universität wien

MASTERARBEIT

Titel der Masterarbeit

„Subtyping of livestock-associated methicillin-resistant *Staphylococcus aureus* CC398 isolates by next generation sequencing“

verfasst von

Sarah Lepuschitz, BSc

angestrebter akademischer Grad

Master of Science (MSc)

Wien, 2015

Studienkennzahl lt. Studienblatt:	A 066 834
Studienrichtung lt. Studienblatt:	Masterstudium Molekulare Biologie
Betreut von:	PD Mag. Dr. Werner Ruppitsch

Declaration

“I confirm that this thesis is entirely my own work. All sources and quotations have been fully acknowledged in the appropriate places with adequate footnotes and citations. Quotations have been properly acknowledged and marked with appropriate punctuation. The works consulted are listed in the bibliography. This paper has not been submitted to another examination panel in the same or a similar form, and has not been published. I declare that the present paper is identical to the version uploaded.”

Place, Date

Signature

Acknowledgements

I am deeply grateful to my supervisor, PD Mag. Dr. Werner Ruppitsch, for offering me such an interesting topic of research. Without his outstanding help, good guidance and advices throughout the process, this master-thesis would not have been achievable.

Furthermore, I would like to thank Univ.Prof. Dr. Franz Allerberger and PD Mag. Dr. Alexander Indra, for their helpful suggestions and support I received throughout the completion of my thesis.

Finally, but from the bottom of my heart, I would like to thank my parents, Judith and Robert Lepuschitz, for their patience and understanding. Without their love and support over the years, none of this would have been possible.

Index

1. Abstract

With the beginning of the new millennium a new methicillin-resistant *Staphylococcus aureus* (MRSA) clonal complex 398 (CC398) has emerged as a cause of infections in humans with contact to livestock (livestock-associated MRSA, LA-MRSA). In Austria, first infections caused by CC398 isolates were observed in 2004 when the National Reference Laboratory for Staphylococci (NRLS) at the Austrian Agency for Health and Food Safety (AGES) started typing MRSA isolates.

From 2004 till today the NRLS observed a steady increase of infections caused by CC398 isolates from 0.2% in 2004 to 7.9% in 2015.

All LA-MRSA isolates from the NRLS harboured CC398 corresponding *spa*-types (t011, t034, t108), whereby t011 was the most prevalent *spa*-type (87.6%). Due to the limited discriminatory power of the subtyping methods multilocus sequence typing (MLST) and sequencing of the variable X-region of staphylococcal protein A gene (*spa*-typing), next generation sequencing (NGS) was applied for high resolution typing of 153 LA-MRSA isolates and an arbitrary collection of 47 hospital acquired (HA)-MRSA isolates belonging to 12 classical MLSTs for comparison. A gene-by-gene approach was used for comparison of genomes based on a predefined core genome (cgMLST) comprising 1861 genes. NGS based typing allowed a differentiation between all investigated LA-MRSA and HA-MRSA isolates. HA-MRSA isolates differed in at least 1717 alleles from LA-MRSA isolates. The maximal allelic difference was 1658 among HA-MRSA isolates and 131 among LA-MRSA isolates. In general the genetic difference between *spa*-types was higher than within isolates of a certain *spa*-type.

In conclusion cgMLST is highly suitable for high resolution typing of LA-MRSA isolates with the same *spa*-type.

2. Introduction

2.1 *Staphylococcus aureus*

In 1882 Alexander Ogston first used the designation “staphylococcus” when he described a micrococcus appearing in groups which cause postsurgical sepsis (Ogston, 1882). Due to the golden colour of colonies the species was named *aureus* (Howard and Kloos, 1987).

The genome size of *S. aureus* is about 2.8 Million bases including both, the core and the accessory genome (Shittu et al., 2007). Bacterial species are defined by a set of conserved core genes, conventionally identified as those genes present in all isolates, and additionally by a number of accessory genes; thereby the composite superset of all genes present in a strain is named pan-genome (Segerman, 2012; van Tonder et al., 2014). In the accessory genome a large portion of none vitally important functions exist, such as genes for antibiotic resistance, toxin production and virulence genes, which can be transferred laterally to recreate new trait combinations (Segerman, 2012; Shittu et al., 2007).

Staphylococci are Gram-positive bacteria, they form grape-like clusters, are non-motile and behave facultative anaerobic (Harris et al., 2002). Within staphylococci, only *S. aureus* and *S. intermedius* are coagulase positive and able to coagulate blood plasma (Foster, 1996). Therefore pathogenic staphylococci can be distinguished from mostly “apathogenic” opportunistic pathogens by their ability to produce coagulase (Harris et al., 2002). The host´s prothrombin, together with coagulase (extracellular protein), produce the staphylothrombin complex, whereby fibrinogen changes into fibrin (Foster, 1996).

Staphylococci are relatively resistant to heat (Kloos and Lambe, 1991) and tolerate high salt concentrations (Wilkinson, 1997). *S. aureus* has a cell wall and a cytoplasmic membrane which encloses the cytoplasm (Shockman and Barrett, 1983). The cell wall basically consists of 50% peptidoglycan (Waldvogel, 1990) and is composed of a multi-layered cell wall mesh (Wilkinson, 1997), 40% teichoic acids (Knox and Wicken, 1973), surface proteins, exoproteins and autolysins (Thakker et al., 1998). *S. aureus* has the competence to adapt to environmental changes such as in case of infection (Harris et al., 2002). This adaptation includes the production of virulence regulators, such as *agr*

(accessory gene regulator), *sar* (staphylococcal accessory regulator) and *sae* (*S. aureus* exoprotein) (Harris et al., 2002).

S. aureus is a natural resident of the skin flora. Based on disruption of the natural invasion barrier, the skin, the bacteria are able to enter the underlying tissue (Elek and Stephen, 1956). It can also cause infections of skin, nose, urethra and gastrointestinal tract in immunosuppressed patients (Shulman and Nahmias, 1972). Enterotoxins A-E, toxic shock syndrome toxin and exofoliative toxins are extracellular toxins produced as a defence mechanism (Projan and Novick, 1997).

In the 1940s, the treatment of *S. aureus* infections with penicillin began and considerably improved the outcome. As a consequence first penicillin resistances emerged already in the late 1940s (Eickhoff, 1972). The use of antibiotics led to multiple drug resistance of *S. aureus* due to the uptake of resistance genes in mobile genetic elements (MGE) (Morris et al., 1998). In 1961 methicillin was introduced as a new antibiotic. Within a year first methicillin resistant *S. aureus* (MRSA) were reported (Jevons, 1961). Resistant strains encode a distinct PBP, PBP2a, encoded by *mecA* which has a low affinity to β-lactam antibiotics (Chamber, 1997) and PBP2a can still execute its function in cell wall synthesis despite the presence of β-lactam-antibiotics (Hiramatsu, 1995).

MRSA can be categorized corresponding to the origin of infection: hospital-, community- or livestock-associated MRSA. Epidemiology and spread of *S. aureus* can be studied by the use of sequence-based typing methods such as multilocus sequence typing (MLST) (Enright et al., 2000) or *spa*-typing (Harmsen et al., 2003). The *S. aureus* population pathogenic for humans consists of approximately ten dominant lineages (Lindsay, 2010): Seven major *S. aureus* lineages belong to hospital-acquired MRSA (CC5, CC12, CC15, CC22, CC25, CC45, CC51) (Feil et al., 2003) and three belong to community-acquired MRSA (CC1, CC8, CC30). Lineage CC398 detected in 2004, which belongs to livestock-associated MRSA is no predominant lineage at the moment but appears to be disseminating rapidly to many parts of the world (Lindsay, 2010).

2.2 Hospital-acquired MRSA

Until the 1980's MRSA was exclusively related to healthcare settings (HA-MRSA). The distribution of MRSA in healthcare settings and the probability to acquire an infection is dependent on various factors such as the use of antibiotics, present disease patterns, time duration of hospitalization, surgery, stay in intensive care units and the level of infection control (hygiene) (Tacconelli et al., 2008).

As a consequence of appropriate, inordinate and unnecessary use of antibiotics successful pathogenic *S. aureus* lineages evolved due to uptake of mobile genetic elements encoding resistance and virulence factors (Muto et al. 2003). Worldwide the majority of HA-MRSA infections is caused by clones of five major lineages (Robinson and Enright, 2003).

The prudent use of antibiotics constitutes an important factor to prevent and decrease MRSA colonization (Tacconelli et al., 2008).

2.3 Community-acquired MRSA

A significant change in the epidemiology of MRSA happened in the 1980s with the emergence of a new clone causing infections in people in the community without previous contact to healthcare facilities (Stegger et al., 2014). Community-acquired (CA)-MRSA are characterized by their ability to infect healthy and often young individuals, assuming that this clone has an increased virulence as well as the power to spread easily (DeLeo et al., 2010; Uhlemann et al., 2014). Typical molecular markers for CA-MRSA include SSCmec IVa (type of methicillin resistance locus) and a gene coding for Panton-Valentin leucocidin (PVL) (Vandenesch et al., 2003). In general CA-MRSA are less resistant to antibiotics than HA-MRSA. European clones seem to be more resistant than U.S. and Oceanian isolates (Stegger et al., 2014; Vandenesch et al., 2003). Usually CA-MRSA causes skin and soft tissue infections, but also necrotizing pneumonia and sepsis have been reported (Uhlemann et al., 2014). Since its occurrence CA-MRSA clones have spread worldwide. Geographically different dominant lineages can be found (Stegger et al., 2014) such as CC80 in Europe, CC1 and CC8 in the United States, CC59 in Asia and CC30 in the south west

Pacific (Lindsay, 2010). The genetic background of CA-MRSA does not correlate with local HA-MRSA, suggesting that CA-MRSA did not originate from HA-MRSA in the same continent. In contrast CA-MRSA showed similarities to HA-MRSA from other continents, whereby an intercontinental transmission seems to be very probable (Vandenesch et al., 2003). Meanwhile highly successful lineages already entered the healthcare settings and therefore became an important reason for healthcare-associated bacteremia (Uhlemann et al., 2014).

2.4 Livestock-associated MRSA

In the beginning of the new millennium a new category of MRSA was identified due to its association to livestock (Voss et al., 2005). Single reports of MRSA in animals occurred already in 1972 (Devriese et al., 1972). The colonization and infection of humans with contact to livestock lead to the discovery of clonal complex 398 in the Netherlands in 2004 (Cuny et al., 2009). This clone attracted attention because it was non-typeable by *Sma*I pulsed-field gel electrophoresis, the gold standard typing method (Huijsdens et al., 2006). Livestock-associated (LA) MRSA include mostly strains belonging to sequence type (ST) 398, harboring several corresponding *spa*-types such as t011, t034, t108 and t1793 (Grisold et al., 2010). So far CC398 isolates are prevalent in pigs but can be isolated from most domestic animals (Loncaric et al., 2014). Since its emergence it has been reported in Europe, Asia and the United States. The proportion of LA-MRSA isolates seems to correlate with the density of pig and calf farming within a country (van Cleef et al., 2011).

LA-MRSA isolates seem to be less virulent, due to the lack of typical virulence determinants (Price et al., 2012). However, they show an increasing gain of antibiotic resistance profiles, especially tetracycline resistance, which can be explained by high antibiotic use in animal husbandry (Larsen et al., 2015; Price et al., 2012).

An infection with LA-MRSA causes mostly skin and soft-tissue infections, but also sepsis and even necrotizing pneumonia have been reported (Cuny et al., 2013).

LA-MRSA CC398 cause human colonisation and disease especially in people with direct exposure to livestock (farmers, veterinarians). As a consequence human to human transmissions between family members have been described (Larsen et al., 2015). Dust from pig holdings and contaminated meat is a possible reservoir for LA-

MRSA infections in unexposed people and the spread into the community and healthcare settings (Larsen et al., 2015; Monaco et al., 2013).

As described by Price et al. (2012) LA-MRSA CC398 probably emerged from human MSSA strains (methicillin sensitive *S. aureus*) and as a consequence of transmission reached livestock, where methicillin and tetracycline resistance was acquired.

A short overview on the historical evolution of MRSA is presented in Figure 1.

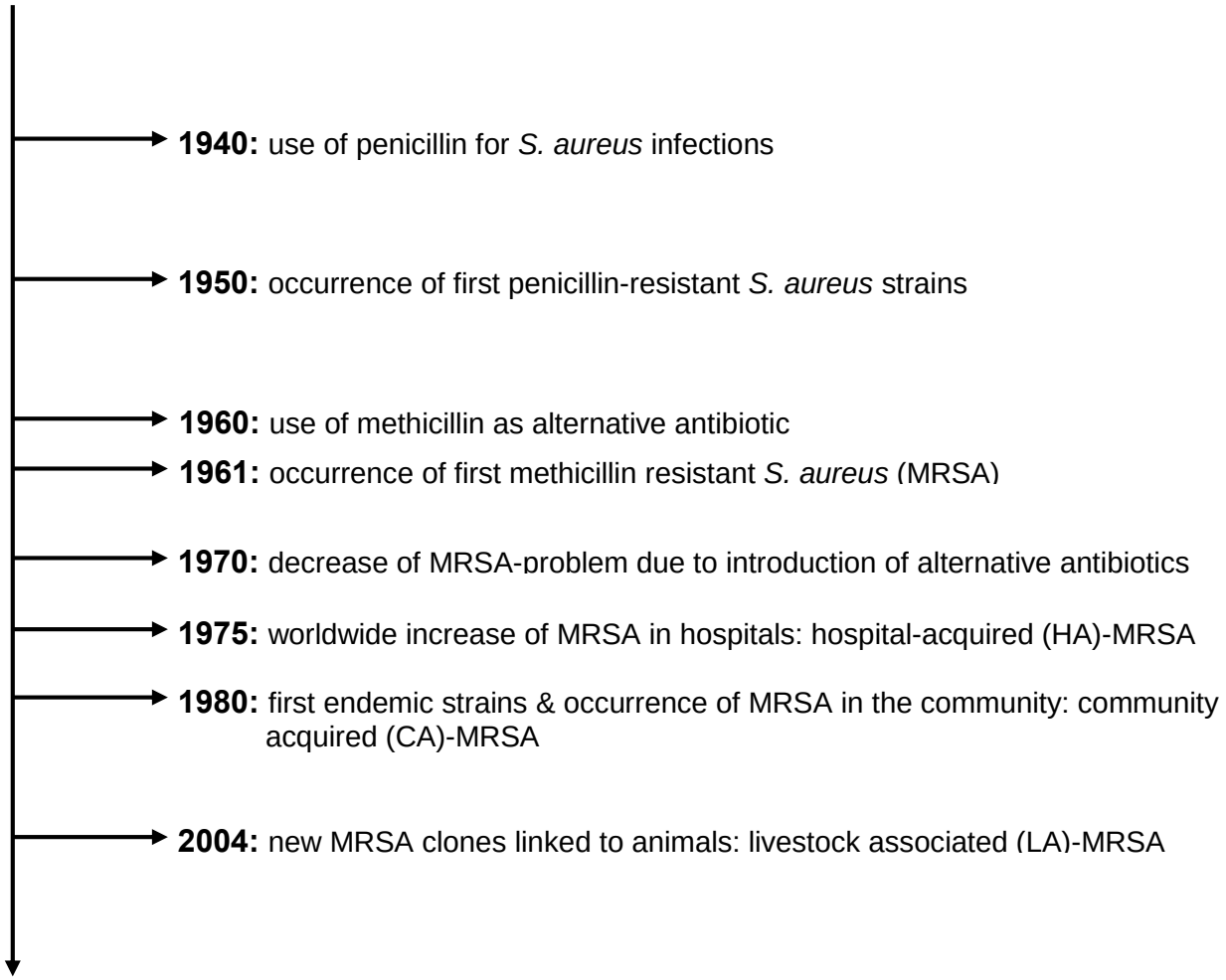

Figure 1. Time lapse of the evolution of methicillin resistant *Staphylococcus aureus*

2.5 Typing methods

2.5.1 Pulsed-field gel electrophoresis (PFGE)

In case of an outbreak pulsed-field gel electrophoresis has been the gold standard method for molecular typing of several pathogens (Leopold et al., 2014). The method developed in 1983 yielded highly discriminatory DNA fingerprints (Schwartz et al., 1983). The relatedness between isolates was defined by a difference of less than four fragments, which assigns isolates to the same strain; a difference of four to six fragments assigns isolates to different strains but to the same genetic lineage (Enright et al., 2000).
The major drawback of this technique are the costs in terms of hands on-time, need of a special software for evaluation and the poor reproducibility of results between laboratories. However, due to the high discriminatory power it is still a valuable typing tool (Harmsen et al., 2003).

2.5.2 *Spa*-typing

Spa-typing characterizes the sequence differences of the polymorphic X region of staphylococcal protein A gene (*spa*) (Cai et al., 2007) consisting of at least two to 27 repeat units of 21 to 27 base pairs in length. The *spa*-type is defined by the sequence, the number and order of repeats, representing the final *spa*-type (Shopsin et al., 1999). The surface immune globulin (Ig)-binding protein A acts as virulence factor enabling *S. aureus* to escape the host's immune response. To circumvent phagocytosis it binds IgG molecules in the incorrect orientation to support the survival of *S. aureus* (Votintseva et al., 2014). *Spa*-typing is one of the most advanced typing methods due to its low expenditure of time, low costs, high discriminatory power and the ability to share and accordingly compare results via an online database (http://www.spaserver.ridom.de) (Cai et al., 2007). In comparison to DNA fragment based typing methods *spa*-typing has the advantage to produce reliable and correct typing information (Harmsen et al., 2003). The database presently contains 15507 *spa*-types (08.12.2015, www.spaserver.ridom.de).

2.5.3 Multilocus sequence typing

Multilocus sequence typing (MLST) distinguishes bacterial isolates due to their differences in the sequences of seven housekeeping genes (Enright et al., 2000). The obtained allelic profile of these genes determines the sequence type (ST) of an isolate. Global MLST databases (http://www.mlst.net/databases/) were established for several important pathogens and sequence data can be submitted for ST determination (Maiden et al., 1998).

These databases allow a worldwide comparison of isolates and provide phylogenetic and epidemiological information (Jolley and Maiden, 2014). Due to the fact that housekeeping genes have a slow mutation rate it is possible to determine if isolates originate from a common ancestor (Enright et al., 2000). MLST provides a high reproducibility and portability and is therefore the method of choice for long-term epidemiological investigations. For MRSA the method is not suitable for outbreak investigation due to limited resolution between very closely related isolates (Jolley and Maiden, 2014).

2.5.4 Next generation sequencing

Next generation sequencing is the method of highly parallel sequencing of short fragments and has revolutionised our understanding of bacterial evolution (Price et al., 2013). Direct analysis of clinical samples provides a high discriminatory power within a short time (Köser et al., 2012). Strains which have the same *spa*-, MLST-, PFGE-type can appear genetically similar but by NGS it is possible to distinguish epidemiologically unrelated strains (Leopold et al., 2014). Therefore NGS based typing is the ultimate tool for outbreak investigations and allows to determine if genetically similar isolates belong to an outbreak or not (Mellmann et al., 2011).

A species genome assembly, which is a highly complex computation, needs about eight copies of a piece of a genome to produce good results. There are many possibilities why errors can appear, such as low coverage, elimination of pieces which are incorrectly identified as mistakes or repeats, or pieces which are inserted in wrong places or in wrong orientation. Therefore, if a gene is missing after the assembly, this absence is not a confirmation of an in fact missing gene (Baker, 2012).

With future developments NGS will allow to predict resistances, identify virulence and type isolates at the highest possible resolution (Price et al., 2013). The development of fully automated software solutions makes it possible to use NGS in the daily routine and to analyse, store and exchange large volumes of genomic data (Kohl et al., 2014; Leopold et al., 2014; Ruppitsch et al., 2015).

In Figure 2 the NGS workflow is presented schematically for the Illumina platform (Illumina, Inc., San Diego, CA, USA). Briefly the workflow starts with DNA isolation, which is followed by a library preparation and final analysis on a NGS platform. The flow cell is a glass slide containing lanes coated with a lawn of surface bound adapter-complimentary oligos, where cluster generation takes place. Cluster generation starts with the binding of DNA molecules to the surface and a subsequent so called bridge amplification. Thus, a cluster is characterized by a clone of the same DNA fragment. Approximately 1500 clusters/mm² can be obtained resulting in up to 15 Gigabases per run. Sequencing is performed in both directions (paired-end sequencing) obtaining up to 25 million reads and read lengths of 2 x 300 bp (Illumina Pub. No. 770-2011-001 as of 25 September 2015 and Illumina Pub. No. 770-2008-016 as of 30 January 2009).

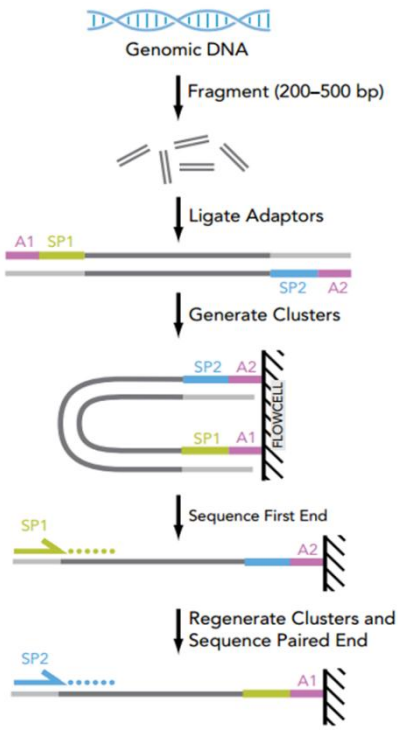

Figure 2. Procedure of paired-end sequencing. (Illumina Pub. No. 770-2008-016 as of 30 January 2009)

2.5.4.1 Core genome MLST

The core genome MLST (cgMLST) is a gene-by-gene approach suitable for comparison of bacterial isolates. This approach is in principle an extension of the classical MLST scheme to hundreds or thousands of genes representing the previously defined so-called core genome of a bacterial species. With cgMLST an efficient analysis of whole genome data and the establishment of a steady increasing database is possible (Jolley and Maiden, 2014).

2.5.4.2 Single nucleotide polymorphisms

For specific questions the resolution of cgMLST may be still insufficient and a method which provides higher resolutions is required. Single nucleotide polymorphisms (SNPs) can be used to show the genetic distance between closely related isolates at the highest possible resolution. For identification of an outbreak SNP analysis can give more accurate information than cgMLST about potential transmission events or definition of an outbreak cluster. However, despite the high resolution of NGS the interpretation of genetic distance data alone still has uncertainties caused by mutation rates and population dynamics. Therefore epidemiological data are essential for interpretation of NGS based typing data used for transmission and outbreak investigation (Worby et al., 2014).

2.6 Aim and task description

The prevalent LA-MRSA clone CC398 shows a steady increase in European countries with intensive livestock production since its emergence in 2003 in the Netherlands (Larsen et al., 2015). This trend was also observable at the NRLS in Austria (Allerberger et al., 2013). The aims of this study were to evaluate the suitability of NGS to determine the diversity of the Austrian LA-MRSA population and its virulence and resistance status. Typing performance of NGS was further evaluated by analysis of a cluster of six *spa*-type t011 isolates involved in a nosocomial transmission from 2010 to 2011 (Schmid et al., 2012).

3. Material and Methods

3.1 Bacterial isolates

For this study a total of 153 LA-MRSA isolates and 47 HA-MRSA isolates were arbitrarily selected. All LA-MRSA isolates were collected in Austria from 2004 to 2015. Isolates were chosen due to their previously detected *spa*-types t011 (n=134), t034 (n=14) or t108 (n=5). Eighty-seven isolates of this collection were obtained from different hospitals in Austria. Of the 153 LA-MRSA isolates, 14 isolates had animal origin, 20 were isolated from food, 20 were environmental isolates, and 12 were from inter-laboratory quality-control tests. For comparison 47 HA-MRSA were selected, which were collected in Salzburg from 2012 to 2014.

3.2 *Spa*-typing

Spa-typing was performed as described (Harmsen et al., 2003). Briefly, the variable x region of the *spa* gene was amplified with primers containing an M13 primer sequence on the 5'-end of the specific priming sequence *spa*-1113F (M13univ.): 5´TGTAAAACGACGGCCAGTTAAAGACGATCCTTCGGTGAG-3´ and *spa*-1514R (M13rev.): 5´CAGGAAACAGCTATGACCCAGCAGTAGTGCCGTTTGCTT-3´. Conditions were : 95°C 5.00 min [95°C 15 s, 55°C 30 s, 72°C 45 s]x45, 72°C 10.00 min.

DNA sequencing was performed as described (Platt et al., 2007) using standard M13 sequencing primers M13univ.: 5´TGTAAAACGACGGCCAGT-3´, and M13rev.: 5´CAGGAAACAGCTATGACC-3´ under following conditions: 96°C 1.00 min [96°C 10 s, 50°C 5 s, 60°C 1.15 min]x15, [96°C 10 s, 50°C 5 s, 60°C 1.30 min]x5, [96°C 10 s, 50°C 5 s, 60°C 2.00 min]x5 using the BigDye Terminator v3.1 kit (applied biosystems by Thermo Fisher Scientific, Waltham, MA, USA). The final analysis was done on a Genetic Analyzer 3500DX (applied biosystems). Ridom StaphTypeTM Software (Ridom GmbH, Münster, Germany) was used for determination of a *spa*-type.

3.3 Triplex PCR (*mecA*, *femA*, PVL)

Triplex PCR was performed as described (Stöger et al., 2007). All isolates were tested for *mecA* (methicillin resistance), *femA* (specific for *S. aureus* and essential for the expression of high-level methicillin resistance) and *luk*-PV (specific for Panton-Valentine Leukocidin (PVL) toxin) using following primer pairs: *mecA* 147 fwd.: 5´GTGAAGATATACCAAGTGATT-3´, *mecA* 147 rev.: 5´ATGCGCTATAGATTGAAAGGA-3´, *femA* fwd.: 5´AACTGTTGGCCACTATGAGT-3´, *femA* rev.: 5´CCAGCATTACCTGTAATCTCG-3´, *LUK*-PV fwd.: 5´ATCATTAGGTAAAATGTCTGGACATGATCCA-3´, *LUK*-PV rev.: 5´GCATCAAGTGTATTGGATAGCAAAAGC-3´ under following conditions : 95°C 5min [95°C 30 s, 55°C 45 s, 72°C 45 s]x45, 72°C 10 min.

3.4 Disk diffusion methodology

The agarose diffusion test was done with 34 arbitrarily selected LA-MRSA isolates according to the method of EUCAST (Antimicrobial susceptibility testing, Version 5.0, January 2015). Isolates were cultured on blood agar (bioMérieux, Inc., Hazelwood, MO, USA, Ref: 43 049) overnight at 37°C. The inoculum was prepared by selecting several morphologically similar colonies and adjusted to McFarland 0.5. For the agar diffusion test Müller Hinton agar plates (bioMérieux, Ref: 43309) were used with different antibiotic discs (Table 1). The plates were incubated at 35 ± 1 °C for 18 ± 2 hours. After incubation the zone diameter [mm] was measured (R= resistant, S= sensitive) (EUCAST Clinical Breakpoint Table v. 5.0, valid from 2015-01-01).

Table 1. Breakpoints in mm for disk diffusion test for *Staphylococcus aureus*

Antibiotic	Disc concentration [µg]	R <	S ≥
Gentamicin	10	18	18
Erythromycin	15	18	21
Clindamycin	2	19	22
Trimethoprim/ Sulfamethoxazol	25	14	17
Ciprofloxacin	5	20	20
Tetracycline	30	19	22

(R= resistant; S= susceptible)

3.5 Next generation sequencing

3.5.1 DNA Isolation

All 200 isolates were cultured on Columbia blood agar plates (bioMérieux, Ref: 43 049) overnight at 37°C. Extraction of genomic DNA (gDNA) was performed with commercial DNA extraction kits QuickExtract™ DNA Extraction Solution 1.0 (Epicentre, Illumina, Inc.) and MagAttract HMW DNA Kit (Qiagen, Hilden, Germany), according to the manufacturer´s instructions. Twenty units lysostaphin (Sigma-Aldrich Corporation, St. Louis, MO, USA) were added to each sample for better lysis of *S. aureus* isolates. To test the reproducibility and influence of DNA quality, both DNA extraction methods were carried out for one isolate (ID = case 6a & 6b).

3.5.2 Whole-genome sequencing

For DNA library preparation 1 ng of purified gDNA was used as recommended by the manufacturer (Illumina, sample preparation guide). gDNA was quantified on a Qubit® 2.0 Fluorometer using the dsDNA BR Assay Kit (Invitrogen by Thermo Fisher Scientific, Waltham, MA, USA) according to the manufacturer's instructions and diluted to 0.2 ng/µl as recommended by the manufacturer (Illumina sample preparation guide).

The Nextera XT DNA Library Preparation Kit (Illumina) was used for preparation of ready-to-sequence libraries of bacterial genomes. The optimal number of isolates per run was calculated with Sequencing Coverage Calculator (www.illumina.com/CoverageCalculator) to obtain a desired coverage of at least 50-fold. MiSeq Reagent Kit v3 containing the reagent cartridge and the flow cell was used for sequencing on an Illumina Miseq. Pooled libraries were loaded on the reagent cartridge. Sequencing was performed in both directions (paired-end sequencing) using a read length of 2 x 300 bp.

3.5.2.1 Assembly and data analysis

For sequence analysis the obtained reads (FASTQ files) were quality trimmed and de novo assembled using Velvet (http://www.ebi.ac.uk/~zerbino/velvet/). Reads were trimmed at their 5′ and 3′ ends until an average base quality of 30 was reached in a window of 20 bases, and the assembly was performed with Velvet version 1.1.04 using optimized k-mer size and coverage cutoff values based on the average length of contigs with >1,000 bp (Jünemann et al., 2013). Assembled genomes were uploaded into SeqSphere$^+$ (Ridom GmbH, Münster, Germany) and CLC Genomics Workbench (CLC bio, Qiagen) for final analysis. With SeqSphere$^+$, genomes were compared by a gene-by-gene approach based on a predefined core genome (cgMLST). The defined core genome using strain *S. aureus* (SA)COL (accession no. NC_002951.2) as reference comprised 1862 core genome genes and 706 accessory genome genes (Leopold et al., 2014). The obtained allelic profiles were used to visualize the population structure and a cluster type was defined with a maximum of 15 allelic differences (Alexander Mellmann, personal communication). Diversities of the investigated bacterial collections were displayed as minimal spanning trees (MST) using SeqSphere$^+$. In addition analysis in SeqSphere$^+$ revealed the classical multilocus sequence type (MLST), *spa*-type (protein A), the accessory genome and allowed the comparison of isolates based on single nucleotide polymorphisms (SNPs). AlignIR software (LI-COR, Lincoln, NE, USA) was used for sequence alignment comparisons. New MLST alleles were converted into BAM files for submission to the classical *S. aureus* MLST database using CLC Genomics Workbench. Arbitrary chosen antibiotic resistance-, virulence- and toxin-genes of interest were particularly searched in the

reference strain SACOL gene list (http://www.uniprot.org/). In addition, two task templates were created for toxin genes and virulence genes (Supplemental file 2) (Joensen et al., 2014), and antimicrobial resistance genes (Supplemental file 3) (Zankari et al., 2012) (sequence data were obtained from http://www.genomicepidemiology.org/).

4. Results

4.1 *Spa*-typing

4.1.1 LA-MRSA

Sanger sequence based *spa*-typing of 153 LA-MRSA isolates identified *spa*-types t011 (n=134), t034 (n=14) and t108 (n=5). All isolates were negative for *luk*-PV. For 134 isolates a positive result for *mecA* and *femA* could be obtained. Seven isolates were positive for *mecA* and negative for *femA* and 12 isolates were negative for *mecA* and *femA*.

4.1.2 HA-MRSA

Sanger sequence based *spa*-typing of 47 HA-MRSA isolates identified *spa*-types t002, t003, t005, t008, t012, t015, t019, t026, t032, t044, t056, t084, t085, t091, t127, t162, t216, t272, t284, t318, t449, t583, t789, t1626, t2216, t13492. All isolates were positive for *mecA* and *femA*. One *spa*-type t044 isolate and two *spa*-type t008 isolates were positive for *luk*-PV.

4.2 NGS data results

4.2.1 Population structure

4.2.1.1 cgMLST HA-MRSA and LA-MRSA

The obtained coverage for all MRSA isolates was from 24 to 116-fold. The percentage of good cgMLST targets ranged from 91.0 to 99.6%.

Based on the defined cgMLST scheme a minimum spanning tree was calculated for all HA-MRSA and LA-MRSA isolates and displayed in relation to the Austrian provinces (Figure 3).

Based on the defined cluster threshold (CT) of 15 allelic differences 18 different complexes were obtained. HA-MRSA isolates differed from LA-MRSA isolates by at least 1472 alleles and maximal 1743 alleles (average distance 1573 alleles). The allelic differences within LA-MRSA isolates ranged from a minimum distance of zero to a maximum distance of 193 (average distance 54 alleles). HA-MRSA isolates showed a minimum distance of zero and a maximum distance of 1748 alleles (average distance 1464 alleles).

One hundred-twelve LA-MRSA isolates could be assigned to 15 complexes, 41 were singletons; 10 HA-MRSA isolates could be assigned to three complexes, 37 were singletons (Supplemental file 1).

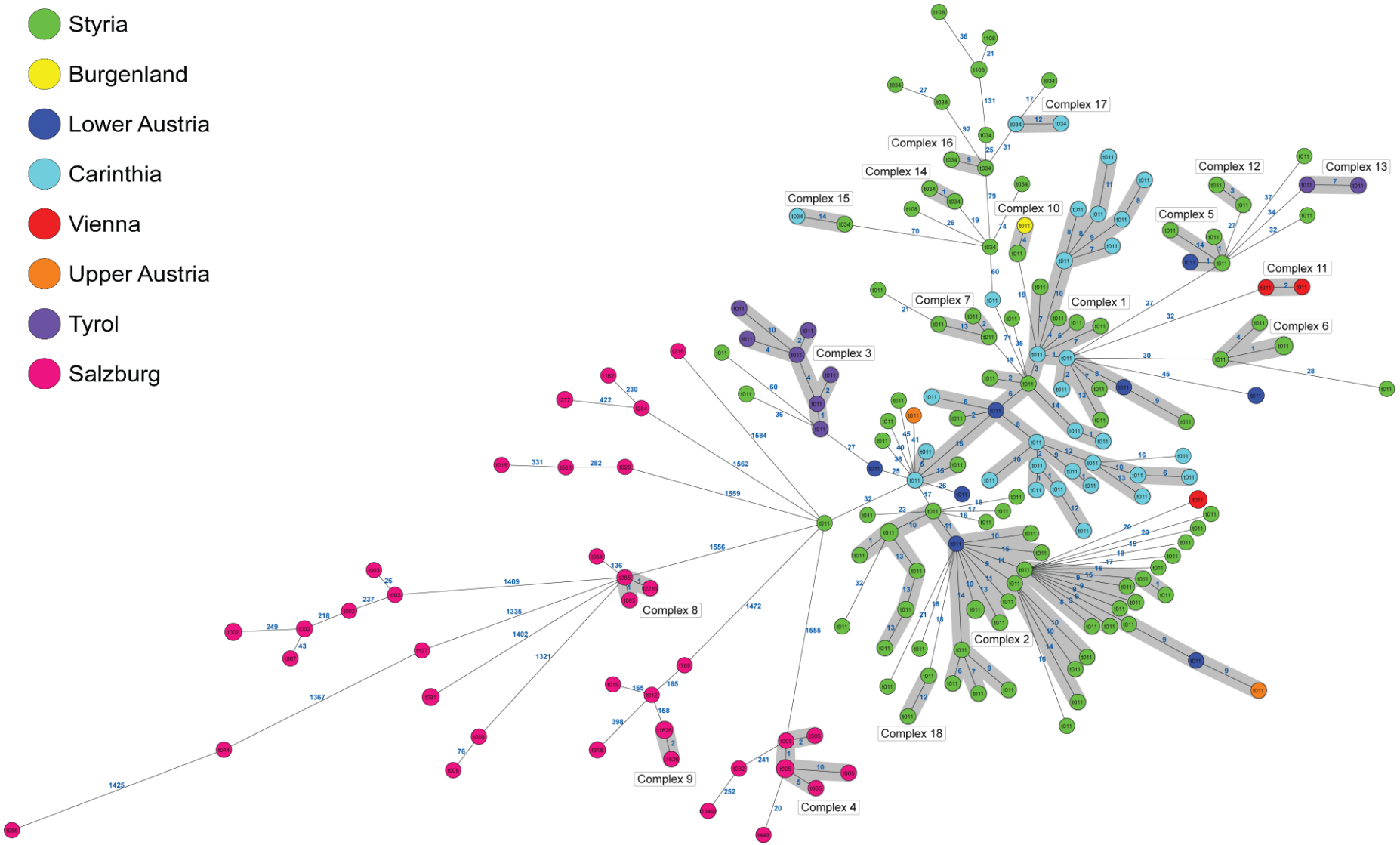

Figure 3. Minimum spanning tree for HA-MRSA and LA-MRSA isolates based on the cgMLST of *S. aureus*. Colours correspond to the country of origin in Austria, whereby all isolates from Salzburg belonged to HA-MRSA. Each circle represents isolates with an allelic profile based on the sequence of the core genome which consists out of 1862 alleles. Blue numbers accord to the allelic differences between two isolates. Isolates with closely related genotypes were identified with a maximum of 15 allelic differences and are shaded in grey.

4.2.1.2 cgMLST LA-MRSA

For a more specific analysis minimum spanning trees (MST) were calculated for all LA-MRSA isolates and are shown based on the province of origin in Austria (Figure 4) and on the sample material (i.e. animal, environmental, food, human) (Figure 5). Circles represent isolates with characteristic allelic profiles based on the core genome genes. Cluster analysis distinguished 15 different complexes with one major cluster (complex 1) comprising 40 isolates from Carinthia, Lower Austria and Styria. Forty-one isolates could not be assigned to a complex due to more than 15 allelic differences (Figure 4). On average LA-MRSA isolates differed by 54 alleles. Isolates belonging to *spa*-type t011 differed from isolates of *spa*-type t034 by at least 60 alleles and from isolates of *spa*-type t108 by at least 86 alleles. Isolates belonging to *spa*-type t034 showed from *spa*-type t108 an allelic difference of 26. The allelic differences were up to 71 between *spa*-type t011 isolates, up to 92 between *spa*-type t034 isolates and up to 36 between *spa*-type t108 isolates.

LA-MRSA isolates from horses formed complex 3 (7 samples) and complex 11 (2 samples). These horse isolates differed by a least 35 (complex 3) and 27 (complex 11) alleles from human LA-MRSA isolates. Three out of five LA-MRSA isolates from pigs showed high similarity (2-5 allelic differences) to human LA-MRSA isolates (complex 1). Two pig isolates were singletons unassignable to human LA-MRSA isolates. Three food LA-MRSA isolates (complex 2) from the province Styria were similar to several patient samples from Styria (9-14 allelic differences). Eleven LA-MRSA isolates derived from dust were closely related to human isolates of complexes 1, 2, 4, 6 and 7 with a maximum of 15 allelic differences (Figure 5).

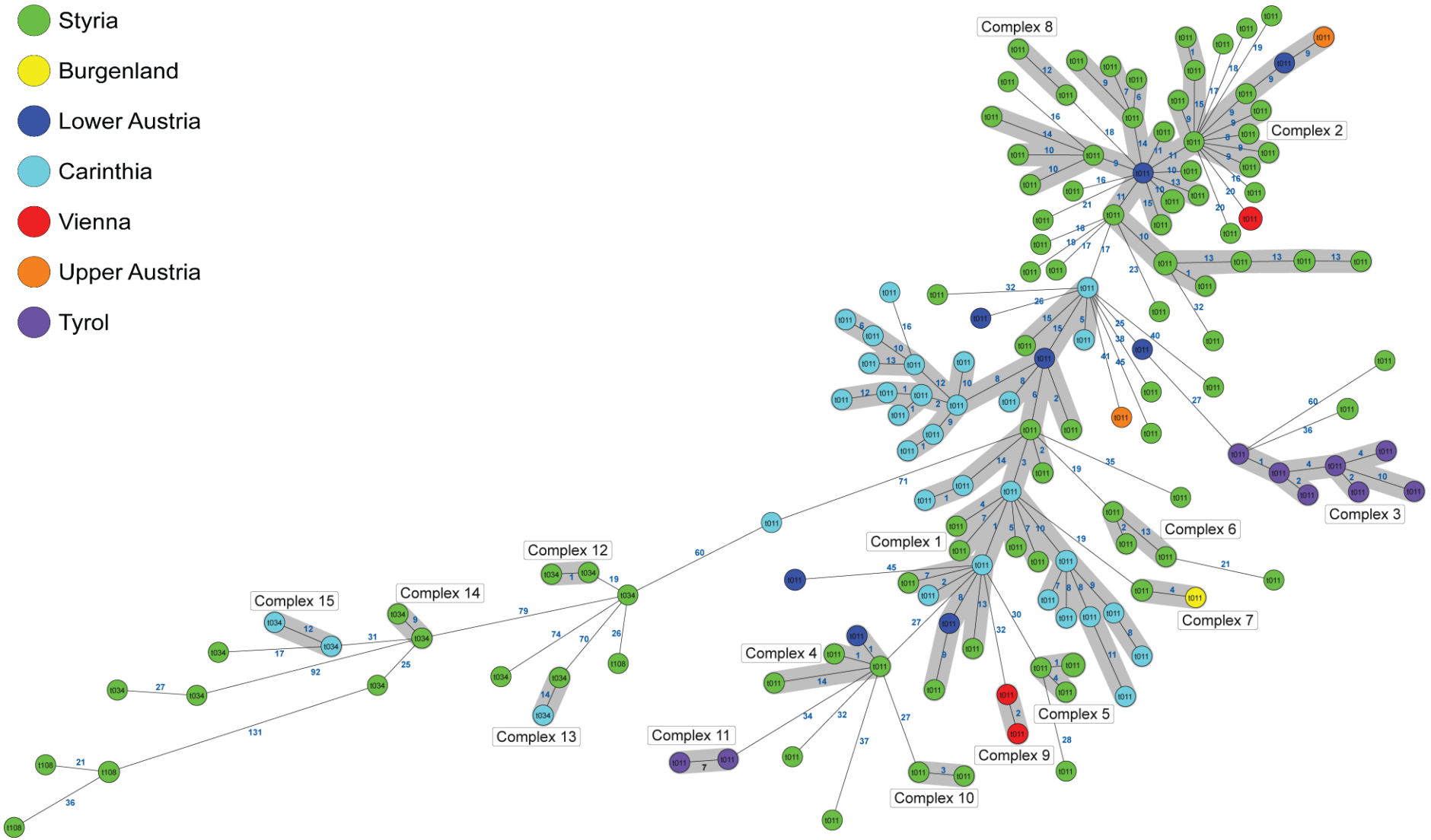

Figure 4. Minimum spanning tree for LA-MRSA isolates based on the cgMLST of *S. aureus*. Colours correspond to the country of origin in Austria. Each circle represents isolates with an allelic profile based on the sequence of the core genome which consists out of 1862 alleles. Blue numbers accord to the allelic differences between two isolates. Isolates with closely related genotypes were identified with a maximum of 15 allelic differences and are shaded in grey.

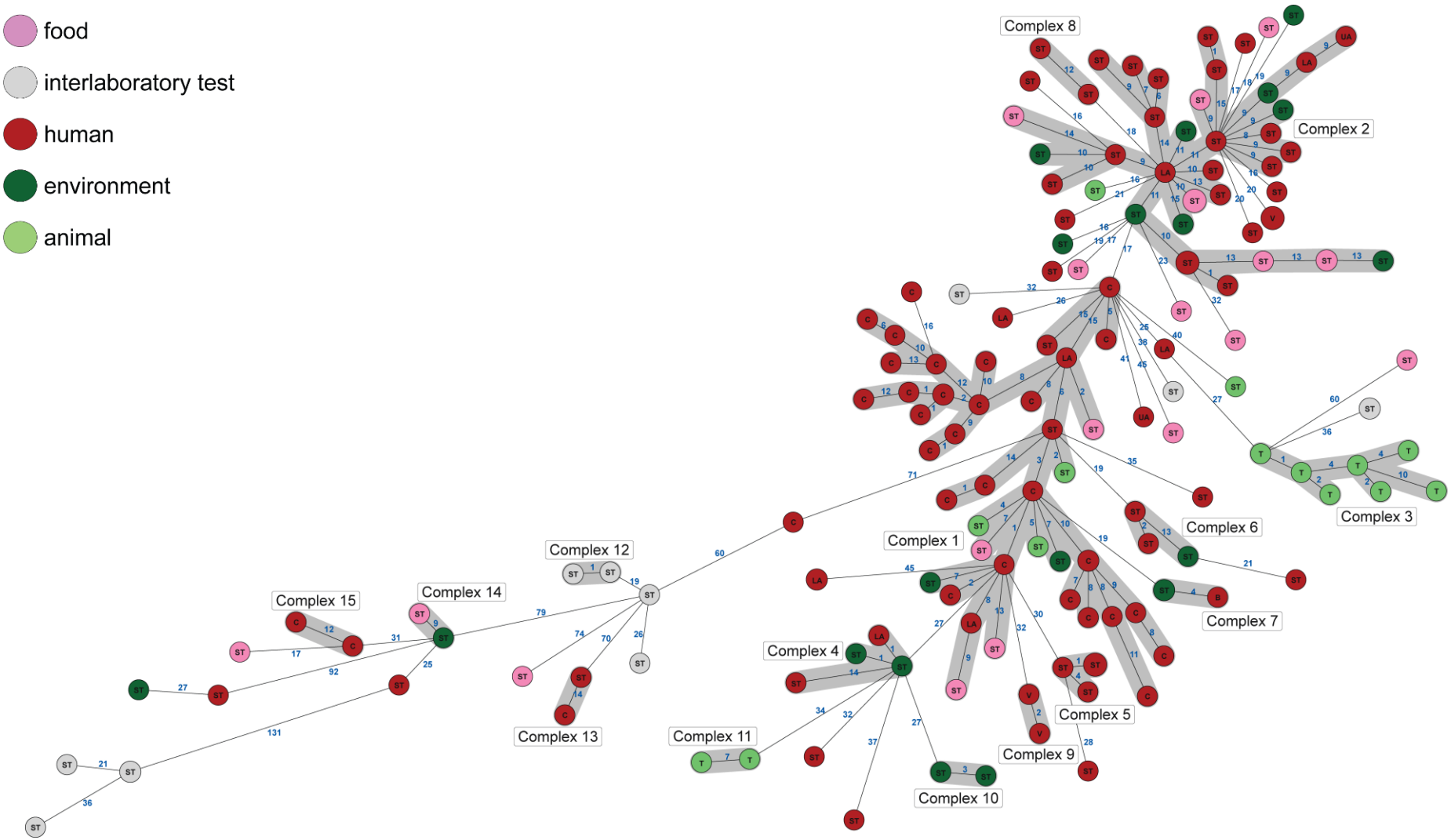

Figure 5. Minimum spanning tree for LA-MRSA isolates based on the cgMLST of *S. aureus*. Colours correspond to the sample material of every isolate. Each circle represents isolates with an allelic profile based on the sequence of the core genome which consists out of 1862 alleles. Blue numbers accord to the allelic differences between two isolates. Isolates with closely related genotypes were identified with a maximum of 15 allelic differences and are shaded in grey. B=Burgenland, C=Carinthia, LA= Lower Austria, ST=Styria, T=Tyrol, UA= Upper Austria, V=Vienna

4.2.1.3 cgMLST HA-MRSA

Minimum spanning trees (MST) were calculated from cgMLST data for all HA-MRSA isolates and were presented in correlation to the classical MLST clonal complex (CC) or sequence type (ST), where an assignment to a CC was unavailable (Figure 6). Classical MLST analysis defined five CCs and seven single STs. Based on the defined cluster threshold (CT) of 15 allelic differences cgMLST cluster analysis determined three different complexes (complex 4, 8 and 9 as determined in Figure 3) and 37 isolates unassignable to a complex. Complex 4, which was assigned to CC22, comprised isolates with *spa*-type t005. Three additional isolates with *spa*-types t032, t449 and t13492 were assigned to CC22 but not included in complex 4 due to more than 15 allelic differences. Complex 8, which was assigned to ST15, comprised two isolates with *spa*-type t085 and one isolate with *spa*-type t2216. One additional isolate with *spa*-type t084 was assigned to ST15 but not included in complex 8 due to more than 15 allelic differences. Complex 9, which was assigned to CC30, comprised two isolates with *spa*-type t1626. Four additional isolates with *spa*-types t012, t019, t318 and t789 belonging to CC30 were not included in complex 9 due to at least 161 allelic differences. Six isolates with *spa*-types t002 (n=3), t003 (n=2) and t067 all belonging to CC5 differed in 26 to 249 alleles. Two isolates with *spa*-type t008 belonging to CC8 differed in 76 alleles. Three isolates with *spa*-types t015, t026 and t583 belonging to CC45 differed in 282 to 331 alleles. Three isolates with *spa*-types t162, t272 and t284 all ST121 differed in 230 to 422 alleles.

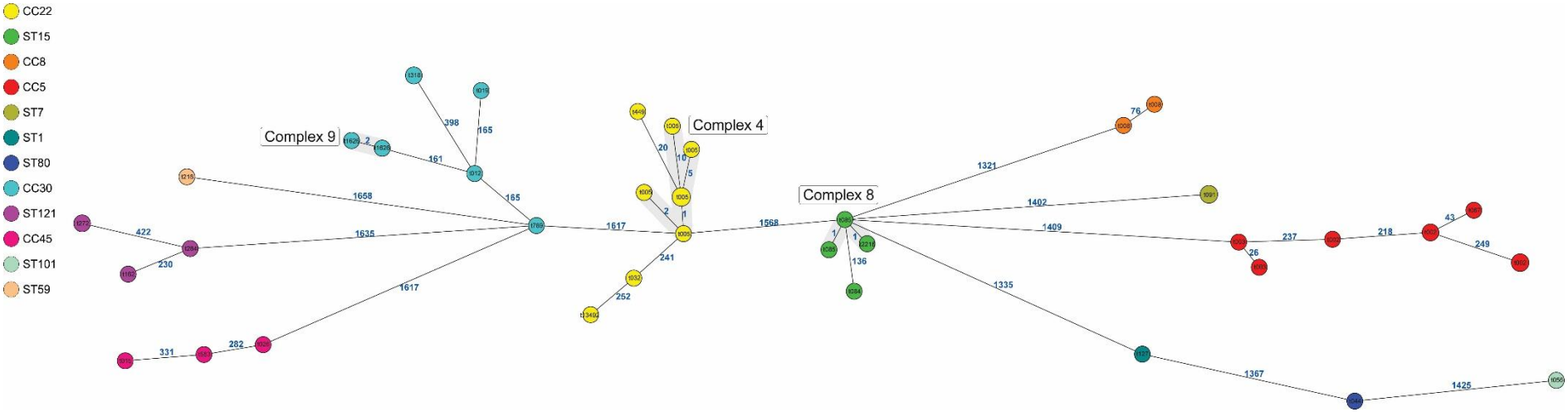

Figure 6. Minimum spanning tree for HA-MRSA isolates based on the cgMLST of *S. aureus*. Colours correspond to the classical MLST clonal complex (CC) or sequence type (ST). Each circle represents isolates with an allelic profile based on the sequence of the core genome which consists out of 1862 alleles. Blue numbers accord to the allelic differences between two isolates. Isolates with closely related genotypes were identified with a maximum of 15 allelic differences and are shaded in grey.

4.2.1.4 Pan genome LA-MRSA

For a more specific analysis minimum spanning trees (MST) were calculated for all LA-MRSA isolates based on the defined pan genome (core and accessory genome) and are shown based on the province of origin in Austria (Figure 7) and on the sample material (Figure 8).

Pan genome (pg) MLST increased the number of allelic differences within LA-MRSA isolates. The minimum distance remained at zero alleles, the maximum distance increased from 193 to 247 alleles, and the average distance increased from 54 to 77 alleles. The cluster type threshold was not changed for comparability with cgMLST analysis of LA-MRSA isolates (Figure 4, Figure 5). Pan genome MLST cluster analysis determined eight identical complexes as cgMLST (complex 3, 5, 8-10 and 11-13 in Figure 4).

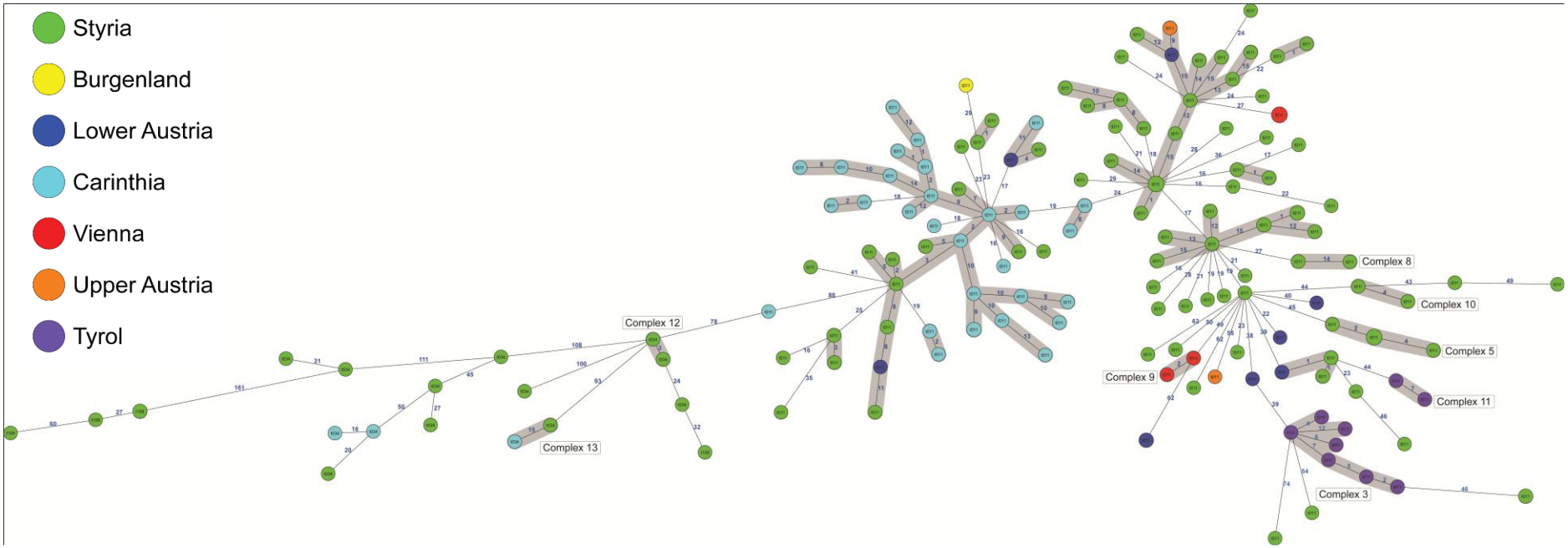

Figure 7. Minimum spanning tree for LA-MRSA isolates based on the pgMLST of *S. aureus*. Colours correspond to the country of origin in Austria. Each circle represents isolates with an allelic profile based on the sequence of the core (1862 alleles) and accessory (706 alleles) genome. Blue numbers accord to the allelic differences between two isolates.

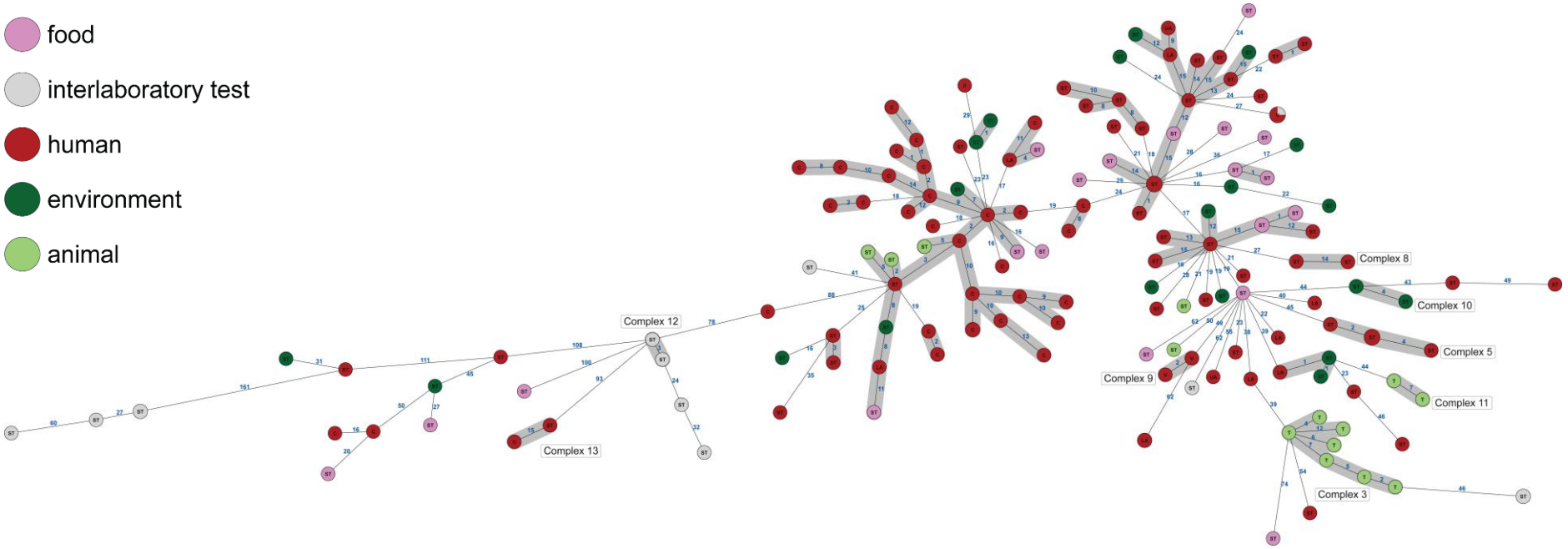

Figure 8. Minimum spanning tree for LA-MRSA isolates based on the pgMLST of *S. aureus*. Colours correspond to the sample material of every isolate. Each circle represents isolates with an allelic profile based on the sequence of the core (1862 alleles) and accessory (706 alleles) genome. Blue numbers accord to the allelic differences between two isolates. B=Burgenland, C=Carinthia, LA= Lower Austria, ST=Styria, T=Tyrol, UA= Upper Austria, V=Vienna

4.2.1.5 Investigation of LA-MRSA transmission chains by NGS

A minimum spanning tree comprising 32 LA-MRSA isolates including a cluster of six *spa*-type t011 isolates involved in a nosocomial transmission from 2010 to 2011 was calculated based on cgMLST (Figure 9). Cluster analysis distinguished four different complexes. The majority of isolates (23) belonged to complex 1. Complex 1-3 comprised isolates with *spa*-type t011, complex 4 comprised isolates with *spa*-type t034. The *spa*-type t011 (case II) and *spa*-type t034 (case 23) isolates differed in at least 87 alleles. Further in complex 1 human-to-human transmission are likely for patient case 17 and case 18, which are family members, and show an allelic difference of six.

Retrospective analysis (Lepuschitz et al., 2016) of the first documented nosocomial transmission of *spa*-type t011 isolates in Austria (Figure 10) revealed that case II-isolate differed in at least 73 alleles from case V- and case VI-isolate excluding cluster case II from this outbreak. Transmission from case III to case IV was confirmed by the NGS findings and from case III to case V newly identified, as it was not indicated by epidemiological findings. The case III-, IV- and V-isolates differed from each other in three alleles. Case V-isolate differed from the case VI-isolate (nurse-case) in 12 alleles, which falsify former assumption of a case V to case VI transmission. Case VI-isolate and case VII-isolate had identical allele profiles indicating strongly the transmission between these two cases. The closest related case VI isolate is from case 2 from another hospital within the same city. Since there was no obvious epidemiological link, no further investigations to former transmissions were done. Analysis of nosocomial transmission of *spa*-type t011 isolates is shown in Table 2.

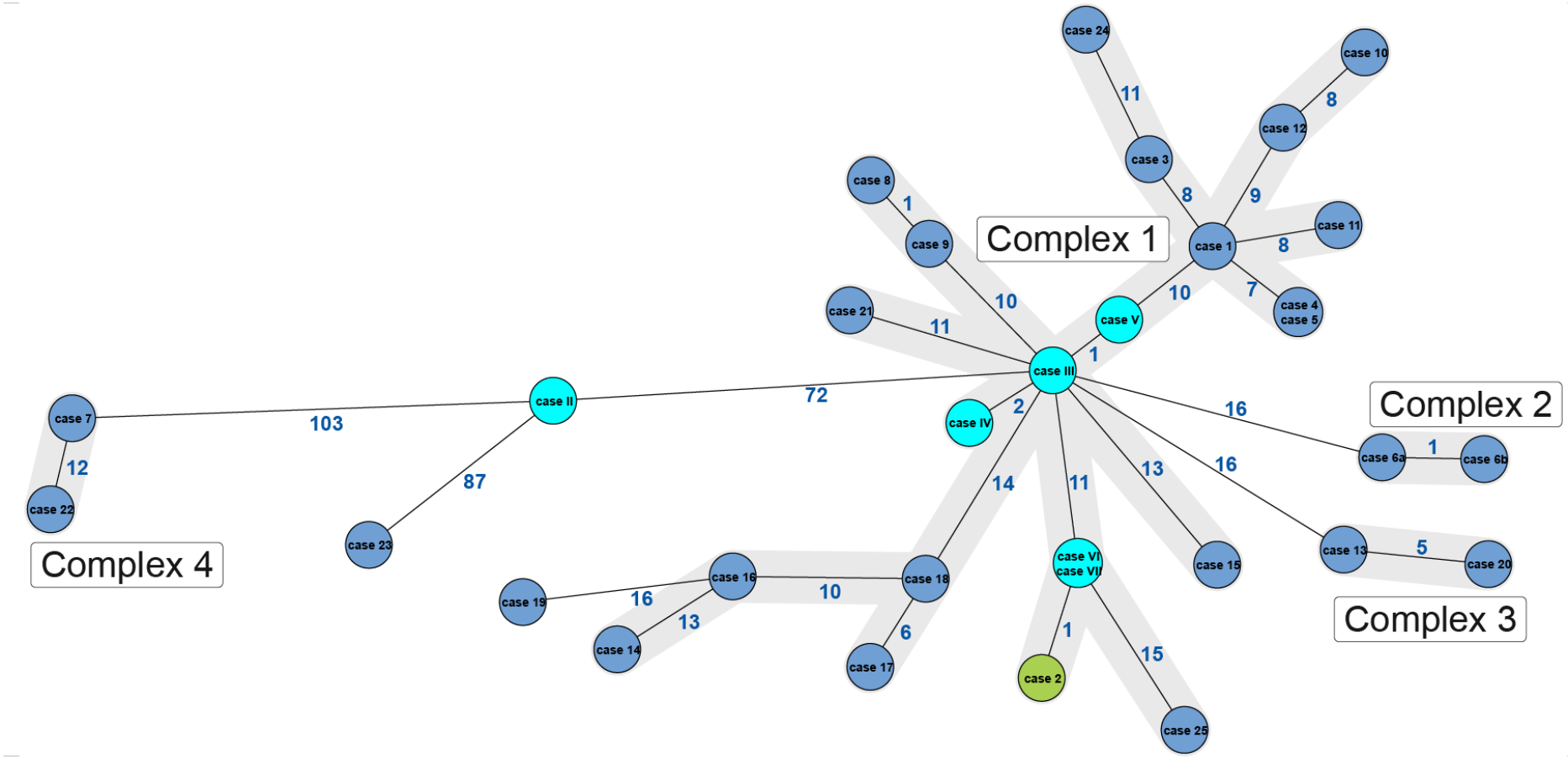

Figure 9. Minimum spanning tree for LA-MRSA isolates from Carinthia based on the cgMLST of *S. aureus*. Each circle represents isolates with an allelic profile based on the sequence of the core genome which consists out of 1862 alleles. Blue numbers accord to the allelic differences between two isolates. Isolates with closely related genotypes were identified with a maximum of 15 allelic differences and are shaded in grey.

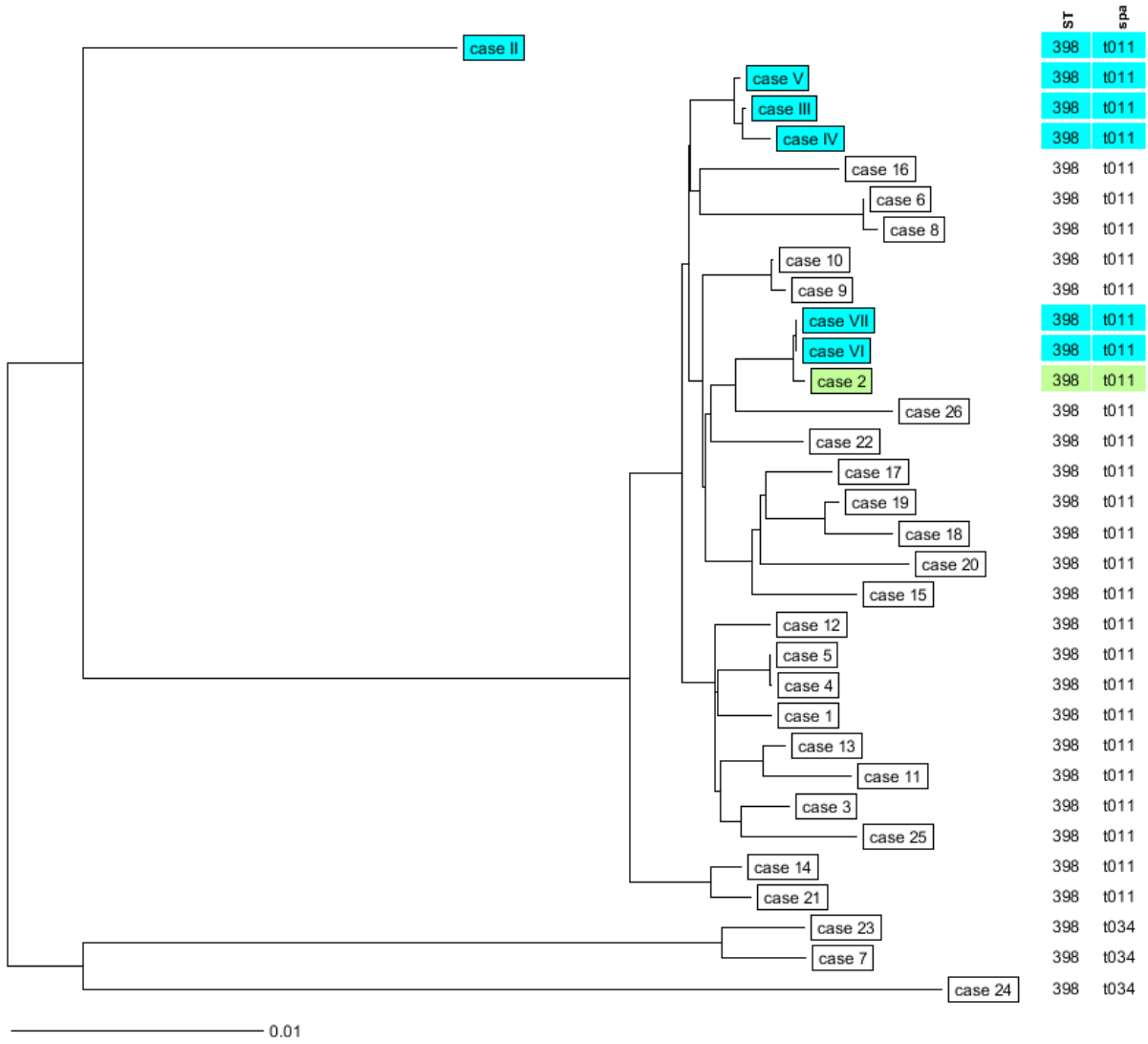

Figure 10. Phylogenetic tree for LA-MRSA isolates from Carinthia based on the cgMLST of *S. aureus*. Each square represents isolates with an allelic profile based on the sequence of the core genome which consists out of 1862 alleles.

Table 2. Analysis of nosocomial transmission of seven *spa*-type t011 isolates

Case (F/M; pat/nurse)	Date of MRSA detection	Site of detection	Department of stay	Contact pattern	NGS defined contact pattern	Investigation of pig farm
II (M pat)	10.8.10	nasopharynx	Thoracic surgery	Contact to pig farm	excluded	refused
III (M pat)	15.11.10	chronic skin lesion nasopharynx	Neurology	Contact to pig farm	Contact to pig farm	positive for t011
IV (M pat)	16.11.10	nasopharynx	Neurology	Contact to case III	Contact to case III	n.a
V (M pat)	27.12.10	chronic skin lesion nasopharynx	Thoracic surgery	Contact to case II (via ward)	Contact to case III	n.a
VI (M nurse)	04.01.11	nasopharynx	Thoracic surgery	Contact (direct) to case V	Contact to case VII Contact to case 2?	n.a
VII (F pat)	05.01.11	nasopharynx	Thoracic surgery	Contact to case II (via ward)	Contact to case VI	n.a

Further investigations for a subset of isolates, including the cluster from 2010-2011 (case II-VII) and case 2, two different preparations of one isolate (case 6a and 6b) and two similar isolates (case 8 and 9), which had no identified epidemiological link, were done. Minimum spanning trees (Figure 11) were calculated based on the cgMLST, based on the pgMLST, and based on SNP analysis. Cluster 2010-2011 isolates including case 2 isolate revealed identical results for cgMLST and pgMLST analysis. Except case III which differed in one allele from case V by cgMLST and in two alleles by pgMLST. The allelic difference from case II to case III increased from 72 cgMLST alleles to 91 pgMLST alleles and to 180 SNPs and showed an increased separation from this outbreak cluster.

Two preparations of isolate 6 (case 6a and case 6b) as well as case 8 and case 9 isolates differed by one allele in their core genome, by two alleles in the pgMLST, and by 27 SNPs respectively 17 SNPs due to recombination events (Figure 12 and Figure 13). Case 6a and case 6b isolates differed by 16 SNPs in SACOL0653 (accessory genome) and by 11 SNPs in SACOL1362 (core genome). Case 8 and case 9 isolates differed by 16 SNPs in SACOL0653 (accessory genome) and 1 SNP in SACOL0806 (core genome).

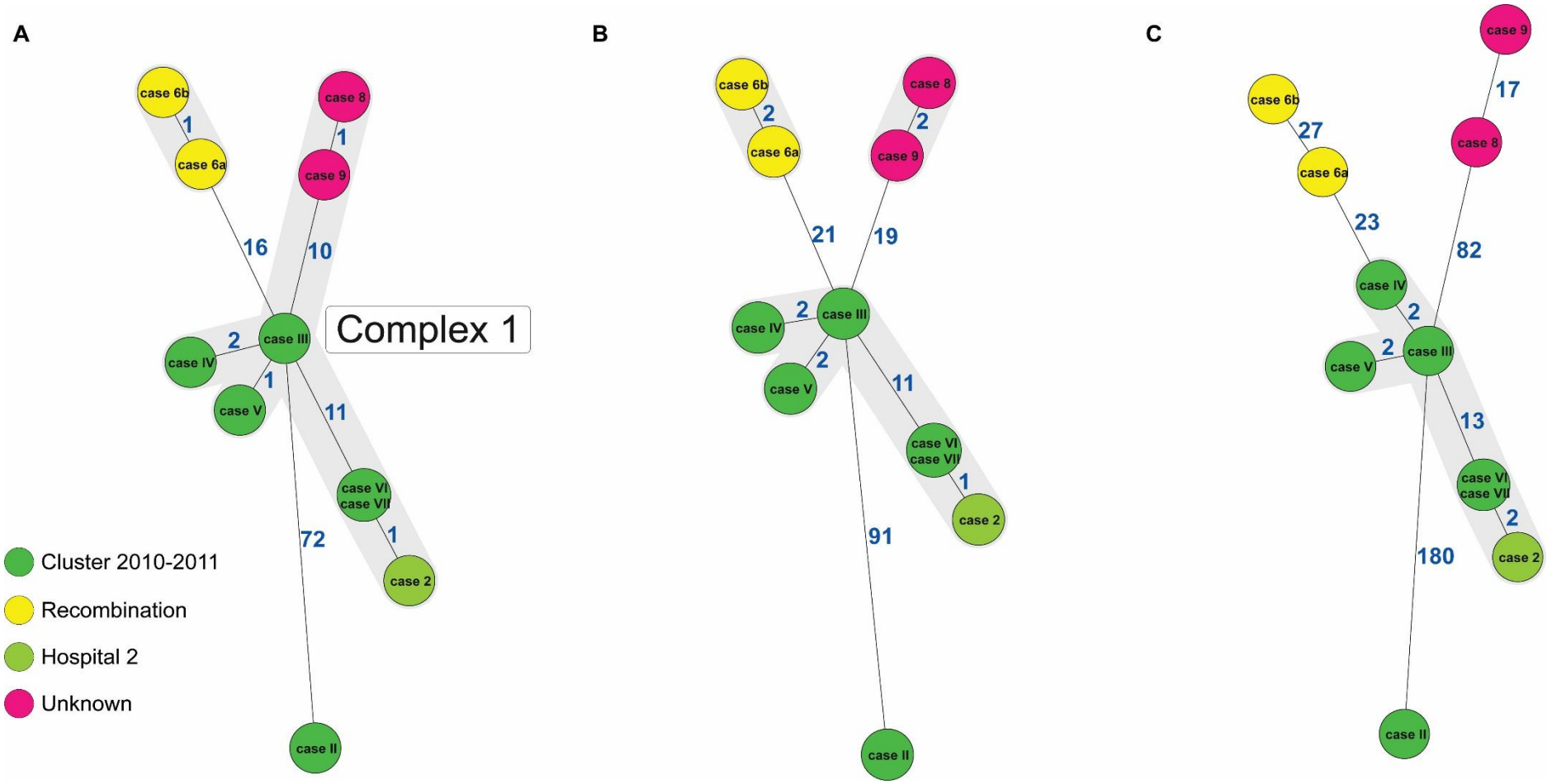

Figure 11. Minimum spanning trees for a subset of LA-MRSA isolates from Carinthia calculated for cgMLST, pgMLST and SNP analysis. Blue numbers accord to the allelic differences between two isolates. A= cgMLST (1862 alleles), B= pgMLST (core (1862 alleles) and accessory (706 alleles), C= SNP (core (1862 alleles) and accessory (706 alleles).

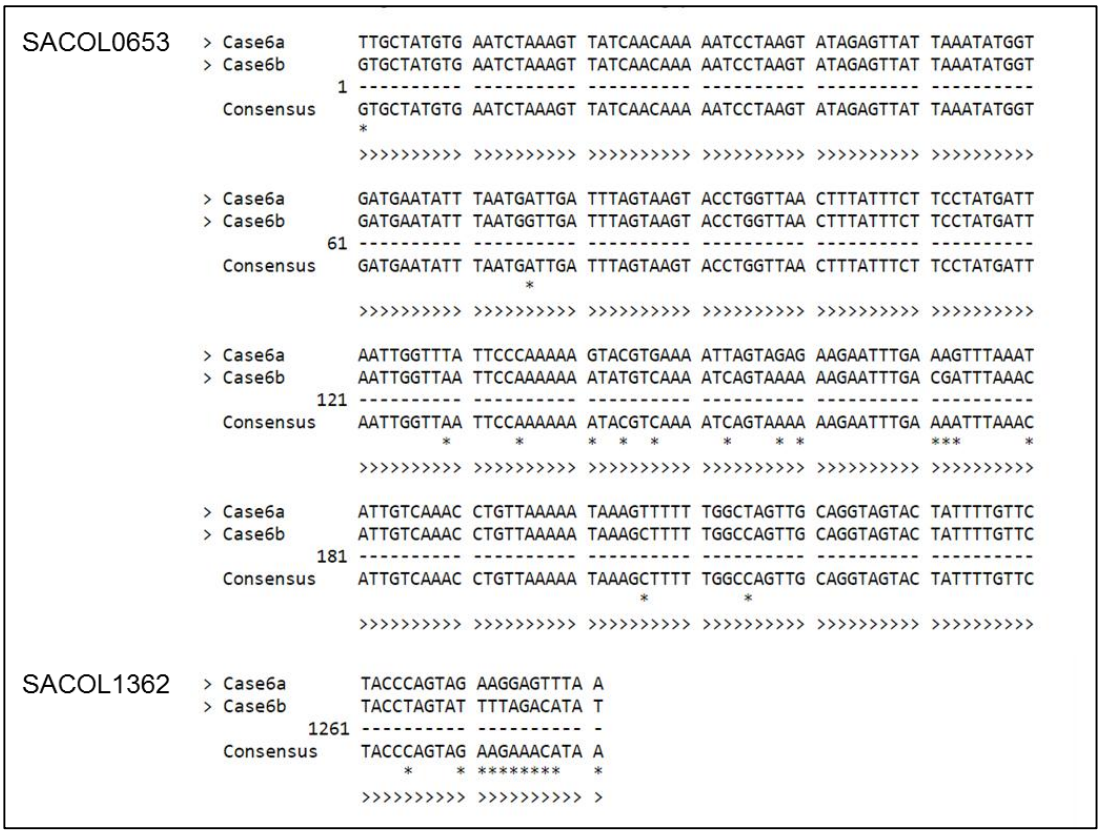

Figure 12. Sequence alignment for SACOL0653 and SACOL1362 of case 6a and case 6b. Sequence differences are marked by asterisks.

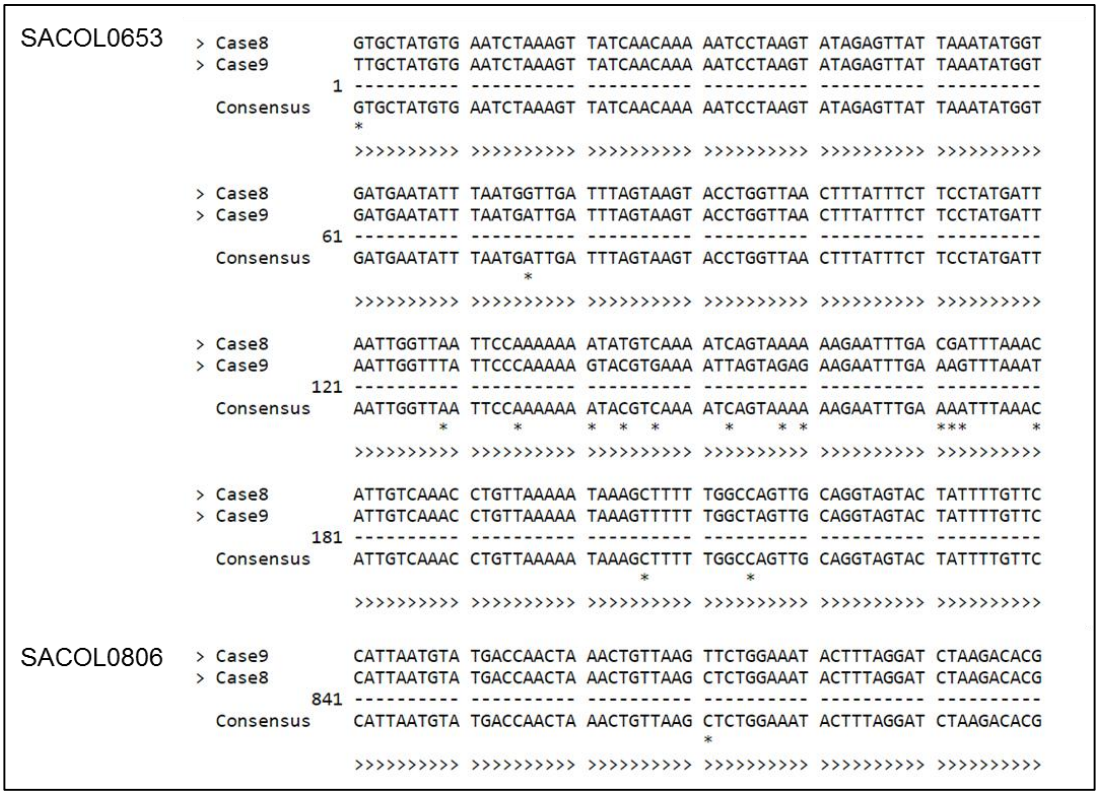

Figure 13. Sequence alignment for SACOL0653 and SACOL0806 of case 8 and case 9. Sequence differences are marked by asterisks.

4.2.2 *Spa*-type, PVL, *femA*, MLST

Spa-type, PVL, *femA* and MLST were extracted from NGS data (Supplemental file 1).

4.2.2.1 LA-MRSA

Identification of the *spa*-type by NGS and analysis in SeqSphere^{+} confirmed the previous Sanger sequencing results (t011 (n=134), t034 (14), t108 (5)). All isolates were negative for PVL. 148 LA-MRSA isolates harboured *femA* (SACOL1410) and *femB* (SACOL1411), two isolates were negative for *femA* and *femB*, two were negative for *femB*, and a single isolate was negative for *femA*. The most prevalent multilocus sequence type (MLST) was ST398 (94.8%, 145/153). One isolate (t011) belonged to ST1744. Seven isolates had new sequence types (ST3226 (n=1), ST3227 (n=2)), three STs of four isolates are still waiting for curation (Table 3).

Table 3. New detected sequence types in seven LA-MRSA isolates

***spa*-type**	**arcC**	**aroE**	**glpF**	**gmk**	**pta**	**tpi**	**yqil**	**ST**
t011, t034, t108	3	35	19	2	20	26	39	398
t011	195	35	19	2	20	26	39	1744
t011	3	35	19	2	20	352*****	39	3226 (n=1)
t011	3	463*****	19	2	20	26	39	3227 (n=2)
t011	x*****	35	19	2	20	26	39	STX1 (n=2)
t011	x*****	35	19	2	20	26	39	STX2 (n=1)
t011	3	35	19	2	20	26	x*****	STX3 (n=1)

For each allele a particular allele number was analysed and the resulting allelic profile produced the specific sequence type (ST). New allele numbers are marked with an asterisk. (n= number of isolates)

4.2.2.2 HA-MRSA

Forty-seven HA-MRSA isolates were selected and analysed by NGS. Identification of the *spa*-type confirmed the previous Sanger sequencing results and sixteen different MLSTs were found (Table 4). One *spa*-type t044 isolate and two *spa*-type t008 isolates were positive for PVL. The detection of *femA* (SACOL1410) and *femB* (SACOL1411) was positive for 45 isolates. One isolate was negative for *femA* (t005) and one for *femB* (t005).

Table 4. Analysis of *spa*-type and MLST for 47 HA-MRSA isolates

MLST	*spa*-type	No. of isolates
ST1	t127	1
ST5	t002 (n=5), t008	6
ST7	t091	3
ST8	t008	2
ST15	t084, t085 (n=2), t2216	4
ST22	t005 (n=8), t032, t449, t13492	11
ST30	t012, t019, t318 (n=2), t789, t1626 (n=2)	7
ST45	t015, t583	2
ST59	t216	1
ST80	t044	1
ST101	t056	1
ST121	t162, t272 (n=2), t284	4
ST146	t002	1
ST225	t003	1
ST508	t026	1
ST3328	t003	1

(n= number of isolates)

4.2.3 Antibiotic resistance genes

4.2.3.1 LA-MRSA

Several genes encoding resistance to antimicrobial agents were identified (Table 5). Out of 153 isolates 144 carried Penicillin binding Proteins (PBP) 1-4, one isolate (t011) had no PBP1, four isolates (t011 (n=2)), t034 (n=2)) had no PBP2, two isolates (t011) had no PBP3 and two isolates (t011) had no PBP4.
Identification of the *mecA* region (SACOL0033) revealed a positive result for 149 isolates. Four isolates (t011 (n=2), t034 (n=2)) were *mecA* negativ. All except two isolates (t011) carried tetracycline resistance (SACOL0122).

Analysis via self-made resistance task templates showed that two isolates (t034) had penicillin resistance encoded by *blaZ* and two isolates (t011) had new alleles for β-lactamases. Erythromycin resistance genes encoded by *ermA* were found in two isolates (t011, t034), *ermB* in thirteen isolates (t011), *ermC* in twenty-two isolates (t011 (n=21), t034 (n=1)) and for seven isolates (t011 (n=2), t034 (n=5)) new allele types were found.
Aminoglycoside resistance genes were present in 51 isolates. Thirty-two isolates (t011 (n=27), t034 (n=4), t108 (n=1)) carried streptomycin resistance, six (t034) spectinomycin resistance, four (t011) gentamycin and kanamycin resistance, two (t034) streptothricin resistance and eight (t011 (n=5), t108 (n=3)) had new allele types. Trimethoprim resistance was detected in 33 isolates (t011 (n=21), t034 (n=11), t108 (n=1)). For 106 isolates (t011 (n=21), t034 (n=13), t108 (n=5)) new alleles were found for quinolone resistance. All 153 isolates had no genes encoding fosfomycin, fusidic acid, nitromidazole, oxazolidinone, phenicol, rifampicin, sulphonamide and vancomycin.

Table 5. Genotypic resistance profiles of 153 LA-MRSA isolates

Antibiotics	t011 (n)	t034 (n)	t108 (n)
aminoglycoside, macrolide, quinolone, tetracycline, trimethoprim, methicillin	4	5	-
aminoglycoside, macrolide, quinolone, tetracycline, trimethoprim, penicillin	-	1	-
aminoglycoside, macrolide, quinolone, tetracycline, penicillin	-	1	-
aminoglycoside, quinolone, tetracycline, trimethoprim, methicillin	9	6	1
aminoglycoside, macrolide, quinolone, tetracycline, methicillin	14	-	-
macrolide, quinolone, tetracycline, trimethoprim, methicillin	3	-	-
macrolide, quinolone, tetracycline, methicillin	17	1	-
aminoglycoside, quinolone, tetracycline, methicillin	7	-	4
quinolone, tetracycline, trimethoprim, methicillin	3	1	-
quinolone, tetracycline, methicillin	28	-	-
tetracycline, trimethoprim, methicillin	1	-	-
tetracycline, methicillin	45	-	-
methicillin	2	-	-

(n= number of isolates)

4.2.3.2 Phenotypic and genotypic resistance test results

Analysis of the disk diffusion test was carried out for 34 arbitrarily chosen isolates (t011 (n=31), t034 (n=2), t108 (n=1)) for comparison with NGS performance. All of this isolates except one (t034) were positive for methicillin. Nineteen (t011 (n=18), t108 (n=1)) were resistant against tetracycline and twelve (t011 (n=10), t034 (n=2)) were resistant against erythromycin, clindamycin and tetracycline. Two isolates (t011) were resistant against gentamicin and tetracycline, one isolate (t011) against erythromycin and tetracycline. All isolates were sensitive for trimethoprim/ sulfamethoxazol.

Comparison of the disk diffusion test with the NGS typing using nine antibiotics (Table 6) revealed matching results for six isolates (No. 2-4, 6, 23, 28) for methicillin, macrolides (M), aminoglycosides (A), trimethoprim (T), quinolones (Q), tetracycline

(Tet), clindamycin (C), penicillin (P), and linconsamide (L). The remaining 28 isolates showed different resistance results against one to six antibiotics. Fourteen of these 28 isolates showed matching resistance results against two antibiotics and the remaining 14 isolates had matching resistance results against one antibiotic (Table 6).

Thirty-three isolates were phenotypically resistant against methicillin and carried *mecA.* One methicillin sensitive isolate (t034) was *mecA* negative but carried a penicillin resistance gene (*blaZ*). Tetracycline resistance was found for all isolates using both typing methods.

Table 6. Comparison of antibiotic resistance typing methods

No.	sample-ID	*spa*-type	Phenotypic and genotypic resistance		Other phenotypic resistance	Other genotypic resistance
			methicillin	tetracycline		
1	**ha1726**	t011	+/+	+/+	M, L	M, Q
2	**WG1**	t011	+/+	+/+	n.d.	n.d.
3	**WG2**	t011	+/+	+/+	n.d.	n.d.
4	**WG3**	t011	+/+	+/+	n.d.	n.d.
5	**201305538**	t011	+/+	+/+	n.d.	Q
6	**802618**	t011	+/+	+/+	n.d.	n.d.
7	**ab3552**	t011	+/+	+/+	M, L	A, M, Q
8	**at3345**	t011	+/+	+/+	M, L	A, M, Q
9	**201312475**	t011	+/+	+/+	A	A, M, Q
10	**150091**	t011	+/+	+/+	n.d.	Q, T
11	**150040**	t011	+/+	+/+	n.d.	Q
12	**150017**	t011	+/+	+/+	n.d.	Q
13	**150063**	t011	+/+	+/+	M, L	M, Q
14	**150035**	t011	+/+	+/+	M, L	A, M, Q
15	**150112**	t011	+/+	+/+	M, L	A, M, Q
16	**150113**	t011	+/+	+/+	M, L	A, M, Q
17	**150138**	t011	+/+	+/+	M, L	M, Q
18	**150111**	t011	+/+	+/+	M, L	A, M, Q, T
19	**150127**	t011	+/+	+/+	n.d.	A, Q
20	**150132**	t011	+/+	+/+	M, L	A, M, Q
21	**150144**	t011	+/+	+/+	n.d.	Q
22	**150157**	t011	+/+	+/+	n.d.	Q
23	**61213**	t011	+/+	+/+	n.d.	n.d.
24	**ST50030**	t011	+/+	+/+	n.d.	Q
25	**54843**	t011	+/+	+/+	n.d.	A, Q
26	**MRSA 13**	t011	+/+	+/+	n.d.	Q
27	**ST50086**	t011	+/+	+/+	M	M, Q
28	**31835**	t011	+/+	+/+	n.d.	n.d.
29	**ST50042**	t011	+/+	+/+	A	A, Q, T
30	**ST50039**	t011	+/+	+/+	n.d.	A, Q, T
31	**84/08**	t011	+/+	+/+	n.d.	A, Q, T
32	**ST50034**	t034	-/(*blaZ*)	+/+	M, L	A, M, Q, P
33	**MRSA 178**	t034	+/+	+/+	M, L	A, Q, T
34	**EQAS 7.8**	t108	+/+	+/+	n.d.	A, Q

(n.d= not detected; M= macrolides; A= aminoglycosides; T= trimethoprim; Q= quinolones; Tet= tetracycline; C= clindamycin; P= penicillin; L= linconsamide)

4.2.3.3 HA-MRSA

Several genes encoding resistance to antimicrobial agents were identified. Penicillin binding proteins PBP1-PBP4 were positive in 46 out of 47 isolates.
The *mecA* region of the reference strain *S. aureus* COL revealed a positive result for 20 isolates (42,6%), eight isolates with new allele types and twelve with *blaZ*. All isolates carried a tetracycline resistance.
Out of 47 isolates all except one (t127) carried a gene encoding for quinolone resistance. Five isolates (t003 (n=2), t008, t012, t13492) were resistant against aminoglycoside and macrolide, three isolates (t002) were resistant against trimethoprim, two (t272) against aminoglycoside, macrolide and trimethoprim, two (t008, t127) against macrolide and one (t044) against fusidic acid.

4.2.4 Virulence genes

4.2.4.1 LA-MRSA

Particular alleles influencing the virulence of *S. aureus* were identified, thereby genes encoding gamma-hemolysin (*hlgA*, *hlgB*, *hlgC*) were found in 143 isolates (93.5%). All isolates had *hlgB* (SACOL2422), 151 isolates had *hlgA* (SACOL2419) and 145 had *hlgC* (SACOL2421).
Leukotoxins *lukD* (SACOL1880) and *lukE* (SACOL1881) were absent in all isolates.
With the exception of two t011 isolates and one t034 isolate all isolates harboured *agrA* (SACOL2026). Eleven isolates had no *agrB* (SACOL2023) (t011 (n=10), t108 (n=1)), 48 isolates had no *agrC* (SACOL2025) (t011 (n=43), t034 (n=5)), two had no *agrD* (SACOL2024) (t011), one isolate (t011) was negative for all agr groups, twelve (t011) were negative for *argB*, *argC*, *argD*, two (t011) were negative for *agrB* and *agrD*, three (t011) were negative for *agrB* and *agrC* and one isolate (t011) was negative for *agrC* and agrD.
One-hundred and forty isolates carried a putative *eta* (exofoliative toxin A, SACOL1184).
One-hundred and forty-four isolates had a putative *sea* locus (SACOL1657). Typing with the self-made task template showed that one isolate (t011) carried *atl*.

4.2.4.2 HA-MRSA

All isolates carried *hlgB*, all but two isolates (t085, t13492) carried *hlgA,* and all but one isolate (t1626) carried *hlgC*.

Identification of leukotoxins *lukD* (SACOL1880) and *lukE* (SACOL1881) revealed twenty-three isolates which were positive for *lukD* and *lukE*, and two isolates which were positive for *lukD*.

Forty-seven isolates carried *eta* (SACOL1184), 35 isolates carried *eta* and *atl* (SACOL1062).

Genes encoding for enterotoxin *sea* could be found in 31 isolates. Three isolates carried *seb*, one carried *sek*, one isolate carried *seb* and *sei,* and one isolate carried *sei* and *sek*.

5. Discussion

Since its emergence in 2003 and the first occurrence 2004 in Austria a steady increase of LA-MRSA CC398 has been observed at the NRLS (Springer et al., 2009). In this study a total of 153 LA-MRSA isolates comprising different *spa*-types (t011 (n=134), t034 (n=14), t108 (n=5)) were examined to determine differences between these isolates by next generation sequencing.

The costs and the expenditure of time for NGS have fallen in the last years and a huge amount of data and information can be obtained. NGS enables the better understanding of molecular epidemiology and evolution of isolates.

As described for other countries the majority of the investigated LA-MRSA CC398 collection belonged to ST398 (Bosch et al., 2013; Chroboczek et al., 2013; Dahms et al., 2014; McCarthy et al., 2012; Monaco et al., 2013; van Wamel et al., 2010). Among our collection with nearly 90% *spa*-type t011 was the most prevalent subtype. Seven isolates with *spa*-type t011 had new sequence types – all CC398 - and were submitted to the *S. aureus* MLST database (http://saureus.mlst.net/).

In contrast to *spa*-typing and MLST, which are inappropriate to assess the transmission of LA-MRSA isolates (Bosch et al., 2013; Price et al., 2012), NGS based typing using a gene-by-gene approach (Leopold et al., 2014; Maiden et al., 2013) allowed the differentiation of all LA-MRSA t011 isolates as well as a clear separation between isolates of related *spa*-types t034 and t108. To improve the resolution for closely related isolates analysis can be extended from the core to the pan genome. A disadvantage of pan genome based typing is that the essential definition of a cluster threshold remains difficult or is impossible, due to the variability of the accessory genome.

In addition, NGS successfully verified *spa*-typing results as determined by Sanger sequencing although repeat regions are challenging for current 3rd generation sequencing technologies due to short read lengths. This drawback has been solved by the development and introduction of 4th generation sequencing technologies generating longer read lengths. Thus, the high resolution of NGS based cgMLST allows now the precise differentiation of isolates even within the same *spa*-type. In my study all isolates but 19 sharing the identical allelic profile were differentiable by

cgMLST analysis. Due to the defined CT 112 isolates could be assigned to 15 different complexes and 41 isolates were singletons not assignable to any complex.

Five complexes harboured exclusively isolates from humans, two complexes harboured exclusively isolates from animals, one complex harboured exclusively environmental samples. Three complexes harboured human and environmental isolates and the two major complexes contained isolates from animals, environment, food and human suggesting that these are widely distributed types. On average inter-laboratory test samples from other countries differed significantly from Austrian isolates. Distinct complexes were obtained from horse samples from Tyrol and patient samples from Carinthia. In contrast patient isolates from Burgenland, Lower- Austria, Upper-Austria and Vienna were not assignable to specific complexes.

Samples from Styria, which represented the majority of samples in this study, originated from different isolation sources (animal, environment, food, human). The similarity of isolates from different sources indicate human to human, animal to human, food to human and dust to human transmissions. High similarity between food and environmental samples indicate the source of contamination. Dust from pig farms is also a source for colonisation of unexposed patients (Monaco et al., 2013).

Unfortunately little epidemiological data were available to confirm a possible linkage of similar samples demonstrating that despite the high resolution of NGS epidemiological data are still essential.

Nevertheless, the superiority of NGS based typing could be demonstrated by re-analysis of an LA-MRSA cluster in a hospital from 2010-2011 originally investigated by *spa*-typing and epidemiological data analysis (Schmid et al., 2012). The re-investigation of epidemiological and *spa*-typing results by NGS revealed new knowledge about transmission pathways. First, NGS revealed that the patient previously identified as the origin of this outbreak is unrelated to all other patients of this cluster. Second, this former cluster definition was false because NGS identified two separate clusters and therefore at least two distinct transmission routes of LA-MRSA into this hospital. Third, NGS identified an additional patient with a nearly identical cgMLST profile. This patient was epidemiologically excluded from the former cluster investigation because of his stay in another hospital. However, due to the high similarity of these two isolates a contact of these two patients must be considered. But, as described above, without epidemiological data this assumption can’t be confirmed.

Although in principle the differences are justified due to the core genome in which circa 75% of genes are present in all strains, the accessory genome is more variable and may provide more information about its relative diversity (Shittu et al., 2007).
Therefore, the cluster was analysed by inclusion of the accessory genome i.e. pgMLST. The way of transmission could be confirmed through the existence of still high related samples.

Finally, this cluster was investigated by SNP calling to obtain the highest possible resolution. SNP analysis is the best way to reproduce genetic distances between two samples to determine if isolates belong to the same outbreak, cluster or if they are responsible for transmission events (Worby et al., 2014). The expected SNP rate for MRSA is one nucleotide change per 6 weeks, whereby this assumption depends on the evolution of the particular pathogen (Leopold et al., 2014). SNP analysis for the pan genome allowed a clear separation of cluster isolates. But defining a threshold for complex investigations is difficult, due to the lack of standardized procedures. But for investigations on similar samples this methods allows to show a precise differentiation.
I observed a recombination event in two different preparations of the same isolate. The two genomes showed two allelic differences of their pan genomes but differed in 27 SNPs. A second recombination event was observed between case 8 and case 9 isolates.
So in principle correct results were obtained by different analytical methods. But in this example and as described above the SNP based approach reflects not the correct genetic distance between samples. In addition allele based typing compared to SNP typing has two major advantages, namely that SNP and single recombination events are correctly treated as one evolutionary event, and that an allelic based standardized nomenclature such as for classical MLST can be used (Leopold et al., 2014).

Analysis of a HA-MRSA collection from one hospital by next generation sequencing revealed a high distance to LA-MRSA isolates as well as a high allelic difference between isolates of different HA-MRSA lineages.
NGS based on cgMLST clustered HA-MRSA isolates according to classical MLST clonal complex (CC) or sequence type (ST). HA-MRSA isolates with identical *spa*-types are differentiable by cgMLST as already demonstrated for LA-MRSA isolates.

Core genome MLST made it possible to show a linkage between samples, which were before assumed to be unrelated due to different *spa*-types.
Further NGS typing identified two family outbreaks consisting out of two and accordingly three persons. Both families showed an identical cgMLST subtype.

Although LA-MRSA are described as less virulent lacking most *S. aureus* associated virulence factors (Argudin et al., 2011), the emergence of more virulent LA-MRSA clones is a matter of time. Therefore all LA-MRSA isolates were screened for virulence factors to determine the current virulence status. To overcome the limited variations and lack of alleles of the reference strain SACOL an additional virulence task template was created. The constructed virulence task template facilitates surveillance of steady transforming populations by fast detection of virulence genes, representing an important step forward in replacing conventional testing methods by NGS (Price et al., 2013).
The high antibiotic resistances in LA-MRSA stand in contrast to low virulence which is a consequence of the common use of antibiotics in livestock production (Argudin et al., 2011). Analysis of the resistance genes in LA-MRSA with the reference strain revealed typical results, such as methicillin and tetracycline resistance (Argudin et al., 2011). Due to the fact that most MRSA isolates are multiresistant for various antibiotics a task template containing sequences encoding antibiotic resistant genes was created. This allowed the identification of new alleles and detection of further resistance genes improving the usefulness of NGS in diagnostics and epidemiological surveillance (Zankari et al., 2012).

In conclusion NGS based typing allows the differentiation within the highly clonal LA-MRSA population in Austria as determined by *spa*-typing. This improvement in typing will now allow to monitor the evolution of strains, the identification of potential transmission routes, the identification of outbreaks and the identification of resistance and virulence determinants and as a consequence the emergence of more virulent clones (Gordon et al., 2014).
Thus, the application of NGS based typing is highly recommended for the public health sector to improve MRSA surveillance. NGS allows a new view and a more focused investigation of transmission as well as outbreak investigation and will therefore improve public health actions in general.

6. References

Allerberger, F., Brand, J., Neubauer, T., and Ruppitsch, W. (2013). Bakterielle Antibiotikaresistenzen in der Tierhaltung. Klinik *5*.

Argudin, M., Tenhagen, B., Fetsch, A., Sachsenroder, J., Kasbohrer, A., Schroeter, A., Hammerl, J., Hertwig, S., Helmuth, R., Braunig, J.*, et al.* (2011). Virulence and Resistance Determinants of German *Staphylococcus aureus* ST398 Isolates from Nonhuman Sources. Applied and Environmental Microbiology *77*, 3052-3060.

Baker, M. (2012). *De novo* genome assembly: what every biologist should know. Nature Methods *9*, 333-337.

Bosch, T., Verkade, E., van Luit, M., Pot, B., Vauterin, P., Burggrave, R., Savelkoul, P., Kluytmans, J., and Schouls, L. (2013). High Resolution Typing by Whole Genome Mapping Enables Discrimination of LA-MRSA (CC398) Strains and Identification of Transmission Events. PloS one *8*, e66493.

Cai, Y., Kong, F., Wang, Q., Tong, Z., Sintchenko, V., Zeng, X., and Gilbert, G. (2007). Comparison of single- and multilocus sequence typing and toxin gene profiling for characterization of methicillin-resistant *Staphylococcus aureus*. Journal of Clinical Microbiology *45*, 3302-3308.

Chamber, H. (1997). Methicillin Resistance in Staphylococci: Molecular and Biochemical Basis and Clinical Implications. Clinical Microbiology Reviews *10*, 781-791.

Chroboczek, T., Boisset, S., Rasigade, J., Tristan, A., Bes, M., Meugnier, H., Vandenesch, F., Etienne, J., and Laurent, F. (2013). Clonal complex 398 methicillin susceptible *Staphylococcus aureus*: a frequent unspecialized human pathogen with specific phenotypic and genotypic characteristics. PloS one *8* e68462.

Cuny, C., Layer, F., Kock, R., Werner, G., and Witte, W. (2013). Methicillin susceptible *Staphylococcus aureus* (MSSA) of clonal complex CC398, t571 from infections in humans are still rare in Germany. PloS one *8*, e83165.

Cuny, C., Nathaus, R., Layer, F., Strommenger, B., Altmann, D., and Witte, W. (2009). Nasal Colonization of Humans with Methicillin-Resistant *Staphylococcus aureus* (MRSA) CC398 with and without Exposure to Pigs. PloS one *4*, e6800.

Dahms, C., Hubner, N., Cuny, C., and Kramer, A. (2014). Occurrence of methicillin-resistant *Staphylococcus aureus* in farm workers and the livestock environment in Mecklenburg-Western Pomerania, Germany. Acta Veterinaria Scandinavica *56*.

DeLeo, F., Otto, M., Kreiswirth, B., and Chambers, H. (2010). Community-associated meticillin-resistant *Staphylococcus aureus*. The Lancet *375*, 1557-1568.

Devriese, L., Damme, L., and Fameree, L. (1972). Methicillin (Cloxacillin)-Resistant *Staphylococcus aureus* strains Isolated from Bovine Mastitis Cases. Zentralblatt für Veterinärmedizin Reihe B *19*, 598-605.

Eickhoff, T. (1972). Therapy of staphylococcal infection (Wiley, New York).

Elek, S., and Stephen, D. (1956). Experimental staphylococcal infections in the skin of man. Annals of the New York Academy of Sciences *65*, 85-90.

Enright, M., Day, N., Davies, C., Peacock, S., and Spratt, B. (2000). Multilocus Sequence Typing for Characterization of Methicillin-Resistant and Methicillin-Susceptible Clones of *Staphylococcus aureus*. Journal of Clinical Microbiology *38*, 1008–1015.

Feil, E., Cooper, J., Grundmann, H., Robinson, D., Enright, M., Berendt, T., Peacock, S., Smith, J., Murphy, M., Spratt, B.*, et al.* (2003). How clonal is *Staphylococcus aureus*? Journal of Bacteriology *185*, 3307-3316.

Foster, T. (1996). Medical Microbiology 4edn (University of Texas Medical Branch at Galveston).

Gordon, N., Price, J., Cole, K., Everitt, R., Morgan, M., Finney, J., Kearns, A., Pichon, B., Young, B., Wilson, D.*, et al.* (2014). Prediction of *Staphylococcus aureus* Antimicrobial Resistance by Whole-Genome Sequencing. Journal of Clinical Microbiology *52* 1182–1191.

Grisold, A.J., Zarfel, G., Hoenigl, M., Krziwanek, K., Feierl, G., Masoud, L., Leitner, E., Wagner-Eibel, U., Badura, A., and Marth, E. (2010). Occurrence and genotyping using automated repetitive-sequence-based PCR of methicillin-resistant *Staphylococcus aureus* ST398 in Southeast Austria. Diagnostic Microbiology and Infectious Disease *66*, 217-221.

Harmsen, D., Claus, H., Witte, W., Rothganger, J., Claus, H., Turnwald, D., and Vogel, U. (2003). Typing of Methicillin-Resistant *Staphylococcus aureus* in a University Hospital Setting by Using Novel Software for *spa* Repeat Determination and Database Management. Journal of Clinical Microbiology *41*, 5442-5448.

Harris, L., Foster, S., and Richards, R. (2002). An introduction to *Staphylococcus aureus*, and techniques for identifying and quantifying *S. aureus* adhesins in relation to adhesion to biomaterial: review. European Cells and Materials *4*, 39-60.

Hiramatsu, K. (1995). Molecular evolution of MRSA. Microbiology and Immunology *39*, 531-543.

Howard, B., and Kloos, W. (1987). Staphylococci (St Louis. CV Mosby).

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

Jevons, M. (1961). “Celbenin” - resistant Staphylococci. British Medical Journal *1*, 124-125.

Joensen, K., Scheutz, F., Lund, O., Hasman, H., Kaas, R., Nielsen, E., and Aarestrup, F. (2014). Real-Time Whole-Genome Sequencing for Routine Typing, Surveillance, and Outbreak Detection of Verotoxigenic *Escherichia coli*. Journal of Clinical Microbiology *52*, 1501-1510.

Jolley, K., and Maiden, M. (2014). Using multilocus sequence typing to study bacterial variation: prospects in the genomic era. Future Microbiology *9*, 623-630.

Jünemann, S., Sedlazeck, F., Prior, K., Albersmeier, A., John, U., Kalinowski, J., Mellmann, A., Goesmann, A., von Haeseler, A., Stoye, J.*, et al.* (2013). Updating benchtop sequencing performance comparison. Nature Biotechnology *31*, 294-296.

Kloos, W., and Lambe, D. (1991). Staphylococcus, Vol 5 (ASM, Washington, D.C.).

Knox, K., and Wicken, A. (1973). Immunological properties of teichoic acids. Bacteriological Reviews *37*, 215-257.

Kohl, T., Diel, R., Harmsen, D., Rothgänger, J., Walter, K., Merker, M., Weniger, T., and Niemann, S. (2014). Whole-Genome-Based *Mycobacterium tuberculosis* Surveillance: a Standardized, Portable, and Expandable Approach. Journal of Clinical Microbiology *52*, 2479-2486.

Köser, C., Holden, M., Ellington, M., Cartwright, E., Brown, N., Ogilvy-Stuart, A., Hsu, L., Chewapreecha, C., Croucher, N., Harris, S.*, et al.* (2012). Rapid Whole-Genome Sequencing for Investigation of a Neonatal MRSA Outbreak. The New England Journal of Medicine *366*, 2267-2275.

Larsen, J., Petersen, A., Sørum, M., Stegger, M., van Alphen, L., Valentiner-Branth, P., Knudsen, L., Larsen, L., Feingold, B., Price, L.*, et al.* (2015). Meticillin-resistant *Staphylococcus aureus* CC398 is an increasing cause of disease in people with no livestock contact in Denmark, 1999 to 2011. Eurosurveillance *20*.

Leopold, S., Goering, R., Witten, A., Harmsen, D., and Mellmann, A. (2014). Bacterial Whole-Genome Sequencing Revisited: Portable, Scalable, and Standardized Analysis for Typing and Detection of Virulence and Antibiotic Resistance Genes. Journal of Clinical Microbiology *52*, 2365–2370.

Lepuschitz, S., Schmid, D., Zerlauth, U., Springer, B., Indra, A., Allerberger, F., and Ruppitsch, W. (2016). The superiority of NGS in tracing chains of transmission: a retrospective analysis of the first documented nosocomial transmission of LA-MRSA *spa*-type t011 in an Austrian hospital, 2010-2011. In European Congress of Clinical Microbiology and Infectious Diseases (Amsterdam).

Lindsay, J. (2010). Genomic variation and evolution of *Staphylococcus aureus*. International Journal of Medical Microbiology *2*, 98-103.

Loncaric, I., Kunzel, F., Licka, T., Simhofer, H., Spergser, J., and Rosengarten, R. (2014). Identification and characterization of methicillin-resistant *Staphylococcus aureus* (MRSA) from Austrian companion animals and horses. Veterinary Microbiology *168*, 381-387.

Maiden, M., Bygraves, J., Feil, E., Morelli, G., Russell, J., Urwin, R., Zhang, Q., Zhou, J., Zurth, K., Caugant, D.*, et al.* (1998). Multilocus sequence typing: A portable approach to the identification of clones within populations of pathogenic microorganisms. PNAS *95*, 3140-3145.

Maiden, M.C., Jansen van Rensburg, M.J., Bray, J.E., Earle, S., Ford, S., Jolley, K., and McCarthy, N. (2013). MLST revisited: the gene-by-gene approach to bacterial genomics. Nature Reviews Microbiology *11*, 728-736.

McCarthy, A., van Wamel, W., Vandendriessche, S., Larsen, J., Denis, O., Garcia-Graells, C., Uhlemann, A., Lowy, F., Skov, R., and Lindsay, J. (2012). *Staphylococcus aureus* CC398 Clade Associated with Human-to-Human Transmission. Applied and Environmental Microbiology *78*, 8845-8848.

Mellmann, A., Harmsen, D., Cummings, C.A., Zentz, E., Leopold, S., Rico, A., Prior, K., Szczepanowski, R., Ji, Y., Zhang, W.*, et al.* (2011). Prospective Genomic Characterization of the German Enterohemorrhagic *Escherichia coli* O104:H4 Outbreak by Rapid Next Generation Sequencing Technology. PloS one *6*, e22751.

Monaco, M., Pedroni, A., Sanchini, A., Bonomini, A., Indelicato, A., and Pantosti, A. (2013). Livestock-associated methicillin-resistant *Staphylococcus aureus* responsible for human

colonization and infection in an area of Italy with high density of pig farming. BMC Infectious Diseases *13*.

Morris, A., Kellner, J., and Low, D. (1998). The superbugs: evolution, dissemination and fitness. Current Opinion in Microbiology *1*, 524-529.

Ogston, A. (1882). Micrococcus Poisoning. Journal of Anatomy and Physiology *17*, 24-58.

Platt, A., Woodhall, R., and George Jr, A. (2007). Improved DNA sequencing quality and efficiency using an optimized fast cycle sequencing protocol. BioTechniques *43*, 58-62.

Price, J., Gordon, N., Crook, D., Llewelyn, M., and Paul, J. (2013). The usefulness of whole genome sequencing in the management of *Staphylococcus aureus* infections. Clinical Microbiology and Infection *19*, 784-789.

Price, L., Stegger, M., Hasman, H., Aziz, M., Larsen, J., Andersen, P., Pearson, T., Waters, A., Foster, J., Schupp, J*., et al.* (2012). *Staphylococcus aureus* CC398: Host Adaptation and Emergence of Methicillin Resistance in Livestock. mBio *3*, e00305-00311.

Projan, S., and Novick, R. (1997). The molecular basis of pathogenicity (Churchill Livingstone, London).

Robinson, D., and Enright, M. (2003). Evolutionary Models of the Emergence of Methicillin-Resistant *Staphylococcus aureus*. Antimicrobial Agents and Chemotherapy *47*, 3926-3934.

Ruppitsch, W., Pietzka, A., Prior, K., Bletz, S., Fernandez, H., Allerberger, F., Harmsen, D., and Mellmann, A. (2015). Defining and Evaluating a Core Genome Multilocus Sequence Typing Scheme for Whole-Genome Sequence-Based Typing of *Listeria monocytogenes*. Journal of Clinical Microbiology *53*, 2869-2876.

Schmid, D., W., R., Orendi, U., Zerlauth, U., and Allerberger, F. (2012). First documented nosocomial transmission of MRSA *spa*-type t011 in an Austrian hospital, 2010-2011. In European Congress of Clinical Microbiology and Infectious Diseases (London).

Schwartz, D., Saffran, W., Welsh, J., Haas, R., Goldenberg, M., and Cantor, C. (1983). New techniques for purifying large DNAs and studying their properties and packaging. Cold Spring Harbor Symposia on Quantitative Biology *47*, 189-195.

Segerman, B. (2012). The genetic integrity of bacterial species: the core genome and the accessory genome, two different stories. Frontiers in Cellular and Infection Microbiology *2*, 1-8.

Shittu, A., Lin, J., and Udo, E. (2007). Insights on Virulence and Antibiotic Resistance: A Review of the Accessory Genome of *Staphylococcus aureus*. Wounds *19*, 237-244.

Shockman, G., and Barrett, J. (1983). Structure, Function, and Assembly of Cell Walls of Gram-Positive Bacteria. Annual Review of Microbiology *37*, 501-527.

Shopsin, B., Gomez, M., Montgomery, S., Smith, D., Waddington, M., Dodge, D., Bost, D., Riehman, M., Naidich, S., and Kreiswirth, B. (1999). Evaluation of protein A gene polymorphic region DNA sequencing for typing of *Staphylococcus aureus* strains. Journal of Clinical Microbiology *37*, 3556-3563.

Shulman, J., and Nahmias, A. (1972). Staphylococcal infections: clinical aspects (Wiley, New York).

Springer, B., Orendi, U., Much, P., Höger, G., Ruppitsch, W., Krziwanek, K., Metz-Gercek, S., and Mittermayer, H. (2009). Methicillin-resistant *Staphylococcus aureus*: a new zoonotic agent? Wiener klinische Wochenschrift *121*, 86-90.

Stegger, M., Wirth, T., Andersen, P., Skov, R., De Grassi, A., Simoes, P., Tristan, A., Petersen, A., Aziz, M., Kiil, K.*, et al.* (2014). Origin and Evolution of European Community-Acquired Methicillin-Resistant *Staphylococcus aureus*. mBio *5*, e01044-01014.

Stöger, A., Gonano, M., Pietzka, A., Allerberger, F., Wagner, M., and Ruppitsch, W. (2007). Comparative molecular analysis of veterinary, dairy, and clinical *Staphylococcus aureus* isolates by *spa* typing and amplification of the mecA and the PVL genes. Clinical Microbiology and Infection *13*, 359-360.

Tacconelli, E., De Angelis, G., Cataldo, M., Pozzi, E., and Cauda, R. (2008). Does antibiotic exposure increase the risk of methicillin-resistant *Staphylococcus aureus* (MRSA) isolation? A systematic review and meta-analysis. Journal of Antimicrobial Chemotherapy *61*, 26-38.

Thakker, M., Park, J., Carey, V., and Lee, J. (1998). *Staphylococcus aureus* Serotype 5 Capsular Polysaccharide is Antiphagocytic and Enhances Bacterial Virulence in a Murine Bacteremia Model. Infection and Immunity *66*, 5183-5189.

Uhlemann, A., Otto, M., Lowy, F., and DeLeo, F. (2014). Evolution of community- and healthcare-associated methicillin-resistant *Staphylococcus aureus*. Infection, Genetics and Evolution *21*, 563-574.

van Cleef, B., Monnet, D., Voss, A., Krziwanek, K., Allerberger, F., Struelens, M., Zemlickova, H., Skov, R., Vuopio-Varkila, J., Cuny, C.*, et al.* (2011). Livestock-associated methicillin-resistant *Staphylococcus aureus* in Humans, Europe. Emerging Infectious Diseases *17*, 502-505.

van Tonder, A., Mistry, S., Bray, J., Hill, D., Cody, A., Farmer, C., Klugman, K., von Gottberg, A., Bentley, S., Parkhill, J.*, et al.* (2014). Defining the Estimated Core Genome of Bacterial Populations Using a Bayesian Decision Model. PLoS Computational Biology *10*, e1003788.

van Wamel, W., Hansenova Manaskova, S., Fluit, A., Verbrugh, H., de Neeling, A., van Duijkeren, E., and van Belkum, A. (2010). Short term micro-evolution and PCR-detection of methicillin-resistant and -susceptible *Staphylococcus aureus* sequence type 398. European Journal of Clinical Microbiology & Infectious Diseases *29*, 119-122.

Vandenesch, F., Naimi, T., Enright, M., Lina, G., Nimmo, G., Heffernan, H., Liassine, N., Bes, M., Greenland, T., Reverdy, M.*, et al.* (2003). Community-Acquired-Methicillin-Resistant *Staphylococcus aureus* Carrying Panton-Valentine Community-Acquired Methicillin-Leukocidin Genes: Worldwide Emergence. Emerging Infectious Diseases *9*, 978-984.

Voss, A., Loeffen, F., Bakker, J., Klaassen, C., and Wulf, M. (2005). Methicillin-resistant *Staphylococcus aureus* in pig farming. Emerging Infectious Diseases *11*, 1965-1966.

Votintseva, A., Fung, R., Miller, R., Knox, K., Godwin, H., Wyllie, D., Bowden, R., Crook, D., and Walker, A. (2014). Prevalence of *Staphylococcus aureus* protein A (*spa*) mutants in the community and hospitals in Oxfordshire. BMC Microbiology *14*.

Waldvogel, F. (1990). *Staphylococcus aureus* (including toxic shock syndrome). In Principles and Practice of Infectious Disease, G. Mandell, R. Douglas, and J. Bennett, eds. (Churchill Livingston, London), pp. 1489-1510.

Wilkinson, B. (1997). The staphylococci in human disease (Churchill Livingston, London).

Worby, C., Chang, H., Hanage, W., and Lipsitch, M. (2014). The Distribution of Pairwise Genetic Distances: A Tool for Investigating Disease Transmission. Genetics *198*, 1395-1404.

Zankari, E., Hasman, H., Cosentino, S., Vestergaard, M., Rasmussen, S., Lund, O., Aarestrup, F., and Larsen, M. (2012). Identification of acquired antimicrobial resistance genes. Journal of Antimicrobial Chemotherapy *67*, 2640-2644.

7. Appendix

Zusammenfassung

Mit dem Beginn des neuen Jahrtausends kam es zum Auftreten eines neuen Methicillin resistenten *Staphylococcus aureus* (MRSA) Stamm, welcher zum klonalen Komplex 398 gehört (CC398). Dieser führt in Menschen, die Kontakt zu Tieren haben, zu Infektionen (Nutztier-assoziierte (LA) - MRSA). In Österreich wurden im Jahr 2004 die ersten Infektionen durch CC398 beobachtet, als das Nationale Referenzlabor für Staphylokokken (NRLS) in der Österreichischen Agentur für Gesundheit und Ernährungssicherheit (AGES) mit der Typisierung von MRSA Isolaten begann.

Von 2004 bis heute beobachtete das NRLS eine stetige Zunahme an Infektionen durch CC398 Isolate von 0.2% im Jahr 2004 auf 7.9% im Jahr 2015.

Alle LA-MRSA Isolate des NRLS gehörten zu CC398 und hatten dazugehörige *spa*-Typen (t011, t034, t108), wobei t011 der vorherrschende *spa*-Typ (87.6%) war. Aufgrund der limitierten Differenzierungsmöglichkeit von konventionellen Subtypisierungsmethoden, wie multilocus sequence typing (MLST) und der Sequenzierung der variablen X-Region des Protein A Gens von *Staphylocuccus aureus* (*spa*-Typisierung), wurde Next Generation Sequencing (NGS) zur hochauflösenden Typisierung des österreichischen LA-MRSA Bestandes angewendet. Insgesamt wurden 153 LA-MRSA Isolate und 47 zufällig gewählte HA-(Krankenhaus-assoziierte) MRSA Isolate mittels NGS untersucht. Dabei wurde eine Gen-für-Gen Vorgehensweise zum Vergleich der Genome auf der Basis eines vordefinierten Core-Genoms (cgMLST) mit 1861 Genen verwendet. Die Typisierung mit NGS ermöglichte eine Unterscheidung zwischen allen LA-MRSA und HA-MRSA Isolaten. HA-MRSA Isolate unterschieden sich von LA-MRSA Isolaten in mindestens 1717 Allelen. Das Maximum an Allel Unterschieden betrug 1658 Allele zwischen HA-MRSA und 131 zwischen LA-MRSA Isolaten. Allgemein war der genetische Unterschied zwischen unterschiedlichen *spa*-Typen größer, als zwischen Isolaten des Selben *spa*-Typs. Damit wurde bestätigt, dass cgMLST sich hervorragend für die hochauflösende Typisierung von LA-MRSA Isolaten desselben *spa*-Typs eignet.

Supplemental files

- Supplemental file 1:
  Table S1 (List of core genome genes used for the *S. aureus* cgMLST scheme)

- Supplemental file 2:
  Table S2 (List of virulence and toxin genes used to create the task template)

- Supplemental file 3:
  Table S3 (List of resistance genes used to create the task template)

All supplemental files can be downloaded at:
http://www.ages.at/service/forschung-kooperationen/publikationssuche/

List of figures

List of tables

Curriculum Vitae

Sarah Lepuschitz, BSc

EDUCATION AND TRAINING

09/2013–Present
Master of Science (MSc) in Molecular Biology with focus on Molecular Medicine
University of Vienna, Vienna (Austria)

09/2009–07/2013
Bachelor of Science (BSc) in Molecular Biology
Graz University of Technology, Graz (Austria)

09/2000–06/2008
Matura, comparable to Abitur (Germany) or A-Levels (England)
Ingeborg-Bachmann-Gymnasium Klagenfurt, Klagenfurt (Austria)

WORK EXPERIENCE

03/2015–12/2015
Master thesis

AGES (Austrian Agency for Health and Food Safety) - Institute for medical Microbiology and Hygiene, Vienna (Austria)

Title: "Subtyping of livestock-associated methicillin-resistant *Staphylococcus aureus* CC398 isolates by next generation sequencing"

Duties: different DNA isolation methods, next generation sequencing, data analysis with SeqSphere+ (Ridom)

Supervisor: PD Mag. Dr. Werner Ruppitsch

08/2014–09/2014
Internship

Medical University Vienna - Skin & Endothelium Research Division (SERD), Vienna (Austria)

Duties: Histochemical staining of human and mouse tissues, antibody testing, Immunofluorescence and Laser Scanning Microscopy, human melanoma cell culture, western blot, endothelial sprouting assays, migration, invasion and proliferation assays, retroviral infection

11/2013–03/2014
Internship

AGES (Austrian Agency for Health and Food Safety) - Institute for medical Microbiology and Hygiene, Vienna (Austria)

Duties: Several molecular methods for subtyping of livestock-associated methicillin-resistant *Staphylococcus aureus*, PCR, antibiotic resistance profiles

07/2012–08/2012 **Bachelor thesis**

National Institution for Veterinary Examinations, Klagenfurt (Austria)

Title: “Comparison of analytical sensitivity of two qPCR methods”

Duties: Conventional and real time PCR, DNA extraction from biological samples, gelelectrophoresis

Supervisor: Mag. Dr. med. vet. Gunter Vogl

08/2011 **Internship**

Klinikum Klagenfurt am Wörthersee - Institute for Pathology and Institute for Laboratory Diagnosis and Microbiology, Klagenfurt (Austria)

Duties: Molecular diagnostics in a histological and cytological laboratory, basics of histochemistry, DNA extraction, PCR, screening of gynaecological cytological smear

08/2010 **Internship**

Carinthian Institute for Food Analysis and Quality Control, Klagenfurt (Austria)

Duties: Quality control for microbiological diagnosis, general laboratory work in molecular diagnostics, Salmonella and EHEC diagnostics by PCR, DNA hybridisation for detection of EHEC, conventional and real time PCR, work with plasmids, cultivation of Salmonella, DNA extraction from biological samples, Gelelectrophoresis

ADDITIONAL INFORMATION

Conferences and Presentations

25th European Congress of Clinical Microbiology and Infectious Diseases (ECCMID), April 2015, Copenhagen, Denmark (Poster)