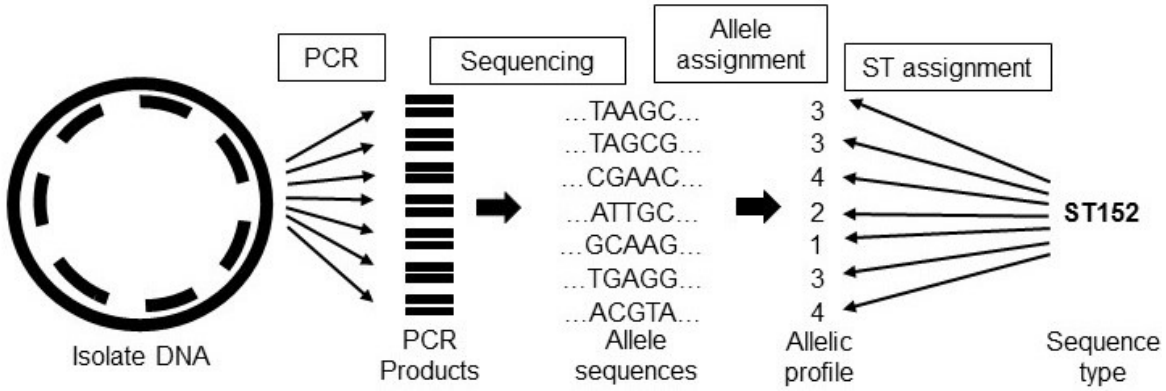


Figure 3: **Principle of multi-locus sequence typing**. Seven housekeeping genes of *Staphylococcus aureus* are amplified and sequenced. Depending on the allele sequences are defined numbers assigned, which determines the allelic profile. Isolates with identical allelic profile of all seven or six alleles belong to the same sequence type (ST). If only five of the seven allele sequences are identical they belong to the same clonal complex (CC).

2.2.2.7 Whole genome sequencing (WGS)

Recently, next generation DNA sequencing (NGS) made WGS possible. WGS is expected to change the practice of clinical microbiology and infection control because of the enhanced understanding of the micro-evolution of SA during long-term carriage and its relationship with different states of disease. The advantage of WGS is that it is a highly discriminative technique that can provide valuable insights into transmission chains. This method allows detection of single nucleotide polymorphisms as well as core genome MLST, among analytical approaches. WGS can be used for MRSA surveillance and outbreak analysis, where the conventional methods such as MLST and *spa*-typing showed an insufficient discriminatory power (Bartels *et al.*, 2015, Leopold *et al.*, 2014, Sabat *et al.*, 2013). In addition, WGS can be applied for the analysis of phenotypical properties on a genotypical basis, e. g. for the determination of antibiotic resistance profiles (Billal *et al.*, 2011). Advances in WGS technology and reduced sequencing costs to approximately 60€ per genome make this method more and more attractive for diagnostic purposes (Price *et al*., 2013).

However, the detailed bioinformatic analysis and limited information technology structure for data storage and exchange, pose a challenge and are some of the major obstacles to be overcome before introduction of WGS in the routine practice. There are several commercial software- and web-based solutions available for the performance of *spa*-typing, MLST, analysis of epidemiologic strain evolution, strain relationships, transmission chains as wells as for screening for predefined genes based on WGS data. Consequently, there will maybe be a future technique for the virulence assignment of clinical isolates based on their genotypical profiles (Laabei *et al.*, 2014).

2.2.2.8 DNA microarray (DNA MA)

DNA MA technology has facilitated the simultaneous analysis of the presence of selected molecular targets based on hybridization reactions since the late 1990s (Bumgarner, 2013). The DNA MA is a collection of DNA probes attached to a solid surface. These DNA probes can be used to detect complementary nucleotide sequences in bacterial isolates. Studies showed that they were already successfully applied for the investigation of the genetic composition of predominant SA lineages and the evolutionary genomics of SA (Lindsay *et al.*, 2006, Fitzgerald *et al.*, 2001). The Alere Technologies DNA SA oligonucleotide microarray used in this study provides 334 immobilized probes corresponding to 233 different genes. The different gene groups of the Alere Technologies DNA MA encompass SA species markers, regulatory genes, antibiotic resistance genes, toxin genes, adhesion genes, biofilm genes, immune evasion genes and other virulence-related genes. Moreover, the classification of the accessory gene regulator, capsule and some SCC*mec* types are included.

DNA of a SA isolate is amplified via a linear PCR and labelled with biotin. In contrast to standard PCR, only one antisense primer per target is used. This leads to single stranded, labeled DNA reaction products and an amplification of a high number of targets but also leads to a lower sensitivity than in a standard PCR. This approach also diminishes the probability of a possible contamination and prevents the competition between probe and antisense strand during the amplification, resulting in an increased probability of the single stranded amplicon binding to the probe (Monecke & Ehricht, 2005). The resulting amplicons are hybridized to the DNA MA and are made visible through a horseradish peroxidase reaction.

The final hybridization profiles are rather specific for different SA strains. It is possible to identify the strain lineage CC or ST of newly collected isolates analyzed by the DNA MA by automatic comparison of the obtained hybridization profiles with the hybridization profiles in the IconoClust (Alere Technologies) software database (Monecke *et al.*, 2008). The software automatically identifies the hybridization patterns by means of digital pictures of the array and provides information on presence, absence and ambiguous signals for the different genes and classifies the strain in concordance with the MLST nomenclature. This method thereby combines the analysis of clinically relevant factors with a globally accepted, epidemiological classification. DNA MA results can be obtained within a day and deliver easy to understand results with a high reproducibility and comparability.

In this work, DNA MA analysis was performed as a fast, commercially available, diagnostic panel for SA typing and as one of the molecular state of the art typing methods with affordable costs, convenient (laboratory) handling and a robust technique with a direct result output without requiring further bioinformatical analysis steps for interpretation. Especially the latter characteristics qualify DNA MA as a suitable method for microbiological laboratories in developing countries.

2.2.2.9 Mass spectrometry using MALDI-TOF

The introduction of matrix-assisted laser desorption ionization - time of flight (MALDI-TOF) as mass spectrometry technique in routine diagnostic initiated a revolution in the identification of microorganism grown on solid media and nowadays also in fluid blood cultures (Christner *et al.*, 2010). It is a rapid, cheap, easily applicable, accepted and highly reliable diagnostic method which is increasingly used in routine diagnostic.

Mass spectrometry detects the mass to charge ratio (m/z) of proteins or protein fragments of a bioanalyte in a few minutes. Therefore, bacteria or bacteria cell extracts are transferred onto single spots of a so-called target. Afterwards, the spots are covered with a matrix. The target with the dried spots is transferred into the mass analyzer (Figure 4). In the analyzer short pulses of laser flashs on the target ionizes the matrix molecules and the bacteria or bacteria cell extracts molecules are ionized by proton transfer from the matrix. After that, the ions are accelerated in an electronic field in a vacuum. During their flight the ions are separated by their mass to charge ratio because small ions reach the detector before large ones. The small ions have a

smaller time of flight. Each detected protein or protein fragment is characterized by its molecular mass (m), charge (z), the ratio of both and its relative signal intensity. The results are displayed as mass spectra with bioanalyte specific peak profiles in the connected data analysis software. This way, the method provides a unique fingerprint of the analyzed microorganism (whole bacteria cell method) or bacterial cell extract (protein extraction method) which can be used for bacterial identification by comparison of the obtained bacteria specific spectra profile against a deposited reference database (Jackson *et al.*, 2005, Seng *et al.*, 2009).

Furthermore, there are attempts to use these fingerprint profiles for the differentiation of MRSA and MSSA as well as the main SA strain lineages (CC and STs) (Josten *et al.*, 2013, Wolters *et al.*, 2011).

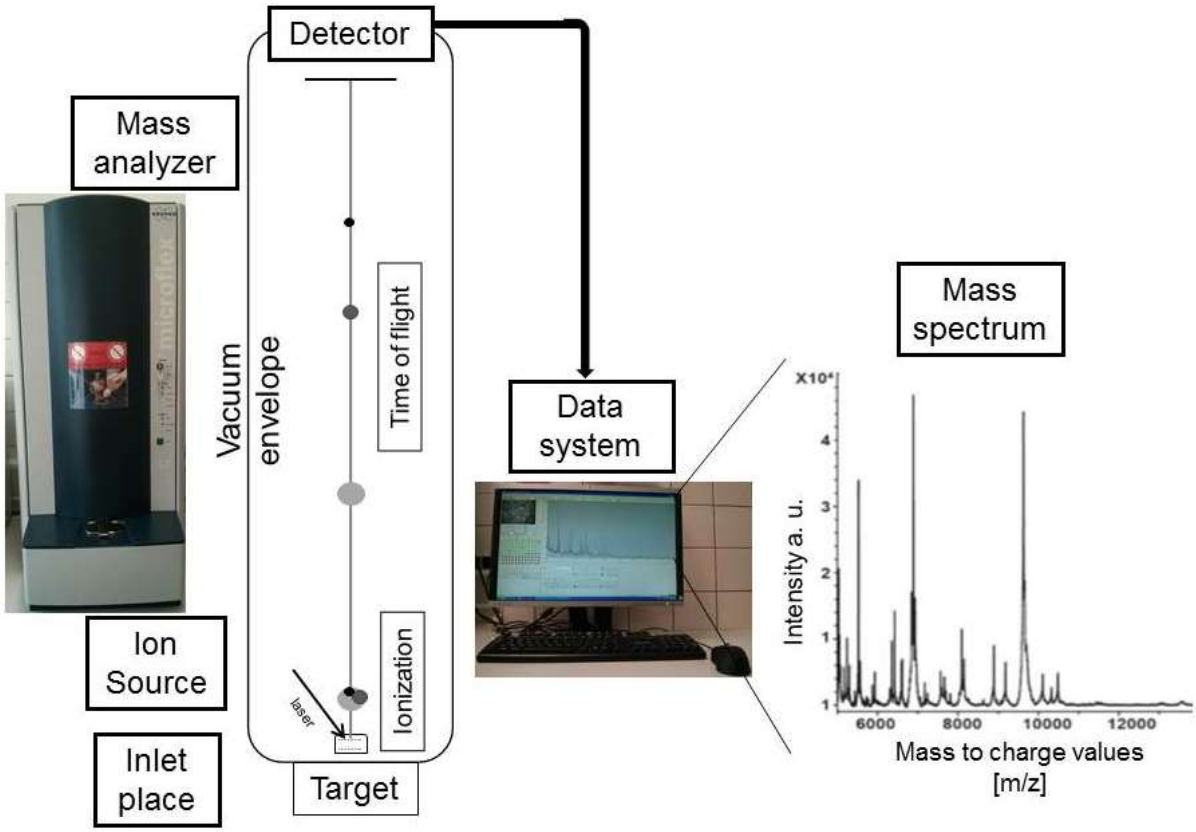


Figure 4: **Scheme of the principle function of the matrix-assisted laser desorption time of flight.** A bacterial colony or cell extract is applied to a target and overlaid with a matrix. The target is inserted into the vacuum envelope. There, a laser ionizes the bacterial molecules which are separated on their way from the target to the detector with respect to their mass and charge. The data from the detector are transferred to the software, so the analysis can be followed in real time.

2.2.3 SA epidemiology

In the early 1940's penicillin was introduced as new antibiotic treatment option. It leads to an improvement of the prognosis of patients with staphylococcal infections. The first penicillin resistant SA was found in 1942, first in hospitals and afterwards also in the community (Lowy, 2003, Rammelkamp & Maxon, 1942). In the 1950s the first reports of large SA outbreaks at maternity wards hospitals were published (Barber & McCartney, 1952). Penicillin resistance rapidly increased and in the late 1960s, about 80% of all staphylococcal isolates were penicillin resistant, regardless of their origin (i.e. hospital- or community-associated). This pattern of resistance, first emerging in hospitals and then spreading to the community, is a pattern recurring with each new wave of antimicrobial resistance (Chambers, 2001).

In 1961, methicillin was introduced as first of the semisynthetic penicillinase-resistant penicillin's. Shortly after the introduction of methicillin the first resistant isolates were identified (Jevons, 1961). In the 1960's the first wave of clonal dissemination of SA with different locally prevalent phage types were reported and were responsible for many infections (Jessen *et al.*, 1969, Parker *et al.*, 1970). The MRSA infections were primarily caused by one SA clone of the phage type 83A, today also known as the archaic MRSA clone ST250 (Aires-de-Sousa & De Lencastre, 2003, Enright *et al.*, 2002, Robinson & Enright, 2003). This so-called "archaic" MRSA clone circulated in hospitals only in some parts of the world as in UK, Denmark and Australia, but the rest of the world was largely unaffected (Crisóstomo *et al*., 2001, Jessen *et al.*, 1969, Rountree & Beard, 1968). The ST250 MRSA descends from the MSSA ST250 genotype, which probably evolved from the MSSA ST8 type (Deurenberg *et al.*, 2009). This first archaic MRSA clone contained the SCC*mec*I (Chambers & De Leo, 2009).

In the mid to the late 1970s, novel pandemic lineages of MRSA emerged, replacing the archaic MRSA clone ST250: ST1, ST8, ST22, ST30 and ST45 (Aires-de-Sousa & De Lencastre, 2003, Robinson & Enright, 2003). These MRSA clones appearing in the mid to late 1970 carried the new SSC*mec* allotypes, SCC*mec*II and SCC*mec*III. This was the beginning of wave II. They started the worldwide pandemic of MRSA in hospitals that is still ongoing (Aires-de-Sousa & De Lencastre, 2003, Robinson & Enright, 2003). Although MRSA were distributed worldwide, they were still restricted to hospitals and other health care institutions.

While the hospital environment still remains a risk factor for SA carriage and nosocomial infections, an increasing number of SA infections is caused by community-associated (CA) MRSA strains (Chuang & Huang, 2013, Otter & French, 2010). This means, in the last two decades started the third wave of MRSA with MRSA invasion into the community as described before for the penicillin resistance. The first cases of CA-MRSA infections in patients arising in the community were published in 1982. However, CA-MRSA were at this time disputed as several of the patients were drug addicts (Saravolatz *et al*., 1982). The earliest reports on CA-MRSA infections were from Australia. Since then, CA-MRSA popped up almost simultaneously in different parts of the world, spread around the globe with different strain types and have become a major health concern, predominantly in the USA (Figure 5). The next and certain emergence of MRSA strains in the community came in the late 1980s in Australian Aboriginals in remote Western Australian communities without previous history of hospitalization (Udo *et al.*, 1993). However, the real awareness came in 1999 with the MMWR (morbidity and mortality weekly report) of four pediatric deaths caused by CA-MRSA infections in Minnesota and North Dakota, USA (MMWR, 1999).

In the USA, CA-MRSA infections were first caused by the USA400 (ST1) strain MW2. USA400 was subsequently replaced by USA300 (ST8), which is now the predominant CA-MRSA clone in the United States (Chua *et al.*, 2011, DeLeo *et al.*, 2010, Takano *et al.*, 2008). In other parts of the world other ST's have dominated. In Europe ST80, which belongs to the CC80 clone is predominant in Europe. Another predominant ST in another part of the world is ST59 belonging to the CC59 clone, being the main sequence type in Taiwan. ST30 belonging to the CC30 clone predominates in Eastern Australia. In Germany the CA-MRSA are more heterogenous and are reported to range between 1% to 2%. The German CA-MRSA are mostly of CC5, CC8, CC30 and CC80 (Köck *et al.*, 2010, Robert Koch Institut, 2018).

Staphylococcal strains, which are prevalent in hospital settings, are so-called hospital-associated (HA) MRSA. HA-MRSA is represented in industrialized countries as Germany by a small number of clones such as CC5, CC8, CC22, CC30 and CC45 which rarely cause infections outside the health-care setting (Chambers & Deleo, 2009, Köck *et al*., 2010). During this second wave of MRSA, MRSA were reported in Germany for the first time in the 1970/1980s (Lenz *et al.*, 1988, Witte *et al.*, 1986). In Germany four major MRSA lineages dominates: t002/ ST5-MRSA-II and t003/ ST225-

MRSA-II (Rhine-Hesse epidemic clone), the Northern-German clone t008/ ST8-MRSA-IV and t032/ ST22-MRSA-IV (Barnim-epidemic-strain) (Cuny & Layer, 2011, Meyer *et al*., 2014).

2.2.3.1 Healthcare-, community- and livestock- associated SA

Division of SA is important to understand the different subgroups. Today the problem is that it is not directly obvious if a MRSA infection is caused by HA- or CA-MRSA. HA-MRSA are characterized by their resistance to β-lactam antibiotics as well as to other classes of antibiotics (Berger-Bachi & Rohrer, 2002) and they are predominantly responsible for infections in hospitalized patients of very young or higher age with typical risk factors. Typical HA-MRSA risk factors are history of hospitalization, residence in a nursing home, skin lesion, hemodialysis as well as indwelling catheters and inserted or implanted foreign bodies (Schaumburg *et al*., 2012).

CA-MRSA strains are the cause of infections in otherwise healthy and young people (Herold *et al.*, 1998) without any typical risk factors. Phenotypic and molecular analysis showed that CA-MRSA differ from the major circulating MRSA clones and are different from country to country. The CA-MRSA strains are not closely related with common endemic hospital strains. They are more virulent and typically characterized by their susceptibility to several non β-lactam antibiotics which hospital strains are usually resistant to (David & Daum, 2010). Common features of outbreaks of CA-MRSA comprise close contact, poor hygienic conditions, shared equipment, high risk of superficial skin abrasions, and lack of access to medical care to treat infections. Moreover, more than 75% of CA-MRSA carry the genes encoding for PVL (Daum *et al.*, 2002, Naimi *et al.*, 2003). PVL is associated with chronic or recurrent skin and soft tissue infections (SSTI) and in rare cases also with severe life-threatening diseases such as necrotizing fasciitis and pneumonia (Lina *et al.*, 1999).

Thus, clear definitions are required to distinguish between HA- or CA-MRSA infections. However, there are several CA-MRSA definitions based on an epidemiologic approach, antibiotic susceptibility patterns and molecular typing which have been thought to distinguish CA-MRSA from HA-MRSA (Witte, 2009). Different definitions, especially based on the last-mentioned characteristics developed in parallel, based on the characteristics of the locally predominant strains.

An epidemiological based definition created from the centers for disease control and prevention is that HA-MRSA infections are all infections that occur after 48 hours following the patient's admission to hospital and/or which fulfill one of the following HA-MRSA risk factors: hemodialysis, surgery, residence in a long-term care facility or antibiotic treatment during the last year, presence of a permanent catheter or percutaneous device or previous MRSA isolation, care in ICU's, and prolonged antibiotic treatment. Conversely, CA-MRSA infection is defined as any infection diagnosed in an outpatient within 24 or 48 hours of hospital admission which doesn't fulfill any of the HA-MRSA risk factors.

Till today, CA-MRSA outbreaks have been reported all over the world (DeLeo *et al.*, 2010, Figure 5).

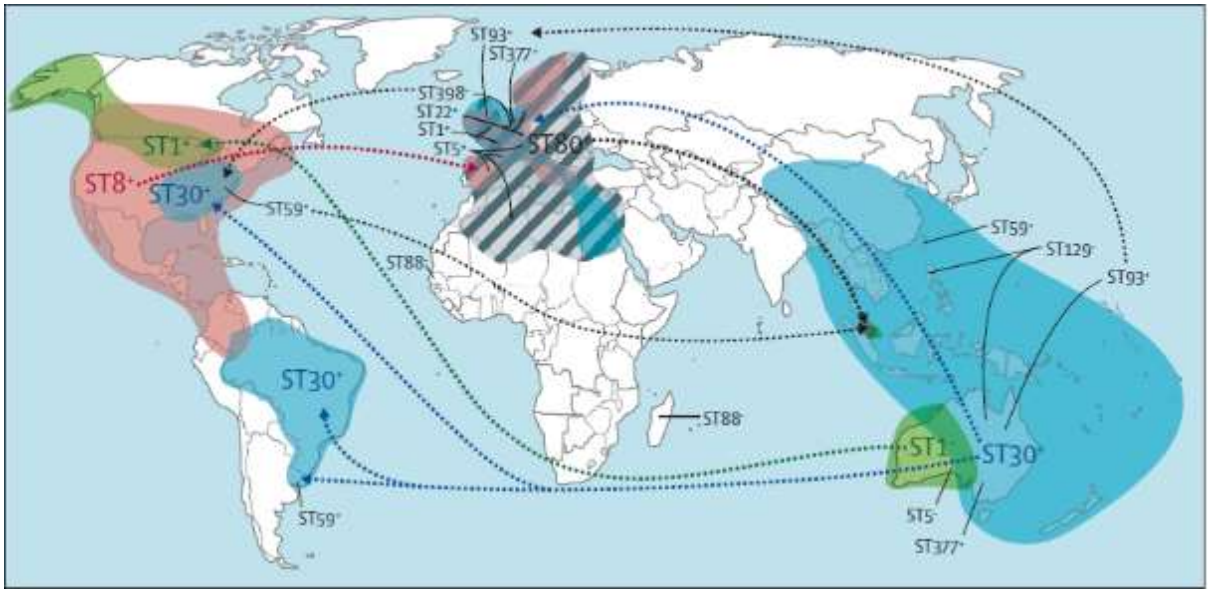


Figure 5: **Global distribution of community-associated (CA) methicillin-resistant *Staphylococcoccus aureus* (MRSA) sequence types** (Figure copy of De Leo *et al*., 2010). The dotted lines represent possible dissemination routes of CA-MRSA strains. Colored regions indicate the estimated distribution area where infections caused by the main strains are known (ST1-green, ST8-red, ST30-blue, ST80-grey hatched). "+" indicates PVL positive strains, "-" indicates PVL negative strains, "±" indicates a combination of PVL positive and negative strains.

Unfortunately, studies including data on CA-MSSA are rare, although they may constitute a major factor in a variety of infections, especially in developing countries.

More recently, since the end of the 1990s, MRSA are also described to be responsible for a range of infections in livestock pigs, cows and companion animals such as dogs and horses (Huijsdens *et al.*, 2006, Weese *et al.*, 2005). Colonization of livestock animals was first described in The Netherlands and France. Since then SA and

especially MRSA become a major challenge on farms and the livestock environment (van Cleef *et al.*, 2011, Köck *et al.*, 2013). Zoonotic transmissions from livestock to farmers and people with close contact to livestock as well as a transmission from livestock workers to their family members are well known. In The Netherlands the number of LA-MRSA isolated from people increased up to 30% of all isolated MRSA and history of livestock contact was included as risk factor to the Dutch screening guidelines (Bosch & Schouls, 2015). The typical livestock MRSA strains belong to the most predominant clonal complexes CC398, CC9 and CC97. MRSA CC398 is responsible for about 2% of all human MRSA infections such as bacteremia, pneumonia and wound infections especially in regions with a high density of livestock.

2.2.3.2 SA in Sub-Saharan Africa

To date, most epidemiological studies on SA were carried out in industrialized countries, although SA also plays an important role as a pathogen in developing countries in Sub-Saharan Africa. In comparison to typically diagnosed tropical diseases such as tuberculosis, malaria and parasite-associated diseases, SA infections are considered neglected tropical diseases. They are frequently underdiagnosed (Herrmann *et al.*, 2013a), probably due to limited availability of microbiology laboratories in these countries. In consequence, there is a lack of molecular, clinical and epidemiological data for this large region. This results in empirical therapy strategies which might entail treatment errors, especially in septic patients as reflected by the high infant mortality rate through bacteremia (Berkley *et al.*, 2005, Mulholland & Adegbola, 2005). Moreover, a concurrence of SA infections with HIV infections or malnutrition has been shown (Kraef *et al.*, 2015, Sigauque *et al.*, 2009, Talbert *et al.*, 2010) as well as a high incidence of bacteremia in African countries, with the highest fatality rates for SA infections (Anah *et al.*, 2008, Mugalu *et al.*, 2006, Obaro *et al.*, 2011).

Most of the African studies are from South African or North African countries such as Marocco, Tunesia, Algeria and Egypt. In contrast, only a limited number of publications is available on SA prevalence, genotypes and prevalence of virulence factors such as PVL (Ako-Nai *et al.*, 1991, Shittu *et al.*, 2015, Phaku *et al.*, 2016), especially in Sub-Saharan Africa. A recent overview article showed that data about MRSA were only available for 15 out of the 54 African countries (Abdulgader *et al.*, 2015). Studies with

focus on CA-MRSA or CA-MSSA in African countries comprise only small isolate numbers of in average up to 150 isolates (Enany *et al.*, 2010, Ramdani-Bouguessa *et al.*, 2006, Ruimy *et al.*, 2008).

Moreover, Schaumburg et al. showed recently that the major African MSSA clones are heterogenously distributed and that ST5 and ST8 MRSA as well as MSSA strains are geographically different distributed, despite the fact that they have the similar genetic background (Schaumburg *et al.*, 2014).

Overall, the predominant strains from the North to the South of Africa are different (Figure 6). ST80-MRSA is prevalent in North Africa with resistance profile differences between the Maghreb and Egypt (Ben Nejma *et al.*, 2013, Djoudi *et al.*, 2013). In comparison, ST8-MSSA is common in Maghreb and Nigeria (Breurec *et al.*, 2011) while ST5-MSSA and ST15-MSSA are more prevalent in West Africa including Cameroon (Breurec *et al.*, 2011, Ghebremedhin *et al.*, 2009). Further, ST88-MRSA is common in West-, Central- and East-Africa. Additionally, ST5-MRSA, ST30, ST121 and ST152 could frequently be detected in West- and Central-Africa while ST8-MRSA is abundant in Central- and South Africa. In South Africa the HA epidemic clones ST22, also called EMRSA-15, and ST36, also called EMRSA-16, are typical STs/CCs (van Rensburg *et al.*, 2010, Oosthuysen *et al.*, 2014). Overall, CC5 as HA strain lineage was the predominant strain in Africa (Abdulgader *et al.*, 2015).

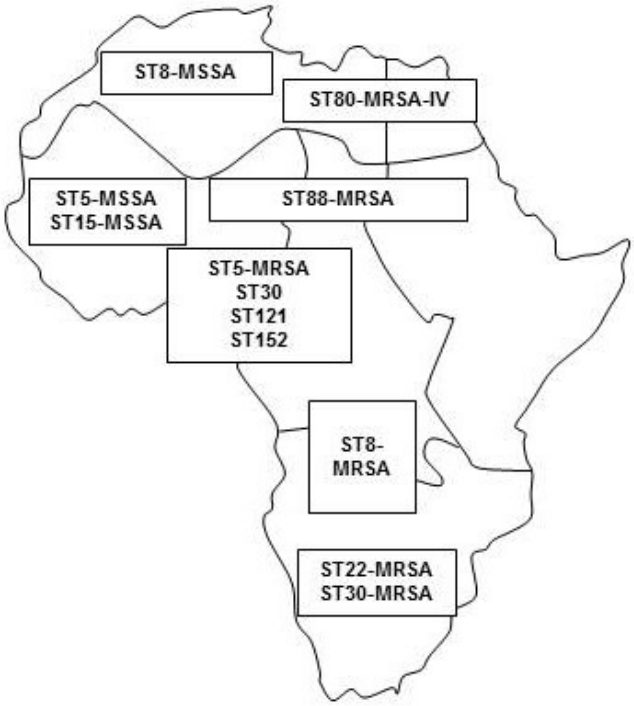


Figure 6: **Distribution of the predominant sequence types in Africa.** Text fields show the predominant sequence types (ST) in the different African regions such as the Maghreb region and Egypt in North Africa, West-, Central Africa as well as in South Africa.

Taken together, only limited data are available on SA distribution and epidemiology, especially for Sub-Saharan Africa. A more detailed knowledge would help to optimize the patients' treatment and surveillance of antibiotic resistance development.

2.3 Objectives of this thesis

The objective of this thesis was a comparison of isolate characteristics of community-associated SA isolates collected from different geographies i.e. in four different African geographic regions and in three federal German states, as well as of isolates obtained by screening (nasal swabs) and from clinical origin. Furthermore, to investigate whether analysis of the gene content / mass spectra could be used to predict defined epidemiological questions and to evaluate different bioinformatic software tools for DNA MA data analysis.

The thesis consists of five different evaluations:

- Evaluation of the potential of the DNA MA array to subtype SA of nasal isolates collected within a 4 week-interval during state-wide routine hospital admission prevalence screenings (APS) in the federal state of Saarland in comparison to *spa*-typing (study A) (Ruffing *et al.*, PLoS One, 2012) and
- Comparison of a standardized approach for DNA MA data analysis by three known bioinformatic tools (study A): splits graph, hierarchical agglomerative clustering (HAC), and principal component analysis (PCA) - methods which were frequently applied in phylogenetic analysis or in genome-wide association studies (Ruffing *et al.*, PLoS One, 2012).
- Evaluation of the suitability of the DNA MA for the analysis of African isolates (study B) (Shittu *et al.*, Frontiers in Microbiology, 2015).
- Determination of the potential of DNA MA to address epidemiological questions, to predict the virulence potential of SA strains and to link specific genotypes with infection types (study C) (Ruffing *et al.*, Scientific Reports, 2017).
- Evaluation of mass spectrometry for SA subtyping with regard to methicillin resistance geographic origin of the isolates, *spa*-type and CC/ST.

3 Material and Methods

3.1 Material

3.1.1 Isolate collection

This work represents a synopsis of four studies (A-D) with the overall aim of epidemiological comparison of SA isolates from four different African geographic regions and three federal German states as well as a comparison of isolates obtained by screening (nasal swabs) and from clinical origin. Study A was a small study based on the previously performed admission prevalence screening (APS) study of the federal state of Saarland (section 4.1) (Herrmann *et al.*, 2013b) funded by the German Ministry of Health and the State of Saarland Government and performed by the State wide German network for the control of MRSA, MRSAarNet (www.mrsaar.net), actually renamed in Infectio^{Saar}. The aim of study A was the comparison of MRSA and MSSA and in advance of study C for evaluation of how to perform the DNA MA data analysis. Study B was an independent study for evaluation of the applicability of the chosen DNA MA for African isolates in advance of study C (Shittu *et al.*, 2015). Study C was a part of the prospective cohort study of the African-German network on staphylococci and staphylococcal diseases (African-German StaphNet) during the collaborative research project “Infection biology and epidemiology of Staphylococci and Staphylococcal diseases in Africa” (DFG, PAK 296). Study D was an independent study for evaluation of the applicability of mass spectrometry for epidemiologic analysis of SA isolates.

The definition of criteria for the isolate inclusion in the studies A to C (e.g. questionnaires, case related forms) were not part of this work.

1. Study A: Nasal isolates of all patients with nasal SA colonization admitted to any of the 24 hospitals in the federal state of Saarland were collected within a 4 week-interval during state-wide routine hospital admission prevalence screenings (APS) (Herrmann *et al.*, 2013b). For each patient clinical data were also obtained (Herrmann *et al.*, 2013b). From 436 MRSA overall, a collection of 46 MRSA isolates from the university clinic of Saarland (one per patient) with assignable clinical data were identified. Additionally, 46 MSSA isolates from patients matched according to the following clinical data of the MRSA patients: gender,

age (<70 vs. ≥70 years), previous hospitalization during the last 6 months were also selected. The study was approved by the ethic commission of Saarland (registration # 127 / 10).

2. Study B: Between March 2009 and April 2010, 52 MSSA (3 nasal and 49 clinical isolates of wound infections) were obtained from samples collected as part of surveillance activities in microbiology laboratories of six health care institutions in Nigeria (convenience sample) (Ado-Ekiti, Ile-Ife, Osogbo, Lagos, Ibadan, Maiduguri) or were included from a previous study (Shittu *et al*., 2015).
3. Study C: A total of 1200 SA isolates were sampled in the prospective cohort study of the African-German network on staphylococci and staphylococcal diseases (African-German StaphNet, DFG PAK 296). From each of six medical centers/ hospitals, three African (Lambaréné, Gabon; Dar-es-Salaam, Tanzania; Ifakara, Mozambique) and three German study sites (Homburg, Saarland; Freiburg, Baden-Württemberg; Münster, North Rhine-Westphalia), 100 consecutive SA isolates from clinical specimens from patients admitted to the medical centers/ hospitals were collected. All specimens were taken at admission or within 24 hours. Only one isolate per patient were included. Furthermore, each study site contributed with a collection of 100 isolates of randomly selected healthy volunteers without hospital employment or other close contact to the health care system. Clinical information as e.g. patient history, clinical presentation risk factors, antibiotic treatment, surgical treatment from all participants and healthy volunteers were further collected by case report forms (CRF) (Appendix, section 6.2). One of the African study sites served as primary care center for children, which was not the case for the German study sites leading to some misbalance in the patient population.

   The 1200 SA isolates were given a unique strain identification code containing the study site (LG= Lambaréné, Gabon, MM= Manhica, Mozambique, IT= Ifakara, Tanzania, HS = Homburg, Saarland, MW= Münster, Westfalen, FR= Freiburg, Baden-Württemberg), the strain origin (N= nasal, B= blood culture, O= wound infection) and a running number e. g. 1-100.

   Ethical approvals for this study was given by the Ministry of Health and Social Welfare of Tanzania, the Institutional Ethics Committee of the Medical Research Unit of the International Foundation of the Albert Schweitzer Hospital

(Lambaréné, Gabon), the Comité Nationale de Bioética para a Saudé (Manhiça, Mozambique), the Ethics Committee of the Medical Association and Medical Department of the Westfälische Wilhelms-University (Münster, North Rhine-Westfalia), the Ethics Committee of the Albert-Ludwigs-University (Freiburg, Baden-Wuerttemberg), and the Ethics Committee of Saarland (Homburg, Saarland). A written informed consent was obtained from all study subjects or their legal guardians.

4. Study D: 219 SA isolates (51 MRSA, 168 MSSA) of study C (n=127, section 4.3) and all isolates of study A (n=92, section 4.1) were used to evaluate the potential of subtyping of SA by MALDI-TOF.

3.1.2 List of all analyzed SA isolates and DNA microarray data

All SA isolates of this work (Study A to Study D) were analyzed by DNA MA. The list of all DNA MA hybridization signals and more detailed analysis of the DNA MA data of study C are shown in the attached electronic excel file (EEF) DNA MA SA data, doctor thesis Ruffing U.

Additional tables in the EEF indicate the percentages of the number of isolates with positive hybridization signals for all DNA MA targets of study C with comparison of i) African versus German isolates (sheet A2), ii) clinical versus nasal isolates (sheet A3), the iii) clinical isolates of African versus German origin (sheet A4), iv) nasal isolates of African versus German origin (sheet A5), and clinical versus nasal isolates in the v) African isolate group (sheet A6), vi) German isolate group (sheet A7) as well as vii) analysis of the absolute numbers of positive DNA MA targets of the African-German comparison study, overall and in the ten most predominant CCs (sheet A8).

3.1.3 Chemicals and consumables

3.1.3.1 Chemicals, media and consumables

Table 1: **List of the used Chemicals, media and consumables**

Chemicals, media, consumables	Manufacturer
Identibac® *S. aureus* Genotyping Kit	Alere GmbH, Köln/Germany
MagAttract HMW DNA Kit	Qiagen GmbH, Hilden/Germany
Qiagen DNeasy® Blood & Tissue Kit (250)	Qiagen GmbH, Hilden/Germany
Acetonitril	Sigma Aldrich, St. Louis/USA
Blue loading buffer	PEQLAB, Erlangen/Germany
Ethanol	Sigma Aldrich, St. Louis/USA
ROTISOLV® HPLC gradient grade water	Sigma Aldrich, St. Louis/USA
100bp DNA ladder	PEQLAB, Erlangen/Germany
HCCA portioned (matrix)	Bruker, Bremen/Germany
Nuclease-Free Water	Promega, Mannheim/Germany
Pastorex™ Latex Test	Bio-Rad, München/Germany
peqGOLD Hot Start-Mix Y	PEQLAB, Erlangen/Germany
QA-Agarose TM	MP Biomedicals, Illkirch/France
SybrGold	Invitrogen, Oregon/USA
Trifluoroacetic acid	Sigma Aldrich, St. Louis/USA
Tris Borat EDTA buffer (10x)	AppliChem, Darmstadt/Germany
Trypticase™ Soy Agar II (TSA) + 5% sheep	Becton Dickinson, Heidelberg/Germany
USB® ExoSAP-IT	Affymetrix, Inc., Ohio/USA
Cryobank™	Mast Diagnostica GmbH, Reinfeld/Germany
eSwab	COPAN Italia S.p.A., Brescia/Italy
inoculation loops	Sarstedt, Nürmbrecht/Germany
pipette tips	Sarstedt, Nürmbrecht/Germany

3.1.3.2 Oligonucleotids

The *spa*-typing primers *spa*-1113f (TAA AGA CGA TCC TTC GGT GAG C) and *spa*-1514r (CAG CAG TAG TGC CGT TTG CTT) were purchased from biomers.net GmbH (Ulm, Germany). The stock solution of the oligonucleotides was prepared with A. dest. according to the manufacturer's instructions to obtain a concentration of 100 $^{pmol}/_{µl.}$ The *spa*-typing master mix consists of 10µl A. dest.,12,5 µl peqGold Hot-Start Mix Y, 0,25µl forward primer and 0,25µl revers primer. The aliquoted stock solution was stored

at -20°C.

3.1.4 Instruments and Software

Table 2: Used instruments and softwares

Instrument/ Software	Manufacturer/Developer
ArrayMate Reader	Alere GmbH, Köln/Germany
Biometra UNO Thermoblock	Biometra GmbH, Göttingen/Germany
Brutschrank	Heraeus, Langenselbold/Germany
FlexCycler	Analytik Jena AG, Jena/Germany
Gel electrophoresis chamber	Bio-Rad, München/Germany
Illumina MiSeq	Illumina, Cambridge/UK
Microflex LT spectrometer	Bruker, Bremen/Germany
MIKRO 20	Hettich Zentrifugen, Tuttlingen/Germany
Minizentrifuge	Eppendorf AG, Hamburg/Germany
Molecular Imager® GelDoc XR+	Bio-Rad, München/Germany
MSP 96 target polished steel BC	Bruker, Bremen/Germany
Nano Drop™ 2000	PEQLAB Biotechnologie GmbH, Erlangen/Germany
Power Pac 1000	Bio-Rad, München/Germany
Thermomixer comfort	Eppendorf AG, Hamburg/Germany
VITEK® 2	BioMérieux, Marcy-l'Etoile/France
Vortex Genie	Scientific Industries, New York/USA
Acrobat Reader	Adobe, Dublin/lireland
Bioconductor packages	(Gentleman *et al.*, 2004)
BioNumerics	Applied Maths, Sint Martens Latem/ Belgium
Biotyper	Bruker, Bremen/Germany
Flexcontrol	Bruker, Bremen/Germany
QuickCalcs	GraphPad Software Inc., California/USA
IconoClust	Alere Technologies GmbH, Jena/Germany
ImageLab™	Bio-Rad, München/Germany
NanoDrop 2000/2000c	PEQLAB Biotechnologie GmbH, Erlangen/Germany
NCBI BLAST package	(Camacho *et al.*, 2009)
R Studio	R Studio Team, 2015 (R Studio: Integrated Development for R. RStudio, Inx. Boston, MA; http://www.rstudio.com/)
Ridom StaphType	Ridom GmbH, Münster/Germany
Ridom™ SeqSphere+	Ridom GmbH, Münster/Germany
SplitsTree	(Huson & Bryant, 2006)
SPSS	IBM, Germany
Stats	(Quackenbush, 2001)
Translate (ExPasy)	(Gasteiger *et al.*, 2003)

Instrument/ Software	Manufacturer/Developer
Word/Excel 2010	Microsoft, Redmond/USA

3.2 Methods

3.2.1 Culture, basic characterization and storage

All isolates were cultured on TSA sheep blood agar plates according to standard procedures and incubated at 37°C. The species' identity of all newly collected SA isolates was confirmed by MALDI-TOF of pure cultures. The presence of the *mecA* gene was determined by DNA MA by the presence of the *mecA* gene and the SCC*mec* cassette (see section 2.1.1). All isolates were stored in Cryobank tubes™ at -80°C (Mast Diagnostica GmbH, Reinfeld/Germany).

3.2.2 DNA microarray

3.2.2.1 DNA-preparation

The DNA preparation was performed according to the manufacturer's instructions of Qiagen in combination with additional information from the Identibac® *S. aureus* Genotyping-Kit.

For every single isolate, the reaction tube (A2) of the Identibac® *S. aureus* Genotyping-Kit containing a lysis enhancer was centrifuged for 1 min at 13226xg to concentrate the pellet at the bottom of the tube. In each reaction tube (A2), 200 µl of the Identibac® *S. aureus* Genotyping-Kit lysis buffer were transferred and incubated at RT for 5 min to activate the lysis enhancer. Three colonies of the overnight culture were inoculated in the suspension. The suspension was then incubated 1 h at 37°C and 550rpm. Afterwards, 200 µl of lysis buffer AL and 40 µl proteinase K were added. This suspension was mixed and incubated for 1 h at 56°C and 550rpm. After the incubation, the tubes were briefly centrifuged and 200 µl of 100% ethanol were added per tube. The whole suspension was transferred onto the provided Qiagen Kit column and centrifuged at 6000xg for 1 min. The flow-through was discarded and the column transferred onto a new tube. Next, different washing procedures followed, beginning with washing of the column with 500 µl of the first washing buffer (AW1) for 1 min at 6000xg. The flow-through was discarded. Then 500 µl of the second washing solution

(AW2) was applied to the column on a new tube and centrifuged at 13226xg for 3 min. Again, the flow-through was discarded and the column was centrifuged for 1 min at 13226xg. The flow-through was discarded and the column was transferred onto a fresh reaction tube. Finally, 40 µl of elution buffer (AE) were pipetted on the column, incubated for 3 min at room temperature and centrifuged for 1 min at 6000xg to elute the DNA.

3.2.2.2 DNA concentration measurement

The quality and quantity of the DNA was spectrophotometrically measured with the NanoDrop2000 (Quiagen GmbH, Hilden, Germany) and the previously used elution buffer as blank value. For the DNA MA analysis, a DNA concentration of at least 80ng/µl and a purity ratio (OD_{260}/OD_{280}) of 1.8 to 2.0 were necessary.

3.2.2.3 DNA microarray hybridization

The original protocol of the manufacturer‘s instructions of the Identibac® *S. aureus* Genotyping-Kits without modifications was carried out (Manual Identibac® *S. aureus* Genotyping-Kit, 2015).

In a linear PCR, one antisense primer per target was used resulting in single stranded DNA amplificates. Additionally, the single stranded DNA was labeled with Biotin-d-UTP of the B1 Master Mix of the *S. aureus* Genotyping-Kit during the linear PCR with approximately 40-fold amplification (Manual Identibac® *S. aureus* Genotyping-Kit, 2015). A master mix of 4.9 µl of labelling buffer B1 and 0.1 µl of labelling enzyme B2 was prepared. For the linear PCR, a DNA concentration of 0.5-2 µg/ 5µl was prepared with RNase free H_2O, and 5µl of this DNA solution mixed with 5 µl of the Mastermix.

Linear PCR protocol:

95°C	5 min	initial denaturation
45 cycles:		
96°C	20 sec	denaturation
50°C	20 sec	annealing
72°C	30 sec	elongation
4°C	∞	cool down

Following the PCR, 90 µl of the hybridization buffer C1 were pipetted into the PCR reaction tubes while the ArrayStrips were rinsed with 200 µl of RNase-free H_2O. Afterwards, the ArrayStrips were incubated with 200 µl of C1 hybridization buffer at 55°C and 550rpm for 2 min. Then, the liquid was removed from the ArrayStrips and all amplificates were transferred onto the ArrayStrips. The ArrayStrips were closed with lids and incubated for 1 h at 55°C and 550rpm.

In parallel, for every isolate a suspension of 1 µl of the C3 horseradish-peroxidase-conjugate and 99 µl of the C4 conjugation buffer were prepared.
After the hybridization step, the ArrayStrips were cleared from the fluids and subsequently washed for three times with 200 µl of C2 washing buffer. Then 100 µl of the C3/C4 horseradish-peroxidase mix was transferred onto the ArrayStrips and incubated without caps for 10 min at 30°C and 550rpm. The C3/C4 horseradish-peroxidase mix was removed and the ArrayStrips were washed twice with 200 µl of the washing buffer C5. During this incubation step, the D1 horseradish-peroxdiase substrate was taken out of the fridge and equilibrated to RT and subsequently centrifuged for 1 min at 13226xg. After 10 min incubation, 100 µl of the D1 were pipetted into the Arrays and incubated at RT for 5 min in darkness. Finally, the whole fluid was removed and the ArrayStrips were measured in the ArrayMate-Reader. The Reader with the Iconoclust software automatically measures the grey signals in comparison to the background signals, interprets the signals according to predefined parameters and gives an exportable output for every single examined isolate.
The signals were interpreted as positive (=target present), negative (=target absent) or ambiguous. For the signal analysis, there were two ways: First ambiguous signals for genes without allelic variants signifies that there is no clear answer about presence or absence of this gene. The reason for this could be poor sample or signal quality, only low copy numbers of plasmids especially in the case of some resistance genes as *aacA-aphD* or gene/allele mutations which influence the hybridization. Second, for genes with allelic variants as e.g. *bbp*, *clfA*, *clfB*, *fnbB*, *set/ssl*, *isaB*, *mprF* and *isdA* the probe with the strongest signal value is regarded as positive, if it exceeded a defined breakpoint. All other allele-specific probes are then regarded as negative, if the signal is below the breakpoint and ambiguous if the signal lies in between the both breakpoints.

3.2.2.4 DNA microarray data analysis by bioinformatic methods

For DNA microarray data analysis five different bioinformatic methods splits graph, affinity propagation, principal component analysis (PCA), hierarchical agglomerative clustering (HAC) and silhouette plot for visualization and cluster analysis were applied. In advance, a latent factor model was applied for DNA MA data preparation for further data analysis by affinity propagation, PCA, HAC and silhouette plot analysis. DNA MA data preparation by latent factor model, affinity propagation, PCA, HAC and silhouette plot for visualization and cluster analysis in study A and study C were technically performed by cooperation partners.

3.2.2.4.1 Splits graph

Splits graphs were used to analyze and visualize the data obtained by *spa*-typing and DNA MA typing. Splits graphs were originally developed to present the evolutionary history of a set of taxa calculated from binary or tertiary sequence data whereby the tree branches have defined branch lengths representing the time of divergence. For the construction of a split network, aligned data sequences are required (Figure 9, 10). These aligned data sequences can be arranged into a two-dimensional matrix of size $N \times N$.

In our case the matrix was constructed from "distances" between N genomic sequences of SA strains. The distance matrix is a two-dimensional array of size $N \times N$, which contains the pairwise distances of a given set of N elements. Each distance between two sequences was set equal to the Euclidian distances between the tertiary sequence data encoding the present, absent or ambiguous signals of a DNA MA. Based on such a matrix the distances between the chosen characteristic can be calculated. In mathematics Euclidean distance is the straight-line distance between two points in Euclidean space. This means that similarity was measured based on the number of differing entries (hybridization signals). For simplicity, let us assume a scenario with 5 probes. Sample A could yield the following hybridization signal: (0 1 1 0 1) and sample B could yield: (1 1 0 0 1). These two samples differ in the first and third element (probe). Thus, their Euclidian distance is 2.

The distance between two isolate sequences is the sum of the splits, representing the differences which separates them. A split is a partition of the set of e.g. hybridization signal sequences into two disjoint groups. Splits are the elements of phylogenetic

trees. Each branch divides the set, of in this case sequences into a split separating the sequences of the one side of the branch to the sequences of the other side. The number of splits represented by the different branches in an unrooted phylogeny contains all the branching information of the phylogeny. This is a so-called set of splits.

In this work the distances are the sum of differences in the sequences i.e. the number of positions in the sequence where the hybridization signals are not identical. In this way, splits graphs visualize characteristic differences between the signals in larger data sets (Huson & Bryant, 2006).

Splits graphs in this work were constructed based on the neighbor-net which is one of the most frequently applied distance-based methods for phylogenetic analyses (Bryant & Moulton, 2004). It constructs splits networks from inferred distance matrices and applies an agglomerative process to create a phylogenetic tree (Huson & Bryant, 2006).

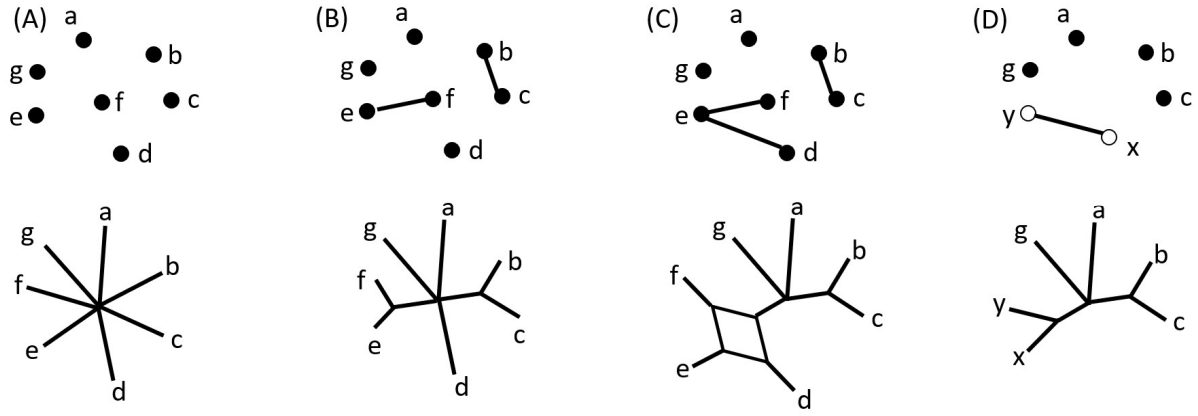


Figure 7: **Visualization of the agglomerative process of the neighbor-joining method**. (A) First each node represents a single hybridization signal sequence. (B) Based on the selection criterion b and c were identified as neighbors. Additionally, e and f were identified as neighbors, too. (C) E has been identified as a neighbor of d and of f. (D) Because e has two neighbors, a reduction is performed, replacing d, e, f by x, y. Figure adapted from Bryan & Moulton, *Molecular Biology and Evolution,* 2003.

The distance matrix was calculated based on the proportion of positions which were different between two sequences or hybridization patterns. In a binary alignment (i.e. an alignment with two different elements p=positive and n=negative), or tertiary alignment (i.e. an alignment with three different elements p =positive, n=negative, and a=ambiguous) every difference (e.g. a “p” instead of an “n” at the equivalent position in a column) leads to a split of the sequences. In the graph, one line splits up into two different lines at the point of difference in the alignment (Figure 8). This point of

separation of the line at the point of difference is named split in graphical presentations (Figure 8).

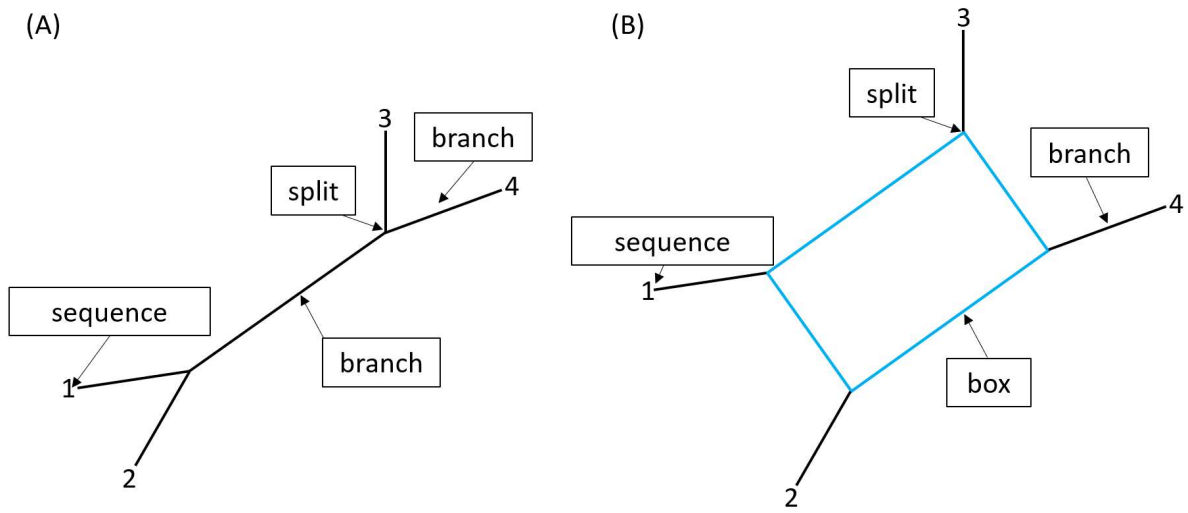


Figure 8: **Two splits graph without (A) and with (B) possible recombination event for four sequences as examples for explanation constructed based on the neighbor-joining method.** A box (highlighted in blue) represents a possible recombination event.

A difference means that there is i.e. a positive instead of a negative signal at the same position of two sequences in an alignment as shown in Figure 9 for gene 4 of isolates one and two.

gene/ isolate	1 2 3 4 5 6 7 8 9 10
1	a p n n p n a n p a
2	a p n p p n a n p a
3	a p n p p n a n p n
4	a p p p p n n n p a
5	a p n n p n n n p a
6	p n n n p n n n p a

Figure 9: **Presentation of a tertiary alignment.** An alignment with three different kinds of hybridization signals is shown, with “a” as symbol for an ambiguous hybridization signal, “p” for a positive hybridization signal and “n” for a negative signal. All splits in the alignment are highlighted by a bar. Columns represent different hybridization signals and rows represent different isolates.

As mentioned before, the evolutionary distance equals the sum of the length of the branches along the way from e.g. hybridization signal 1 to hybridization signal 4. If

there is no possible recombination event, splits are connected by the shortest direct connection line, named branch (Figure 8A). Possible events of recombinations are represented as boxes (Figure 8B, highlighted in blue). Graphically these are parallel lines in the splits graph.

In this work, splits graphs were applied for comparison of DNA MA analysis (study A) to *spa*-typing analysis as reference method routinely used in microbiology labs (section 4.1) and for analysis of DNA MA data of a small selection of Nigerian SA isolates (study B, section 4.2).

The splits graphs were created based on the DNA MA data organized as alignments as shown exemplary in Figure 9 and 10. For analyzing the DNA MA data by splits graphs, the hybridization signals (e.g. positive/negative hybridization signal of a gene/allele) were organized in a matrix. This means that the DNA MA hybridization signals of all analyzed SA isolates were aligned like SA sequences to obtain the required data matrix (Figure 10).

The SplitsTree software (Huson & Bryant, 2006) was used to analyze the similarities between the hybridization patterns of 46 MRSA and 46 MSSA isolates of study A (shown in section 4.1), and of 52 Nigerian MSSA isolates (study B, shown in section 4.2). The DNA MA results were imported into the SplitsTree software (Huson & Bryant, 2006) as alignments in a FASTA format and the DNA MA hybridization results were converted into a “pseudo” nucleotide code consisting of the three symbols “g” (negative hybridization signal), “a” (ambiguous) and “c” (positive hybridization signal) (Figure 10).

Figure 10: **Part of a tertiary alignment in a FASTA format of the DNA microarray hybridization sequences after conversion into a pseudo” nucleotide code consisting of the three symbols “g” (negative hybridization signal), “a” (ambiguous) and “c” (positive hybridization signal)**.

The splits graphs were constructed using SplitsTree versions 4.10 (Figure 15, 17) and

4.11.3 (Figure 20) with default settings (characters transformation, uncorrected P; distance transformation, neighbor-net; variance, ordinary least squares).

Additionally, splits graphs were automatically constructed by the software SplitsTree (Huson & Bryant, 2006) based on the *spa*-typing data of the 46 MRSA and 46 MSSA (Figure 14) and the 43 CC5 SA isolates (Figure 16) obtained from study A. First a cost/distance matrix (Figure 11) was calculated by the based upon repeat pattern (BURP) algorithm of the Ridom StaphType software (Ridom, Münster, Germany) based on the *spa*-typing data. This cost/distance matrix was transferred via an automatically link of the Ridom StaphType *spa*-typing software to SplitsTree.

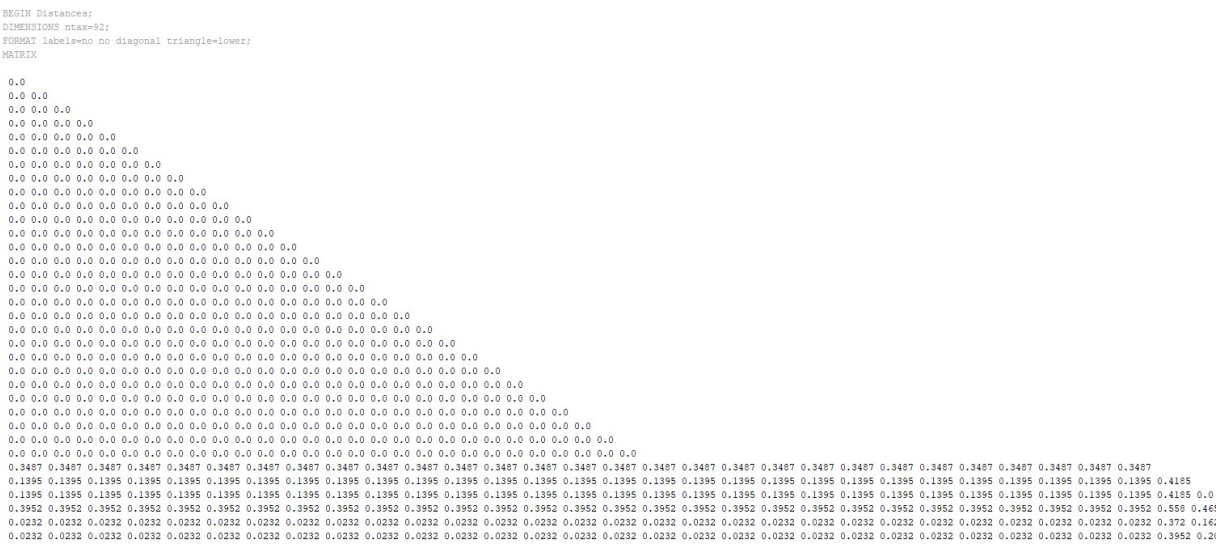


Figure 11: **Part of the cost/distance matrix calculated by the based upon repeat pattern (BURP) algorithm of the Ridom StaphType software (Ridom, Münster, Germany) based on *spa*-typing sequence data of 92 *Staphylococcus aureus* isolates.**

Next, the software SplitsTree calculated the splits and constructed the graph based on the computed export cost/distance matrix (Figure 12) from Ridom StaphType.

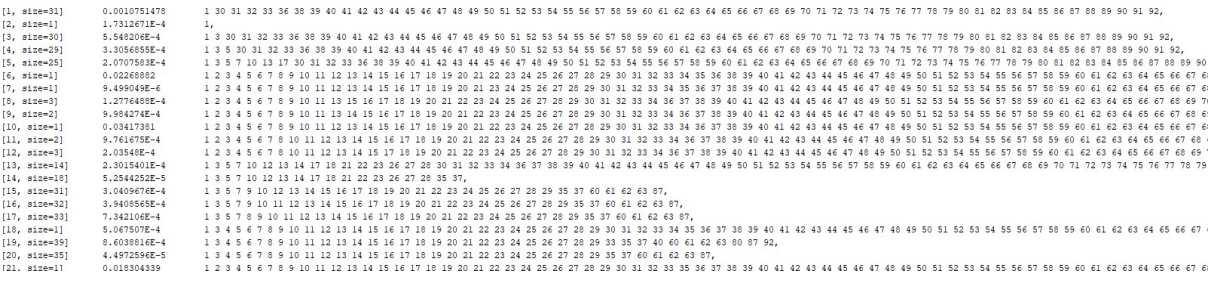


Figure 12: **Part of the calculation of the splits based on a previously computed cost/distance matrix**.

3.2.2.4.2 Imputing ambiguous data points with Latent factor model (LFM)

As described in section 3.2.2.3, the DNA MA hybridization pattern reader automatically

categorized the signals in “absent“, “present“ or “ambiguous” hybridization signal. In our case, "absent" entries were represented as 0 values and "present" entries as 1. "Ambiguous" values are those with ambiguous hybridization signals represented as “-“.

As described below, DNA MA data were processed with methods for dimensionality reduction such as affinity propagation, HAC and PCA. The aim was to identify those hybridization probes (genes) distinguishing rather homogeneous groups of samples. Before doing this, we had to deal with the “ambiguous” values that typically represent about 5% of the measured data points. One possible strategy of dealing with the ambiguous values is to replace these values by an intermediate value i.e. 0.5. However, this would mean that we would treat these values with a definite level of certainty which may be inappropriate because there may exist many different reasons why these values were returned as ambiguous. Another strategy is to filter out all hybridization probes (genes) or samples that contain at least one ambiguous value. However, in our case this would reduce the number of probes or samples to a fairly small number. Finally, we decided on a strategy whereby ambiguous values are replaced by binary “positive” or “negative” values based on a prediction model using Latent Factor Models (LFM) applied by the bioinformatic collaborator Ruslan Akulenko. In the area of data sciences, this type of task is referred to as “data imputation” and is typically solved by predictions based on the values of the remaining data.

LFM are popular methods in predicting missing data for example in image reconstruction, see e.g. (Candès & Recht, 2012, Koren *et al.*, 2009). The mathematical details of the LFM method are described by the collaborators (Akulenko *et al.*, 2016). Briefly, the LFM method is able to complete missing data points in a matrix by constructing two accessory matrices that, when multiplied with each other, closely match the known values of the original matrix. The accuracy of the LFM prediction was tested by random deletion of 5% of the (known) positive (1) and negative (0) entries. Next, these missing entries were predicted by LFM and the calculated results were compared with the original data. This reconstruction of known data had an accuracy of 97%. Thus, it can be assumed that also the predicted values that replace the ambiguous values will have a similar accuracy. As mentioned before, the 5% percentage of deleted entries corresponded with the number of ambiguous signals detected in the DNA MA experiments.

3.2.2.4.3 Affinity propagation

Affinity propagation was applied for DNA MA data analysis of study C. Affinity propagation is a sophisticated clustering algorithm (Frey & Dueck, 2007), which identifies exemplars and clusters. As with any other clustering method, a “cluster” contains a group of samples that are more homogeneous to each other than to the remaining samples. Exemplars are members of the given data set that are representatives for the identified clusters. For each cluster, affinity propagation outputs one exemplar. Affinity propagation uses a measure of similarity between pairs of data points as input (here the similarity of DNA MA hybridization profiles of the isolates). In contrast to other clustering algorithms affinity propagation does not need the number of clusters to be determined or to be estimated for starting the algorithm. The number of clusters is automatically determined by the algorithm in an optimal way.

3.2.2.4.4 Principal component analysis (PCA)

The Principal component analysis (PCA) is essentially a method for dimension reduction to extract relevant information from large complex data sets. The method was used in three studies of this work (study A, section 4.1, study C, section 4.3 and study D, section 4.4). In study A and C, it was applied for more detailed SA DNA MA data analysis. In study D it was applied for the evaluation of the subtyping potential of SA based on their total mass spectra.

For this, PCA transforms (projects) the original data onto a new set of uncorrelated, orthogonal basis vectors (principal components). The first principal component (PC) is directed along the direction with largest variance in the data, the second PC is orthogonal to the first PC and points along the direction of second-largest variance and so forth. In our case, the dimensionality of the original data is equal to the number of DNA MA target sequences. The variance in this data reflects the differences between the binary hybridization signals of the probes. An important property of PCA is that – by way of construction – the first few PCs describe most of the variance in the data (Marques *et al.*, 2008). Thus, the full data set can be projected on the first two or three PCs and visualized in a 2D- (e.g. Figure 18, 24) or 3D-plot (e.g. Figure 25, 26) without losing too much information. This data compression captures most of the variation among the samples and visualizes similarities and differences in a readily accessible fashion. In our case, the PCs will point along those DNA MA probes that show the

largest variation of hybridization among the samples. Samples with similar DNA MA hybridization profiles are arranged closer to each other in the 2D/3D plot compared to samples with larger differences (Caddick *et al.*, 2006).

For the PCA, in study A and study C the data were preprocessed basically as for the cluster dendrogram. Again, all probe IDs containing ambiguous results were replaced by LFM. This was done for all DNA MA data of the 92 isolates of study A and for all 1200 African and German SA isolates of study C. The main PC vectors identified by PCA have the largest coefficients in the fields corresponding to those MA probes were the samples differ most. In this case, we focused our analysis on projections along the first two PC. The PCs were computed by the R function prcomp in the package „stats“.

3.2.2.4.5 Hierarchical agglomerative clustering dendrogram (HAC)

HAC was applied for DNA MA data analysis of study A (Figure 19). This method constructs a dendrogram, a branching tree diagram which represents the hierarchy of clusters based on the degree of similarity or number of shared characteristics of the data. The method was applied in our study with 43 CC5 SA isolates of study A using DNA MA hybridization data (Figure 19). The distance along the tree from one element to the next one represents the relative degree of similarity, while short distances representing high similarities. In this bottom-up approach every character (e.g. hybridization signal or isolate) represents an individual cluster at the beginning. Then, at each step the most similar elements (e.g. isolates with the most similar hybridization pattern) are pooled into a new larger cluster until all entries are connected (Figure 13).

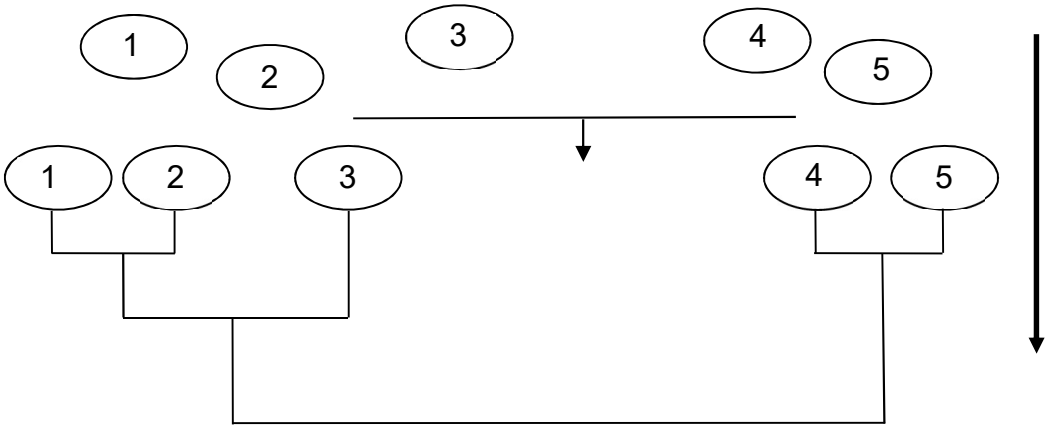


Figure 13: **Presentation how a dendrogram is constructed by the hierarchical agglomerative clustering algorithm.** As indicated by the name, this type of clustering proceeds in a bottom-up fashion. Initially, every element (sample) is considered as a single cluster (shown at the top). Then, the pairs with smallest pairwise distance are connected to each other. In this example, elements (1) and (2) are connected at this step, as well as (4) and (5). Subsequently, element (3) is merged with the cluster containing elements (1) and (2) because its distance to these elements is smaller than to the other elements. In this way, step by step, smaller clusters are merged into larger clusters until one large cluster remains in the end. One important decision during this clustering procedure is to decide at which distance threshold the clustering should be halted. This then determines the final number of obtained clusters.

In our case, similarity was measured by the Euclidian distance of the DNA MA hybridization profiles. In mathematics Euclidean distance is the straight-line distance between two points in Euclidean space. This means that similarity was measured based on the number of differing entries (hybridization signals) between two vectors holding the hybridization signals of two samples. For simplicity, let us assume a scenario with 5 probes. Sample A could yield the following hybridization signal: (0 1 1 0 1) and sample B could yield: (1 1 0 0 1). These two samples differ in the first and third element (probe). Thus, their Euclidian distance is 2.

Before hierarchical clustering, all probe identifications with ambiguous results were replaced by imputed values predicted with LFM method as described in section 3.2.2.4.2. Thereafter, each isolate contains only negative (0) or positive (1) values.

Next, the phylogenetic–like analysis was performed with the HAC algorithm of the program R suite (version 2.13.1) (Ihaka & Gentleman, 1996, R Development Core Team, 2008), together with Bioconductor packages (Gentleman *et al.*, 2004) as well as the software package Stats. Stats computes a Euclidean distance matrix to measure the similarity of the target sequence hybridization profiles in the different SA isolates

with the software tool „dist“ (distance) function in the software package „Stats“ (R, version, 2.13.1). Then, the cluster dendrogram was constructed using the HAC method and the “hclust” function of „Stats“ based on Ward’s method (Pryer *et al.*, 2001, Quackenbush, 2001).

3.2.2.4.6 Silhouette plot for clustered data

The silhouette plot was applied in study C to confirm, if the subjectively assigned isolates to clusters, really belong to these cluster based on a dispassionate mathematical evaluation.

Here, data (e.g. isolates/hybridization signals) is clustered by the K-means clustering algorithm. At the beginning, this algorithm arbitrarily selects a given number of K data points (single isolates of the isolate set) as centers of K clusters. All remaining data points are assigned to the closest one of these cluster centers. These “clusters” will initially not be very homogeneous. Then for each cluster, a new central data point is determined that is really in the center of its current members. Afterwards, all data points are again assigned to the cluster with the closest center point. For some points, nothing will change. Other points may now be assigned to a different cluster. This procedure is iterated until the cluster centers remain stable.

One open point in this is to choose the optimal number K for the number of clusters. This depends on the data set. As optimality criterion the so-called silhouette analysis uses the silhouette value. The silhouette value compares how close each data point (e.g. isolate) of a cluster is to the other points (e.g. isolates) in the same cluster relative to its distance to points in neighboring clusters. In our case distances reflect the Euclidean distance as explained before. Basically, this method measures the strengths of a cluster or how good the assignment of a point (e.g. isolate/hybridization signal) to a cluster (e.g. African isolate group, or adhesion gene group) is. The values range from [-1, 1]. That way, it can be controlled by PCA if a subjectively recognized isolate group or the amount of recognized isolate groups is correct. If the silhouette coefficients are close to 1, the point/isolate is located far away from the next cluster. A value close to 0 indicates that this isolate is close to or is lying on the threshold value between two adjacent clusters. Negative values indicate that these isolates are possibly assigned to the wrong cluster.

The silhouette analysis was performed with the software “R”, version 3.2.2, function

silhouette and package “cluster” version 2.0.3 based on default parameters. Afterwards, K is selected that yields a maximal silhouette value.

3.2.3 MALDI-TOF analysis

3.2.3.1 Culture and MALDI-TOF performance

For the MALDI-TOF analysis of the SA isolates of the fourth study, freshly streaked blood agar plates were prepared from one-day old blood agar plate cultures. Applying the whole cell measurement method, one bacterial colony was picked with a toothpick and transferred onto the predefined spot of the MALDI target plate. For each isolate, eight colonies were streaked on eight individual spots of the target plate. The measurement was repeated on three different days. The dried spots were covered with 1 µl of the matrix (Table 1). The target plate with the different isolates was analyzed in a Microflex LT spectrometer (Bruker Daltonics) using the preinstalled AutoXecute method as soon as every single spot covered with matrix was dried.

3.2.3.2 Mass spectra analysis

Mass spectra analyses were performed in study D using the mass spectra tools as well as mathematical and statistical methods of the software BioNumerics, version 7.5. The spectra data files of all 219 measured SA isolates were imported into BioNumerics. Then all MALDI TOF spectra and the DNA MA data were assigned to the individual isolates using the isolates’ unique identifier.

Additionally, all available information of geographic origin, *spa*-type and CC were imported as a table into BioNumerics. The data of the single isolates could be connected by the isolates’ unique identifier.

3.2.3.2.1 Mass spectra preprocessing and PCA

All mass spectra were preprocessed for further analysis. For this, smoothing and baseline subtraction of the uploaded isolate spectra as well as a peak matching were performed. This consists of a background subtraction using a rolling disc algorithm, a noise removal with a Kaiser window and peak detection with a continuous wavelet algorithm (Vranckx *et al.*, 2017). These steps prepare the spectra for application of similarity coefficients such as Pearson correlation coefficient as well as PCA. These steps are necessary to obtain the same basic level on the y axis of all single spectra

and to order peaks of the same mass to charge ratio at the same position on the x-axis. This step eliminates the “background noise”. Next, different comparisons were performed with the BioNumerics software where the mass spectra of SA isolates of different *spa*-types, different CCs, methicillin-susceptible and methicillin-resistant SA and of different geographic origin were compared using PCA.

3.2.3.2.2 Mass spectra identification projects with Support Vector Machine

With a Support Vector Machine (SVM) so called identification projects were performed (Vranckx *et al.*, 2017). The SVM predicted if the mass spectra were true positively, false negatively or false positively assigned or not assigned to the known, predefined subgroup of the SA isolates (e.g. *spa*-types, CCs, MRSA or MSSA).

A SVM is a specific mathematical algorithm for pattern recognition, also called classifier. It performs an identification or classification. This means the algorithm assigns biological samples or mass spectra to user defined classes e.g. CC or *spa*-type using their mass spectra based on the mass spectra of a training set, for example SA isolates with already known CC or *spa*-type. The information about e.g. CC or *spa*-type were previously imported as well as the mass spectra profiles into BioNumerics and could be assigned by the isolates’ unique identifier.

First the classifier was trained with a training set of mass spectra from isolates with known classes. After training of the classifier, a cross validation analysis was performed. In a cross-validation analysis k mass spectra are removed from the training set, treated as unknown and their class is predicted by the classifier using the remaining k-1 mass spectra of the training set. This is done for a specified percentage of the training set, typically 100% as in this work. The number of false and true positives and false and true negatives as well as unassigned isolates are reported to evaluate the performance of the identification which led to a confusion matrix (error matrix). In the confusion matrix each predefined group e.g., geographic origin is represented as column and the predicted group e.g., African or German as row. This can provide additional information about the identification and help to identify groups that cannot be differentiated from each other but can be differentiated from the remaining groups. This is an internal validation as the complete training set has been used to train the classifier. If enough isolates are available, at least 10% can be left out of training and the classifier can be applied to this 10% and the results can be compared to the true

classes. This external validation was not performed due to an insufficient size of the groups in our study.

3.2.3.2.3 Identification project or pairwise comparisons

Detailed pairwise comparisons or so-called identification projects were performed between two entries of mass spectra of the comparison data sets or in this case the mass spectra subgroups and defined groups (*spa*-type, clonal complexes, resistance group, geographic origin) from the database with previously imported isolate data. The aim was to determine if they were correctly assigned to the previously defined groups. The respective results were displayed as percentage of similarity for all pairwise comparisons.

Then the classifier, in this work the Support Vector Machine (SVM), is chosen. A cross validation analysis was performed as described in section 3.2.3.2.2 for the classifier training but without removing of any mass spectra which led to a confusion matrix (error matrix). The confusion matrix provides the results, the objective estimation of the classifier's performance. In this study it means if the SA isolates were correctly assigned to the predefined categories geographic origin, methicillin resistance, *spa*-type and CC.

3.2.3.2.4 Statistical analysis of peak intensities

Statistical analysis was performed on the peak intensities with the aim to identify characteristic peak classes of different CCs. First, a cluster analysis was performed using Pearson correlation and UPGMA (unweighted pair group with arithmetic mean) as statistical tests. Next, the intensities of the peak classes of the spectra were compared using the Mann-Whitney test with Bonferroni procedure for correction of multiple testing. Peaks with significant values for a specific CC were considered potential discriminative peak candidates. Furthermore, randomly selected spectra of different CCs were visually compared at the identified peak values. These peaks were assumed to be potentially mutated proteins or differentially expressed proteins. To determine which proteins were represented by these peaks, an in-silico databank search was carried out with the ExPASy software tool "Translate" provided by the Swiss Institute of Bioinformatics. The identified peak mass values were entered in the ExPASy software tool and potential proteins for this value were identified according to

a predefined range.

3.2.4 *Spa*-typing

The DNA was extracted by inoculation of 5 colonies from freshly streaked blood agar plates into 200 µl A. dest and incubated for 10 min at 95°C and 550rpm. Next, the tubes were centrifuged for 3 min at 13226xg to sediment cell debris. Then the following master mix was prepared for the *spa*-PCR (per isolate): 10 µl H_2O, 12.5 µl PCR–Mix (peqGOLD Hot-Start-Mix Y), 0.25 µl forward primer *spa*-1113f and 0.25 µl reverse primer *spa*-1514r. Finally, 2.5 µl of the DNA-containing supernatants were added.

spa-PCR protocol:

95°C	5 min	initial denaturation
35 cycles:		
94°C	45 sec	denaturation
60°C	45 sec	annealing
72°C	90 sec	elongation
96°C	20 sec	denaturation
72°C	90 sec	final elongation
4°C	∞	cool down

After the PCR, the amplificate was digested to cut off the primer sequences. To this end, 5 µl of the enzyme ExoSAP-IT and 5 µl of the amplificate were mixed, centrifuged and incubated for 15 min at 37°C and then for 15 min at 80°C. For the PCR control, 3 µl of the amplification product were visualized on a 1% TBE agarose gel. The amplification products were separated on a 1% agarose gel together with a ladder, a negative and a positive control at 150V for 35 min.
Amplificates were sent to GATC Biotech AG (Konstanz) for sequencing of the gene *spa*. Sequence files of *spa* were downloaded and analyzed using the semiautomatic Ridom StaphType software for *spa*-sequence analysis.

3.2.5 Multi locus sequence typing (MLST)

MLST analysis (Enright *et al.*, 2000) of 12 SA isolates of study C for CC determination was carried out by our cooperation partners in Münster.

3.2.6 Whole genome sequencing for DNA MA concordance analysis and analysis of nucleotide sequences of three genes (*panD*, *rpmG,* and *rpmD*)

The accuracy of the DNA MA was validated by WGS analyses of 154 representative isolates (Strauss *et al.*, 2016). A motivation for WGS was that the DNA MA technique does not enable distinction between allelic variants or gene polymorphisms, which are not represented on the array as suspected for the SA adhesion gene *map*/*eap* in study C.

Additionally, analysis of nucleotide sequences of the two ribosomal genes *rpmG, rpmD* and the aspartate 1 decarboxylase *panD* to determine potential sequence mutations was performed in cooperation with the Institute for Hygiene, UKM, Münster using Ridom^{TM} SeqSphere$^{+,}$ as described by Strauss et al., 2016.

3.2.7 Statistical methods

3.2.7.1 Chi square test

Univariate analysis of patient data and of gene distribution in MSSA vs. MRSA of study A (section 4.1) were performed using GraphPad QuickCalcs Web site (GraphPad Software Inc.) (http://graphpad.com/quickcalcs/contingency1/).

P-values of CC and gene distribution of study C with 1200 African and German SA (section 4.3) were calculated by Chi-square tests with Hommel correction for multiple testing using R Studio Team (2015) (R Studio: Integrated Development for R. RStudio, Inx. Boston, MA; http://www.rstudio.com/).

3.2.7.2 Median of age

Median ages of volunteers and patients of the African and German SA study (Appendix, section 6.3) were calculated using SPSS (IBM).

4 Results

4.1 Determination of clonal lineages and gene distribution of MRSA and matched MSSA by DNA MA

Using a collection of 46 MRSA and 46 matched MSSA the MRSA and MSSA clonal lineages of community-associated SA isolates of the federal state of Saarland were analyzed. Further, it was analyzed which one of three different bioinformatic methods (splits graph, hierarchical agglomerative clustering [HAC], and principal component analysis [PCA]) is the most appropriate DNA MA data analysis method. In comparison to DNA MA data analysis by splits graph, *spa*-typing data were also analyzed by splits graph. Furthermore, the presence of resistance and virulence gene distribution in MRSA in comparison to MSSA were regarded.

4.1.1 DNA MA and *spa*-typing data of stem collection

The *spa*-typing and DNA MA analyses identified in total 13 different *spa*-types belonging to 5 CCs among the MRSA isolates and 33 *spa*-types belonging to 13 CC types among the MSSA isolates (Table 3 and 4). CC5 (n=41, 89.1%), the so-called Rhine-Hesse epidemic strain was the predominant strain type among the MRSA with 30 (78%) isolates belonging to t003 (Table 3). In the MSSA group *spa*-types t012 (n=6, 13%) and t015 (n=5, 10.9%) were most commonly found (Table 4). CCs being present in both the MRSA as well as in the MSSA isolate group were the epidemic CCs CC5, CC8, CC22, CC45 and the zoonotic lineage CC398. CCs only found in the MSSA group were CC1, CC7, CC15, CC30, CC78, CC97 and CC101. No MRSA specific CCs were detected. The three *spa*-types t002, t008 and t015 were found in the MRSA as well as in the MSSA group.

DNA MA analysis showed furthermore that 38 (83%) of the MRSA isolates belonged to SCC*mec* type II, four isolates (9%) belong to SCC*mec* type IV and one isolate (2%) belong to SCC*mec* type. However, the SCC*mec* type of three isolates (7%) could not be determined by DNA MA.

Table 3: **List of the CCs, *spa*- and SCC*mec* types of the 46 MRSA collected during the admission prevalence screening of the federal state of Saarland. List is sorted according to CC and *spa*-type.** na=not available

Isolate	**CC/ST**	***spa*-type**	***SCCmec type***	**numbers**
R11	CC5	t002	II	1
R1, R2, R3, R6, R9, R10, R12, R13, R15, R16, R19, R20, R24, R25, R27, R30, R31, R32, R33, R34, R35, R36, R37, R38, R39, R40, R41	CC5	t003	II	27
R21, R22, R23	CC5	t003	na	3
R5	CC5	t010	II	1
R14	CC5	t045	II	1
R29	CC5	t481	II	1
R4, R8, R17, R18	CC5	t504	II	4
R7	CC5	t887	II	1
R26	CC5	t1079	II	1
R28	CC5	t3195	II	1
R42, R43	CC8	t008	IV	2
R46	CC22	t022	IV	1
R45	CC45	t015	IV	1
R44	CC398	t011	V	1

Table 4: **List of the CCs and *spa*-types of the 46 MSSA collected during the admission prevalence screening of the federal state of Saarland. List is sorted according to CC and *spa*-type.**

Isolate	CC	*spa*-type	Numbers
S40	CC1	t8864	1
S42	CC5	t002	1
S43	CC5	t493	1
S41, S44	CC7	t091	2
S28	CC8	t008	1
S38	CC15	t018	1
S34, S37	CC15	t084	2
S36	CC15	t306	1
S35	CC15	t8786	1
S26	CC22	t005	1
S24	CC22	t310	1
S25	CC22	t625	1
S27	CC25	t078	1
S2, S3, S5, S6, S7, S10	CC30	t012	6
S4	CC30	t019	1
S8	CC30	t273	1
S1	CC30	t584	1
S9	CC30	t8831	1
S13, S16, S17, S19, S22	CC45	t015	5
S23	CC45	t026	1
S11	CC45	t040	1
S18	CC45	t050	1
S14	CC45	t073	1
S21	CC45	t339	1
S15	CC45	t620	1
S12	CC45	t1689	1
S20	CC45	t2239	1
S39	CC88	t8863	1
S29	CC97	t131	1
S31, S32, S33	CC97	t267	3
S30	CC97	t8830	1
S45	CC101	t4044	1
S46	CC398	t571	1

4.1.2 Cluster analysis of total stem collection

Splits graph (Huson & Bryant, 2006) were evaluated for visualization of the *spa*-typing data and for the capability of the DNA MA array to further differentiate the 46 MRSA and the 46 MSSA isolates profiles.

4.1.2.1 Splits graph of 46 MRSA and 46 MSSA based on *spa*-typing data

Splits graph of the *spa*-typing data scattered the isolate set into nine different isolate groups (cluster 1, 2, 4, 5, 6, 7, 8, 10, 12), and five single isolates (cluster 3, 9, 12, 13, 14) (Figure 14, Table 5) and indicated how the individual groups are placed in relation to each other.

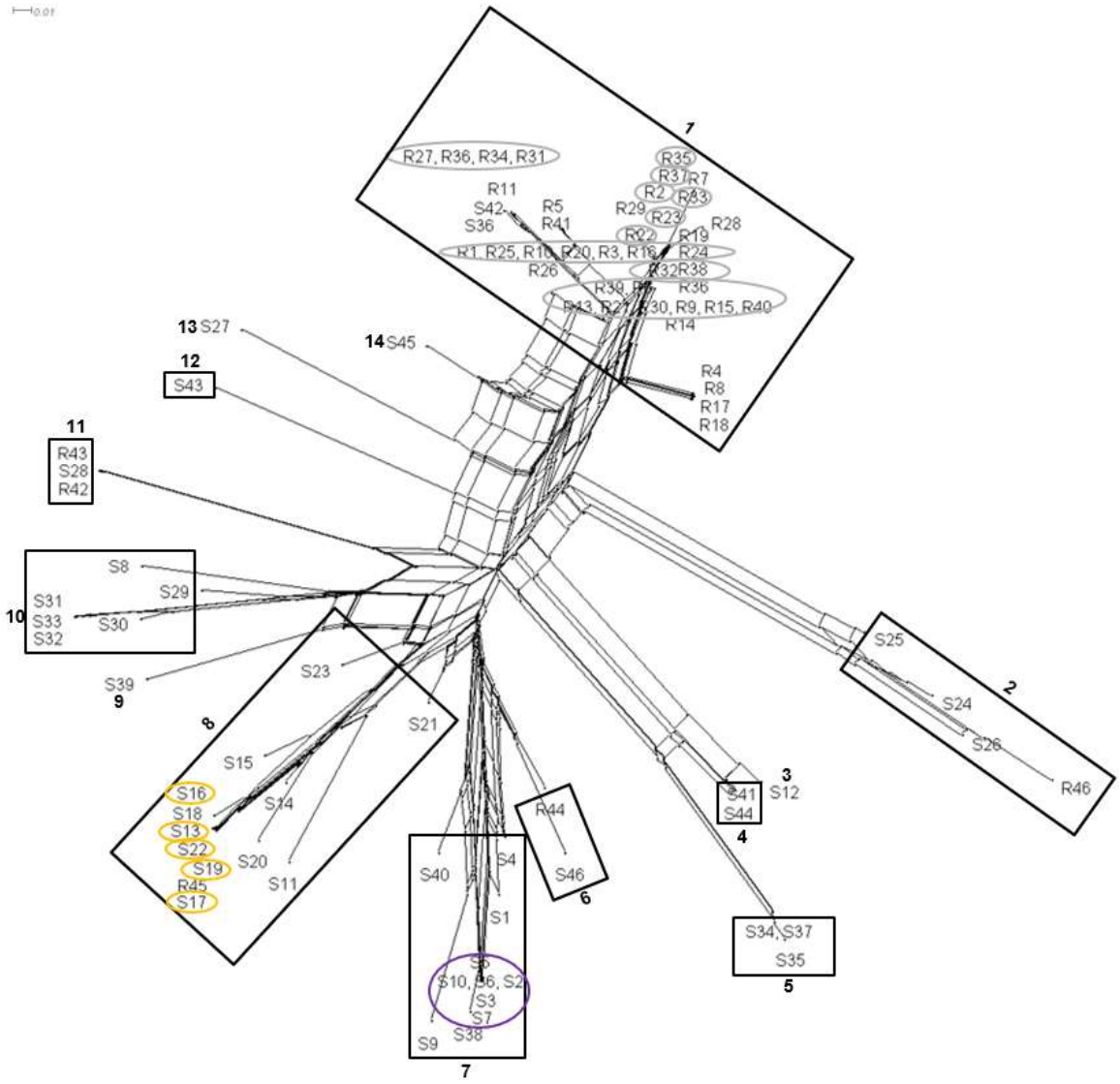


Figure 14: ***Spa*-typing data-based diversity analysis of 46 MSSA (S1-S46) and 46 MRSA (R1-R46) isolates.** Splits graphs were constructed based on *spa*-typing data and a calculated cost distance matrix by direct data transfer from Ridom StaphType to splits tree (Huson & Bryant, 2006). The predominant *spa*-types of this study, t003 (grey circles), t012 (purple circles) and t015 (orange circles) are highlighted. Isolates of the same clonal complexes (CC) are framed by tetragons. All isolates groups were numbered (1-14). Figure adapted from Ruffing et al., PLoS One, 2012.

Based on the Ridom database 61 isolates of 17 *spa*-types could be assigned to a CC or ST (Table 6), while the remaining isolates could not be assigned by this database.

Splits graph analysis of the *spa*-typing data allowed to assign 19 more isolates to the following CC/ST isolate groups: CC5 (n=9), CC22 (n=1), CC30 (n=3), CC45 (n=5), ST15/ST18 (n=1).

In both cases most isolates were assigned to the CC5 group (n=43, 47%) (Table 5).

Table 5: **List of the groups in splits graph based on *spa*-typing data isolates**. na =not available

cluster	no.	*spa*-type	CC/ST assignment described by Ridom spa server database	Additional assignment by splits graph	Isolate
1	43	t002	CC5		R11, S42
		t003	CC5		R1, R2, R3, R6, R9, R10, R12, R13, R15, R16, R19, R20, R21, R22, R23 R24, R25, R27, R30, R31, R32, R33, R34, R35, R36, R37, R38, R39, R40, R41
		t010	CC5		R5
		t045	CC5		R14
		t306	na	CC5	S36
		t481	na	CC5	R29
		t504	na	CC5	R4, R8, R17, R18
		t887	na	CC5	R7
		t1079	na	CC5	R26
		t3195	na	CC5	R28
2	4	t005	CC22	CC22	S26
		t022	CC22	CC22	R46
		t310	CC22	CC22	S24
		t625	na		S25
3	1	t1689	na		S12
4	2	t091	ST7		S41, S44
5	3	t084	ST15, ST18		S34, S37
		t8786	na	ST15, ST18	S35
6	2	t011	na		R44
		t571	na		S46
7	11	t012	CC30	CC30	S2, S3, S5, S6, S7, S10
		t018	CC30	CC30	S38
		t019	CC30	CC30	S4
		t584	na		S1
		t8831	na		S9
		t8864	na		S40
8	13	t015	CC45		R45, S13, S16, S17, S19, S22
		t026	CC45		S23
		t040	na	CC45	S11
		t050	CC45		S18
		t073	na	CC45	S14
		t339	na	CC45	S21
		t620	na	CC45	S15
		t2239	na	CC45	S20
9	1	t8863	na		S39
10	6	t131	na		S29
		t267	na		S31, S32, S33
		t8830	na		S30
		t273	na		S8
11	3	t008	CC8		R42, R43, S28
12	1	t493	na		S43
13	1	t078	ST-26		S27
14	1	t4044	na		S45

4.1.2.2 Splits graph of 46 MRSA and 46 MSSA based on DNA MA data

Splits graph analysis (Huson & Bryant, 2006) based on the DNA MA hybridization profiles identified 13 clusters with the most frequent ones being CC5 (n=43), CC45 (n=14), CC30 (n=9), CC15 (n=5) and CC97 (n=5) (Figure 15, Table 6).

Splits graph based on DNA MA data presented in detail how the single isolates are placed in relation to each other and that isolates of the same CC as S12 of CC45 and S43 of CC5 build one common group (Figure 15, Table 6).

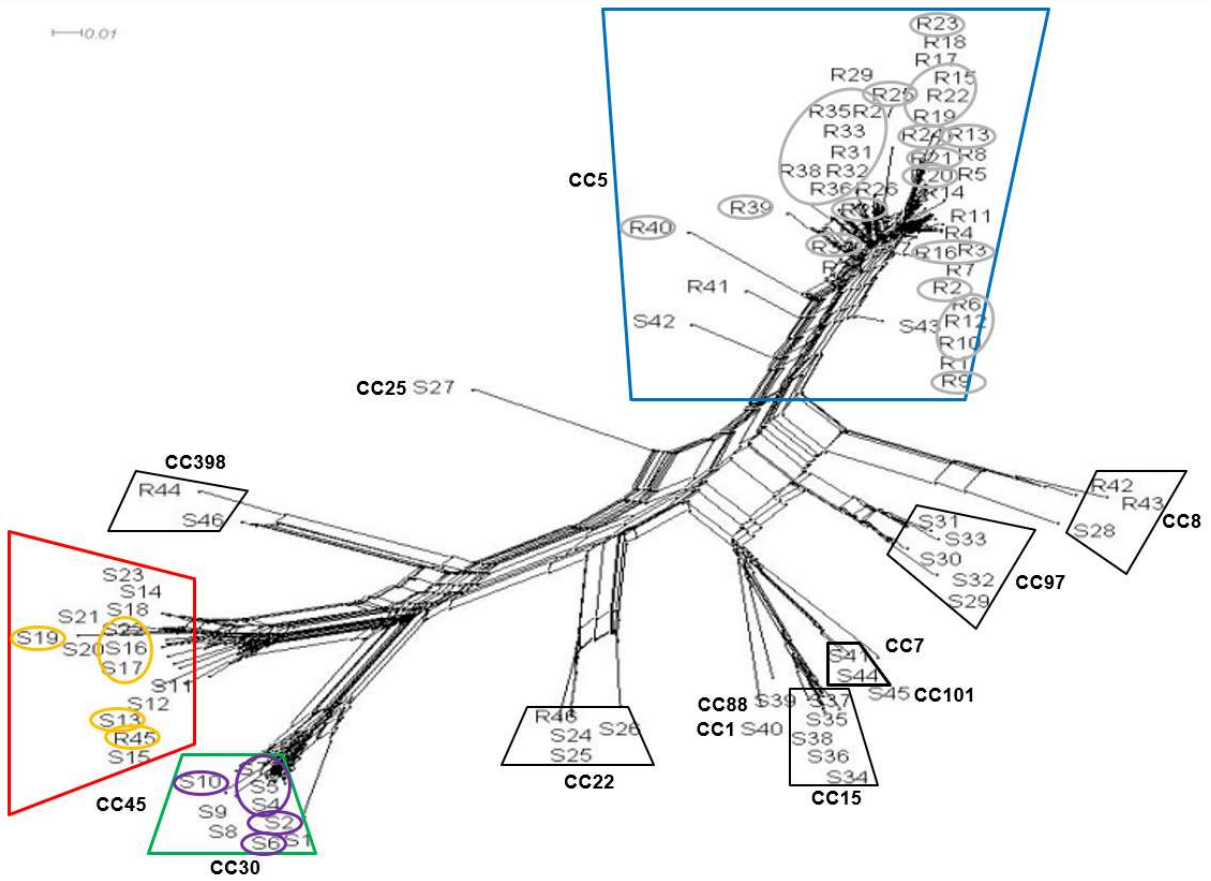


Figure 15: **DNA microarray (MA) data-based diversity analysis of 46 MSSA (S1-S46) and 46 MRSA (R1-R46) isolates**. Splits graphs were constructed based on the DNA MA hybridization profiles in FASTA format on default settings using splits tree. The positive, negative and ambiguous hybridization results represented as 1 (present), 0 (absent) and "-" (ambiguous) in the DNA MA results presentation were converted into a nucleotide code: 1=c, 0=g, "-"=a. Columns represented the target sequences and the rows represent the isolates. Isolates of the same clonal complexes (CC) are framed by tetragons. CCs with the largest isolates numbers (e.g. CC5 [blue], CC45 [red] and CC30 [green]) are encircled by coloured tetragons. The predominant *spa*-types of this study, t003 (grey circles), t012 (purple circles) and t015 (orange circles) are highlighted. Adapted from Ruffing et al., PLoS One, 2012.

Table 6: **List of the CC groups and assigned isolates** based on **the DNA MA data derived splits graph analysis.**

CC group	numbers	Isolate
CC1	1	S40
CC5	43	R1, R2, R3, R4, R5, R6, R7, R8, R9, R10, R11, R12, R13, R14, R15, R16, R17, R18, R19, R20, R21, R22, R23, R24, R25, R26, R27, R28, R29, R30, R31, R32, R33, R34, R35, R36, R37, R38, R39, R40, R41, S42, S43
CC7	2	S41, S44
CC8	3	R42, R43, S28
CC15	5	S34, S35, S36, S37, S38
CC22	4	S24, S25, S26, R46
CC25	1	S27
CC30	10	S1, S2, S3, S4, S5, S6, S7, S8, S9, S10
CC45	14	R45, S11, S12, S13, S14, S15, S16, S17, S18, S19, S20, S21, S22, S23
CC88	1	S39
CC97	5	S29, S30, S31, S32, S33
CC101	1	S45
CC398	2	S46 R44

4.1.2.3 Comparison of the splits graph of 46 MRSA and 46 MSSA based on *spa*-typing and DNA MA data

Comparison of both splits graphs revealed a different assignment of five MSSA isolates in the two splits graphs for S8, S12, S36, S38, and S43 (Table 7). Based on the *spa*-typing data S12 and S43 were singletons whereas DNA MA data grouped S12 with CC45 and S43 with the CC5 isolate group, respectively. Isolate S8 was assigned to CC97 by *spa*-typing data but to CC30 by DNA MA data. Similarly S36 were assigned to CC5 by *spa*-typing data but to CC15 based on DNA MA data, and S38 was assigned to CC30 based on *spa*-typing data but to CC15 based on DNA MA data.

Table 7: **Comparison of the assignment of five MSSA isolates in splits graphs based on *spa*-typing and DNA MA data.**

Isolate	*spa*-type	DNA MA CC	Assignment in splits graph based on *spa*-typing data to cluster of CCx isolates	Assignment in splits graph based on DNA MA data to cluster of CCx isolates
S8	t273	CC30	CC97	CC30
S12	t1689	CC45	no assignment	CC45
S36	t306	CC15	CC5	CC15
S38	t018	CC15	CC30	CC15
S43	t493	CC5	no assignment	CC5

Further, comparison of the splits graphs based on *spa*-typing and DNA MA data showed a different order and number of clusters in the splits graphs as well as of single isolates. Splits graph based on *spa*-typing yielded 9 clusters and six singletons. Splits graph based on DNA MA generated 13 clusters with four singletons. Singletons S27, S39 and S45 are identical in both splits graphs. Only isolate clusters of CC7 and CC15 or CC30 and CC45 were in both splits graphs in direct neighborhood to each other. All other isolate clusters showed an incomparable order with different CCs assembling in both splits graphs, as neighbours as could be seen for example for the CC30 and CC398 clusters.

4.1.3 Cluster analysis of 43 CC5 isolates by splits graphs based on *spa*-typing and DNA MA as well as by PCA and HAC based on DNA MA

Isolates belonging to CC5 dominated the isolate collection. Splits graph analysis of the whole collection of 92 isolates allowed only a rough isolate group assignment. In order to increase the resolution within this CC, splits graphs as well as to two further bioinformatical methods, i.e. PCA and HAC were next used to evaluate isolates belonging to CC5 only.

4.1.3.1 Splits graph of 43 CC5 isolates based on *spa*-typing data

Splits graph analysis of the 43 CC5 isolates (MRSA R1-41 and MSSA S42, S43) based on the *spa*-typing data yielded in *spa*-type specific subgroups of isolates of *spa*-type t002, t003, t010 and t504 (Figure 16).

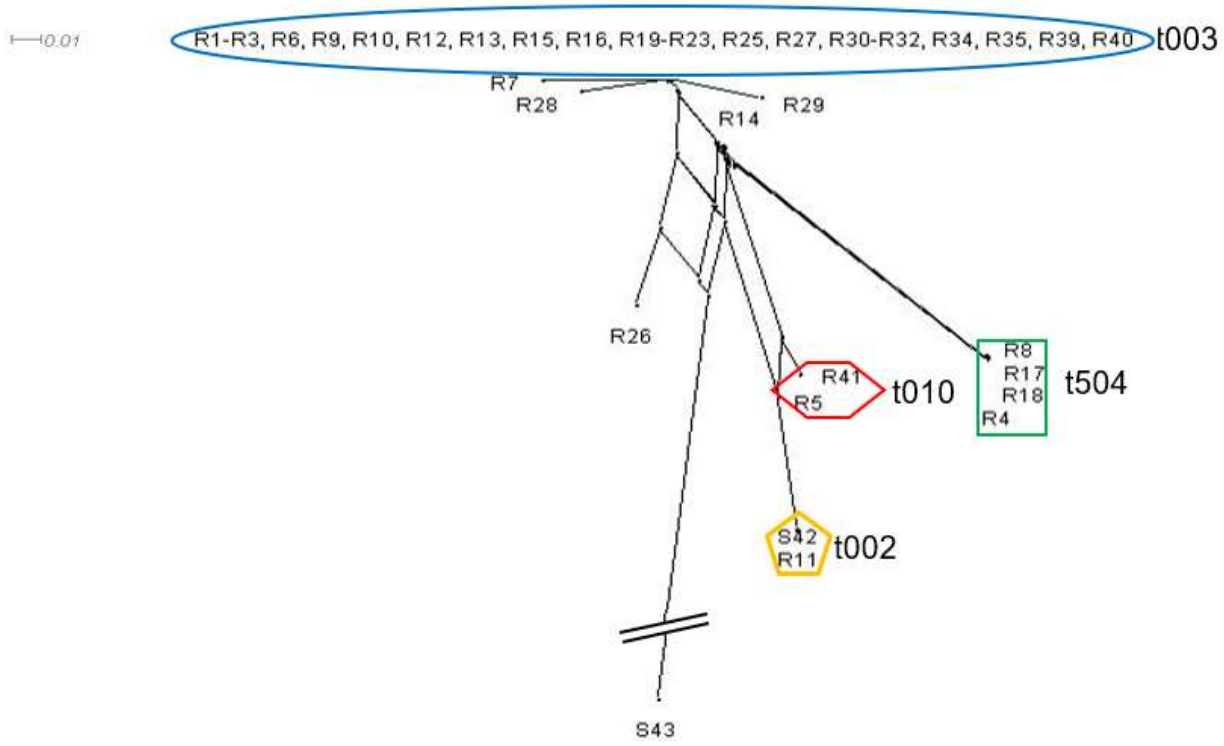


Figure 16: **DNA microarray subclassification analysis of the 43 CC5 isolates, R1-41 (MRSA) and S42, S43 (MSSA)**. Splits graph based on *spa*-typing data and cost distance matrix calculated by direct data transfer from Ridom StaphType to splits tree (Huson & Bryant, 2006). The MRSA *spa*-types t003 (blue circle), t504 (green square), t010 (red hexagon) and t002 (orange, pentagon) which were detected more than once, are highlighted. Figure adapted from Ruffing et al., PLoS One, 2012.

4.1.3.2 Splits graphs of 43 CC5 isolates based on DNA MA data

Splits graph analysis based on the DNA MA hybridization profiles of the 43 CC5 isolates subdivided the CC5 isolates into a number of different groups. However, only for the group of isolates R22, R23, and R24 there was a clear denominator visible – these isolates were the only isolates being positive for the genes encoding the ACME locus (Figure 17). For the remaining isolates the grouping could not be assigned to a single denominator.

Isolates of the predominant *spa*-type t003 (Figure 17, isolates marked by blue circles) were found in all tree branches, while isolates of *spa*-type t504 (Figure 17, isolates marked by green squares) showed an accumulation in one part of the tree.

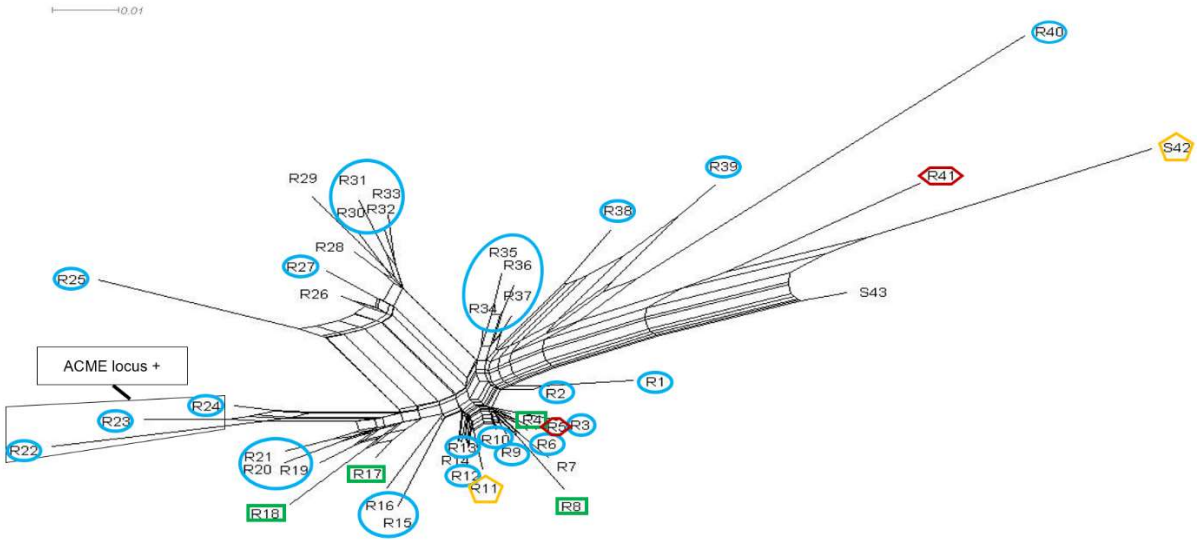


Figure 17: **DNA microarray (MA) subclassification analysis of the 43 CC5 isolates, R1-41 (MRSA) and S42, S43 (MSSA)**. Splits graph based on DNA MA hybridization profiles in FASTA format on default settings using splits tree. The MRSA *spa*-types t003 (blue circles), t504 (green squares), t010 (red hexagons) and t002 (orange, pentagons) which were detected more than once, are highlighted. Three isolates carrying the ACME locus as specific feature were highlighted. Adapted from Ruffing et al., PLoS One, 2012.

4.1.3.3 Comparison of the splits graphs based on *spa*-typing and DNA MA data of 43 CC5 isolates

4.1.3.3.1 DNA MA data analysis of 43 CC5 MRSA/MSSA with PCA

PCA (Figure 18A) based on the hybridization profiles of the 43 CC5 isolates scattered the isolates into two clusters with 30 isolates (70%) grouping together in cluster I and 9 isolates in cluster II (20.9%), respectively, while four isolates (R40, R41, S4 and S43; 9.1%) were displayed individually.

A subsequent PCA of the 30 CC5 isolates, assembled as cluster 1 in the first PCA (Figure 18B) identified four subgroups (Figure 18B, Ia-Id). Subgroup Id isolates showed positive hybridization signals for the gene loci of the ACME locus.

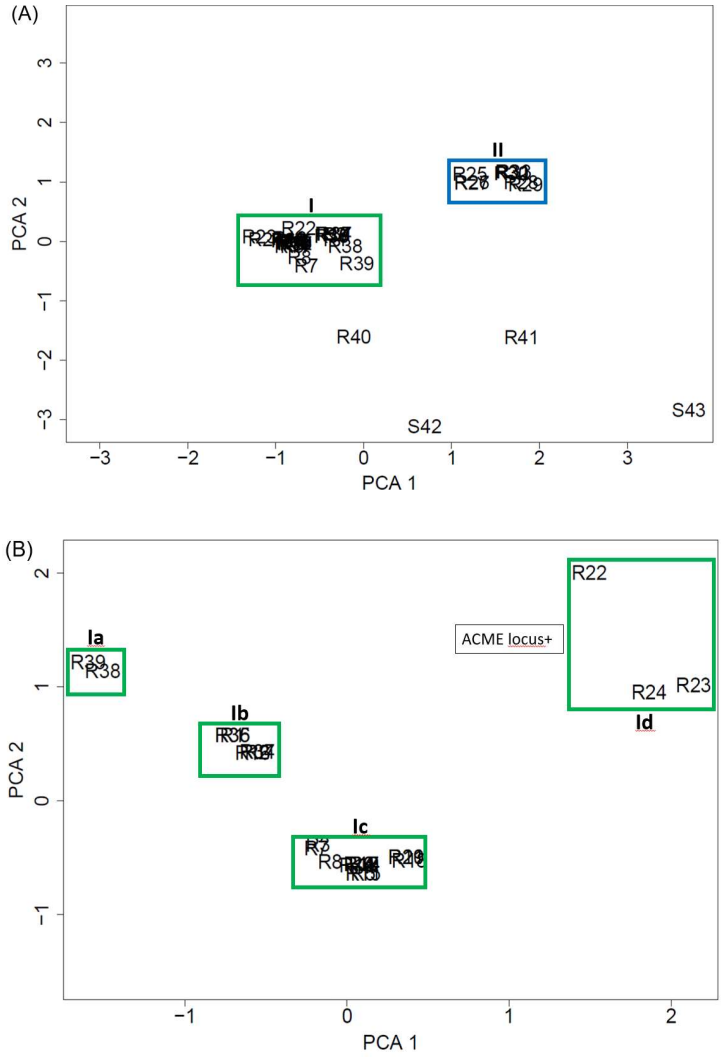


Figure 18: **Principal component analysis (PCA) of 43 CC5 isolates, R1-41 (MRSA) and S42, S43 (MSSA)**. PCA was conducted based on the hybridization profiles of the 334 target sequences of each *Staphylococcus aureus* isolate yielded by DNA microarray analyses. (A) Two clusters I (green) and II (blue) within the 43 CC5 isolates were identified by PCA. (B) A higher resolution PCA plot of the 30 MRSA isolates of cluster I (green) identified four subclusters (Ia-Id). Subcluster Id was characterized by the presence of the ACME locus. Adapted from Ruffing et al., PLoS One, 2012.

In total, five different PCA clusters (Ia to Id, II) and four unassigned isolates could be identified by PCA.

4.1.3.3.2 DNA MA data analysis of 43 CC5 isolates with HAC and bioinformatic methods' data analysis comparison

DNA MA hybridization pattern analysis of the CC5 SA isolates (Figure 19) by HAC showed five isolate clusters (I-V). The HAC provides a high resolution and it is very easy to see how the isolates relate to each other. This means it is easy to understand

which isolates are very similar and which not.

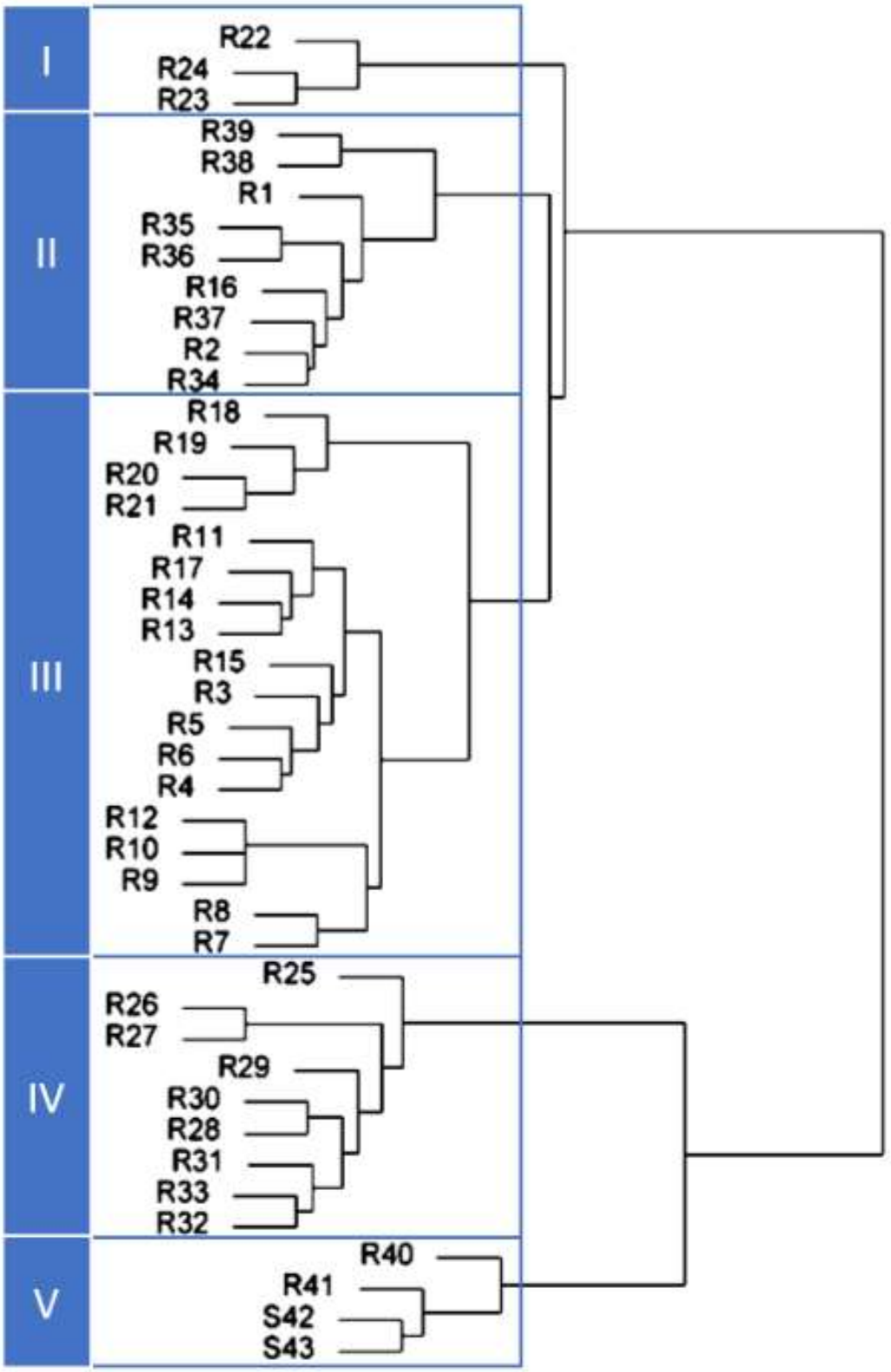


Figure 19: **Hierarchical agglomerative clustering (HAC) of 43 CC5 isolates, R1-41 (MRSA) and S42, S43 (MSSA)**. HAC was conducted based on the hybridization profiles of the 334 target sequences of each *Staphylococcus aureus* isolate yielded by DNA microarray analyses. Five clusters I to V were identified by HAC.

4.1.3.3.3 Comparison of cluster analysis of 43 CC5 isolates with splits graph, PCA and HAC based on DNA MA data

The comparison of all three different bioinformatical analyses methods showed that 12 isolates were assigned to two identical isolate groups of R22-R24 (n=3) and R25-R33 (n=9). All other 31 CC5 isolates (R1, R2, R34-R41, S42, S43) were assigned to different subgroups (Table 8).

Table 8: **Correlation of CC5 *Staphylococcus aureus spa*-typing results with subgroups identified based on the DNA microarray (MA) data analysis using three different bioinformatic methods.** The DNA MA subgroup analyses were performed with principal component analysis (PCA, 3nd column), splits graph (4rd column) and hierarchical agglomerative clustering (5th column) based on the DNA MA hybridization results.

Isolate	*spa*-type	MA PCA cluster (arbitrary group)	MA Splits graph cluster (arbitrary group)	MA Cluster dendrogram (arbitrary group)
R22	t003	I d	1	I
R24	t003	I d	1	I
R23	t003	I d	1	I
R39	t003	I a	4	II
R38	t003	I a	4	II
R1	t003	I b	6	II
R35	t003	I b	3	II
R36	t003	I b	3	II
R16	t003	I b	8	II
R37	t003	I b	3	II
R2	t003	I b	6	II
R34	t003	I b	3	II
R18	t504	I c	9	III
R19	t003	I c	9	III
R20	t003	I c	9	III
R21	t003	I c	9	III
R11	t002	I c	7	III
R17	t504	I c	9	III
R14	t045	I c	7	III
R13	t003	I c	7	III
R15	t003	I c	8	III
R3	t003	I c	7	III
R5	t010	I c	7	III
R6	t003	I c	7	III
R4	t504	I c	7	III
R12	t003	I c	7	III
R10	t003	I c	7	III
R9	t003	I c	7	III
R8	t504	I c	7	III
R7	t887	I c	7	III
R25	t003	II	2	IV
R26	t1079	II	2	IV
R27	t003	II	2	IV
R29	t481	II	2	IV
R30	t003	II	2	IV
R28	t3195	II	2	IV
R31	t003	II	2	IV
R33	t003	II	2	IV
R32	t003	II	2	IV
R40	t003	others	4	V
R41	t010	others	5	V
S42	t002	others	5	V
S43	t493	others	5	V

Comparison of the three applied bioinformatic methods showed that each model may have its specific strengths. PCA enables a direct simple overview of isolate groups of

an almost unlimited number of isolates. Yet, simple assignment of each point in the graph to the corresponding isolate is difficult because of the very close up to overlapping presentation of the samples. For good resolution you need a digital format which allow to zoom in the figure. HAC reveals a more detailed isolate relationship with direct assignment to each isolate and the corresponding isolate cluster. However, HAC is only useful for analysis of up to one hundred isolates. The problem is that for more isolates it is not possible to present the dendrogram on a A4 page with sufficient resolution. Again, for good resolution you need a digital format which allow to zoom in the dendrogram as described before for the PCA. Thus, splits graph seems to be the most appropriate method for cluster analysis for a high number of isolates. Additionally, splits graph is easy to apply.

4.1.4 Gene repertoire analysis

The gene repertoire of the 46 MRSA and 46 MSSA isolates was analyzed for individual genes associated with e.g. antibiotic resistance, toxin production, adhesion, and immune evasion.

The SA species markers (*rrn, gapA, katA, coa, nuc1, spa*) were all positive in the entire strain set. Consequently, species identification of the DNA MA was confirmed.

The analysis of the accessory gene regulator (*agr*) genes, a master regulatory system of virulence regulation in SA (Novick, 2003), showed that all CC5 MRSA isolates (n=41, 89.1%) belonged to *agr*-type II (Table 9). The remaining 5 (10.9%) MRSA isolates were associated with *agr*I (Table 9). 21 MSSA isolates of CC7, CC8, CC22, CC45, CC97, CC101, and CC398 (45.7%) belonged to *agr*-type I, too. Ten MSSA isolates of CC1, CC30, and CC88 harbored *agr*-type III (21.7%) and 6 MSSA of CC5 and CC15 were positive for *agr*-type II (15.2%). Identification of the *agr*-type of nine MSSA isolates was not possible.

Table 9: **Comparison of the *agr*-type distribution in the clonal complexes (CC) of methicillin-resistant *Staphylococcus aureus* (SA) and methicillin-susceptible SA and of the Staphylococcal cassette chromosome(SSC)*mec* types in MRSA.**

	CC	*agr* type I, n (%)	*agr* type II, n (%)	*agr* type III, n (%)	SCC*mec* type I, n (%)	SCC*mec* type II, n (%)	SCC*mec* type IV, n (%)	SCC*mec* type V, n (%)
MRSA	CC5	0 (0)	41 (89.1)	0 (0)	0 (0)	41 (89.1)	1 (2.2)	0 (0)
	CC8	2 (4.3)	0 (0)	0 (0)	0 (0)	0 (0)	1 (2.2)	0 (0)
	CC22	1 (2.2)	0 (0)	0 (0)	0 (0)	0 (0)	1 (2.2)	0 (0)
	CC45	1 (2.2)	0 (0)	0 (0)	0 (0)	0 (0)	1 (2.2)	0 (0)
	CC398	1 (2.2)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (2.2)
		5 (10.9)	41 (89.13)	0 (0)	0 (0)	41 (89.1)	4 (8.7)	1 (2.2)
MSSA	CC1	0 (0)	0 (0)0	1 (2.2)	na	na	na	na
	CC5	0 (0)	1 (2.2)	0 (0)	na	na	na	na
	CC7	2 (4.3)	0 (0)	0 (0)	na	na	na	na
	CC8	1 (2.2)	0 (0)	0 (0)	na	na	na	na
	CC15	0 (0)	5 (10.9)	0 (0)	na	na	na	na
	CC22	3 (6.5)	0 (0)	0 (0)	na	na	na	na
	CC30	0 (0)	0 (0)	0 (0)	na	na	na	na
	CC45	8 (17.4)	0 (0)	0 (0)	na	na	na	na
	CC88	0 (0)	0 (0)	0 (0)	na	na	na	na
	CC97	5 (10.9)	0 (0)	0 (0)	na	na	na	na
	CC101	1 (2.2)	0 (0)	0 (0)	na	na	na	na
	CC398	1 (2.2)	0 (0)	0 (0)	na	na	na	na
		21 (45.7)	6 (15.2)	10 (21.7)				

SCC*mec*-type analysis (Table 9) showed that 41 MRSA isolates (89.1%), possessed a SCC*mec*-type II-cassette while the other MRSA isolates harbored a *SCCmec* type IV cassette (n=4, 8.7%) or a SCC*mec*-type V-cassette (n=1, 2.2)%). The SCC*mec* types of three MRSA could not be determined by DNA MA.

The β-lactamase operon with the three genes *blaZ*, *blaI* and *blaR* were detected in the majority of the MRSA (n=39, 84.8%) and MSSA isolates (n=29, 63%) (Table 10). The gene *erm*(A) conferring resistance against macrolides, lincosamides and streptogramins (MLS_B) was found in significantly more MRSA (n=41, 89.1%) than MSSA isolates (n=3, 6.5%) ($p<0.001$) (Table 10). The putative marker gene for fosfomycin and bleomycin resistance, *fosB,* was also detected significantly more often in the MRSA isolates (n=43, 93.5%) than in MSSA isolates (n=20, 43.5%) ($p<0.001$) (Table 10). Most of the isolates were positive for the *sdrM* gene (n=82/92, 89%). Other resistance genes were only detected in single isolates (Table 10). In general, MRSA isolates showed a higher number of antibiotic resistance genes than MSSA isolates.

Table 10: **Comparison of the DNA microarray target recognition of selected resistance genes of methicillin-susceptible *Staphylococcus aureus* (SA) and methicillin-resistant SA.**

gene	MSSA, n (%)	MRSA, n (%)	p-value
blaZ/I/R	29 (63)	39 (84.8)	<0.05
ermA	3 (6.5)	41 (89.1)	<0.001
ermC	1 (2.2)	0 (0)	ns
fosB	20 (43.5)	43 (93.5)	<0.001
tetK	2 (4.3)	0 (0)	ns
tetM	0 (0)	1 (2.2)	ns
qacA	0 (0)	1 (2.2)	ns
qacC	0 (0)	1 (2.2)	ns
sdrM	40 (87)	42 (91)	ns

p-values were calculated by Chi-square test using GraphPad Quick Calcs Web site (GraphPad Software Inc.) (http://graphpad.com/quickcalcs/contingency1/); ns=not significant

The ACME gene cluster was found in three MRSA isolates of the strain lineage ST2 (6.5%). Analysis of the leukocidin hybridization results shows that none of the isolates were positive for the Panton-Valentine leukocidin (PVL) encoding genes (*lukF-*/*lukS*-PV) (Table 11). The *tst*1 (encoding toxic shock syndrome toxin 1) was mainly found in MSSA of CC30 (n=8, 17.4%). Outside CC30 *tst1* was only found in one isolate of CC45 (2.2%).

The hybridization patterns of the enterotoxin gene cluster (*egc*) genes *seg*, *sei*, *sem*, *sen*, *seo*, and *seu* showed that these genes were significantly more frequently detected in MRSA (n=43, 93.5%) compared to MSSA (n=29, 63%) (p<0.001) (Table 11). Moreover, the *egc* genes were only detected in four CCs, namely CC5, C22, CC30, and CC45.

The coding sequence of the metalloprotease aureolysin (*aur*) was identified significantly more often in MRSA (n=43, 93.5%) than in MSSA (n=30, 65.2%; p<0.001) (Table 11). The same result was visible for the serine protease genes *splA* and *splB,* which were predominantly detected in MRSA isolates of the CCs CC1, CC5, CC7, CC8, CC15 and CC97.

Most MRSA isolates carried the genes encoding factors for capsule type 5 biosynthesis (n=44, 96%) in contrast to MSSA isolates (n=15, 32%) (Table 11). Only one MRSA isolate (2.2%) was positive for the genes of capsule type 8, while the majority of the MSSA isolates harbored the capsule type 8 encoding genes (n=31, 74%) (Table 11).

The microbial surface components recognizing adhesive matrix molecule

(MSCRAMM) encoding gene *cna* (collagen-binding protein) was detected with a higher rate in the MSSA group (n=26, 57%) compared to the MRSA group (n=3, 7%) (Table 11). In the MRSA group, the SERAM (secretable expanded repertoire adhesive molecules) encoding genes *fib* (encoding a 19kDa fibrinogen-binding protein; n=38, 83%), and *sasG* (encoding the cell-wall anchored surface protein SasG; n=43, 93%) were identified more often (Table 11). Additional MSCRAMM encoding genes (Table 11) and two further SERAM genes were found in most of the isolates (≥83%), without methicillin resistance specific differences.

Table 11: **Comparison of the DNA microarray target recognition of selected toxin and adhesion genes of of methicillin-susceptible *Staphylococcus aureus* (SA) versus methicillin-resistant SA.**

gene/ gene cluster	MSSA, n (%)	MRSA, n (%)	p-value
lukF/S-PV	0 (0)	0 (0)	ns
tst1	9 (19.5)	0 (0)	<0.01
egc	29 (63)	43 (93.5)	0.001
aur	30 (65.2)	43 (93.5)	<0.01
splA	19 (41.3)	42 (91.3)	<0.001
splB	20 (44.3)	43 (93.5)	<0.001
*cap*5	15 (32)	45 (98)	<0.001
*cap*8	31 (74)	1 (2.2)	<0.001
cna	26 (57)	3 (6.5)	<0.001
bbp	42 (91)	43 (94)	ns
clfA	46 (100)	46 (100)	ns
clfB	43 (94)	46 (100)	ns
ebh	42 (91)	44 (96)	ns
eno	46 (100)	46 (100)	ns
ebps	46 (100)	46 (100)	ns
fnbA	45 (98)	44 (96)	ns
fnbB	37 (82)	45 (98)	ns
sdrC	46 (100)	46 (100)	ns
fib	20 (44.3)	38 (83)	<0.001
sasG	19 (41.3)	43 (93.5)	<0.001
map/eap	42 (91)	43 (94)	ns
vwb	46 (100)	46 (100)	ns

p-values were calculated by Chi-square test using GraphPad Quick Calcs Web site (GraphPad Software Inc.) (http://graphpad.com/quickcalcs/contingency1/); ns=not significant

4.2 Characterization of Nigerian SA isolates by DNA MA analysis

Analysis of the strain lineages and the gene repertoire of 52 Nigerian MSSA isolates by DNA MA showed that they belonged to 12 different CCs: CC1 (n=6, 11.6%), CC5 (n=9, 17.3%), CC7 (n=1, 1.9%), CC8 (n=4, 7.7%), CC15 (n=7, 13.5%), CC25 (n=2, 3.8%), CC30 (n=8, 15.4%), CC45 (n=2, 3.8%), CC80 (n=1, 1.9%), CC97 (n=1, 1.9%), CC121 (n=8, 15.4%), and CC152 (n=3, 5.8%). More than half of the 52 MSSA belonged to CC1, CC5, CC8, CC30, and CC45.

As shown in Figure 20, isolates of CC7, CC8, CC25, CC45, CC97, and CC152 (n=13, 25%) were associated with *agr*-type I while 16 isolates of CC5 and CC15 (30.2%) belonged to *agr*-type II and 15 isolates of CC1, CC30 and CC80 belonged to *agr*-type III. Combined signals for *agr*-type I and type-IV were observed for one, three and four isolates grouped within CC25, CC152, and CC121 (EEF, sheet A1).

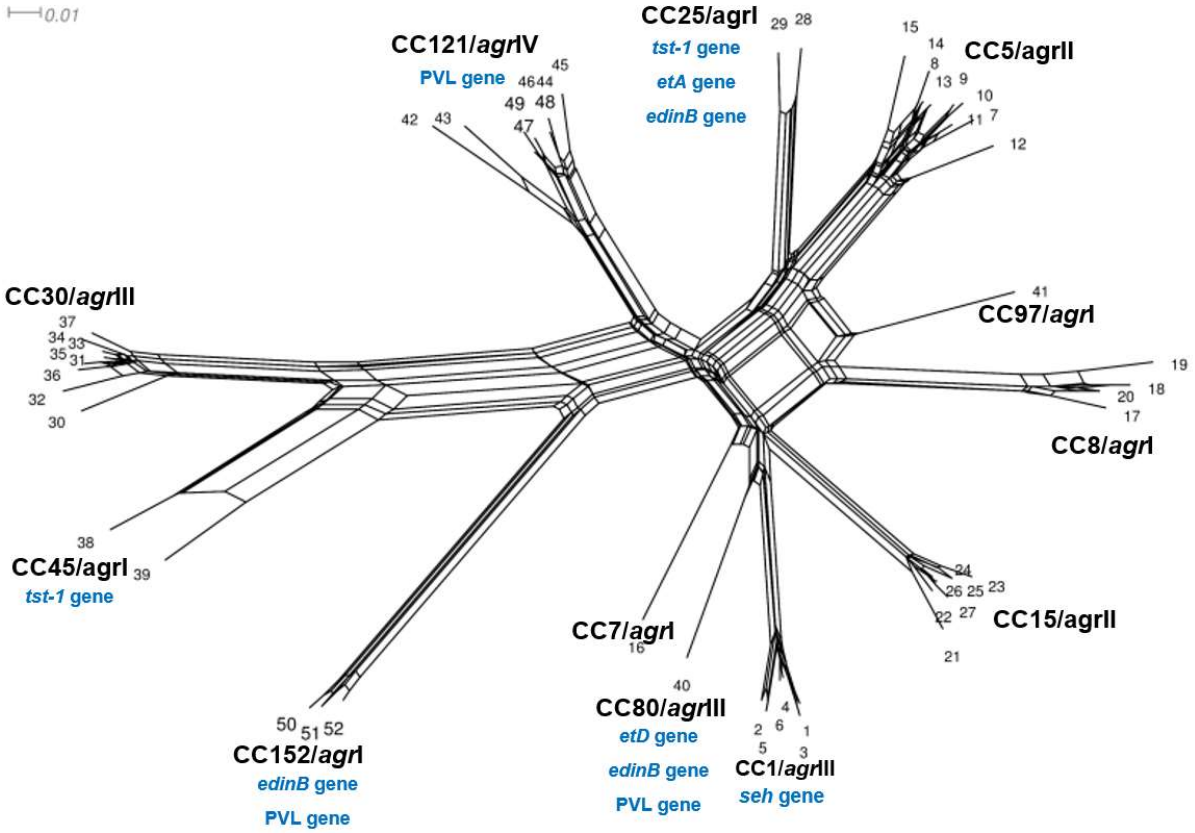


Figure 20: **Similarity analysis of 52 Nigerian methicillin-susceptible SA isolates**. Splits graph are constructed based on the DNA microarray (MA) hybridization profiles on default settings by splits tree (Huson & Bryant, 2006). The hybridization results were arranged in a matrix in a FASTA format. The present, absent and ambiguous hybridization results represented as 1 (present), 0 (absent) and "-" (ambiguous) in the DNA MA results presentation were converted into a nucleotide code: 1=c, 0=g, "-"=a. Columns represented the target sequences and the rows represent the isolates. Clonal complexes (CC) and *agr*-types as well as genes (blue) which were present in all isolates of a CC are indicated. Adapted from Shittu et al., Front Microbiol, 2015.

The detection rates of selected resistance and virulence genes were shown in Table 12. The predominant resistance genes in the Nigerian SA isolates were *sdrM,* encoding for a multidrug efflux channel (n=52, 100%) and *fosB*, encoding for fosfomycin resistance (n=39, 75%) (Table 12).

Table 12: **Detection rates of selected resistance and virulence genes in 52 Nigerian methicillin-susceptible *Staphylococcus aureus* isolates.**

genes	n, (%)
sdrM	52 (100)
fosB	39 (75)
blaZ/I/R	36 (69)
aur	30 (65.2)
tetK/tetM	13 (25)
lukF/S-PV	27 (52)
tst1	3 (6)
etA	2 (4)
etB	0 (0)
etD	3 (6)
cap8	33 (64)
cap5	19 (37)
icaA	50 (96)
icaC	49 (94)
icaD	50 (96)
fnbA	52 (100)
clfA	52 (100)
ebh	52 (100)
eno	52 (100)
ebps	52 (100)
fnbA	52 (100)
cna	27 (52)

The *lukF*- and *lukS-PV* genes encoding the Panton-Valentine leukocidin (PVL) were detected in 27 (52%) isolates belonging to seven CCs (CC1 [n=4, 8%], CC5 [n=5, 10%], CC15 [n=1, 2%], CC30 [n=5, 10%], CC80 [n=1, 2%], CC121 [n=8, 15%], and CC152 [n=3, 6%]; EEF, sheet A1).

The gene *tst1* (the toxic shock syndrome) was found in three isolates of CC1 (n=1, 1.9%) and CC45 (n=2, 3.8%). Presence of the genes *etA*, *etB* and *etD,* encoding exfoliative toxins, was observed in a total of five isolates of CC15 (n=2, 4%), CC25 (n=2, 4%), and CC80 (n=1, 2%). Almost two-third of the analyzed isolates carried the genes encoding for capsule type 8 (n=33, 63.5%; Table 12). The remaining 19 isolates carried capsule type 5 encoding genes (36.5%; Table 12).

Biofilm-associated genes *icaA* (n=50, 96%), *icaC* (n=49, 94%) and *icaD* (n=50, 96%) (intracellular adhesion genes) were found in almost all isolates (Table 12). The MSCRAMM encoding genes *clfA* (clumping factor A), *ebh* (cell wall-associated fibronectin-binding protein), *eno* (enolase binding protein), *ebps* (cell surface elastin binding protein), and *fnbA* (fibronectin binding protein A) were detected in all 52

isolates (Table 12; EEF, sheet A1). In contrast, *cna* (collagen binding adhesin) was only found in isolates of CC1 (n=6, 52%), CC30 (n=8, 15%), CC121 (n=8, 15%) and CC152 (n=3, 6%; EEF, sheet A1).

4.3 Characterization and comparison of African and German SA isolates

4.3.1 Collection of African and German SA isolates

The data of all participants were documented in case related forms (CRFs) (see Appendix, section 6.2) and checked of fulfilling the requirements for inclusion into the study cohort. Characteristics of African and German healthy volunteers and patients are shown in the Appendix, section 6.3.

4.3.2 Clonal complexes and sequence types analyses

The DNA MA data assigned 1,185 isolates of the 1,200 isolates to 32 known CCs and three STs based on the hybridization profiles. For the remaining 15 isolates, CC grouping could be resolved for three of the isolates (0.25%) by affinity propagation of the DNA MA data, whereas for 12 isolates assignment by DMA MA data was not possible. A subsequent MLST analysis (Enright *et al.*, 2000) was carried out by our cooperation partners in Münster. The 12 isolates were assigned to five different STs (ST2370, ST2678, ST2733, ST2735, and ST2744) that had not been described so far.

4.3.2.1 Analysis of clonal complexes with African or German SA origin

A comparison of the CC distribution in the African and German study sites is shown in Figure 21 and Table 13. The comparison identified several significant differences in CC representations between the African and the German study sites. Four CCs were either exclusively found in Africa (CC80 and CC88) or in Germany (CC50 and CC398), all other CCs were detected both in Africa and in Germany. Furthermore, significant geographic differences were identified for 15 of the 40 detected CCs and STs between the African and German study sites (Figure 21). The 22 (55%) most common CC's were statistically, mainly present either in Africa or in Germany (Figure 21, Table 13). On the other hand isolates of CC5, CC8, CC9, CC25 as well as CC707 were observed in Africa and Germany, respectively (Figure 21).

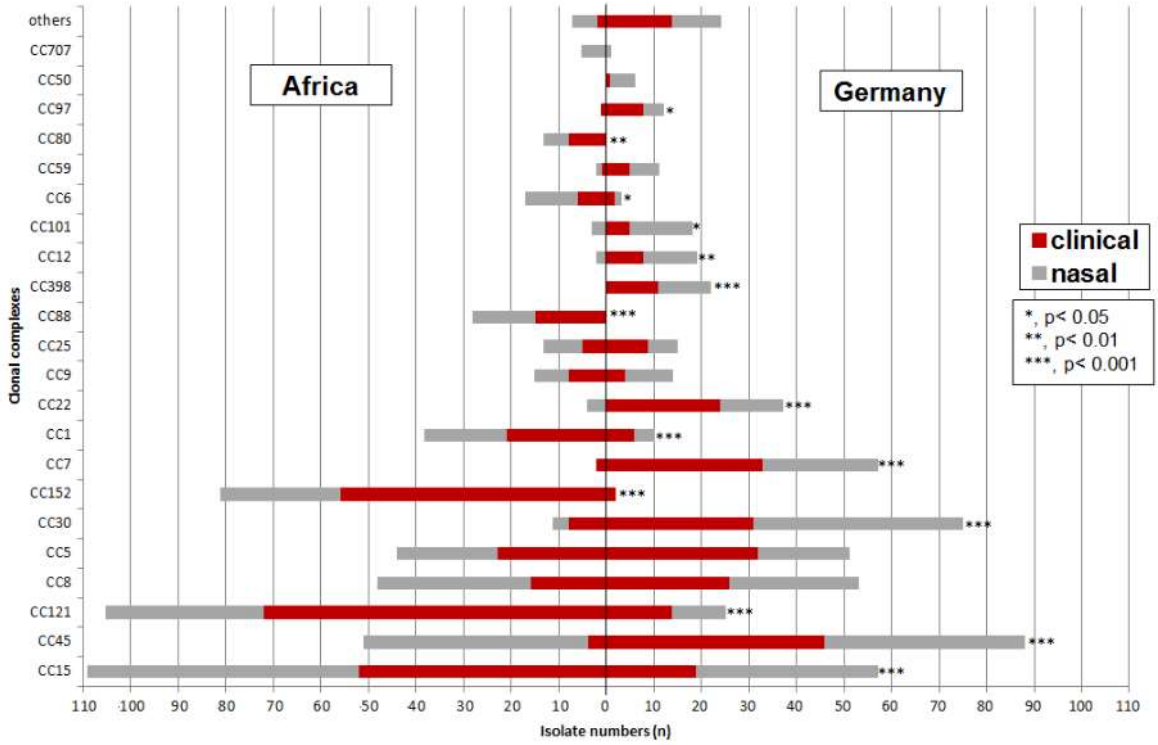


Figure 21: **Comparison of the frequencies of the 22 most prevalent clonal complexes (CCs) of community-associated *Staphylococcus aureus* isolates in the African and German study sites.** CCs and sequence types (STs) with less than six isolates were summarized as group others. The CCs on the y-axis were ordered according to the total number of identified isolates. Prevalence of clinical (red) and nasal (grey) isolates in the African (left side) and German group (right side) is shown. Statistical analyses for association of CCs with African or German origin were performed by Chi square test with Hommel correction for multiple testing using the program R Studio; *: p<0.05, **: p<0.01, ***: p<0.001. Figure adapted from Ruffing et al., 2017.

Predominant African CCs were CC80 (n=13, 100%, p<0.001), CC88 (n=28, 100%, p<0.001), CC152 (n=79, 98%, p<0.001), and CC121 (n=105, 81%, p<0.001) (Figure 21, Table 13). The main CCs in Germany were CC398 (n=22, 100%, p<0.001), CC22 (n=37, 90%, p<0.001), CC7 (n=57, 97%, p<0.001), CC30 (n=75, 87%, p<0.001,) and CC45 (n=88, 63%, p=0.01) (Figure 21, Table 13).

Table 13: **Clonal complex distribution in Africa and Germany with total numbers and percentages of isolates found in African and German study sites for each CC group.**

	Total [n]	Africa [n]	African isolates [%]	Germany [n]	German isolates [%]	p-value*
CC1	48	38	79	10	21	<0.001
CC5	95	44	46	51	54	ns
CC6	20	17	77	3	23	0.03
CC7	59	2	3	57	97	<0.001
CC8	101	48	4	53	53	ns
CC9	29	15	52	14	48	ns
CC12	21	2	9	19	91	<0.01
CC15	166	109	66	57	34	<0.001
CC22	41	4	10	37	90	<0.001
CC25	28	13	46	15	54	ns
CC30	86	11	13	75	87	<0.001
CC45	139	51	37	88	63	0.01
CC50	6	0	0	6	100	ns
CC59	13	2	15	11	85	ns
CC80	13	13	100	0	0	<0.001
CC88	28	28	100	0	0	<0.001
CC97	13	1	92	12	8	<0.05
CC101	21	3	14	18	86	0.05
CC121	130	105	81	25	19	<0.001
CC152	81	79	98	2	2	<0.001
CC398	22	0	0	22	100	<0.001
CC707	6	5	83	1	17	ns
others	34	10	29	24	71	ns
sum	1200	600		600		

Statistical analyses were performed by Chi square test with Hommel correction for multiple testing using the program R Studio; ns= not significant

4.3.2.2 Association analysis of clonal complexes with clinical or nasal isolate origin

In addition to the analysis of the CC distribution in the African and German study sites, the CC distribution within the clinical and nasal isolate groups were compared (Figure 22). CC121 ($p<0.01$) and CC152 ($p<0.01$) were identified significantly more often in clinical isolates, whereas CC45 ($p<0.05$) were significantly more often detected in nasal isolates (Figure 22). The other CC's did not show any predominance for isolates of either nasal or clinical origin.

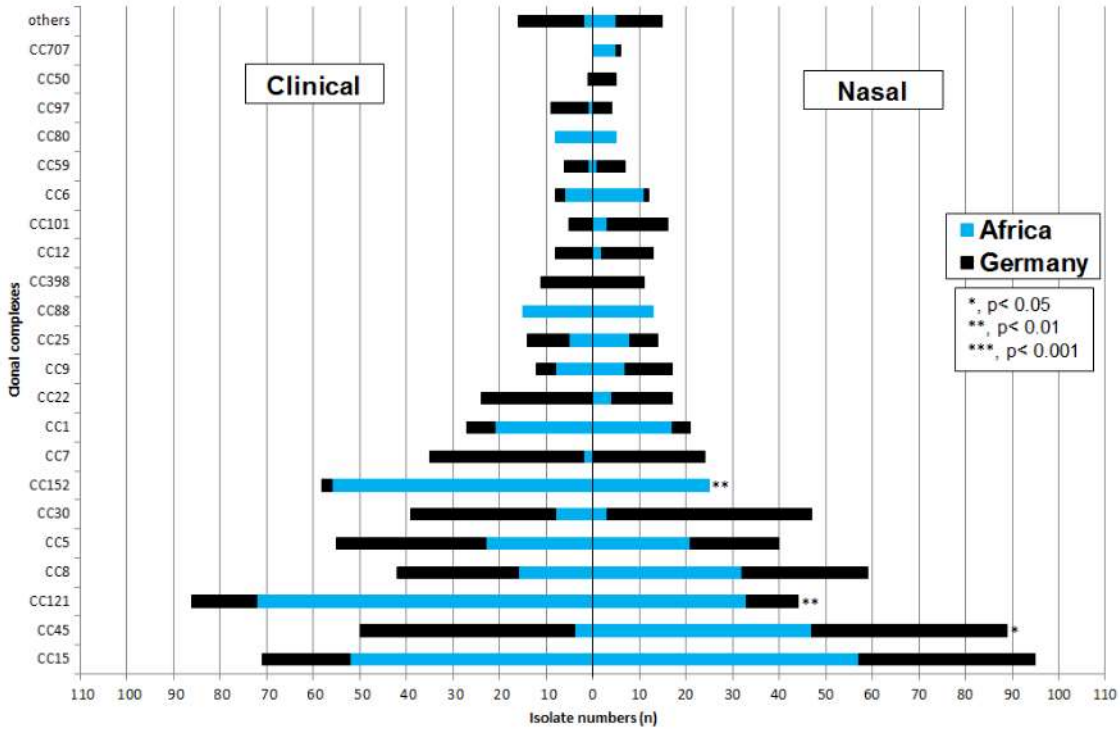


Figure 22: **Comparison of the prevalence of the nasal and clinical community-associated *Staphylococcus aureus* isolates in the 22 most prevalent clonal complexes (CCs).** CCs and sequence types (STs) with less than six isolates were summarized as group others. The CCs on the y-axis were ordered with respect to the total number of identified isolates from the bottom to the top. Prevalence of German (black) and African (blue) isolates in the clinical (left side) and nasal group (right side) is indicated. Statistical analyses for association of CCs with clinical or nasal isolate origin were performed by Chi square test with Hommel correction for multiple testing using R Studio; *: $p<0.05$, **: $p<0.01$, ***: $p<0.001$.

Comparison of the DNA MA data of 300 clinical and 300 nasal African strains (Figure 22, Table 14) revealed also for the African isolates alone that a significantly larger proportion of CC121 (yellow section, $p<0.001$) and CC152 (black, $p<0.001$) isolates were of clinical origin. In contrary, isolates of CC8 (orange, $p<0.05$) and CC45 (red, $p<0.001$) were more frequently detected in nasal isolates. Among the German isolates

(Figure 23A, second and fourth column) CC15 were found significantly more frequent in the nasal isolate group (dark blue section, p<0.05).

Table 14: **Comparison of the clonal complex (CC) distribution of community-associated *Staphylococcus aureus* in the individual African and German study sites with differentiation of the clinical or nasal isolate origin.** The study sites were represented by abbreviations: IT= Ifakara, Tanzania; LG= Lambaréné, Gabon; Manhiça, Mozambique; FR=Freiburg, Baden-Württemberg; HS=Homburg, Saarland; MW= Münster, Nordrhein-Westfalen.

	clinical isolates						nasal isolates					
	Africa			Germany			Africa			Germany		
	IT	LG	MM	FR	HS	MW	IT	LG	MM	FR	HS	MW
CC15	5	27	19	3	9	7	17	24	16	11	11	16
CC45	0	0	4	13	12	21	10	35	2	13	12	17
CC121	24	7	42	6	4	4	6	4	23	4	2	5
CC8	4	6	6	8	11	7	14	6	12	12	6	9
CC5	9	8	6	11	15	6	11	6	4	5	10	4
CC30	4	2	2	10	7	14	2	1	0	18	11	15
CC152	13	38	4	2	0	0	0	14	11	0	0	0
CC7	0	0	2	10	14	9	0	0	0	5	13	6
CC1	9	9	3	2	2	2	6	1	10	0	4	0
CC22	0	0	0	10	5	9	1	0	3	8	0	5
CC9	7	1	0	1	2	1	6	1	0	2	4	4
CC25	2	0	3	5	2	2	1	0	7	2	2	2
CC88	10	1	4	0	0	0	7	1	5	0	0	0
CC398	0	0	0	3	5	3	0	0	0	5	4	2
CC12	0	0	0	2	4	2	0	0	2	3	6	2
CC101	0	0	0	1	1	3	2	1	0	3	8	2
CC6	3	1	2	1	0	1	9	1	1	1	0	0
CC59	0	0	1	2	2	1	1	0	0	2	4	0
CC80	6	0	2	0	0	0	3	0	2	0	0	0
CC97	1	0	0	5	0	3	0	0	0	0	2	2
CC50	0	0	0	1	0	0	0	0	0	2	0	3
CC707	0	0	0	0	0	0	0	4	1	0	0	1
others	3	0	0	4	5	5	4	1	1	4	1	5
sum	100	100	100	100	100	100	100	100	100	100	100	100

The analysis of the relative CC frequencies of isolates of clinical (Table 14; Figure 23A) or nasal origin (Table 14; Figure 23B) identified the following: Within the Ifakara (Tanzania, IT) isolate group, CC121 (yellow, p<0.001) and CC152 (black, p<0.001) were significantly more often identified in the group of clinical origin, while CC8 (orange, p<0.001), CC15 (dark blue, p<0.05) and CC45 (red, p<0.001) were found more often in the group of nasal isolates. In the clinical strain group of Lambaréné (Gabon, LG) CC1 (grey, p<0.05) and CC152 isolates (black, p<0.001) were

dominating, while CC45 isolates (red, $p<0.001$) were only detected among isolates of nasal origin. In Manhiça (Mozambique, MM) isolates of CC121 (yellow, $p<0.01$) were observed more often in the clinical isolate group, similar to the situation visible for Ifakara. Compared to the African study sites, the relative frequencies of CC's in the clinical and nasal German groups as well as in-between the three German institutions were more homogenous. Only isolates of CC15 (blue, $p<0.05$) were found more often in the nasal isolates than in the clinical isolates sampled in Freiburg (FR) (Figure 23). In addition, the distribution of CCs in the African clinical and African nasal isolate group was more diverse than in the German isolate groups.

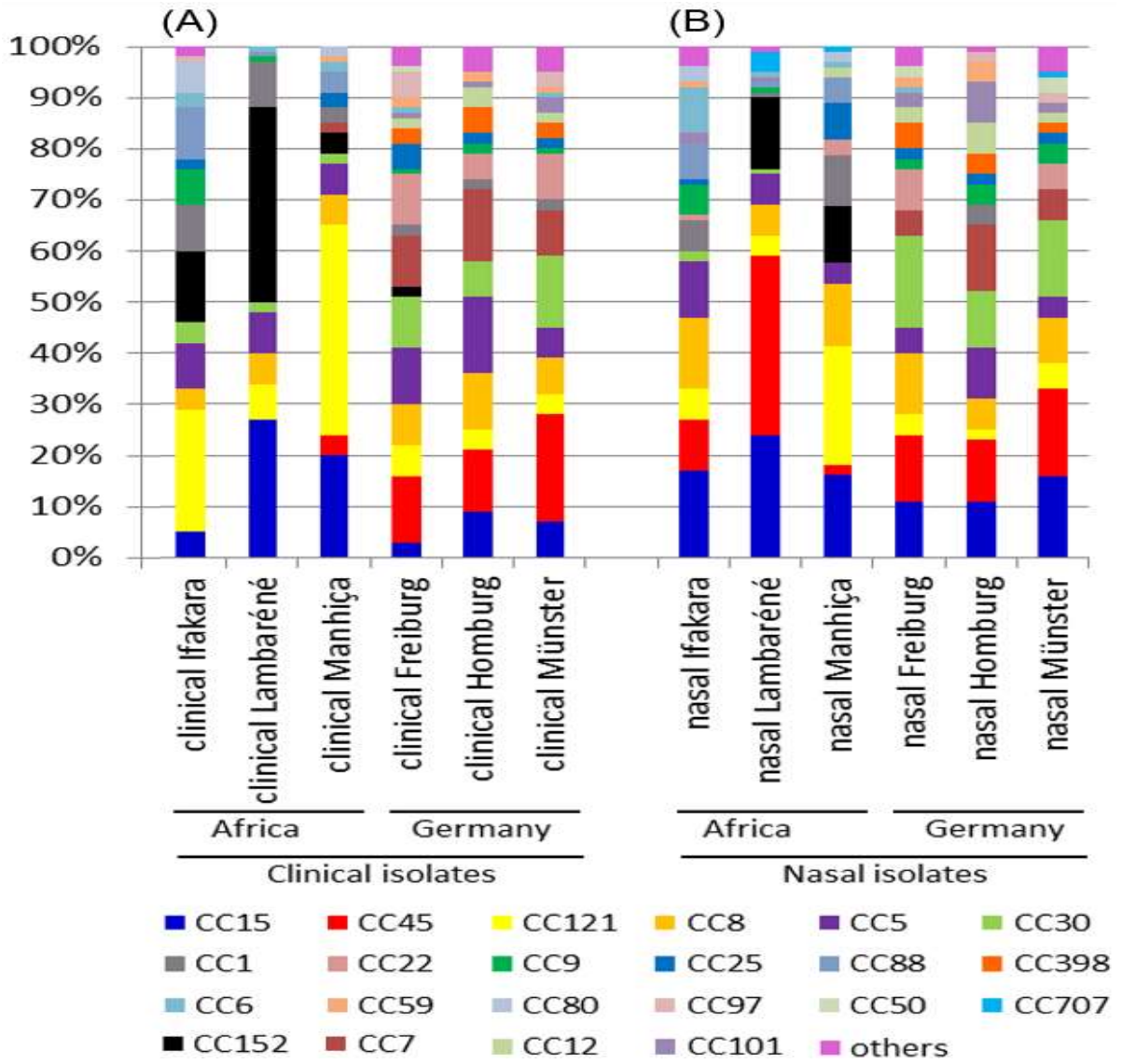


Figure 23: **Relative distribution of the 22 most prevalent clonal complexes (CC) of community-associated *Staphylococcus aureus*.** The percentage distribution of the individual African and German study sites with discrimination of (A) clinical isolates and (B) nasal isolates are shown. CCs and sequence types with less than six isolates were summarized as group others.

4.3.2.3 Antibiotic resistance and virulence-related gene presence as a function of geographical origin and clinical significance of SA isolates

The following paragraphs summarize the comparisons of the hybridization results of selected DNA MA targets of regulatory SA genes (Table 15), resistance genes (Table 16), toxin and virulence genes (Table 17), and adhesion genes (Table 18) with respect to the geographic or clinical/nasal isolate origin.

The SA species markers *rrn*, *gap*A, *kat*A, *coa*, *nuc*1, *spa*, and *sbi* were positive in virtually all isolates tested in this work (≥99.8%), thus providing an internal quality control for the species identification and for the performance of the DNA MA (Table 15).

The genes encoding for major regulatory factors of the virulence factor synthesis in SA such as SarA, SaeS and the sensor protein VraS*,* a member of the two-component regulatory system VraS/VraR, were detected in all except of two isolates (Table 15). The accessory gene regulator-type I (*agr*I) locus was more prevalent in German isolates (n=331 [55%] vs. n=209 [35%], p<0.001), while the accessory gene regulator-type IV (*agr*IV) encoding genes were detected more often in African isolates (n=221 [37%] vs. n=38 [6%], p<0.001).

Table 15: **Comparison of the DNA microarray target recognition of community-associated *Staphylococcus aureus* genes encoding regulatory factors with respect to the geographic or clinical isolate origin.**

	All isolates			**Clinical isolates**			**Nasal isolates**		
genes	**African, n**	**German, n**	**p***	**African, n**	**German, n**	**p***	**African, n**	**German, n**	**p***
sarA	600	599	ns	300	300	ns	300	299	ns
saeS	600	600	ns	300	300	ns	300	300	ns
vraS	599	600	ns	300	300	ns	299	300	ns
agrI total.	**209**	**331**	**<0.001**	**99**	**179**	**<0.001**	110	152	ns
agrII.total	161	151	ns	78	68	ns	83	83	ns
agrIII.total	93	84	ns	50	36	ns	43	48	ns
***agrIV* total**	**221**	**38**	**<0.001**	**128**	**20**	**<0.001**	**93**	**18**	**<0.001**

*p-values were calculated by the Chi-square test with Hommel correction for multiple testing using R Studio; ns= not significant

In Table 16 the number of resistance genes found in Africa and Germany for the nasal

and clinical cohort is shown, respectively.

Screening for the staphylococcal methicillin resistance gene cassette SCC*mec*, carrying the methicillin resistance gene *mecA*, revealed that a total of 40/1200 (3%) strains were found to carry this cassette. MRSA strains were equally distributed in isolates from Africa (n=17, 3%) and from Germany (n=23, 4%; Table 16). In Africa the MRSA isolates were equally distributed between clinical (n=9, 3%) and nasal isolates (n=8, 3%), whereas in Germany significantly more MRSA isolates could be identified in clinical (n=21, 7%) than in nasal isolates (n=2, 1%; p<0.0001) (Table 16).

The erythromycin resistance genes *erm*(A) and *erm*(C) were detected in isolates belonging to 12 different CCs (n=43, 4%) and 21 CCs (n=134, 11%), respectively. *Erm*(A) was more often observed in German isolates than in African isolates (n=40 [7%] vs. n=3 [1%], p<0.001; Table 16) both for clinical isolates (n=24 [8%] vs. n=2 [1%], p<0.001) and for nasal isolates (n=16 [5%] vs. n=1 [0,3%], p<0.001). On the other hand *erm*(C) was identified in higher numbers in the African isolate set than in the German isolate set (n=91 [15%] vs. n=43 [7%], p<0.001); Table 16). This was again visible both for the clinical subgroup (n=41 [14%] vs. n=24 [8%], p<0.05) and for the nasal subgroup (n=50 [17%] vs. n=19 [6%], p<0.001; Table 16). The tetracycline resistance gene *tet*(K) was found in all CCs in Africa except for CC30. In Germany it was detected occasionally in CC5 (n=1/51), CC7 (n=3/57), CC8 (n=1/53), CC15 (n=4/57), CC97 (n=1/12), CC121 (n=1/25), CC188 (n=1/5), and CC398 (n=3/22). A significantly larger proportion of *tet*(K) and *tet*(M) was detected in African isolates irrespective of their clinical or nasal origin (n=211 [35%] vs. n=15 [3%] for *tet*(K), and n=47 [8%] vs. n=7 [1%] for *tet*(M); p<0.001; Table 16). A high prevalence for the fosfomycin resistance gene *fosB* was identified in African and German clinical/nasal isolates (n=193 [64%] vs. n=158 [53%] for African and German clinical isolates, and n=181 [60%] vs. n=185 [62%] for African and German nasal isolates; Table 16).

Table 16: **Comparison of the DNA microarray target recognition of selected resistance genes of community-associated *Staphylococcus aureus* isolates with respect to the geographic or clinical/nasal origin.**

	All isolates			Clinical isolates			Nasal isolates		
genes	**African, n**	**German, n**	**p***	**African, n**	**German, n**	**p***	**African, n**	**German, n**	**p***
mecA	17	23	ns	9	21	ns	8	2	ns
***erm*(A)**	**3**	**40**	**<0.01**	2	24	ns	1	16	ns
***erm*(C)**	**91**	**43**	**<0.001**	**41**	**24**	**<0.01**	50	19	ns
***tet*(K)**	**211**	**15**	**<0.001**	**113**	**9**	**<0.001**	**98**	**6**	**<0.001**
***tet*(M)**	**47**	**7**	**<0.001**	29	7	ns	**18**	**0**	**<0.05**
fosB	374	343	ns	193	158	ns	181	185	ns
qacA	0	4	ns	0	4	ns	0	0	ns
qacC (total)	7	16	ns	2	10	ns	5	6	ns
sdrM	**594**	**558**	**<0.001**	299	273	ns	295	285	ns

*p-values were calculated by the Chi-square test with Hommel correction for multiple testing using R Studio; ns= not significant

The distribution of selected virulence genes encoding for toxins, leukocidins, capsule syntheses factors and adhesins in African and German isolates was examined (Table 17).

A major difference between African and German isolates was observed for the Panton-Valentine leukocidin (PVL) encoding genes (*lukF-*/*lukS-PV*). Overall, *lukF-PV* and *lukS-PV* were found in 287/1200 (24%) isolates belonging to 13 different CCs, of which 272 (45%) were detected in Africa and only 15 (3%) in Germany (EEF, sheet A1). On the CC level, most *lukF-PV* and *lukS-PV* positive isolates belonged to the following CCs: CC15 (n=28), CC88 (n=20), CC121 (n=94), and CC152 (n=78). Statistical analysis revealed a significant association of the PVL encoding genes with clinical vs. nasal origin (n=202 [34%] vs. n=85 [14%], p<0.0001; Table 17). Occurrence of *lukF-/lukS-PV* was highly associated in CC5 with African origin (13/44 [30%] vs. 0/51 [0%] for African vs. German isolates, p<0.0001) and a similar situation was seen for CC15 (27/109 [25%] vs. 1/57 [2%], p<0.0001), CC22 (3/4 [75%] vs. 1/37 [3%], p<0.0015), CC30 (8/11 [73%] vs. 2/75 [3%], p<0.0001), and CC121 (87/105 [83%] vs. 7/25 [28%], p<0.0001) (EEF, sheet A8).

Table 17: **Comparison of the DNA microarray target recognition of selected community-associated *Staphylococcus aureus* virulence genes with respect to the geographic or clinical/nasal isolate origin.**

	All isolates			Clinical isolates			Nasal isolates		
genes	**African, n**	**German, n**	**p***	**African, n**	**German, n**	**p***	**African, n**	**German, n**	**p***
tst1	36	67	ns	12	23	ns	24	44	ns
sea	92	68	ns	44	28	ns	48	40	ns
seb	**114**	**45**	**<0.001**	**72**	**23**	**<0.001**	**42**	**22**	ns
sec	**49**	**92**	**<0.05**	**19**	**57**	**<0.01**	**30**	**35**	ns
sed	**21**	**52**	**<0.001**	**9**	**35**	**<0.05**	**12**	**17**	ns
seh	34	26	ns	18	12	ns	16	14	ns
sej	25	41	ns	10	27	ns	15	14	ns
sek	56	27	ns	27	14	ns	29	13	ns
sel	**50**	**92**	**<0.05**	**20**	**57**	**<0.01**	**30**	**35**	ns
***egc* total**	**253**	**332**	**<0.01**	**120**	**173**	**<0.001**	**133**	**159**	ns
seq	56	27	ns	27	14	ns	29	13	ns
ser	20	37	ns	8	24	ns	12	13	ns
lukF.PV	**272**	**15**	**<0.001**	**187**	**15**	**<0.001**	**85**	**0**	**<0.001**
lukS.PV	**273**	**15**	**<0.001**	**188**	**15**	**<0.001**	**85**	**0**	**<0.001**
etA	39	24	ns	19	9	ns	20	15	ns
etB	21	7	ns	12	4	ns	9	3	ns
etD	21	17	ns	10	9	ns	11	8	ns
cap5	237	224	ns	107	126	ns	130	98	ns
cap8	362	376	ns	193	174	ns	169	202	ns

*p-values were calculated by Chi-square with Hommel correction for multiple testing using R Studio; ns= not significant

Analysis of the capsule gene distribution revealed that the type 5 capsule genes (*cap5*) were associated with CC5 (n=92/95, [97%]), CC8 (n=101/101, [100%]), CC22 (n=39/41, [95%]) and CC152 (n=79/81, [98%]), while all other CCs, except CC1, carried type 8 capsule genes (*cap8*) (EEF, sheet A8). The capsule type for CC1 isolates was found to be inhomogeneous, with 38 of 48 (79%) isolates being *cap8*-positive and the remaining 10 isolates (21%) being *cap5*-positive (EEF, sheet A8). In nasal samples, the capsule type 5 encoding genes were predominant in isolates from African volunteers compared to isolates from German volunteers (n=130/238, [54%] vs. n=98/224, [44%], p<0.01). In comparison, *cap8* harboring isolates were more frequent among nasal samples collected in Germany in comparison to nasal samples in Africa (n=202/376, [54%] vs. n=169/362, [47%], p<0.01; Table 17).

For other genes such as *fib* (encoding a fibrinogen binding protein; n=448, 75% vs. n=372, 62%), *ebh* (encoding the cell wall-associated fibronectin-binding protein;

n=594, 99% vs. n=553, 92%), *fnbB* (encoding a fibronectin binding protein; n=587, 98% vs. n=522, 87), and *sasG* (SA surface protein G; n=370, 62% vs. n=253, 42%) higher frequencies were detected in African compared to German isolates (p<0.001 for all comparisons; Table 18). In contrast, the *map* (major histocompatibility complex class II analog protein) gene, also referred to as *eap* (extracellular adhesion protein), which is exclusively found in SA, was more often identified in German isolates than in African isolates (n=582, 97% vs. n=494, 82%; Table 18). In detail, 18% (n=106) of the *map/eap* negative isolates belong to the African cohort and only 3% (n=18) to the German cohort. Notably, all CC152 isolates (n=82, with 80 isolates of African origin and only 2 isolates of German origin) lacked a *map/eap* specific signal, pointing towards the possibility that the *eap* nucleotide sequence of CC152 isolates might not match with the *map/eap* target sequence provided by the DNA MA. A limited number of *map/eap* positive isolates was also observed among African CC1 isolates (n=16, 42%), while the *map/eap* recognition in the other CCs approached 100% irrespective of the isolate origin.

Table 18: **Comparison of the DNA microarray target recognition of selected adhesion genes of community-associated *Staphylococcus aureus* with respect to the geographic or clinical isolate origin.**

	All isolates			**Clinical isolates**			**Nasal isolates**		
	African, n	**German, n**	**p***	**African, n**	**German, n**	**p***	**African, n**	**German, n**	**p***
fib	**448**	**372**	**<0.001**	231	184	ns	217	188	ns
ebh	**594**	**553**	**<0.001**	299	270	ns	295	283	ns
fnbB	**587**	**522**	**<0.001**	294	262	ns	293	260	ns
sasG	**370**	**253**	**<0.001**	161	134	ns	**209**	**119**	**<0.001**
map/eap	**494**	**582**	**<0.001**	233	290	ns	261	292	ns

*p-values were calculated by Chi-square with Hommel correction for multiple testing using R Studio; ns= not significant

4.3.3 Isolate cluster analysis

The combination of 2D-PCA and silhouette analysis identified nine isolate clusters (labelled 1 to 9, Figure 24) that correspond with the CC attribution of the isolates. In addition to the nine identified isolate clusters, a heterogeneous isolate group comprising isolates of different CCs was seen on the left side, in the center of the plot. Further, the silhouette analysis showed that each isolate's position in the PCA plot matched well within the identified cluster areas with a silhouette value of 0.59. A silhouette coefficient close to 0 means that this isolate is close or on the threshold value between two adjacent clusters.

Three of the nine clusters (#1, 2, and 3) were located on the left plot side of the PCA panel and six separate isolate clusters (#4-9) on the right side. The clusters on the left plot side preferentially consisted of 'African' isolates (clusters #1 [CC8], #2 [CC15], #3 [CC121]). In comparison, the clusters on the right side of the plot were prevalently formed by isolates of 'German' origin (clusters #5 [CC22], #6 [CC398], and #9 [CC30]). The CC121 cluster (#3) was closely associated with the heterogeneous group in the central part of the two-dimensional PCA plot. Remarkably, the PCA separated the CC45 isolates into two distinct clusters (cluster #7 and cluster #8) of which cluster #8 include isolates of African and German origin (Figure 24A), while cluster #7 consisted only of nasal African isolates (Figure 24A, B).

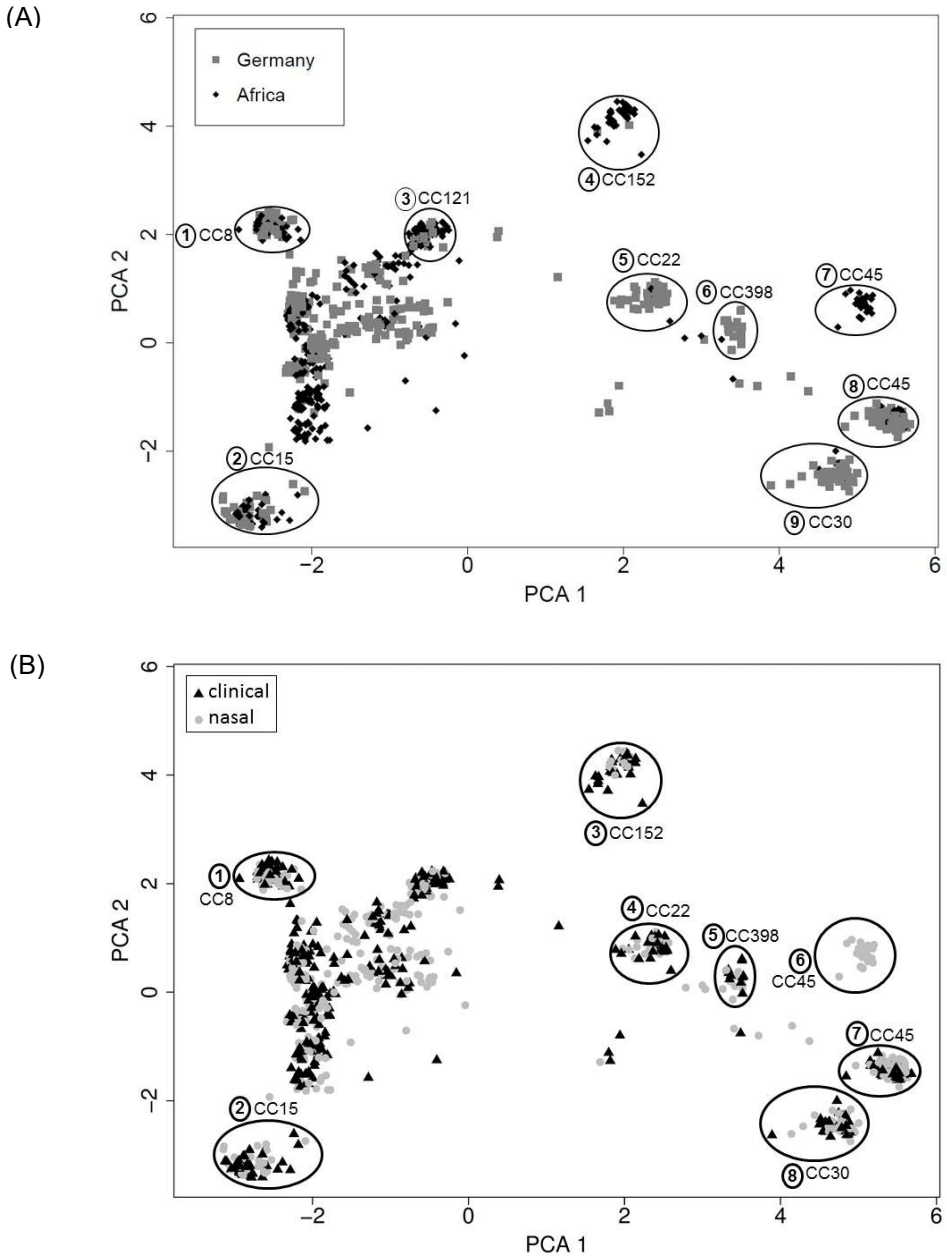


Figure 24: **Two-dimensional principal component analysis (PCA) of (A) 600 African and 600 German or (B) 600 clinical and 600 nasal community associated *Staphylococcus aureus* isolates.** PCA is based on the presence or absence of hybridization profiles of the 334 target sequences of each isolate determined by DNA microarray. (A) Each black diamond represents an African isolate and each grey square represents a German isolate. (B) Clinical isolates were denoted by black triangles and nasal isolates by grey circles. Clonal complex (CC) isolate clusters are indicated and numbered. Close proximity of triangles/squares indicates high similarity between these isolates.

4.3.4 DNA MA and whole genome sequencing concordance analysis

WGS was performed because DNA MA technique can't detect allelic variants or gene polymorphisms, which are not represented on the array by the sequence targets and thus resulting in false negative results. The existence of such an allelic variant was assumed for the gene *map/eap* because previous studies indicated *map/eap* to be present in the clear majority of SA. A lack of *map/eap* target recognition was observed in 124/1200 isolates, with 18% (n=106) *map/eap* negative isolates in the African cohort and 3% (n=18) in the German cohort. Absence of an *eap/map* signal was especially recognized for CC152 isolates, which all were negative for this signal. Therefore, WGS was performed for exemplary isolates of the CC152. An aberrant *map/eap* gene were identified in this CC type, which lacked the DNA region represented on the DNA MA (Strauss *et al.*, 2016).

However, when results of WGS were compared with the DNA MA data, it was found that both techniques were highly (but not fully) reliable with respect to the gene/allele identification (with 1.7% WGS errors and 1.8% MA errors) (Strauss *et al.*, 2016).

4.4 Subtyping of SA isolates by MALDI-TOF

To test if it is possible to subtype SA strains based on their mass spectra with regard to methicillin resistance, geographic origin of the isolates, *spa*-type and strain lineage (CC), the mass spectra profiles of 219 SA isolates (51 MRSA, 168 MSSA) of the African-German-StaphNet (study C; n=127, section 4.3) and the admission prevalence screening of Saarland (study A; n=92, section 4.2) were compared.

4.4.1 Evaluation of correct assignment of SA based on their total mass spectra to methicillin resistance, geographic origin, *spa*-type and strain lineage (CC/ST)

The 219 isolates belonged to 20 different CCs with the predominant CCs being CC5 (n=58, 26:5%), CC8 (n=20; 9:1%), CC15 (n=17; 7:8%), CC30 (n=11, 5%), CC45 (n=28, 12.8%), as well as CC121 (n=26, 11.9%). Further, the isolates were of 108 different *spa*-types, dominated by *spa*-types t003 (n=29, 13.2%) and t002 (n=11, 5.0%) (Figure 25A). For 6 *spa*-types only three isolates, for 12 *spa*-types only two isolates and for 77 *spa*-types only a single isolate were found (Table 19).

Table 19: **Number of different *spa*-types and assigned *Staphylococcus aureus* isolates.**

***spa*-type**	**number of different *spa*-types**	**number of isolates for each *spa*-type**
t003	1	29
t002	1	11
t084, t015	2	9
t012, t272	2	6
t701, t1476	2	5
t008, t314, t504, t645, t939	5	4
t148, t267, t304, t311, t1849, t4499	6	3
t026, t085, t091, t131, t159, t346, t714, t941, t4198, t5122, t11586, t12385	12	2
t005, t010, t011, t018, t019, t021, t022, t040, t045, t050, t065, t073, t078, t098, t127, t197, t237, t273, t279, t306, t310, t339, t435, t437, t451, t481, t493, t571, t584, t620, t625, t630, t690, t702, t711, t729, t887, t934, t940, t2393, t2720, t2771, t3092, t3195, t3772, t4044, t4333, t5562, t6172, t6208, t6431, t8786, t8830, t8831, t8863, t8864, t10333, t10419, t1079, t1081, t1114, t1171, t2164, t2239 t1510, t1579, t1689, t1991, t11282, t11311, t11312, t11313, t11314, t11318, t11375, t11583, t11584	77	1

To analyze if the total mass spectra-based isolate clusters correspond with *spa*-type (Figure 25A), CC (Figure 26), resistance profile (MRSA/MSSA) (Figure 25B) or geographic origin (Figure 25C) PCAs for each of these three categories were performed.

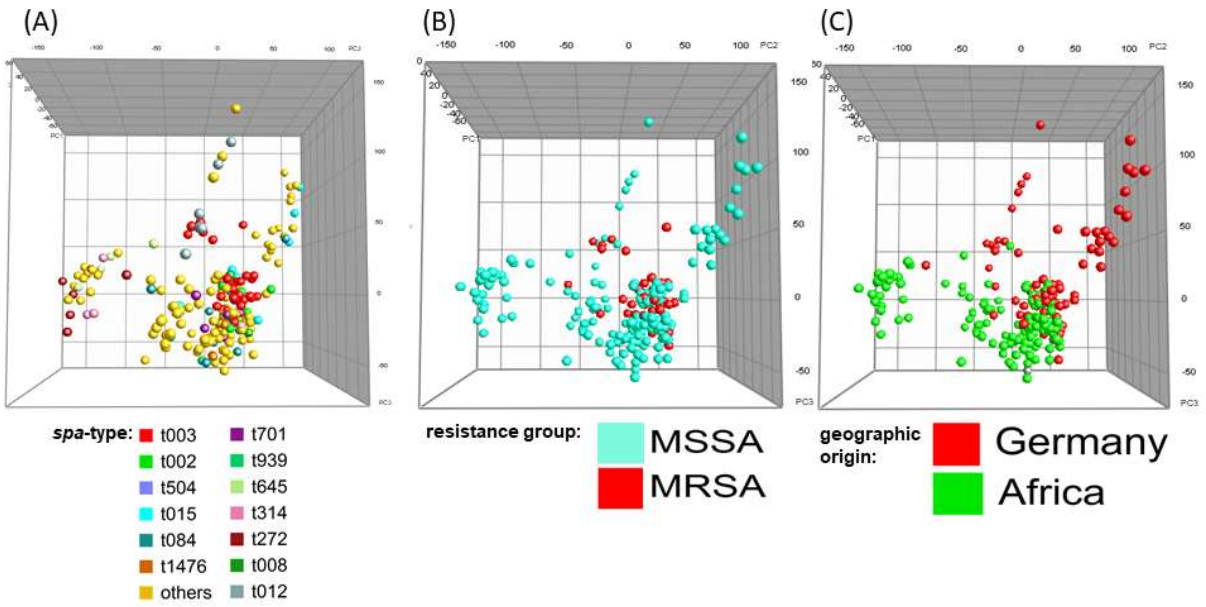


Figure 25: **Three-dimensional principal component analysis (PCA) of mass spectra of 219 African and German methicillin-susceptible *Staphylococcus aureus* (MSSA) and methicillin-resistant *Staphylococcus aureus* (MRSA) isolates**. PCA based on the mass spectra profiles of 219 *Staphylococcus aureus* genotypically well characterized isolates, projected onto the three principal components PC1, PC2 and PC3. Each spot represents an individual isolate that is highlighted according to (A) the *spa*-type, (B) the resistance profile (MSSA or MRSA) (C) the geographic origin (African or German). The PCA clustering pattern reflect the similarity between isolates. Higher similarity is represented by a higher proximity of the isolates.

The PCA showed that isolates of the same *spa*-type (Figure 25A) or carrying methicillin resistance (Figure 25B) were not separated into different groups based on their mass spectra. Separation into different groups would mean that in the PCA dots of the same colour, representing isolates of the same group, build separate clusters.

In addition, PCAs were performed to identify if the SA mass spectra correspond with the CCs they belong to (Figure 26). The results of the PCA identified that there were no CC5, CC8, CC45 or CC88 specific isolate groups based on the isolates' mass spectra profiles (Figure 26B/C, 26E/F). Mass spectra profiles of these four CCs could not be distinguished from mass profiles of all other CCs. However, a different picture emerged for the mass spectra of the CC121 isolates. The spectra of the CC121 isolates formed a separate group (Figure 26D, red) in comparison to the mass spectra of the isolates of all other CCs. This strongly suggests that CC121 isolates can be identified based on their mass spectra profile, while there is no possibility to identify isolates of other CCs based on their mass spectra profile.

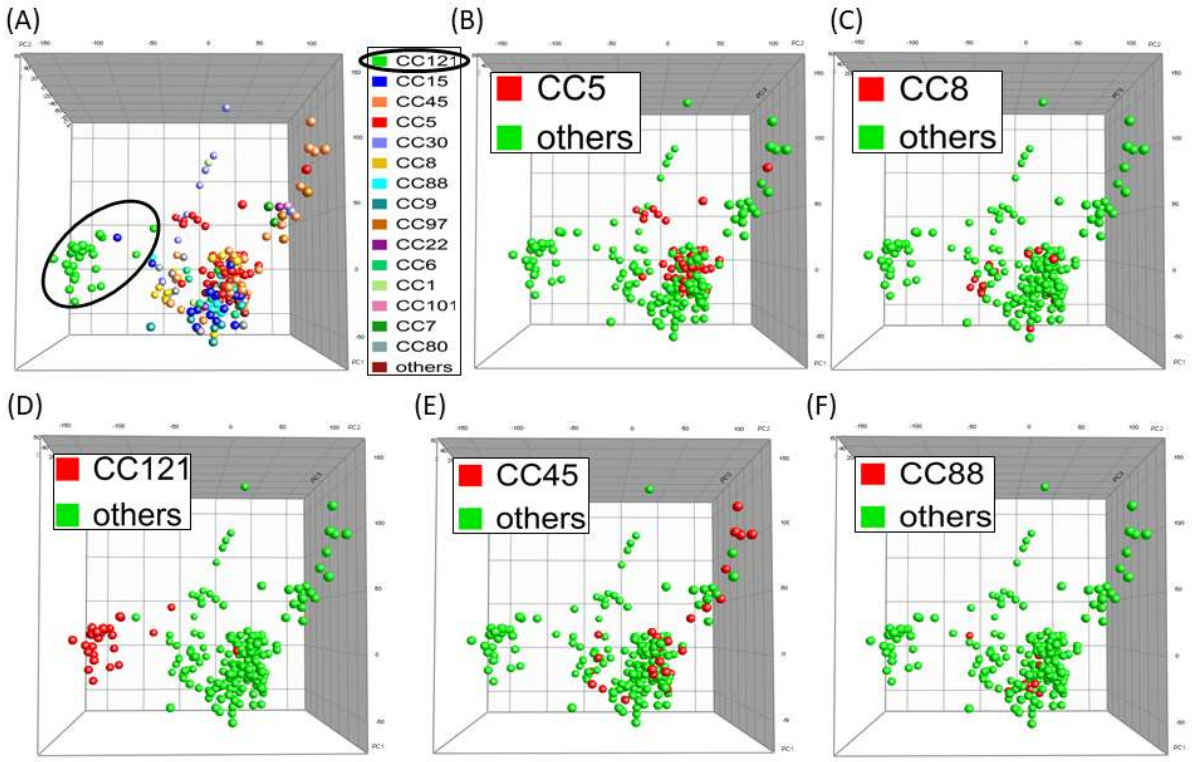


Figure 26: **Three-dimensional principal component analysis (PCA) of 219 *Staphylococcus aureus* (SA) mass spectra with color codes highlighting different clonal complexes (CCs).** The PCA was projected onto the first three principal components (PC1, PC2 and PC3) that separates the data set**.** (A) Clustering of the SA mass spectra with color codes for 15 different CCs contributing at least two isolates, and three single isolates summarized as group “others”. (B-F) Clustering of the mass spectra of specific CCs (red), (B) CC5, (C) CC8, (D) CC121, (E) CC45 and (F) CC88 in comparison to all other isolates (green).

The PCA results were confirmed in single identification projects for each of the four previous mentioned categories (Table 20). Results are shown as performance of the classifier support vector machine for the four predefined groups in the four identification projects (Table 20).

Table 20: **Collected results of the performance of the classifier support vector machine for the four predefined groups in the four identification projects.** The table shows the results of the objective estimation of the classifiers performance for calculation of the likelihood of correct or incorrect assigned *Staphylococcus aureus* (SA) isolates to the predefined groups (geographic origin [African/German)] resistance profile [methicillin resistant/susceptible SA {MRSA/MSSA}], *spa*-type [t003 vs. all other *spa*-types] and clonal complex [CC121 vs. all other isolates]). True positive assigned isolates are e.g. German isolates correctly assigned to the German group. False negative assigned isolates are e.g. German isolates which were assigned to the African group. False positive assigned isolates are e.g. African isolates which were falsely assigned to the German group.

Class name	Isolate numbers	True positive assigned isolates (%)	False negative assigned isolates (%)	#False positive assigned isolates (%)
Geography	219	90.87	9.13	
Germany	103	90.29	9.71	9.71
Africa	116	91.38	8.62	8.62
Resistance	219	84.93	15.07	
MRSA	168	89.29	10.71	8.93
MSSA	51	70.59	29.41	35.29
spa-types	219	89.04	10.96	
t003	29	62.07	37.93	44.83
others	190	93.16	6.84	5.79
Clonal complexes	219	97.72	2.28	
CC121	26	92.31	7.69	11.54
others	193	98.45	1.55	1.04

#The right column shows the percentages of additional isolates the algorithm had assigned to the different classes i.e. 9.7% of the African isolates were judged to be of German origin.

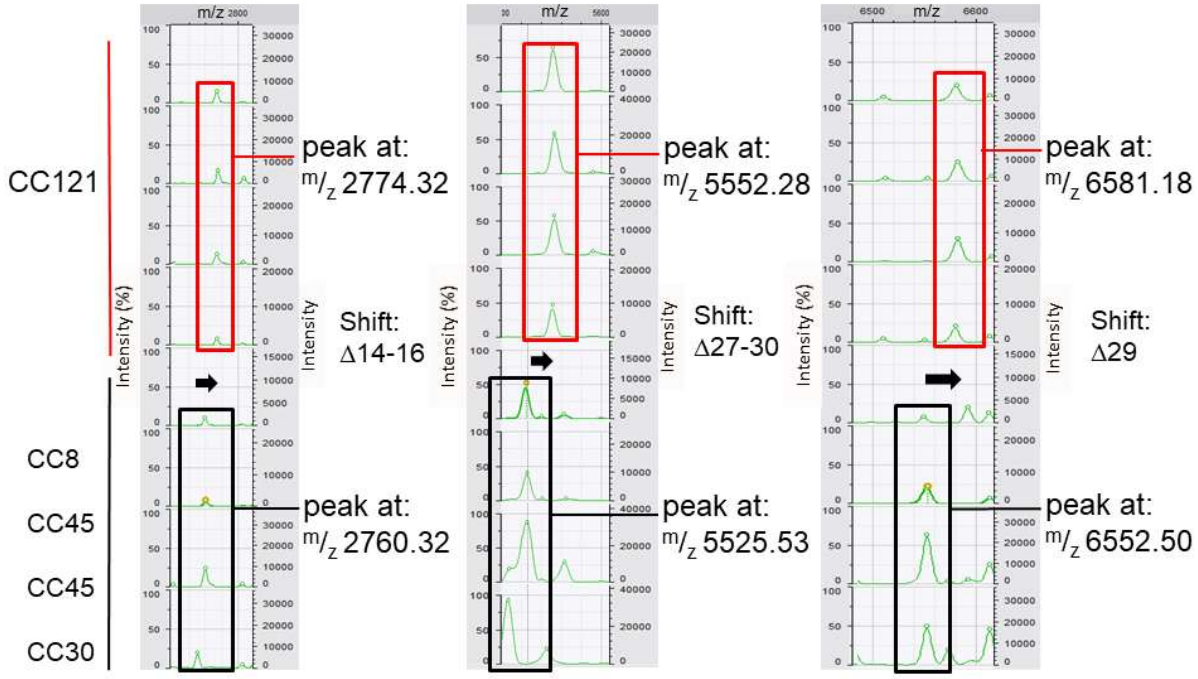
Total mass spectra analysis of African versus German isolate origin indicated a tendency of geographically dependent isolate clustering with true positive assignment of about 90% of the German SA isolates and 91% of the African SA. A high number of MSSA isolates was misclassified (29% false negative and 35% false positive), while 89% of MRSA isolates were assigned to the correct group based on their total mass spectra. A substantial part of the predominant *spa*-type t003 isolates were also misclassified with 38% false negative and 45% false positive (Table 20). In the assignment analysis of total mass spectra isolates of CC121 versus total mass spectra belonging to all other CCs in the study showed a tendency of true positive assignment of CC121 isolates of about 92% and of SA isolates of other CCs of about 98%.

In conclusion, the identification projects identified no reliable correspondence of the SA isolates' clusters based on the total mass spectra with the geographic origin, methicillin resistance profile or *spa*-type. However, a positive tendency for correct assignment of CC121 isolates were detected.

4.4.2 Analysis of potential CC121 specific total mass spectra peaks

Next, it was examined if there are characteristic mass spectra peaks discriminating CC121 mass spectra from all other SA mass spectra by visual comparison (Figure 27). Therefore, the mass spectra profiles of four randomly selected CC121 isolates and four other isolates of CC8 (n=1), CC30 (n=1) and CC45 (n=2) (Figure 27) were regarded. The visual comparison showed that all four CC121 isolates showed three peak shifts in comparison to the spectra obtained from non-CC121 isolates from a mass to charge ratio (m/z) of $^{m}/_{z}$ 2760.32 to $^{m}/_{z}$ 2774.32, from $^{m}/_{z}$ 5525.53 to $^{m}/_{z}$ 5552.28 and from $^{m}/_{z}$ 6552.50 to $^{m}/_{z}$ 6581.18 (Figure 27). This let assume that these peaks are potentially characteristic CC121 peak shifts ($p<0.001$).

These potentially characteristic CC121 peak shifts were statistically confirmed as significantly characteristic CC121 peaks ($p<0.001$) with the Mann-Witney test and Bonferroni correction for multiple correction based on the total mass spectra peak list of all 219 SA isolates.



m/z= mass to charge ratio

Figure 27: **Mass spectra sections of three CC121 specific peak shifts**. Comparison of four CC121 (red) isolates and four isolates of three other clonal complexes (CCs), namely CC8 (n=1), CC30 (n=1) and CC45 (n=2), (black). The mass to charge ratio (m/z) of three peaks were shifted in the CC121 isolates.

A web based in-silico database search was performed to identify the corresponding proteins to these peaks. The best hits with the highest probabilities were the aspartate 1 decarboxylase PanD (~2806Da), and the two ribosomal proteins RpmG (5375Da) and RpmD (6554Da), respectively. In order to evaluate if the peak shifts might be caused by mutations, the nucleotide sequences of the corresponding genes *panD*, *rpmG,* and *rpmD* of the CC121, CC45, CC30 and CC8 SA isolates (data not shown) were compared by our cooperation partners in Münster. However, this comparison showed no nucleotide alterations in any of the three genes so that mutations in these genes can be excluded as reason for the differences in the CC121 mass spectra. An explanation could be posttranscriptional modifications such as phosphorylation.

5 Discussion

5.1 Differentiation of closely related SA strains of a hospital admission prevalence screening study in Saarland by DNA microarray

In this study of a SA cohort of the tertiary care university hospital of Saarland (section 4.1), the DNA MA platform of Identibac® (Alere Technologies GmbH, Jena, Germany) was applied as a state-of-the-art genotyping method with affordable costs and technical robustness for genotyping of matched nasal MSSA and MRSA isolates collected upon patient admission. The obtained DNA MA data of this small study were used to evaluate three bioinformatic methods to analyze which is the most appropriate DNA MA data analysis method in comparison to *spa*-typing as reference method. Further, the MRSA and MSSA clonal lineages of community-associated SA isolates of the federal state of Saarland and the resistance and virulence gene distribution in MRSA in comparison to MSSA were evaluated.

The analysis of the MRSA and MSSA revealed that the MSSA cohort revealed a large CC diversity (n=12) in comparison to the analysis of the MRSA isolate set displaying a limited number of CCs (n=5). The MRSA cohort further showed a high prevalence for *spa*-type t003 in CC5. Overall higher diversity of MSSA compared with MRSA has been described in previous studies (Deurenberg & Stobberingh, 2008, Ghasemzadeh-Moghaddam *et al.*, 2011, Goering *et al.*, 2008).

The low CC diversity, the predominance of CC5 within the MRSA strain set might be considered at first sight as a limitation of this study. However, it might also be regarded as a strength, as the discrimination of such closely related strains is a major challenge when analyzing HA-MRSA isolate cohorts of geographically restricted studies. Such high-resolution strain differentiation is of special importance to confirm potential outbreak situations from accumulated MRSA detections to identify the starting point and to interrupt the responsible transmission routes.

The DNA MA from Alere Technologies (Jena, Germany) provides information for a comprehensive number of genetic loci that are associated with resistance formation and infectivity of SA. The chip contains 334 gene fragments of genes encoding for virulence factors, resistance genes as well as regulatory elements, and hybridizing DNA of a specific isolate. This chip allows to test whether a gene is present or absent

in this isolate. The hybridization pattern can be furthermore used to assign the isolate to a known CC or ST (based on the image of the microarray, analyzed by the designated reader and iconoclust software provided by Alere Technologies GmbH, Jena). A major advantage of this method is that isolates with identical *spa*-types and/or CCs/STs might be further distinguishable, when strains differ in the genes tested by this array. However, due to the large amount of data that is generated by this method, this information is not easily visible and often requires advanced bioinformatic evaluation methods, especially if a larger set of isolates should be compared.

This study suggested that this method is particularly well suited in answering detailed questions related to SA typing, infection control and SA epidemiology when the DNA MA data is analyzed by appropriate software tools, for example to investigate if two SA isolates are genetically identical or how they differ and/or if they are from the same strain lineage. While splits graph and HAC are bioinformatic methods typically used in phylogenetic analyses, PCA is a dimension reduction model that is especially used in genome wide association analyses. Based on this evaluation, an analysis algorithm has been developed in this work for further DNA MA genotyping projects for analysis and best approach for result presentation.

Additionally, splits graph based on DNA MA provided a better resolution based on the high number of analyzed loci with a more detailed view how the single isolates or isolates groups are placed in relation to each other compared to splits graph based on *spa*-typing data.

This means, that it is easy to evaluate or recognize in outbreak situations if isolates/ isolate groups belong to the same strain or give a hint that isolates even not being totally identical, but may descend from the same mother strain. This allowed in potential outbreak situations to decide if there is a real outbreak situation to handle with or if the observed accumulation in fact represents different introductions. Furthermore, DNA MA data can be used to select isolates where further typing is necessary i.e. WGS.

Comparison of the both splits graph based on *spa*-typing and DNA MA for the predominating isolates of CC5 showed that the splits graph based on DNA MA has a much higher resolution. It reveals several isolates subgroups not presented by splits graph based on *spa*-typing data and how each single isolate is placed in relation to the other isolates as could be seen for R3-R14, R17-R21 and R22-R24.

Splits graph based on *spa*-typing data allowed a correct assignment of five more isolates (S36 of *spa*-type t306, R4, R8, R17 and R18 of *spa*-type t504) to CC5 than described till yet by the Ridom *spa*-server database but as described before in literature (Deurenberg *et al*., 2009, Cuny *et al*., 2015).

Comparison of splits graph based on *spa*-typing and DNA MA showed that splits graph based on DNA MA was able to assign two more isolates to a isolate group CC (S12 to CC45 and S43 to CC5 cluster) than splits graph based on *spa*-typing data.

These findings showed, that splits graph presentation of DNA MA data make it relatively easy to see how isolates may be related even without knowledge of an expert and can thus be an important tool for infection control nurses.

Further, DNA MA data seemed to overcome some of the limitations of *spa*-typing as established SA typing system.

MLST still proved the framework for epidemiological understanding. However, MLST is expensive and laborious because of the higher number of analyzed genes. MLST is based on the sequencing of fragments of seven housekeeping genes that, in evolutionary terms, slowly accumulate genetic variations (Aanensen & Spratt, 2005). This could attribute to an insufficient discriminatory potential in epidemiologically unrelated strains because of the low or missing number of differences in the analyzed genes. In this case epidemiologically unrelated strains might falsely be assigned to the same ST or CC. MLST is unsuitable for the direct tracking of fast changes. This is the reason why MLST is not suited for outbreak analysis – but provide reliable information as a basis for SA epidemiology and is the method of choice for larger epidemiological studies.

In comparison, *spa*-typing is based on the sequencing of a single gene. It is a relatively cheap method, which could be transferred and directly applied in other labs leading to identical, unambiguous results. A remarkable degree of *spa* gene repeat polymorphisms within specific clonal lineages often provides a higher discriminatory power of this method if compared to MLST (Strommenger *et al*., 2006). However, a disadvantage of *spa*-typing is the risk of evolutionary convergence, reflecting homoplasies because of the high mutation rate of the *spa*-locus. This may not allow a correct discrimination or may lead to incorrect assignment of isolates to clones as it might be the case for five MSSA isolates in this study with regard to epidemiological

analysis (Basset *et al.*, 2012, Harris *et al.*, 2010, Nübel *et al.*, 2011, Nübel *et al.*, 2008). Reason could be the exchange or deletion of *spa*-repeats leading to *spa*-types typically not found and described in literature for isolates of these CCs. In literature, *spa*-type t273 were found in SA isolates of CC1, *spa*-type t306 were described in CC5 isolates, t018 in CC30 isolates and t493 in CC182 isolates (Boswihi *et al*., 2016, Krupa *et al.*, 2015, Nulens *et al.*, 2008, Rijnders *et al*., 2009) while in this study isolates of these *spa*-types clustered with isolates of other CCs. Thus, in this single locus typing method evolutionary convergence may lead to indifferent clustering of these isolates by *spa*-typing in comparison to the DNA MA analyzing more loci.

Another disadvantage is that only one gene region is analyzed and that putative changes in other regions are not detectable by this method. Genetic changes may accidently blur the picture. Although *spa*-typing showed a high mutation rate, the resolution of this method is not discriminatory enough when single clones as t003 dominated as seen this work. Thus, *spa*-typing is a highly useful method for MSSA but only with minor extent for monitoring spreads of MRSA because in MRSA cohorts single *spa*-types dominate often (Ruffing *et al.*, 2012, Witte *et al.*, 2008). In these cases additional methods are often needed.

In comparison to splits graph based on DNA MA data, you need to be an expert to understand how similar *spa*-types based on one single gene are for the correct interpretation of splits graph based on *spa*-typing data. The resolution of splits graphs is too low. You must be able to understand how similar the single repeat sequences are to see if they are closely related to each other. That means you need to know if there is only a single nucleotide mutation or if they are totally different repeat patterns resulting in another *spa*-type.

Although, as could be seen, the higher number of analyzed loci by DNA MA provided a better resolution in comparison to *spa*-typing. Analysis of the whole isolate cohort of 92 isolates by splits graph based on DNA MA showed limitation for the resolution of the CC5 isolates. Therefore, it was needed to perform an analysis only of the CC5 isolates separately by three selected bioinformatic methods, splits graph HAC, PCA for better resolution and to determine which method provides the best result presentation.

Notably, the analysis of the DNA MA results obtained with the closely related CC5 isolates revealed different isolate subgroups by the three independently applied

bioinformatic methods. This is a critical point. The selection of the applied method could influence the result interpretation. The reason could be the different mathematical algorithm of the different methods which are based on different weightings of differences (splits graph, PCA) or similarities (HAC) of the analyzed isolates.

Splits graphs in this work were constructed based on the neighbour-joining method (Saitou & Nei, 1987) which is one of the most frequently applied distance-based methods for phylogenetic analyses. It constructs splits networks from inferred distance matrices and applies an agglomerative process to create a phylogenetic tree (Huson & Bryant, 2006). The distance matrix was calculated based on the proportion of positions which were different between two sequences or hybridization patterns.

In the PCA the full data set can be projected on the first two or three PCs without losing too much information. The data compression captures most of the variation among the samples and visualizes similarities and differences. In this work, the PCs will point along those MA probes that show the largest variation of hybridization among the samples. Samples with similar DNA MA hybridization profiles are arranged closer to each other in the 2D/3D plot compared to samples with larger differences (Caddick *et al.*, 2006).

HAC constructs a dendrogram or branching tree diagram representing the hierarchy of clusters based on the degree of similarity or number of shared characteristics of the data. The distance along the tree from one element to the next one represents the relative degree of similarity, while short distances representing high similarities.

Thus, it has to be kept in mind that independent of the chosen method the obtained result must not reflect the total reality but shows a potential structure for a better understanding of the data.

Additionally, it could be seen that each model may have its specific strengths and that the number of analyzed samples and their degree of diversity should be considered to choose the optimal method. For larger numbers of isolates, PCA appeared as the best-suited method for a direct sample overview visualized in 2- or 3-dimensional graphs. A disadvantage in this case was that closely related isolates are overlapping in the presentation of all isolates so that is was difficult to identify isolates in a group. However, it is possible to zoom in the figure with the disadvantage that not all isolates could be seen in a glance. For smaller sample sizes (e.g. <100), HAC revealed a more

detailed relationship of the isolates and a directly visible high-resolution subgrouping. This is also the disadvantage of this method because the high-resolution needs a large amount of space for presentation. This makes it an inappropriate method for construction of figures for publications with a size of maximally a DIN A4 page.

Splits graph analysis seemed to be the most appropriate method for diversity analysis in this study because of its easy applicability, direct assignment of each isolate to the branched subgroups in a 2-dimensional graph, and its free availability as open source software tool (Huson & Bryant, 2006). Moreover, also for a large amount of analyzed isolates it is possible to construct a figure which could be easily included in publications.

DNA MA revealed substantial differences in the genetic repertoires of MRSA isolates in comparison to the MSSA isolates, which were related to antimicrobial resistance genes, *agr*- and capsule-types, and virulence genes. Specifically, the β-lactamase operon was detected in 21% more MRSA (39, 84.8%) than MSSA isolates (29, 63%). This implies that 37% of MSSA are penicillinase susceptible. Practically, this means that for a high number of MSSA penicillin could be a treatment alternative. Moverover, the MRSA cohort was characterized by a significantly higher number of virulence genes of the leukocidin-, enterotoxin-, hemolysin-, protease- and adhesion gene families. However, the highly prevalent German epidemic MRSA clone (Rhine-Hesse) dominated the genetic profile of the MRSA cohort, which might have caused some bias. Therefore, evaluation of the conclusions in other studies is needed. Certainly, it may also suggest that this Rhine-Hesse virulence gene repertoire is of relevance for the successful epidemic spreading of this specific MRSA clone. Yet, since all examined study isolates were of nasal origin, no association of virulence genes and disease could be drawn.

The findings of this study indicate that implementation of the DNA MA in routine microbiological diagnostics might be a useful first line analysis tool for outbreak surveillance in an easy and fast interpretable way. Further, it is an alternative to WGS that requires both advanced hardware and software as well as complex bioinformatic post processing of the sequencing results (Strauss *et al.*, 2016). By using DNA MA analysis, the number of isolates requiring WGS (e.g. in case of an outbreak) could be reduced to a small number of closely related isolates with identical DNA MA profiles clustering in the same genetic subgroup.

5.2 Characterization of Nigerian SA isolates by DNA MA

A substantial amount of information exists for prevalent SA lineages circulating in the Western hemisphere and this information was utilized to define the gene regions that are used by the DNA MA for identification. In contrary, only very little information on the gene composition of African isolates was available at the time of the study. Therefore, it could not be excluded that the sequence targets provided by the DNA MA were not functional with African isolates. In order to address this question, a small Nigerian (Sub-Saharan African) cohort of 52 SA isolates (49 MSSA, 3 MRSA) were investigated by DNA MA. Most of the isolates were of clinical origin but a clear distinction between hospital/community-associated strains could not be made. Based on the results of a previous study, splits graph analysis was chosen for a subgroup analysis given the small study size.

The analysis showed that DNA MA appeared to be an adequate tool for molecular characterization and analysis of the clonal distribution also of African SA showing a heterogeneous and divergent nature of the examined Nigerian isolates. Nevertheless, attribution of CCs by the DNA MA software is based on the specific isolate hybridization pattern on the DNA MA rather than gene sequencing. ST/CC hybridization pattern that is not included in the database, could not be detected or differentiated from other STs/CCs by the DNA MA. This was the case for ST2427 which could not be discriminated from ST8. Thus, the DNA MA hybridization pattern database continuously needs to be updated to detect all known and new STs/CCs.

The identified heterogeneous and divergent nature of the examined Nigerian MSSA isolates corresponds to the overall higher diversity of MSSA compared with MRSA as described in previous studies (Blomfeldt *et al.*, 2013, Deurenberg & Stobberingh, 2008, Ghasemzadeh-Moghaddam *et al.*, 2011, Goering *et al.*, 2008, Rasmussen *et al.*, 2014, Ruffing *et al.*, 2012).

The potential for individual statistical comparisons was limited due to the low number of isolates included. However, as the isolates were obtained from patients of various medical institutions in Nigeria it enables us to observe some tendencies.

PVL-positive isolates were identified in half of the isolates (52%) and seven of the 12 CCs. The high prevalence of PVL correspond to the high numbers of soft and deep skin infections as described for other African countries (Breurec *et al.*, 2011,

Schaumburg *et al.*, 2014, Sina *et al.*, 2013). In contrast, an association of toxin encoding genes such as the enterotoxin genes and *tst1* was only seen for clonal types such as CC30. This is in agreement with previous investigations showing that *tst1* has a strong linkage with CC30 and that the enterotoxin gene cluster *egc* is often associated with CC5, CC22 and CC45 (Holtfreter *et al.*, 2007, Monecke *et al.*, 2008).

Summarizing, the data showed that DNA MA is an adequate tool for molecular characterization and analysis of the clonal distribution of African SA. The easiness of use of the DNA MA make this method a possible technique for outbreak investigation in labs without access to advanced molecular technologies. Further, the study also showed the need for further trials employing well-controlled, prospectively collected clinical isolates and clinical data to delineate the virulence gene repertoire of Sub-Saharan African SA in conjunction with the clinical disease presentation.

5.3 Genotypic characterization and comparison of Sub-Saharan African and German isolates

SA or rather MSSA are a ubiquitous worldwide distributed, opportunistic pathogen. However, epidemiological and molecular data on SA derived from infections and nasal isolates are rare for developing countries such as in Sub-Saharan Africa. This study provides a comprehensive data set on molecular bacterial characteristics of strictly non-nosocomial SA isolates of volunteers/patients. The isolates were collected under controlled conditions in three Sub-Saharan African and three German study centers with Germany as a representative for industrialized countries of a temperate climate zone. The data of all participants were documented in case related forms (CRFs) (see Appendix, section 6.2) and checked of fulfilling the requirements for inclusion into the study cohort. The criteria for the isolate collection was not part of this work.

The DNA MA based CC and ST distribution analyses showed that the cluster distribution among African and German isolates was inhomogeneous with certain CCs being more prevalent among either African or German patients or volunteers. As there was some heterogeneity in population age and comorbidities between the German and African cohorts this may, however, have distorted the ‘true’ distribution of clones and genes between isolates from the different geographic regions.

In a landmark study from 2010, Grundmann et al. determined the distribution of STs

derived from the *spa*-types of SA isolates that were recovered from patients with invasive disease in 2006/2007 in Europe (Grundmann *et al.*, 2010). Approximately one half of the 565 MSSA isolates was attributed to the following CCs (in descending order): CC45, CC5, CC15, CC7, CC30, CC8, CC22 and CC12. In a strictly population-based survey from Germany, SA was recovered from 85 out of 389 participants (5 of them MRSA). The most prevalent MSSA clones belonged to CC5, CC8, CC1, CC15, and CC30 (Holtfreter *et al.*, 2016, Mehraj *et al.*, 2014). In a population-based study in the Northeast of Germany the most common SA lineage was CC30, followed by CC45, CC15, CC8 and CC22. In line with these observations, the same CCs were also prevalent in our study, with similar distributions of CC5, CC8, and CC45 in the African and German isolate groups. They also belong to the 10 most prevalent CCs in this study.

The population structure of major MSSA clones in Africa has been described in a recent review (Schaumburg *et al.*, 2014) as well as in individual studies (Breurec *et al.*, 2011, Conceicao *et al.*, 2015, Ruimy *et al.*, 2008). They show that CC5, CC15 and CC30 are largely distributed in Africa and Europe. In line with this, these three CCs also belonged to the 10 most prevalent CCs of this study. However, CC15 isolates were more prevalent in the African cohort and CC30 isolates were almost exclusively found in the German group. In comparison, CC8 MSSA were mainly reported from Maghrebian regions, whereas CC8 MRSA were predominantly detected in Central and South Africa and CC121 was more frequently found in Sub-Saharan countries (Schaumburg *et al.*, 2014). MSSA-CC80 has more often been identified in North Africa, while the CA MRSA clone ST80 is prevalent in Europe (Otter & French, 2010, Stegger *et al.*, 2014). In this study all CC80 isolates were of African origin.

According to Schaumburg et al. CC88 isolates are typically methicillin-resistant and isolates of this CC were almost exclusively detected in West, Central and East African regions. Because of this, the cluster was given the acronym ‘African clone’ (Schaumburg *et al.*, 2014). In accordance with Schaumburg and colleagues (Schaumburg *et al.*, 2014), CC88 was only detected at African study sites. However, in contrast to Schaumburg and colleagues in this study CC88 isolates were in most cases MSSA (75%).

CC152, which was almost exclusively found among African isolates in this study, has been considered to be a CA-MRSA emerging from the Balkan region (Monecke *et al.*,

2007; Egyir *et al.* 2014). It is presumed that this strain lineage originates from African MSSA, which first spread through central Europe and then acquired methicillin resistance (Okon *et al.*, 2009, Otter & French, 2010). A study from Mali showed that CC152 was the second most frequent MSSA lineage isolated from healthy carriers and that all tested isolates were PVL positive. Similar results were found in a study of PVL positive MSSA infections of Eritrean and Ethiopian asylum seekers in Switzerland, where 7 out of 15 infections were caused by PVL positive CC512 isolates (Jaton *et al.*, 2016). When summarizing the results of different studies, CC152 can be considered as the major PVL positive CC in Sub-Saharan Africa (Ruimy *et al.*, 2008, Schaumburg *et al.*, 2014), which is also supported by the findings presented in this thesis.

Previous reviews indicate that there are “typical” CC clusters, such as CC5, CC15 and CC30, which are prevalent in Europe as well as “typical” African strain lineages i.e. CC80-MSSA, CC88 or CC152 (Abdulgader *et al.*, 2015, Schaumburg *et al.*, 2014). These observations correspond with the results of this study for the “African” clones CC80-MSSA, CC88 and CC152. In comparison, the typical in Europe prevalent CC5 was almost equally distributed in the African and German isolate groups, while CC15 isolates also reported to constitute a typical European CC and were more often found in the African group at all three study sites.

Clear-cut studies allowing for frequency comparison between European and African clusters are lacking. A case study from travelers returning from Africa demonstrated that SA of African origin may have a different genetic background and/or harbor virulence determinants that are atypical for European isolates (Zanger *et al.*, 2012). This is shown by African resistance data and the unexpected presence of atypical virulence factors such as the Panton-Valentine leukocidin, as could be seen in this study, too (Zanger *et al.*, 2012).

Taken together, in this comparative study, certain CCs were found to be significantly enriched (CC15, CC121, CC152) or even specific (CC88, CC80) for the African cohort. In the German cohort, CCs such as CC45, CC30, CC7 and CC22 were found significantly more often, and CC398 and CC50 were only found at German sites, although the overall numbers of the latter two CCs were comparably low. Additionally, compared to the African study sites, the relative frequencies of CC’s in the clinical and nasal German groups as well as in-between the three German institutions were more homogenous.

Regarding the *agr* subtypes, *agr*-type IV was identified with an over-representation in African and *agr*I in German isolates. In comparison, *agr*-type II (about 25%) and *agr*-type III (14%) were found in similar amounts in the German and African isolates. This is consistent with previous studies demonstrating that *agr*-type IV is the predominant *agr*-type in African CC121 isolates (Kolawole *et al.*, 2013, Oosthuysen *et al.*, 2014). Similarly, capsule types 5 and 8 were found to be prevalent in Africa and Germany to a different extent (Budimir *et al.*, 2010, Lattar *et al.*, 2009, Melles *et al.*, 2008). In African nasal isolates capsule type 5 was predominant, while capsule type 8 was predominant in German nasal samples. A previous study on remote African pygmies showed similar results with respect to the capsule type distribution compared to data from European studies (Schaumburg *et al.*, 2011).

The previously reported differences in carriage rates of the PVL encoding genes were clearly confirmed in this study. Significantly more African SA, almost half of the African isolates (45%) and only single German isolates (3%), carried the PVL encoding genes. Previous African field studies showed that Africa is a PVL-endemic region with high rates of PVL, both in pandemic CA-MRSA- and CA-MSSA-lineages ranging from 17 to 74% (Breurec *et al.*, 2011, Campbell *et al.*, 2008, Okon *et al.*, 2009, Ruimy *et al.*, 2008, Schaumburg *et al.*, 2011, van der Meeren *et al.*, 2014). In clear contrast, in developed regions the PVL prevalence rates in MSSA have been reported to be between 1 to 5% (Lowy, 1998, Monecke *et al.*, 2009, von Eiff *et al.*, 2004).

Remarkably, the MRSA prevalence in our study was very low in comparison (1-3%) to an admission prevalence screening in Germany of the EUREGIO MRSA-net with a MRSA prevalence of 6.5% (n=762 MRSA out of 23566 SA isolates) and a state wide admission prevalence screening study of Saarland with 12% of all collected SA (n=3558) being MRSA (Herrmann *et al.*, 2013b, Köck *et al.*, 2009). Unfortunately, no publications of African SA admission studies were available. The prevalences reported in this thesis probably better reflect the situation seen in the community in Africa and Germany, as this is the only study including isolates from healthy volunteers or from community-associated disease and strictly excluding isolates of hospital-associated infections.

We found a significant enrichment of the tetracycline encoding genes *tet*(K) and *tet*(M) in African isolates. This is in accordance with data of other studies from African countries finding tetracycline resistance of 21.8-92% (Djoudi *et al.*, 2013, Mariem *et*

al., 2013, Phaku *et al.*, 2016). A likely reason for this might be the use of these antibiotics as first line treatment, self-medication, incorrect dosage and unavailable and/or expensive alternative antibiotics as second line treatment because of their higher costs, after first line treatment failure (Fasehun, 1999).

In this study the fosfomycin resistance gene *fosB* was frequently detected in African isolates (62%). In comparison, a recent review compared the antibiotic susceptibility profiles of MRSA in Africa and showed that 84-99% of the MRSA isolates were fosfomycin susceptible (Falagas *et al.*, 2013). These results are the opposite of the findings of the present thesis. The discrepancy might be driven by the fact that in this thesis almost exclusively MSSA isolates (97%) were analyzed. This finding might also indicate that *fosB* is more distributed in African CA-MSSA than in African CA-MRSA isolates. An alternative explanation might be that fosfomycin, as an old antibiotic used for the treatment of uncomplicated urinary tract infections, is frequently used in the community setting in the context of increasing fluoroquinolone resistances (Abubakar *et al*., 2018, Popovic *et al.*, 2010). Unfortunately, no literature of fosfomycin resistance in African MSSA and no data about treatment guidelines, antibiotic stewardship or investigations of typical genetic resistance mechanisms detected in Sub-Saharan African countries were found to confirm the latter hypothesis.

The observed distribution differences of the adhesion genes can be explained by the CC distribution of the isolates. This means, for example, that CC121 and CC152 were mainly found in the African group, while CC50 and CC398 were only found in the German group as well as that single adhesion genes were not recognized in specific CCs (missing *fib* gene recognition in CC22, comprising 37 German and only 4 African isolates and missing *eap* gene recognition in CC152, comprising 79 African and only 2 German isolates). The missing recognition might be caused by DNA mutations, allelic variants or polymorphisms that prevented a hybridization with the DNA MA sequence target albeit of the fact that the respective gene is present (as shown for example for *map/eap*). The lack of *fib* in CC22 influenced the overall results towards larger *fib* positivity among African isolates. Similarly, the higher ratio of *map/eap* positive isolates in the German group is most likely caused by the inability of the CC152 *eap* variant to hybridize with the respective target sequence presented on the DNA MA, as nearly all CC152 isolates were of African origin (80/82). Previous studies indicated *map/eap* to be present in the clear majority of SA clinical isolates (Hussain *et al.*, 2001; Hussain *et*

al., 2008), and that different versions of this gene may exist (Joost *et al*., 2009).

Individual gene detection by DNA MA analysis and their comparison between clinical/nasal and African/German origin of isolates was informative but a more complex analysis was needed to address potential correlations of the hybridization patterns and the isolate origin. PCA gave an overall impression of the hybridization patterns of the 1200 isolates, which could partly be related to isolate characteristics such as geographic origin or clinical significance. PCA allowed the presentation of a combination of all SA specific markers assessed by DNA MA, thus avoiding multiple comparisons of single target recognition. This way, the clustering of the 'predominant African' isolates of CC8, CC15, CC121 and CC152 and the 'German' clusters CC22, CC398 and CC30 were shown. Such an association analysis of isolates with different geographic origin to CCs has not been performed before. A further significant cluster attribution to the 'clinical' versus 'nasal' subgroup was found for three CCs (CC45, CC121 and CC152).

Previous studies already examined the association of CCs and genes with specific clinical presentation (Fowler *et al.*, 2007, Nienaber *et al.*, 2011, Tristan *et al.*, 2012, Nethercott *et al*., 2013). Given the major importance of such an association for clinicians, this was one of the main objectives of this study, too. However, this attempt failed. One reason was that the numbers of specific clinical presentations in the African and German patient group were to small, especially for a comparison of the African and German differences, or similarities regarding e.g. HIV infection. A second reason might be the inequal CC distribution in the different disease type groups to be compared. This failure somehow reflects the literature for this topic which is a kind of contradictive, as some studies report an association of CCs with specific clinical presentation (Fowler *et al.*, 2007, Nienaber *et al.*, 2011, Nethercott *et al*., 2013) while other failed to do so (Feil *et al.*, 2003, Tristan *et al.*, 2012).

In conclusion, the DNA MA analysis of the prospectively collected, community-associated SA isolate set obtained from volunteers and patients demonstrates clear and significant differences between German and Sub-Saharan African localizations with respect to clonal cluster attribution and prevalence of certain virulence- and resistance-related genes.

5.4 Potential of MALDI-TOF as strain differentiation method

The mass spectrometric method MALDI-TOF allows the differentiation of bacteria to the species level in routine microbiology laboratories, fulfilling the criteria of standardization, low turn-around time, convenience in handling, and low costs (Lasch *et al.*, 2014, Dingle & Butler-Wu, 2013, Benagli *et al.*, 2011). However, in outbreak situations the report to the species level do not inform on the association of a given isolate to the outbreak. Molecular subtyping methods such as PFGE, *spa*-typing, MLST, DNA MA and WGS are relatively time-consuming, laborious and very often unavailable for routine diagnostic laboratories as well as in part still connected with high costs. Therefore, MALDI-TOF based subtyping methods could be of paramount clinical value in outbreak situations for fast analysis and calculation of the transmission risk potential of specific strains through the knowledge of the ST or CC they belong to. Earlier work demonstrated a successful subtyping by MALDI-TOF up to the subspecies level for different bacterial species such as *Streptococcus pneumoniae* and *Streptococcus agalactiae* (Williamson *et al.*, 2008, Lartigue *et al.*, 2009), *Salmonella enterica* (Dieckmann *et al.*, 2008), *Listeria monocytogenes* (Barbuddhe *et al.*, 2008), and *Francisella tularensis* (Seibold *et al.*, 2010). One major goal of SA subtyping is not only the identification up to the subspecies level but also the differentiation between MRSA and MSSA strains, for example to prevent MRSA transmission in the clinical setting by prevention strategies such as isolation of MRSA patients. However, the effectiveness of SA subtyping by MALDI-TOF is still a matter of debate. The present literature is dominated by studies with low isolate numbers and differential methods for MRSA identification reporting contradictive results (Du *et al.*, 2002, Edwards-Jones *et al.*, 2000, Josten *et al.*, 2013, Lasch *et al.*, 2014, Szabados *et al.*, 2012, Wang *et al.*, 2013, Wolters *et al.*, 2011). Altogether, a differentiation of SA strains based on their mass spectra in MSSA or MRSA could not be achieved in the majority of studies (Wang *et al.*, 2013, Sparbier *et al.*, 2013, Edwards-Jones *et al.*, 2000, Bernardo *et al.*, 2002). However, two studies reported successful discrimination between MRSA and MSSA mass spectra. First, success was achieved by incorporation of nonradioactive isotopes into newly synthesized proteins during cell growth and functional testing in the presence of antibiotics (oxacillin) (Sparbier *et al.*, 2013) and second based on the selection of MRSA specific peaks in combination with the commercially available software tools of Bruker Daltonik GmbH (Flex Analysis software and MALDI Biotyper

Version 3.02, ClinPro Tools Version 2.2) (Ueda *et al.*, 2015). In this study similarity of mass spectra between two isolates can be directly compared by head-to-head analysis of the individual peak lists. In comparison to this work, in a recent study a support vector machine classification using Flex analysis 3.0 (Bruker Daltonics) was performed leading to identification rates of 93.3% for MSSA and 86.7% for MRSA (Sogawa *et al.*, 2017). This latter approach was similar to the approach in this work analyzing the whole mass spectra and peaks with bioinformatical standardized statistical analysis as support vector machine for MRSA identification by the commercially available software BioNumerics (Applied Maths) and MRSA identification of 89.3% and MSSA of 70.6%. MRSA identification was not 100% successful in both studies and thus not applicable in routine. Maybe further refinements of the methods and databases could lead to applicability in clinical routine.

A second major goal of SA subtyping is the differentiation of strains into known genotypes such as CC/STs or *spa*-types. This objective was investigated in a number of recent studies (Bohme *et al.*, 2012, Camoez *et al.*, 2015, Josten *et al.*, 2013, Lasch *et al.*, 2014, Ostergaard *et al.*, 2015, Wolters *et al.*, 2011, Zhang *et al.*, 2015) and different lineage-related peaks were identified as markers of the established CCs or *spa*-types (Ostergaard *et al.*, 2015, Zhang *et al.*, 2015). These studies demonstrated that MALDI-typing was more or less successful for the discrimination between a couple of typical outbreak strains such as CC1, CC5 and CC22, and more recently also for the LA-MRSA CC398 (Bohme *et al.*, 2012, Josten *et al.*, 2014, Wolters *et al.*, 2011, Ostergaard *et al.*, 2015, Monecke *et al.*, 2010, Sauget *et al.*, 2016). All studies had in common that they looked for strain lineage-specific peaks of CCs and *spa*-types as biomarkers. If identified, these peaks were subsequently used for the identification of the CC or *spa*-type of newly collected isolates. However, most studies worked with limited isolate numbers derived from a single center, which might have caused a certain isolate selection bias. Another drawback of these studies was that different strategies and software tools were used to accomplish their task, which restricts the comparability of these studies.

In this thesis, a high variety of genotypes from six different study sites located in two continents were examined. The primary aim of this study was to determine whether it was possible to establish an automated software tool for MALDI-typing that can be used in the daily diagnostic routine for strain subtyping without bias through manual

evaluation of different users. Automated software (BioNumerics, Applied Maths) was used for statistical analysis and grouping of MALDI-types by PCA for previously defined SA groups [CC, *spa*-type], resistance profile [MRSA, MSSA] or geographic origin [African, German]).

By using the automated universal approach of this study, it was not possible to correctly identify *spa*-types or CCs based on the SA mass spectra, except for CC121. Remarkably, isolates of the latter CC formed a distinct MALDI-type subgroup. Statistical analysis by Mann-Whitney-U test over all spectra identified a significant association of three peaks (*m/z* 2774, *m/z* 5552 and *m/z* 6581) to the CC121 MALDI-TOF spectra. Two of the identified characteristic CC121 peaks with a mass to charge ratio of *m/z* 2774 and *m/z* 5552 had not previously been described. In contrast, the single peak *m/z* 6580/1 was described before as biomarker for CC130 in two studies (Josten *et al.*, 2013, Ostergaard *et al.*, 2015). In the present study, CC130 isolates were not included and the question of similarity between CC121 and CC130 MALDI-TOF spectra remains to be addressed in future studies. These findings let suggest that MALDI-TOF analysis may be a potential method for specific CCs in local outbreaks. However, it has to be considered that not only single peaks but rather a combination of peak patterns should be regarded for a correct CC identification or CC differentiation. The reason is that till yet the available data are limited and thus a high risk of misidentification is given.

5.5 Conclusion

The data shown in this thesis demonstrate that DNA MA is an adequate tool for SA characterization of German as well as of African SA. It allows the detection of antibiotic resistance cassettes, virulence genes and molecular epidemiological analysis of SA isolates. Major advantages of the DNA MA application are the fast-available results (within 6 h), the objective result interpretation and the easy to understand results (automated read out of signals). In addition, the comparison of *spa*-typing and DNA MA showed a higher subtyping potential of the DNA MA especially for closely related strains as SA CC5 isolates.

However, a disadvantage is that DNA MA can only detect what is covered by the sequence targets on the DNA MA, which can result in false negative results. Furthermore, large datasets require the use of bioinformatical tools. Notably, the

analysis of the DNA MA results revealed different isolate subgroups by the different bioinformatic methods. This is a critical point. The selection of the applied method and its mathematical algorithm has to be taken into account in the result interpretation. Thus, in the future a fast, detailed and standardized strain comparison with bioinformatic analysis tools is required for DNA MA as well as WGS data. Especially with regard to the growing amount of molecular and clinical data the development of fast, easy to handle and easy to understand software tools are necessary for data evaluation and interpretation.

Testing a comprehensive set of SA isolates with MALDI-TOF revealed that it was not possible to determine the methicillin resistance profile, geographic origin or the strain lineage of SA isolates based on their mass spectra profiles with one exception. Only CC121 isolates could be correctly identified. Together with findings of previous studies this show that general identification of methicillin resistance and strain lineages based on mass spectra is not possible, however by applying homemade rules it can be used to identify specific strain types and even particular CC types, which may be an advantage for routine laboratories in situations with local outbreaks.

In the future, fast and reliable SA characterization data may help to close the white spots of missing SA data of low- and middle-income countries, may allow a better traceability of SA transmission and a better understanding of the success of specific SA strains. DNA MA and/or WGS may become the leading molecular techniques for isolate differentiation and characterization in diagnostics with regard to dropping costs of WGS and the development of automatic data analysis systems. However, to be really effective, a compilation of DNA MA/WGS data in a large database will be needed as is already available for MLST and *spa*-typing data.

6 Appendix

6.1 Manufacturer information of Alere Technologies to the DNA microarray targets

Table 21: List of all DNA microarray target groups with written out of the manufacturer information of Alere Technologies.

Gene group	Gene abbreviation	Gene encoding for
SPECIES MARKER	*rrnD1* (*S. aureus*)	Domain 1 of 23S-rRNA
	gapA	glyceraldehyde 3-phosphate dehydrogenase, locus 1
	katA	katalase A
	coA	Coagulase
	nuc1	thermostable extracellular nuclease 1
	spa	staphylococcal protein A
	sbi	IgG-binding protein Sbi
REGULA-TORY GENES	*sarA*	staphylococcal accessory regulator A
	saeS	histidine protein kinase, sae locus
	vraS	sensor protein
	*agr*I (total)	accessory gene regulator allele I
	agrB-I	
	agrC-I	
	agrD-I	
	*agr*II (total)	accessory gene regulator allele II
	agrB-II	
	agrC-II	
	agrD-II	
	*agr*III (total)	accessory gene regulator allele III
	agrB-III	
	agrC-III	
	agrD-III	
	*agr*IV (total)	accessory gene regulator allele IV
	agrB-IV	
	agrC-IV	
	hld	haemolysin delta
METHICILLIN RESISTANCE GENOTYPE AND SCCmec TYPING	*mecA*	alternate penicillin binding protein 2, defining MRSA
	delta_mecR	truncated signal transducer protein MecR1
	ugpQ	glycerophosphoryl diester phosphodiesterase, associated with mecA
	ccrA-1	cassette chromosome recombinase genes A-1
	ccrB-1	cassette chromosome recombinase genes B-1
	plsSCC (COL)	plasmin-sensitive surface protein
	Q9XB68-dcs	hypothetical protein from SCCmec elements
	ccrA-2	cassette chromosome recombinase gene A-2
	ccrB-2	cassette chromosome recombinase gene B-2
	kdpA-SCC	potassium-translocating ATPase A, chain 2

METHICILLIN RESISTANCE GENOTYPE AND SCCmec TYPING	*kdpB-SCC*	potassium-transporting ATPase B, chain 1
	kdpC-SCC	potassium-translocating ATPase C, chain 2
	kdpD-SCC	sensor kinase protein
	kdpE-SCC	KDP operon transcriptional regulatory protein
	mecI	methicillin-resistance gene regulatory protein
	mecR	*mecA* repressor
	xylR	homolog of xylose repressor, associated with SCCmec-elements
	ccrA-3	cassette chromosome recombinase gene A-3
	ccrB-3	cassette chromosome recombinase gene B-3
	merA	mercury resistance gene operon, Hg(II) reductase
	merB	mercury resistance gene operon, alkylmercury lyase
	ccrAA (MRSAZH47)_probe 1	cassette chromosome recombinase gene (hypothetical)
	ccrAA (MRSAZH47)_probe 2	cassette chromosome recombinase gene (hypothetical)
	ccrC (85-2082)	cassette chromosome recombinase gene C
	ccrA-4	cassette chromosome recombinase gene A-4
	ccrB-4	cassette chromosome recombinase gene B-4
RESISTANCE GENOTYPE: PENICILLINASE	*blaZ*	beta-lactamase gene
	blaI	beta-lactamase repressor (inhibitor)
	blaR	beta-lactamase regulatory protein
RESISTANCE GENOTYPE: MLS-ANTIBIOTICS	*erm*(A)	rRNA methyltransferase associated with macrolide/lincosamide resistance
	erm(B)	rRNA methyltransferase associated with macrolide/lincosamide resistance
	erm(C)	rRNA methyltransferase associated with macrolide/lincosamide resistance
	lnu(A)	lincosaminide nucleotidyltransferase (=linA)
	msr(A)	macrolide efflux pump
	mef(A)	macrolide efflux protein A
	mph(C)	macrolide phosphotransferase II (=mpbBM)
	vat(A)	virginiamycin A acetyltransferase
	vat(B)	acetyltransferase inactivating streptogramin A
	vga(A)	ABC transporter conferring resistance to streptogramin A and related compounds
	vga(A) (BM 3327)	vga(A) allele from strain BM 3327
	vgB(A)	virginiamycin B hydrolase (=vgb)
RESISTANCE GENOTYPE: AMINOGLY-COSIDES	*aacA-aphD*	aminoglycoside adenyl-/phosphotransferase (gentamicin, tobramycin)
	aadD	aminoglycoside adenyltransferase (neo-/ kanamycin, tobramycin)
	aphA3	aminoglycoside phosphotransferase (neo-/ kanamycin)

RESISTANCE GENOTYPE : MISCELLANEOUS GENES	*sat*	streptothricin acetyltransferase
	dfrS1	dihydrofolate reductase mediating trimethoprim resistance (=dfrA)
	far1	fusidic acid resistance gene (= far1)
	fusC (*Q6GD50*)	fusidic acid resistance gene (= Q6GD50)
	mupA	isoleucyl-tRNA synthetase associated with mupirocin resistance (=mupR)
	tet(K)	tetracycline efflux protein
	tet(M)	ribosomal protection protein associated with tetracycline resistance
	cat (total)	chloramphenicol acetyltransferase
	cat (pC221)	
	cat (pc223)	
	cat (pMC524/pC194)	
	cat (pSBK203R)	
	cfr	23S rRNA methyltransferase (phenicols, lincosamides, oxazolidinones, pleuromutilins, streptogramin A)
	fexA	chloramphenicol/florfenicol exporter
	fosB	metallothiol transferase
	fosB (plasmid)	
RESISTANCE GENOTYPE: EFFLUX SYSTEMS	*qacA*	quaternary ammonium compound / multidrug efflux protein C
	qacC (total)	quaternary ammonium compound / multidrug efflux protein A
	qacC (consensus)	
	qacC (equine)	
	qacC (SA5)	
	qacC (Ssap)	
	qacC (ST94)	
	sdrM	putative transport protein (=tetEfflux)
RESISTANCE GENOTYPE: GLYCOPEPTIDES	*vanA*	vancomycin resistance gene
	vanB	vancomycin resistance gene from enterococci and Clostridium
	vanZ	teicoplanin resistance gene from enterococci
VIRULENCE: TOXIC SHOCK TOXIN	*tst1* (consensus)	toxic shock syndrome toxin 1
	tst1 ("human" allele)	
	tst1 ("bovine" allele)	

VIRULENCE: ENTERO-TOXINS	*sea*	enterotoxin A (=entA)
	sea (320E)	enterotoxin A, allele from strain 320E
	sea (N315) / *sep*	enterotoxin A, allele from strain N315 = enterotoxin P
	seb	enterotoxin B (=entB)
	sec	enterotoxin C (=entC)
	sed	enterotoxin D (=entD)
	see	enterotoxin E (=entE)
	seg	enterotoxin G (=entG)
	seh	enterotoxin H (=entH)
	sei	enterotoxin I (=entI)
	sej	enterotoxin J (=entJ)
	sek	enterotoxin K (=entK)
	sel	enterotoxin L (=entL)
	selm	enterotoxin-like gene/protein M (=sem, entM)
	seln (consensus)	enterotoxin-like gene/protein N (=sen, entN), consensus probe
	seln (other than RF122)	
	selo	enterotoxin-like gene/protein O (=seo, entO)
	egc (total)	enterotoxin gene cluster, consisting of seg, sei, selm, seln, selo, selu
	seq	enterotoxin Q (=entQ)
	ser	enterotoxin R (=entR)
	selu	enterotoxin-like gene/protein U (=seu, entU)
	ORF CM14_ probe1	enterotoxin-like protein ORF CM14
	ORF CM14_ probe2	enterotoxin-like protein ORF CM14
VIRULENCE: HLG AND LEUKOCI-DINS	*lukF*	haemolysin gamma / leukocidin, component B (F)
	lukS	haemolysin gamma / leukocidin, component C (S)
	lukS (ST22+ST45)	haemolysin gamma / leukocidin, component C (S), allele from ST22 and ST45
	hlgA	haemolysin gamma, component A
	lukF-PV	Panton-Valentine leukocidin F component
	lukS-PV	Panton-Valentine leukocidin S component
	lukF-PV (P83)	F component of leukocidin from ruminants
	lukM	S component of leukocidin from ruminants
	lukD	leukocidin D component
	lukE	leukocidin E component
	lukX	leukocidin/haemolysin toxin family protein (= lukB or lukG)
	lukY	leukocidin/haemolysin toxin family protein (= lukA or lukH)
	lukY (ST30+ST45)	leukocidin/haemolysin toxin family protein (= lukA or lukH), allele from ST30 and ST45

VIRULENCE: HAEMO-LYSINS	*hl*	putative membrane protein
	hla	haemolysin alpha
	hl /// (consensus)	putative membrane protein
	hl /// (other than RF122)	
	*hlb*_probe 1	haemolysin beta
	*hlb*_probe 2	haemolysin beta
	*hlb*_probe 3	haemolysin beta
	un-disrupted hlb	haemolysin beta without phage insertion
VIRULENCE: HLB-CONVERTING PHAGES	*sak*	Staphylokinase
	chp	chemotaxis-inhibiting protein (CHIPS)
	scn	staphylococcal complement inhibitor
VIRULENCE: EXFOLIATIVE TOXINS	*etA*	exfoliative toxin serotype A
	etB	exfoliative toxin serotype B
	etD	exfoliative toxin D
VIRULENCE: EPIDERMAL CELL DIFFERENTION INHIBITORS	*edinA*	epidermal cell differentiation inhibitor A
	edinB	epidermal cell differentiation inhibitor B
	edinC	epidermal cell differentiation inhibitor C
VIRULENCE: ACME LOCUS	ACME (total)	Arginine Catabolic Mobile Element
	arcA-SCC	ACME-locus: arginine deiminase
	arcB-SCC	ACME-locus: ornithincarbamoyltransferase
	arcC-SCC	ACME-locus: carbamatkinase
	arcD-SCC	ACME-locus: arginine/ornithine-antiporter
VIRULENCE: PROTEASES	*aur* (consensus)	Aureolysin
	aur (other than MRSA252)	
	aur (MRSA252)	
	splA	serinprotease A
	splB	serinprotease B
	splE	serinprotease E
	sspA	glutamylendopeptidase
	sspB	staphopain B, protease
	sspP (consensus)	staphopain A (staphylopain A), protease
	sspP (other than ST93)	
VIRULENCE: STAPHYLO-COCCAL SUPERANTI-GEN/ ENTEROTO-XIN-LIKE GENES (SET/SSL)	*setC*	staphylococcal exotoxin-like protein/SAg gene homolog, SAUSA300_0370
	*ssl01/set6*_probe1_11	staphylococcal superantigen-like protein 1 (probes)
	*ssl01/set6*_probe2_11	
	*ssl01/set6*_probe1_12	
	*ssl01/set6*_probe2_12	
	*ssl01/set6*_probe4_11	
	*ssl01/set6*_ probeRF122	
	ssl01/set6 (COL)	staphylococcal superantigen-like protein 1 (interpretation/alleles)
	ssl01/set6 (Mu50+N315)	

VIRULENCE: STAPHYLO-COCCAL SUPERANTI-GEN/ ENTEROTO-XIN-LIKE GENES (SET/SSL)	*ssl01/set6* (MW2+MSSA476)	
	ssl01/set6 (MRSA252)	
	ssl01/set6 (RF122)	
	ssl01/set6 (other alleles)	
	ssl02/set7	staphylococcal superantigen-like protein 2
	ssl02/set7 (MRSA252)	
	*ssl03/set8*_probe 1	staphylococcal superantigen-like protein 3
	*ssl03/set8*_probe 2	
	ssl03/set8 (MRSA252, SAR0424)	
	ssl04/set9	staphylococcal superantigen-like protein 4
	ssl04/set9 (MRSA252, SAR0425)	
	*ssl05/set3*_probe 1	staphylococcal superantigen-like protein 5
	ssl05/set3 (RF122, probe-611)	
	*ssl05/set3*_probe 2 (612)	
	ssl05/set3 (MRSA252)	
	ssl06/set21	staphylococcal superantigen-like protein 6
	ssl06 (NCTC8325+MW2)	
	ssl07/set1	staphylococcal superantigen-like protein 7
	ssl07/set1 (MRSA252)	
	ssl07/set1 (AF188836)	
	*ssl08/set12*_probe 1	staphylococcal superantigen-like protein 8
	*ssl08/set12*_probe 2	
	*ssl09/set5*_probe 1	staphylococcal superantigen-like protein 9
	*ssl09/set5*_probe 2	
	ssl09/set5 (MRSA252)	
	ssl10/set4	staphylococcal superantigen-like protein 10
	ssl10 (RF122)	
	ssl10/set4 (MRSA252)	
	ssl11/set2 (COL)	staphylococcal superantigene-like protein 11
	ssl11+set2(Mu50+N315)	
	ssl11+set2(MW2+ MSSA476)	
	ssl11/set2 (MRSA252)	
	setB3	staphylococcal exotoxin-like protein, second locus
	setB3 (MRSA252)	
	setB2	
	setB2 (MRSA252)	
	setB1	

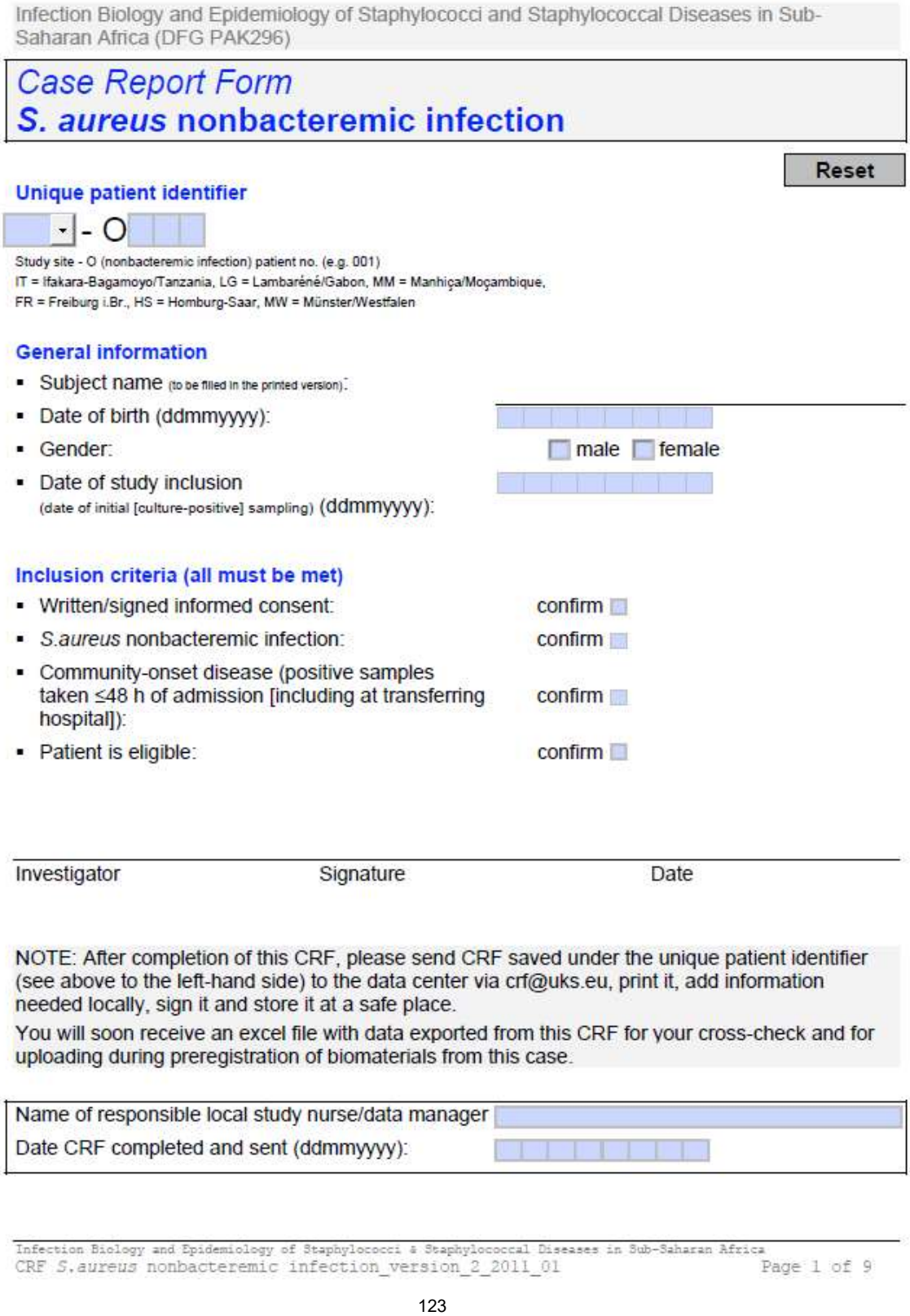
CAPSULE- AND BIOFILM-ASSOCIATED GENES	*cap 1* (total)	Capsule type 1
	capH1	capsular polysaccharide synthesis enzyme
	capJ1	O-antigen polymerase
	capK1	capsular polysaccharide biosynthesis protein
	cap 5 (total)	Capsule type 5
	capH5	capsular polysaccharide synthesis enzyme
	capJ5	O-antigen polymerase
	capK5	capsular polysaccharide biosynthesis protein
	cap 8 (total)	Capsule type 8
	capH8	capsular polysaccharide synthesis enzyme
	capI8	capsular polysaccharide biosynthesis protein
	capJ8	O-antigen polymerase
	capK8	capsular polysaccharide biosynthesis protein
	icaA	intercellular adhesion protein A
	icaC	intercellular adhesion protein C
	icaD	biofilm PIA synthesis protein D
	bap	surface protein involved in biofilm formation
ADHAESION FACTORS / GENES ENCODING MICROBIAL SURFACE COMPO-NENTS RECOGNI-ZING ADHESIVE MATRIX MOLECULES (MSCRAMM GENES)	*bbp* (total)	bone sialoprotein-binding protein
	bbp (consensus)	
	bbp (COL+MW2)	
	bbp (MRSA252)	
	bbp (Mu50)	
	bbp (RF122)	
	bbp (ST45)	
	clfA (total)	clumping factor A
	clfA (consensus)	
	clfA (COL+RF122)	
	clfA (MRSA252)	
	clfA (Mu50+MW2)	
	clfB (total)	clumping factor B
	clfB (consensus)	
	clfB (COL+Mu50)	
	clfB (MW2)	
	clfB (RF122)	
	cna	collagen-binding adhesin
	ebh (consensus)	cell wall associated fibronectin-binding protein
	ebpS (total)	cell surface elastin binding protein
	*ebpS*_probe 612	
	*ebpS*_probe 614	
	ebpS (01-1111)	
	ebpS (COL)	
	eno	Enolase

ADHAESION FACTORS / GENES ENCODING MICROBIAL SURFACE COMPO-NENTS RECOGNI-ZING ADHESIVE MATRIX MOLECULES (MSCRAMM GENES)	*fib*	fibrinogen binding protein (19kDa)
	fib (MRSA252)	
	fnbA (total)	fibronectin-binding protein A
	fnbA (consensus)	
	fnbA (COL)	
	fnbA (MRSA252)	
	fnbA (Mu50+MW2)	
	fnbA (RF122)	
	fnbB (total)	fibronectin-binding protein B
	fnbB (COL)	
	fnbB (COL+Mu50+MW2)	
	fnbB (Mu50)	
	fnbB (MW2)	
	fnbB (ST15)	
	fnbB (ST45-2)	
	map (total)	major histocompatibility complex class II analog protein (=Extracellular adherence protein, eap)
	map (COL)	
	map (MRSA252)	
	map (Mu50+MW2)	
	sasG (total)	Staphylococcus aureus surface protein G
	sasG (COL+Mu50)	
	sasG (MW2)	
	sasG (other than MRSA252+RF122)	
	sdrC (total)	Ser-Asp rich fibrinogen-/bone sialoprotein-binding protein C
	sdrC (consensus)	
	sdrC (B1)	
	sdrC (COL)	
	sdrC (Mu50)	
	sdrC (MW2+ MRSA252+RF122)	
	sdrC (other than MRSA252+RF122)	
	sdrD (total)	Ser-Asp rich fibrinogen-/bone sialoprotein-binding protein D
	sdrD (consensus)	
	sdrD (COL+MW2)	
	sdrD (Mu50)	
	sdrD (other)	
	vwb (total)	von Willebrand factor binding protein
	vwb (consensus)	
	vwb (COL+MW2)	
	vwb (MRSA252)	
	vwb (Mu50)	
	vwb (RF122)	

IMMUNODO-MINANT ANTIGEN B	*isaB*	immunodominant antigen B
	isaB (MRSA252)	
DEFENSIN RESISTANCE GENOTYPE	*mprF* (COL+MW2)	defensin resistance gene protein
	mprF (Mu50+MRSA252)	
TRANSFERRIN BINDING PROTEIN	*isdA* (consensus)	transferrin-binding protein
	isdA (MRSA252)	
	isdA (other than MRSA252)	
PUTATIVE TRANS-PORTER	*lmrP* (other than RF122)_probe1	hypothetical protein, similar to integral membrane protein LmrP
	lmrP (other than RF122)_probe2	
	lmrP (RF122)_probe1	
	lmrP (RF122)_probe2	
TYPE I RESTRIC-TION-MODIFICATION SYSTEM, SINGLE SEQUENCE SPECIFICITY PROTEIN	*hsdS1* (RF122)	type I site-specific deoxyribonuclease subunit, 1st locus
	hsdS2 (Mu50+N315+COL+USA300 +NCTC8325)	type I site-specific deoxyribonuclease subunit, 2nd locus
	hsdS2 (MW2+MSSA476)	
	hsdS2 (RF122)	
	hsdS2 (MRSA252)	
	hsdS3 (all other than RF122+ MRSA252)	type I site-specific deoxyribonuclease subunit, 3rd locus
	hsdS3 (COL+USA300+NCTC8325 +MW2+MSSA476+RF122)	
	hsdS3 (Mu50+N315)	
	hsdS3 (CC51+ MRSA252)	
	hsdS3 (MRSA252)	
	hsdSx (CC25)	type I site-specific deoxyribonuclease subunit, unknown locus
	hsdSx (CC15)	
	hsdSx (*etd*)	
MISCELLA-NEOUS GENES	*Q2FXC0*	hypothetical protein, located next to serine protease operon
	Q2YUB3	unspecific efflux/transporter
	Q7A4X2	hypothetical protein
HYALURO-NATE LYASE	*hysA1* (MRSA252)	hyaluronate lyase, first / second locus
	hysA1 (MRSA252+RF122) and/or *hysA2* (consensus)	
	hysA1 (MRSA252+RF122) and/or *hysA2* (COL+USA300)	
	hysA2 (all other than MRSA252)	hyaluronate lyase, second locus
	hysA2 (COL+USA300+NCTC8325)	
	hysA2 (all other than COL+USA300+ NCTC8325)	
	hysA2 (all other than COL+USA300+ NCTC8325),1	
	hysA2 (MRSA252)	

6.2 Case report forms

Figure 28: **Case report form of *Staphylococcus aureus* (SA) isolates of nonbacteremic infection.** The case report form defined the isolate number of the collected SA isolate of a nonbacteremic infection, collected general patient information, inclusion criteria to exclude SA isolates of hospital acquired infections, clinical patient data and microbiologic data of the SA isolates.



Infection Biology and Epidemiology of Staphylococci and Staphylococcal Diseases in Sub-Saharan Africa (DFG PAK296)

Case Report Form
***S. aureus* nonbacteremic infection**

Reset

Unique patient identifier

- O

Study site - O (nonbacteremic infection) patient no. (e.g. 001)
IT = Ifakara-Bagamoyo/Tanzania, LG = Lambaréné/Gabon, MM = Manhiça/Moçambique,
FR = Freiburg i.Br., HS = Homburg-Saar, MW = Münster/Westfalen

General information

- Subject name (to be filled in the printed version):
- Date of birth (ddmmyyyy):
- Gender: male female
- Date of study inclusion (date of initial [culture-positive] sampling) (ddmmyyyy):

Inclusion criteria (all must be met)

- Written/signed informed consent: confirm
- *S.aureus* nonbacteremic infection: confirm
- Community-onset disease (positive samples taken ≤48 h of admission [including at transferring hospital]): confirm
- Patient is eligible: confirm

Investigator Signature Date

NOTE: After completion of this CRF, please send CRF saved under the unique patient identifier (see above to the left-hand side) to the data center via crf@uks.eu, print it, add information needed locally, sign it and store it at a safe place.
You will soon receive an excel file with data exported from this CRF for your cross-check and for uploading during preregistration of biomaterials from this case.

Name of responsible local study nurse/data manager
Date CRF completed and sent (ddmmyyyy):

Infection Biology and Epidemiology of Staphylococci & Staphylococcal Diseases in Sub-Saharan Africa
CRF *S.aureus* nonbacteremic infection_version_2_2011_01 Page 1 of 9

Infection Biology and Epidemiology of Staphylococci and Staphylococcal Diseases in Sub-Saharan Africa (DFG PAK296)

Case Report Form
S. aureus nonbacteremic infection

Demographic data

- Place of birth (country):
- Ethnicity:
- Place of residence
  - name of city or village/county (or region)
  - country
  - postal code (if available):
  - GPS coordinates:
    (to be filled in the printed version if possible; for later reporting during biomaterials preregistration)

Patient history

- Previous hospitalization during the last 6 months prior to the present admission: ☐ yes ☐ no ☐ unknown
- Close health care contact (e.g. parenteral therapy, wound care, hemodialysis) within 30 days prior to SAB: ☐ yes ☐ no ☐ unknown
- Continuous residency in a nursing home before nonbacteremic infection: ☐ yes ☐ no ☐ unknown
- History of tuberculosis in the last 6 months: ☐ yes ☐ no ☐ unknown
- Antituberculous drugs within the last 4 weeks: ☐ yes ☐ no ☐ unknown
  - if, yes, was rifampicin including the regimen: ☐ yes ☐ no ☐ unknown
- Antibiotics within the last 4 weeks: ☐ yes ☐ no ☐ unknown

Clinical presentation

- Presentation with or development of severe sepsis 48 h following admission/presentation: ☐ yes ☐ no ☐ unknown
- Presentation with or development of septic shock 48 h following admission/presentation: ☐ yes ☐ no ☐ unknown
- Clinical site of infection:
  - superficial [skin & skin structure] ☐
  - other (deep) sites to be specified: ☐
    - bone ☐
    - joint ☐
    - deep skin abscess ☐

Case Report Form
S. aureus nonbacteremic infection

- muscle: thigh ☐
- muscle: regions other than thigh ☐
- fasciitis ☐
- respiratory tract/lungs incl. pleura ☐
- heart/heart valve ☐
- CSF/brain ☐
- urinary tract ☐
- other, please specify

Day 14 outcome variables

- Was the patient admitted to hospital care within 3 days after disease onset because of staphylococcal disease: ☐ yes ☐ no ☐ unknown
  - if yes, date of discharge (dd.mm.yyyy) ☐ unknown
- Did the patient die within 14 days after disease onset: ☐ yes ☐ no ☐ unknown
  - if yes, date of death (ddmmyyyy):
  - specify (probable) cause of death:
  - if not: last date known to be alive (ddmmyyyy)

Risk factors/comorbidities/underlying disease

- Intravenous drug abuse (IVDA): ☐ yes ☐ no ☐ unknown
- Intravascular catheter(s) present prior to disease onset: ☐ yes ☐ no ☐ unknown
- Other intravascular foreign bodies (i.e.artificial heart valves): ☐ yes ☐ no ☐ unknown
  - if yes, please specify
- Other foreign bodies (i.e. arthroplasty): ☐ yes ☐ no ☐ unknown
  - if yes, please specify

Infection Biology and Epidemiology of Staphylococci and Staphylococcal Diseases in Sub-Saharan Africa (DFG PAK296)

Case Report Form
S. aureus nonbacteremic infection

- HIV infection: ☐yes ☐no ☐unknown
- Charlson comorbidity score items
  - Myocardial infarction/coronary heart disease ☐yes ☐no ☐unknown
  - Congestive heart failure ☐yes ☐no ☐unknown
  - Peripheral vascular disease ☐yes ☐no ☐unknown
  - Cerebrovascular disease ☐yes ☐no ☐unknown
  - Dementia ☐yes ☐no ☐unknown
  - Chronic obstructive pulmonary disease ☐yes ☐no ☐unknown
  - Connective tissue disease ☐yes ☐no ☐unknown
  - Peptic ulcer disease ☐yes ☐no ☐unknown
  - Mild liver disease ☐yes ☐no ☐unknown
    OR moderate-severe liver disease ☐yes ☐no ☐unknown
  - Diabetes mellitus ☐yes ☐no ☐unknown
    OR diabetes mellitus with organ damage ☐yes ☐no ☐unknown
  - Hemiplegia ☐yes ☐no ☐unknown
  - Moderate-severe renal disease ☐yes ☐no ☐unknown
  - Any tumour (within last 5 years) ☐yes ☐no ☐unknown
    Lymphoma ☐yes ☐no ☐unknown
    Leukemia ☐yes ☐no ☐unknown
    Metastatic solid tumor ☐yes ☐no ☐unknown
  - AIDS ☐yes ☐no ☐unknown
- McCabe-Jackson underlying disease prognosis score items
  - Does the patient have an underlying acute, fulminant disease expected to be rapidly (weeks/months) fatal ☐yes ☐no ☐unknown
  - Does the patient have a chronic, incurable condition that is expected to be fatal within the next few years? ☐yes ☐no ☐unknown
  - Does the patient have a chronic condition not expected to be fatal within the next 4 years? ☐yes ☐no ☐unknown

Infection Biology and Epidemiology of Staphylococci and Staphylococcal Diseases in Sub-Saharan Africa (DFG PAK296)

Case Report Form
***S. aureus* nonbacteremic infection**

- Current parasitic diseases (results within 7 days prior or subsequent to initial positive *S. aureus* sampling)
  - Positive blood smear for *Plasmodium* sp:
    if positive: date of sampling (ddmmyyyy)
    number of trophozoites [/µL]
    or negative/not done/unknown ☐
  - Positive blood smear for microfilariae:
    if positive: date of sampling (ddmmyyyy)
    species of *Filaria*
    or negative/not done/unknown ☐
  - Positive stool sampling for intestinal helminths / ova:
    if positive: date of sampling (ddmmyyyy)
    species species 1
    species 2
    species 3
    or negative/not done/unknown ☐
  - Positive urine microscopy for *S. haematobium*:
    if positive: date of sampling (ddmmyyyy)
    or negative/not done/unknown ☐

Infection Biology and Epidemiology of Staphylococci and Staphylococcal Diseases in Sub-Saharan Africa (DFG PAK296)

Case Report Form
S. aureus nonbacteremic infection

Microbiology (initial isolate)

- Antibiogram
  - penicillin ☐ S ☐ R ☐ unknown
  - cefoxitin ☐ S ☐ R ☐ unknown
  - tetracycline ☐ S ☐ I ☐ R ☐ unknown
  - erythromycin ☐ S ☐ I ☐ R ☐ unknown
  - clindamycin ☐ S ☐ I ☐ R ☐ unknown
  - gentamicin ☐ S ☐ I ☐ R ☐ unknown
  - chloramphenicol ☐ S ☐ I ☐ R ☐ unknown
  - trimethoprim-sulfamethoxazole ☐ S ☐ I ☐ R ☐ unknown
- E-test results (MIC in µg/mL)
  - Cefoxitin ☐☐☐.☐☐☐
  - Clindamycin ☐☐☐.☐☐☐
  - Linezolid ☐☐☐.☐☐☐
  - Vancomycin ☐☐☐.☐☐☐
  - Daptomycin ☐☐☐.☐☐☐
  - Tigecycline ☐☐☐.☐☐☐
- Other tests
  - MRSA detection ☐ yes ☐ no ☐ unknown

    if yes, describe method/result
  - Induced clindamycin resistance ☐ yes ☐ no ☐ unknown
- Were there multiple samples from identical site(s) positive for *S.aureus*: ☐ yes ☐ no ☐ unknown
- Were there multiple samples from different sites positive for *S.aureus*: ☐ yes ☐ no ☐ unknown

  if yes, please specify the sites

NOTE: Please store and send up to three isolates (if available) per case including the initial isolate (mandatory) to the study center

Infection Biology and Epidemiology of Staphylococci and Staphylococcal Diseases in Sub-Saharan Africa (DFG PAK296)

Case Report Form
S. aureus nonbacteremic infection

Antibiotic treatment against staphylococci disease (day 0 = day of initial [culture-positive] sampling)

- Antibiotic(s) given on day 01:

name: application:
daily dose: , g
name: application:
daily dose: , g

- Antibiotic(s) given on day 02:

name: application:
daily dose: , g
name: application:
daily dose: , g

- Antibiotic(s) given on day 03:

name: application:
daily dose: , g
name: application:
daily dose: , g

- Did the patient receive more than 3 full days of antibacterial therapy for staphylococcal disease? ☐ yes ☐ no ☐ unknown

Surgical treatment

- Incision/drainage within 3 days after disease onset: ☐ yes ☐ no ☐ unknown
- Removal of foreign body (if any) within 3 days after disease onset: ☐ yes ☐ no ☐ unknown/not applicable
- Other surgical procedures within 3 days after disease

  onset because of staphylococcal disease: ☐ yes ☐ no ☐ unknown

  if yes, type of surgery

  and date of surgery (ddmmyyyy)

Infection Biology and Epidemiology of Staphylococci and Staphylococcal Diseases in Sub-Saharan Africa (DFG PAK296)

Case Report Form
S. aureus nonbacteremic infection

Sampling *S. aureus* isolates

Initial isolate

- Date of sampling (ddmmyyyy):
- Site of sampling:
- Local laboratory identifier/number *S. aureus* isolate:
- Unique *S. aureus* study isolate label: -O -I1 Study site/subject no.

Second isolate

- Not done
- Date of sampling (ddmmyyyy):
- Site of sampling:
- Local laboratory identifier/number *S. aureus* isolate:
- Unique *S. aureus* study isolate label: -O -I2 Study site/subject no.

Third isolate

- Not done
- Date of sampling (ddmmyyyy):
- Site of sampling:
- Local laboratory identifier/number *S. aureus* isolate:
- Unique *S. aureus* study isolate label: -O -I3 Study site/subject no.

NOTE: Please send appropriately labelled materials as soon as possible to central storage facility. Note that you need to preregister at the StaphNet-Portal all specimens to be shipped; preregistration includes uploading the CRF data in an exported excel file as provided by the data center.

Name of responsible lab technician:

Do you confirm that the above data on sampling and labelling are correct:

Infection Biology and Epidemiology of Staphylococci and Staphylococcal Diseases in Sub-Saharan Africa (DFG PAK296)

Case Report Form
***S. aureus* nonbacteremic infection**

Plasma/serum sampling

First serum / plasma

- Not done ☐
- Date of serum / plasma sampling (ddmmyyyy): ☐☐☐☐☐☐☐☐
- Local laboratory identifier/number serum/plasma sample: ☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐
- Unique serum label: ☐☐ -O ☐☐☐ -S1 Study site/subject no.
- Unique plasma label: ☐☐ -O ☐☐☐ -P1 Study site/subject no.

Second serum / plasma

- Not done ☐
- Date of serum / plasma sampling (ddmmyyyy): ☐☐☐☐☐☐☐☐
- Local laboratory identifier/number serum/plasma sample: ☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐
- Unique serum label: ☐☐ -O ☐☐☐ -S2 Study site/subject no.
- Unique plasma label: ☐☐ -O ☐☐☐ -P2 Study site/subject no.

Third serum / plasma

- Not done ☐
- Date of sampling (ddmmyyyy): ☐☐☐☐☐☐☐☐
- Local laboratory identifier/number serum/plasma sample: ☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐
- Unique serum label: ☐☐ -O ☐☐☐ -S3 Study site/subject no.
- Unique plasma label: ☐☐ -O ☐☐☐ -P3 Study site/subject no.

NOTE: Please send appropriately labelled materials as soon as possible to central storage facility. Note that you need to preregister at the StaphNet-Portal all specimens to be shipped; preregistration includes uploading the CRF data in an exported excel file as provided by the data center.

Name of responsible lab technician: ☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐

Do you confirm that the above data on sampling and labelling are correct: ☐

Figure 29: **Case report form of *Staphylococcus aureus* (SA) isolates bloodstream infection.** The case report form defined the isolate number of the collected SA isolate of a bloodstream infection, collected general patient information, inclusion criteria to exclude SA isolates of hospital acquired infections, clinical patient data and microbiologic data of the SA isolates.

Infection Biology and Epidemiology of Staphylococci and Staphylococcal Diseases in Sub-Saharan Africa (DFG PAK296)

Case Report Form
S. aureus bloodstream infection

Reset

Unique patient identifier

[ ] - B [ ]

Study site - B (bloodstream infection) patient no. (e.g. 001)
IT = Ifakara-Bagamoyo/Tanzania, LG = Lambaréné/Gabon, MM = Manhiça/Moçambique,
FR = Freiburg i.Br., HS = Homburg-Saar, MW = Münster/Westfalen

General information

- Subject name (to be filled in the printed version):
- Date of birth (ddmmyyyy): [ ]
- Gender: ☐ male ☐ female
- Date of study inclusion (date of initial [culture-positive] sampling) (ddmmyyyy): [ ]

Inclusion criteria (all must be met)

- Written/signed informed consent: confirm ☐
- *S.aureus* bloodstream infection/bacteremia: confirm ☐
- Community-onset disease (positive samples taken ≤48 h of admission [including at transferring hospital]): confirm ☐
- Patient is eligible: confirm ☐

Investigator Signature Date

NOTE: After completion of this CRF, please send CRF saved under the unique patient identifier (see above to the left-hand side) to the data center via crf@uks.eu, print it, add information needed locally, sign it and store it at a safe place.
You will soon receive an excel file with data exported from this CRF for your cross-check and for uploading during preregistration of biomaterials from this case.

Name of responsible local study nurse/data manager
Date CRF completed and sent (ddmmyyyy): [ ]

Infection Biology and Epidemiology of Staphylococci and Staphylococcal Diseases in Sub-Saharan Africa (DFG PAK296)

- Place of birth (country):
- Ethnicity:
- Place of residence
  - name of city or village/county (or region)
  - country
  - postal code (if available):
  - GPS coordinates:
    (to be filled in the printed version if possible; for later reporting during biomaterials preregistration)

- Previous hospitalization during the last 6 months prior to the present admission: ☐ yes ☐ no ☐ unknown
- Close health care contact (e.g. parenteral therapy, wound care, hemodialysis) within 30 days prior to SAB: ☐ yes ☐ no ☐ unknown
- Continuous residency in a nursing home before SAB onset: ☐ yes ☐ no ☐ unknown
- History of tuberculosis in the last 6 months: ☐ yes ☐ no ☐ unknown
- Antituberculous drugs within the last 4 weeks: ☐ yes ☐ no ☐ unknown
  if, yes, was rifampicin including the regimen: ☐ yes ☐ no ☐ unknown
- Antibiotics within the last 4 weeks: ☐ yes ☐ no ☐ unknown

- Presentation with or development of severe sepsis 48 h following admission/presentation: ☐ yes ☐ no ☐ unknown
- Presentation with or development of septic shock 48 h following admission/presentation: ☐ yes ☐ no ☐ unknown
- Clinical site of infection:
  - superficial [skin & skin structure] ☐
  - other (deep) sites to be specified: ☐
    - bone ☐
    - joint ☐
    - deep skin abscess ☐

Infection Biology and Epidemiology of Staphylococci and Staphylococcal Diseases in Sub-Saharan Africa (DFG PAK296)

- muscle: thigh ☐
- muscle: regions other than thigh ☐
- fasciitis ☐
- respiratory tract/lungs incl. Pleura ☐
- heart/heart valve ☐
- CSF/brain ☐
- urinary tract ☐
- other, please specify ☐

- New metastatic lesions detected 8 or more days after disease onset:
  - no new metastatic lesions identified ☐
  - bone ☐
  - joint ☐
  - deep skin abscess ☐
  - muscle: thigh ☐
  - muscle: regions other than thigh ☐
  - fasciitis ☐
  - respiratory tract/lungs incl. Pleura ☐
  - heart/heart valve ☐
  - CSF/brain ☐
  - urinary tract ☐
  - other ☐

Infection Biology and Epidemiology of Staphylococci and Staphylococcal Diseases in Sub-Saharan Africa (DFG PAK296)

- Intravenous drug abuse (IVDA): ☐yes ☐no ☐unknown
- Intravascular catheter(s) present prior to disease onset: ☐yes ☐no ☐unknown
- Other intravascular foreign bodies (i.e.artificial heart valves): ☐yes ☐no ☐unknown
  - if yes, please specify
- Other foreign bodies (i.e. arthroplasty): ☐yes ☐no ☐unknown
  - if yes, please specify
- HIV infection: ☐yes ☐no ☐unknown
- Charlson comorbidity score items
  - Myocardial infarction/coronary heart disease ☐yes ☐no ☐unknown
  - Congestive heart failure ☐yes ☐no ☐unknown
  - Peripheral vascular disease ☐yes ☐no ☐unknown
  - Cerebrovascular disease ☐yes ☐no ☐unknown
  - Dementia ☐yes ☐no ☐unknown
  - Chronic obstructive pulmonary disease ☐yes ☐no ☐unknown
  - Connective tissue disease ☐yes ☐no ☐unknown
  - Peptic ulcer disease ☐yes ☐no ☐unknown
  - Mild liver disease ☐yes ☐no ☐unknown
    OR moderate-severe liver disease ☐yes ☐no ☐unknown
  - Diabetes mellitus ☐yes ☐no ☐unknown
    OR diabetes mellitus with organ damage ☐yes ☐no ☐unknown
  - Hemiplegia ☐yes ☐no ☐unknown
  - Moderate-severe renal disease ☐yes ☐no ☐unknown
  - Any tumour (within last 5 years) ☐yes ☐no ☐unknown
    Lymphoma ☐yes ☐no ☐unknown
    Leukemia ☐yes ☐no ☐unknown
    Metastatic solid tumor ☐yes ☐no ☐unknown
  - AIDS ☐yes ☐no ☐unknown

- McCabe-Jackson underlying disease prognosis score items
  - Does the patient have an underlying acute, fulminant disease expected to be rapidly (weeks/months) fatal ☐ yes ☐ no ☐ unknown
  - Does the patient have a chronic, incurable condition that is expected to be fatal within the next few years? ☐ yes ☐ no ☐ unknown
  - Does the patient have a chronic condition not expected to be fatal within the next 4 years? ☐ yes ☐ no ☐ unknown
- Current parasitic diseases (results within 7 days prior or subsequent to initial positive sampling)
  - Positive blood smear for _____ sp:
    if positive: date of sampling (ddmmyyyy)
    number of trophozoites [/µL]
    or negative/not done/unknown
  - Positive blood smear for microfilariae:
    if positive: date of sampling (ddmmyyyy)
    species of
    or negative/not done/unknown
  - Positive stool sampling for intestinal helminths / ova:
    if positive: date of sampling (ddmmyyyy)
    species species 1
    species 2
    species 3
    or negative/not done/unknown
  - Positive urine microscopy for _____:
    if positive: date of sampling (ddmmyyyy)
    or negative/not done/unknown

Infection Biology and Epidemiology of Staphylococci and Staphylococcal Diseases in Sub-Saharan Africa (DFG PAK296)

- Antibiogram
  - penicillin ☐S ☐R ☐unknown
  - cefoxitin ☐S ☐R ☐unknown
  - tetracycline ☐S ☐I ☐R ☐unknown
  - erythromycin ☐S ☐I ☐R ☐unknown
  - clindamycin ☐S ☐I ☐R ☐unknown
  - gentamicin ☐S ☐I ☐R ☐unknown
  - chloramphenicol ☐S ☐I ☐R ☐unknown
  - trimethoprim-sulfamethoxazole ☐S ☐I ☐R ☐unknown
- E-test results (MIC in µg/mL)
  - Cefoxitin ☐☐☐.☐☐
  - Clindamycin ☐☐☐.☐☐
  - Linezolid ☐☐☐.☐☐
  - Vancomycin ☐☐☐.☐☐
  - Daptomycin ☐☐☐.☐☐
  - Tigecycline ☐☐☐.☐☐
- Other tests
  - MRSA detection ☐yes ☐no ☐unknown

    if yes, describe method/result

  - Induced clindamycin resistance ☐yes ☐no ☐unknown
- Were there multiple samples from identical site(s) positive for : ☐yes ☐no ☐unknown
- Were there multiple samples from different sites positive for : ☐yes ☐no ☐unknown

  if yes, please specify the sites

NOTE: Please store and send up to three isolates (if available) per case including the initial blood culture isolate (mandatory) to the study center

Infection Biology and Epidemiology of Staphylococci and Staphylococcal Diseases in Sub-Saharan Africa (DFG PAK296)

- Antibiotic(s) given on day 01:

name: application:
daily dose: g
name: application:
daily dose: g

- Antibiotic(s) given on day 02:

name: application:
daily dose: g
name: application:
daily dose: g

- Antibiotic(s) given on day 03:

name: application:
daily dose: g
name: application:
daily dose: g

- Antibiotic(s) given on day 04:

name: application:
daily dose: g
name: application:
daily dose: g

- Antibiotic(s) given on day 05:

name: application:
daily dose: g
name: application:
daily dose: g

- Antibiotic(s) given on day 06:

name: application:
daily dose: g
name: application:
daily dose: g

Infection Biology and Epidemiology of Staphylococci and Staphylococcal Diseases in Sub-Saharan Africa (DFG PAK296)

- Antibiotic(s) given on day 07:

name: application:
daily dose: , g
name: application:
daily dose: , g

- Antibiotic(s) given on day 08:

name: application:
daily dose: , g
name: application:
daily dose: , g

- Antibiotic(s) given on day 09:

name: application:
daily dose: , g
name: application:
daily dose: , g

- Antibiotic(s) given on day 10:

name: application:
daily dose: , g
name: application:
daily dose: , g

- Antibiotic(s) given on day 11:

name: application:
daily dose: , g
name: application:
daily dose: , g

- Antibiotic(s) given on day 12:

name: application:
daily dose: , g
name: application:
daily dose: , g

Infection Biology and Epidemiology of Staphylococci and Staphylococcal Diseases in Sub-Saharan Africa (DFG PAK296)

- Antibiotic(s) given on day 13:

  name: application:
  daily dose: . g
  name: application:
  daily dose: . g

- Antibiotic(s) given on day 14:

  name: application:
  daily dose: . g
  name: application:
  daily dose: . g

- Did the patient receive more than 14 full days of antibacterial therapy for staphylococcal disease? ☐ yes ☐ no ☐ unknown

- Incision/drainage
  - within first 3 days after disease onset: ☐ yes ☐ no ☐ unknown
  - within 4-7 days after disease onset: ☐ yes ☐ no ☐ unknown
  - within 8-14 days after disease onset: ☐ yes ☐ no ☐ unknown
- Removal of intravascular catheter (if any):
  - within first 3 days after disease onset: ☐ yes ☐ no ☐ unknown/not applicable
  - within 4-7 days after disease onset: ☐ yes ☐ no ☐ unknown/not applicable
  - within 8-14 days after disease onset: ☐ yes ☐ no ☐ unknown/not applicable
- Removal of foreign body (other than iv catheter, if any):
  - within first 3 days after disease onset: ☐ yes ☐ no ☐ unknown/not applicable
  - within 4-7 days after disease onset: ☐ yes ☐ no ☐ unknown/not applicable
  - within 8-14 days after disease onset: ☐ yes ☐ no ☐ unknown/not applicable
- Other surgical procedures within 14 days after disease onset because of staphylococcal disease: ☐ yes ☐ no ☐ unknown

  if yes, type of surgery

  and date of surgery (ddmmyyyy)

Infection Biology and Epidemiology of Staphylococci and Staphylococcal Diseases in Sub-Saharan Africa (DFG PAK296)

- Was the patient admitted to hospital care because of staphylococcal disease: ☐ yes ☐ no ☐ unknown
  - if yes, date of discharge (dd.mm.yyyy) ☐ unknown
- Did the patient die within 14 days after disease onset: ☐ yes ☐ no ☐ unknown
  - if yes, date of death (ddmmyyyy):
  - specify (probable) cause of death:
  - if not: last date known to be alive (ddmmyyyy)

- left-sided endocarditis ☐
- right-sided endocarditis ☐
- pneumonia ☐
- vertrebral osteomyelitis/spondylodiscitis ☐
- other bone/joint infection ☐
- skin/soft-tissue infection ☐
- iv catheter-/device-associated bacteremia without secondary lesion(s) ☐
- iv catheter-/device-associated bacteremia with secondary lesion(s) ☐
- disseminated infection (multiple metastatic lesions) ☐
- other, please specify

Infection Biology and Epidemiology of Staphylococci and Staphylococcal Diseases in Sub-Saharan Africa (DFG PAK296)

Initial isolate

- Date of sampling (ddmmyyyy):
- Confirm blood site of sampling:
- Local laboratory identifier/number isolate:
- Unique study isolate label: Study site/subject no.

Second isolate

- Not done
- Date of sampling (ddmmyyyy):
- Site of sampling:
- Local laboratory identifier/number isolate:
- Unique study isolate label: Study site/subject no.

Third isolate

- Not done
- Date of sampling (ddmmyyyy):
- Site of sampling:
- Local laboratory identifier/number isolate:
- Unique study isolate label: Study site/subject no.

NOTE: Please send appropriately labelled materials as soon as possible to central storage facility. Note that you need to preregister at the StaphNet-Portal all specimens to be shipped; preregistration includes uploading the CRF data in an exported excel file as provided by the data center.

Name of responsible lab technician:

Infection Biology and Epidemiology of Staphylococci and Staphylococcal Diseases in Sub-Saharan Africa (DFG PAK296)

First serum / plasma

- Not done
- Date of serum / plasma sampling (ddmmyyyy):
- Local laboratory identifier/number serum/plasma sample:
- Unique serum label: Study site/subject no.
- Unique plasma label: Study site/subject no.

Second serum / plasma

- Not done
- Date of serum / plasma sampling (ddmmyyyy):
- Local laboratory identifier/number serum/plasma sample:
- Unique serum label: Study site/subject no.
- Unique plasma label: Study site/subject no.

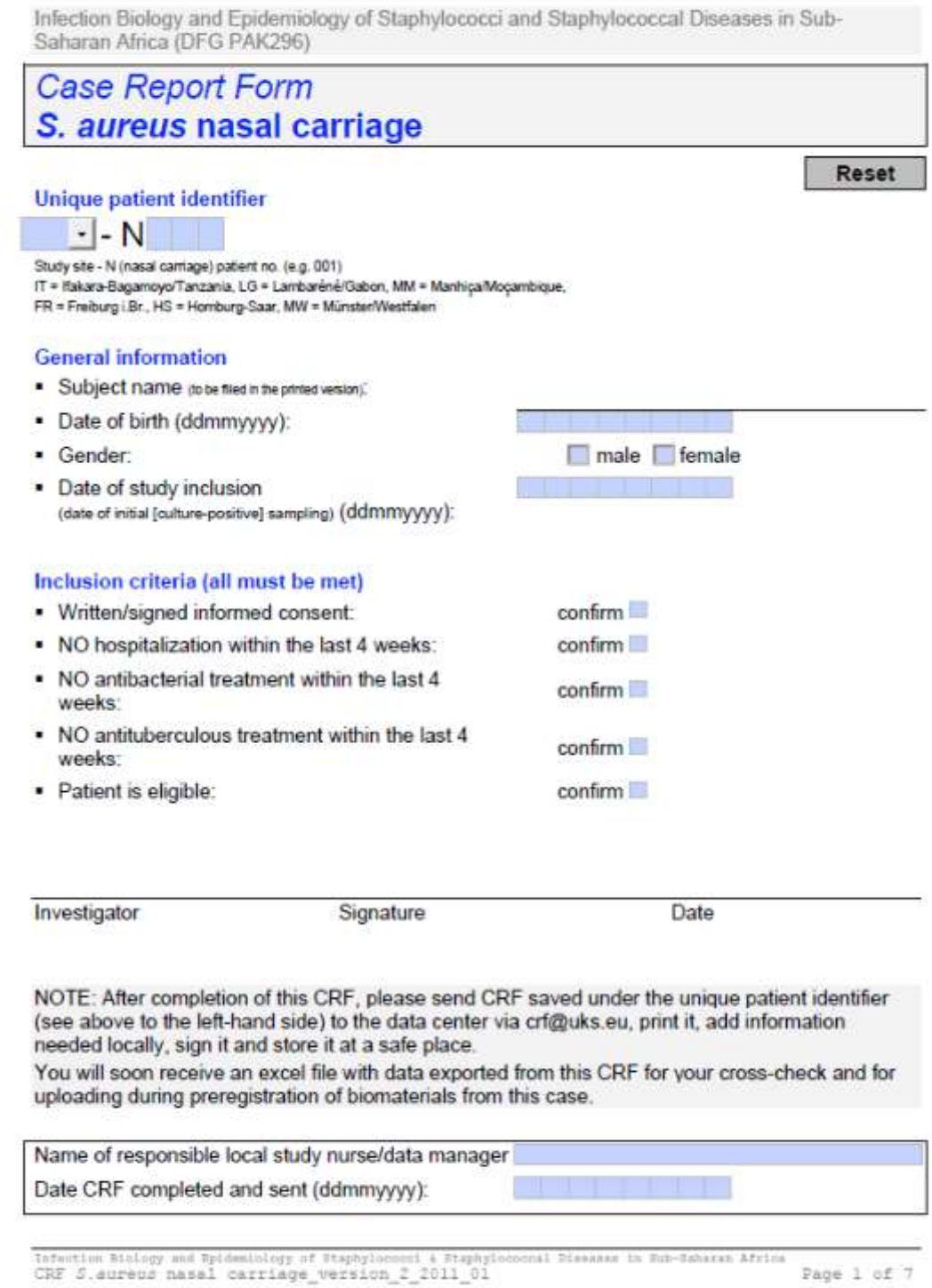
Third serum / plasma

- Not done
- Date of sampling (ddmmyyyy):
- Local laboratory identifier/number serum/plasma sample:
- Unique serum label: Study site/subject no.
- Unique plasma label: Study site/subject no.

NOTE: Please send appropriately labelled materials as soon as possible to central storage facility. Note that you need to preregister at the StaphNet-Portal all specimens to be shipped; preregistration includes uploading the CRF data in an exported excel file as provided by the data center.

Name of responsible lab technician:

Figure 30: **Case report form of *Staphylococcus aureus* (SA) isolates of nasal carriage.** The case report form defined the isolate number of the collected SA isolate of a nasal swab of a healthy volunteer, collected general patient information, inclusion criteria to exclude SA isolates of hospital acquired infections, clinical patient data and microbiologic data of the SA isolates.



Infection Biology and Epidemiology of Staphylococci and Staphylococcal Diseases in Sub-Saharan Africa (DFG PAK296)

Case Report Form
***S. aureus* nasal carriage**

Reset

Unique patient identifier

- N

Study site - N (nasal carriage) patient no. (e.g. 001)
IT = Ifakara-Bagamoyo/Tanzania, LG = Lambaréné/Gabon, MM = Manhiça/Moçambique,
FR = Freiburg i.Br., HS = Homburg-Saar, MW = Münster/Westfalen

General information

- Subject name (to be filled in the printed version):
- Date of birth (ddmmyyyy):
- Gender: male female
- Date of study inclusion (date of initial [culture-positive] sampling) (ddmmyyyy):

Inclusion criteria (all must be met)

- Written/signed informed consent: confirm
- NO hospitalization within the last 4 weeks: confirm
- NO antibacterial treatment within the last 4 weeks: confirm
- NO antituberculous treatment within the last 4 weeks: confirm
- Patient is eligible: confirm

Investigator Signature Date

NOTE: After completion of this CRF, please send CRF saved under the unique patient identifier (see above to the left-hand side) to the data center via crf@uks.eu, print it, add information needed locally, sign it and store it at a safe place.
You will soon receive an excel file with data exported from this CRF for your cross-check and for uploading during preregistration of biomaterials from this case.

Name of responsible local study nurse/data manager
Date CRF completed and sent (ddmmyyyy):

Infection Biology and Epidemiology of Staphylococci & Staphylococcal Diseases in Sub-Saharan Africa
CRF S.aureus nasal carriage_version_2_2011_01 Page 1 of 7

Infection Biology and Epidemiology of Staphylococci and Staphylococcal Diseases in Sub-Saharan Africa (DFG PAK296)

Case Report Form
S. aureus nasal carriage

Demographic data

- Place of birth (country):
- Ethnicity:
- Place of residence
  - name of city or village/county (or region)
  - country
  - postal code (if available):
  - GPS coordinates:
    (to be filled in the printed version if possible; for later reporting during biomaterials preregistration)

Patient history

- Had the subject been hospitalized 6 months until 4 weeks prior to sampling: ☐ yes ☐ no ☐ unknown

NOTE: Patients newly admitted are eligible, but patients previously hospitalized within the last 4 weeks are ineligible.

- Diseased person: ☐ yes ☐ no
  - if yes, describe (pathology / organ system):
  - if yes, newly admitted to hospital: ☐ yes ☐ no ☐ unknown
- Continuous residency in a nursing home before sampling: ☐ yes ☐ no ☐ unknown

Risk factors/comorbidities/underlying disease

- Intravenous drug abuse (IVDA): ☐ yes ☐ no ☐ unknown
- Intravascular catheter(s) present prior to disease onset: ☐ yes ☐ no ☐ unknown
- Other intravascular foreign bodies (i.e.artificial heart valves): ☐ yes ☐ no ☐ unknown
  - if yes, please specify

Infection Biology and Epidemiology of Staphylococci and Staphylococcal Diseases in Sub-Saharan Africa (DFG PAK296)

Case Report Form
S. aureus nasal carriage

- Other foreign bodies (i.e. arthroplasty): ☐ yes ☐ no ☐ unknown
  - if yes, please specify
- HIV infection: ☐ yes ☐ no ☐ unknown
- Charlson comorbidity score items
  - Myocardial infarction/coronary heart disease ☐ yes ☐ no ☐ unknown
  - Congestive heart failure ☐ yes ☐ no ☐ unknown
  - Peripheral vascular disease ☐ yes ☐ no ☐ unknown
  - Cerebrovascular disease ☐ yes ☐ no ☐ unknown
  - Dementia ☐ yes ☐ no ☐ unknown
  - Chronic obstructive pulmonary disease ☐ yes ☐ no ☐ unknown
  - Connective tissue disease ☐ yes ☐ no ☐ unknown
  - Peptic ulcer disease ☐ yes ☐ no ☐ unknown
  - Mild liver disease ☐ yes ☐ no ☐ unknown
    OR moderate-severe liver disease ☐ yes ☐ no ☐ unknown
  - Diabetes mellitus ☐ yes ☐ no ☐ unknown
    OR diabetes mellitus with organ damage ☐ yes ☐ no ☐ unknown
  - Hemiplegia ☐ yes ☐ no ☐ unknown
  - Moderate-severe renal disease ☐ yes ☐ no ☐ unknown
  - Any tumour (within last 5 years) ☐ yes ☐ no ☐ unknown
    Lymphoma ☐ yes ☐ no ☐ unknown
    Leukemia ☐ yes ☐ no ☐ unknown
    Metastatic solid tumor ☐ yes ☐ no ☐ unknown
  - AIDS ☐ yes ☐ no ☐ unknown
- McCabe-Jackson underlying disease prognosis score items
  - Does the patient have an underlying acute, fulminant disease expected to be rapidly (weeks/months) fatal ☐ yes ☐ no ☐ unknown
  - Does the patient have a chronic, incurable condition that is expected to be fatal within the next few years? ☐ yes ☐ no ☐ unknown
  - Does the patient have a chronic condition not expected to be fatal within the next 4 years? ☐ yes ☐ no ☐ unknown

Infection Biology and Epidemiology of Staphylococci and Staphylococcal Diseases in Sub-Saharan Africa (DFG PAK296)

Case Report Form
S. aureus nasal carriage

- Current parasitic diseases (results within 7 days prior or subsequent to initial positive *S. aureus* sampling)
  - Positive blood smear for *Plasmodium* sp:
    if positive: date of sampling (ddmmyyyy)
    number of trophozoites [/µL]
    or negative/not done/unknown
  - Positive blood smear for microfilariae:
    if positive: date of sampling (ddmmyyyy)
    species of *Filaria*
    or negative/not done/unknown
  - Positive stool sampling for intestinal helminths / ova:
    if positive: date of sampling (ddmmyyyy)
    species
      species 1
      species 2
      species 3
    or negative/not done/unknown
  - Positive urine microscopy for *S. haematobium*:
    if positive: date of sampling (ddmmyyyy)
    or negative/not done/unknown

Infection Biology and Epidemiology of Staphylococci and Staphylococcal Diseases in Sub-Saharan Africa (DFG PAK296)

Case Report Form
***S. aureus* nasal carriage**

Microbiology (initial isolate)

- Antibiogram
  - penicillin ☐S ☐R ☐unknown
  - cefoxitin ☐S ☐R ☐unknown
  - tetracycline ☐S ☐I ☐R ☐unknown
  - erythromycin ☐S ☐I ☐R ☐unknown
  - clindamycin ☐S ☐I ☐R ☐unknown
  - gentamicin ☐S ☐I ☐R ☐unknown
  - chloramphenicol ☐S ☐I ☐R ☐unknown
  - trimethoprim-sulfamethoxazole ☐S ☐I ☐R ☐unknown
- E-test results (MIC in µg/mL)
  - Cefoxitin ☐☐☐.☐☐
  - Clindamycin ☐☐☐.☐☐
  - Linezolid ☐☐☐.☐☐
  - Vancomycin ☐☐☐.☐☐
  - Daptomycin ☐☐☐.☐☐
  - Tigecycline ☐☐☐.☐☐
- Other tests
  - MRSA detection ☐yes ☐no ☐unknown

    if yes, describe method/result
  - Induced clindamycin resistance ☐yes ☐no ☐unknown

Infection Biology and Epidemiology of Staphylococci and Staphylococcal Diseases in Sub-Saharan Africa (DFG PAK296)

Case Report Form
S. aureus nasal carriage

Sampling *S. aureus* isolates

Initial isolate

- Date of sampling (ddmmyyyy):
- Confirm nasal site of sampling:
- Local laboratory identifier/number *S. aureus* isolate:
- Unique *S. aureus* study isolate label: -N -I1 (Study site/subject no.)

Second isolate

- Not done
- Date of sampling (ddmmyyyy):
- Site of sampling:
- Local laboratory identifier/number *S. aureus* isolate:
- Unique *S. aureus* study isolate label: -N -I2 (Study site/subject no.)

Third isolate

- Not done
- Date of sampling (ddmmyyyy):
- Site of sampling:
- Local laboratory identifier/number *S. aureus* isolate:
- Unique *S. aureus* study isolate label: -N -I3 (Study site/subject no.)

NOTE: Please send appropriately labelled materials as soon as possible to central storage facility. Note that you need to preregister at the StaphNet-Portal all specimens to be shipped; preregistration includes uploading the CRF data in an exported excel file as provided by the data center.

Name of responsible lab technician:

Do you confirm that the above data on sampling and labelling are correct:

Infection Biology and Epidemiology of Staphylococci and Staphylococcal Diseases in Sub-Saharan Africa (DFG PAK296)

Case Report Form
S. aureus nasal carriage

Plasma/serum sampling

First serum / plasma

- Not done ☐
- Date of serum / plasma sampling (ddmmyyyy): ☐☐☐☐☐☐☐☐
- Local laboratory identifier/number serum/plasma sample: ☐☐☐☐☐☐☐☐☐☐
- Unique serum label: ☐ -N ☐☐☐ -S1 Study site/subject no.
- Unique plasma label: ☐ -N ☐☐☐ -P1 Study site/subject no.

Second serum / plasma

- Not done ☐
- Date of serum / plasma sampling (ddmmyyyy): ☐☐☐☐☐☐☐☐
- Local laboratory identifier/number serum/plasma sample: ☐☐☐☐☐☐☐☐☐☐
- Unique serum label: ☐ -N ☐☐☐ -S2 Study site/subject no.
- Unique plasma label: ☐ -N ☐☐☐ -P2 Study site/subject no.

Third serum / plasma

- Not done ☐
- Date of sampling (ddmmyyyy): ☐☐☐☐☐☐☐☐
- Local laboratory identifier/number serum/plasma sample: ☐☐☐☐☐☐☐☐☐☐
- Unique serum label: ☐ -N ☐☐☐ -S3 Study site/subject no.
- Unique plasma label: ☐ -N ☐☐☐ -P3 Study site/subject no.

NOTE: Please send appropriately labelled materials as soon as possible to central storage facility. Note that you need to preregister at the StaphNet-Portal all specimens to be shipped; preregistration includes uploading the CRF data in an exported excel file as provided by the data center.

Name of responsible lab technician: ☐☐☐☐☐☐☐☐☐☐

Do you confirm that the above data on sampling and labelling are correct: ☐

6.3 Characteristics of African and German healthy volunteers and patients

Table 22: **Characteristics of African and German healthy volunteers with respect to the standardized questionnaire from whom nasal *Staphylococcus aureus* isolates were obtained.** Only participants were included without hospitalization, antibacterial and antituberculous treatment within the last 4 weeks.

	Total	Africa (n=300)	Germany (n=300)	p-value*
Median# age (range), yrs	22 (0-89)	18 (0-61)	23 (0-89)	<0.001
Male gender, n (%)	280	132 (44%)	148 (49%)	ns
History of hospital admission 6 months	14 (2%)	1 (<1%)	13 (4%)	<0.01
Known HIV infection	15 (3%)	15 (5%)	0	<0.001
History of AIDS	7 (1%)	7 (2%)	0	0.02
History of peripheral vascular disease	6 (1%)	0	6 (2%)	0.03
History of connective tissue disease	8 (1%)	0	8 (3%)	0.01
Known diabetes	8 (1%)	0	8 (3%)	0.01

*p-values were calculated by Chi-square corrected for multiple testing using R Studio

Median was calculated using SPSS IBM

Table 23: **Characteristics of African and German patients from whom clinical community-associated *Staphylococcus aureus* isolates of wound infections and blood cultures were obtained.**

	Total	Africa (n=300)	Germany (n=300)	p-value*
Age (yrs), median (range)	29 (0-98)	3 (0-71)	53 (0-98)	<0.001
Male gender, n (%)	348 (58%)	160 (53%)	188 (63%)	ns
History of hospital admission between 6 months and 4 weeks prior to to sampling	157 (26%)	29 (10%)	128 (43%)	<0.01
Healthcare contact in the past 4 weeks	129 (22%)	36 (12%)	93 (31%)	<0.001
Continuous residency in a nursing home before sampling	4 (<1%)	2 (<1%)	2 (<1%)	ns
Antibiotic treatment in the past 4 weeks	133 (22%)	55 (18%)	78 (26%)	ns
Tuberculosis in the past 6 months	4 (<1%)	4 (1%)	0	ns
Antituberculous drugs in the past 4 weeks	1 (<1%)	1 (<1%)	0	ns
Risk factors for staphylococcal disease				
History of intravenous drug abuse	7 (1%)	0	7 (2%)	0.02
Intravenous catheter in place	23 (4%)	0	23 (8%)	<0.001
Other intravascular devices in place	28 (5%)	0	28 (9%)	<0.001
Non-vascular foreign body/device in place	73 (12%)	3 (1%)	70 (23%)	<0.001
Known HIV infection	26 (4%)	26 (9%)	0	<0.001
Charlson comorbidity score, median (range)	0 (0-12)	0 (0-0)	1 (0-12	ns
History of AIDS	8 (1%)	8 (3%)	0	0.01
History of myocardial infarction	37 (6%)	0	37 (12%)	<0.001
History of congestive heart failure	26 (4%)	0	26 (9%)	<0.001
History of peripheral vascular disease	65 (11%)	0	65 (22%)	<0.001
History of cerebral vascular disease	65 (11%)	0	65 (22%)	<0.001
Dementia	7 (1%)	0	7 (2%)	0.02
Chronic obstructive lung disease	19 (3%)	2 (<1%)	17 (6%)	<0.001
History of connective tissue disease	30 (5%)	0	30 (10%)	<0.001
History of peptic ulcer disease	14 (2%)	2 (<1%)	12 (4%)	0.01
Mild liver disease	21 (4%)	0	21 (7%)	<0.001
Moderate to severe liver disease	7 (1%)	0	7 (2%)	0.02
Known diabetes	64 (11%)	4 (1%)	60 (20%)	<0.001
Diabetes with organ damage	15 (3%)	0	15 (5%)	<0.001
Moderate to severe renal disease	25 (4%)	0	25 (8%)	<0.001
Solid tumor	51 (9%)	1 (<1%)	50 (17%)	<0.001
Metastatic solid tumor	20 (3%)	0	20 (7%)	<0.001
Lymphoma	6 (1%)	0	6 (2%)	0.03

	Total	Africa (n=300)	Germany (n=300)	p-value*
Severity of chronic underlying disorder(s) according to McCabe & Jackson classification				
Rapidly fatal underlying disease	30 (5%)	19 (6%)	11 (4%)	ns
Ultimately fatal disease [< 5 years]	58 (10%)	17 (6%)	41 (14%)	ns
Non-fatal underlying disease	132 (22%)	22 (7%)	110 (37%)	<0.001
Type of infection				
Bloodstream infection	79 (13%)	29 (10%)	50 (17%)	ns
Non-bacteremic infection	521 (87%)	271 (90%)	250 (83%)	ns
Acute clinical presentation				
Severe sepsis	38 (6%)	18 (6%)	20 (7%)	ns
Septic shock	7 (1%)	2 (<1%)	5 (2%)	ns
Clinical site(s) of infection				
Superficial skin infection	348 (58%)	191 (64%)	157 (52%)	ns
Deep skin abscess	121 (20%)	82 (27%)	39 (13%)	0.002
Bone	22 (4%)	0	22 (7%)	<0.001
Joint	20 (3%)	3 (1%)	17 (6%)	ns
Muscle	11 (2%)	4 (1%)	7 (2%)	ns
Fascia	2 (<1%)	1 (<1%)	1 (<1%)	ns
Respiratory tract	25 (4%)	6 (2%)	19 (6%)	ns
Heart	5 (1%)	0	5 (2%)	ns
Central nervous system	1 (<1%)	1 (<1%)	0	ns
Urinary tract	7 (1%)	0	7 (2%)	ns
Other	87 (15%)	20 (7%)	67 (22%)	<0.001
New metastatic lesions	54 (9%)	25 (8%)	29 (10%)	ns

*p-values were calculated by Chi-square corrected for multiple testing

Median were calculated using SPSS IBM

7 References

1. Aanensen, D.M. & B.G. Spratt, (2005) The multilocus sequence typing network: mlst.net. *Nucleic acids research* **33**: W728-733.

2. Abdulgader, S.M., A.O. Shittu, M.P. Nicol & M. Kaba, (2015) Molecular epidemiology of Methicillin-resistant *Staphylococcus aureus* in Africa: a systematic review. *Frontiers in microbiology* **6**: 348.

3. Abubakar, U. & A.S. Syed, (2018) Prevalence, trend and antimicrobial susceptibility of methicillin resistant *Staphylococcus aureus* in Nigeria: a systematic review. *Journal of Infection and Public Health* **11**: 763–770.

4. Aires-de-Sousa, M., K. Boye, H. de Lencastre, A. Deplano, M.C. Enright, J. Etienne, A. Friedrich, D. Harmsen, A. Holmes, X.W. Huijsdens, A.M. Kearns, A. Mellmann, H. Meugnier, J.K. Rasheed, E. Spalburg, B. Strommenger, M.J. Struelens, F.C. Tenover, J. Thomas, U. Vogel, H. Westh, J. Xu & W. Witte, (2006) High interlaboratory reproducibility of DNA sequence-based typing of bacteria in a multicenter study. *Journal of clinical microbiology* **44**: 619-621.

5. Aires-de-Sousa, M. & H. De Lencastre, (2003) Evolution of sporadic isolates of methicillin-resistant *Staphylococcus aureus* (MRSA) in hospitals and their similarities to isolates of community-acquired MRSA. *Journal of clinical microbiology* **41**: 3806–3815.

6. Ako-Nai, A.K., A.D. Ogunniyi, A. Lamikanra & S.E. Torimiro, (1991) The characterisation of clinical isolates of *Staphylococcus aureus* in Ile-Ife, Nigeria. *Journal of medical microbiology* **34**: 109-112.

7. Akulenko, R., M. Merl & V. Helms, (2016) BEclear: Batch Effect Detection and Adjustment in DNA Methylation Data. *PLoS One* **11**: e0159921.

8. Alonzo, F., 3rd & V.J. Torres, (2014) The bicomponent pore-forming leucocidins of *Staphylococcus aureus*. *Microbiology and molecular biology reviews: MMBR* **78**: 199-230.

9. Anah, M.U., J.J. Udo, S.O. Ochigbo & L.N. Abia-Bassey, (2008) Neonatal septicaemia in Calabar, Nigeria. *Tropical doctor* **38**: 126-128.

10. Archer, G.L., (1998) *Staphylococcus aureus*: a well-armed pathogen. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* **26**: 1179-1181.

11. Archer, G.L. & C.G. Mayhall, (1983) Comparison of epidemiological markers used in the investigation of an outbreak of methicillin-resistant *Staphylococcus aureus* infections. *Journal of Clinical Microbiology* **18**: 395-399.

12. Arciola, C.R., D. Campoccia, P. Speziale, L. Montanaro & J.W. Costerton, (2012) Biofilm formation in *Staphylococcus* implant infections. A review of molecular mechanisms and implications for biofilm-resistant materials. *Biomaterials* **33**: 5967-5982.

13. Argudin, M.A., M.C. Mendoza & M.R. Rodicio, (2010) Food poisoning and *Staphylococcus aureus* enterotoxins. *Toxins* **2**: 1751-1773.

14. Baba, T., F. Takeuchi, M. Kuroda, H. Yuzawa, K. Aoki, A. Oguchi, Y. Nagai, N. Iwama, K. Asano, T. Naimi, H. Kuroda, L. Cui, K. Yamamoto & K. Hiramatsu, (2002) Genome and virulence determinants of high virulence community-acquired MRSA.

Lancet **359**: 1819-1827.

15. Baig, S., T.B. Johannesen, S. Overballe-Petersen, J. Larsen, A.R. Larsen & M. Stegger, (2018) Novel SCC*mec* type XIII (9A) identified in an ST152 methicillin-resistant *Staphylococcus aureus*. *Infection, genetics and evolution: journal of molecular epidemiology and evolutionary genetics in infectious diseases* **61**: 74-76.

16. Bailey, A.L., M.A. Monck & P.R. Cullis, (1997) pH-induced destabilization of lipid bilayers by a lipopeptide derived from influenza hemagglutinin. *Biochimica et biophysica acta* **1324**: 232-244.

17. Barber, M. & J.E. McCartney, (1952) *S. aureus* infection in a maternity hospital. *Edinburgh Medicinal Journal* **59**: 200–207.

18. Barbuddhe, S.B., T. Maier, G. Schwarz, M. Kostrzewa, H. Hof, E. Domann, T. Chakraborty & T. Hain, (2008) Rapid identification and typing of listeria species by matrix-assisted laser desorption ionization-time of flight mass spectrometry. *Applied and environmental microbiology* **74**: 5402-5407.

19. Bartels, M.D., H. Larner-Svensson, H. Meiniche, K. Kristoffersen, K. Schonning, J.B. Nielsen, S.M. Rohde, L.B. Christensen, A.W. Skibsted, J.O. Jarlov, H.K. Johansen, L.P. Andersen, I.S. Petersen, D.W. Crook, R. Bowden, K. Boye, P. Worning & H. Westh, (2015) Monitoring meticillin resistant *Staphylococcus aureus* and its spread in Copenhagen, Denmark, 2013, through routine whole genome sequencing. *Euro surveillance: bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin* **20**.

20. Basset, P., U. Nübel, W. Witte, D.S. Blanc, (2012) *Evaluation of adding a second marker to overcome Staphylococcus aureus spa typing homoplasies. Journal of Clinical Microbiology* **50**:1475–1477.

21. Bayer, M.G., J.H. Heinrichs & A.L. Cheung, (1996) The molecular architecture of the sar locus in *Staphylococcus aureus*. *Journal of bacteriology* **178**: 4563-4570.

22. Becker K., S. van Alen, E.A. Idelevich, N. Schleimer, J. Seggewiß, A. Mellmann, U. Kaspar, G. Peters, (2018) Plasmid-encoded transferable *mecB*-mediated methicillin resistance in S*taphylococcus aureus*. *Emerging Infectious Diseases* **24**: 242-248.

23. Becker, K., A.W. Friedrich, G. Lubritz, M. Weilert, G. Peters & C. von Eiff, (2003) Prevalence of genes encoding pyrogenic toxin superantigens and exfoliative toxins among strains of *Staphylococcus aureus* isolated from blood and nasal specimens. *Journal of clinical microbiology* **41**: 1434-1439.

24. Ben Nejma, M., M. Mastouri, B. Bel Hadj Jrad & M. Nour, (2013) Characterization of ST80 Panton-Valentine leukocidin-positive community-acquired methicillin-resistant *Staphylococcus aureus* clone in Tunisia. *Diagnostic microbiology and infectious disease* **77**: 20-24.

25. Benagli, C., V. Rossi, M. Dolina, M. Tonolla & O. Petrini, (2011) Matrix-assisted laser desorption ionization-time of flight mass spectrometry for the identification of clinically relevant bacteria. *PLoS One* **6**: e16424.

26. Berger-Bachi, B. & S. Rohrer, (2002) Factors influencing methicillin resistance in staphylococci. *Archives of microbiology* **178**: 165-171.

27. Berkley, J.A., B.S. Lowe, I. Mwangi, T. Williams, E. Bauni, S. Mwarumba, C. Ngetsa, M.P. Slack, S. Njenga, C.A. Hart, K. Maitland, M. English, K. Marsh & J.A. Scott, (2005) Bacteremia among children admitted to a rural hospital in Kenya. *The New*

England journal of medicine **352**: 39-47.

28. Bernardo, K., N. Pakulat, M. Macht, O. Krut, H. Seifert, S. Fleer, F. Hunger &. Kronke, (2002) Identification and discrimination of *Staphylococcus aureus* strains using matrix-assisted laser desorption/ionization-time of flight mass spectrometry. *Proteomics* **2**: 747-753.

29. Billal, D.S., J. Feng, P. Leprohon, D. Legare & M. Ouellette, (2011) Whole genome analysis of linezolid resistance in *Streptococcus pneumoniae* reveals resistance and compensatory mutations. *BMC genomics* **12**: 512.

30. Blair, J.E. & R.E.0. Williams, (1961) Phage typing of staphylococci. *Bulletin of the World Health Organization* **24**: 771-784.

31. Blanc, D.S., P. Francioli & P.M. Hauser, (2002) Poor value of pulsed-field gel electrophoresis to investigate longterm scale epidemiology of methicillin-resistant *Staphylococcus aureus*. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases* **2**: 145-148.

32. Blomfeldt, A., H.V. Aamot, A.N. Eskesen, F. Muller & S. Monecke, (2013) Molecular characterization of methicillin-sensitive *Staphylococcus aureus* isolates from bacteremic patients in a Norwegian University Hospital. *Journal of clinical microbiology* **51**: 345-347.

33. Bohme, K., S. Morandi, P. Cremonesi, I.C. Fernandez No, J. Barros-Velazquez, B. Castiglioni, M. Brasca, B. Canas & P. Calo-Mata, (2012) Characterization of *Staphylococcus aureus* strains isolated from Italian dairy products by MALDI-TOF mass fingerprinting. *Electrophoresis* **33**: 2355-2364.

34. Bosch, T. & L.M. Schouls, (2015) Livestock-associated MRSA: innocentor serious health threat? *Future Microbiology* **10**: 445–447.

35. Boswihi, S.S., E.E. Udo, N. Al-Sweih, (2016) Shifts in the clonal distribution of methicillin-resistant *Staphylococcus aureus* in Kuwait Hospitals: 1992-2010. *PLos One* **11**: e0162744.

36. Breurec, S., S.B. Zriouil, C. Fall, P. Boisier, S. Brisse, S. Djibo, J. Etienne, M.C. Fonkoua, J.D. Perrier-Gros-Claude, R. Pouillot, C.E. Ramarokoto, F. Randrianirina, A. Tall, J.M. Thiberge, i. Working Group on *Staphylococcus aureus*, F. Laurent & B. Garin, (2011) Epidemiology of methicillin-resistant *Staphylococcus aureus* lineages in five major African towns: emergence and spread of atypical clones. *Clinical microbiology and infection: the official publication of the European Society of Clinical Microbiology and Infectious Diseases* **17**: 160-165.

37. Bronner, S., H. Monteil & G. Prevost, (2004) Regulation of virulence determinants in *Staphylococcus aureus*: complexity and applications. *FEMS microbiology reviews* **28**: 183-200.

38. Bryant, D. & V. Moulton, (2004) Neighbor-Net: An Agglomerative Method for the Construction of Phylogenetic Networks. *Molecular Biology and Evolution* **21**: 255–265.

39. Budimir, A., R.H. Deurenberg, Z. Bosnjak, E.E. Stobberingh, H. Cetkovic & S. Kalenic, (2010) A variant of the Southern German clone of methicillin-resistant *Staphylococcus aureus* is predominant in Croatia. *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases* **16**: 1077-1083.

40. Bumgarner, R., (2013) Overview of DNA microarrays: types, applications, and their

future. *Current protocols in molecular biology* **Chapter 22**: Unit 22.21.

41. Caddick, J.M., A.C. Hilton, R.A. Armstrong, P.A. Lambert, T. Worthington & T.S. Elliott, (2006) Description and critical appraisal of principal components analysis (PCA) methodology applied to pulsed-field gel electrophoresis profiles of methicillin-resistant *Staphylococcus aureus* isolates. *Journal of microbiological methods* **65**: 87-95.

42. Cadena, J., J. Thinwa, E.A. Walter & C.R. Frei, (2016) Risk factors for the development of active methicillin-resistant *Staphylococcus aureus* (MRSA) infection in patients colonized with MRSA at hospital admission. *American journal of infection control* **44**: 1617-1621.

43. Camacho, C., G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer & T.L. Madden, (2009) BLAST+: architecture and applications. *BMC bioinformatics* **10**: 421.

44. Camoez, M., J.M. Sierra, M.A. Dominguez, M. Ferrer-Navarro, J. Vila & I. Roca, (2015) Automated categorization of methicillin-resistant *Staphylococcus aureus* clinical isolates into different clonal complexes by MALDI-TOF mass spectrometry. *Clinical microbiology and infection: the official publication of the European Society of Clinical Microbiology and Infectious Diseases* **22**: 161.e1-161.e7.

45. Campbell, S.J., H.S. Deshmukh, C.L. Nelson, I.G. Bae, M.E. Stryjewski, J.J. Federspiel, G.T. Tonthat, T.H. Rude, S.L. Barriere, R. Corey & V.G. Fowler, Jr., (2008) Genotypic characteristics of *Staphylococcus aureus* isolates from a multinational trial of complicated skin and skin structure infections. *Journal of clinical microbiology* **46**: 678-684.

46. Candès, E. & B. Recht, (2012) Exact Matrix Completion via Convex Optimization. *Foundations of Computational Mathematics* **55**: 712-772.

47. Chambers, H.F. & F.R. DeLeo, (2009) Waves of resistance: *Staphylococcus aureus i*n the antibiotic era. *Nature reviews. Microbiology* **7**: 629-641.

48. Chambers, H.F., (2001) The changing epidemiology of *Staphylococcus aureus*? *Emerging Infectious Diseases*. **7**:178-182.

49. Chavakis, T., K.T. Preissner & M. Herrmann, (2007) The anti-inflammatory activities of *Staphylococcus aureus*. *Trends in immunology* **28**: 408-418.

50. Chavakis, T., K. Wiechmann, K.T. Preissner & M. Herrmann, (2005) *Staphylococcus aureus* interactions with the endothelium: the role of bacterial "secretable expanded repertoire adhesive molecules" (SERAM) in disturbing host defense systems. *Thrombosis and haemostasis* **94**: 278-285.

51. Cheung, A.L., M.G. Bayer & J.H. Heinrichs, (1997) sar Genetic determinants necessary for transcription of RNAII and RNAIII in the agr locus of Staphylococcus aureus. *Journal of bacteriology* **179**: 3963-3971.

52. Christner, M., H. Rohde, M. Wolters, I. Sobottka, K. Wegscheider & M. Aepfelbacher, (2010) Rapid identification of bacteria from positive blood culture bottles by use of matrix-assisted laser desorption-ionization time of flight mass spectrometry fingerprinting. *Journal of clinical microbiology* **48**: 1584-1591.

53. Chua, K., F. Laurent, G. Coombs, M.L. Grayson & B.P. Howden, (2011) Antimicrobial resistance: Not community-associated methicillin-resistant *Staphylococcus aureus* (CA-MRSA)! A clinician's guide to community MRSA - its evolving antimicrobial resistance and implications for therapy. *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America* **52**: 99-114.

54. Chuang, Y.Y. & Y.C. Huang, (2013) Molecular epidemiology of community-associated meticillin-resistant *Staphylococcus aureus* in Asia. *The Lancet. Infectious diseases* **13**: 698-708.

55. Coia, J.E., F. Thomson-Carter & D. Baird, J. Platt, (1990) Characterisation of methicillin- resistant *Staphylococcus aureus* by biotyping, immunoblotting and restriction enzyme fragment patterns. *Journal of Medical Microbiology* **31:** 125-32.

56. Collen, D., (1998) Staphylokinase: a potent, uniquely fibrin-selective thrombolytic agent. *Nature medicine* **4**: 279-284.

57. Conceicao, T., C. Coelho, I.S. Silva, H. de Lencastre & M. Aires-de-Sousa, (2015) *Staphylococcus aureus* in former Portuguese colonies from Africa and the Far East: missing data to help fill the world map. *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases* **21**: 842 e841-842 e810.

58. Cooper, J.E. & E.J. Feil, (2004) Multilocus sequence typing--what is resolved? *Trends in microbiology* **12**: 373-377.

59. Crisóstomo, M.I., H. Westh, A. Tomasz, M. Chung, D.C. Oliveira & H. de Lencastre, (2001) The evolution of methicillin resistance in *Staphylococcus aureus*: Similarity of genetic backgrounds in historically early methicillin-susceptible and -resistant isolates and contemporary epidemic clones. *Proceedings of the National Academy of Sciences of the United States of America* **98**: 9865–9870.

60. Cucarella, C., M.A. Tormo, E. Knecht, B. Amorena, I. Lasa, T.J. Foster & J.R. Penades, (2002) Expression of the biofilm-associated protein interferes with host protein receptors of *Staphylococcus aureus* and alters the infective process. *Infection and immunity* **70**: 3180-3186.

61. Cue, D., M.G. Lei & C.Y. Lee, (2012) Genetic regulation of the intercellular adhesion locus in staphylococci. *Frontiers in cellular and infection microbiology* **2**: 38.

62. Cuny, C., F. Layer, G. Werner, D. Harmsen, I. Daniels-Haardt, A. Jurke, A. Mellmann, W. Witte & R. Köck, (2015) State-wide surveillance of antibiotic resistance patterns and *spa* types of methicillin-resistant Staphylococcus aureus from blood cultures in North Rhine-Westphalia, 2011-2013. *Clinical microbiology and infection* **21**: 750-757.

63. Cuny, C. & F. Layer, (2011) Auftreten und Verbreitung von MRSA in Deutschland 2010. *Epidemiologisches Bulletin* **26**: 233-241.

64. Daum, R.S., T. Ito, K. Hiramatsu, F. Hussain, K. Mongkolrattanothai, M. Jamklang & S. Boyle-Vavra, (2002) A novel methicillin-resistance cassette in community-acquired methicillin-resistant *Staphylococcus aureus* isolates of diverse genetic backgrounds. *The Journal of infectious diseases* **186**: 1344-1347.

65. David, M.Z. & R.S. Daum, (2010) Community-associated methicillin-resistant *Staphylococcus aureus*: epidemiology and clinical consequences of an emerging epidemic. Clinical Microbiology reviews **23**: 616-687.

66. DeLeo, F.R., M. Otto, B.N. Kreiswirth & H.F. Chambers, (2010) Community-associated meticillin-resistant *Staphylococcus aureus*. *Lancet* **375**: 1557-1568.

67. Deurenberg, R.H., E. Nulens, H. Valvatne, S. Sebastian, C. Driessen, J. Craeghs, E. De Brauwer, B. Heising, Y.J. Kraat, J. Riebe, F.S. Stals, T.A. Trienekens, J. Scheres, A.W. Friedrich, F.H. van Tiel, P.S. Beisser & E.E. Stobberingh, (2009) Cross-border dissemination of methicillin-resistant Staphylococcus aureus, Euregio Meuse-Rhin

region. *Emerging Infectious Diseases* **15**: 727–734.

68. Deurenberg, R.H. & E.E. Stobberingh, (2008) The evolution of *Staphylococcus aureus*. *Infection, genetics and evolution: journal of molecular epidemiology and evolutionary genetics in infectious diseases* **8**: 747-763.

69. Dieckmann, R., R. Helmuth, M. Erhard & B. Malorny, (2008) Rapid classification and identification of salmonellae at the species and subspecies levels by whole-cell matrix-assisted laser desorption ionization-time of flight mass spectrometry. *Applied and environmental microbiology* **74**: 7767-7778.

70. Dinges, M.M., P.M. Orwin & P.M. Schlievert, (2000) Exotoxins of *Staphylococcus aureus*. *Clinical microbiology reviews* **13**: 16-34, table of contents.

71. Dingle, T.C. & S.M. Butler-Wu, (2013) Maldi-tof mass spectrometry for microorganism identification. *Clinics in laboratory medicine* **33**: 589-609.

72. Djoudi, F., C. Bonura, S. Benallaoua, A. Touati, D. Touati, A. Aleo, C. Cala, T. Fasciana & C. Mammina, (2013) Panton-Valentine leukocidin positive sequence type 80 methicillin-resistant *Staphylococcus aureus* carrying a staphylococcal cassette chromosome mec type IVc is dominant in neonates and children in an Algiers hospital. *The new microbiologica* **36**: 49-55.

73. Du, Z., R. Yang, Z. Guo, Y. Song & J. Wang, (2002) Identification of *Staphylococcus aureus* and determination of its methicillin resistance by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Analytical chemistry* **74**: 5487-5491.

74. Dubin, G., (2002) Extracellular proteases of *Staphylococcus* spp. *Biological chemistry* **383**: 1075-1086.

75. Duim, B, T.M. Wassenaar, A. Rigter & J. Wagenaar, (1999) High-resolution genotyping of Campylobacter strains isolated from poultry and humans with amplified fragment length polymorphism fingerprinting. *Applied and environmental microbiology* **65**: 2369–2375.

76. Edwards-Jones, V., M.A. Claydon, D.J. Evason, J. Walker, A.J. Fox & D.B. Gordon, (2000) Rapid discrimination between methicillin-sensitive and methicillin-resistant *Staphylococcus aureus* by intact cell mass spectrometry. *Journal of medical microbiology* **49**: 295-300.

77. Egyir, B., L. Guardabassi, M. Sørum, S.S. Nielsen, A. Kolekang, E.Frimpong, K.K. Addo,M. Jemima Newman & A.R. Larsen, (2014) Molecular epidemiology and antimicrobial susceptibility of clinical *Staphylococcus aureus* from healthcare institutions in Ghana. *PLoS One* **9**: e89716.

78. Enany, S., E. Yaoita, Y. Yoshida, M. Enany & T. Yamamoto, (2010) Molecular characterization of Panton-Valentine leukocidin-positive community-acquired methicillin-resistant *Staphylococcus aureus* isolates in Egypt. *Microbiological research* **165**: 152-162.

79. Enright, M.C., D.A. Robinson, G. Randle, E.J. Feil, H. Grundmann, & B.G. Spratt, (2002) The evolutionary history of methicillin-resistant *Staphylococcus aureus* (MRSA). *Proceedings of the National Academy of Sciences of the United States of America* **99**: 7687–7692.

80. Enright, M.C., N.P. Day, C.E. Davies, S.J. Peacock & B.G. Spratt, (2000) Multilocus sequence typing for characterization of methicillin-resistant and methicillin-

susceptible clones of *Staphylococcus aureus*. *Journal of clinical microbiology* **38**: 1008-1015.

81. Falagas, M.E., D.E. Karageorgopoulos, J. Leptidis & I.P. Korbila, (2013) MRSA in Africa: filling the global map of antimicrobial resistance. *PLoS One* **8**: e68024.

82. Fasehun, F., (1999) The antibacterial paradox: essential drugs, effectiveness, and cost. *Bulletin of the World Health Organization* **77**: 211-216.

83. Feil, E.J. & M.C. Enright, (2004) Analyses of clonality and the evolution of bacterial pathogens. *Current opinion in microbiology* **7**: 308-313.

84. Feil, E.J., J.E. Cooper, H. Grundmann, D.A. Robinson, M.C. Enright, T. Berendt, S.J. Peacock, J.M. Smith, M. Murphy, B.G. Spratt, C.E. Moore & N.P. Day, (2003) How clonal is *Staphylococcus aureus*? *Journal of bacteriology* **185**: 3307-3316.

85. Feng, Y., C.J. Chen, L.H. Su, S. Hu, J. Yu & C.H. Chiu, (2008) Evolution and pathogenesis of *Staphylococcus aureus*: lessons learned from genotyping and comparative genomics. *FEMS microbiology reviews* **32**: 23-37.

86. Fitzgerald, J.R. & J.M. Musser, (2001) Evolutionary genomics of pathogenic bacteria. *Trends in Microbiology* **9**: 547–553.

87. Foster, T.J., (2016) The remarkably multifunctional fibronectin binding proteins of *Staphylococcus aureus*. *European journal of clinical microbiology & infectious diseases : official publication of the European Society of Clinical Microbiology* **35**:1923-1931.

88. Foster, T.J., J.A. Geoghegan, V.K. Ganesh & M. Hook, (2014) Adhesion, invasion and evasion: the many functions of the surface proteins of *Staphylococcus aureus*. *Nature reviews. Microbiology* **12**: 49-62.

89. Foster, T.J., (2005) Immune evasion by staphylococci. *Nature reviews. Microbiology* **3**: 948-958.

90. Foster, T.J. & M. Hook, (1998) Surface protein adhesins of *Staphylococcus aureus*. *Trends in microbiology* **6**: 484-488.

91. Fowler, V.G., Jr., C.L. Nelson, L.M. McIntyre, B.N. Kreiswirth, A. Monk, G.L. Archer, J. Federspiel, S. Naidich, B. Remortel, T. Rude, P. Brown, L.B. Reller, G.R. Corey & S.R. Gill, (2007) Potential associations between hematogenous complications and bacterial genotype in *Staphylococcus aureus* infection. *The Journal of infectious diseases* **196**: 738-747.

92. Frenay, H.M., A.E. Bunschoten, L.M. Schouls, W.J. van Leeuwen, C.M. Vandenbroucke-Grauls, J. Verhoef & F.R. Mooi, (1996) Molecular typing of methicillin-resistant *Staphylococcus aureus* on the basis of protein A gene polymorphism. *European journal of clinical microbiology & infectious diseases: official publication of the European Society of Clinical Microbiology* **15**: 60-64.

93. Frey, B.J. & D. Dueck, (2007) Clustering by passing messages between data points. *Science* **315**: 972-97*6.*

94. Friães, A., C. Resina, V. Manuel, L. Lito,M. Ramirez & J. Melo-Cristino, (2015) Epidemiological survey of the first case of vancomycin-resistant *Staphylococcus aureus* infection in Europe. *Epidemiology and infection* **143**: 745-748.

95. Gasteiger, E., A. Gattiker, C. Hoogland, I. Ivanyi, R.D. Appel & A. Bairoch, (2003) ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic acids research* **31**: 3784-3788.

96. Gentleman, R.C., V.J. Carey, D.M. Bates, B. Bolstad, M. Dettling, S. Dudoit, B. Ellis, L. Gautier, Y. Ge, J. Gentry, K. Hornik, T. Hothorn, W. Huber, S. Iacus, R. Irizarry, F. Leisch, C. Li, M. Maechler, A.J. Rossini, G. Sawitzki, C. Smith, G. Smyth, L. Tierney, J.Y. Yang & J. Zhang, (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome biology* **5**: R80.

97. Ghasemzadeh-Moghaddam, H., E. Ghaznavi-Rad, Z. Sekawi, L. Yun-Khoon, M.N. Aziz, R.A. Hamat, D.C. Melles, A. van Belkum, M.N. Shamsudin & V. Neela, (2011) Methicillin-susceptible *Staphylococcus aureus* from clinical and community sources are genetically diverse. *International journal of medical microbiology : IJMM* **301**: 347-353.

98. Ghaznavi-Rad, E., M. Nor Shamsudin, Z. Sekawi, A. van Belkum & V. Neela, (2010) A simplified multiplex PCR assay for fast and easy discrimination of globally distributed staphylococcal cassette chromosome mec types in meticillin resistant S*taphylococcus aureus*. *Journal of Medical Microbiology*, **59**: 1135–1139.

99. Ghebremedhin, B., M.O. Olugbosi, A.M. Raji, F. Layer, R.A. Bakare, B. Konig & W. Konig, (2009) Emergence of a community-associated methicillin-resistant *Staphylococcus aureus* strain with a unique resistance profile in Southwest Nigeria. *Journal of clinical microbiology* **47**: 2975-2980.

100. Gill, S.R., D.E. Fouts, G.L. Archer, E.F. Mongodin, R.T. Deboy, J. Ravel, I.T. Paulsen, J.F. Kolonay, L. Brinkac, M. Beanan, R.J. Dodson, S.C. Daugherty, R. Madupu, S.V. Angiuoli, A.S. Durkin, D.H. Haft, J. Vamathevan, H. Khouri, T. Utterback, C. Lee, G. Dimitrov, L. Jiang, H. Qin, J. Weidman, K. Tran, K. Kang, I.R. Hance, K.E. Nelson & C.M. Fraser, (2005) Insights on evolution of virulence and resistance from the complete genome analysis of an early methicillin-resistant *Staphylococcus aureus* strain and a biofilm-producing methicillin-resistant *Staphylococcus epidermidis* strain. *Journal of bacteriology* **187**: 2426-2438.

101. Gillaspy, A.F., C.Y. Lee, S. Sau, A.L. Cheung & M.S. Smeltzer, (1998) Factors affecting the collagen binding capacity of *Staphylococcus aureus*. *Infection and immunity* **66**: 3170-3178.

102. Gillet, Y., B. Issartel, P. Vanhems, J.C. Fournet, G. Lina, M. Bes, F. Vandenesch, Y. Piemont, N. Brousse, D. Floret & J. Etienne, (2002) Association between *Staphylococcus aureus* strains carrying gene for Panton-Valentine leukocidin and highly lethal necrotising pneumonia in young immunocompetent patients. *Lancet* **359**: 753-759.

103. Goering, R.V., R.M. Shawar, N.E. Scangarella, F.P. O'Hara, H. Amrine-Madsen, J.M. West, M. Dalessandro, J.A. Becker, S.L. Walsh, L.A. Miller, S.F. van Horn, E.S. Thomas & M.E. Twynholm, (2008) Molecular epidemiology of methicillin-resistant and methicillin-susceptible *Staphylococcus aureus* isolates from global clinical trials. *Journal of clinical microbiology* **46**: 2842-2847.

104. Gordon, R.J. & F.D. Lowy, (2008) Pathogenesis of methicillin-resistant *Staphylococcus aureus* infection. *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America* **46**: 350-359.

105. GraphPad Software Inc., http://www.graphpad.com/quickcalcs/contingency1/.

106. Grumann, D., U. Nubel & B.M. Broker, (2014) *Staphylococcus aureus* toxins--their functions and genetics. *Infection, genetics and evolution: journal of molecular epidemiology and evolutionary genetics in infectious diseases* **21**: 583-592.

107. Grundmann, H., D.M. Aanensen, C.C. van den Wijngaard, B.G. Spratt, D.

Harmsen, A.W. Friedrich & G. European Staphylococcal Reference Laboratory Working, (2010) Geographic distribution of *Staphylococcus aureus* causing invasive infections in Europe: a molecular-epidemiological analysis. *PLoS medicine* **7**: e1000215.

108. Guss, B., M. Uhlen, B. Nilsson, M. Lindberg, J. Sjoquist & J. Sjodahl, (1984) Region X, the cell-wall-attachment part of staphylococcal protein A. *European journal of biochemistry / FEBS* **138**: 413-420.

109. Hallin M., A. Deplano, O. Denis, R. De Mendonça, R. De Ryck, & M. J. Struelens, (2007) Validation of Pulsed-Field Gel Electrophoresis and *spa* Typing for Long-Term, Nationwide Epidemiological Surveillance Studies of *Staphylococcus aureus* Infections. *Journal of Clinical Microbiology* **45**: 127–133.

110. Hallin, M., O. Denis, A. Deplano, R. De Ryck, S. Crevecoeur, S. Rottiers, R. de Mendonca & M.J. Struelens, (2008) Evolutionary relationships between sporadic and epidemic strains of healthcare-associated methicillin-resistant *Staphylococcus aureus*. *Clinical microbiology and infection: the official publication of the European Society of Clinical Microbiology and Infectious Diseases* **14**: 659-669.

111. Hanaki, H., K. Kuwahara-Arai, S. Boyle-Vavra, R.S. Daum, H. Labischinski & K. Hiramatsu, (1998) Activated cell-wall synthesis is associated with vancomycin resistance in methicillin-resistant *Staphylococcus aureus* clinical strains Mu3 and Mu50. *The Journal of antimicrobial chemotherapy* **42**: 199-209.

112. Harmsen, D., H. Claus, W. Witte, J. Rothganger, H. Claus, D. Turnwald & U. Vogel, (2003) Typing of methicillin-resistant *Staphylococcus aureus* in a university hospital setting by using novel software for spa repeat determination and database management. *Journal of clinical microbiology* **41**: 5442-5448.

113. Harris, S.R., E.J. Feil, M.T.G. Holden, M.A. Quail, E.K. Nickerson, N. Chantratita, S. Gardete, A. Tavares, N. Day, J.A. Lindsay, J.D. Edgeworth, H. de Lencastre, J. Parkhill, S.J. Peacock & S.D. Bentley, (2010) Evolution of MRSA during hospital transmission and intercontinental spread. *Science* **327**: 469 – 474.

114. He, L., Z. Hong-Xiang, Y. Wang, K.Y. Le, Q. Liu, J. Shang, Y. Dai, H.Meng, X. Wang, T. Li, Q. Gao, J. Qin, H. Lu, M. Otto & M. Li, (2018) Detection and analysis of Methicillin resistant human-adapted sequence type 398 allows insight into community associated methicillin-resistant *Staphylococcus aureus* evolution. *Genome Medicine* **10**: 5.

115. Herold, B.C., L.C. Immergluck, M.C. Maranan, D.S. Lauderdale, R.E. Gaskin, S. Boyle-Vavra, C.D. Leitch & R.S. Daum, (1998) Community-acquired methicillin-resistant *Staphylococcus aureus* in children with no identified predisposing risk. *JAMA* **279**: 593-598.

116. Herrmann, M., S. Abdullah, A. Alabi, P. Alonso, A.W. Friedrich, G. Fuhr, A. Germann, W.V. Kern, P.G. Kremsner, I. Mandomando, A.C. Mellmann, G. Pluschke, S. Rieg, U. Ruffing, F. Schaumburg, M. Tanner, G. Peters, H. von Briesen, C. von Eiff, L. von Muller & M.P. Grobusch, (2013a) Staphylococcal disease in Africa: another neglected 'tropical' disease. *Future microbiology* **8**: 17-26.

117. Herrmann, M., C. Petit, A. Dawson, J. Biechele, A. Halfmann, L. von Muller, S. Graber, S. Wagenpfeil, R. Klein & B. Gartner, (2013b) Methicillin-resistant *Staphylococcus aureus* in Saarland, Germany: a statewide admission prevalence screening study. *PLoS One* **8**: e73876.

118. Hiramatsu, K., N. Aritaka, H. Hanaki, S. Kawasaki, Y. Hosoda, S. Hori, Y. Fukuchi & I. Kobayashi, (1997) Dissemination in Japanese hospitals of strains of *Staphylococcus aureus* heterogeneously resistant to vancomycin. *Lancet* **350**: 1670-1673.

119. Hiramatsu, K., I. Teruyo, S. Tsubakishita, T. Sasaki, F. Takeuchi, Y. Morimoto, Y. Katayama, M. Matsuo, K. Kuwahara-Arai, T. Hishinuma & T. Baba, (2013) Genomic basis for methicillin resistance in *Staphylococcus aureus*. *Infection & Chemotherapy* **45**: 117-136.

120. Holtfreter, S., D. Grumann, V. Balau, A. Barwich, J. Kolata, A. Goehler, S. Weiss, B. Holtfreter, S.S. Bauerfeind, P. Döring, E. Friebe, N.Haasler, K. Henselin, K. Kühn, S. Nowotny, D. Radke, K.Schulz, S.R. Schulz, P. Trübe, C.H. Vu, B. Walther, S. Westphal, C.Cuny, W. Witte, H. Völzke, H.J. Grabe, T. Kocher, I. Steinmetz & B.M. Bröker, (2016) Molecular Epidemiology of *Staphylococcus aureus* in the General Population in Northeast Germany: Results of the Study of Health in Pomerania (SHIP-TREND-0). *Journal of clinical microbiology* **54**: 2774–2785.

121. Holtfreter, S., D. Grumann, M. Schmudde, H.T. Nguyen, P. Eichler, B. Strommenger, K. Kopron, J. Kolata, S. Giedrys-Kalemba, I. Steinmetz, W. Witte & B.M. Broker, (2007) Clonal distribution of superantigen genes in clinical *Staphylococcus aureus* isolates. *Journal of clinical microbiology* **45**: 2669-2680.

122. Howden, B.P., A.Y. Peleg & T.P. Stinear, (2014) The evolution of vancomycin intermediate *Staphylococcus aureus* (VISA) and heterogenous-VISA. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases* **21**: 575-582.

123. Hu, D.L., K. Omoe, F. Inoue, T. Kasai, M. Yasujima, K. Shinagawa & A. Nakane, (2008) Comparative prevalence of superantigenic toxin genes in meticillin-resistant and meticillin-susceptible *Staphylococcus aureus* isolates. *Journal of medical microbiology* **57**: 1106-1112.

124. Hu, Q., H. Peng & X. Rao, (2016) Molecular Events for Promotion of Vancomycin Resistance in Vancomycin Intermediate *Staphylococcus aureus*. *Frontiers in microbiology* **7**: 1601.

125. Huijsdens, X.W., B.J. van Dijke, E. Spalbur, M.G. van Santen-Verheuvel, M. EOC Heck, G.N. Pluister, A. Voss, W.J.B. Wannet & A.J. de Neeling, (2006) Community-acquired MRSA and pig-farming. *Annals of Clinical Microbiology and Antimicrobials* **5**: 26.

126. Huson, D.H. & D. Bryant, (2006) Application of phylogenetic networks in evolutionary studies. *Molecular biology and evolution* **23**: 254-267.

127. Hussain, M., C. von Eiff, B. Sinha, I. Joost, M. Herrmann, G. Peters & K. Becker, (2008) *eap* gene as novel target for specific identification of *Staphylococcus aureus*. *Journal of clinical microbiology* **46**: 470-476.

128. Hussain, M., K. Becker, C. von Eiff, G. Peters & M. Herrmann, (2001) Analogs of *eap* protein are conserved and prevalent in clinical *Staphylococcus aureus* isolates. *Clinical and diagnostic laboratory immunology* **8**: 1271-1276.

129. Ihaka, R. & R. Gentleman, (1996) R: A Language for Data Analysis and Graphics. *Journal of Computational and Graphical Statistics* **5**: 299-314.

130. Ito, T., K. Kuwahara-Arai, Y. Katayama, Y. Uehara, X. Han, Y. Kondo & K. Hiramatsu, (2014) Staphylococcal Cassette Chromosome *mec* (SCC*mec*) analysis of

MRSA. *Methods in Molecular Biology* **1085**: 131-148.

131. Ito, T., K. Hiramatsu, A. Tomnasz, H. de Lencastre, V. Perreten, M.T.G. Holden, D.C. Coleman, R. Goering P.M. Giffard, R.L. Skov, K. Zhang, H. Westh, F. O'Brien, F.C. Tenover, D.C. Oliveira, S. Boyle-Vavra, F. Laurent, A.M. Kearns, B. Kreiswirth, K. Soo Ko, H. Grundmann, J.E. Sollid, J.F. John, Jr., R. Daum, B. Soderquist & G. Buistx on behalf of the International Working Group on the Classification of Staphylococcal Cassette Chromosome Elements (IWG-SCC), (2012) Guidelines for Reporting Novel *mecA* Gene Homologues. *Antimicrobial agents and chemotherapy* **56**: 4997-4999.

132. Jackson, K.A., V. Edwards-Jones, C.W. Sutton & A.J. Fox, (2005) Optimisation of intact cell MALDI method for fingerprinting of methicillin-resistant *Staphylococcus aureus*. *Journal of microbiological methods* **62**: 273-284.

133. Jaton, L., T. Pillonel, K. Jaton, E. Dory, G. Prod'hom, D.S. Blanc, F. Tissot, P. Bodenmann & G. Greub, (2016) Common skin infection due to Panton–Valentine leucocidin-producing *Staphylococcus aureus* strains in asylum seekers from Eritrea: a genome-based investigation of a suspected outbreak. *Clinical Microbiology and Infection* **22**: 739.e5-8.

134. Jessen, O., K. Rosendal, P. Bulow, V. Faber & K.R. Eriksen, (1969) Changing staphylococci and staphylococcal infections. A ten-year study of bacteria and cases of bacteremia. *The New England Journal of Medicine* **281**: 627-635.

135. Jevons, M., (1961) "Celbenin" - resistant staphylococci. *British Medical Journal* **1**: 124-125.

136. Joost I., D. Blass,M. Burian, C. Goerke, C. Wolz, L. von Müller, K. Becker, K. Preissner, M. Herrmann & M. Bischoff, (2009) Transcription analysis of the extracellular adherence protein from *Staphylococcus aureus* in authentic human infection and in vitro. *The journal of infectious diseases* **199**: 1471-1478.

137. Josten, M., J. Dischinger, C. Szekat, M. Reif, N. Al-Sabti, H.G. Sahl, M. Parcina, I. Bekeredjian-Ding & G. Bierbaum, (2014) Identification of agr-positive methicillin-resistant *Staphylococcus aureus* harbouring the class A mec complex by MALDI-TOF mass spectrometry. *International journal of medical microbiology: IJMM* **304**: 1018-1023.

138. Josten, M., M. Reif, C. Szekat, N. Al-Sabti, T. Roemer, K. Sparbier, M. Kostrzewa, H. Rohde, H.G. Sahl & G. Bierbaum, (2013) Analysis of the matrix-assisted laser desorption ionization-time of flight mass spectrum of *Staphylococcus aureus* identifies mutations that allow differentiation of the main clonal lineages. *Journal of clinical microbiology* **51**: 1809-*1817.*

139. Kirkliauskienė A., A. Ambrozaitis A., R.L. Skov & N. Frimodt-Moller, (2010) The prevalence of *Staphylococcus aureus* nose and throat carriage by healthy adults. „Vis uomenės sveikata" **2**: 124-131.

140. Kluytmans, J., A. van Belkum & H. Verbrugh, (1997) Nasal carriage of *Staphylococcus aureus*: epidemiology, underlying mechanisms, and associated risks. *Clinical microbiology reviews* **10**: 505-520.

141. Köck, R., F. Schaumburg, A. Mellmann, M. Köksal, A. Jurke, K. Becker, A.W. Friedrich, (2013) Livestock-associated methicillin-resistant *staphylococcus aureus* (MRSA) as causes of human infection and colonization in Germany. *PLoS One* **8**: e55040.

Köck, R., K. Becker, B. Cookson, J.E. van Gemert-Pijnen, S. Harbarth, J. Kluytmans, M.

Mielke, G. Peters, R.L. Skov, M.J. Struelens, E. Tacconelli, A. Navarro Torne, W. Witte & A.W. Friedrich, (2010) Methicillin-resistant *Staphylococcus aureus* (MRSA): burden of disease and control challenges in Europe. *Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin* **15**: 19688.

142. Köck, R., L. Brakensiek, A. Mellmann, F. Kipp, M. Henderikx, D. Harmsen, I. Daniels-Haardt, C. von Eiff, K. Becker, M.G. Hendrix & A.W. Friedrich, (2009) Cross-border comparison of the admission prevalence and clonal structure of meticillin-resistant *Staphylococcus aureus*. *Journal of Hospital Infection* **71**: 320-326.

143. Kolawole, D.O., A. Adeyanju, F. Schaumburg, A.L. Akinyoola, O.O. Lawal, Y.B. Amusa, R. Kock & K. Becker, (2013) Characterization of colonizing *Staphylococcus aureus* isolated from surgical wards' patients in a Nigerian university hospital. *PLoS One* **8**: e68721.

144. Koren, Y., R. Bell & C. Volinsky, (2009) Matrix Factorization Techniques for Recommender Systems. *Computer* **42**: 30-37.

145. Kraef, C., A.S. Alabi, G. Peters, K. Becker, P.G. Kremsner, E.G. Rossatanga, A. Mellmann, M.P. Grobusch, P. Zanger & F. Schaumburg, (2015) Co-detection of Panton-Valentine leukocidin encoding genes and cotrimoxazole resistance in *Staphylococcus aureus* in Gabon: implications for HIV-patients' care. *Frontiers in microbiology* **6**: 60.

146. Koreen, L., S.V. Ramaswamy , E.A. Graviss , S. Naidich, J.M. Musser & B.N. Kreiswirth, (2004) *Spa* typing method for discriminating among *Staphylococcus aureus* isolates: implications for use of a single marker to detect genetic micro- and macrovariation. *Journal of clinical microbiology* **42**:792-799.

147. Krupa, P., J. Bystroń, M. Podkowik, J. Empel, A. Mroczkowska & J. Bania, (2015) Population structure and oxacillin Resistance of *Staphylococcus aureus* from pigs and pork meat in South-West of Poland. *BioMed research international* **2015**: 141475.

148. Laabei, M., M. Recker, J.K. Rudkin, M. Aldeljawi, Z. Gulay, T.J. Sloan, P. Williams, J.L. Endres, K.W. Bayles, P.D. Fey, V.K. Yajjala, T. Widhelm, E. Hawkins, K. Lewis, S. Parfett, L. Scowen, S.J. Peacock, M. Holden, D. Wilson, T.D. Read, J. van den Elsen, N.K. Priest, E.J. Feil, L.D. Hurst, E. Josefsson & R.C. Massey, (2014) Predicting the virulence of MRSA from its genome sequence. *Genome research* **24**: 839-849.

149. Lacey, K.A., J.A. Geoghegan & R.M. McLoughlin, (2016) The Role of *Staphylococcus aureus* Virulence Factors in Skin Infection and Their Potential as Vaccine Antigens. *Pathogens* **5**.

150. Lartigue, M.F., G. Hery-Arnaud, E. Haguenoer, A.S. Domelier, P.O. Schmit, N. van der Mee-Marquet, P. Lanotte, L. Mereghetti, M. Kostrzewa & R. Quentin, (2009) Identification of *Streptococcus agalactiae* isolates from various phylogenetic lineages by matrix-assisted laser desorption ionization-time of flight mass spectrometry. *Journal of clinical microbiology* **47**: 2284-2287.

151. Lasch, P., C. Fleige, M. Stammler, F. Layer, U. Nubel, W. Witte & G. Werner, (2014) Insufficient discriminatory power of MALDI-TOF mass spectrometry for typing of *Enterococcus faecium* and *Staphylococcus aureus* isolates. *Journal of microbiological methods* **100**: 58-69.

152. Lattar, S.M., L.P. Tuchscherr, R.L. Caccuri, D. Centron, K. Becker, C.A. Alonso, C. Barberis, G. Miranda, F.R. Buzzola, C. von Eiff & D.O. Sordelli, (2009) Capsule expression and genotypic differences among *Staphylococcus aureus* isolates from

patients with chronic or acute osteomyelitis. *Infection and immunity* **77**: 1968-1975.

153. Lemichez, E., M. Lecuit, X. Nassif & S. Bourdoulous, (2010) Breaking the wall: targeting of the endothelium by pathogenic bacteria. *Nature reviews. Microbiology* **8**: 93-104.

154. Lemon, K.P., V. Klepac-Ceraj, H.K. Schiffer, E.L. Brodie, S.V. Lynch & R. Kolter, (2010) Comparative analyses of the bacterial microbiota of the human nostril and oropharynx. *mBio* **1**.

155. Lenz, W., E. Eilers & U. Lehmacher, (1988) Characterization of methicillin-resistant *Staphylococcus aureus* strains isolated from 1974 to 1983 in West Germany with respect to the results of lysotyping. *Zentralblatt für Bakteriologie, Mikrobiologie und Hygiene* **268**: 277-293.

156. Leopold, S.R., R.V. Goering, A. Witten, D. Harmsen & A. Mellmann, (2014) Bacterial whole-genome sequencing revisited: portable, scalable, and standardized analysis for typing and detection of virulence and antibiotic resistance genes. *Journal of clinical microbiology* **52**: 2365-2370.

157. Limbago, B.M., A.J. Kallen, W. Zhu, P. Eggers, L.K. McDougal & V.S. Albrecht, (2014) Report of the 13th Vancomycin-resistant *Staphylococcus aureus* isolate from the United States. *Journal of clinical microbiology* **52**: 998 –1002.

158. Lina, G., Y. Piemont, F. Godail-Gamot, M. Bes, M.O. Peter, V. Gauduchon, F. Vandenesch & J. Etienne, (1999) Involvement of Panton-Valentine leukocidin-producing *Staphylococcus aureus* in primary skin infections and pneumonia. *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America* **29**: 1128-1132.

159. Lindsay, J.A. & M.T. Holden, (2004) *Staphylococcus aureus*: superbug, super genome? *Trends in microbiology* **12**: 378-385.

160. Lindsay, J.A., C.E. Moore, N.P. Day, S.J. Peacock, A.A. Witney, R.A. Stabler, S.E. Husain, P.D. Butcher & J. Hinds, (2006) Microarrays reveal that each of the ten dominant lineages of *Staphylococcus aureus* has a unique combination of surface-associated and regulatory genes. *Journal* of bacteriology **188**: 669-676.

161. Liu, Q., W.S. Yeo & T. Bae, (2016) The SaeRS Two-Component System of *Staphylococcus aureus*. *Genes* **7**.

162. Lowy, F.D., (1998) *Staphylococcus aureus* infections. *The New England journal of medicine* **339**: 520-532.

163. Lowy, F.D., (2003) Antimicrobial resistance: the example of *Staphylococcus aureus*. *The Journal of clinical investigation* **111**: 1265-1273.

164. Maiden, M.C., J.A. Bygraves, E. Feil, G. Morelli, J.E. Russell, R. Urwin, Q. Zhang, J. Zhou, K. Zurth, D.A. Caugant, I.M. Feavers, M. Achtman & B.G. Spratt, (1998) Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proceedings of the National Academy of Sciences of the United States of America* **95**: 3140-3145.

165. Manual Identibac® *S. aureus* Genotyping-Kit, (2015) Manual S. aureus Genotyping Kit 2.0. *Alere-technologies GmbH*.

166. Mariem, B.J., T. Ito, M. Zhang, J. Jin, S. Li, B.B. Ilhem, H. Adnan, X. Han & K. Hiramatsu, (2013) Molecular characterization of methicillin-resistant Panton-valentine

leukocidin positive *Staphylococcus aureus* clones disseminating in Tunisian hospitals and in the community. *BMC microbiology* **13**: 2.

167. Marples, R. R. & V.T. Rosdahl, (1997) International quality control of phage typing of staphylococcus aureus. *Journal of medical microbiology* **46:** 511-6.

168. Marques, R.C., J.V. Bernardi, J.G. Dorea, W.R. Bastos & O. Malm, (2008) Principal component analysis and discrimination of variables associated with pre- and post-natal exposure to mercury. *International journal of hygiene and environmental health* **211**: 606-614.

169. McAdow, M., D.M. Missiakas & O. Schneewind, (2012) *Staphylococcus aureus* secretes coagulase and von Willebrand factor binding protein to modify the coagulation cascade and establish host infections. *Journal of innate immunity* **4**: 141-148.

170. McDevitt, D., P. Francois, P. Vaudaux & T.J. Foster, (1994) Molecular characterization of the clumping factor (fibrinogen receptor) of *Staphylococcus aureus*. *Molecular microbiology* **11**: 237-248.

McGuinness, W.A., S.D. Kobayashi & F.R. DeLeo, (2016) Evasion of Neutrophil Killing by Staphylococcus aureus. *Pathogens* **5**.

171. Mehraj, J., M.K. Akmatov, J. Strompl, A. Gatzemeier, F. Layer, G. Werner, D.H. Pieper, E. Medina, W. Witte, F. Pessler & G. Krause, (2014) Methicillin-sensitive and methicillin-resistant *Staphylococcus aureus* nasal carriage in a random sample of non-hospitalized adult population in northern Germany. *PLoS One* **9**: e107937.

172. Melles, D.C., K.L. Taylor, A.I. Fattom & A. van Belkum, (2008) Serotyping of Dutch *Staphylococcus aureus* strains from carriage and infection. *FEMS immunology and medical microbiology* **52**: 287-292.

173. Meyer, E., C. Schröder, P. Gastmeier & C. Geffers, (2014) The reduction of nosocomial MRSA infection in Germany—an analysis of data from the *Hospital Infection Surveillance System (KISS) between 2007 and 2012. Deutsches Ärzteblatt International* **111**: 331–336.

174. Monecke, S., H. Kanig, W. Rudolph, E. Muller, G. Coombs, H. Hotzel, P. Slickers & R. Ehricht, (2010) Characterisation of Australian MRSA strains ST75- and ST883-MRSA-IV and analysis of their accessory gene regulator locus. *PLoS One* **5**: e14025.

175. Monecke, S., R. Ehricht, P. Slickers, N. Wiese & D. Jonas, (2009) Intra-strain variability of methicillin-resistant *Staphylococcus aureus* strains ST228-MRSA-I and ST5-MRSA-II. *European journal of clinical microbiology & infectious diseases: official publication of the European Society of Clinical Microbiology* **28**: 1383-1390.

176. Monecke, S., P. Slickers & R. Ehricht, (2008) Assignment of *Staphylococcus aureus* isolates to clonal complexes based on microarray analysis and pattern recognition. *FEMS immunology and medical microbiology* **53**: 237-251.

177. Monecke, S., B. Berger-Bächi, G. Coombs, A. Holmes, I. Kay,A. Kearns,H.J. Linde, F. O'Brien, P. Slickers & R. Ehricht. (2007) Comparative genomics and DNA array-based genotyping of pandemic *Staphylococcus aureus* strains encoding Panton-Valentine leukocidin. *Clinical Microbiology and Infection* **13**: 236–249.

178. Monecke, S. & R. Ehricht, (2005) Rapid genotyping of methicillin-resistant *Staphylococcus aureus* (MRSA) isolates using miniaturised oligonucleotide arrays. *Clinical microbiology and infection: the official publication of the European Society of Clinical Microbiology and Infectious Diseases* **11**: 825-833.

Moran, G.J., M.D., A. Krishnadasan, R.J. Gorwitz, G.E. Fosheim, L.K. McDougal, R.B. Carey, D.A. Talan & the EMERGEncy ID Net Study Group, (2006) Methicillin-resistant *S. aureus* infections among patients in the emergency department. *The new england journal of medicine* **355**: 666-674.

179. Morbidity and mortality weekly report, (1999) Four pediatric deaths from community-acquired methicillin-resistant *Staphylococcus aureus*—Minnesota and North Dakota. *JAMA* **282**: 1123-1125.

180. Mortimer, P. & C. Arnold C, (2001) FAFLP: last word in microbial genotyping? *Journal of Medical Microbiology* **50**: 393-5.

181. Mugalu, J., M.K. Nakakeeto, S. Kiguli & D.H. Kaddu-Mulindwa, (2006) Aetiology, risk factors and immediate outcome of bacteriologically confirmed neonatal septicaemia in Mulago hospital, Uganda. *African health sciences* **6**: 120-126.

182. Mulholland, E.K. & R.A. Adegbola, (2005) Bacterial infections-a major cause of death among children in Africa. *The New England journal of medicine* **352**: 75-77.

183. Murchan S., M.E. Kaufmann, A. Deplano, R. de Ryck, M. Struelens, C.E. Zinn, V. Fussing, S. Salmenlinna, J. Vuopio-Varkila, N. El Solh, C. Cuny, W. Witte, P.T. Tassios, N. Legakis, W. van Leeuwen, A. van Belkum, A. Vindel, I. Laconcha, J. Garaizar, S. Haeggman, B. Olsson-Liljequist, U. Ransjo, G. Coombes & B. Cookson, (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. *Journal of clinical microbiology* **41**: 1574-1585.

184. Naimi, T.S., K.H. LeDell, K. Como-Sabetti, S.M. Borchardt, D.J. Boxrud, J. Etienne, S.K. Johnson, F. Vandenesch, S. Fridkin, C. O'Boyle, R.N. Danila & R. Lynfield, (2003) Comparison of community- and health care-associated methicillin-resistant *Staphylococcus aureus* infection. *JAMA* **290**: 2976-2984.

185. Nelson J.L., K. C. Rice, S. R. Slater, P. M. Fox, G.L. Archer, K. W. Bayles, P.D. Fey, B.N. Kreiswirth & G.A. Somerville, (2007) Vancomycin-Intermediate *Staphylococcus aureus* Strains Have Impaired Acetate Catabolism: Implications for Polysaccharide Intercellular Adhesin Synthesis and Autolysis. *Antimicrobial agents and Chemotherapy* **51**: 616–622

186. Nethercott C., A. N. Mabbett, M. Totsika, P. Peters, J.C. Ortiz, G. R. Nimmo, G.W. Coombs, M.J. Walker & M.A. Schembria, (2013) Molecular Characterization of Endocarditis-Associated *Staphylococcus aureus*. *Journal of clinical microbiology* **51**: 2131–2138.

187. Ni Eidhin, D., S. Perkins, P. Francois, P. Vaudaux, M. Hook & T.J. Foster, (1998) Clumping factor B (ClfB), a new surface-located fibrinogen-binding adhesin of *Staphylococcus aureus*. *Molecular microbiology* **30**: 245-257.

188. Nienaber, J.J., B.K. Sharma Kuinkel, M. Clarke-Pearson, S. Lamlertthon, L. Park, T.H. Rude, S. Barriere, C.W. Woods, V.H. Chu, M. Marin, S. Bukovski, P. Garcia, G.R. Corey, T. Korman, T. Doco-Lecompte, D.R. Murdoch, L.B. Reller, V.G. Fowler, Jr. & I. International Collaboration on Endocarditis-Microbiology, (2011) Methicillin-susceptible *Staphylococcus aureus* endocarditis isolates are associated with clonal complex 30 genotype and a distinct repertoire of enterotoxins and adhesins. *The Journal of infectious diseases* **204**: 704-713.

189. Nishifuji, K., M. Sugai & M. Amagai, (2008) Staphylococcal exfoliative toxins: "molecular scissors" of bacteria that attack the cutaneous defense barrier in mammals. *Journal of dermatological science* **49**: 21-31.

190. Novick, R.P., (2003) Autoinduction and signal transduction in the regulation of staphylococcal virulence. *Molecular microbiology* **48**: 1429-1449.

191. Nübel, U., B. Strommenger, F. Layer & W. Witte, (2011) From types to trees: Reconstructing the spatial spread of *Staphylococcus aureus* based on DNA variation. *International Journal of Medical Microbiology: IJMM* **301**: 614– 618.

192. Nübel ,U.,P. Roumagnacc, M. Feldkampc, J. Songe, K.S.Kof, Y. Huangg, G. Coombsh, M.I.H. Westh, R. Skov, M.J. Struelensm, R.V. Goering, B. Strommenger, A. Weller, W. Witte & M. Achtman, (2008) Frequent emergence and limited geographic dispersal of methicillin-resistant *Staphylococcus aureus*. *Proceedings of the National Academy of Sciences* **105**: 14130 –14135.

193. Nulens, E., E.E. Stobberingh, H. van Dessel, S. Sebastian, F.H. van Tiel, P.S. Beisser, R.H. Deurenberg, (2008) Molecular characterization of *Staphylococcus aureus* bloodstream isolates collected in a Dutch University Hospital between 1999 and 2006. *Journal of clinical microbiology* **46**: 2438–2441.

194. O'Riordan, K. & J.C. Lee, (2004) *Staphylococcus aureus* capsular polysaccharides. *Clinical microbiology reviews* **17**: 218-234.

195. Obaro, S., L. Lawson, U. Essen, K. Ibrahim, K. Brooks, A. Otuneye, D. Shetima, P. Ahmed, T. Ajose, M. Olugbile, D. Idiong, D. Ogundeji, C. Ochigbo, G. Olanipekun, W. Khalife & R. Adegbola, (2011) Community acquired bacteremia in young children from central Nigeria-a pilot study. *BMC infectious diseases* **11**: 137.

196. Ogston, A., (1882) *Micrococcus* Poisoning. *Journal of anatomy and physiology* **17**: 24-58.

197. Okon, K.O., P. Basset, A. Uba, J. Lin, B. Oyawoye, A.O. Shittu & D.S. Blanc, (2009) Cooccurrence of predominant Panton-Valentine leukocidin-positive sequence type (ST) 152 and multidrug-resistant ST 241 *Staphylococcus aureus* clones in Nigerian hospitals. *Journal of clinical microbiology* **47**: 3000-3003.

198. Oosthuysen, W.F., H. Orth, C.J. Lombard, B. Sinha & E. Wasserman, (2014) Population structure analyses of *Staphylococcus aureus* at Tygerberg Hospital, South Africa, reveals a diverse population, a high prevalence of Panton-Valentine leukocidin genes, and unique local methicillin-resistant *S. aureus* clones. *Clinical microbiology and infection: the official publication of the European Society of Clinical Microbiology and Infectious Diseases* **20**: 652-659.

199. Ostergaard, C., S.G. Hansen & J.K. Moller, (2015) Rapid first-line discrimination of methicillin-resistant *Staphylococcus aureus* strains using MALDI-TOF MS. *International journal of medical microbiology: IJMM* **305**: 838-847.

200. Otter, J.A. & G.L. French, (2012) Community-associated meticillin-resistant *Staphylococcus aureus*: the case for a genotypic definition. *Journal of Hospital Infection* **81**: 143e148.

201. Otter, J.A. & G.L. French, (2010) Molecular epidemiology of community-associated meticillin-resistant *Staphylococcus aureus* in Europe. *The Lancet. Infectious diseases* **10**: 227-239.

202. Otto, M., (2008) Staphylococcal biofilms. *Current topics in microbiology and*

immunology **322**: 207-228.

203. Parker, M.T. & J.H. Hewitt, (1970) Methicillin resistance in *Staphylococcus aureus. Lancet* **295**: 800-804.

204. Phaku, P., M. Lebughe, L. Strauss, G. Peters, M. Herrmann, D. Mumba, A. Mellmann, J.J. Muyembe-Tamfum & F. Schaumburg, (2016) Unveiling the molecular basis of antimicrobial resistance in *Staphylococcus aureus* from the Democratic Republic of the Congo using whole genome sequencing. *Clinical microbiology ancd infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases* **22**: 644 e641-645.

205. Popovic, M., D. Steinort, S. Pillai & C. Joukhadar (2010) Fosfomycin: an old, new friend? *European journal of clinical microbiology & infectious diseases : official publication of the European Society of Clinical Microbiology* **29**: 127-142.

206. Prévost G, B. Pottecher, M. Dahlet, M. Bientz, J.M. Mantz & Y. Piémont, (1991) Pulsed field gel electrophoresis as a new epidemiological tool for monitoring methicillin-resistant *Staphylococcus aureus* in an intensive care unit. *Journal of Hospital Infection* **17**: 255-269.

207. Price, J., N.C. Gordon, D. Crook, M. Llewelyn & J. Paul, (2013) The usefulness of whole genome sequencing in the management of *Staphylococcus aureus* infections. *Clinical microbiology and infection: the official publication of the European Society of Clinical Microbiology and Infectious Diseases* **19**: 784-789.

208. Pryer, J.A., R. Nichols, P. Elliott, B. Thakrar, E. Brunner & M. Marmot, (2001) Dietary patterns among a national random sample of British adults. *Journal of epidemiology and community health* **55**: 29-37.

209. Quackenbush, J., (2001) Computational analysis of microarray data. *Nature reviews. Genetics* **2**: 418-427.

210. R Development Core Team, (2008) R: A language and environment for statistical computing. *R Foundation for Statistical Computing, http://R-project.org*.

211. Ramdani-Bouguessa, N., M. Bes, H. Meugnier, F. Forey, M.E. Reverdy, G. Lina, F. Vandenesch, M. Tazir & J. Etienne, (2006) Detection of methicillin-resistant *Staphylococcus aureus* strains resistant to multiple antibiotics and carrying the Panton-Valentine leukocidin genes in an Algiers hospital. *Antimicrobial agents and chemotherapy* **50**: 1083-1085.

212. Rammelkamp, Ch. & T. Maxon, (1942) Resistance of *Staphylococcus aureus* to the action of penicillin. *Experimental Biology and Medicine* **51**(3): 386-389.

213. Rasmussen, G., S. Monecke, O. Brus, R. Ehricht & B. Soderquist, (2014) Long term molecular epidemiology of methicillin-susceptible *Staphylococcus aureus* bacteremia isolates in Sweden. *PLoS One* **9**: e114276.

214. Redpath, M.B., T.J. Foster & C.J. Bailey, (1991) The role of the serine protease active site in the mode of action of epidermolytic toxin of *Staphylococcus aureus*. *FEMS microbiology letters* **65**: 151-155.

215. Rijnders M.I.A., R.H. Deurenberg, M.L.L. Boumans, J.A.A. Hoogkamp-Korstanje, P.S. Beisser, the Antibiotic Resistance Surveillance Group & E.E. Stobberingh, (2009) Population structure of *Staphylococcus aureus* strains isolated from intensive care unit patients in the Netherlands over an 11-Year Period (1996 to 2006). *Journal of clinical microbiology*, **47**: 4090–4095.

216. Robert Koch Institut, (2018) Ständige Impfkommission: Empfehlungen der Ständigen Impfkommission (STIKO) am Robert Koch-Institut. *Epidemiologisches Bulletin* **34**: 335 – 382.

217. Robert Koch Institut, (1997) MRSA mit nur noch intermediärer Glykopeptidempfindlichkeit in Japan und in den USA. *Epidemiologisches Bulletin* **45**: 314-315.

218. Robinson, D. A. & M.C. Enright, (2003) Evolutionary Models of the Emergence of Methicillin-Resistant *Staphylococcus aureus*. *Antimicrobial agents and chemotherapy* **47**: 3926-3934.

219. Rolo, J., P. Worning, J.B. Nielsen, R. Bowden, O. Bouchami, P. Damborg, L. Guardabassi, V. Perreten, A. Tomasz, H. Westh, H. de Lencastre & M. Miragaia, (2017) Evolutionary origin of the staphylococcal cassette chromosome *mec* (SCC*mec*). *Antimicrobial agents and chemotherapy* **61**: e02302- 16. 208.

220. Rosenbach, F.J., (1884). Mikro-Organismen bei den. Wund-Infections-Krankheiten des Menschen. JF Bergmann's Verlag, Wiesbaden, **18**: 1–122.

221. Rountree, P.M. & M.A. Beard, (1968) Hospital strains of *Staphylococcus aureus* with particular reference to methicillin-resistant strains. M*edical Journal of Australia* **2:**1163–1168.

222. Ruffing, U., R. Akulenko, M. Bischoff, V. Helms, M. Herrmann & L. von Muller, (2012) Matched-cohort DNA microarray diversity analysis of methicillin sensitive and methicillin-resistant *Staphylococcus aureus* isolates from hospital admission patients. *PLoS One* **7**: e52487.

223. Ruffing, U., A. Alabi, T. Kazimoto, D.C. Vubil, R. Akulenko, S. Abdulla, P. Alonso, M. Bischoff, A. Germann, M.P. Grobusch, V. Helms, J. Hoffmann, W.V. Kern, P.G. Kremsner, I. Mandomando, A. Mellmann, G. Peters, F. Schaumburg, S. Schubert, L. Strauß, M. Tanner, H. von Briesen, L. Wende, L. von Muller & M. Herrmann, (2017) community-associated *Staphylococcus aureus* from Sub-Saharan Africa and Germany: a cross-sectional geographic correlation Study. *Scientific Reports* **7**: 154.

224. Ruimy, R., A. Maiga, L. Armand-Lefevre, I. Maiga, A. Diallo, A.K. Koumare, K. Ouattara, S. Soumare, K. Gaillard, J.C. Lucet, A. Andremont & E.J. Feil, (2008) The carriage population of *Staphylococcus aureus* from Mali is composed of a combination of pandemic clones and the divergent Panton-Valentine leukocidin-positive genotype ST152. *Journal of bacteriology* **190**: 3962-3968.

225. Sabat, A.J., A. Budimir, D. Nashev, R. Sa-Leao, J. van Dijl, F. Laurent, H. Grundmann, A.W. Friedrich & E.S.G.o.E. Markers, (2013) Overview of molecular typing methods for outbreak detection and epidemiological surveillance. *Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin* **18**: 20380.

226. Saitou, N. & M. Nei, (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular biology and evolution* **4**: 406-425.

227. Saravolatz, L.D., D.J. Pohlod & L.M. Arking, (1982) Community-acquired methicillin-resistant *Staphylococcus aureus* infections: a new source for nosocomial outbreaks. *Annals of Internal Medicine* **97**: 325–329.

228. Sauget, M., N. van der Mee-Marquet, X. Bertrand & D. Hocquet, (2016) Matrix-assisted laser desorption ionization-time of flight Mass spectrometry can detect

Staphylococcus aureus clonal complex 398. *Journal of microbiological methods* **127**: 20-23.

229. Schaumburg, F., A.S. Alabi, G. Peters & K. Becker, (2014) New epidemiology of *Staphylococcus aureus* infection in Africa. *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases* **20**: 589-596.

230. Schaumburg, F., R. Köck, A. Mellmann, L. Richter, F. Hasenberg, A. Kriegeskorte, A.W. Friedrich, S. Gatermann, G. Peters, C. von Eiff, K. Becker & study group, (2012) Population Dynamics among methicillin-tesistant *Staphylococcus aureus i*solates in Germany during a 6-year period. *Journal of clinical microbiology* **50**:3186-3192.

231. Schaumburg, F., R. Köck,, A.W. Friedrich, S. Soulanoudjingar, U.A. Ngoa, C. von Eiff, S. Issifou, P.G. Kremsner, M. Herrmann, G. Peters & K. Becker, (2011) Population structure of *Staphylococcus aureus* from remote African Babongo Pygmies. *PLoS neglected tropical diseases* **5**: e1150.

232. Seibold, E., T. Maier, M. Kostrzewa, E. Zeman & W. Splettstoesser, (2010) Identification of *Francisella tularensis* by whole-cell matrix-assisted laser desorption ionization-time of flight mass spectrometry: fast, reliable, robust, and cost-effective differentiation on species and subspecies levels. *Journal of clinical microbiology* **48**: 1061-1069.

233. Seng, P., M. Drancourt, F. Gouriet, B. La Scola, P.E. Fournier, J.M. Rolain & D. Raoult, (2009) Ongoing revolution in bacteriology: routine identification of bacteria by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* **49**: 543-551.

234. Shallcross, L.J., E. Fragaszy, A.M. Johnson & A.C. Hayward, (2013) The role of the Panton-Valentine leucocidin toxin in staphylococcal disease: a systematic review and meta-analysis. *The Lancet. Infectious diseases* **13**: 43-54.

235. Shanson, D.C., (1981) Antibiotic resistant *Staphylococcus aureus*. *Journal of Hospital Infection* **2**: 11-36.

236. Shittu, A.O., O. Oyedara, K. Okon, A. Raji, G. Peters, L. von Muller, F. Schaumburg, M. Herrmann & U. Ruffing, (2015) An assessment on DNA microarray and sequence-based methods for the characterization of methicillin-susceptible *Staphylococcus aureus* from Nigeria. *Frontiers in microbiology* **6**: 1160.

237. Shopsin, B., M. Gomez, S.O. Montgomery, D.H. Smith, M. Waddington, D.E. Dodge, D.A. Bost, M. Riehman, S. Naidich & B.N. Kreiswirth, (1999) Evaluation of protein A gene polymorphic region DNA sequencing for typing of *Staphylococcus aureus* strains. *Journal of clinical microbiology* **37**: 3556-3563.

238. Sigauque, B., A. Roca, I. Mandomando, L. Morais, L. Quinto, J. Sacarlal, E. Macete, T. Nhamposa, S. Machevo, P. Aide, Q. Bassat, A. Bardaji, D. Nhalungo, M. Soriano-Gabarro, B. Flannery, C. Menendez, M.M. Levine & P.L. Alonso, (2009) Community-acquired bacteremia among children admitted to a rural hospital in Mozambique. *The Pediatric infectious disease journal* **28**: 108-113.

239. Sina, H., T.A. Ahoyo, W. Moussaoui, D. Keller, H.S. Bankole, Y. Barogui, Y. Stienstra, S.O. Kotchoni, G. Prevost & L. Baba-Moussa, (2013) Variability of antibiotic susceptibility and toxin production of *Staphylococcus aureus* strains isolated from skin,

soft tissue, and bone related infections. *BMC microbiology* **13**: 188.

240. Smith, T.L., M.L. Pearson, K.R. Wilcox, C. Cruz, M.V. Lancaster, B. Robinson-Dunn, F.C. Tenover, M.J. Zervos, J.D. Band, E. White & W.R. Jarvis, (1999) Emergence of vancomycin resistance in *Staphylococcus aureus*. Glycopeptide-Intermediate *Staphylococcus aureus* Working Group. *The New England journal of medicine* **340**: 493-501.

241. Sogawa, K., M. Watanabe, T. Ishige, S. Segawa, A. Miyabe, S. Murata, T. Saito, A. Sanda, K. Furuhata & F. Nomura, (2017) Rapid discrimination between methicillin-sensitive and methicillin-resistant *Staphylococcus aureus* using MALDI-TOF mass spectrometry. *Biocontrol Science* **22**: 163-169.

242. Somerville, G.A. & R.A. Proctor, (2009) At the crossroads of bacterial metabolism and virulence factor synthesis in staphylococci. *Microbiology and molecular biology reviews: MMBR* **73**: 233-248.

243. Sparbier, K., C. Lange, J. Jung, A. Wieser, S. Schubert & M. Kostrzewa, (2013) MALDI biotyper-based rapid resistance detection by stable-isotope labeling. *Journal of clinical microbiology* **51**: 3741-3748.

244. Spaulding, A.R., W. Salgado-Pabon, P.L. Kohler, A.R. Horswill, D.Y. Leung & P.M. Schlievert, (2013) Staphylococcal and streptococcal superantigen exotoxins. *Clinical microbiology reviews* **26**: 422-447.

245. Stegger, M., T. Wirth, P.S. Andersen, R.L. Skov, A. De Grassi, P. Martins Simões, A. Tristan, A. Petersen, M. Aziz, K. Kiil, I. Cirković, E.E. Udo, R. del Campo, J. Vuopio-Varkila, N. Ahmad, S. Tokajian, G. Peters, F. Schaumburg, B.Olsson-Liljequist, M.Givskov, E.E. Driebe, H.E. Vigh, A. Shittu, N. Ramdani-Bougessa, J.-P. Rasigade, L.B. Price, F. Vandenesch, A.R. Larsen & F. Laurent, (2014) Origin and evolution of european community-acquired methicillin-resistant *Staphylococcus aureus. mBio* **5**: e01044-14.

246. Strauss, L., U. Ruffing, S. Abdulla, A. Alabi, R. Akulenko, M. Garrine, A. Germann, M.P. Grobusch, V. Helms, M. Herrmann, T. Kazimoto, W. Kern, I. Mandomando, G. Peters, F. Schaumburg, L. von Muller & A. Mellmann, (2016) Detecting *Staphylococcus aureus* Virulence and Resistance Genes - a Comparison of Whole Genome Sequencing and DNA Microarray Technology. *Journal of clinical microbiology*.

247. Strommenger, B., C. Kettlitz, T. Weniger, D. Harmsen, A.W. Friedrich & W. Witte, (2006) Assignment of *Staphylococcus* isolates to groups by *spa* typing, SmaI macrorestriction analysis, and multilocus sequence typing. *Journal of clinical microbiology* **44**: 2533-2540.

248. Szabados, F., M. Kaase, A. Anders & S.G. Gatermann, (2012) Identical MALDI TOF MS-derived peak profiles in a pair of isogenic SCC*mec*-harboring and SCC*mec*-lacking strains of *Staphylococcus aureus*. *The Journal of infection* **65**: 400-405.

249. Tacconelli, E., G. De Angelis, C. de Waure C, M. A. Cataldo, G. La Torre, R. Cauda, (2009) Rapid screening tests for meticillin-resistant *Staphylococcus aureus* at hospital admission: systematic review and meta-analysis. *The Lancet. Infectious diseases* **9**: 546-554.

250. Taglialegna, A., S. Navarro, S. Ventura, J.A. Garnett, S. Matthews, J.R. Penades, I. Lasa & J. Valle, (2016) Staphylococcal Bap Proteins Build Amyloid Scaffold Biofilm Matrices in Response to Environmental Signals. *PLoS pathogens* **12**: e1005711.

251. Takano, T., W. Higuchi, H. Zaraket, T. Otsuka, T. Baranovich, S. Enany, K. Saito,

H. Isobe, S. Dohmae, K. Ozaki, M. Takano, Y. Iwao, M. Shibuya, T. Okubo, S. Yabe, D. Shi, I. Reva, L.J. Teng & T. Yamamoto, (2008) Novel characteristics of community-acquired methicillin-resistant *Staphylococcus aureus* strains belonging to multilocus sequence type 59 in Taiwan. *Antimicrobial agents and chemotherapy* **52**: 837-845.

252. Takeuchi, F., S. Watanabe, T. Baba, H. Yuzawa, T. Ito, Y. Morimoto, M. Kuroda, L. Cui, M. Takahashi, A. Ankai, S. Baba, S. Fukui, J.C. Lee & K. Hiramatsu, (2005) Whole-genome sequencing of *Staphylococcus haemolyticus* uncovers the extreme plasticity of its genome and the evolution of human-colonizing staphylococcal species. *Journal of bacteriology* **187**: 7292-7308.

253. Talbert, A.W., M. Mwaniki, S. Mwarumba, C.R. Newton & J.A. Berkley, (2010) Invasive bacterial infections in neonates and young infants born outside hospital admitted to a rural hospital in Kenya. *The Pediatric infectious disease journal* **29**: 945-949.

254. Tristan, A., J.P. Rasigade, E. Ruizendaal, F. Laurent, M. Bes, H. Meugnier, G. Lina, J. Etienne, M. Celard, P. Tattevin, S. Monecke, V. Le Moing, F. Vandenesch & A.s.G.o.I.E. French, (2012) Rise of CC398 lineage of *Staphylococcus aureus* among Infective endocarditis isolates revealed by two consecutive population-based studies in France. *PLoS One* **7**: e51172.

255. Tsubakishita, S., K. Kuwahara-Arai, T. Sasaki & K. Hiramatsu, (2010) Origin and molecular evolution of the determinant of methicillin resistance in staphylococci. *Antimicrobial agents and chemotherap* **54**: 4352-4359.

256. Udo, E.E., M.A. Al-Bustan, L.E. Jacob & T.D. Chugh, (1999) Enterotoxin production by coagulase-negative staphylococci in restaurant workers from Kuwait City may be a potential cause of food poisoning. *Journal of Medical Microbioogy* **48**: 819–823.

257. Udo, E.E, J.W. Pearman & W.B. Grubb; (1993) Genetic analysis of community isolates of methicillin-resistant Staphylococcus aureus in Western Australia. *The Journal of Hospital Infection* **25**: 97–108.

258. Ueda, O., S. Tanaka, Z. Nagasawa, H. Hanaki, T. Shobuike & H. Miyamoto, (2015) Development of a novel matrix-assisted laser desorption/ionization time-of-flight mass spectrum (MALDI-TOF-MS)-based typing method to identify meticillin-resistant *Staphylococcus aureus* clones. *The Journal of hospital infection* **90**: 147-155.

259. Uhlen, M., B. Guss, B. Nilsson, S. Gatenbeck, L. Philipson & M. Lindberg, (1984) Complete sequence of the staphylococcal gene encoding protein A. A gene evolved through multiple duplications. *The Journal of biological chemistry* **259**: 1695-1702.

260. Urwin, R. & M.C. Maiden, (2003) Multi-Locus sequence typing: a tool for global epidemiology. *Trends in Microbiology,* **10**: 479-487.

261. van Belkum, A., D.C. Melles, J. Nouwen, W.B. van Leeuwen, W. van Wamel, M.C. Vos, H.F. Wertheim & H.A. Verbrugh, (2009) Co-evolutionary aspects of human colonisation and infection by *Staphylococcus aureus*. *Infection, genetics and evolution: journal of molecular epidemiology and evolutionary genetics in infectious diseases* **9**: 32-47.

262. van Belkum A., W. van Leeuwen, M.E. Kaufmann, B. Cookson, F. Forey, J. Etienne, R. Goering, F. Tenover, C. Steward, F. O'Brien, W. Grubb, P. Tassios, N. Legakis, A. Morvan, N. El Solh, R. de Ryck, M. Struelens, S. Salmenlinha, J. Vuopio-Varkila, M. Kooistra, A. Talens, W. Witte & H. Verbrugh, (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. *Journal of clinical microbiology* **36**: 1653-1659.

263. van Cleef, B.A.G.L., D. L. Monnet, A. Voss, K. Krziwanek, F. Allerberger, M. Struelens, H. Zemlickova, R.L. Skov, J. Vuopio-Varkila, C. Cuny, A.W. Friedrich, I. Spiliopoulou, J. Pászti, H. Hardardottir, A. Rossney, A. Pan, A. Pantosti, M. Borg, H. Grundmann, M. Mueller-Premru, B. Olsson-Liljequist, A. Widmer, S. Harbarth, A. Schweiger, S. Unal, & J.A.J.W. Kluytmans, (2011) Livestock associated Methicillin-Resistant *Staphylococcus aureu*s in Humans, Europe. *Emerging Infectious Diseases* **17**: 502-505.

264. van der Meeren, B.T., P.S. Millard, M. Scacchetti, M.H. Hermans, M. Hilbink, T.B. Concelho, J.J. Ferro & P.C. Wever, (2014) Emergence of methicillin resistance and Panton-Valentine leukocidin positivity in hospital- and community-acquired *Staphylococcus aureus* infections in Beira, Mozambique. *Tropical medicine & international health: TM & IH* **19**: 169-176.

265. van Rensburg, B.W., A.M. van Staden, G.J. Rossouw & G. Joubert, (2010) The profile of adult nephrology patients admitted to the Renal Unit of the Universitas Tertiary Hospital in Bloemfontein, South Africa from 1997 to 2006. *Nephrology, dialysis, transplantation: official publication of the European Dialysis and Transplant Association - European Renal Association* **25**: 820-824.

266. Vandenesch, F., G. Lina & T. Henry, (2012) *Staphylococcus aureus* hemolysins, bi-component leukocidins, and cytolytic peptides: a redundant arsenal of membrane-damaging virulence factors? *Frontiers in cellular and infection microbiology* **2**: 12.

267. Vandenesch, F., T. Naimi, M.C. Enright, G. Lina, G.R. Nimmo, H. Heffernan, N. Liassine, M. Bes, T. Greenland, M.E. Reverdy & J. Etienne, (2003) Community-acquired methicillin-resistant *Staphylococcus aureus* carrying Panton-Valentine leuko-cidin genes: worldwide emergence. *Emerging Infectious Diseases* **9**: 978–984.

268. Vazquez, V., X. Liang, J.K. Horndahl, V.K. Ganesh, E. Smeds, T.J. Foster & M. Hook, (2011) Fibrinogen is a ligand for the *Staphylococcus aureus* microbial surface components recognizing adhesive matrix molecules (MSCRAMM) bone sialoprotein-binding protein (Bbp). *The Journal of biological chemistry* **286**: 29797-29805.

269. von Eiff, C., A.W. Friedrich, G. Peters & K. Becker, (2004) Prevalence of genes encoding for members of the staphylococcal leukotoxin family among clinical isolates of *Staphylococcus aureus*. *Diagnostic microbiology and infectious disease* **49**: 157-162.

270. Vos, P., R. Hogers, M. Bleeker, M. Reijans, T. van de Lee, M- Hornes, A. Frijters, J. Pot, J. Peleman, M.Kuiper & M. Zabeau, (1995) AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research* **23**: 4407-14.

271. Vranckx K., K. De Bruyne & B. Pot, (2017) Analysis of MALDI-TOF MS Spectra using the BioNumerics Software, *Wiley online library*, 10.1002/9781118960226.ch21, Editors: H.N. Shah, Saheer E. Gharbia.

272. Wang, Y.R., Q. Chen, S.H. Cui & F.Q. Li, (2013) Characterization of *Staphylococcus aureus* isolated from clinical specimens by matrix assisted laser desorption/ionization time-of-flight mass spectrometry. *Biomedical and environmental sciences: BES* **26**: 430-436.

273. Weese J.S., M. Archambault, B.M. Willey, H. Dick, P. Hearn, B.N. Kreiswirth, B. Said-Salim, A. McGeer, Y. Likhoshvay, J.F. Prescott & D.E. Low, (2005) Methicillin-

resistant *Staphylococcus aureus* in horses and horse personnel, 2000–2002. *Emerging Infectious Diseases* **11**: 430-435.

274. Weller, T.M., (2000) Methicillin-resistant *Staphylococcus aureus* typing methods: which should be the international standard? *The Journal of hospital infection* **44**: 160-172.

275. Wertheim, H.F., M.C. Vos, A. Ott, A. van Belkum, A. Voss, J.A. Kluytmans, P.H. van Keulen, C.M. Vandenbroucke-Grauls, M.H. Meester & H.A. Verbrugh, (2004) Risk and outcome of nosocomial *Staphylococcus aureus* bacteraemia in nasal carriers versus non-carriers. *Lancet* **364**: 703-705.

276. Williamson, Y.M., H. Moura, A.R. Woolfitt, J.L. Pirkle, J.R. Barr, G. Carvalho Mda, E.P. Ades, G.M. Carlone & J.S. Sampson, (2008) Differentiation of *Streptococcus pneumoniae* conjunctivitis outbreak isolates by matrix-assisted laser desorption ionization-time of flight mass spectrometry. *Applied and environmental microbiology* **74**: 5891-5897.

277. Witte, W., (2009) Community-acquired methicillin-resistant *Staphylococcus aureus*: what do we need to know? *Clinical microbiology and infection: the official publication of the European Society of Clinical Microbiology and Infectious Diseases* **15**: 17-25.

278. Witte, W., I. Klare, U. Nübel, B. Strommenger & G. Werner, (2008) Emergence and spread of antibiotic resistant Gram positive bacterial pathogens. *International journal of medical microbiology: IJMM* **298**: 365–377.

279. Witte, W., V.D. Nguyen & K. Dunnhaupt, (1986) Methicillin-resistant *Staphylococcus aureus* (MRSA) in the German Democratic Republic. Incidence and strain characteristics. *Journal of hygiene, epidemiology, microbiology, and immunology* **30**: 185-193.

280. Wolters, M., H. Rohde, T. Maier, C. Belmar-Campos, G. Franke, S. Scherpe, M. Aepfelbacher & M. Christner, (2011) MALDI-TOF MS fingerprinting allows for discrimination of major methicillin-resistant *Staphylococcus aureus* lineages. *International journal of medical microbiology: IJMM* **301**: 64-68.

281. Wu S., C. Pscitelli, H. de Lencastre & A. Tomasz, (1996) Tracking the evolutionary origin of the methicillin resistance gene: cloning and sequencing of a homologue of *mecA* from a methicillin susceptible strain of S*taphylococcus sciuri*. *Microbial drug resistance* **2**:435-441.

282. Zanger, P., D. Nurjadi, R. Schleucher, H. Scherbaum, C. Wolz, P.G. Kremsner & B. Schulte, (2012) Import and spread of Panton-Valentine Leukocidin-positive *Staphylococcus aureus* through nasal carriage and skin infections in travelers returning from the tropics and subtropics. *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America* **54**: 483-492.

283. Zhang, T., J. Ding, X. Rao, J. Yu, M. Chu, W. Ren, L. Wang & W. Xue, (2015) Analysis of methicillin-resistant *Staphylococcus aureus* major clonal lineages by Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry (MALDI-TOF MS). *Journal of microbiological methods* **117**: 122-127.

284. Zhao, S., S.E. Mitchell, J. Meng, S. Kresovich, M.P. Doyle, R.E. Dean, , A. M. Casa & J.W. Weller, (2000) Genomic typing of Escherichia coli O157:H7 by semi-automated fluorescent AFLP analysis. *Microbes and Infection* **2**: 107-13.

8 Publications / congress contributions

Publications:

- **Ruffing U**, Alabi A, Kazimoto T, Vubil D, Akulenko R, Abdulla S, Alonso P, Bischoff M, Germann A, Grobusch M, Helms V, Hoffmann J, Kern W, Kremsner P, Mandomando I, Mellmann A, Peters G, Schaumburg F, Schubert S, Strauß L, Tanner M, von Briesen H, Wende L, von Müller L &Herrmann M. Community-associated *Staphylococcus aureus* from Sub-Saharan Africa and Germany: A cross-sectional geographic correlation study. **Sci Rep**. 2017 Mar; 7(1):154.
- Strauß L, **Ruffing U**, Abdulla S, Alabi A, Akulenko R, Garrine M, Germann A, Grobusch M, Helms V, Herrmann M, Kazimoto T, Kern W, Mandomando I, Peters G, Schaumburg F, von Müller L, Mellmann A. Detecting *Staphylococcus aureus* virulence and resistance genes – a comparison of whole genome sequencing and DNA microarray technology. **J Clin Microbiol**. 2016 April; 54(4): 1008-1116.
- Shittu AO, Oyedara O, Okon K, Raji A, Peters G, von Müller L, Schaumburg, Herrmann M and **Ruffing U.** An assessment on DNA microarray and sequence-based methods for the characterization of methicillin-susceptible *Staphylococcus aureus* from Nigeria. **Front Microbiol**. 2015 Oct; 6: 1160.
- **Ruffing U,** Akulenko R, Bischoff M, Helms V, Herrmann M, von Müller L. Matched-cohort DNA microarray diversity analysis of methicillin-sensitive and methicillin-resistant *Staphylococcus aureus* isolates from hospital admission patients. **PLoS One**. 2012; 7(12):e52487.

Congress presentations:

- **Ruffing U**, Genomic Comparison of a Large Cohort of Non-Nosocomial Commensal and Clinical *Staphylococcus aureus* of African and German origin. 17th International Symposium on Staphylococci and Staphylococcal Infections (ISSSI), Seoul, 30 August – 2 September, 2016. Invited Speaker.
- Strauß L, **Ruffing U**, Akulenko R, Mellmann A and the StaphNet Consortium. Inferring *Staphylococcus aureus* Virulence and Resistance Traits from Whole Genome Sequences – Comparison of the Alere Indentibac© Microarray and the Respective *In*

Silico Typing Scheme. 67th Annual Meeting of the German society for Hygiene and Microbiology (DGHM), Münster, 22-25 September 2015. International Journal of Medical Microbiology, DVV314, p120.

- **Ruffing U**, von Müller L, Herrmann M and the StaphNet consortium. Comparison of the molecular staphylococcal epidemiology in Sub-Saharan Africa vs. Germany: a prospective, multicentre cohort study on community associated *Staphylococcus aureus*. 9th European Congress on Tropical Medicine and International Health (ECTMIH), Basel, 6-10 September 2015 O.LB2.002, p145.
- **Ruffing U**, Vranckx K, Herrmann M, von Müller L. Identification of clonal complex 121 *Staphylococcus aureus* strains by MALDI-TOF mass spectrometry, StaphNet consortium. 25th European Congress of Clinical Microbiology and Infectious Diseases (ECCMID), Copenhagen, 25-28 April 2015, www.escmid.org, O234.
- **Ruffing U**, Alab A, Abdulla S, Akulenko R, von Briesen H, Germann A, Grobusch MP, Helms V, Kazimoto T, Kern W, Mandomando I, Mellmann A, von Müller L, Peters G, Schaumburg F, Schubert S, Vubil D, Wende L, Herrmann M. Molecular genotyping of 1200 community acquired nasal and clinical associated *Staphylococcus aureus* isolates of the African-German StaphNet multicenter study. 66th Annual Meeting of the German society for Hygiene and Microbiology (DGHM), and Annual Conference for the Associaton for General and Applied Microbiology (VAAM), Dresden, 05-08 Octobre 2014. BIOs*pectrum*, KMV06, p121.
- **Ruffing U**, Wende L, Schubert S, Kazimoto T, Abdulla S, Akulenko R, Alabi A, Bischoff M, Germann A, Grobusch MP, Helms V, Hofman J, Kern W, Mandomando I, Mellmann A, Peters G, Schaumburg F, Vubil D, Herrmann M, von Müller L. African and German multicentre screening study of 1200 *Staphylococcus aureus* isolates. 65th Annual Meeting of the German society for Hygiene and Microbiology (DGHM), and German Society of Infectious Diseases (DGI), Rostock, 22-25 September 2013. International Journal of Medical Microbiology, MSV01, p77.

Published Abstracts and poster presentations:

- **Ruffing U**, Roth LE, Touré FS, Monemo P, Tia H, Cissé A, Diané B, Herrmann M, Akoua-Koffi C., Becker S. Comparative genotypic characterization of community associated *Staphylococcus aureus* from clinical infections and asymptomatic carriers in central Côte d'Ivoire. 13th Congress of Infectious diseases and tropical medicine (KIT), Würzburg, 15-18 june, 2016, www.kit2016.de/PDF/Abstract_Book.pdf, P-059.
- **Ruffing U**, Halfmann A, Bischoff M, Herrmann M, von Müller L. MRSA risk factors associated with predominant *spa*-types in hospitals – experience from a statewide MRSA admission screening (MRSAarNet). 66th Annual Meeting of the German society for Hygiene and Microbiology (DGHM), and Annual Conference for the Associaton for General and Applied Microbiology (VAAM), Dresden, 05-08 Octobre 2014. BIOs*pectrum*, KMP18, p121.
- **Ruffing U**, Akulenko R, Wende L, Schubert S, Kazimoto T, Delfino V, Alabi A, Herrmann M, Schaumburg F, Kern W, Herrmann M, von Müller L and the StaphNet consortium. Molecular Epidemiology of *Staphylococcus aureus* Infections in Germany and Africa – a Multicenter Cohort Study using DNA Microarray to Characterize Invasive and Commensal Community-associated Isolate. Interscience Conference on Antimicrobial Agents and Chemotherapy (ICAAC). 5-9 September 2014, www.icaaconline.com, C-814.
- **Ruffing U**, Akulenko R, Wende L, Schubert S, Kazimoto T, Delfino V, Alabi A, Herrmann M, Schaumburg F, Kern W. DNA microarray based genotyping of 1200 *Staphylococcus aureus* isolates of the African-German StaphNet multicentre study. 25th European Congress of Clinical Microbiology and Infectious Diseases (ECCMID), Barcelona, 10-13 May 2014, www.escmid.org, P1539.
- **Ruffing U**, Wende L, Schubert S, Kazimoto T, Abdulla S, Akulenko R, Alabi A, Bischoff M, Germann A, Grobusch MP, Helms V, Hofmann J, Kern W, Mandomando I, Mellmann A, Peters G, Schaumburg F, Vubil D, Herrmann M, von Müller L on behalf of the German-African study group of *Staphylococcus aureus*. Characterisation of *Staphylococcus aureus* isolates of an African and German multicentre screening by DNA microarray-based genotyping, 24th European Congress of Clinical Microbiology and Infectious Diseases (ECCMID), Barcelona, 27-30 April 2013, www.escmid.org, R2682.

- Wende L, Schubert S**,** **Ruffing U**, Bischoff M, Kern W, Hofmann J, Abdulla S, Schaumburg F, Peters G, Alabi A, Grobusch MP, Mandomando I, Mellmann A, Germann A, Herrmann M, von Müller L. Characterization of *Staphylococcus aureus* isolates of an African and German multicentre screening by DNA microarray based genotyping. 64th Annual Meeting of the German society for Hygiene and Microbiology (DGHM), Hamburg, 30 September to 03 October 2012. International Journal of Medical Microbiology, MSP07, p87.
- **Ruffing U**, Akulenko R, Petit C, Bischoff M, Helms V, Herrmann M, von Müller L. Comparison of microarray-based genotyping and *spa*-typing for routine typing of *Staphylococcus aureus*. 64th Annual Meeting of the German society for Hygiene and Microbiology (DGHM), Hamburg, 30 September - 03 October 2012. International Journal of Medical Microbiology, KMP10, p49.
- **Ruffing U**, Petit C, Halfmann A, Bischoff M, Herrmann M, von Müller L. Characterization of MRSA and MSSA strains of a federal state multicenter screening by *spa*-typing and microarray-based genotyping. 63th Annual Meeting of the German society for Hygiene and Microbiology (DGHM), Essen 25-28 September 2011. International Journal of Medical Microbiology, KMP03, p50/51.

9 Acknowledgement

My special thanks goes to Prof. Dr. Mathias Herrmann for the opportunity of this versatile and interesting topic, the steady willingness to engage in discussion and stimulating criticism, the trust in my work as well as for the excellent and diverse job opportunities at the institute.

I give my warmest regards to my supvervisor Prof. Dr. Markus Bischoff for the supervision, inspiring discussion and helpful suggestions and especially to Prof. Dr. Lutz von Müller, my project supervisor and mentor, for his suggestions, positive motivation, constructive criticism and steady engagement.

I want to thank Univ.-Prof. Dr. Martina Sester for the friendly willingness to write the second opinion of this work.

Thank you to all our African and German cooperation partners, especially of the African-German-StaphNet, for provision of isolates, interesting discussions, suggestions and fruitful collaboration.

I thank Univ.-Prof. Dr. Volkhard Helms and especially Dr. Ruslan Akulenko for their bioinformatic support and their patience in explaining me everything.

Thanks to my first medical students, Sabine Schubert and Laura Stichter, for the isolate collection work at the university clinic of Homburg. Thank you to our scientific guestworkers, Theckla Kazimoto and Delfino Vubil. Thanks to Adebayo Shittu, my brother, who supports me in my first “real” English steps. Especially, thank you all for the great time and your patience with your unexperienced teacher.

I would like to thank Eloise Müller-Schulte and Sarah Wendlandt for proofreading.

Thanks a lot to Rosmarie Gaupp and Robert Skov not only for proofreading, but also for his suggestions, constructive criticism and support.

Thank you to all colleagues of the institute, especially Susanne Loibl, Sabine Freis and Karin Hilgert, for the friendly working atmosphere and support in the diagnostic and research lab.

Thanks to my parents and my friends for their support and interest in my work.

Finally, special thank to my spouse, his support and patience.