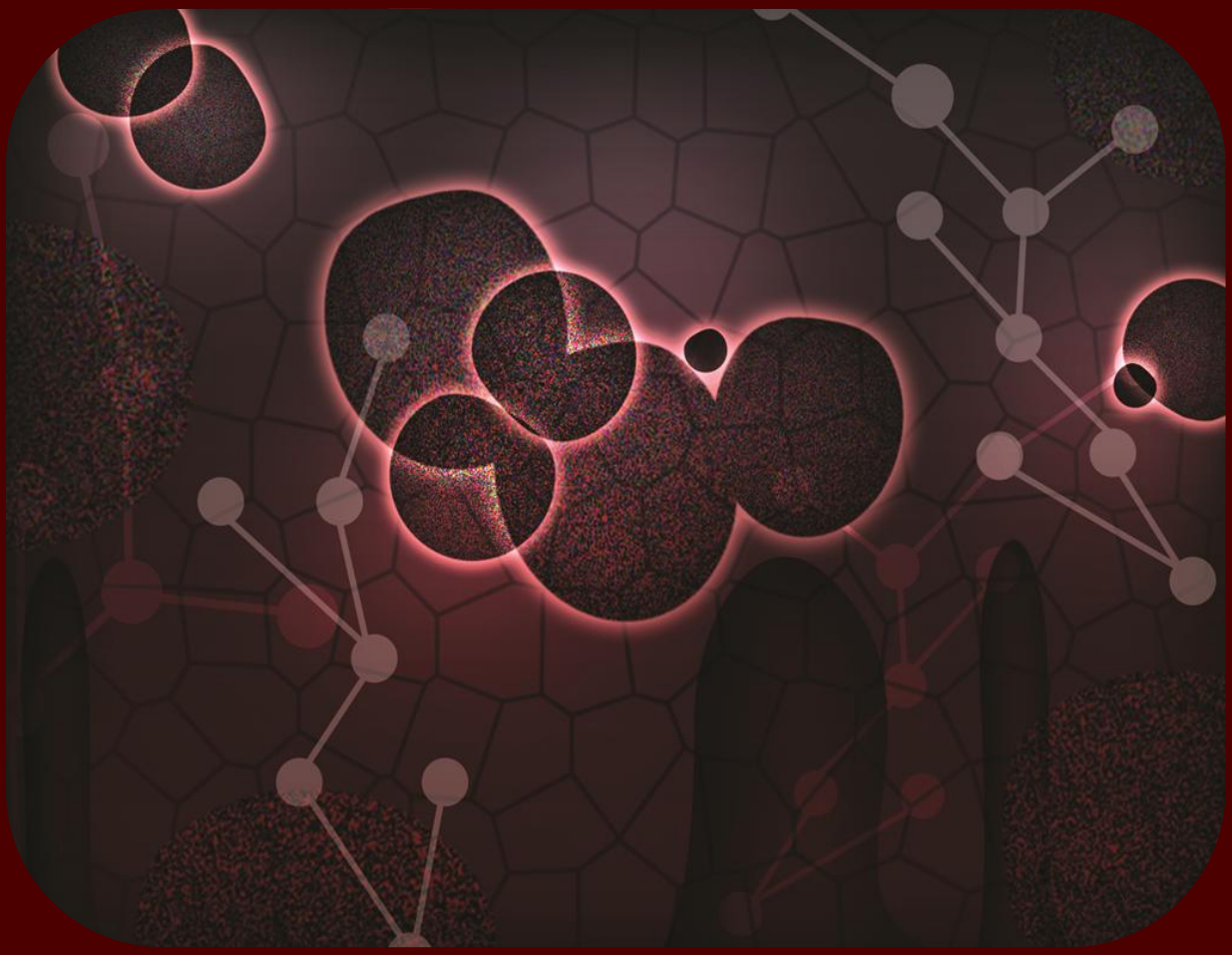
Origin and evolution of the β-lactam resistance determinant in staphylococci

Joana Rolo



Dissertation presented to obtain the Ph.D degree in Molecular Biology

Instituto de Tecnologia Química e Biológica António Xavier | Universidade Nova de Lisboa

Oeiras,
April, 2016

INSTITUTO
DE TECNOLOGIA
QUÍMICA E BIOLÓGICA
ANTÓNIO XAVIER /UNL

Knowledge Creation

Origin and evolution of the ß-lactam resistance determinant in staphylococci

Joana Rolo

Dissertation presented to obtain the Ph.D degree in Molecular Biology

Instituto de Tecnologia Química e Biológica António Xavier | Universidade Nova de Lisboa

Oeiras, April, 2016

INSTITUTO
DE TECNOLOGIA
QUÍMICA E BIOLÓGICA
ANTÓNIO XAVIER /UNL

Knowledge Creation

Second edition, © Joana Rolo

Cover by Hélio Mateus

ISBN 978-989-20-6397-3

Supervisors:

Dr. Maria Miragaia

Professor Hermínia de Lencastre

Dr. Rita Sobral

Chairman of examiners:

Professor Miguel Teixeira

Examiners:

Dr. Matthew Holden

Dr. Stefan Schwarsz

Dr. João Paulo Gomes

ACKNOWLEDGMENTS

To Dr. Maria Miragaia, my supervisor and Head of the Laboratory of Bacterial Evolution and Molecular Epidemiology, at Instituto de Tecnologia Química e Biológica António Xavier (ITQB-AX) da Universidade Nova de Lisboa. Maria was one of my very first scientific supervisors, who very patiently taught me how to work in a lab. I thank her for all the patience and guidance in those first years. I also thank her for her criticism, for pushing me forward and for sharing my enthusiasm about Science. Finally, I thank her for her kindness and endless support, for never letting me give up and most importantly for trusting and believing in my abilities.

To Professor Hermínia de Lencastre, my co-supervisor and Head of the Laboratory of Molecular Genetics, at the Instituto de Tecnologia Química e Biológica António Xavier (ITQB-AX) da Universidade Nova de Lisboa, for accepting me in her lab as an undergraduate and for introducing me to a scientific career. I thank her also for her criticism, her careful revision of my writing and her guidance in my scientific career. I learned a lot from her, both scientifically and personally.

To Dr. Rita Sobral, also my co-supervisor and Head of the Laboratory of Molecular Microbiology of Bacterial Pathogens at UCIBIO-REQUIMTE, Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa, for teaching me some methodologies at the bench, for her valuable suggestions that have always helped this thesis to move forward and for her detailed revisions of my writing. A word of thanks also for her friendship and support.

To Dr. Henrik Westh, my supervisor at the Department of Microbiology, at the Hvidovre Hospital, Denmark, where part of the studies were performed. I felt very welcomed at Copenhagen; I thank him for his kindness and thoughtfulness. A word of thanks also for his interest in my work and helpful critics that many times helped to move my research forward while in Copenhagen.

To Professor Alexander Tomasz, head of the Laboratory of Microbiology and Infectious Diseases at The Rockefeller University, New York, USA, for the valuable critique and thorough revisions of my manuscripts. I also thank him for the unique opportunity to discuss my results and to be able to share my enthusiasm with my work.

To the members of the thesis committee, Dr. Isabel Couto, Assistant Professor at Instituto de Higiene e Medicina Tropical and Dr. Isabel Gordo, head of the Evolutionary Biology group at Instituto Gulbenkian de Ciência, for their helpful critics and suggestions. A special additional acknowledgment to Dr. Isabel Couto, for her previous work with *S. sciuri* strains that helped to select a representative collection for the studies that are part of this thesis.

To Fundação para a Ciência e Tecnologia (FCT, Portugal), for my PhD fellowship SFRH/BD/72675/2010 and for partially funding the studies in this thesis. To the Federation of European Societies of Microbiology (FEMS) for the Research Grant FEMS-RG-2014-0078, that supported one of my visits to Copenhagen.

To Instituto de Tecnologia e Química Biológica António Xavier (ITQB-AX), for providing excellent work facilities.

To the Department of Microbiology, at the Hvidovre Hospital, Copenhagen, Denmark, for accepting me as a foreign student and for providing financial support during my visits to Copenhagen.

To all current and former members of the Laboratory of Molecular Genetics, the Laboratory of Bacterial Evolution and Molecular Epidemiology and the Laboratory of Microbiology of Human Pathogens, for all the help and suggestions that made this thesis go forward. Thank you also for the great relaxing moments during coffee breaks and lunches (especially the ones involving my beloved grilled sardines!). To Sra D. Isilda Gueifão and Sra D. Manuela Nogueira, for the precious technical and administrative support, respectively. I also thank their friendship, kindness and care. A special word of thanks to Professor Ana Madalena Ludovice, who introduced me to the field of molecular genetics, and Dr. Raquel Sá-Leão, head of the Laboratory of Microbiology of Human Pathogens, for their helpful critics.

To my fellow PhD students, Carina Valente, Inês Grilo and Débora Tavares, for walking with me through this long and whining road! We did it!

For all the support at the lab and outside of it, I thank my friends Catarina Milheiriço, Nelson Frazão and Ana Tavares. Thank you for tirelessly and patiently supporting me all the way. You accompanied my scientific growth and are definitely a part of it – and a part of me. I am truly thankful to you and will never forget all the small (and not so small) things you did for me.

Part of this thesis was performed in Denmark, in the incredible city of Copenhagen. A huge “tusind tak” to Peder Worning and Jesper Boye Nielsen, who introduced me to the field of bioinformatics and very kindly gave me a hand whenever I needed. A special thanks to Frederik Hertz, Jesper Boye Nielsen and Klaus Gamst Hansen, who very patiently put up with me all day long during my three stays in Copenhagen. I am proud to say (despite all the complaints and longing, especially for Portugal’s sunny weather) that I enjoyed every single minute in your company and we eventually became friends. I must thank all the people at the Department of Microbiology, at the Hvidovre Hospital, and especially Louise Christensen, Mette Pinholt, Mette Bartels and Kit Boye, for your kindness and for your efforts in making me feel integrated and part of the team. Finally, I thank Torkil Svensgaard, who very patiently solved all the hardware issues. Of course outside the hospital, life went on and for making it so enjoyable and stimulating, I have to thank my friends and fellow PhD students “the mob” Lara Marcos da Silva, Diana Campos and Rita Pinto. Thank you for always remembering to bring me along in all adventures - I am truly blessed to have had the opportunity to share this experience with you! A special thanks also to João Martins and Daniel Carvalho with whom I had the pleasure of sharing housing with. You had the power of cheering me up every single day, whenever I came home, and you were always there for anything that I needed. Finally, I would like to thank all the people I met in Copenhagen and made my stay there truly unforgettable: Thor, Alex, Natalia, Timo and Claudio. Thank you all for your incredible support and stimulating discussions!

I could never have pulled this off without a lot of help from my friends. A special thanks to Bruno Guerra and Catarina Rodrigues (and Lara, of course!). You were always there and kept pushing me forward. I will never forget that. Thank you so much for your friendship!

For keeping me sane and with my feet on the ground, I thank my dearest friends Inês Vieira and Liliana Amaral. You never let me gave up and supported me blindly all the way! I truly could never have made it without you. Thank you for your love and support.

To the great women that were part of my life and helped this thesis move forward: my grandmother, my sister, my mother and my baby daughter. To my grandmother, Ermelinda, for teaching me the most valuable lessons ever. You are missed every day. To my sister, Margarida, for being such an inspiration. You have always been the wise one amongst us two – thank you for all the advice. For having proud in me and pushing me forward in a way that only you know how to do. Finally, to my incredible mother, Teresa. For having always a word of confidence and friendship, even in the darkest of moments. For teaching me to be strong. A

big thank you, to all three of you, for all the love and care throughout these years. To my daughter Alice, for bringing joy and hope in a time when it was most needed and for pushing me forward to finish this thesis.

Finally, I dedicate this thesis to the two most important men in my life and the ones that most inspired me: my father and my husband.

To my father, Carlos, I can only say thank you. You did everything right. As a child, you inspired me to be curious. As an adult, you inspired me to be ambitious. Thank you for all the good advice you have always gave me, for all the interest you have always demonstrated in my career, and for all the inspiration you provided. The truth is I feel I am never truly alone - you are with me every minute of my life, guiding me as you always did.

To my husband, Hélio, for pushing me forward and supporting me all the way. For patiently listening to me and always having a kind word, even through a computer screen. For being so understanding and helpful. And of course, for walking this path with me – you were there in every single step, every curve, every bump on the road.

This thesis is as much yours as it is mine.

A todos os restantes familiares e amigos que me apoiaram ao longo deste anos, um grande obrigado e bem-hajam! Estão no meu coração!

ABSTRACT

In staphylococci, resistance to methicillin and to all β-lactam antibiotics is provided by the *mecA* gene, which encodes a penicillin-binding protein with low affinity to β-lactams (PBP2a). The *mecA* is carried by a mobile genetic element, the staphylococcal cassette chromosome *mec* (SCC*mec*), one of the most widely studied bacterial pathogenicity islands. SCC*mec* carries *mecA* and its regulators (the *mec* complex), as well as cassette chromosome recombinases encoded by *ccr* genes that form the *ccr* complex. These recombinases assure the mobility of the cassette. In addition, SCC*mec* cassettes carry joining regions (J regions) that link the *orfX* to the *mec* complex (J3); the *mec* complex to the *ccr* complex (J2) and the *ccr* complex to the end of the cassette (J1). The J regions can carry additional antibiotic resistance determinants, transposons, insertion sequences and plasmids. The SCC*mec* element always inserts at the same site in the bacterial chromosome, downstream *orfX* (which encodes a RNA methyltransferase), located 500 kb downstream the origin of replication. SCC*mec* is a very diverse element; so far eleven different types have been identified in *Staphylococcus aureus* and many more are probably carried by coagulase-negative staphylococcal species. SCC*mec* is transferred horizontally among strains and species of *Staphylococcus*, through an unknown molecular mechanism.

The crucial event leading to the emergence and dissemination of methicillin-resistant *S. aureus* (MRSA) and methicillin-resistant *Staphylococcus epidermidis* (MRSE) was the acquisition of SCC*mec* in the genetic background of these bacteria, shortly after the introduction of methicillin in clinical practice. However little is known regarding the origin of this element. In this Thesis we aimed to contribute to the understanding of the origin and steps of assembly of SCC*mec* and to assess its impact in the emergence and evolution of MRSE as a nosocomial pathogen.

Previous studies indicated that *mecA* has originated in the most primitive group of staphylococcal species, the phylogenetic *sciuri* group. The *sciuri* group comprises five species that are widespread in nature and only rarely colonize humans. Noteworthy, *mecA* homologues with different degrees of nucleotide identity with *S. aureus mecA* have been identified among three species of the *sciuri* group:

Staphylococcus sciuri (*mecA1*, 80%), *Staphylococcus vitulinus* (*mecA2*, 90%) and *Staphylococcus fleurettii* (*mecA*, 99%). In this Thesis we showed that the *mecA* homologues carried by these species have the same chromosomal location (the native location, 200 kb downstream *orfX*), suggesting that these native penicillin-binding proteins have evolved from a common ancestor, the most primordial one being *mecA1*. In addition, we studied the evolution of the *mecA* homologues towards the expression of resistance in their own native host, in species belonging to the *sciuri* group. We found that the development of resistance in the species of the *sciuri* group was achieved by distinct mechanisms: diversification of the promoter region of the gene, accumulation of single-nucleotide polymorphisms (SNPs) in the non-binding domain of *mecA1*-encoded PBP4, and adaptation of the genetic background. Moreover, we found that major diversification of *mecA* homologues begun during the introduction of β-lactam antibiotics, namely penicillin, in human clinical practice and in animal feeding as additives. We have also described, for the first time, the high frequency and diversity of *ccr* genes among *S. sciuri* isolates collected in different time periods, different hosts and different geographic locations; the results indicate that the *ccr* complex originated in *S. sciuri*. We suggest that the most primitive *Staphylococcus* species, *S. sciuri*, is the source of the *mecA* gene and the *ccr* complex, the building blocks of SCC*mec*.

Besides focusing in *ccr* and *mecA*, we have also searched for elements carried in the J regions of SCC*mec* in order to understand the contribution of the three species of the *sciuri* group to the assembly of this element. Using whole-genome sequencing analysis, we propose for the first time, a model for the assembly of SCC*mec*. According to our data, the *mec* complex (*mecR2-mecR1-mecI-mecA*-IS*431mec*) in native location evolved in parallel with SCC in the *orfX* region. The *mecR2* and genes within J2-J3 regions originated in *S. sciuri* and evolved over phylogeny (*S. sciuri*-> *S. vitulinus* -> *S. fleurettii*), but the remaining regulators were only added in *S. fleurettii*. Once formed, the *mec* complex from *S. fleurettii* was mobilized to *S. vitulinus*, probably by recombination. The first SCC in *S. sciuri* was formed by the creation of direct/inverted repeats (DR/IR) around *ccr* genes and housekeeping genes resident within the *orfX* region. The final assembly step of SCC*mec* occurred in *S. sciuri*, wherein the *ccr*

complex and the adjacent J1 region located within a resident SCC element incorporated the *mec* complex with adjacent J2-J3 regions from *S. vitulinus*. Moreover, we showed that the very first SCC*mec* type was an ancestral of SCC*mec* III, a cassette that has spread to a large number of different staphylococcal species and is responsible for the emergence of the MRSA ST239-III clone, associated to one of the most important MRSA pandemics.

Overall our data showed that *S. sciuri*, the most primitive staphylococcal species and also the most widespread in nature, was the origin of *mecA*, the *ccr* complex and of a primordial SCC*mec* element. Our studies reinforce the importance of animal-associated staphylococcal species as a source of antimicrobial resistance determinants and the use of antibiotics in treatment and animal feeding additives as the driving force for their emergence. We suggest that the detailed study of these primitive staphylococcal species could help to anticipate the emergence of other antibiotic resistance determinants.

Our results have also shown that similarly to *S. sciuri*, other coagulase-negative staphylococci, were important players in the assembly of SCC*mec*. This was the case of *S. epidermidis*, which we established to be involved in the assembly of SCC*mec* IV. We studied a collection of *S. epidermidis* isolates obtained in Denmark in 1965, and found that a structure resembling SCC*mec* IV cassettes, a ψSCC*mec*-IV-like element (a SCC element lacking the *ccr* complex), was already carried by these early *S. epidermidis*. This structure did not contain a *ccr* complex and carried a *mecA* copy disrupted by an IS*431* and thus did not provide resistance to β-lactams. In addition, early methicillin-susceptible *S. epidermidis*, lacking *mecA*, carried in the *orfX* vicinity genes with high homology with genes that are part of SCC*mec* IV. Overall, the data provided in this Thesis, highly suggest that SCC*mec* IV emerged in *S. epidermidis* through a mechanism similar to that described for *S. sciuri*, wherein a resident SCC containing specific *ccr* complexes incorporated heterologous *mec* complexes. In addition, we unraveled a mechanism that can be used by early bacteria to accommodate *mecA* in the chromosome: the interruption of *mecA* coding frame by a copy of the insertion sequence IS*431*. This observation could correspond to a first step in the domestication

of *mecA* that allows for the necessary adaptation of the genome to the expression of resistance to β-lactam antibiotics.

Besides being involved in the assembly of SCC*mec* IV, we showed that *S. epidermidis* were also active players in the diversification of SCC*mec*, when the contact with the hospital environment increased. Actually, when we compared *S. epidermidis* strains of the same genetic background collected in the community and the hospital in the same time period and geographic origin, we found that strains collected in the hospital environment presented a higher frequency and diversity of SCC*mec* than their community counterparts.

The impact of the contact of *S. epidermidis* with the hospital environment in its development as a pathogen was further evidenced when we compared the genomes of the early *S. epidermidis* with contemporary isolates collected in Denmark. We found that the core genomes of highly virulent contemporary isolates were related with the ones found among early isolates, although the distribution of mobile genetic elements varied greatly between the two collections. Specifically, contemporary isolates carried insertion sequences and SCC*mec* in higher frequency and diversity, while prophages were more abundant in early genomes. Of note, we have also found that the number and frequency of biofilm-associated genes, like *ica*, *aap*, *bap*, *sdrF* and ACME-I, was higher among contemporary genomes. Taking our results together, we suggest that adaptation towards the hospital environment has favored the accumulation of genes involved in biofilm, β-lactam resistance and genome plasticity. Our findings allowed a better understanding of the molecular evolutionary pathways used by opportunistic nosocomial pathogens, like *S. epidermidis*, during adaptation to the hospital environment and might help to design new strategies for treatment and infection control of *S. epidermidis*.

In conclusion, in this Thesis we were able to:

i. Identify *S. sciuri* as the original source of the *ccr* complex and the *mecA* determinant;

ii. Identify the molecular mechanisms involved in the development of β-lactam resistance in primitive staphylococcal species;

iii. Propose a model for SCC*mec* assembly, which occurred in the phylogenetic *sciuri* group;

iv. Produce evidence that SCC*mec* IV originated in *S. epidermidis*;

v. Establish that *S. epidermidis* adaptation to the hospital environment involved multiple mechanisms, namely acquisition and diversification of SCC*mec*, acquisition of insertion sequences, loss of phages and accumulation of genes involved in biofilm formation.

RESUMO

No género *Staphylococcus*, a resistência à meticilina e a todos os antibióticos β-lactâmicos é conferida pelo gene *mecA*, que codifica uma proteína envolvida na síntese da parede celular, denominada PBP2a (PBP de, *penicillin-binding protein*) com afinidade reduzida para os antibióticos β-lactâmicos. O gene *mecA* está incluído no elemento genético móvel SCC*mec* (de, *staphylococcal cassette chromosome mec*), considerado uma das mais importantes ilhas de patogenicidade bacterianas. O elemento SCC*mec* transporta o gene *mecA* e os seus reguladores (denominado complexo *mec*), bem como recombinases que asseguram a mobilidade da cassette, denominadas *ccr* (de, *cassette chromosome recombinases*) e que formam o complexo *ccr*. Adicionalmente, as cassettes SCC*mec* contém regiões de junção (*joining* ou regiões J) que ligam a *orfX* ao complexo *mec* (J3); o complexo *mec* e o complexo *ccr* (J2); e o *ccr* ao final da cassette (J1). As regiões de junção podem conter outros determinantes de resistência a antibióticos, inseridos em transposões, sequências de inserção e plasmídeos. O elemento SCC*mec* insere-se sempre no mesmo local cromosómico, a jusante do gene *orfX* (que codifica uma metiltransferase de RNA), localizado a 500 kb da origem de replicação. O SCC*mec* é um elemento genético extremamente diverso; existem onze tipos diferentes descritos em *Staphylococcus aureus* e numerosos tipos adicionais foram identificados em *Staphylococcus* coagulase-negativos. A transferência do SCC*mec* ocorre horizontalmente entre estirpes e espécies de *Staphylococcus*, por via de um mecanismo molecular ainda desconhecido.

O acontecimento-chave que levou à emergência e disseminação de MRSA (de, *methicillin-resistant Staphylococcus aureus*) e de MRSE (de, *methicillin-resistant Staphylococcus epidermidis*), foi a aquisição do elemento SCC*mec*, que se pensa possa ter ocorrido imediatamente após a introdução da meticilina na prática clínica. Contudo, a origem deste elemento é ainda desconhecida. Neste Tese de Doutoramento, os nossos principais objectivos foram o estudo da origem e dos passos que levaram à construção do SCC*mec* e ainda a avaliação do seu impacto na emergência e evolução de *S. epidermidis*.

Vários estudos anteriores demonstraram que a origem do gene *mecA* pode ser encontrada em espécies mais primitivas do género *Staphylococcus*; estas espécies formam um grupo designado, grupo *sciuri,* que engloba espécies que estão disseminadas na natureza e que apenas raramente colonizam o ser humano. De notar, genes homólogos do *mecA* com diferentes graus de identidade nucleotídica foram identificados em diversas espécies dentro deste grupo, nomeadamente em *Staphylococcus sciuri* (*mecA1*, 80%), *Staphylococcus vitulinus* (*mecA2*, 90%) e *Staphylococcus fleurettii* (*mecA*, 99%).

Nos nosso estudos verificámos que todos os genes homólogos de *mecA* destas espécies têm a mesma vizinhança genética e estão localizados no mesmo *locus* cromossómico (local nativo, 200 kb a jusante da *orfX*), o que sugere que evoluíram de um ancestral comum. O estudo da evolução dos genes homólogos do *mecA* no seu hospedeiro, em espécies pertencentes ao grupo filogenético *sciuri* permitiu-nos também descrever que a resistência nestas espécies emergiu através de diversos mecanismos: diversificação do promotor, acumulação de mutações pontuais no domínio *non-binding* das proteínas PBPs nativas (codificada pelos genes nativos homólogos *do mecA*) e adaptação do património genético. Verificámos também que o momento no qual ocorreu a maior diversificação do gene *mecA1* em *S. sciuri* coincide com a introdução dos antibióticos β-lactâmicos como forma de tratamento de infecções bacterianas em humanos e como aditivos em rações para animais. Identificámos também, pela primeira vez, uma elevada frequência e diversidade de genes *ccr* numa colecção de isolados de *S. sciuri* obtidos em diferentes períodos, diversos hospedeiros e países distintos. Os resultados sugerem que a espécie estafilocócica mais primitiva, *S. sciuri,* foi não só a origem do gene *mecA* mas também do complexo *ccr*, que juntos constituem os dois elementos centrais do SCC*mec*.

Para além de estabelecermos qual a origem dos principais elementos do SCC*mec*, complexo *mec* e complexo *ccr*, estudámos também a origem de elementos presentes nas regiões J do SCC*mec* e os passos evolutivos que levaram à construção do SCC*mec*. O estudo da sequência dos genomas de um grande número de isolados de *S. sciuri*, *S. vitulinus* e *S. fleurettii* permitiu desenhar um modelo para a construção do SCC*mec*. De acordo com os nossos resultados, a construção do complexo *mec* no local

nativo ocorreu em paralelo com a construção de elementos SCC na região *orfX*. O gene *mecR2* e genes pertencentes à região J2/J3 tiveram origem em *S. sciuri* evoluíram no local nativo ao longo da filogenia (*S. sciuri*-> *S. vitulinus* -> *S. fleurettii*), mas os restantes genes reguladores do *mecA* só foram adicionados ao complexo *mec* mais tarde na evolução filogenética, em *S. fleurettii*. Uma vez formado, o *mec* complex terá sido transferido para *S. vitulinus* onde recombinou com o *mecA2* no local nativo. Por outro lado a construção do elemento SCC terá ocorrido na *orfX* através da criação de DR/IR (de, *direct/inverted repeats*) à volta de recombinases e genes *housekeeping* residentes nesta região (pertencentes à região J1). O passo final de construção do elemento SCC*mec* terá ocorrido em *S. sciuri*, onde o *mec* complex de *S. vitulinus* terá sido introduzido num elemento SCC já previamente formado. Os nossos dados indicaram, também, que o primeiro tipo SCC*mec* foi um elemento ancestral do SCC*mec* III, uma cassette que se disseminou num elevado número de espécies estafilocócicas diferentes e é responsável pela emergência do clone MRSA ST239-III, associado a uma das mais importantes pandemias mundiais de MRSA.

Em resumo, os nossos dados revelaram que *S. sciuri*, a espécie estafilocócica mais primitiva e também a mais distribuída na natureza, foi a origem do *mecA*, do complexo *ccr* e de uma estrutura primordial do SCC*mec*. Os estudos presentes nesta Tese de Doutoramento, vêm reforçar a importância de espécies estafilócocicas associadas aos animais como fontes de genes de resistência a antibióticos, e o uso de antibióticos no tratamento de infecções bacterianas em humanos e como aditivos alimentares em rações, como causas da emergência desta resistência. O estudo detalhado destas espécies poderá ajudar a antecipar a ocorrência de emergência de resistência a outras classes de antibióticos.

Os nossos resultados demonstraram também que, à semelhança de *S. sciuri*, outras espécies de *Staphylococcus* coagulase-negativos, estiveram envolvidas na construção de elementos SCC*mec*. Uma destas bactérias foi *S. epidemidis*, que verificámos ter contribuído para a construção do SCC*mec* IV. Em particular verificámos que uma estirpe arcaica de *S. epidermidis* obtida na Dinamarca em 1965, transportava um elemento muito semelhante ao SCC*mec* IV, denominada Ψ-SCC*mec* IV. Esta estrutura não contém o complexo *ccr*, e tem o *mecA* interrompido por uma sequência

de inserção (IS*431*), pelo que não confere resistência aos antibióticos β-lactâmicos. Para além disso, verificou-se que outros isolados de *S. epidermidis* recolhidos na mesma data e local geográfico continham, na região da *orfX*, genes com elevada homologia com genes pertencentes à cassete SCC*mec* IV. Em conclusão, os nossos dados demonstraram que o SCC*mec* IV surgiu em *S. epidermidis*, provavelmente por um mecanismo semelhante àquele descrito para *S. sciuri*, onde um elemento SCC residente terá incorporado o *mec* complex. Adicionalmente, identificou-se um mecanismo que poderá ter sido utilizado pelos isolados de *S. epidermidis* de 1965 para acomodar o gene *mecA* no seu cromossoma: a interrupção do *mecA* por uma cópia intacta da sequência de inserção IS*431*. Este fenómeno poderá corresponder a um passo intermédio de domesticação do *mecA*, que permitirá a adaptação do genoma à expressão da resistência aos antibióticos β-lactâmicos.

Para além de estar envolvido na construção do SCC*mec*, os nossos estudos indicam que a espécie *S. epidermidis* contribuiu para a diversificação do SCC*mec*, principalmente no ambiente hospitalar. Na realidade, a comparação da estrutura populacional de isolados de *S. epidermidis* obtidos na comunidade e no hospital no mesmo período temporal e com a mesma origem geográfica demonstrou que as estirpes isoladas nos dois ambientes pertenciam à mesma linhagem genética. No entanto verificámos que quando isoladas nos hospitais as estirpes apresentaram uma maior frequência do SCC*mec* e um maior número de tipos diferentes de SCC*mec*.

O impacto do hospital na adaptação, evolução e emergência de *S. epidermidis* como agente patogénico foi adicionalmente evidenciado quando comparámos os genomas dos isolados *S. epidermidis* obtidos em 1965 com o genoma de isolados *S. epidermidis* contemporâneos obtidos também na Dinamarca. Verificou-se que as estirpes arcaicas pertenciam à mesma linhagem genética que as estirpes contemporâneas; no entanto variavam quanto à distribuição de elementos genéticos móveis e conteúdo genético. Em particular, os isolados contemporâneos apresentaram maior frequência de sequências de inserção, maior frequência e diversidade de SCC*mec*, e menor conteúdo de profagos. Adicionalmente, verificou-se que os isolados contemporâneos continham um maior número e frequência de genes associados à formação de biofilme, nomeadamente os genes *ica*, *aap*, *bap* e *sdrF*, bem como uma

maior frequência do elemento ACME-I. Assim, os nossos dados sugerem que a adaptação de *S. epidermidis* ao hospital ocorreu através da acumulação de determinantes genéticos associados à formação de biofilme, resistência aos antibióticos β-lactâmicos e plasticidade genética.

Em conclusão, os nossos estudos sugerem que o contacto de *S. epidermidis* com o ambiente hospitalar, favoreceu a emergência da resistência aos antibióticos β-lactâmicos, a acumulação de genes envolvidos na formação de biofilme e a sua plasticidade genética. Estes resultados contribuíram para o conhecimento das estratégias usadas por microorganismos oportunistas na sua adaptação ao ambiente hospitalar, o que poderá ajudar no desenho de estratégias de tratamento e controlo da infecção.

Resumindo, nesta Tese de Doutoramento:

i. Identificámos *S. sciuri* como a origem do complexo *ccr* e o determinante genético *mecA*;

ii. Identificámos os mecanismos moleculares envolvidos na aquisição da resistência aos antibióticos β-lactâmicos nas espécies estafilocócicas mais primitivas;

iii. Elaborámos um modelo para a construção do SCC*mec*, que ocorreu no grupo filogenético *sciuri*;

iv. Produzimos evidências que suportam a origem do SCC*mec* IV em *S. epidermidis*;

v. Estabelecemos que a adaptação de *S. epidermidis* ao ambiente hospitalar envolveu diversos mecanismos, nomeadamente, a aquisição e diversificação do SCC*mec*, a aquisição de sequências de inserção, perda de profagos e a acumulação de genes associados à formação de biofilmes.

THESIS OUTLINE

This Thesis focuses on the origin and assembly of SCC*mec*, in particular: (i) the role of *Staphylococcus sciuri* as the origin of the methicillin resistance determinant *mecA* and the *ccr* complex; (ii) the assembly of a primordial SCC*mec* III structure in the phylogenetic *sciuri* group; (iii) the assembly of SCC*mec* IV in *S. epidermidis*; (iv) the impact of hospital contact in SCC*mec* and *S. epidermidis* evolution.

In **Chapter I**, a general introduction providing background on the relevant literature regarding SCC*mec* origin, assembly and impact in the evolution of staphylococcal species, is provided. In addition, the gaps in the current knowledge that we aim to address with the results of this Thesis are highlighted.

The findings and results of this Thesis are organized in two parts. In **Part I**, the role of the most primitive staphylococcal species in the origin of SCC*mec* is discussed. Part I comprises three chapters. **Chapter II** focuses on the evolution of the recombinases responsible for the mobilization of SCC*mec* (the *ccr* genes) and the role of *S. sciuri* in the construction of the first *ccr* complexes. In **Chapter III** the evolution of the ancestral form of the central element of methicillin resistance, *mecA* and the molecular mechanisms that led to the development of β-lactam resistance in primitive staphylococcal species are assessed. Finally, in **Chapter IV**, the role of *S. sciuri* and related species in the assembly of a primordial SCC*mec* is discussed in detail.

Part II focuses on the impact of hospital contact in the evolution of *S. epidermidis* and SCC*mec*. In **Chapter V**, the study of a collection of *S. epidermidis* isolates obtained in the early antibiotic era sheds light on the evolutionary origin of MRSE and SCC*mec* IV and their comparison with contemporary isolates showed the impact of contact with hospital environment on the genetic content of *S. epidermidis*. In **Chapter VI,** the comparison of contemporary *S. epidermidis* isolates obtained in the hospital and in the community allowed to identify additional strategies developed by this species to adapt to the hospital environment.

Finally, in **Chapter VII**, the results of this Thesis are discussed; a model for assembly of SCC*mec*, including several species of the phylogenetic *sciuri* group, is

presented. In addition, the role of the hospital environment in the evolution of SCC*mec* and *S. epidermidis* is discussed.

LIST OF ABBREVIATIONS

A

AAP – accumulation-associated protein

ACME – arginine catabolic mobile element

AMP – antimicrobial peptide

Atl – autolysin

B

Bap – biofilm-associated protein

bp – base pairs

BURST – based upon related sequence types

C

CA – community-associated

CC – clonal complex

ccr – cassette chromosome recombinase

CDS – coding sequences

CI – confidence interval

CoNS – coagulase-negative staphylococci

D

DLV – double-locus variant

DNA – deoxyribonucleic acid

dNTP – deoxynucleotide

DR – direct repeat

H

HA – hospital-associated

HVR – hypervariable region

I

IR – inverted repeat

IS – insertion sequence

M

MIC – minimum inhibitory concentration

MLST – multilocus sequence typing

MRCoNS – methicillin-resistant coagulase-negative staphylococci

MRS – methicillin-resistant staphylococci

MRSA – methicillin-resistant *Staphylococcus aureus*

MRSE – methicillin-resistant *Staphylococcus epidermidis*

MSSA – methicillin-susceptible *Staphylococcus aureus*

MSSE – methicillin-susceptible *Staphylococcus epidermidis*

N

NA – not available

ND – not determined

NGS – next generation sequencing

NT – non-typeable

O

ORF – open reading frame

P

PCR – polymerase chain reaction

PBP – penicillin-binding protein

PIA – polyssaccharide intercellular adhesin

PFGE – pulsed-field gel electrophoresis

PSM – phenol-soluble modulins

R

RNA – ribonucleic acid

S

SID – Simpson's index of diversity

SCC – staphylococcal cassette chromosome

SCC – CI – staphylococcal cassette chromosome composite island

SCC*mec* – staphylococcal cassette chromosome mec

SLV – single-locus variant

SNP – single-nucleotide polymorphism

ST – sequence type

T

Tn – transposon

U

UPGMA – unweighted pair group method with arithmetic mean

W

WGS – whole-genome sequence

TABLE OF CONTENTS

Chapter I

General Introduction

1. *Staphylococcus* in the clinical setting: a worldwide concern

Since their discovery in the late 1800s (1), staphylococci, and particularly *Staphylococcus aureus* have been widely studied due to their importance in the clinical setting. Despite being primarily commensals, staphylococci can cause a wide range of infections that are often difficult to treat, due to their adaptive power to environmental stresses such as the ones in the hospital environment, and increased resistance to antimicrobials. Actually, the burden caused by staphylococcal infections is increasing worldwide (2), and recent advances in molecular typing techniques have revealed some of the secrets behind the success of these highly adaptable pathogens. In the following sections, the available literature on this subject is reviewed, with a focus on the opportunistic pathogen *Staphylococcus epidermidis*.

1.1. The genus *Staphylococcus* and clinically relevant species

The genus *Staphylococcus* is part of the family *Staphylococcaceae,* that belongs to the order *Bacillales* of the class *Bacilli* (3). *Bacilli* belong to the phylum *Firmicutes*, which comprises Gram-positive bacteria that typically have a low G+C DNA content (30-40%) (2). The genus *Staphylococcus* comprises 52 species and 28 subspecies (2, 4) and the bacteria belonging to *Staphylococcus* have a typical round shape (cocci) and agglomerate as “grape-like” clades (5). Phenotypically, these bacteria are non-motile and facultative anaerobes. They can be easily identified by their biochemical phenotypic properties: non-production of oxidase, production of catalase and tolerance to high concentrations of salt (6). Regarding their ability to produce coagulase the genus can be subdivided into two large groups, one comprising 38 species that do not produce coagulase, so-called coagulase-negative staphylococci, CoNS (2) and 14 which produce coagulase, coagulase-positive staphylococci. Moreover, *Staphylococcus* cells are enveloped by a characteristic cell wall enriched in O-acetylated peptidoglycan. The peptidoglycan layer in *Staphylococcus* cell wall is highly crossed-linked, composed by linear glycan chains linked by short peptides, which together with teichoic acids, constitute the cell wall of *Staphylococcus* (7).

Recently, the genus *Staphylococcus* has been re-classified, due to the advances and availability of molecular data (in particular whole-genome sequencing data) that allowed the establishment of more accurate phylogenetic relationships. The 52 *Staphylococcus* species have been organized into distinct cluster groups, which, together with their phenotypic and epidemiological characteristics, were divided into six groups of species (Auricularis, Hyicus-Intermedius, Epidermidis-Aureus, Saprophyticus, Simulans and *Sciuri*) (2, 8)(Figure 1).

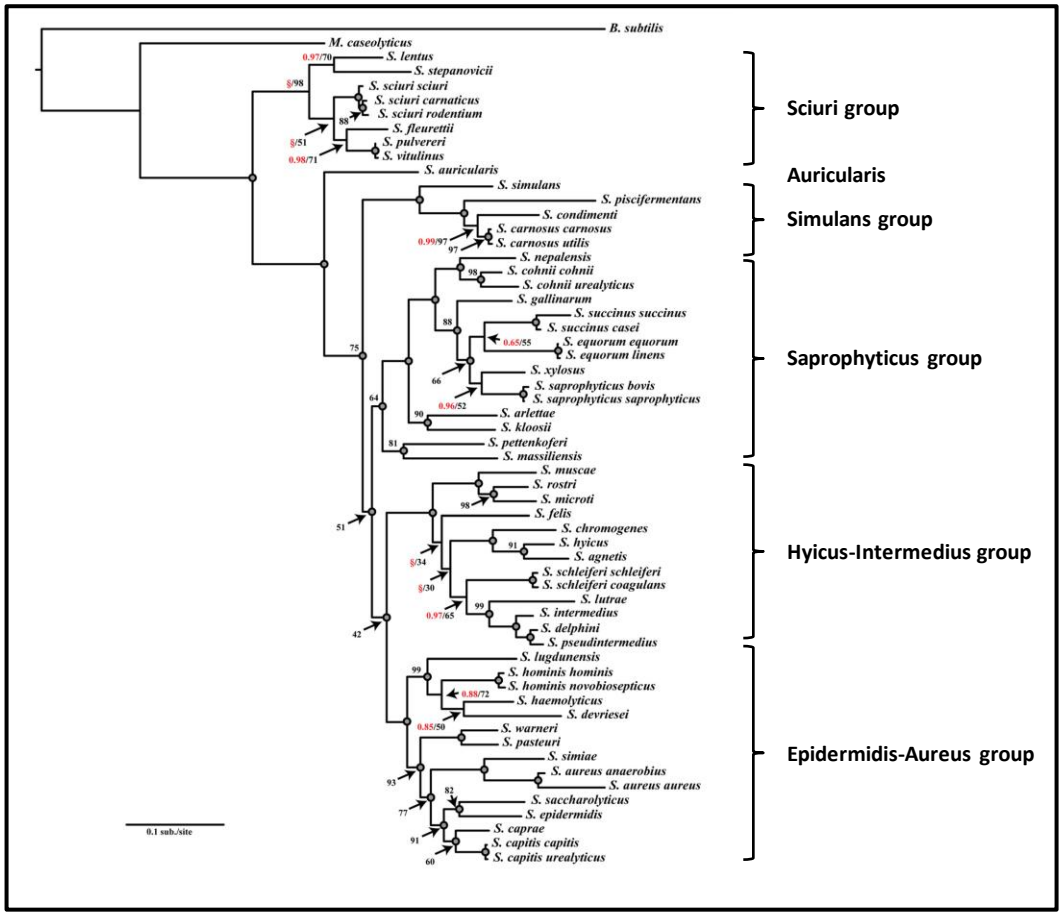


Figure 1. Maximum likelihood phylogram based on the nucleotide sequence of 16S ribosomal genes of 46 staphylococcal species. The consensus phylogram was generated from 200 bootstrap replicates with five maximum likelihood search replicates per bootstrap. Adapted from (8).

The vast majority of the most clinically relevant species belong to the Epidermidis-Aureus group, such as *Staphylococcus aureus*, *Staphylococcus epidermidis*, *Staphylococcus haemolyticus*, *Staphylococcus hominis*, *Staphylococcus capitis* and *Staphylococcus lugdunensis*. Other clinically relevant species, which are not so frequently recovered from clinical specimens, are part of the Saprophyticus group, such as *Staphylococcus saprophyticus*, *Staphylococcus cohnii* and *Staphylococcus xylosus.* The Auricularis group includes a single species, *Staphylococcus auricularis* that is found exclusively as a colonizer of the human external ear. The Simulans group comprises species that only transiently cause disease in humans, and are typically isolated from meat and food (2). Finally, the Hyicus-Intermedius and the Sciuri group includes species that are most frequently found colonizing animals and only rarely found colonizing or causing infections in humans (2).

1.2. Molecular typing techniques and whole-genome sequencing analysis in *Staphylococcus*

The molecular epidemiology of *Staphylococcus* has been extensively studied through the application of multiple molecular typing methods and, more recently, the analysis of whole-genome sequences (WGS) revealed *Staphylococcus* population structure with unprecedented resolution. The most commonly used molecular typing techniques, pulsed-field gel electrophoresis (PFGE) and multilocus sequence typing (MLST), as well as the recent advances in WGS analysis will be briefly reviewed in this section.

1.2.1. Pulsed-field gel electrophoresis (PFGE)

The separation of macrorestriction DNA fragments using PFGE was one of the first typing methods used to study the molecular epidemiology of *Staphylococcus*. In this technique, a restriction enzyme is used to digest the chromosomal bacterial DNA into a relatively low number (15 to 25) of fragments (9). These fragments are then separated in an agarose gel that is submitted to an electric field, which changes in

orientation, direction and intensity during the electrophoresis. The result is a DNA macrorestriction profile of fragments that are distributed according to their molecular weight (9). Since this profile corresponds to the complete genetic content of a strain, the variations observed between different PFGE patterns reflect genetic events that occurred recently, such as insertions, deletions and mutations occurring in the enzyme's restriction site. Due to its high discriminatory power, this method has been previously considered as the "gold standard" technique to be used to study outbreaks (9, 10). The degree of relatedness among isolates as estimated by their macrorestriction patterns has been defined for *S. aureus* (11, 12) and *S. epidermidis* (13). In general, two isolates are considered to be closely related if the differences in their macrorestriction patterns are consistent with the occurrence of a single genetic event, which results in 2-3 band differences in the macrorestriction patterns (11). When the macrorestriction patterns presented by the isolates differ by 4-6 bands, then the isolates are considered to be possibly related and when more than 7 band differences are observed, the isolates are considered unrelated (11). This analysis has been classically performed by visual inspection of the band profiles but in recent years it has been replaced by automatic statistical analysis, using informatics tools, such as the program BioNumerics (Applied Maths, Kortrijk, Bélgica), that uses image analysis parameters. Briefly, the program allows the definition of tolerance and optimization values for the clustering of the patterns of bands obtained for each strain. The tolerance is the measurement of "movement" of each band, while the optimization limits the "movement" of each fingerprint as a whole. These parameters have been set for each species. For example, for *S. epidermidis*, it has been proposed the use of the Dice algorithm to determine the similarity between the restriction profiles, with 1.3% optimization and 0.8% tolerance; the profiles are then clustered with the unweighted pair group method with arithmetic mean (UPGMA) and profiles are considered to belong to the same PFGE type with a cut-off value of 79% of similarity between their band pattern (13).

1.2.2. Multilocus sequence typing (MLST)

Multilocus sequence typing (MLST) relies on the sequencing of internal fragments of seven housekeeping genes, which are scattered along the chromosome. To each different allele in the population, a number is attributed and the resulting combination of numbers for the seven genes is called the allelic profile; a number is assigned to each unique allelic profile, which corresponds to the sequence type (ST). Since this method focus on housekeeping genes, it reflects the accumulation of genetic mutations that occur at a slow evolutionary rate, being appropriate to compare isolates collected in different geographic regions and time periods (14). MLST data is commonly analyzed with the eBURST algorithm (15), which is based on a simple model of evolution and diversification. It defines groups of STs related to one another, called the clonal complexes (CC) and its predicted founder (the ST that is most frequent and presents a higher number of related STs). The *S. epidermidis* MLST scheme present in the MLST database (www.mlst.net) (16) has been widely used to study the molecular epidemiology of this bacterium. For *S. epidermidis*, it has been defined that STs belonging to the same CC should share at least six alleles (out of seven) with another ST of the same group (17).

1.2.3. Whole-genome sequence analysis

1.2.3.1. Historical prespective

Since its emergence in the market in 2005, the technology behind the first next-generation sequencing (NGS) devices has evolved substantially, together with the broadening of its applications. Several NGS technologies emerged, based on different sequencing chemistries, such as the 454 sequencing technology, the Illumina technology, the Solid technology and the Ion torrent technology, providing a wide range of amplification strategies and having different number of Mb sequenced per run, of read length and of cost per run (Table 1). These technologies overcame the limitation of Sanger sequencing (developed in the 1970s)(18), that was the need to

amplify the number of target DNA fragments before sequencing, which was done usually by cloning into specific bacterial hosts (19).

The very first NGS technology developed was 454 sequencing technology, that was released by 454 Life Sciences in 2005 (20), a company that was later acquired by Roche (Basel, Switzerland). The development of this technology was ground-breaking since it was based on a new DNA amplification method called the emulsion PCR (19) that consists in the shearing of the genome into DNA fragments which are then ligated to streptavidin beads. The beads are then captured into separate emulsion droplets, which act as individual amplification reactors (19). Each droplet is then transferred into a well of a microtiter plate and sequenced by pyrosequencing technology. The pyrosequencing technology is based on the measurement of the release of inorganic pyrophosphate by chemiluminescence upon incorporation of labeled dNTPs in the nascent DNA chain (21). This technology produces reads of around 300 bp and when it was introduced in the market, was very costly (Table 1). In addition, it has the disadvantage of producing a high number of sequencing errors.

Shortly after the release of the first 454 apparatus, a new NGS technology was introduced in the market by Illumina (Hiseq, San Diego, USA). Briefly, the Illumina technology consists in shearing the genome into small single-stranded DNA fragments (the libraries). These are then attached to a flowcell and sequencing is achieved through solid-phase bridge amplification (19). During this process, one end of a single DNA molecule is attached to the flowcell using an adapter; the molecules subsequently bend over and hybridize to complementary adapters, thereby forming the template for the synthesis of their complementary strands (19). The templates are sequenced in a massively parallel fashion using the basic principles of Sanger sequencing (18). This technology produces short sequence reads (50-250 bp) (19) (Table 1) but due to its sequencing chemistry, it produces less sequencing errors than 454 technology. Furthermore, the cost of each run decreased dramatically due to the low cost of the reagents used in the sequencing reaction, which promoted the use of this technology, particularly for sequencing bacterial genomes. The development of a benchtop sequencer by Illumina, MiSeq, in 2011 has further increased the use of this NGS technology for the control of outbreaks and molecular epidemiology of pathogens,

namely staphylococcal species. In fact, for the two major staphylococcal pathogens *S. aureus* and *S. epidermidis*, 4369 and 106 whole-genome sequences have been deposited in the NCBI database by August 2015 (www.ncbi.nlm.nih.gov/), respectively, illustrating the growing interest in this high throughput technology.

Several different NGS technologies were introduced in the market shortly after the release of Hiseq by Illumina (Table 1). Briefly, in the Solid technology (Applied Biosystems, ThermoFisher, Waltham, Massachusetts, United States) a library of DNA fragments linked to clonal beads is prepared by emulsion PCR. Then adapter primers hybridize to the template beads and a set of four fluorescently labeled di-base probes are added to the mixture, that compete for ligation to the sequencing primer. The sequencing is achieved after multiple cycles of ligation, detection and cleavage. On the other hand, the Ion Torrent technology (Life Technologies, ThermoFisher, Waltham, Massachusetts, United States) is based on the release of pyrophosphate and a positively charged hydrogen ion, when incorporation of dNTPs occurs into a nascent DNA strand. Briefly, the DNA polymerase is immobilized on a semiconductor chip that is then flooded with many copies of single-stranded DNA template and each of the four dNTPs, sequentially. If a certain dNTP is incorporated, the hydrogen ion is released and the pH of the solution changes, which is detected by the chip. Both of these technologies have been used to study bacterial genomes (19), but their low throughput power and elevated cost have not made them first choices to study small genomes.

Table 1. Summary of the most used sequencing technologies and their main characteristics. Adapted from (19, 22, 23).

Technology (Company)	Year of development	Approach	Read length	bp per run	Cost per Mb (USD)
Sanger sequencing	1975	Synthesis in the presence of dye terminators	Up to 900 bp	1 Mb	$10000
454 (Roche)	2005	Pyrosequencing on solid support	200-300 bp	80-120 Mb	$1000
Hiseq/Miseq (Illumina)	2006	Sequencing by synthesis with reversible terminators	30-40 bp 150-250 bp	500-800 Mb 4-8 Gb	$25-$120
Solid (ThermoFisher)	2007	Massively parallel sequencing by ligation	35 bp	1-3 Gb	$60
Ion torrent (ThermoFisher)	2010	Sequencing by measurement of the H+ ions released during base incorporation	200-400 bp	1-2 Gb	$100-$600
PacBio RS (Pacific Biosciences)	2011	Single molecule real-time sequencing	10-60 kb	500 Mb-1 Gb	$200
Nanopore sequencing (Oxford Nanopore Technologies)	2014	Real-time nanopore sequencing by measurement of changes in ionic current produced by base incorporation	5-6 kb	Defined by the user	$60

The principles and chemistry of each NGS technology are different but all of these technologies produce short reads. Therefore, the reads obtained produce highly fragmented assemblies and a closed genome is hard to achieve. To overcome this problem, third generation sequencing technologies have been introduced in the market, which produce longer reads. PacBio sequencing (Pacific Biosciences, San Francisco, USA) is a high throughput method that has been recently released, that produces sequencing reads by reading in real-time a continuous sequence from the molecular template. Briefly, a specialized cell, called the SMRT cell is used, that contains a single DNA polymerase molecule immobilized in each well. The libraries are prepared by shearing the DNA into double-stranded fragments (2-5 kb) and ligating hairpin adapters to each end of the fragments (http://www.pacificbiosciences.com/), thus creating a single-stranded end. The sequencing reads are obtained in real-time, by monitoring the incorporation of fluorescent-dyed nucleotides in the nascent DNA chain. This method allows obtaining long sequence reads (3-9 kb) (Table 1); the drawback is that it is prone to sequencing errors and it is still very costly (Table 1). More recently, a third generation sequencing technology developed by Oxford Nanopore Technologies (Oxford, UK) has been introduced in the market (Table 1). This technology is incorporated in a small machine called MinION. Briefly, the technology is based in the measurement of changes in the electronic current as single molecules of DNA are passed through the DNA polymerase, immobilized in a biological nanopore. By using a hairpin adapter, each molecule is read twice and the resulting reads are long. Although it is suitable to assemble closed genomes, the error rate has been reported to be high (24). This technology is promising though, because of the low cost of sequencing a genome in comparison with PacBio sequencing (23)(Table 1).

1.2.3.2. Whole-genome sequencing analysis

The limiting step of whole-genome sequencing (WGS) has been the low availability of user-friendly software to analyze the huge amounts of data produced by this technology. However, the recent application of WGS to the molecular epidemiology of bacteria, including *Staphylococcus*, is trending towards an increase, with the emergence of improved analysis methods and user-friendly software. The fast development of accessible informatics tools to analyze clinically relevant targets has allowed the application of WGS in clinical laboratories.

1.2.3.2.1. Closed reference genome-based analysis

The first reports that applied WGS to the study of *Staphylococcus* epidemiology used a method called single-nucleotide polymorphisms (SNP) analysis and was developed for *S. aureus* (25). This method involves aligning the reads produced by sequencing the genome of the isolate, with a fully sequenced and annotated reference genome. A phylogenetic reconstruction with this data is performed with appropriate algorithms, the end result being a SNP tree (25). This method has proven to be very useful to detect bacterial lineages, as well as SNP variation in the part of the genome that is shared between all isolates tested and the reference strain (25-28). However, interpretation of the results is not yet optimal, since the variation of the number and genetic location of SNPs among clonal lineages and even inside the same clonal lineage is not uniform. Therefore, an appropriate cut-off value for clonal relatedness in case of outbreaks is difficult to define and depends on the type of collection analyzed. Thus interpretation of the SNP tree should be done only locally, using appropriate epidemiological information (29). The recent advances in phylogeny reconstruction with a Bayesian inference, such as BEAST (30), that takes into consideration also the epidemiological data, are promising and have proven to be suitable to use in surveillance and epidemiological studies (31, 32).

1.2.3.2.2. *De novo* assembly

An appropriate closed reference genome is not always available. Therefore, assembling the short reads into long contigs, in a reference-free manner, stimulated the interest of researchers and several algorithms have been developed for this purpose, such as VELVET (33). The so-called assembly *de novo* has been widely used in several epidemiological studies. The drawback of this method is the end result, which consists of hundreds of contigs with no overlapping genomic regions that represent different segments of the genome (22). To overcome this difficulty, alignment of the contigs, produced for each strain by VELVET, is achieved with powerful iteractive alignment algorithms, such as Mauve (34). Mauve can align a reasonable number of genomes, partitioned in hundreds of contigs, while accounting for rearrangements (22).

Another useful approach to deal with the high number of genomic contigs of a given isolate is to perform a hybrid assembly, which consists in combining the reads obtained with Illumina (or other NGS technologies) with the reads obtained with PacBio sequencing (35) or nanopore sequencing reads (23) . In this approach, the hybrid assembly method allows for correction of sequencing errors produced with third generation sequencing technologies, since it incorporates the short reads obtained by NGS technologies (35). Powerful assemblers that allow error corrections, have been developed to work with reads from different technologies, such as SPAdes (36) and CELERA (37), among others.

Finally, a commonly used strategy to analyze the high number of contigs produced by the *de novo* assembly is to perform a gene-by-gene analysis, or whole-genome MLST (38). This “super MLST” utilizes the assembly *de novo* data and consists in extracting alleles of genes present in the predicted core genome of all strains (38, 39). Following the rationale for MLST data, an allele number is assigned to each gene and an allelic profile is attributed to each strain. This strategy was first implemented in an open-source database system called the BIGSDB (bacterial isolate genome sequence database)(38). A similar approach has been recently described and incorporated in an automated program, called the SeqSphere (Ridom GmbH,

Germany). This method has proven to be valuable to infer epidemiological relatedness between isolates in outbreak situations (39). Another method consists on predicting the core genome of the strains, by selecting the genes that are present in all strains (22). An alignment can then be produced with the concatenated gene sequences of the predicted core genome, by using algorithms that have been widely used before for single gene alignments, such as CLUSTALW (40) and MUSCLE (41). Phylogenetic trees are then performed using the resulting alignment. This strategy has been recently used to study the core genomes of *S. aureus* and *S. epidermidis* (42).

1.2.3.2.3. Downstream analysis

The contigs obtained in the assembly *de novo* can be used for many objectives. Several tools have been developed to study draft genomes, particularly of bacterial origin. For instance, different algorithms and databases have been recently developed to rapidly detect antibiotic resistance genes, with the interest of aiding clinicians in the course of treatment of infectious diseases (43). Other databases and web-based servers have been developed, which identify virulence genes, plasmid genes, prophages and insertion sequences, among many other functional categories; examples are the antibiotic resistance genes online database ResFinder, https://www .cge.cbs.dtu.dk/services/ResFinder/, the insertion sequence online database ISFinder, https://www-is.biotoul.fr, among others. The great majority of these platforms work with the BLAST algorithm that identifies homologues of the genes in the contigs by comparing these with the ones deposited in a database.

One useful and most used approach is functional annotation of genomes or genomic regions. In these cases, the annotation is usually confined to the protein-coding sequences (CDS) present in contigs containing regions of interest of the genome. Briefly, the annotation process is usually based in programs that use *ab initio* algorithms trained on gene models from related species (for instance, AUGUSTUS (44)) or gene alignments (using for example tblastx) and databases (such as KEGG) that complement the predicted gene models (45). However, as the evidence available for some species is mostly incomplete and sometimes contradictory, this is a difficult task

that often benefits from manual curation. A complete genome annotation represents a considerable effort and requires bioinformatic proficiency; furthermore, it strongly depends on the quality of the genome assembly. Only near-complete genomes interrupted only by small gaps will yield satisfactory results (45).

1.3. The opportunistic pathogen *Staphylococcus epidermidis*

1.3.1. *Staphylococcus epidermidis* as a commensal

Staphylococcus epidermidis is mainly a commensal of the human skin, being one of the first bacteria that colonize the skin of newborns after birth (46). Its main niches are the humid areas of the human skin, such as the anterior nares, the axillae, the inginal and perineal areas, the toe webs and also the conjunctiva (2), where it forms populations of cells, which are often composed of multiple strains of this species (47). *S. epidermidis* can be also found colonizing the skin of non-human mammals, but this colonization is thought to be transient and of human origin (48, 49). Nevertheless, relatively high frequencies of carriage of *S. epidermidis* have been described, particularly in production animals in Europe (40%)(50).

While colonizing, *S. epidermidis* cells are believed to adhere to the human skin (51). In fact, *S. epidermidis* genome is well equipped with genes for adhesion, such as adhesins (*aae* and *sdrG*, among many others, see below) (52). The role of *S. epidermidis* as part of the protection barrier provided by the skin is well documented. In fact, some studies suggest that *S. epidermidis* has an antagonistic effect against more virulent species, like *S. aureus*. One example of this is the inhibition of *S. aureus* by *S. epidermidis* through the excretion of the serine protease Esp; specifically, it has been found that purified Esp inhibits biofilm formation by *S. aureus* and destroys pre-existing biofilms (53). Furthermore, it was shown that Esp enhances the susceptibility of *S. aureus* biofilms to immune system components. Other exoproteins produced by *S. epidermidis* that have antimicrobial activity against *S. aureus* are the PSMs (phenol-soluble modulins, see below) (54, 55) (Figure 2). In addition, it has been showed that *S.*

epidermidis strains carrying specific alleles of the quorum-sensing system *agr* (one of the most important virulence regulons in staphylococci) inhibit the colonization of the skin of healthy volunteers by *S. aureus* strains carrying specific *agr* alleles, which suggests that cross-interference between *agr* alleles affects negatively the colonization of these bacteria of the same niche (56). Furthermore, *S. epidermidis* strains with the ability to produce bacteriocins have been also identified (57). In particular, epidermicidin produced by a *S. epidermidis* strain recovered from a human skin swab, was shown to have an inhibitory activity against *S. aureus*, other CNS species and enterococci (58) and a thiopeptide produced by *S. epidermidis* of avian origin, was reported to have activity against several different gram-positive pathogens (including *S. aureus*, enterococci, *Clostridium difficile* and others) (59). Finally, several studies have showed that *S. epidermidis* can interact with the human immune system. Specifically, the main component of *S. epidermidis* biofilms, PIA (polysaccharide intercellular adhesin, see below), is able to stimulate the immune system through binding to toll-like receptor 2 (TLR2) present in human cells (60) (Figure 2). PIA was found to be able to induce the production of pro-inflammatory cytokines via TLR2. *S. epidermidis* can also interact with antimicrobial peptides (AMPs). AMPs are small peptides that have antimicrobial activity and are part of the innate immune system. A study has showed that a PSM produced by *S. epidermidis* can boost the production of AMPs by the host, leading to an increased killing capacity of human neutrophils against pathogenic bacteria such as group A streptococci (55). Altogether, these data suggest that the presence of *S. epidermidis* in the human skin contributes to its homeostasis and prevents colonization by pathogenic species.

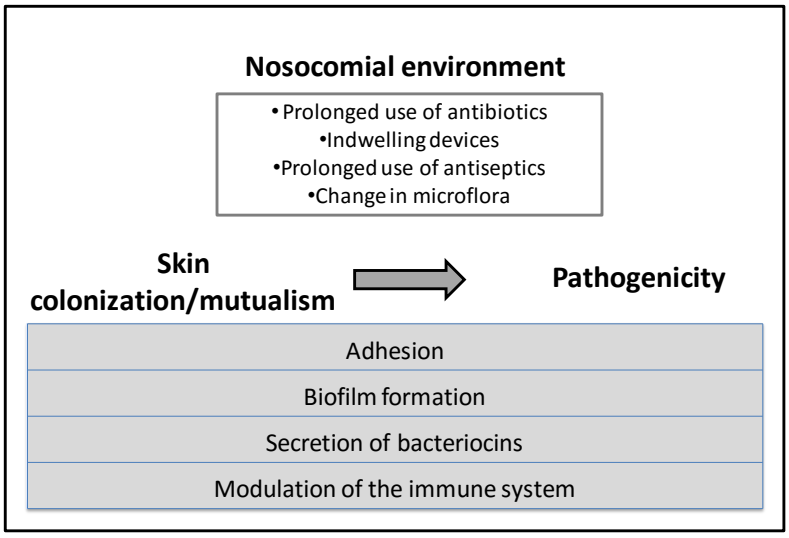


Figure 2. Phenotypic characteristics of *Staphylococcus epidermidis* that contribute to its mutualism with the human skin and/or its pathogenic potential, as well as the conditions of the nosocomial environment that prompt the change in *S. epidermidis* lifestyle. Adapted from (51, 61).

1.3.2. *Staphylococcus epidermidis* as a pathogen

S. epidermidis is also an important opportunistic pathogen, particularly once the skin barrier is compromised. The same characteristics that make *S. epidermidis* such a good colonizer and protector of its niche, can also turn it into a powerful pathogen (Figure 2). In particular, the capacity to form biofilm and the production of PSMs (see below) are important for both colonization and infection lifestyles of *S. epidermidis* (51).

S. epidermidis is one of the most relevant pathogens in the hospital setting, being responsible for around 70% of native valve endocarditis, 50% of neonatal sepsis and 30-50% of infections associated with indwelling devices in hospitals worldwide (2, 61). Most of the infections in the hospital environment are thought to be of endogenous origin, being caused by the strains that colonize the patient. However, transmission between the nose and hands of staff personnel and patients (62), healthcare clothing (63), medical equipament (64), and hospital surfaces (65) have also been observed.

S. epidermidis infections are usually associated with an extended hospitalization stay and therefore constitute an overall burden for patients and healthcare providers. In addition, with the increasing use of indwelling devices in modern medicine, the incidence of *S. epidermidis* infections has the tendency to increase. On the other hand, in the community*, S. epidermidis* infections are rare (<5%), but urinary tract infections, particularly in children, have been reported (66), as well as community-acquired native valve endocarditis (2). In addition, peritonitis associated to peritoneal dyalisis, considered to be a community-onset infection, increased from 7 to 19% in the last decade, in Germany (67).

The treatment of *S. epidermidis* infections, often associated with biofilm formation, is further complicated by the high rates of resistance to multiple classes of antimicrobials that this pathogen exhibits, particularly in the hospital environment. One of the most successful classes of antibiotics for the treatment of *S. epidermidis* infections have been β-lactams, but emergence of resistance to these antibiotics has been hindering its use in clinical practice.

Penicillin-resistant *S. epidermidis* has been reported in 1949 (68), shortly after the introduction of penicillin in the clinical practice. Methicillin, a semi-synthetic antibiotic resistant to the action of β-lactamase was introduced in the clinical practice in 1959, to overcome the problem of penicillin-resistant strains. Nonetheless, only two years after, the first methicillin resistant staphylococcal strain, a methicillin-resistant *Staphylococcus aureus* (MRSA) strain, emerged in the UK (69). In addition, by 1963, about 10 % of *S. epidermidis* isolates collected from infections in Denmark were resistant to methicillin (methicillin-resistant *S. epidermidis*, MRSE isolates) (70). In recent studies, 90% of *S. epidermidis* isolates in hospitalized patients are resistant to penicillin (71) and the oxacillin resistance reached 80% (2, 72). The inefficacy of β-lactams in the treatment of nosocomial *S. epidermidis* infections is worrisome since other antibiotics are not so well tolerated, are more expensive, and frequently require intravenous administration (like for instance, the glycopeptide vancomycin).

In addition to β-lactams, resistance to ciprofloxacin, clindamycin, erythromycin, chloramphenicol, fusidic acid, fosfomycin, rifampin and tetracycline has been

described in *S. epidermidis* recovered from bloodstream and catheter-related infections worldwide (73, 74). The frequency of isolates exhibiting resistance to each of these antibiotics has been lower than 20% (2). More rarely observed, but already reported, is resistance to glycopeptides (75) and linezolid (76). In addition, a high increase on the frequency of multiple antibiotic resistant *S. epidermidis* has been observed in the last decade, from 30% to 60% in the USA (2, 74). This steady increase in the frequency of multidrug-resistant *S. epidermidis* is related with the increase of the use of antibiotics and invasive procedures in the hospital environment.

1.3.3. Molecular epidemiology and population structure

Several studies have focused on the molecular epidemiology of *S. epidermidis* recovered from the hospital environment. The great majority of these studies have used molecular tools and, more recently, whole-genome sequencing to elucidade the population structure of nosocomial *S. epidermidis.*

One study where a representative collection comprising isolates from 17 countries worldwide was characterized by MLST and eBURST analysis, showed that most of the sequence types identified among hospital-associated isolates analyzed (31%) belonged to a single clonal complex, CC2 (Figure 3) (17). CC2 was found to be highly diverse; isolates belonging to this CC show diverse PFGE types and a high number of different sequence types (39 STs) (17). The clonal lineage CC2 was firstdescribed as being composed of two different clusters; cluster I includes ST2 and related sequence types, while cluster II comprises all the other sequence types organized in subgroups, each with a subgroup founder (ST5, ST6, ST57, ST85 and ST89). In addition to the complex and diverse CC2, nine other clonal complexes were identified (CC1, CC2, CC11, CC21, CC23, CC33, CC42, CC49 and CC66) (17). CC2 isolates have been described to be particularly virulent and invasive, usually strong-biofilm producers (77) and to carry insertion sequences that allow for the efficient modulation of this phenotype (78). However, the genes underlying CC2 success have not yet been identified or confirmed in a large number of isolates or using genome-wide studies.

In a different study, *S. epidermidis* isolates with exactly the same PFGE macrorestriction pattern have been identified in different geographic regions, which indicate that *S. epidermidis* isolates are able to disseminate (79). On the other hand, using the nucleotide sequence of an internal fragment of the seven housekeeping MLST genes, it has been estimated that the recombination/mutation rate of *S. epidermidis* was 2.5:1 (17). Since *S. epidermidis* was found to have a high capacity of clonal dissemination associated to a high recombination rate, the authors suggested that the structure of *S. epidermidis* population is of the epidemic type where recombination is very frequent, and that epidemic clones would emerge only sporadically (17). In a recent report, the core genome of *S. epidermidis* has been studied in detail, including nosocomial and commensal isolates (42). In this study, a phylogenetic tree constructed using the results of the whole-genome MLST approach showed that the *S. epidermidis* population was clustered in three distinct clusters, A, B and C (Figure 3). Most of the hospital-associated isolates, belonging to CC2, fell into clusters A/C while the commensal isolates, which belonged in its majority to other minor clonal complexes, were included in cluster B (42). Recombination events were observed in a large number of genes between isolates belonging to the same cluster; specifically, an estimated 40% of the core genome genes showed evidence of homologous recombination in at least one lineage (42). Interestingly, annotation of these genes has revealed that the great majority seemed to be involved in adaptation to the hospital (metal toxicity resistance, cell wall and capsular synthesis, and antimicrobial resistance). Therefore, the authors suggested that *S. epidermidis* adaptation to the nosocomial environment was most probably achieved through recombination.

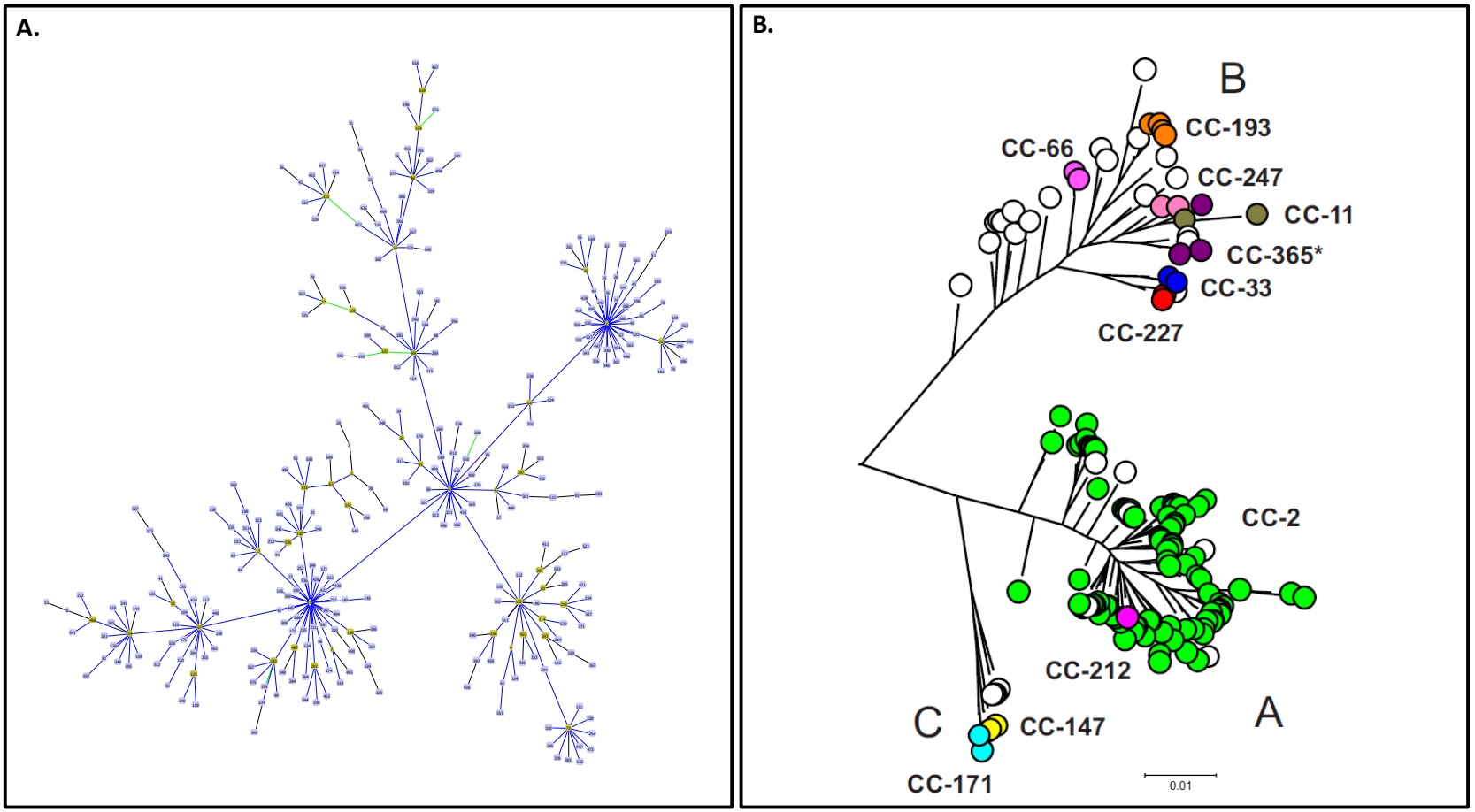


Figure 3. Population structure of *S. epidermidis*. (**A.**) eBURST analysis of the sequence types (STs) deposited on the dabatase, July 2015 (www.mlst.net). For the sake of simplicity, only clonal complex 2 (CC2) is shown. The most prevalent STs are represented by a bigger font. Light green STs indicate probable ancestors (group founders) and dark green STs constitute subgroup founders. Blue STs correspond to STs that share the same background (CC) (**B.**). Alignment of the concatenated genes predicted to be part of the core genome (42).

At the beginning of the work presented in this Thesis, the population structure of *S. epidermidis* in the healthy community was poorly assessed. The few studies available at the time indicated that antimicrobial resistance was low and that the genetic background of the isolates presented a high genetic diversity (80, 81). In addition, cross-colonization of *S. epidermidis* isolates between individuals sharing the same households had been observed (82). However, it was not known whether strains colonizing individuals in the community were the same or different from those causing infections in the hospital. Moreover, a comprehensive study comparing the hospital-associated and the community-associated *S. epidermidis* populations had not been performed.

1.3.4. Virulence genes

S. epidermidis virulence and modes of infection have not been extensively studied. Whole genome sequencing of *S. epidermidis* show few virulence genes, such as genes encoding high molecular weight toxins or superantigens. However, *S. epidermidis* carries a large number of genes involved in biofilm formation, immune evasion, bacteriocins and exoproteases (51). These genes are thought to be associated with colonization, but due to the opportunistic character of *S. epidermidis* life style, they have also been considered as virulence factors (Table 2).

Table 2. *Staphylococcus epidermidis* virulence factors. Adapted from (51, 73).

Virulence factor	Genetic determinant	Function/mechanism
Proteins and polysaccharides	*icaA, icaB, icaC, icaD, aap, bhp*	Biofilm formation
Adhesins	*aae, atlE, bhp, ebp, embB, gehD, scaA, scab, sdrF, sdrG, tagF, ybiD*	Adherence to biomaterials/host proteins
Capsule	*capA, capB, capC, capD*	Protection against phagocitose
Toxins	*psmα, psmβ1, psmβ2, psmδ, psmε, hld*	Cytolitic activity
Enzymes and proteases	*gehC, gehD, sspA, sspB, sspC, sepA, esp*	Host tissue destruction Colonization Bactericidals Biofilm formation

Biofilm formation

S. epidermidis infections are mainly associated with the use of indwelling devices primarily due to its ability to produce biofilm. Biofilms are composed by a cellular mass with the bacterial cells disposed in layers. Besides providing the cells protection against external aggressions, the biofilm cells present a low activity metabolic state, which limits the action of most antibiotics that act on live and metabolically active cells (51, 83, 84).

The formation of a biofilm is a multifactorial process, which involves four different steps (Figure 4): initial attachment, cellular aggregation, maturation and detachment. In the initial attachment, the bacterial cells adhere to a surface, through the action of adhesins and specific autolysins (AtlE, Aae and SdrF/SdrG). Next, the cellular aggregation occurs, resulting in the formation of a multicellular structure,

mediated by teichoic acids, extracellular DNA and exopolyssacharides (51). On the step of maturation of the biofilm, proteases are thought to be involved, such as Esp, that have been described to inhibit the colonization by other species during biofilm maturation (53). In the detachment step, the last phase of the biofilm formation, some cells return to the planktonic lifestyle and eventually start a new biofilm elsewhere in the human host. This step is regulated by the *agr* system, a recognized regulator of virulence factors in *Staphylococcus* (51). Little is known regarding the proteins that are responsible for the development of this step, but proteases and phenol-soluble modulins (PSMS) might be involved (51, 85).

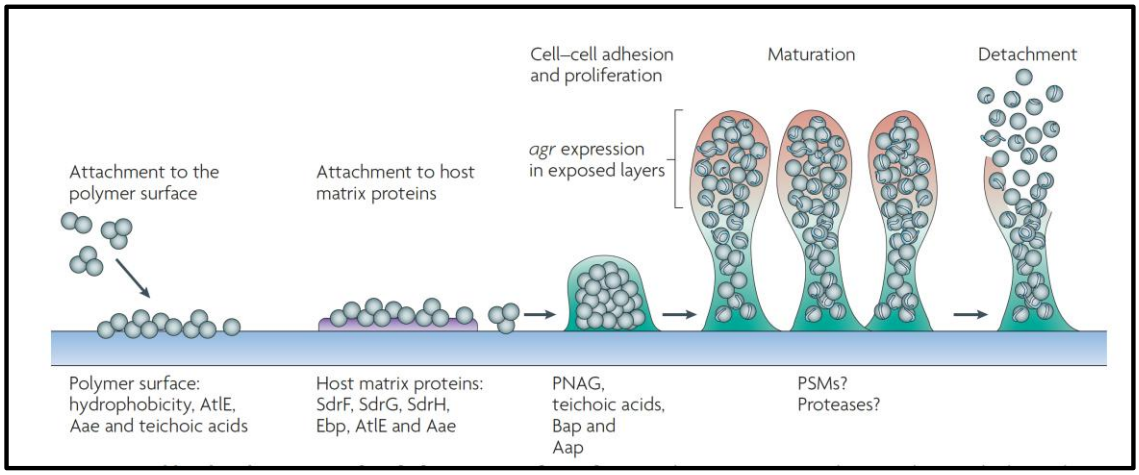


Figure 4. Steps that lead to *S. epidermidis* biofilm formation and proteins thought to be involved (51).

One of the most important molecules for biofilm formation in *S. epidermidis* is the exopolyssacharide called PIA (polysaccharide intercellular adhesion), a polymer of N-acetyl-glucosamine (PNAG). PIA is the main component of the extracellular matrix, allowing for the growth and providing stability to the biofilm. This polymer is encoded by the *ica* operon, composed by five genes (*icaA, icaB, icaC, icaD* e *icaR*). The *icaA* and *icaD* encode proteins involved in the production of N-acetyl-glucosamine; *icaB* is involved in the deacetylation of these monomers and *icaC* in their translocation. Finally, *icaR* encodes a repressor of the *ica* operon (84). An alternative regulatory system of this operon has been described, through the insertion/excision of the IS element IS*256* on the *icaC* gene (86). In addition, irreversible switching from biofilm-

positive to biofilm-negative phenotype has been observed, that has been suggested to occur through spontaneous deletion of the *ica* operon, caused by deregulation of the *recA* expression, a gene that has been implicated in recombination (87).

The frequency of the *ica* operon is high among isolates associated with nosocomial infections (50-85%), particularly in isolates from catheter-associated infections (88), orthopedic transplants (89, 90) and bloodstream infections (91), which illustrates its clinical relevance. However, some reports describe the occurrence of infections caused by *S. epidermidis* isolates that did not carry the *ica* operon and for which the production of biofilm was mediated by the surface proteins Aap and Bap (51, 92), although these were rare.

Phenol-soluble modulins (PSMs)

The phenol-soluble modulins (PSMs) are small proteins, with an amphipathic α-helix structure. PSMs can be of the α type, that have a neutral or positive net charge, and are usually small in size (20-25 aminoacids); or of the β-type, that have a negative net charge and are longer (43-45 aminoacids)(93). The *psm* genes are encoded either on the core genome or on mobile genetic elements (93). These small proteins were described to act upon the human immune system (51), through the induction of cytokines release by macrophages, thus promoting inflammation (94). Additionally, they can have cytolytic activity towards leukocytes (95). *S. epidermidis* can produce cytolytic PSMs such as PSMδ (96), but due to its mutualistic lifestyle, the greatest amounts of PSMs produced by *S. epidermidis* correspond to less aggressive PSMs such as PSMβ peptides (51), which have been described to be involved in the structuring of biofilm and detachment (85).

The arginine catabolic mobile element (ACME)

The arginine catabolic mobile element (ACME) was initially described in *S. aureus* strain USA300, isolated in the community (97). This element is composed by the *arc* operon, encoding genes involved in the metabolism of arginine and by the *opp*-

3 operon, that encodes oligopeptide permeases (98). ACME is considered to be a ψSCC (pseudo-staphylococcal cassette chromosome, see below*).* An element similar, in structure and sequence, to *S. aureus* ACME, was identified in *S. epidermidis* strain ATCC12228 – ACME-I (99) and among hospital-associated *S. epidermidis* (51%), particularly in isolates belonging to the main clonal lineage, CC2 (100). Furthermore, the most frequent ACME allotype identified in hospital-associated *S. epidermidis* was very similar in the nucleotide sequence to the one carried by the *S. aureus* USA300 strain (only 11 nucleotides of difference), which served as evidence to suggest that ACME was originated in *S. epidermidis* (97, 100).

ACME is believed to confer to the staphylococcal isolates a higher capacity of dissemination and host colonization (100), being involved in the capacity of adaptation and competition of the isolates that carry it. In a rabbit model, *S. aureus* isolates carrying ACME-I had a higher capacity of survival and dissemination, than isogenic isolates with a deletion of ACME (98). The fact that ACME-I is carried in high frequency by the main clonal lineage of *S. epidermidis*, suggests that this element also contributes to a higher epidemicity in *S. epidermidis* (101). In fact, it has been demonstrated that *S. epidermidis* carrying ACME have a higher resistance to acidic environments, such as the human skin, during biofilm growth (102). This is partly due to the detoxification of polyamines by the *speG* enzyme, present on the ACME island, as demonstrated with studies with the *S. aureus* USA300 type strain in a murine model (103).

Other virulence factors

Other virulence factors carried by *S. epidermidis* include proteins that seem to be important for the colonization and proliferation of this pathogen in the host, namely adhesins, toxins and extracellular enzymes (Table 2). The adhesins, act mainly in the initial step of the biofilm formation, adhesion to surfaces (AtlE, AAe) and also host proteins (SdrF and SdrG).

The *S. epidermidis* genome also contains elements that allow the evasion of the human immune system. An example is the *cap* operon that encodes a capsule (CapA, a polymer of poly-γ-glutamic acid) that confers resistance to phagocytosis by macrophages (104). Another example is the study where the authors described that PIA protects *S. epidermidis* cells against the action of the hosts AMPs (105).

Finally, the extracellular enzymes produced by *S. epidermidis* are still poorly studied, but they are thought to have a cytolytic activity in host tissues cells and a role in the degradation of fibrinogen and complement proteins. In addition, they might be involved in the maturation step of the biofilm (51).

2. Methicillin resistance: a crucial event in the evolution of *Staphylococcus*

In the last fifty years, epidemics of methicillin-resistant staphylococci (MRS), and particularly MRSA have been frequently reported in the nosocomial setting, with devastating outcomes for patients and hospitals (106). MRS are resistant to all antibiotics of the large class of β-lactams, which constitute one of the first lines of defense against staphylococcal infections. Furthermore, MRS often carries additional antibiotic resistance genes which further complicates the treatment of infections caused by these pathogens.

The staphylococcal cassette chromosome *mec* (SCC*mec*), the central mobile element that carries the genetic determinant responsible for β-lactam resistance (*mecA* gene), and its impact in the epidemiology of staphylococci have been well studied. However, little is known regarding its origin or the steps leading to the construction of this mobile genetic element. In this chapter, the main findings regarding the origin and assembly of SCC*mec* are reviewed.

2.1. Mechanism of resistance

Peptidoglycan synthesis in *S. aureus*, is carried out by four different membrane-bound penicillin-binding proteins (native PBPs 1-4) that catalyse two key reactions of the final steps of cell wall assembly: formation of the linear glycan chains (transglycosilation) and formation of the peptide cross-bridges (transpeptidation) (7, 107). The β-lactam antibiotics are a class of antibiotics that derived from penicillin; they all have in its composition a β-lactam ring. In the presence of β-lactam antibiotics, the transpeptidase domain of native PBPs is inactivated, which impairs peptidoglycan cross-linking (107). These antibiotics acylate, irreversibly, the native PBPs which leads to the interruption of the cell wall synthesis and to bacterial death (7).

The most clinically relevant genetic determinant responsible for methicillin resistance is the *mecA* gene. This gene encodes a 78 kDa high molecular mass penicillin-binding protein, called PBP2a (penicillin-binding protein 2a). PBP2a has two main domains: the non-binding domain and the transpeptidase domain, which contains the active site (108). Resistance is associated to the fact that PBP2a has low affinity for the β-lactam ring (109). Contrarily to susceptible strains, in strains carrying PBP2a and in the presence of β-lactam antibiotics, the transpeptidase domain of PBP2a remains active, which along with the cooperative action of the transglicosylase domain of the native PBP2 assure the final assembly of the cell wall (107). Further insights into this mechanism of resistance have been obtained when the crystal structure of PBP2a was solved. Specifically, it was observed that a torsion of PBP2a, at the active site, was responsible for the low acylation rate by the β-lactam ring (108).

2.2. The staphylococcal cassette chromosome *mec* (SCC*mec*)

2.2.1. Historical perspective

Only two years after the introduction of methicillin (a semi-synthetic β-lactam antibiotic), in clinical practice, the first methicillin resistant staphylococcal strain, a MRSA, emerged in 1959 in the UK (69). Not much is known regarding the earliest MR-coagulase-negative staphylococci (CoNS). The unique data available showed that approximately 10% of *S. epidermidis* isolates collected in hospitalized patients in Denmark in the early 60s were resistant to methicillin (70), but no studies on the genetic background of those strains were performed. The report of the earliest methicillin-resistant *S. epidermidis* (MRSE) strain described corresponds to MRSE isolated in 1973, in Canada, which carried SCC*mec* IV, one of the smallest SCC*mec* cassettes (110).

2.2.2. Basic structure and diversity

SCC*mec* is composed of two structural elements: the *mec* complex and the *ccr* complex (111). The *mec* complex includes *mecA* and its regulators, the repressor *mecI* and the inducer *mecRI*; the *ccr* complex encodes site-specific recombinases that assure the mobility of the cassette. So far, eleven different types of SCC*mec* ranging from 20 to 70 kb have been described in *S. aureus* (111-114), that vary on the combination of the class of *mec* complex and the type of *ccr* that each cassette carries (111) (Figure 5).

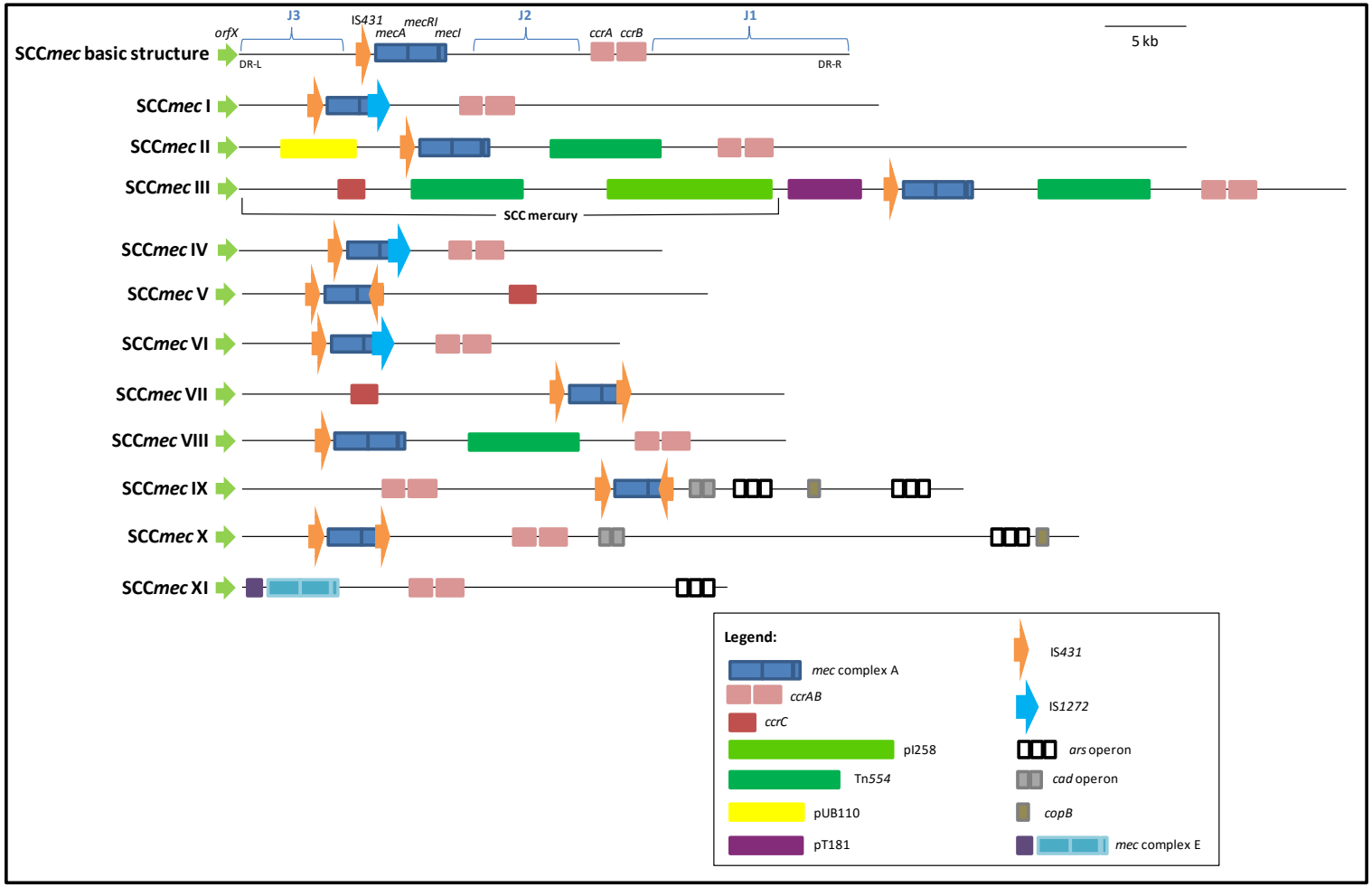


Figure 5. Schematic representation of the structural elements carried by the eleven SCC*mec* types (SCC*mec* I-XI) identified in *S. aureus* (Genbank accession numbers: SCC*mec* I, AB033763.2; SCC*mec* II, D86934; SCC*mec* III, AB037671; SCC*mec* IV,AB0631722; SCC*mec* V, AB121219; SCC*mec* VI, AF411935.3; SCC*mec* VII, AB373032; SCC*mec* VIII, FJ390057; SCC*mec* IX, AB505628.1; SCC*mec* X, AB505630; SCC*mec* XI, whole-genome sequencing data of the strain LGA251, FR821779.1).

The *mec* complex and the *ccr* complex are linked by three joining regions (J1-J3) that vary in size, according to the number of genes they carry. In these regions, most cassettes carry a combination of different insertion sequences, transposons, plasmid sequences, antibiotic resistance genes and heavy metal resistance genes.

2.2.3. Transfer and mobility

SCC*mec* is a mobile genetic element; its insertion in the chromosome occurs always at the same locus, downstream of an open reading frame, originally called *orfX* (109). The *orfX* (or *rlmH*) has been recently described to encode a RNA methyltransferase (115). Upon integration of SCC*mec* downstream the 3' end of *orfX*, at the attachment site *attB*, 15 bp direct and/or inverse repeats (DR/IR) are created, that delimit the element in the chromosome (DR-left downstream *orfX*, *attL*; DR-right at the end of SCC*mec*, *attR*). When SCC*mec* is excised, an extrachromosomal circular intermediate is formed and the recognition sites *attSCC* are created in the chromosome and the intermediate circular form (109). These insertion/excision events are catalyzed by serine recombinases encoded by the cassette chromosome recombinases (*ccr*) genes that recognize these *att* recognition sites.

SCC*mec* is carried by different strains and species, which suggests it is frequently transferred. Some studies suggest that SCC*mec* is probably transferred by phages, since encapsidation of the element and transduction to sensitive strains has been previously observed (116-118). Other authors have suggested that SCC*mec* could be transferred by natural transformation. Evidence for that come from the observation that natural competent *S. aureus* (induced by activation of SigH) could incorporate SCC*mec* and become methicillin resistant (119). Finally, an homologue of *mecA* has been found in a plasmid in *Macrococcus caseolyticus*, a species phyogenetically related to *Staphylococcus*, which might indicate that plasmids might also be a vehicle of transmission of this locus (120). However, this is still a matter of debate. Likely several mechanisms of transfer can be responsible for SCC*mec* dissemination.

2.2.4. The *mec* complex

The *mec* complex is classified into classes, according to the structure and type of its composing elements. To date, six different classes of *mec* complex have been defined (A-E)(111)(www.sccmec.org) (Table 3).

Table 3. Structure of each SCC*mec* type, as defined by the combination of the class of *mec* complex and the type of *ccr*. The main staphylococcal species for which the SCC*mec* type has been identified are also shown.

SCC*mec* type	*mec* complex	*ccr* complex	Staphylococcal species	References
I	B	1 (A1B1)	*S. aureus, S. hominis*	(121, 122)
II	A	2 (A2B2)	*S. aureus, S.epidermidis*	(123, 124)
III	A	3 (A3B3)	*S. aureus, S.epidermidis*	(123, 124)
IV	B	2 (A2B2)	*S.epidermidis, S. aureus*	(110, 125)
V	C2	5 (C1)	*S. haemolyticus, S. aureus*	(126, 127)
VI	B	4 (A4B4)	*S. aureus*	(128)
VII	C1	5 (C1)	*S. aureus*	(129)
VIII	A	4 (A4B4)	*S. aureus, S. hominis*	(130, 131)
IX	C2	1 (A1B1)	*S. aureus*	(112)
X	C1	6 (A1B6)	*S. aureus*	(112)
XI	E	7 (A1B3)	*S. aureus, S. xylosus, S. sciuri*	(113, 132, 133)

The *mec* complex A corresponds to a complete *mec* complex, with intact forms of the regulators *mecI* and *mecRI*. The expression of *mecA* is regulated by the repressor protein MecI (encoded by the *mecI* gene) and the inducer MecRI (encoded by the *mecRI* gene)(134). In addition, *mec* complex A carries an anti-repressor of *mecA* that has been described recently, called *mecR2* (135), which interacts with MecI, leading to its proteolytic cleavage. The *mec* complex A also harbors the so-called hypervariable region (HVR), downstream *mecA*. The HVR region includes a truncated *mvaS* gene encoding the 3-hydroxy-3-methylglutaryl-coenzyme A (HMG-CoA) synthase, an intact *ugpQ* gene encoding glycerophosphoryl diester phosphodiesterase, and an intact *maoC* gene encoding the acyl dehydratase MaoC. Another region included in the HVR is the *dru* (direct repeat unit) locus (136), that is a variable region, composed of repeats of 40 bp in tandem. Upstream HVR, *mec* complex A (and all classes of *mec*

complex, with the exception of class E) carries a copy of IS*431* (frequently called IS*431mec*). The *mec* complex A includes an ORF, located immediately downstream *mecA* that encodes a PSM. This ORF has been called *psm-mec* and has been associated with an enhanced toxicity ability of *S. aureus*, since it has been shown that has cytolytic activity against white and red human cells and was able to induce inflammation (137). This *mec* complex class is part of SCC*mec* types II, III and VIII (130, 138) (Figure 5).

The elements present in *mec* complex A are not always present in the remaining classes of *mec* complex (139) (Table 4).

Table 4. Structure of the classes of *mec* complex and the types of *ccr* complex described to date (August 2015) (www.sccmec.org).

SCC*mec* structural element	Structure
***mec* complex**	
A	IS*431-mecA-mecRI-mecI*
B	IS*431-mecA-ΔmecRI*-IS*1272*
C1	IS*431-mecA-ΔmecRI*-IS*431*
C2	IS*431-mecA-ΔmecRI*-IS*431**
D	IS*431-mecA-ΔmecRI*
E	*blaZ-mecC-mecRI*$_{LGA251}$-*mecI*$_{LGA251}$
***ccr* complex**	
Type 1	A1B1
Type 2	A2B2
Type 3	A3B3
Type 4	A4B4
Type 5	C1
Type 6	A5B3
Type 7	A1B6
Type 8	A1B3

*IS*431* arranged in the opposite direction.

The *mec* complex B comprises a deletion of *mecRI*; in this class of *mec* complex, *mecRI* is interrupted by a copy of IS*1272* (111, 140). SCC*mec* types I, IV and VI all carry *mec* complex B (128, 138, 141). The *mec* complex C also carries deletion of *mecRI* but, in this case, interrupted by IS*431*. This class of *mec* complex carries therefore two copies of IS*431* (IS*431* and IS*431mec*). In *mec* complex C1, the two copies of IS*431* are arranged in the same direction of transcription, while in *mec* complex C2, the two copies are in opposite directions. The *mec* complex C1 is part of SCC*mec* types VII and X (112, 129), and *mec* complex C2 is part of SCC*mec* types V and IX (112, 127). The *mec* complex D is the smallest class of *mec* complex: it contains only a small part of *mecRI* truncated, but with a deletion of IS elements downstream *mecRI*. It has not been associated to any SCC*mec* type, but it has been found sporadically in the *orfX* region of CoNS isolates, inside delimited SCC*mec* elements (140). Finally, the recently described *mec* complex E, which is part of SCC*mec* XI (113), carries the *mec* homologue *mecC* and intact forms of its regulators *mecI*$_{LGA251}$ and *mecRI*$_{LGA251}$, as well as a copy of *blaZ*$_{LGA251}$ upstream *mecC*.

2.2.5. The *ccr* complex

The *ccr* complex comprises the *ccr* genes, which encode site and orientation specific recombinases of the invertase-resolvase family (109), and surrounding ORFs of unknown function. Three different *ccr* genes have been described in *S. aureus*, *ccrA*, *ccrB* and *ccrC*, that share nucleotide sequence similarities below 50% (111). Two distinct groups of *ccr* complexes have been described: *ccrAB*, which includes *ccrA* and *ccrB* genes, which are usually found adjacent to each other, and *ccrC*. In addition, sequence polymorphisms in these genes have been identified, that led to the definition of different allotypes of *ccrA* and *ccrB* genes. In general, the *ccrA and ccrB* genes belonging to the same allotype have over 85% of nucleotide sequence identity (111). For each allotype of *ccrA* and *ccrB*, several different alleles have been identified (142). The *ccrC* genes are more conserved, but several different alleles of this gene have also been described. To date, eight different *ccr* complexes have been identified, that vary in the combination of *ccr* allotypes (www.sccmec.org). The most

disseminated *ccr* complexes are *ccr* complex 1 (A1B1), present in SCC*mec* types I and IX (112, 138); *ccr* complex 2 (A2B2), part of SCC*mec* types II and IV (138, 141); *ccr* complex 4 (A4B4), present in SCC*mec* types VI and VIII (128, 130) and finally *ccr* complex 5 (*ccrC1*) which is part of SCC*mec* types V and VII (127, 129) (Tables 3 and 4).

Both *ccrA* and *ccrB* genes are required for the precise excision and integration of SCC*mec* in the chromosome (109). Several studies have been performed with the aim of assessing the functionality of the Ccr proteins. Data gathered so far suggests that while CcrB binds DNA, the physical interaction between CcrA and CcrB, in a precise ratio, is required for *attB* site-specific recognition and SCC*mec* chromosomal insertion (143). The integration of SCC*mec* in the chromosome is dependent on the recognition of sequences flanking the *att* site and by the presence of more than one *att* site on either the chromosome or the integration sequences (144). In addition, it has already been demonstrated that CcrC can also promote the precise excision of SCC*mec* V (127).

The origin of the *ccr* complex is not clear, but homologues of these genes have been identified in other gram-positive bacteria. In *Enterococcus* species (*Enterococcus faecium, Enterococcus durans, Enterococcus hirae, Enterococcus casseliflavus* and *Enterococcus gallinarum*), homologues of *ccrA* and *ccrB* genes share less than 40% of nucleotide sequence identity with *S. aureus ccrA* and *ccrB* (145). However, the chromosomal location of these genes or their functionality has not been assessed. A different set of *ccrA* and *ccrB* homologues has been identified in *Macrococcus caseolyticus*, in a SCC*mec*-like structure, that shared less than 52% of sequence identity with *S. aureus ccrAB* genes (146). In this species, these homologues were able to excise the SCC*mec*-like element from the chromosome of *M. caseolyticus* and, like in *S. aureus*, give rise to a circular intermediate (146).

2.2.6. SCC elements and pseudo SCC

SCC elements not carrying the *mec* complex but carrying other genes relevant for staphylococcal survival, competition and virulence have been described. They have been called SCC non-*mec* or simply SCC, since they carry a *ccr* complex, also insert at the *orfX* 3' end and are flanked by DR/IR at the chromosome. These SCC elements (Table 5) can carry heavy metal resistance genes, like SCC*mer*, that carries the *mer* operon (encoding resistance to mercury) and the *cad* operon (encoding resistance to cadmium) (147); capsule genes, like SCC*cap* (148); or even cell-wall synthesis genes, such as *pbp* and *tagF* (SCC*pbp4*) (149). Frequently, two or more SCC – named composite elements - are inserted in tandem, downstream *orfX*. An example is the composite element widely disseminated in the pandemic MRSA Brazilian clone and in some nosocomial *S. epidermidis* isolates, SCC*mec* III-SCC*mercury* (17, 147).

Other structures, the so-called ψSCC, have been identified inserted in the *orfX* region and flanked by DR/IR. However these structures do not have the *ccr* complex and therefore theoretically they cannot be mobilized by themselves, but can be mobilized by adjacent SCC. Different ψSCC structures have been described in *Staphylococcus* (Table 5). One of the most epidemiologically relevant ψSCC element is the arginine catabolic mobile element (ACME)(98), which is usually found downstream SCC*mec* IV, which together constitute a composite element (98).

Table 5. Examples of SCC and ψSCC structures identified in staphylococci. Only the species where the element was first identified is showed.

SCC	*ccr*	Other structural elements	Species	Reference
SCC$_{mer}$	*ccrC*	Tn554, pI258	*S. aureus*	(147)
SCC-CI	*ccrAB2*	*mer, cad* operons	*S. epidermidis*	(149)
SCC$_{pbp4}$	*ccrAB4*	*tagF, pbp4*	*S. epidermidis*	(149)
SCC$_{12263}$	*ccrAB1*	M.*St*sl modification methylase	*S. hominis*	(150)
SCC$_{h1435}$	*ccrC*	*ars* operon	*S. haemolyticus*	(151)
SCC$_{cap1}$	*ccrC*	*cap* operon	*S. aureus*	(148)
SCC$_{15305cap}$	*ccrC*	*cap*, *hsd* operon	*S. saprophyticus*	(152)
SCC$_{15305RM}$	*ccrAB*	*repA*, DNA methylase	*S. saprophyticus*	(152)
SCC$_{MSSA476}$	*ccrAB1*	*hsd* operon	*S. aureus*	(27)
SCC$_{SH32}$	*ccrA5B3*	*hsd*, DNA helicase	*S. haemolyticus*	(153)
ψSCC	***ccr***	**Other structural elements**	**Species**	**Reference**
ACME-I	-	*arcA, opp3*	*S. aureus*	(98)
ACME-II	-	*arcA*	*S. epidermidis*	(101)
ACME-III	-	*opp3*	*S. epidermidis*	(101)
ψSCC*mec* SH32	-	*mecA*	*S. haemolyticus*	(153)
ψSCC*mec* WCH1	-	*mecA*	*S. haemolyticus*	(154)
ψSCC$_{19A}$	-	*ars, cad, copB* operons	*S. haemolyticus*	(155)
CC6082	-	*hsdM*	*S. aureus*	(156)

2.2.7. Joining regions (J regions)

Three joining regions are part of all SCC*mec* types: J1, spanning the region between the DR-right and the *ccr* complex; J2, which includes the region between *ccr* and the *mec* complex; and J3, that includes the region between the *mec* complex and the DR-left, downstream *orfX* (Figure 5). These regions are very diverse in the number and type of genes that they carry, but most correspond to genes conferring resistance to non-β-lactam antibiotics and to heavy metals. In each SCC*mec* type a large number of ORFs encoding hypothetical proteins of unknown function is also present in these regions,. In addition, genes encoding lipases or exported proteins have also been identified in the regions J1-J3 (113, 130).

The combination of *mecA*, encoding resistance to β-lactam antibiotics and genes encoding resistance to other classes of antimicrobials, as well as to heavy metals, has been associated with the success of MRS in the clinical setting. Specifically, some of the most successful hospital-associated MRSA and MRSE genetic backgrounds carry the SCC*mec* cassettes II and III (123, 124), which are the ones that carry the highest number of genes encoding resistance to antimicrobials and heavy metals. SCC*mec* type II carries the transposon Tn*554* in the J2 region, conferring resistance to erythromycin and spectinomycin and the plasmid pUB110 in the J3 region, which confers resistance to kanamycin. SCC*mec* type III carries the transposon ψTn*554* in the J2 region, which confers resistance to cadmium and the plasmid pT181 in the J3 region, conferring resistance to tetracycline. Besides these elements, downstream SCC*mec* III, SCC*mercury* is usually found, that carries pI258 (that encodes resistance to mercury) and Tn*554* (138) (Figure 5).

The recently described SCC*mec* types IX, X and XI, mostly associated with colonization and infection in livestock and farm animals, carry a large number of genes encoding heavy metal resistance in their J regions (112, 113). Specifically, the J1 region of type IX SCC*mec* contains a *cadDX* operon, encoding resistance to cadmium, a *copB* gene, which encodes resistance to copper and two arsenate resistance operons, *arsRBC* and *arsDARBC*. SCC*mec* X contains a *cadDX* operon, a *copB* gene, and an *arsRBC* operon in the J1 region and an *arsRBC* operon in the J3 region (112). Finally, SCC*mec* XI

contains an *arsRBC* operon in the J1 region (113) (Figure 5). Heavy metals and β-lactams are widely used in agricultural productions as growth promoters which is believed to be the driving force for the dissemination of these SCC*mec* types among staphylococci from livestock (112, 157, 158).

2.2.8. SCC*mec* typing

SCC*mec* typing schemes have been extensively developed and used due to the epidemiological relevance of this mobile genetic element in *Staphylococcus*. The identification of the SCC*mec* type has been useful for clone definition (13, 159) and have been crucial for the study of outbreaks, particularly those associated to MRSA.

SCC*mec* typing is based in the amplification of multiple loci that are characteristic of each type using polymerase chain reaction (PCR). Several approaches have been developed, targeting only the *mec* complex and the *ccr* complex (125, 127), or also targeting the J regions (160-164). For some SCC*mec* types, like SCC*mec* IV, typing schemes that allow the discrimination of subtypes have been also developed (165). In addition, a strategy based on the amplification and sequencing of an internal fragment of *ccrB* has also been developed with the objective of identifying simultaneously the SCC*mec* type carried by clinical isolates and the phylogenetic relationships between them (166).

However, with the recent description of new SCC*mec* types, these schemes are no longer updated. Furtheremore, with the advent of low-cost whole-genome sequencing, SCC*mec* typing will probably be performed by analysis of WGS data.

2.2.9. SCC*mec* distribution in staphylococci: CoNS as the origin of specific SCC*mec* types

Eleven SCC*mec* types have been identified in *S. aureus* collected from different environments (types I-VIII are mostly carried by human isolates; while types IX-XI have been mostly associated to animal isolates). The genetic diversity of SCC*mec* carried by CoNS appears to be much higher than in *S. aureus*. Actually, CoNS can carry in addition to all the SCC*mec* types described in *S. aureus* also new combinations of *mec* complex and *ccr* complex (124, 131, 167, 168). In spite of this diversity, the frequency of each SCC*mec* type appears to be species-specific, particularly in human-associated species. In *S. epidermidis*, the most frequent SCC*mec* type found is SCC*mec* IV (79, 110, 124), in *S. haemolyticus* is SCC*mec* V (126, 167) and in *S. hominis* are SCC*mec* IA (a combination of *mec* complex A with *ccrAB1*) and VIII (122, 131). Interestingly, methicillin-susceptible isolates of these same species were found to carry the *ccr* allotypes characteristic of each one of these SCC*mec* types: *ccrAB2* seems to be the most frequent *ccr* type among MS-*S. epidermidis* (Rolo and Miragaia, unpublished), *ccrC* in MS-*S. haemolyticus* (167) and *ccrAB1*/*ccrAB4* in MS-*S. hominis* (131). These data suggests that *S. epidermidis*, *S. haemolyticus* and *S. hominis* might have played a role in the assembly of SCC*mec* IV, SCC*mec* V and SCC*mec* I/VIII, respectively. Examples of additional elements carried by SCC*mec* elements that have probably originated in these CoNS include: the insertion sequences IS*1272* and IS*Sha1* from *S. haemolyticus*, that are part of *mec* complex B (138) and SCC*mec* X (112), respectively; and hypothetical proteins from *S. epidermidis* which are found within SCC*mec* VIII (130).

In addition, different associations of genetic background and SCC*mec* type are usually found among MR-CoNS, indicating that its acquisition is frequent (17). In fact, in a study where a large collection of *S. epidermidis* isolates obtained from inpatients in hospitals worldwide has shown that *S. epidermidis* has acquired SCC*mec* at least 56 times, and the great majority of these acquisitions occurred preferentially within the main genetic lineage CC2 (17). CoNS and particularly *S. epidermidis* constitute therefore a reservoir of specific SCC*mec* elements, particularly in the hospital environment.

2.3. The *mec* homologues in animal-related species: the origin of *mecA*

The rapid emergence of methicillin resistance among staphylococci, shortly after the introduction of methicillin in the clinical practice, has raised the hypothesis that by then *mecA* already existed in the staphylococcal population. Several studies have addressed the search for the origin of *mecA* and different *mec* homologues, with diverse nucleotide sequence identities with *mecA*, have been identified in the most primitive staphylococcal species, which colonize wild and domesticated mammals (169) (Figure 6). However, the role of each *mec* homologue in the evolutionary history of *mecA* has remained elusive.

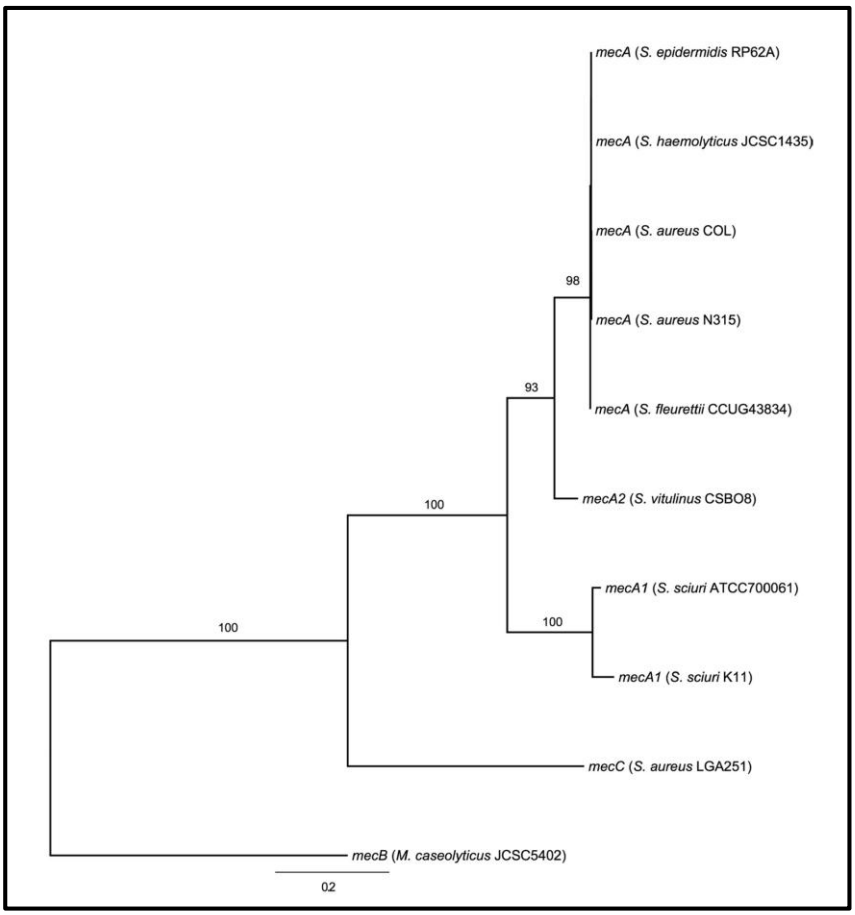


Figure 6. Maximum likelihood phylogenetic tree of *mec* homologues. The numbers at the tree branches are percentage bootstrap values indicating the confidence levels. The bar length indicates the number of substitutions per site (bar, 20 per 100 sites)(169).

2.3.1. *Staphylococcus sciuri mecA1*

The first *mec* homologue identified was *mecA1*, in *Staphylococcus sciuri*. This species is the most widespread staphylococcal species in nature; it has been isolated from the skin of virtually every mammal, but rarely in humans (170). Phylogenetic analysis of the 16S rRNA and other genes has shown that *S. sciuri* belongs to the most primitive group of staphylococcal species, the so-called *sciuri* group (8) and DNA-DNA hybridization profiles and biochemical properties showed that *S. sciuri* was composed of three subspecies: *S. sciuri sciuri*, *S. sciuri rodentius* and *S. sciuri carnaticus* (171).

In a study where a large representative collection of *S. sciuri* isolates were screened for the presence of *mecA* by dot-blot hybridization, it was found that *mecA1* was ubiquitous in this species (172). Moreover, a recent study, where the sequencing of the vicinity of *mecA1* was performed for one isolate of each *S. sciuri* subspecies, showed that *mecA1* in the *S. sciuri*, is flanked upstream by the *mva* locus (mevalonate metabolism), and downstream, by the *xyl* operon (only in *S. sciuri carnaticus*) (173). Due to its ubiquity and given the fact that *mecA1* was flanked in the chromosome by housekeeping genes that participate in metabolism, *mecA1* was suggested also to be a housekeeping gene (173).

The *mecA1* gene shares 80% nucleotide sequence identity with MRSA *mecA* (174) and encodes a penicillin-binding protein, that was called PBP4 (82% aminoacid sequence identity with PBP2a) (175). The identity of PBP4 domains with those of S. aureus PBP2a was different: while the transpeptidase domain was highly conserved (96% nucleotide sequence identity; 91% aminoacid sequence identity), the non-binding domain was more diverse (80% nucleotide sequence identity; 68% aminoacid sequence identity)(174).

Like *mecA*-encoded PBP2a, PBP4 was found to have low affinity to β-lactams (175). Despite the fact that the cell wall composition of S. *sciuri* and *S. aureus* is different (175), when introduced into a methicillin-susceptiple *S. aureus* genetic background which is known to be able to express resistance (COL-S), PBP4 was able to produce a cell wall similar to the one of *S. aureus*. Moreover, it was able to fully replace the physiological functions of the native PBP2a in a *S. aureus* in the presence of

antibiotic (176) and to confer β-lactam resistance to the recipient strain (176). Altogether, these findings strongly suggest that *mecA1* is the evolutionary precursor of *mecA*.

In spite of the low affinity of PBP4 for β-lactams, the great majority of natural *S. sciuri* isolates are susceptible to β-lactams (172); however, methicillin-resistant *S. sciuri* strains have been described. Molecular characterization of resistant strains has revealed that the resistant phenotype can be achieved with one of two different mechanisms: the strains can either carry a copy of *mecA* in their genome, besides *mecA1* (172) or have alterations in the promoter of *mecA1* that lead to overproduction of the *mecA1*-encoded PBP4 (177). These alterations are associated to the insertion of IS*256* or accumulation of SNPs in the *mecA1* promoter region that can lead to a higher rate of transcription (177). However, how frequently this type of events occur in the natural *S. sciuri* population, is not known.

2.3.2. *Staphylococcus vitulinus mecA2*

A second *mec* homologue (*mecA2*) was identified in another species belonging to the *sciuri* phylogenetic group that is called *Staphylococcus vitulinu*s, which like *S. sciuri* colonizes mainly animals, like cattle and horses (8, 170). The *mecA2*, has 90% nucleotide/aminoacid sequence identity with MRSA *mecA*, and was first identified in *S. vitulinus* strains recovered from horses in Switzerland (178). The sequencing of *mecA2* vicinity in one *S. vitulinus* strain, showed that like *mecA1*, *mecA2* was flanked in the chromosome by the *mva* and *xyl* operons (173). In addition, similarly to *mecA1* in its native host, *mecA2*-carrying *S. vitulinus* were found to be susceptible to β-lactams (173, 178). As no further studies regarding *mecA2* are available, the role of this *mec* homologue in the evolutionary history of *mecA* and its distribution in *S. vitulinus* population are still unknown.

2.3.3. *Staphylococcus fleurettii mecA*

A third *mec* homologue has been identified in another species of the phylogenetic *sciuri* group, (173), *Staphylococcus fleurettii*, which is mostly recovered from horses, pigs and cows (170, 179). Contrarily to *S. sciuri* and *S. vitulinus*, *S. fleurettii* was found to carry a *mecA* form that is almost identical (99-100% nucleotide identity) to that of MRSA *mecA*. Moreover, like in *S. aureus* SCC*mec*, in this species *mecA* was found to be together with its regulators, *mecRI* and *mecI* and the hypervariable (HVR) region was located upstream *mec* complex (173). However, this *mec* complex is not ubiquitous in *S. fleurettii*, since strains with deletion of this locus have been identified (173).

Sequencing of the vicinity of *mecA* in *S. fleurettii* has revealed that, as in the case of *mecA2* in *S. vitulinus* and *mecA1* loci in *S. sciuri carnaticus* strains, the *mva* and *xyl* operons were located in the vicinity of the *mec* gene (173). The topology of the phylogenetic tree regarding the sequence analysis of the *mvaS* gene, part of the *mva* operon, located upstream the *mec* homologues in *S. sciuri*, *S. vitulinus* and *S. fleurettii*, was the same as the one obtained for the nucleotide sequence of the *mec* homologues (173). This finding suggests that *mvaS* evolved with the *mec* homologue, indicating that *mecA* in *S. fleurettii* was inherited by vertical transmission and not recently acquired.

As opposed to what was found for *mecA1* in *S. sciuri* and *mecA2* in *S. vitulinus*, *mecA*-carrying *S. fleurettii* are β-lactam resistant (173). Moreover, since it carries other elements found in the same relative position as the ones found in SCC*mec* (*mecRI*, *mecI* and the HVR region), it was suggested that *S. fleurettii* was the last donor of *mec* complex A to an assembled SCC in another species, which probably gave rise to SCC*mec* (173). However, evidence to support this hypothesis was never produced.

2.3.4. *Staphylococcus lentus*

Another species of *sciuri* group include *Staphylococcus lentus* which has been recovered from different domestic animals, such as sheeps, horses and poultry (170). The few reports available indicate that *S. lentus* methicillin resistant strains also carry *mec* complex A (180, 181). In another study, a PCR screening failed to identify *mecA* or *mec* homologues among a collection of S*. lentus*; in addition, the phylogenetic analysis of the *mvaS* gene indicated that it was unrelated with *mvaS* gene from the other members of the *sciuri* group (173). Therefore, the authors suggested that the archaic mec gene might have been carried by the ancestral of *S. sciuri*, *S. fleurettii* and *S. vitulinus*, but not *S. lentus* (173).

2.3.5. Other *mec* homologues (*mecB*, *mecC*)

In recent years, *mec* homologues with lower nucleotide identity with MRSA *mecA* were identified, in *Macrococcus caseolyticus* and in staphylococcal species most frequently isolated from non-human mammals. *Macrococccus caseolyticus* is a species of the genus *Macrococcus*, the genus more phylogenetically related with the *Staphylococcus* genus (182). It is frequently isolated from raw meat and milk and, unlike *Staphylococcus*, was never isolated from human samples (183).

The *mecA* homologue identified in *M. caseolyticus* (183) had only 62% nucleotide sequence identity and 51% aminoacid sequence identity with MRSA *mecA* and therefore was called *mecB* (169). It was found along with homologues of its regulators *mecRI* and *mecI* (53% aminoacid sequence identity and 64% aminoacid sequence identity, respectively, with the ones carried by the prototype MRSA strain N315), as well as the penicillinase-encoded *blaZ* gene (183). *M. caseolyticus* strains *mecB*-carrying were resistant to β-lactams; introduction of *mecB* in a susceptible *S. aureus* genetic background was able to provide a resistance phenotype, confirming that *mecB* should encode a PBP with low-affinity to β-lactams (183). Noteworthy, in *M. caseolyticus*, *mecB* and its regulatory locus have been found in a plasmid (183), in a transposon integrated at the *orfX* region (120) and in SCC*mec* (146); in SCC*mecB*,

homologues of *ccrA* and *ccrB* genes, sharing less than 52% of nucleotide sequence identity with MRSA *ccrAB*, were also identified (146). These *ccrAB* homologues were able to properly excise SCC*mecB* from the chromosome (146). The findings suggest that in *M. caseolyticus*, *mecB* is exclusively part of the mobilome and not of the core chromosome.

A third *mec* homologue, called *mecC*, has been recently identified, which shared 69% nucleotide sequence identity and 63% aminoacid sequence identity with MRSA *mecA* (169). The *mecC* was firstly identified in *S. aureus* strains collected from dairy cattle (113, 114), but since then it has been identified among several genetic backgrounds of *S. aureus* collected in a wide range of domestic and wild animals (184), as well as humans (185, 186). A retrospective study identified *mecC* in *S. aureus* isolates collected in Denmark as early as 1975 (113). Nonetheless, its prevalence nowadays is still low, comparing with the frequency of *mecA*-carrying *S. aureus* isolates (184).

Similarly to *mecB*, homologues of the regulators *mecRI* and *mecI* (with 45% of aminoacid sequence identity and 60% of aminoacid sequence identity, respectively, with the ones carried by the prototype MRSA strain N315), as well as *blaZ*, were found in close vicinity to *mecC* (113). These elements were located in the *orfX* region and inside a SCC*mec* element (Figure 5). This element was named SCC*mec* XI and *mecC* and its regulators were considered to be a new *mec* complex, called *mec* complex E (113). Like *mecA1* and *mecB*, *mecC* conferred low-level resistance in its native host, but when introduced in a *S. aureus* genetic background was able to provide high level resistance and was able to participate in the cell wall synthesis, as PBP2a (187).

Besides being found in *S. aureus*, *mecC* has been also identified in isolates of the *sciuri* phylogenetic clade, such as *S. sciuri carnaticus* (133) and the recently described species *Staphylococcus stepanovicii* (188), and also in the poultry associated species *Staphylococcus xylosus* (132).

3. The missing links in SCC*mec* assembly and its impact in *Staphylococcus* evolution

To date, different *mec* homologues have been described, particularly in strains originated in animals, such as *mecA1* in *S. sciuri* (172), *mecA2* in *S. vitulinus* (178), *mecA* in *S. fleurettii* (173), *mecB* in *M. caseolyticus* (183) and *mecC* in livestock-associated *S. aureus* (113, 114). Several evidences suggest that the *S. sciuri mecA1* corresponds to an ancestral form of *mecA* (172, 175-177). However, the evolutionary history of *mecA* has never been addressed in a comprehensive way and evolutionary links between these *mecA* homologues are still poorly understood. Moreover, the steps that led to the assembly of the mobile genetic element that received *mecA* are still elusive.

On the other hand, the impact of SCC*mec* in the evolution of staphylococcal species, in particular CoNS species such as *S. epidermidis*, had not yet been addressed. The earliest MRSE strain described carried SCC*mec* IV and was isolated in 1973, in Canada (110). Despite being reported in Denmark in the early 60s (70), early MRSE have never been characterized. Therefore, it is still unclear when SCC*mec* was acquired by *S. epidermidis*. Furthermore, it has been described that the main hospital-associated *S. epidermidis* clonal lineage, CC2, has acquired SCC*mec* several times (17); nevertheless, the role of nosocomial *S. epidermidis* in SCC*mec* evolution and conversely, the role of SCC*mec* in *S. epidermidis* adaptation to the hospital environment has never been addressed.

In this thesis we aim to understand the role of each species of the *sciuri* phylogenetic group in the origin and evolution of SCC*mec* and its impact in the evolution and adaptation of staphylococcal species, in particular *S. epidermidis*.

PART I

Role of the most primitive staphylococcal species in the origin and assembly of SCC*mec*

High frequency and diversity of cassette chromosome recombinases (*ccr*) in methicillin-susceptible *Staphylococcus sciuri*

Joana Rolo, Hermínia de Lencastre, and Maria Miragaia

Journal of Antimicrobial Chemotherapy. 2014 Jun;69(6):1461-9.

All the experimental work was performed by J. Rolo.

ABSTRACT

Objectives: Previous studies produced evidence that *mecA*, the determinant of β-lactam resistance in methicillin resistant *Staphylococcus aureus* (MRSA), may have originated in the most primitive and widespread animal commensal species - *Staphylococcus sciuri*. But how the *mecA* homologue (*mecA1/pbpD*) was captured from *S. sciuri* into the staphylococcal cassette chromosome *mec* (SCC*mec*) has remained unclear.

Methods: To understand the role of *S. sciuri* in the assembly of SCC*mec*, we screened 118 methicillin-susceptible (MS) *S. sciuri* isolates for SCC*mec* central elements - *ccr* and *mec* complex (*ccrAB*, *ccrC*, *mecA, mecI, mecRI*), by dot-blot. In addition, isolates were typed by PFGE and the chromosomal proximity of SCC*mec* elements was determined by Southern hybridization. *ccr* typing was performed by nucleotide sequencing.

Results: *ccrAB* were identified in 35% of the isolates (n=41) represented by 24 PFGE types, but *ccrC* was not found. None of the isolates carried *mecA* or its regulators, but all isolates carried *mecA1/pbpD*. In the majority of isolates, *ccr* and *mecA1* were located near *orfX*, the SCC*mec* integration site. Moreover in 31% (13) of the *ccrAB*-carrying strains, *ccrAB, mecA1* and *orfX* co-localized in the chromosome. The nucleotide sequence of *ccrA*/*ccrB* was highly diverse, including *ccr* genes closely related (80 -97%) to those found in MRSA.

Conclusions: Our results suggest that *S. sciuri* was a natural recipient and a rich reservoir of *ccr* for the assembly of SCC*mec*. The chromosomal location of *mecA1*, near *orfX*, the recognition site of *ccr*, was probably crucial for its mobilization out of *S. sciuri* species into SCC*mec*.

INTRODUCTION

Resistance to methicillin and to the entire class of β-lactam antibiotics is associated with the acquisition of the exogenous *mecA* gene, which encodes an "extra" penicillin-binding protein (PBP2a) with low affinity for β-lactams. All *S. sciuri* strains – including fully susceptible isolates – were shown to carry a *mecA* homologue (*mecA1/pbpD*), coding for penicillin-binding protein 4 (PBP4) of *S. sciuri* (172, 175). When transduced into *S. aureus, mecA1*/*pbpD* was able to reconstruct the methicillin resistance phenotype, conferring high level, homogeneous and broad spectrum β-lactam resistance; the activity of *mecA1*/*pbpD* was controlled by *mecA* regulators *(mecI, mecRI*) and it was able to participate in cell wall biosynthesis, producing a peptidoglycan typical of *S. aureus* – providing additional evidence that the *S. sciuri mecA1*/*pbpD* is the evolutionary precursor of *mecA* (176, 189).

The methicillin resistance determinant is carried by the mobile element staphylococcal cassette chromosome *mec* (SCC*mec*), which promoted its spread on a global scale (109). SCC*mec* is composed of two central elements: the *mec* complex that carries either intact or truncated forms of the regulators (*mecI*/*mecRI*), and the *ccr* complex, which encodes recombinases responsible for the mobility of the cassette. The recombinases encoded by the *ccr* genes catalyze the precise insertion/excision of the cassette at the 5' end of *orfX* gene. (109) The *orfX* located near the *oriC* genes encodes for a SPOUT methyltransferase (190). So far as many as eleven different types of SCC*mec* have been described in *S. aureus* (111-114) that vary in the combination of class of *mec* complex and *ccr* allotypes as well as in the composition of the junction regions (111). However, the evolutionary steps in the assembly of this complex structure are not known.

Data gathered so far indicate that species of the *S. sciuri* group (a primordial phylogenetic clade), like *S. sciuri*, *Staphylococcus fleurettii* and *Staphylococcus vitulinus* were possibly involved in this process. Several different structural elements of SCC*mec* were found in these species, independently of SCC*mec*. This the case for *pls*, *ccrAB* (191), Tn*554* (180) and part of the hyper-variable region (HVR) (192), in *S. sciuri*; and *mecA* regulators (*mecI* and *mecR1*) in *S. fleurettii* (173). Additionally, distant

homologues of SCC*mec* structural elements were also found outside the genus *Staphylococcus*, in *Macrococcus* and *Enterococcus* (*ccrAB*) (120, 145). However it is still not clear how all these findings can be put together to reconstruct the evolutionary history of the primordial SCC*mec*.

In this study we further explored the contribution of *S. sciuri* to the evolution of SCC*mec*. We demonstrate that *ccr* are frequent and highly diverse in methicillin susceptible *S. sciuri*. Moreover, we found that the original location of *mecA1*/*pbpD* is in the vicinity of the integration site of *ccr* genes in *S. sciuri* chromosome, which is within the *orfX* region.

METHODS

Bacterial collection and isolation: a representative collection of 118 *mecA1*-positive, methicillin susceptible *S. sciuri* isolates, that were negative for the *S. aureus mecA* gene, were selected in order to assemble a collection the most diverse as possible in terms of host range (28 different animal and humans), isolation period (1972-2010), demographic and epidemiological data (Supplementary Table S1). Identification at the species level was performed by 16S RNA ribotyping (171) and API-Staph (Biomerieux, France).

DNA preparation: DNA was extracted with the isothiocyanate/guanidine method, as previously described (124). Probes for *ccrAB2*, *ccrC, mecA1*/*pbpD, mecI* and *mecRI* (membrane-spanning domain) were amplified by PCR as described before (192) (Supplementary Table S2) and purified using DNA clean & concentrator kit (ZYMO Research, Irvine, USA), according with the instructions of the manufacturer. For the preparation of *orfX* probe, the PCR reaction conditions were the following: 1x Gotaq buffer (Promega, Wisconsin, USA), 200 µM dNTPS (Bioron, Ludigshafen, Germany), 15 Mm $MgCl_2$ (Promega, Wisconsin, USA), 0.4 µM each primer, 1.25 U of Gotaq enzyme (Promega, Wisconsin, USA) and 1 µg DNA; initial denaturation at 94°C for 1 minute, followed by 30 cycles (94°C 1 min, 50°C 1 min, 72°C 1 min) and final extension at 72°C for 10 minutes. The DNA was purified as specified above. The following control strains were used as a source for the preparation of probes: *S. aureus* N315 (*ccrAB2, mecI, mecRI*) (193), *S. aureus* WIS (*ccrC*) (127) and *S. sciuri* K11 (*mecA1*/*pbpD, orfX*) (172).

Assessment of methicillin susceptibility: all isolates were screened for phenotypic resistance to oxacillin with Etest (Oxoid, Cambridge, United Kingdom), following the Clinical and Laboratory Standards Institute (CLSI) criteria (194). In addition, all isolates were screened for the presence of *mecA* and *mecA1* by PCR with specific primers (Supplementary Table S2). The PCR reaction conditions were the following: 1x Gotaq buffer (Promega, Wisconsin, USA), 200 µM dNTPS (Bioron, Ludigshafen, Germany), 15 Mm $MgCl_2$ (Promega, Wisconsin, USA), 0.4 µM each primer, 1.25 U of Gotaq enzyme (Promega, Wisconsin, USA) and 1 µg DNA; initial denaturation at 94°C for 1 minute,

followed by 30 cycles (94°C 1 min, 50°C 1 min, 72°C 1 min) and final extension at 72°C for 10 minutes. Isolates were considered to be susceptible to oxacillin if the MIC to oxacillin was < 2 μg/ml and in addition did not carry *mecA*, as proposed by CLSI criteria (194).

Screening for SCC*mec* structural elements: dot blot hybridization was performed as described before (172). Briefly, total DNA was fixed in nitrocellulose membranes and sequentially hybridized with specific probes for *ccrAB2*, *ccrC, mecI* and *mecRI* using 1:1 and 1:10 dilutions. Labeling and detection of probes was performed with the ECL Labeling and Detection system as suggested by the manufacturer (GE Healthcare Lifesciences, Little Chalfont, UK). The following controls were used: *S. aureus* COL (*ccrAB1*),(52) *S. aureus* N315 (*ccrAB2*),(193) *S. aureus* ANS46 (*ccrAB3*),(142) *S. aureus* HDE288 (*ccrAB4*),(128) *S. aureus* WIS (*ccrC*),(127) *S. sciuri* K3 (*mecA, mecA1*/*pbpD, ccrAB3*)(172) and *S. sciuri* K11 (*mecA1*/*pbpD*)(172).

Assessment of genetic backgrounds: the isolates that were positive for the presence of SCC*mec* structural elements by dot-blot hybridization were further characterized for their genetic background by analysis of SmaI macrorestriction patterns of chromosomal DNA obtained after pulsed-field gel electrophoresis. Agarose plugs with embedded DNA (DNA disks) were prepared as previously described (195) and restricted with SmaI (20U/disk) following the instructions of the manufacturer (New England Biolabs, Beverley, USA). Analysis of the macrorestriction patterns was achieved with Bionumerics software (Applied Maths, Saint-Martens-Latem, Belgium). Dendrograms were generated with the Dice algorithm (1.1% optimization and 1.3 tolerance), using the Unweighted Pair Group Method with Arithmetic Mean (UPGMA) method and PFGE types were defined with 75% similarity between clusters. Each different PFGE type was designated by a different capital letter. ApaI restriction and running conditions were exactly the same as those used with SmaI enzyme.

***ccr* typing:** for strains carrying *ccr* gene homologs by dot-blot, the presence of *ccrA*, *ccrB* and *ccrC* was confirmed by PCR as previously described (Supplementary Table S2) (125, 127). In addition, *ccrA* and *ccrB* were typed by sequencing an internal fragment

obtained by PCR, as described before (142). Resulting sequences were aligned and compared with the *ccrA* and *ccrB* sequences available on www.ncbi.nlm.nih.gov/ and www.ccrbtyping.net. Alignment of the sequences was performed with ClustalW algorithm and phylogenetic analysis was performed with Neighbour-Joining method, using MEGA5 software (196).

The *ccrA* and *ccrB* were considered to belong to the same *ccr* allotype when the sequences obtained shared more than 85% of nucleotide identity (111). The *ccr* genes were considered to be non-typeable when no amplification product by PCR or multiple superimposed peaks in sequencing reads were repeatedly obtained. The following controls were used for *ccr* screening by PCR: *S. aureus* COL (*ccrAB1*) (52) and WIS (*ccrC*)(127).

Identification of chromosomal location of SCC structural elements: the chromosomal location of SCC*mec* structural elements was analyzed by Southern blotting of the rare cutting enzymes SmaI and ApaI (50U/DNA disk) restriction fragments with probes for *ccrAB2, orfX* and *mecA1*/*pbpD*, using ECL Labeling and Detection system as suggested by the manufacturer (GE Healthcare Lifesciences, Little Chalfont, UK). The following controls were used for Southern blotting: *S. aureus* COL (*ccrAB1*)(52), *S. aureus* HU25 (*ccrAB3, ccrC*)(142), *Staphylococcus epidermidis* RP62A (*ccrAB2*)(52), *S. epidermidis* ATCC12228 (*ccrAB2*, *ccrAB4*) (99), *S. sciuri* K1 (*mecA1*/*pbpD*, *orfX*)(172), *S. sciuri* K3 (*mecA, mecA1*/*pbpD, ccrAB3, orfX*)(172) and *S. sciuri* K11 (*mecA1*/*pbpD*, *orfX*)(172).

RESULTS

The *ccr* genes are frequent among methicillin-susceptible *S. sciuri* strains

In order to understand how common the central elements of SCC*mec* were in *S. sciuri*, we analyzed 118 isolates that did not carry *S. aureus mecA*, but carried the native *mec* homologue *mecA1*. Although all isolates had a MIC equal or superior to the limit of resistance defined for CoNS (between 0.5-2 µg/mL), since they did not carry the *mecA*, they were all considered to be susceptible, as suggested by the CLSI guidelines. All the isolates were screened for the presence of genes within the *mec* complex (*mecI*, *mecRI*) and the *ccr* complex (*ccrAB* and *ccrC*) by dot-blot hybridization and PCR. Although we did not find any strain carrying *mecI* or *mecRI*, we found that 35% of *S. sciuri* strains (n=41) analyzed carried *ccr* genes (Figure 1A).

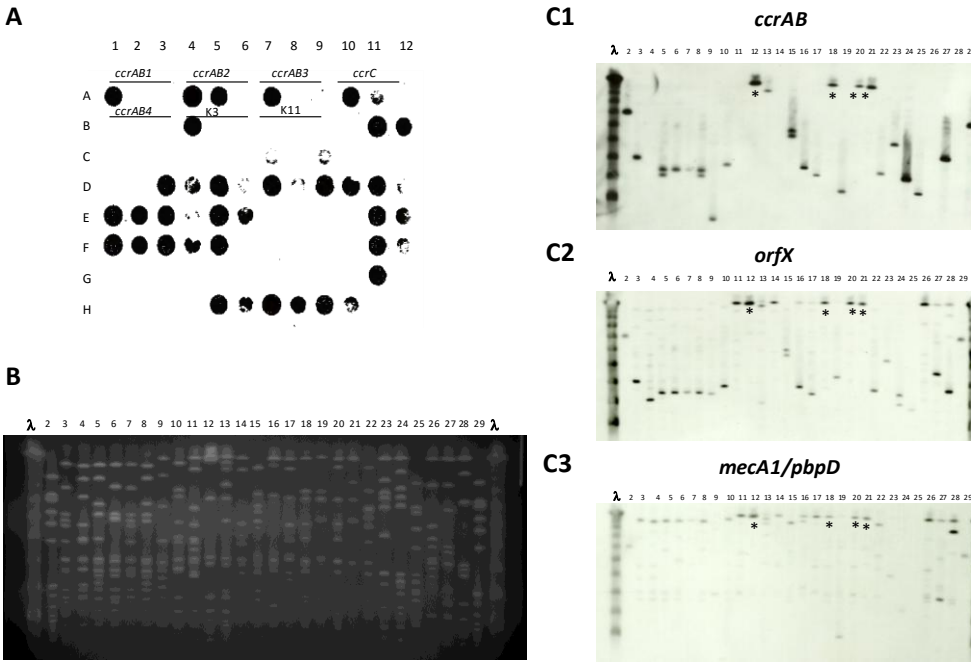


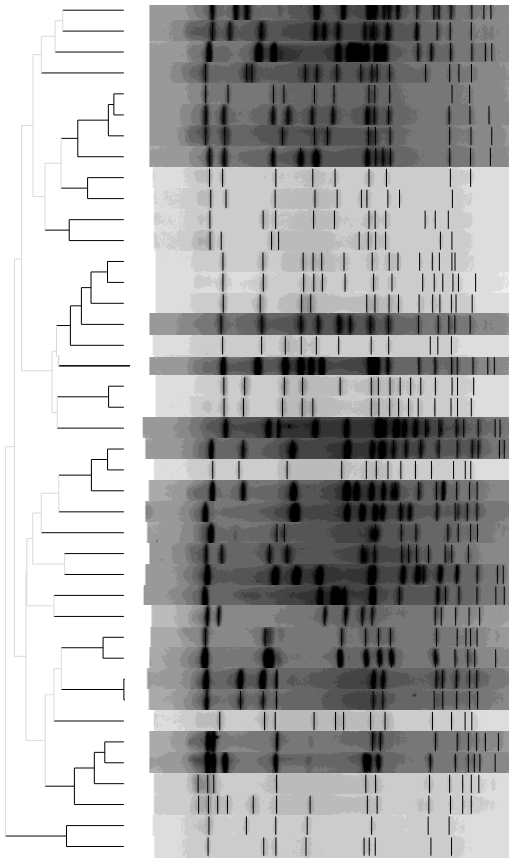
Figure 1. Detection of *ccr* and its location on the chromosome of *S. sciuri*. **A.** Dot blot hybridization of total DNA with a *ccrAB/ccrC* probe. A1-A12, B1-B9 – positive control strains (*S. aureus* COL, *ccrAB1*; *S. aureus* N315,*ccrAB2*; *S. aureus* ANS46, *ccrAB3*; *S. aureus* WIS, *ccrC*; *S. aureus* HDE288, *ccrAB4*; *S. sciuri* K3, *ccrAB3*; *S. sciuri* K11, *ccrAB* negative control), 1:10, 1:100 total DNA dilutions in rows; remaining dots – *S. sciuri* isolates, 1:10 total DNA dilutions. **B.** SmaI macrorestriction patterns of *S. sciuri* isolates separated by PFGE. Control strains are also shown: 1, 30 – Lambda ladder; 2, 29 -RP62A; 22-COL; 23-N315; 24-HU25; 25-ATCC12228; 26-K1; 27-K3; 28-K11. **C.** Southern hybridization of the SmaI PFGE DNA restriction fragments obtained in B with specific probes: *ccrAB2* **(C1)**, *orfX* **(C2)** and *pbpD* **(C3)**. Asterisks indicate the cases in which SmaI hybridization bands with *pbpD*, *orfX* and *ccrAB2* probes were the same.

Moreover, we observed that the *ccr*-carrying strains belonged to 24 different PFGE types; were isolated from different hosts (28 wild and domestic animals and humans), during different time periods (1972 to 2010) and have originated in distant geographic areas (Czech Republic, Sweden, Portugal, USA and the former Yugoslavia) (Supplementary Table S3). In addition, the same *ccr* type was found associated to two or more PFGE types (Supplementary Table S3 and Figure 2).

The *ccrAB* allotypes of *S. sciuri* are similar to those carried by MRSA strains

To understand the relatedness of the *S. sciuri ccr* genes to *ccr* genes carried by contemporary methicillin-resistant *S. aureus* (MRSA) strains, we studied the *ccr* genes identified in the 41 *ccr*-positive strains using PCR and sequencing.

All *S. sciuri ccr* genes belonged to allotypes A and B; no strains carrying *ccrC* were found. *S. sciuri* strains carried three main *ccrA* types that accounted for more than half of the strains (52%) analyzed: *ccrA5* (11 isolates, 26%), *ccrA1* (six isolates, 14%) and a new *ccr* type related with *ccrA2* (80 % nucleotide sequence identity, five isolates, 12%) (Supplementary Table S4). The alleles *ccrA1* and *ccrA2* are usually associated to SCC*mec* types I and II/IV. Among the remaining strains we identified several new *ccrA* allotypes that shared between 61-78% sequence identity with contemporary *ccrA1-A5* types (n=19); and strains (n=4) carrying non-typeable *ccrA* (Supplementary Table S3 and Figure 3A).



Strain	Node index (%)	SmaI-PFGE type	*ccrAB* type	Host	Geographic origin	Year of isolation
K162	73	Y	A5B13	Bottlenose dolphin	USA	1992
K20	73	N	A5B13	Bottlenose dolphin	USA	1992
K148	70	P	NT	Bottlenose dolphin	USA	1992
K21	64.5	L	A5B5	Pilot whale	USA	1993
K164	95.6	M	A5B13	Bottlenose dolphin	USA	1992
K165	95.6	M	New *ccrA* type 1/B13	Hairy-footed hamster	USA	1992
KLO56	93.3	M	New *ccrA* type 1/B13	Opossum	USA	1972
K156	79.9	M	NT	Beef tongue	USA	1992
HA4	84	O	A5B9	Snake	Czech Republic	2002
K113	84	O	New *ccrA* type 3/B9	Norway rat	USA	1992
HSM805	76	B	A5B13	Human	Portugal	2010
K52	76	B	A5B13	European red squirrel	USA	1992
SS24	92.8	G	A1B1	Human	Portugal	1997
SS3	92.8	G	NT	Human	Portugal	1996
SS23	85.7	G	A1B13	Human	Portugal	1997
K69	81.5	G	NT	Human	USA	1986
HA22	76.8	G	New *ccrA* type 3/B9	Snake	Czech Republic	2002
K141	70.6	H	New *ccrA* type 3/B9	California mouse	USA	1992
SS18	93	I	NT	Human	Portugal	1996
SS27	93	I	NT	Human	Portugal	1998
01/02	70.9	J	NT	Barn environment	Czech Republic	2002
K29	93	Q	New *ccrA* type 3/B13	Norway rat	USA	1992
SS34	93	Q	New *ccrA* type 1/B5/7	Human	Portugal	1996
CCM4743	85.8	Q	New *ccrA* type 1/B5/7	Human	Sweden	2002
K27	71.7	R	A1B13	Norway rat	USA	1992
CNCTC M62/89	64	V	New *ccrA* type 1/B5/7	Cattle	Czech Republic	2002
K171	74	T	New *ccrA* type 2/B5/7	Opossum	USA	1972
K83	74	S	NT	Human	Czech Republic	1992
K125	69.6	X	A5B13	Cotton rat	USA	1992
K30	69.6	U	A5B13	Jersey cattle heifer	USA	1992
11/01	90.9	C	New *ccrA* type 1/B13	Human	Czech Republic	2002
CCUG38359	90.9	C	New *ccrA* type 1/B13	Human	Sweden	2002
12/01	72.7	D	New *ccrA* type 1/B5/7	Human	Czech Republic	2002
2485/IV/01	72.7	D	NT	Human	Czech Republic	2002
CCUG37410	69.7	K	New *ccrA* type 3/B9	Human	Sweden	2002
17/01	91.6	E	New *ccrA* type 1/B13	Human	Czech Republic	2002
CNCTC M32/86	91.6	E	A5B13	Pig skin	Czech Republic	2002
JUG4	87.1	E	New *ccrA* type 2/B13	Dog	Yugoslavia	2002
JUG9	78.4	F	New *ccrA* type 2/B13	Dog	Yugoslavia	2002
KRU5	75	A	A5B13	Turkey	Czech Republic	2002
KU2	75	A	New *ccrA* type 2/B13	Chicken	Czech Republic	2002

Figure 2. Phylogenetic relationship of the ccr-positive S. *sciuri* strains based on the analysis of SmaI-restricted chromosomes separated by pulsed-field gel electrophoresis. The dendrogram was performed with the Dice algorithm (1.1% optimization and 1.3 tolerance), using the Unweighted Pair Group Method with Arithmetic Mean (UPGMA) method; PFGE types were defined with 75% similarity between clusters. In addition, PFGE type, % similarity in each node of the dendrogram, *ccrAB* type and epidemiological data are also shown for each *ccr*-positive isolate. *ccrAB* types were considered new when either *ccrA* or *ccrB* alleles were new.

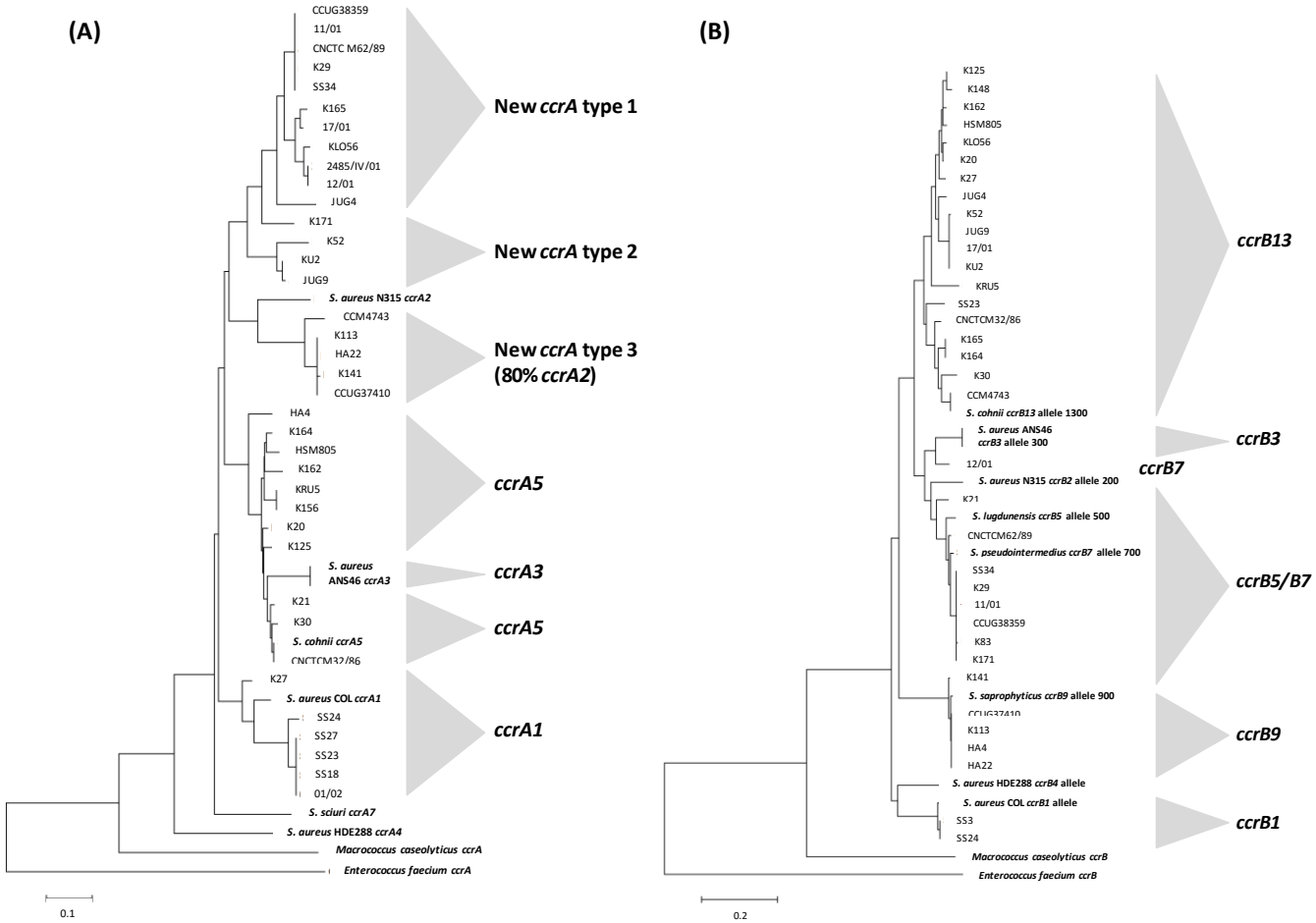


Figure 3. Phylogenetic trees constructed from *ccrA* (**A**) or *ccrB* (**B**) *n*ucleotide sequences. Alleles carried by *S. sciuri* isolates and prototype strains (in bold) are shown. Alignments of the sequences were performed with ClustalW algorithm and dendrograms were predicted based on nucleotide percent identity with the Neighbour-joining method, using MEGA5 software [196]. Allotypes were defined considering a cut-off of 85% identity (111).

Regarding *ccrB*, the great majority of strains carried *ccrB13* (n=19, 45%) and *ccrB5/B7* (n=9, 21%). The remaining isolates carried either *ccrB9* (n=5) or *ccrB1* (n=2) (Supplementary Tables S3, S4 and Figure 3B). The *ccrB13* was never found associated to any SCC*mec* type, but *ccrB1* was found associated to SCC*mec* type I and *ccrB5* and *B7* had 85-86.6% nucleotide sequence identity with *S. aureus ccrB3*, which was previously found associated to SCC*mec* type III.

Only one of the five combinations of *ccrA* and *ccrB* types found in *S. sciuri* isolates analyzed in this study was previously seen in MRSA. This corresponds to *ccrA1B1* complex found in the SCC*mec* type I. The remaining *ccr* complexes were new,

and included new combinations of previously described *ccrA* and *ccrB* types (*A1B13, A5B13, A5B5* and *A5B9*); the new *ccrA* types (n=19 isolates); and non-typeable *ccr* types (n=9 isolates). In addition, we observed that none of the *ccr* complexes found here were restricted to a unique PFGE type (Supplementary Table S3).

mecA1/*pbpD, ccrAB* and *orfX* co-localize in the *S. sciuri* chromosome

In order to find out if *ccrAB* of *S. sciuri* was also located in the *orfX* (like in other staphylococcal species), we sequentially hybridized the SmaI macrorestriction fragments of *S. sciuri* isolates with probes for *ccrAB2* and *orfX*. We found that for the great majority of isolates (34, 83%), *ccr* and *orfX* hybridized in the same SmaI fragment, ranging from 75 to 400 kb, suggesting that the *ccr* of *S. sciuri* also recognizes the *orfX* region as the site for integration and excision from the chromosome (Supplementary Table S3 and Figure 1B, 1C).

Although *mecA1*/*pbpD* was suggested to be the origin of the methicillin resistance determinant *mecA*, its location in the chromosome was never explored. To test this, we hybridized sequentially the SmaI macrorestriction PFGE fragments with probes for *mecA1*/*pbpD* and *orfX*. In addition, we sequentially hybridized also ApaI macrorestriction PFGE fragments, to discard the possibility of the existence of a SmaI cutting site between *mecA1*/*pbpD* and *orfX*. We found that in the majority of the isolates tested (76%), *mecA1*/*pbpD* and *orfX* co-localized in the same SmaI (n=14, SmaI fragments 200-400 kb) or ApaI (n=17, ApaI fragments 300-400 kb; data not shown) fragments (Supplementary Table S3 and Figure 1B and 1C). These results indicate that the location of *mecA1*/*pbpD* on the *S. sciuri* chromosome is in the vicinity of *orfX*.

To understand how close the *mecA* precursor (*mecA1*/*pbpD*) was to *ccr*, the proximity of these elements was assessed using Southern hybridization of SmaI restriction fragments using specific probes. We found that in 31% of the *ccrAB* carrying strains, *ccrAB*, *orfX* and *mecA1*/*pbpD* co-localized in the same SmaI fragment (200-400 kb) indicating that all three genes are in the vicinity of one another in the chromosome (Supplementary Table S3 and Figure 1B and 1C).

DISCUSSION

It was previously proposed that *S. sciuri mecA1*/*pbpD* was the evolutionary precursor of the methicillin-resistance determinant *mecA*, (172, 174-176) but the steps that lead to the introduction of this native gene into a mobile genetic element remained unknown. In this study we provided further evidences of the key role of *S. sciuri* in the assembly of the core structural elements of SCC*mec*.

The extremely high frequency of *ccr* in susceptible *S. sciuri*, represents the highest described so far in any methicillin-susceptible staphylococci (not carrying the *S. aureus mecA*) (197, 198). These recombinases were found widely distributed in the *S. sciuri* population analyzed, including several different PFGE types, originating in different hosts, and distant geographic regions as well as periods of isolation. These results demonstrate the wide distribution of *ccr* in the *S. sciuri* population and the existence of multiple independent acquisitions of these genes by *S. sciuri,* suggesting either a high rate of *ccr* acquisition or a high recombination occurring at the *ccr* chromosomal locus. Furthermore, such high frequency and wide distribution imply that this species might have been the most natural *ccr* recipient among staphylococci. The *S. sciuri ccrAB* genes could have been imported or be descendents of the distant *ccr* homologues found in *M. caseolyticus* and *Enterococcus* (120, 145).

Besides being widely distributed in *S. sciuri*, the *ccrAB* found in this species were highly diverse, including relatives of *ccrA* and *ccrB* allotypes described in MRSA, but also different *ccrA* and *ccrB* combinations. This high diversity in the *ccr* locus and the occurrence of several new combinations of *ccrA* and *ccrB* is in contrast to what was observed before for other methicillin-susceptible coagulase negative staphylococcal (CoNS) species, such as *S. epidermidis*, *Staphylococcus haemolyticus* and *Staphylococcus hominis,* which appear to carry only specific *ccr* allotypes and specific *ccrAB* combinations. (17, 197-199). This level of genetic diversity suggests that *ccrAB* may have been present in *S. sciuri* for a long time and this species might have been the source of all *ccr* allotypes presently carried by SCC*mec* in other staphylococci. The reasons for *S. sciuri* being so receptive to such a different array of *ccr* genes are not known, but might be related to its ubiquitous nature in the environment or to a non-

identified specificity of its epidemiology, associated to an enhanced ability to acquire and maintain mobile genetic elements or to an increased recombination rate.

In particular, the high level of identity found between *S. sciuri ccrA1* (87%) *and ccrB1* (97%) with those *ccr* genes found in SCC*mec* type I, suggests that *S. sciuri* could have been the source for the assembly of *ccrA1B1* carried by SCC*mec* type I of the historically early strains of MRSA (52, 200). Moreover, *ccrA1B1* was the unique *ccrAB* combination that was observed both in *S. sciuri* and MRSA. Similarly, *S. sciuri* may also have been the origin of *ccrA3B3*, since close homologues of these allotypes and combination of these allotypes (*ccrA5*, *ccrB5*, *ccrB7*) were also frequently found in this species. Interestingly, the *ccrA3B3* complex is part of SCC*mec* type III, which is the most frequently found SCC*mec* type in methicillin-resistant *S. sciuri* (201). However, since *ccr* from *S. sciuri* were not 100% identical to those found within SCC*mec*, we cannot exclude the existence of intermediate *ccr* structures between the two.

We found that the nucleotide sequence of *ccrA* and *ccrB* genes found in *S. sciuri* was similar to those of SCC*mec* carried by contemporary MRSA strains. In addition the location of these genes in the chromosome was near *orfX*. According to our data *orfX* is also the location of the *mecA1*/*pbpD* gene on the *S. sciuri* chromosome, in the majority of the isolates tested (76%). In the remaining 24% of the isolates the hybridization of *ccr* and *orfX* with different SmaI/ApaI fragments could be explained by the presence of a SmaI cutting site between *ccr* and o*rfX*. However we cannot exclude that *ccr* might be located in other region of the chromosome in these particular strains.

The location of *mecA1*/*pbpD*, near *orfX*, is unique. In species such as *S. aureus, S. haemolyticus*, *S. epidermidis* and *S. saprophyticus*, *pbp* genes are scattered throughout the genome (http://www.ncbi.nlm.nih.gov) and are far away from the *orfX* region, except for the *mecA* gene that encodes PBP2A. Moreover, we found a high variability in the *orfX* region as illustrated by the variation in the size of SmaI-*orfX* hybridizing fragments, like was previously observed for *S. epidermidis* (124)*,* suggesting the frequent insertion and excision of elements in this site. The recent finding in *S. sciuri* of a hybrid SCC*mec*-SCC structure carrying the recently described *mecC*, near *orfX,* evidences well the level of recombination occurring in the region (133). These data suggests that the original location of *mecA1*/*pbpD* near *orfX*, a hot spot for

variation in staphylococci, was probably important for its assembly into SCC*mec* in *S. sciuri*.

Altogether our results allow us to propose that *S. sciuri* may have been the species in which "the capture" of *mecA1*/*pbpD* by a SCC-like element might have occurred. Studies are being conducted, including the parallel analysis of *orfX* vicinity and spontaneous excision of putative SCC*mec* elements in *S. sciuri* and other species of the *S. sciuri* group (*S. fleurettii*, *S. vitulinus*) that we hope will help to identify the species in which the assembly of a primordial mobile SCC*mec* structure could have occurred.

In conclusion, we showed that the chromosomal location of *mecA1*/*pbpD* in *S. sciuri* was probably critical for its integration into the SCC*mec*. Moreover, we provided evidence that suggest that *S. sciuri* might have contributed not only with the *mecA* precursor but also with *ccr* genes for the assembly of SCC*mec*.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Alexander Tomasz for helpful discussions and for critical reading of the manuscript. We wish to acknowledge Drs. Wesley Kloos, Oto Melter, Ivo Sedlácek, Petr Jezek, Petr Petras, Srdjan Stepanović, José Melo-Cristino, Gabriel Olim and Alberto Pereira for providing the *S. sciuri* strains used in this study. In addition, we would like to acknowledge Isabel Couto for extensive characterization of part of the strains that were studied. Finally, we thank Keichii Hiramatsu and Teruyo Ito for the control strains N315 and WIS and the Network on Antimicrobial Resistance in *Staphylococcus aureus* (NARSA) for strain MW2.

FUNDING

This study was supported by projects Ref. PTDC/BIA-EVF/117507/2010 from Fundação para a Ciência e Tecnologia and European Society of Clinical Microbiology and Infectious Diseases (ESCMID) Research Grants 2010. Additionally, we also acknowledge support from FCT (Portugal) through grant # PEst-OE/EQB/LA0004/2011 (ITQB). J. Rolo was supported by fellowship SFRH /BD/ 72675 / 2010 from Fundação para a Tecnologia, Portugal.

TRANSPARENCY DECLARATION

None to declare.

SUPPLEMENTARY DATA

Supplementary Table S1: Epidemiological and molecular characteristics of the 118 *Staphylococcus sciuri* isolates studied. When applicable, bibliographic reference of the isolates is cited.

Strain	Animal/Human origin	Geographic origin	Date of isolation	Reference	Oxacillin MIC (µg/ml)	PCR *mecA*	Dot blot (*ccr*)	Dot blot (*mecI*)	Dot blot (*mecRI*)	SmaI-*ccr* (kb)	SmaI-*pbpD* (kb)	SmaI-*orfx* (kb)	ApaI-*pbpD* (kb)	ApaI-*orfX* (kb)	SmaI-PFGE type	*ccrAB* type
SS3	Human	Portugal	1996	(202)	1	-	+	-	-	150 kb	> 400 kb	150 kb	no digestion	no digestion	G	NT
SS5	Human	Portugal	1996	(202)	1	-	+	-	-	-	> 400 kb	100 kb	nd	nd	nd	nd
SS16	Human	Portugal	1996	(202)	0,75	-	-	-	-	-	2 bands; > 400, 150 kb	> 400 kb	nd	nd	nd	nd
SS18	Human	Portugal	1996	(202)	0,75	-	+	-	-	2 bands; 80, 100 kb	>400 kb	100 kb	2 bands; > 400 kb **	200 kb	I	NT
SS23	Human	Portugal	1997	(202)	1	-	+	-	-	100 kb	>400 kb	100 kb	> 400 kb	> 400 kb	G	A1B13
SS24	Human	Portugal	1997	(202)	0,75	-	+	-	-	100 kb	> 400 kb	100 kb	> 400 kb	> 400 kb	G	A1B1
SS27	Human	Portugal	1998	(202)	0,75	-	+	-	-	2 bands; 80, 100 kb	> 400 kb	100 kb	2 bands; > 400 kb **	200 kb	I	NT
SS34	Human	Portugal	1996	(202)	0,75	-	+	-	-	< 50 kb	150 kb	100 kb	> 400 kb	300 kb	Q	New
CCUG37410	Human	Sweden	2002	This study	0,75	-	+	-	-	125 kb	> 400 kb	125 kb	no digestion	no digestion	K	New
03/00	Human	Czech Republic	2002	This study	1	-	-	-	-	nd	nd	nd	nd	nd	nd	nd
Jug14	Human	Yugoslavia	2002	This study	1	-	+	-	-	-	> 400 kb	> 400 kb	nd	nd	nd	nd
Jug2	Dog	Yugoslavia	2002	This study	0,75	-	-	-	-	nd	nd	nd	nd	nd	nd	nd
Jug3	Dog	Yugoslavia	2002	This study	0,75	-	+	-	-	-	> 400 kb	> 400 kb	nd	nd	nd	nd
Jug4	Dog	Yugoslavia	2002	This study	0,38	-	+	-	-	> 400 kb	> 400 kb	> 400 kb	2 bands; >400, 300 kb	150 kb	E	New
Jug5	Dog	Yugoslavia	2002	This study	0,5	-	-	-	-	nd	nd	nd	nd	nd	nd	nd
Jug9	Dog	Yugoslavia	2002	This study	0,75	-	+	-	-	> 400 kb	2 bands; > 400, 400 kb	400 kb	2 bands; > 400 kb **	2 bands; > 400 kb **	F	New
Jug21	Dog	Yugoslavia	2002	This study	1	-	-	-	-	nd	nd	nd	nd	nd	nd	nd
Jug22	Dog	Yugoslavia	2002	This study	1	-	-	-	-	nd	nd	nd	nd	nd	nd	nd
Jug23	Dog	Yugoslavia	2002	This study	1	-	-	-	-	nd	nd	nd	nd	nd	nd	nd

Strain	Animal/Human origin	Geographic origin	Date of isolation	Reference	Oxacillin MIC (µg/ml)	PCR *mecA*	Dot blot (*ccr*)	Dot blot (*mecI*)	Dot blot (*mecRI*)	SmaI-*ccr* (kb)	SmaI-*pbpD* (kb)	SmaI-*orfx* (kb)	ApaI-*pbpD* (kb)	ApaI-*orfX* (kb)	SmaI-PFGE type	*ccrAB* type
K41	Patas monkey	USA	1976	(172)	0,75	-	-	-	-	nd	nd	nd	nd	nd	nd	nd
K45	Squirrel	Panama	1976	(172)	1	-	-	-	-	nd	nd	nd	nd	nd	nd	nd
K46	Howler monkey	Panama	1976	(172)	0,5	-	-	-	-	nd	nd	nd	nd	nd	nd	nd
K52	European red squirrel	USA	1992	(172)	0,5	-	+	-	-	150 kb	2 bands; > 400 kb **	150 kb	> 400 kb	> 400 kb	B	A5B13
K77	Neonatal ward	Mozambique	1992	(172)	0,5	-	-	-	-	nd	nd	nd	nd	nd	nd	nd
K81	Howler monkey	Panama	1976	(172)	1	-	-	-	-	nd	nd	nd	nd	nd	nd	nd
K113	Norway rat	USA	1992	(172)	1	-	+	-	-	125 kb	> 400 kb	125 kb	> 400 kb	> 400 kb	O	New
K122	Howler monkey	Panama	1976	(172)	1	-	-	-	-	nd	nd	nd	nd	nd	nd	nd
K138	Howler monkey	Panama	1976	(172)	0,75	-	-	-	-	nd	nd	nd	nd	nd	nd	nd
HSM805	Human	Portugal	2010	This study	0,75	-	+	-	-	> 400 kb	> 400 kb	> 400 kb	> 400 kb	> 400 kb	B	A5B13
Jug15	Human	Yugoslavia	2002	This study	1	-	-	-	-	nd	nd	nd	nd	nd	nd	nd
KU2	Chicken	Czech Republic	2002	(172)	0,75	-	+	-	-	100 kb	50 kb	150 kb	2 bands; > 400 kb **	2 bands; > 400 kb **	A	New
KU12	Chicken	Czech Republic	2002	(172)	1	-	-	-	-	nd	nd	nd	nd	nd	nd	nd
HA3	Snake	Czech Republic	2002	This study	1,5	-	-	-	-	nd	nd	nd	nd	nd	nd	nd
HA4	Snake	Czech Republic	2002	This study	1,5	-	+	-	-	> 400 kb	> 400 kb	> 400 kb	> 400 kb	> 400 kb	O	A5B9
HA11	Snake	Czech Republic	2002	This study	1	-	-	-	-	nd	nd	nd	nd	nd	nd	nd
KRU5	Turkey	Czech Republic	2002	(172)	0,5	-	+	-	-	> 400 kb	> 400 kb	> 400 kb	2 bands; > 400 kb **	> 400 kb	A	A5B13
KRU6	Turkey	Czech Republic	2002	(172)	0,5	-	-	-	-	nd	nd	125 kb	nd	nd	nd	nd
HA17	Snake	Czech Republic	2002	This study	1	-	-	-	-	nd	nd	nd	nd	nd	nd	nd
HA22	Snake	Czech Republic	2002	This study	1	-	+	-	-	125 kb	> 400 kb	125 kb	> 400 kb	> 400 kb	G	New
HA14	Snake	Czech Republic	2002	This study	1	-	-	-	-	nd	nd	nd	nd	nd	nd	nd
HAD12	Snake	Czech Republic	2002	This study	1	-	-	-	-	nd	nd	nd	nd	nd	nd	nd
HAD28	Snake	Czech Republic	2002	This study	1,5	-	-	-	-	nd	nd	nd	nd	nd	nd	nd
K30	Jersey cattle heifer	USA	1992	(172)	1,5	-	+	-	-	200 kb	> 400 kb	200 kb	> 400 kb	> 400 kb	U	A5B13
K31	Jersey cattle calf	USA	1992	(172)	0,5	-	-	-	-	-	> 400 kb	> 400 kb	nd	nd	nd	nd
K32	Jersey cattle calf	USA	1992	(172)	0,75	-	-	-	-	-	> 400 kb	> 400 kb	nd	nd	nd	nd
K33	Jersey cattle heifer	USA	1992	(172)	0,75	-	-	-	-	-	> 400 kb	> 400 kb	nd	nd	nd	nd
K61	Pilot whale	USA	1992	(172)	1,5	-	+	-	-	-	> 400 kb	175 kb	nd	nd	nd	nd
K116	Beef lips	USA	1992	(172)	0,75	-	+	-	-	-	> 400 kb	> 400 kb	nd	nd	nd	nd

Strain	Animal/Human origin	Geographic origin	Date of isolation	Reference	Oxacillin MIC (µg/ml)	PCR *mecA*	Dot blot (*ccr*)	Dot blot (*mecI*)	Dot blot (*mecRI*)	SmaI-*ccr* (kb)	SmaI-*pbpD* (kb)	SmaI-*orfx* (kb)	ApaI-*pbpD* (kb)	ApaI-*orfX* (kb)	SmaI-PFGE type	*ccrAB* type
K163	Holstein cow	USA	1992	(172)	0,75	-	+	-	-	-	> 400 kb	150 kb	nd	nd	nd	nd
K10	European red squirrel	USA	1992	(172)	0,75	-	+	-	-	-	150 kb	100 kb	nd	nd	nd	nd
CCM4743	Human	Sweden	2002	This study	1,5	-	+	-	-	175 kb	100 kb	175 kb	2 bands; > 400 kb **	> 400 kb	Q	New
2485/IV/01	Human	Czech Republic	2002	This study	1	-	+	-	-	300 kb	2 bands; > 400, 350 kb	300 kb	2 bands; > 400 kb **	> 400 kb	D	NT
12/01	Human	Czech Republic	2002	This study	0,75	-	+	-	-	300 kb	2 bands; > 400, 350 kb	300 kb	2 bands; > 400 kb **	> 400 kb	D	New
CNCTC M62/89	Cattle	Czech Republic	2002	This study	0,75	-	+	-	-	2 bands; 50, 75 kb	2 bands; > 400, 100 kb	75 kb	2 bands; > 400 kb **	> 400 kb	V	New
K27	Norway rat	USA	1992	(172)	0,75	-	+	-	-	250 kb	2 bands; 250, 150 kb	250 kb	200 kb	350 kb	R	A1B13
K29	Norway rat	USA	1992	(172)	0,75	-	+	-	-	< 50 kb	2 bands; 250, 150 kb	100 kb	200 kb	300 kb	Q	New
K83	Human	Czech Republic	1992	(172)	0,75	-	+	-	-	250 kb	250 kb	100 kb	> 400 kb	> 400 kb	S	NT
K125	Cotton rat	USA	1992	(172)	1	-	+	-	-	150 kb	200 kb	150 kb	2 bands; > 400 kb **	> 400 kb	X	A5B13
K5	Neonatal ward	Mozambique	1992	(172)	2	-	+	-	-	-	150 kb	100 kb	nd	nd	nd	nd
01/02	Barn environment	Czech Republic	2002	This study	1	-	+	-	-	2 bands; 100, 125 kb	400 kb	100 kb	2 bands; > 400 kb **	200 kb	J	NT
04/00	Human	Czech Republic	2002	This study	0,75	-	+	-	-	-	> 400 kb	> 400 kb	nd	nd	nd	nd
05/00	Human	Czech Republic	2002	This study	1	-	+	-	-	-	> 400 kb	> 400 kb	nd	nd	nd	nd
06/00	Human	Czech Republic	2002	This study	1	-	+	-	-	-	2 bands; 150, 300 kb	150 kb	nd	nd	nd	nd
07/00	Human	Czech Republic	2002	This study	0,75	-	+	-	-	-	> 400 kb	> 400 kb	nd	nd	nd	nd
11/01	Human	Czech Republic	2002	This study	0,5	-	+	-	-	200 kb	2 bands; > 400, 200 kb	125 kb	> 400 kb	> 400 kb	C	New
15/01	Human	Czech Republic	2002	This study	0,75	-	-	-	-	nd	nd	nd	nd	nd	nd	nd
17/01	Human	Czech Republic	2002	This study	0,38	-	+	-	-	> 400 kb	> 400 kb	> 400 kb	2 bands; > 400, 250 kb	150 kb	E	New
CCUG38359	Human	Sweden	2002	This study	0,75	-	+	-	-	200 kb	2 bands; > 400, 200 kb	125 kb	> 400 kb	> 400 kb	C	New

Strain	Animal/Human origin	Geographic origin	Date of isolation	Reference	Oxacillin MIC (µg/ml)	PCR *mecA*	Dot blot (*ccr*)	Dot blot (*mecI*)	Dot blot (*mecRI*)	SmaI-*ccr* (kb)	SmaI-*pbpD* (kb)	SmaI-*orfx* (kb)	ApaI-*pbpD* (kb)	ApaI-*orfX* (kb)	SmaI-PFGE type	*ccrAB* type
CNCTC M32/86	Pig skin	Czech Republic	2002	This study	1	-	+	-	-	150 kb	2 bands; >400, 400 kb	150 kb	2 bands; > 400, 250 kb	200 kb	E	A5B13
CNCTC M59/89	Domestic fowl	Czech Republic	2002	This study	0,75	-	-	-	-	nd	nd	nd	nd	nd	nd	nd
CNCTC M60/89	Cattle	Czech Republic	2002	This study	1	-	-	-	-	nd	nd	nd	nd	nd	nd	nd
CNCTC M61/89	Cattle	Czech Republic	2002	This study	0,75	-	+	-	-	-	> 400 kb	> 400 kb	nd	nd	nd	nd
K105	Human	USA	1971	(172)	1	-	-	-	-	nd	nd	nd	nd	nd	nd	nd
K12	Arabian horse	USA	1992	(172)	1	-	-	-	-	nd	nd	nd	nd	nd	nd	nd
K13	Eastern grey squirrel	USA	1992	(172)	0,5	-	-	-	-	nd	nd	nd	nd	nd	nd	nd
K132	Howler monkey	Panama	1976	(172)	0,5	-	-	-	-	nd	nd	nd	nd	nd	nd	nd
K139	Holstein cow	USA	1992	(172)	1	-	-	-	-	nd	nd	nd	nd	nd	nd	nd
K14	Eastern harvest mouse	USA	1992	(172)	0,75	-	-	-	-	nd	nd	nd	nd	nd	nd	nd
K140	Opossum	USA	1972	(172)	1	-	-	-	-	nd	nd	nd	nd	nd	nd	nd
K141	California mouse	USA	1992	(172)	1	-	+	-	-	250 kb	> 400 kb	250 kb	> 400 kb	> 400 kb	H	New
K142	Horse	USA	1992	(172)	1	-	-	-	-	nd	nd	nd	nd	nd	nd	nd
K143	Racoon	USA	1972	(172)	1	-	-	-	-	nd	nd	nd	nd	nd	nd	nd
K144	Jersey calf	USA	1992	(172)	1	-	-	-	-	nd	nd	nd	nd	nd	nd	nd
K147	Jersey calf	USA	1992	(172)	0,5	-	-	-	-	nd	nd	nd	nd	nd	nd	nd
K148	Bottlenose dolphin	USA	1992	(172)	0,38	-	+	-	-	2 bands; > 400, 150 kb	> 400 kb	150 kb	2 bands; > 400, 300 kb	300 kb	P	NT
K149	Eastern grey squirrel	USA	1972	(172)	0,75	-	-	-	-	nd	nd	nd	nd	nd	nd	nd
K150	Pilot whale	USA	1992	(172)	1	-	-	-	-	nd	nd	nd	nd	nd	nd	nd
K152	Holstein cow	USA	1992	(172)	1	-	-	-	-	nd	nd	nd	nd	nd	nd	nd
K154	Norway rat	USA	1992	(172)	1,5	-	-	-	-	nd	nd	nd	nd	nd	nd	nd
K155	Pilot whale	USA	1992	(172)	1,5	-	-	-	-	nd	nd	nd	nd	nd	nd	nd
K156	Beef tongue	USA	1992	(172)	0,75	-	+	-	-	2 bands; > 400, 125 kb	> 400 kb	125 kb	> 400 kb	100 kb	M	NT
K16	Jersey calf	USA	1992	(172)	0,75	-	-	-	-	nd	nd	nd	nd	nd	nd	nd
K161	Opossum	USA	1972	(172)	0,75	-	-	-	-	nd	nd	nd	nd	nd	nd	nd

Strain	Animal/Human origin	Geographic origin	Date of isolation	Reference	Oxacillin MIC (µg/ml)	PCR *mecA*	Dot blot (*ccr*)	Dot blot (*mecI*)	Dot blot (*mecRI*)	SmaI-*ccr* (kb)	SmaI-*pbpD* (kb)	SmaI-*orfx* (kb)	ApaI-*pbpD* (kb)	ApaI-*orfX* (kb)	SmaI-PFGE type	*ccrAB* type
K162	Bottlenose dolphin	USA	1992	(172)	1	-	+	-	-	100 kb	> 400 kb	100 kb	> 400 kb	> 400 kb	Y	A5B13
K164	Bottlenose dolphin	USA	1992	(172)	1	-	+	-	-	>400 kb	> 400 kb	>400 kb	2 bands; > 400 kb **	> 400 kb	M	A5B13
K165	Hairy-footed hamster	USA	1992	(172)	0,75	-	+	-	-	>400 kb	> 400 kb	>400 kb	> 400 kb	> 400 kb	M	New
K167	Beef tongue	USA	1992	(172)	0,75	-	+	-	-	-	> 400 kb	>400 kb	nd	nd	nd	nd
K170	Bottlenose dolphin	USA	1992	(172)	1	-	-	-	-	nd	nd	nd	nd	nd	nd	nd
K171	Opossum	USA	1972	(172)	0,5	-	+	-	-	50 kb	150 kb	100 kb	> 400 kb	> 400 kb	T	New
K172	Jersey calf	USA	1992	(172)	0,75	-	-	-	-	nd	nd	nd	nd	nd	nd	nd
K173	Eastern grey squirrel	USA	1972	(172)	0,75	-	-	-	-	nd	nd	nd	nd	nd	nd	nd
K174	Opossum	USA	1972	(172)	0,75	-	+	-	-	-	2 bands; 350, 200 kb	150 kb	nd	nd	nd	nd
K175	Bottlenose dolphin	USA	1992	(172)	0,75	-	-	-	-	nd	nd	nd	nd	nd	nd	nd
K18	Beef lips	USA	1992	(172)	1	-	-	-	-	nd	nd	nd	nd	nd	nd	nd
K2	Beef tongue	USA	1992	(172)	1	-	-	-	-	nd	nd	nd	nd	nd	nd	nd
K20	Bottlenose dolphin	USA	1992	(172)	1	-	+	-	-	200 kb	> 400 kb	200 kb	> 400 kb	> 400 kb	N	A5B13
K21	Pilot whale	USA	1993	(172)	0,5	-	+	-	-	> 400 kb	> 400 kb	> 400 kb	> 400 kb	> 400 kb	L	A5B5
K22	Morgan horse	USA	1992	(172)	1,5	-	-	-	-	-	> 400 kb	200 kb	nd	nd	nd	nd
K23	Red kangaroo	USA	1992	(172)	0,75	-	-	-	-	-	> 400 kb	> 400 kb	nd	nd	nd	nd
K24	Morgan horse	USA	1992	(172)	0,75	-	-	-	-	-	> 400 kb	150 kb	nd	nd	nd	nd
K25	Prairie vole	USA	1992	(172)	0,5	-	-	-	-	-	> 400 kb	> 400 kb	nd	nd	nd	nd
K51	Human	USA	1971	(172)	1,5	-	-	-	-	nd	nd	nd	nd	nd	nd	nd
K69	Human	USA	1986	(172)	1	-	+	-	-	200 kb	400 kb	200 kb	> 400 kb	> 400 kb	G	NT
KLO56	Opossum	USA	1972	(172)	0,75	-	+	-	-	> 400 kb	>400 kb	> 400 kb	> 400 kb	150 kb	M	New
KLO58	Squirrel monkey	USA	1972	(172)	0,75	-	-	-	-	-	>400 kb	200 kb	nd	nd	nd	nd
KLO59	Domestic dog	USA	1972	(172)	2	-	-	-	-	-	>400 kb	150 kb	nd	nd	nd	nd
KLO63	Eastern grey squirrel	USA	1972	(172)	1	-	-	-	-	-	>400 kb	> 400 kb	nd	nd	nd	nd
KLO64	Southern flying squirrel	USA	1972	(172)	2	-	-	-	-	-	>400 kb	150 kb	nd	nd	nd	nd

+, positive; -, negative; nd, not determined; **, two bands with a very similar fragment size hybridized.

Table S2. Sequence and predicted amplicon size (bp) of the primers used in this study. In addition, the locations of the primers in the gene are indicated. The bibliographic source is also shown.

Gene	Primers (5´-3´)	Location of the primers (gene size), bp	Amplicon size, bp	Source
ccrAB	αc: ATCTATTTCAAAAATGAACCA	1288 (*ccrA*, 1350)	560	(125)
	βc: ATTGCCTTGATAATAGCCITCT	475 (*ccrB*, 1629)	560	(125)
ccrAB2	α2:TAAAGGCATCAATGCACAAACACT	915 (*ccrA*, 1350)	1000	(125)
	βc: ATTGCCTTGATAATAGCCITCT	475 (*ccrB*, 1629)	1000	(125)
ccrC	γF: CGTCTATTACAAGATGTTAAGGATAAT	190 (1676)	520	(125)
	γR:CCTTTATAGACTGGATTATTCAAAATAT	710 (1676)	520	(125)
ccrA	ccrAF1: YCCWAAYTAYTGTGGYCGTGT	642 (1350)	296	(142)
	ccrAR1: TKYTKGTGCRTTKATNCCT	938 (1350)	296	(142)
ccrB	ccrBF1: CGWYTRGCWMGWAAYACHTC	340 (1629)	496	(142)
	ccrBR1: CTTTTCGWCKYTTWTCRYTCC	836 (1629)	496	(142)
mecA	MRSA GL F: TATGAGATAGGCATCGTTCC	505 (2007)	334	This study
	MRSA GL R: TTACCAATAACTGCATCATC	839 (2007)	334	This study
mecI	mI3: CAAAAGGACTGGACTGGAGTCCAAA	115 (372)	180	(125)
	mI4: CAAGTGAATTGAAACCGCCT	295 (372)	180	(125)
mecRI	mecR3: GTCTCCACGTTAATTCCATT	133 (1758)	310	(125)
(MS	mecR4:GTCGTTCATTAAGATATGACG	443 (1758)	310	(125)
domain)				
orfX	orfX-F1: CGCTTAGGTGCATATACAAAGAC	76 (480)	369	This study
	orfX- R1: ACGCACTATCACATTGTTCC	445 (480)	369	This study
pbpD	SAMECA165: CGATAATGGTGAAGTAGA	162 (2001)	1308	(174)
	SAMECA1482: TATATCTTCACCAACACC	1470 (2001)	1308	(174)

Table S3 - *ccr* types and SmaI hybridization fragments found in isolates of each PFGE type. Epidemiological data is also shown.

PFGE type (No isolates)	Year of isolation	Geographic origin	Animal/Human source	*ccrAB* types	Hybridization fragment, kb (number of isolates)		
					SmaI-*ccr*	SmaI-*mecA1*/ApaI-*mecA1*	SmaI-*orfX*/ApaI-*orfX*
A (2)	2002	Czech Republic	Chicken, turkey	New *ccrA* type 2/B13	100 (1)	50/>400 (1)	150/>400 (1)
				A5B13	>400 (1)	>400/>400 (1)	400/>400 (1)
B (2)	1992,2010	USA, Portugal	European red squirrel, human	A5B13	150 (1)	>400/>400 (1)	150/>400 (1)
				A5B13	>400 (1)	>400/>400 (1)	>400/>400 (1)
C (2)	2002	Czech Republic, Sweden	Human	New *ccrA* type 1/B13	200 (2)	>400, 200/>400 (2)	125/>400 (2)
D (2)	2002	Czech Republic	Human	NT, New *ccrA* type 1/B5/B7	300 (2)	350/>400 (2)	300/>400 (2)
E (3)	2002	Czech Republic, Yugoslavia	Pig, human, dog	New *ccrA* type 2/B13	>400 (1)	>400/>400 (1)	>400/150 (1)
				New *ccrA* type 1/B13	>400 (1)	>400/>400,250 (1)	>400/150 (1)
				A5B13	150 (1)	>400/>400,250 (1)	150/200 (1)
F (1)	2002	Yugoslavia	Dog	New *ccrA* type 2/B13	>400 (1)	> 400,400/>400 (1)	400/>400 (1)
G (5)	1986, 1996, 1997, 2002	Portugal, Czech Republic, USA	Human, snake	A1B13, A1B13	100 (2)	> 400/>400 (2)	100/>400 (2)
				NT	200 (1)	400/>400 (1)	400/>400 (1)
				NT	150 (1)	>400/* (1)	150/*(1)
				New *ccrA* type 3/B9	125 (1)	>400/>400 (1)	125/>400 (1)
H (1)	1992	USA	California mouse	New *ccrA* type 3/B9	250 (1)	>400/>400 (1)	250/>400 (1)
I (2)	1996,1998	Portugal	Human	NT	80,100 (2)	>400/>400 (2)	100/200 (2)
J (1)	2002	Czech Republic	Barn environment	NT	100,125 (1)	400/> 400 (1)	100/200 (1)
K (1)	2002	Sweden	Human	New *ccrA* type 3/B9	125 (1)	>400 /* (1)	125/*(1)
L (1)	1993	USA	Pilot whale	A5B5	>400 (1)	>400/>400 (1)	>400/>400 (1)
M (4)	1972, 1992	USA	Bottlenose dolphin, hairy-footed hamster, opossum, beef tongue	A5B13, New *ccrA* type 1/B13	>400 (2)	>400/>400 (2)	>400/>400 (2)
				New *ccrA* type 1/B13	>400 (1)	>400/>400 (1)	>400/150 (1)
				NT	>400,125 (1)	>400/>400 (1)	125/200 (1)

PFGE type (No isolates)	Year of isolation	Geographic origin	Animal/Human source	*ccrAB* types	Hybridization fragment, kb (number of isolates)		
					SmaI-*ccr*	SmaI-*mecA1*/ApaI-*mecA1*	SmaI-*orfX*/ApaI-*orfX*
N (1)	1992	USA	Bottlenose dolphin	A5B13	200 (1)	>400/>400 (1)	200/>400 (1)
O (2)	1992,	Czech Republic,	Norway rat, snake	New *ccrA* type 3/B9	125 (1)	>400/>400 (1)	125/>400 (1)
	2002	USA		A5B9	>400 (1)	>400/>400 (1)	>400/>400 (1)
P (1)	1992	USA	Bottlenose dolphin	NT	>400, 150 (1)	>400/>400, 300 (1)	150/300 (1)
Q (3)	1992,	Portugal,	Human, Norway rat	New *ccrA* type 1/B5/7	175 (1)	100 />400 (1)	175/>400 (1)
	1996,	Sweden, USA		New *ccrA* type 1/B5/7	<50 (1)	150/>400 (1)	100/300 (1)
	2002			New *ccrA* type 3/B13	<50 (1)	250,150/200 (1)	100/300 (1)
R (1)	1992	USA	Norway rat	A1B13	250 (1)	250/200 (1)	250/350 (1)
S (1)	1992	Czech Republic	Human	NT	250 (1)	250/>400 (1)	100/>400 (1)
T (1)	1972	USA	Opossum	New *ccrA* type 2/B5/7	50 (1)	150/>400 (1)	100/>400 (1)
U (1)	2002	Czech Republic	Cattle	A5B13	50,75 (1)	>400,100/>400 (1)	75/>400 (1)
V (1)	1992	USA	Jersey cattle heifer	New *ccrA* type 1/B5/7	200 (1)	>400/>400 (1)	200/>400 (1)
X (1)	1992	USA	Cotton rat	A5B13	150 (1)	200/>400 (1)	150/>400 (1)
Y (1)	1992	USA	Bottlenose dolphin	A5B13	100 (1)	>400/>400 (1)	100/>400 (1)

* no digestion was obtained; NT: non-typable; *ccr* types were considered new when its sequence had < 85% homology with known *ccr* types.

Table S4. Highest nucleotide identities (%) between *ccr* allotype carried by *S. sciuri* and *ccr* allotypes carried by prototype *S. aureus* or coagulase-negative staphylococci (CoNS) strains. Ten *S. sciuri* strains carried non-typable *ccrA* (4) or *ccrB* (6) alleles. A representative sequence of each cluster defined by the phylogenetic tree (Figures 3 and 4) was used in the analysis.

Allotype (no strains)	**Similarity with known *ccr* types**		**SCC*mec*/SCC***	**Reference**
	% identity	**Species/Strain**		
***ccrA5* (11)**	94 % *ccrA5*	*S. cohnii*/WC28	*ccr*A5B3/*mec* complex A	(168)
New *ccrA* type 1 (11)	82 % *ccrA5*	*S. cohnii*/WC28	*ccr*A5B3/*mec* complex A	(168)
***ccrA1* (6)**	87% *ccrA1*	*S. aureus*/COL	SCC*mec* I	www.sccmec.org
New *ccrA* type 3 (5)	80 % *ccrA2*	*S. aureus*/N315	SCC*mec* II/IV	www.sccmec.org
New *ccrA* type 2 (4)	74 % *ccrA5*	*S. cohnii*/WC28	*ccr*A5B3/*mec* complex A	(168)
***ccrB13* (19)**	90 % *ccrB13*	*S. cohnii*/M10F1	MR; non-typable	www.ccrbtyping.net
***ccrB5/B7* (8)**	98% *ccrB7*	*S. pseudointermedius*/KM241	Not available	www.ccrbtyping.net
	93% *ccrB5*	*S. lugdunensis*/S22S5	Not available	www.ccrbtyping.net
***ccrB9* (5)**	99% *ccrB9*	*S. saprophyticus*/TSU33	Not available	www.ccrbtyping.net
***ccrB1* (2)**	97% *ccrB1*	*S. aureus*/COL	SCC*mec* I	www.sccmec.org
***ccrB7* (1)**	91% *ccrB7*	*S. pseudointermedius*/KM241	Not available	www.ccrbtyping.net

*SCC*mec* types or SCC elements previously described carrying the *ccr* types with the highest % identity with each *ccr* type found in *S. sciuri*. When a SCC*mec* was non-typeable, the combination of type of *ccr* complex and the class of *mec* complex is shown, when available; MR – methicillin-resistan

Evidence for the evolutionary steps leading to β-lactam resistance in staphylococci

Joana Rolo, Peder Worning, Jesper Boye Nielsen, Rory Bowden, Ons Bouchami, Peter Damborg, Luca Guardabassi ,Vincent Perreten, Alexander Tomasz, Henrik Westh, Hermínia de Lencastre , Maria Miragaia

Manuscript in preparation

All the experimental work and the whole-genome sequencing analysis – with the exception of SNP analysis – was performed by J. Rolo

ABSTRACT

The most epidemiologically important mechanism of antibiotic resistance in staphylococci is associated with *mecA* – an acquired gene which encodes an extra penicillin-binding protein (PBP2a) with low affinity to all β-lactam antibiotics. In this study we aim to understand the evolutionary steps linking *mecA* precursors, identified in species of *Staphylococcus sciuri* group - the most primitive staphylococci - to the β-lactam resistance gene *mecA* and the resistant phenotype.

For this purpose, we sequenced the genome of 106 isolates of *S. sciuri*, *S. vitulinus* and *S. fleurettii* and determined their β-lactam susceptibility profiles. Single-nucleotide polymorphisms (SNP) analysis of core genome was performed to assess isolates genetic relatedness and detailed phylogenetic analysis of the *mecA* homologues was achieved through nucleotide/aminoacid sequence analysis. In addition, the predicted structure of *mecA* homologue-encoded PBPs of β-lactam susceptible and resistant strains were compared.

We showed for the first time that β-lactam resistance emerged several times and by different mechanisms during evolution of the most primitive staphylococcal species. The development of resistance involved different strategies including structural diversification of native PBPs in the non-binding domain, changes in the *mecA* homologues promoter and adaptation of the genetic background of the bacteria. Moreover, our data suggests that jump to the human host and antibiotic exposure were the forces driving evolution towards a resistance phenotype.

These results highlighted the numerous resources available to bacteria to adapt to antibiotic pressure. Moreover, the molecular strategies associated to β-lactam resistance revealed here could help to predict the emergence of new antimicrobial resistance genes.

INTRODUCTION

The most important antibiotic resistance mechanism in staphylococci is associated with the *mecA* gene, which forms the basis of resistance to the large class of β-lactam antibiotics. The *mecA* gene is not native to *Staphylococcus aureus* and other pathogenic staphylococci; it is transferred horizontally among staphylococcal strains, in a complex cassette called staphylococcal cassette chromosome *mecA* (SCC*mec*) (109), which inserts always at the same locus in the chromosome, downstream *orfX* (which encodes a RNA methyltransferase) (109, 115). Several studies have demonstrated that acquisition of *mecA* confers a competitive advantage in the nosocomial, community and veterinary settings (203, 204). In particular, introduction of the *mecA* determinant into the *S. aureus* genome, has led to the emergence of methicillin-resistant *S. aureus* (MRSA), driving one of the most important pandemics worldwide (204).

The *mecA* determinant encodes an extra penicillin-binding protein (PBP2a) that has low affinity for β-lactams (7). The expression of resistance is achieved by a slow rate of acylation of PBP2a as well as a low affinity of the enzyme for β-lactams (205). Structural studies have revealed that the poor acylation rate that PBP2a presents when in contact with β-lactams is due to a distorted active site (108). Distortion is thought to be achieved by the flexibility of the non-binding domain and regions surrounding the active site groove in the transpeptidase domain (composed by the residues Ser403, Lys406, Tyr446, Ser462, Asn464, Ser598 and Thr600)(108). Furthermore, it was found that the position of Ser403 is crucial for a nucleophilic attack of the β-lactam ring, which leads to acylation of the protein (108). The displacement of Ser403 residue is therefore a critical key feature of PBP2a.

The first MRSA isolates were identified in the UK among clinical isolates in 1961, shortly after the introduction of methicillin into clinical practice (200, 206). Early MRSA were found to present a heterogeneous profile of resistance to β-lactams (121). Further studies have revealed that mutations in genes associated with the cell division as well as central metabolism influence the expression of β-lactam resistance and the resulting phenotype (the so-called auxiliary genes) (207). In addition, expression of

homogeneous high level resistance has been associated with the activation of the bacterial stringent response, provoked by mutations in the *relA* system (208, 209) and related regulons and genes (210). These findings evidence the important role of the genetic background in the expression of β-lactam resistance in *S. aureus*.

The rapid emergence of such an efficient resistance mechanism raised the hypothesis that *mecA* was already present in one of the numerous staphylococcal species prior to the introduction of the antibiotic. In fact, an homologue named *mecA1*, with 80% nucleotide identity to *mecA* has been identified in the primitive coagulase-negative *Staphylococcus sciuri*, where this gene was found to be ubiquitous (172). Several lines of evidence suggested that *mecA1* was the *mecA* precursor. While *mecA1* does not confer resistance to β-lactams in *S. sciuri*, there are reports of β-lactam resistant strains that have alterations in the promoter region of this gene (177). When introduced – in the laboratory - into a *S. aureus* genetic background, *mecA1* was able to confer β-lactam resistance and produce a protein with properties resembling that of MRSA PBP2a (176, 189). Additional *mecA* homologues have also been identified in related species. In particular, in *Staphylococcus vitulinus*, a *mecA* homologue with 90% nucleotide identity with MRSA *mecA* has been identified (*mecA2*)(178). Additionally, the *mecA* along with its regulators, *mecI* and *mecRI* has been identified in a small number of *Staphylococcus fleurettii* isolates (173).

Despite the importance of *mecA* in the epidemiology of antibiotic resistant staphylococci, the evolutionary history of this gene has remained unclear. The purpose of the study described here was to shed light on evolutionary steps linking the several native *mecA* homologues identified in primitive coagulase negative staphylococci to the β-lactam resistance gene *mecA* and the resistant phenotype. This knowledge could serve as a model for the evolution of other antibiotic resistance genes and could help to predict the emergence of new antibiotic resistance determinants.

METHODS

Bacterial strain collection: a collection of 106 staphylococcal isolates, comprising 76 *S. sciuri*, 18 *S. vitulinus* and 12 *S. fleurettii* was assembled. Regarding *S. sciuri*, 28 isolates were obtained from humans, while the remaining 45 isolates were recovered from both wild and domesticated mammals (supplementary table S1). Isolates were collected in different countries (Czech Republic, Denmark, Portugal, Switzerland, Sweden, former Yoguslavia, Mozambique, Panama and USA) over a large sampling period-1972-2012. *S. vitulinus* and *S. fleurettii* isolates were collected from horses and bovine mastitis milk samples, in Denmark, Switzerland and the Netherlands, in 2004, 2005 and 2010. *S. sciuri* isolates were identified at the species level by 16S RNA ribotyping and API-Staph (Biomerieux, France). *S. fleurettii* and *S. vitulinus* were identified at the species level by 16S sequencing.

β-lactam susceptibility: was assessed by oxacillin Etest (Biomérieux, France). The breakpoint for defining susceptibility isolate was evaluated as suggested by EUCAST (www.eucast.org): isolates were considered susceptible when MIC < 3 µg/ml. Population analysis profiles (PAPs) for oxacillin were determined for representative isolates (28/60 *S. sciuri* exclusively carrying *mecA1*, 23/37 isolates carrying *mecA*, 9/9 isolates carrying *mecA2*) as previously described (211). The PAP results of *S. sciuri* isolates have already been published (172, 202).

Whole-genome sequencing and *de novo* assembly: DNA was extracted with the phenol/chlorophorm extraction method (*S. sciuri*) and the Qiagen DNEasy Blood & Tissue Kit (*S. vitulinus* and *S. fleurettii*) (Qiagen, Limburg, The Netherlands). The sequencing was performed in a HiSeq with an estimated coverage of 40x. The reads were assembled *de novo* using VELVET (33).

Reference genome *S. fleurettii* 402567: DNA of a *S. fleurettii* 402567 was prepared by phenol/chlorophorm extraction and was sequenced using a PacBio RS apparatus. De

novo assembly was performed using HGAP 3 (https://github.com/PacificBiosciences/Bioinformatics-Training/wiki/HGAP-in-SMRT-Analysis).

A reference genome was produced by combining Illumina and PacBio sequencing data for a single strain, *S. fleurettii* isolate 402567. PacBio reads were combined with Illumina reads obtained for each isolate in CLC Genomics Workbench, using the Genome Finishing module. The resulting contigs were ordered using *Staphylococcus xylosus* closed genome (average nucleotide identity with *S. sciuri*, 78.01 %; *S. vitulinus*, 77.12%; and *S. fleurettii*, 78.54%). Estimated gaps were closed by mapping Illumina data of remaining *S. fleurettii* strains to the contigs. The resulting closed genome was annotated with RAST (http://rast.nmpdr.org/).

Estimation of *S. sciuri* strain-to-strain phylogenies: The reference genome *S. fleurettii* 402567 was used to perform a SNP analysis of the predicted core genome of *S. sciuri* isolates. SNP analysis was performed using Stampy (version 1.0.11) where reads were mapped to the reference genome. SNP calling was performed using SAMtools (version 0.1.12) and Neighbor Joining (NJ) analysis was used to assess the phylogeny. Trees were drawn in FigTree (http://tree.bio.ed.ac.uk/software/figtree/).

Phylogenetic analysis of *mecA* homologues: nucleotide sequences of *mecA* homologues and two core-associated genes (*aroE* and *gyrB*) were identified by BLAST analysis and were extracted from the sequence of the contigs. Alignments with the entire gene or regions corresponding to specific domains were performed with ClustalW. Phylogenetic trees were constructed with a neighbor-joining algorithm. To estimate the rates of evolution of *mecA* homologues, BEAST was used (30). The parameters were set as described by Gray *et al* (212). We used the Bayesian framework implemented in the BEAST software package under the general time reversible nucleotide substitution model. The molecular clock was calibrated under a strict molecular clock (which assumes the same evolutionary rates for all branches in

the tree), with the constant population size coalescent prior. The Markov chain Monte Carlo (MCMC) analysis was run up to 10^7 generations. Results were visualized in Tracer v.1.5, and proper mixing of the MCMC was assessed by calculating the effective sampling size (ESS) for each parameter. All ESS values were >200. The maximum clade credibility (MCC) tree, which is the tree with the largest product of posterior clade probabilities, was selected from the posterior tree distribution using the program TreeAnnotator (available as part of the BEAST package). Final trees were annotated with FigTree (http://tree.bio.ed.ac.uk/software/figtree/). The BEAST analysis was performed for all *mec* homologues sequences and for the sequences belonging to each *mec* gene separately.

Estimation of recombination/mutation rates: RDP4 (213) was used to predict which part of the *mecA* homologues sequences were under recombination. The recombination/mutation rate of *mecA1* was estimated with this program. The sequence of two *S. sciuri* core-associated genes (*aroE* and *gyrB*) were also used for comparison.

Modelling of protein structure: the structure of representative proteins encoded by the *meA* homologues was predicted using ModWeb (https://modbase.compbio.ucsf.edu/modweb/) (214). The predicted structure encoded by one *mecA* allele, one *mecA2* allele and six *mecA1* alleles (representing each major clade of the phylogenetic tree, 0.015 distance cut off) were obtained. Alignments of the structures modeled with PBP2a were produced in PyMol (The PyMOL Molecular Graphics System, Version 1.5.0.3 Schrödinger, LLC) and visually inspected for relevant alterations of the protein structure.

Assessment of genetic diversity: The degree of genetic diversity of the different *mecA* homologues was assessed by the Simpson's index of diversity (SID)(215), using a

confidence interval of 95%. The online tool available at http://darwin.phyloviz.net/ComparingPartitions/ was used.

Data deposition: the reference genome obtained was deposited in NCBI database and raw reads in ENA.

RESULTS

The *mecA* homologues were ubiquitous in *S. sciuri*, *S. vitulinus* and *S. fleurettii*

In order to understand the distribution and chromosomal location of *mecA* homologues in *S. sciuri*, *S. vitulinus* and *S. fleurettii*, we identified *mecA* homologues in the contigs obtained for the 106 strains with BLAST analysis. We observed that all strains carried at least one copy of *mecA* homologues in their chromosome, although their location could vary. These could be found either in the native location (200 kb from *orfX*) or in the *orfX*, the SCC*mec* insertion site. We confirmed that *mecA1* was ubiquitous in *S. sciuri* (172) and *mecA* was carried by all twelve *S. fleurettii* isolates (173) in the native location. *S. vitulinus* was different from the other species in the fact that the *mecA* homologue carried in the native location varied. Half of the strains (n=9) carried *mecA2* as previously reported (178) and the remaining strains either carried *mecA* (n=6) or did not carry any *mecA* homologue in this region (n=3). The *mecA* was the only *mecA* homologue that was found near the *orfX* inside complete SCC*mec* elements (data not shown). This was observed in 16 *S. sciuri* isolates carrying *mecA1* in the native location and in the three *S. vitulinus* that did not carry any *mecA* homologue in the native location.

High genetic diversity in *mecA1* contrasts with conservation of *mecA2* and *mecA*

To assess the level of genetic diversity in *mecA* homologues, we aligned all *mecA1*, *mecA* and *mecA2* sequences identified in our study (Supplementary Figure 1A). The *mecA1* was extremely diverse, including a total of 44 different alleles (Simpson's Index of Diversity (SID) = 97.2%, CI= 95.7%-98.7%) that varied between 93-100% in nucleotide identities (Supplementary Figure 1B). In contrast, *mecA2* and *mecA* were very conserved. The *mecA* had ten different alleles (SID=70.4%, CI=60.5%-80.2%), varying from 99.75 to 100% in nucleotide identity and *S. vitulinus* showed only two alleles (SID=21.6%, CI=9.7%-33.6%) that were 99.95% identical (Supplementary Table 1).

To further understand if the *mecA1* genetic diversity was associated to the diversity of *S. sciuri* genetic backgrounds or if it was a result of a faster evolutionary rate, we compared the SID observed for this gene with SID of two other *S. sciuri* housekeeping genes scattered in the chromosome, *aroE* and *gyrB*. The SID of both housekeeping genes in *S. sciuri* was similar between *gyrB* (34 alleles, SID=91%, CI= 86.1%-96.1%) and *aroE* (31 alleles, SID=92.7%, CI= 89.1%-96.4%), but lower than that obtained for *mecA1* (SID=97.2%, CI= 95.7%-98.7%). This difference was even clearer when the sequence of aminoacids was predicted and compared. While the SID of *mecA1*-encoded PBP4 was still high, 96.2% (40 different aminoacid sequences, CI= 94.5%-97.9%), the ones obtained for *gyrB*-encoded DNA gyrase subunit B and *aroE*-encoded Shikimate 5-dehydrogenase were lower (8 different aminoacid sequences, SID=75.3%, CI= 70.1%-80.6% and 22 different aminoacid sequences, SID=88.6%, CI= 83.6%-93.6%, respectively). This result suggests that *mecA1* is accumulating single-nucleotide polymorphisms (SNPs) and non-synonymous mutations faster than other *S. sciuri* housekeeping genes.

Non-synonymous mutations were mainly observed in the non-binding domain of *S. sciuri* PBP4

The protein encoded by *mecA*, PBP2a, is a protein of 668 aminoacids containing two domains: the non-binding domain (NB, 163 aminoacids), which is located in the positions 147-309 of the protein and the transpeptidase domain (TP, 314 aminoacids), corresponding to the positions 345-658 of the protein. The TP domain is the one responsible for the full expression of β-lactam resistance in methicillin-resistant staphylococcal isolates carrying PBP2a (108). Like in PBP2a, also in PBP4 of *S. sciuri*, the protein encoded by *mecA1*, these two domains were identified when a BLASTP analysis was performed. The NB domain corresponded to positions 146-308 of PBP4 and the TP domain was identified in positions 343-656 of PBP4. To understand if differences in aminoacid sequence observed in *mecA1* occurred in NB or TP regions, both regions were aligned for all strains. In addition, the corresponding nucleotide sequence was extracted from the WGS data and aligned. The NB domain corresponded to 489

nucleotides (*mecA1* pos 436-924; *mecA* pos 439-927; *mecA2* pos 439-927) and the TP domain corresponded to 942 nucleotides (*mecA1* pos 1027-1968; *mecA* pos 1033-1974; *mecA2* 1033-1974).

A close inspection of the alignments showed that the nucleotide sequence of the NB domain of *mecA1*-encoded PBP4 had 165/489 (34%) SNPs and the TP domain had 123/942 (13%) SNPs. Moreover, when a similar analysis was performed for the aminoacid sequence of PBP4, we found that the NB domain accumulated many more substitutions (59/163, 36%) than the TP domain (26/314, 8%). The results indicate that most of the SNPs observed in the NB domain corresponded to non-synonymous mutations, while those observed in the TP domain were synonymous.

The number of SNPs and aminoacid substitutions was much lower for *mecA/mecA2* and corresponding PBPs than that found in *mecA1*. The *mecA* NB nucleotide sequence showed only 8/489 SNPs (1.6%) and 3/163 aminoacid substitutions (1.8%) while the TP domain showed 20/942 SNPs (2.11%) and 7/314 (2.22%) aminoacid substitutions. Regarding *mecA2*-encoded domains they were very conserved: no SNPs were observed in the NB domain and a single SNP was observed in the TP domain (0.1%), resulting in a single aminoacid change (0.31%).

Evidences for the occurrence of recombination in *S. sciuri mecA1*

In order to understand the mechanism associated to genetic diversification of *mecA* homologues, allele numbers were assigned to each NB and TP domain nucleotide sequence, for each gene. A total of 25 alleles in NB domain of *mecA1* and 32 alleles in the TP domain were identified. The *mecA* showed only 6 NB alleles and 6 TP alleles and *mecA2* had a single NB allele and two TP alleles. Strikingly, we observed that there were cases where the same NB allele could be found in combination with different TP alleles or the other way around. This phenomenon was observed both in *mecA* and *mecA1*, but was more evident in *mecA1* in which a larger number of different combinations were observed (Supplementary Table 1). This result led us to

hypothesize that recombination between the two domains might be frequent in *S. sciuri*.

To further test this hypothesis, we used a pairwise scanning approach, using the RDP4 program. Briefly, the program identifies portions of the gene under recombination by aligning different portions of the gene individually, with the aid of a maximum-likelihood tree. A ClustalW alignment of all *mecA1*, *mecA2* and *mecA* alleles was used to verify the occurrence of recombination events. The analysis was also performed separately for all genes. We found that there were no estimated recombination events between the three *mecA* homologues (data not shown). The *mecA* and *mecA2* alleles also showed no recombination events within each gene individually. On the other hand, in the *mecA1*-carrying *S. sciuri* population, we found the occurrence of five different recombination events (Supplementary Figure 1C). Three of these events were predicted to have occurred in the N-terminal encoding region of the gene, one in the NB domain of the gene and one in the TP domain. RDP estimated that the average recombination rate/site was 0.0049 and the average mutation rate/site was 0.03191. Taken together, the average site recombination/mutation rate observed in *mecA1* gene was 0.15:1. Moreover, we observed that recombinant *mecA1* alelles were distributed among 46 isolates, representing 60.5% of the *S. sciuri* population.

To understand if recombination was specific of *mecA1* or if it was otherwise a frequent event occurring in *S. sciuri*, we searched for the occurrence of recombination events in the two core-associated *S. sciuri* genes, *gyrB* and *aroE*. We found a single recombination event among *gyrB* alleles and no recombinant events were detected in *aroE* alleles (data not shown). The *gyrB* like *mecA1* is located near the *oriC*, suggesting that recombination events might be favored during DNA replication in this region of the chromosome.

Different levels of β-lactam resistance are associated to specific *mecA* homologues

To assess if diversity in *mecA* homologues could have impact on the susceptibility to β-lactams we determined susceptibility to oxacillin for all isolates of *S. sciuri*, *S. fleurettii* and *S. vitulinus*, by Etest. In addition, susceptibility to oxacillin was evaluated by population analysis profiles in a representative collection that included 28/60 isolates carrying *mecA1* only; 23/37 carrying *mecA*; 9/9 carrying *mecA2*. Current MIC breakpoints were defined only for clinically significant *Staphylococcus* species. Since these cannot be directly applied to the staphylococcal species analyzed in this study, we decided to determine MIC breakpoints for the *S. sciuri* group of species. The MIC breakpoints for oxacillin were defined for this set of species by Etest as suggested by EUCAST. The MIC values obtained by Etest were compared for resistant strains carrying *mecA* and susceptible strains carrying *mecA* homologues *mecA1* and *mecA2*. The breakpoint for resistance was set at 3 μg/ml oxacillin and strains were considered resistant if they presented a MIC to oxacillin >3 μg/ml either by Etest or population analysis profiles (MIC presented by more than 90% of the population).

The great majority of *S. sciuri* strains carrying *mecA1* only (54/60) had an MIC to oxacillin that was lower than 3 μg/ml. The remaining six strains had a MIC >3 μg/ml or exactly 3 μg/ml: K4, K5, K7, Jug17, SS37 and SS41. Jug17 was the exclusive *mecA* negative strain carrying *mecA1* allele 4. The resistant strains K4, K5, K7, SS37 and SS41 shared the same *mecA1*, allele 17. Overall, there was a consistency between the *mecA1* allele carried by *S. sciuri* and the MIC of the isolates (Supplementary Table 1), in the absence of *mecA*. The 16 *S. sciuri* strains that carried *mecA* in addition to *mecA1* had MIC values, ranging from 16 to >256 μg/ml (Table 1).

Population analysis profiles of resistance to oxacillin of two *mecA*-positive isolates confirmed that these isolates were heterogeneously resistant, being able to survive up to 800 μg/ml of oxacillin (Figure 1A). We performed oxacillin population analysis profiles for 24 additional *mecA*-negative isolates that were susceptible to oxacillin by Etest. The great majority of these isolates produced heterogeneous profiles, with MIC lower than 3 μg/ml, but half of the isolates (11 out of 24) could grow up to 6-100 μg/ml (Figure 1A).

Table 1. Main characteristics of β-lactam resistant strains.

Strain	MIC µg/ml (eTest/PAP)	Mechanism of resistance	Phylogenetic group	Date of isolation	*mec* allele (date of emergence in population)	Recombinant *mec* allele
M1234	>256	SCC*mec*	*S. sciuri* new 2	2009	*mecA* 9 (2002)	+
M692	96	SCC*mec*	*S. sciuri* new 2	2007	*mecA* 9 (2002)	+
M2590	>256	SCC*mec*	*S. sciuri* new 2	2012	*mecA* 9 (2002)	+
M2276	>256	SCC*mec*	*S. sciuri* new 2	2011	*mecA* 9 (2002)	+
D573	>256	SCC*mec*	*S. sciuri* new 2	2007	*mecA* 9 (2002)	+
M1653	>256	SCC*mec*	*S. sciuri* new 2	2010	*mecA* 9 (2002)	+
CH17	>256	SCC*mec*	*S. sciuri* new 2	2010	*mecA* 8 (1982)	+
CH18	>256	SCC*mec*	*S. sciuri* new 2	2010	*mecA* 8 (1982)	+
M2710	>256	SCC*mec*	*S. sciuri* new 2	2012	*mecA* 7 (1977)	+
HSM851	16	SCC*mec*	*S. sciuri* new 2	2010	*mecA* 7 (1977)	+
Jug17	>256	Altered PBP4	*S. sciuri* new 2	2002	*mecA1* 4 (1982)	+
K3	>256	SCC*mec*	*S. sciuri rodentius*	1992	*mecA* 10 (1967)	+
K4	>256	Alterations in *mecA1* promoter	*S. sciuri rodentius*	1992	*mecA1* 17 (1982)	+
K5	25	Alterations in *mecA1* promoter	*S. sciuri rodentius*	1992	*mecA1* 17 (1982)	+
K7	>256	Genetic background?	*S. sciuri rodentius*	1992	*mecA1* 17 (1982)	+
SS37	25	Alterations in *mecA1* promoter	*S. sciuri rodentius*	1996	*mecA1* 17 (1982)	+
SS41	3	Alterations in *mecA1* promoter	*S. sciuri rodentius*	1996	*mecA1* 17 (1982)	+
CH16	24	SCC*mec*	*S. sciuri rodentius*	2010	*mecA* 7 (1977)	+
K6	>256	SCC*mec*	*S. sciuri rodentius*	1992	*mecA* 7 (1977)	-
M1640	96	SCC*mec*	*S. sciuri sciuri*	2010	*mecA* 7 (1977)	-
Jug1	>256	SCC*mec*	*S. sciuri* new 1	2002	*mecA* 7 (1977)	+
M1886	64	SCC*mec*	*S. sciuri* new 1	2011	*mecA* 7 (1977)	-
CH5	>256	Genetic background?	*S. vitulinus*	2005	*mecA* 4 (1987)	-
CH15	>256	Alterations in *mecA2* promoter	*S. vitulinus*	2004	*mecA2* 2 (1972)	-
CH19	8	*mecA* native location	*S. fleurettii*	2010	*mecA* 1 (1962)	-
CH20	6	*mecA* native location	*S. fleurettii*	2010	*mecA* 3 (1982)	-
CH21	4	*mecA* native location	*S. fleurettii*	2010	*mecA* 2 (1992)	-
CH23	4	*mecA* native location	*S. fleurettii*	2010	*mecA* 2 (1992)	-
CH24	>256	*mecA* native location	*S. fleurettii*	2010	*mecA* 2 (1992)	-
CH25	4	*mecA* native location	*S. fleurettii*	2010	*mecA* 6 (1972)	-
CH26	>256	*mecA* native location	*S. fleurettii*	2010	*mecA* 2 (1992)	-
CH27	4	*mecA* native location	*S. fleurettii*	2010	*mecA* 2 (1992)	-
CH29	4	*mecA* native location	*S. fleurettii*	2010	*mecA* 2 (1992)	-
402567	>256	*mecA* native location	*S. fleurettii*	2004	*mecA* 5 (1982)	-

In *S. vitulinus*, the level of resistance to oxacillin was similar, irrespective of the fact that strains carried *mecA* or *mecA2*. The great majority of strains carrying the *mecA2* (8/9) were susceptible having MIC values ranging from 0.1 to 3 μg/ml. The level of β-lactam resistance was the same for strains carrying *mecA2* allele 1 (n=2) and *mecA2* allele 2 (n=7). The only exception was a single strain, CH15, that carried *mecA2* allele 2, and had an MIC >400 μg/ml and a homogeneous resistant profile to oxacillin (Figure 1B). The *S. vitulinus* strains carrying *mecA* in the native location instead of *mecA2* (9 isolates) carried the *mecA* allele 4. Almost all these isolates presented a susceptible phenotype with MICs ranging from 0.2-1.5 μg/ml but had a heterogeneous profile in which cells could grow up to 100-400 μg/ml in PAPs (Figure 1B). The only exception was strain CH2, with MIC of 4 μg/ml and an heterogeneous profile and strain CH5 with a MIC of 400 μg/ml and a homogeneous profile. The three *S. vitulinus* isolates carrying *mecA* in SCC*mec* (CH1, H91 and CH3), carried *mecA* allele 7 and showed similar levels of resistance as those observed for isolates carrying *mecA* in the native location. Strains H91 and CH1 presented an Etest MIC of 2 and 1 μg/ml, respectively and had subpopulations capable of growing up to 200 to 400 μg/ml of oxacillin. Strain CH3 had a even more susceptible phenotype with MIC of 0.1 μg/ml and subpopulations that grew up to 0.4 μg/ml of oxacillin (Figure 1B).

In *S. fleurettii* the great majority of isolates were resistant to oxacillin with Etest MIC of 4->256 μg/ml. Five different *mecA* alleles were identified among the isolates; the great majority of isolates carried *mecA* allele 2 (8/12 isolates). Oxacillin population analysis profiles of these *S. fleurettii* isolates showed that almost all strains were heteroresistant and could grow up to 25-400 μg/ml (Figure 1C). The only exceptions were two strains, CH22 and CH28 (MIC=1 μ/ml) that could grow only up to 6 μg/ml and 0.75 μg/ml, respectively. Both these strains carried *mecA* allele 2; therefore, in this case there was not a direct correlation between their increased susceptibility to oxacillin and the *mecA* allele that these strains carried.

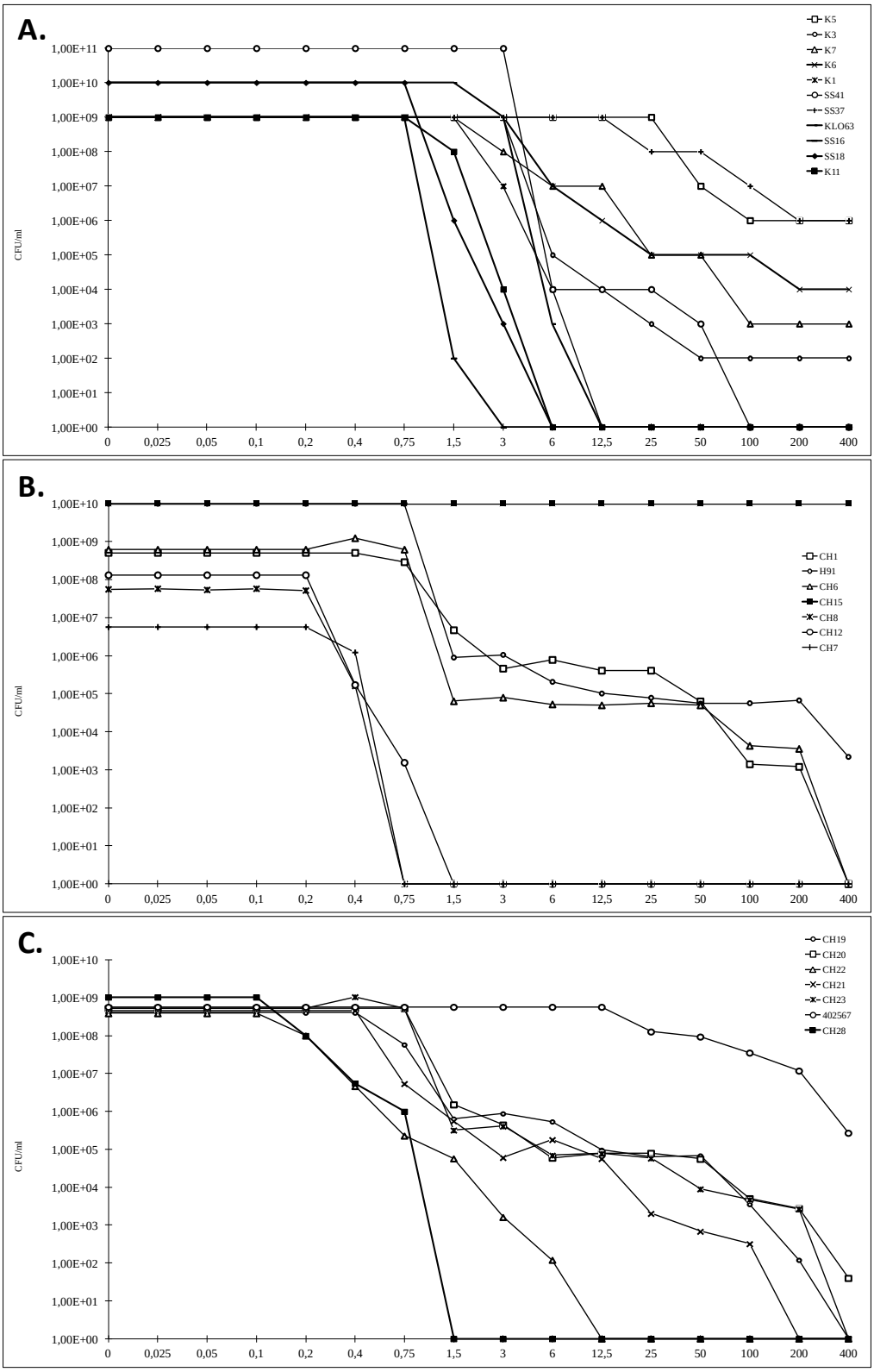


Figure 1. Oxacillin susceptibility population analysis profiles (PAPs) for representative *S. sciuri* (A)(172, 202), *S. vitulinus* (B) and *S. fleurettii* (C).

Phenotypic resistance to oxacillin is associated to mutations in the *mecA* homologues promoter

Mutations in the *mecA1* promoter of *S. sciuri* strains SS37 and SS41, namely the insertion of IS*256* and a single-nucleotide polymorphism (SNP), respectively, were previously suggested to be associated to a resistant phenotype (177). To further investigate the role of the *mecA* homologues promoter in the expression of β-lactam resistance in the population of *S. sciuri* group of species, we analyzed and compared the promoter region of isolates expressing resistance.

The *S. sciuri* strains carrying only *mecA1* and that showed phenotypic resistance to oxacillin were strains Jug17, K4, K5 and K7 (besides SS37 and SS41 that were already published, (177)). Our phylogenetic analysis (see below) has revealed that K4, K5 and K7 strains belong to the *S. sciuri* subspecies *S. sciuri rodentius*. We compared the sequence of the *mecA1* promoter in these strains with the sequence of methicillin susceptible type strain *S. sciuri rodentius* K3w (177). In strains K4 and K5 we found similar alterations to the promoter as the ones identified in strain SS41. We were able to identify the same -10 sequence and a very similar -35 sequence that differed from SS41 in a single nucleotide, corresponding to a change from thymine to cytosine. On the other hand, the promoter of *mecA1* of strain K7 had no alterations when compared to the promoter of *mecA1* in the susceptible strain K3w. Strain Jug17 belonged to a new phylogenetic group of *S. sciuri* (group 2, see below). We compared the promoter of this strain with a susceptible strain, SS16, that belonged to the same phylogenetic group and we found that the promoter region of the gene was very conserved between both strains (no alterations).

To understand if like in *S. sciuri* alterations in the promoter could be responsible for a resistant phenotype in *S. vitulinus*, we aligned the nucleotide sequence 200 bp upstream *mecA2* in two strains carrying the same allele but being either susceptible (CH10) or resistant to β-lactams (CH15). The single difference found was that the ribosome binding site (RBS) sequence, GGGAGGG, was located immediately upstream *mecA2* in strain CH15, in position -3, while in strain CH10 and remaining *mecA2*-carrying *S. vitulinus* strains, this sequence was located in position -6.

To further assess the impact of the promoter in the expression of oxacillin resistance in *S. vitulinus*, we compared the promoter of high-level homogeneously resistant CH5, a strain carrying *mecA* in the native locus, with the promoter of low-level heterogeneously resistant H91, that carried *mecA* in SCC*mec*. Analysis of the region located upstream *mecA* (200 bp) of strain CH5 showed the occurrence of two SNPs, in comparison with strain H91: in position -19, a change from a thymine to a cytosine and in position -97, a change from a thymine to guanidine. Nonetheless, the -10 sequence TATACT and the -35 sequence TTGACA were conserved between the two strains. These results suggest that CH5 increased homogeneous resistance to oxacillin might be related with alterations in the genetic background.

In *S. fleurettii*, we compared the promoter of strain 402567, expressing high level resistance with those of strains having low level resistance (CH22) or susceptible strains (CH28). Interestingly, we found that CH28 had a deletion of 16 bp in position -29 upstream *mecA* that may correspond to a deletion of the -10 sequence of the promoter (TATACT), which might explain its susceptible profile. On the other hand, strain CH22, had no SNPs in a 200 bp nucleotide sequence located upstream *mecA* when compared with the oxacillin-resistant strain 402567; we suggest that the increased susceptibility to oxacillin of CH22 strain might be related with alterations in the genetic background.

The structure of the active site in *mecA* homologue-encoded PBPs is associated to β-lactams susceptibility level

The finding of a high diversity in the nucleotide and aminoacid sequence of *mecA* homologues associated to the variability in susceptibility to β-lactams, led as to hypothesize that alterations in the structure of *mecA* homologues with impact in protein activity could explain different levels of β-lactam resistance. To test this hypothesis we predicted the structure of the proteins encoded by *mecA* homologues (*mecA* allele 5, *mecA2* allele 2, *mecA1* allele 4, *mecA1* allele 17, *mecA1* allele 21, *mecA1* allele 22, *mecA1* allele 25 and *mecA1* allele 42) using ModWeb (https://modbase.compbio.ucsf.edu/modweb/). Modeller has identified that the

overall predicted structure was either 100% similar to PPB2a, PDB code 1MWU (*mecA* allele 5), 90% similar (*mecA2* allele 2) or 82% similar (*mecA1* alleles).

The alignment of each structure predicted by Modeller with PBP2a in Pymol and the visual inspection of the protein active center showed that the protein encoded by *mecA* allele 5, carried by *S. fleurettii* had the exact same residues as *S. aureus* PBP2a. Although these residues were not in the exact same position, their orientation was the same (Figure 2A), suggesting that the acylation rate of these two proteins would be similar. Accordingly, the *S. fleurettii* isolates carrying this allele showed a high MIC to oxacillin, like is observed for strains carrying PBP2a (Supplementary Table 1).

In what respects to the protein encoded by *S. vitulinus, mecA2* allele 2, which is associated to susceptibility to β-lactams in almost all strains, it had two residues that were in positions different from those found in PBP2a: Ser403 and Thr600 (Figure 2B). In particular, Thr600 was not located in a β-sheet like is the case in PBP2a. These two differences might be enough to expose Ser403 to the β-lactam ring, leading to an acylation of the protein and to the susceptible phenotype observed. The only strain carrying *mecA2* allele 2 that was resistant to β-lactams (CH15) had an alteration in the promoter of the protein.

Finally, we studied in detail the active center of proteins encoded by *mecA1* (Ser401, Lys404, Tyr444, Ser460, Asn464, Ser596 and Thr598). The protein encoded by *mecA1* allele 42, represented by the *S. carnaticus* type strain K11 (171), that has a susceptible phenotype to β-lactams, had two residues, Ser596 and Thr598, that were in positions that were completely different from those found in PBP2a (Figure 2C). Modeller predicted that unlike PBP2a these two residues were not in a β-sheet conformation. The position of these residues, which surround the active site groove, was more relaxed, suggesting that the access to Ser401 is facilitated, which could explain the susceptible phenotype observed. The structures predicted for the remaining *mecA1* alleles analyzed were similar to the one described for strain K11, having the same position of the residues and structure of the active site. The only exception were the *S. sciuri* strains Jug17, HSM851 and M2710, carrying allele 4 (Figure

2D), in which Thr598 was much closer to Tyr444, thus "closing" the active site groove and putatively protecting Ser401 from interacting with the β-lactam ring. The low rate of acylation that should be provided by this distinctive structure might explain the high resistance to β-lactams observed in these particular strains.

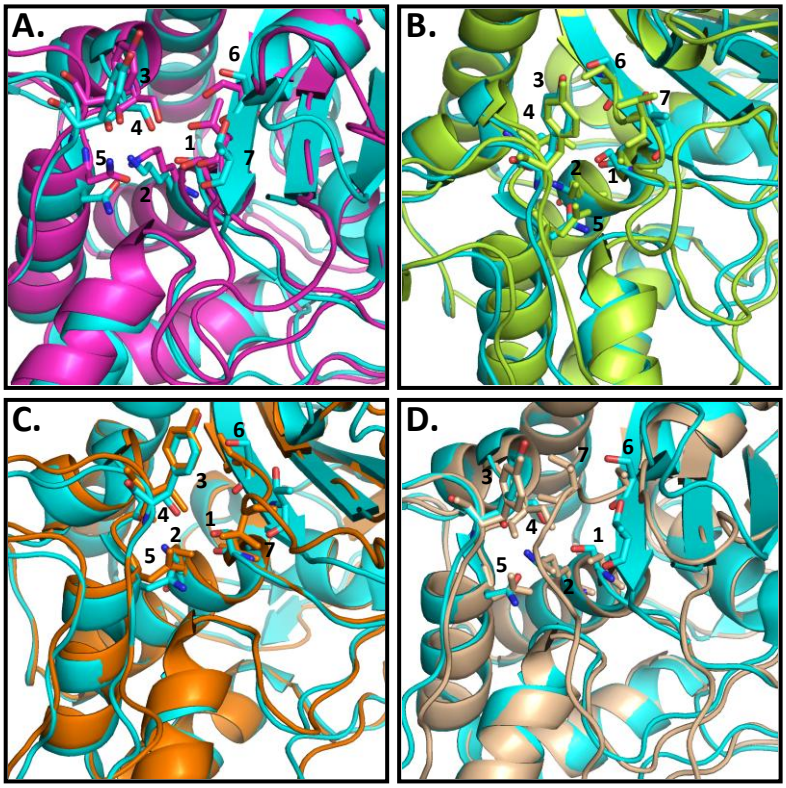


Figure 2. Alignment of the active centre of PBP2a (highlighted in cyan blue) and representative PBPs putatively encoded by *mecA* homologues. The structure of the PBP was predicted by Modeller and the alignment was produced in Pymol. **A.** 402567 *mecA* allele 5/PBP2a. **B.** CH10 *mecA* allele 2/PBP2a. **C.** K11 *mecA1* allele 42/PBP2a.1. **D.** JUG17 *mecA1* allele 4/PBP2a. Ser401/Ser403. 2. Lys404/Lys406. 3. Tyr444/Tyr446. 4. Ser460/Ser462. 5. Asn464/Asn466. 6. Ser596/Ser598. 7. Thr598/Thr600.

The majority of mutations conferring β-lactam resistance have emerged in the beginning of the antibiotic era

We investigated the evolutionary rate of each *mecA* homologue, using BEAST. The MCC tree using a strict molecular clock has shown that the most recent common ancestor of all *mecA* homologues alleles was estimated to have emerged in 1947, seven years after the introduction of penicillin into clinical practice, and that two different clades split in 1952 (Figure 3A). One of these clades originated *mecA1* and the other originated *mecA2* and *mecA*. According to our results, the first *mecA* allele emerged in 1962, in *S. fleurettii*. On the other hand, the first *mecA* allele carried by SCC*mec* emerged in the *S. sciuri rodentius* type strain, K3 (171) in 1967 and the first *mecA2* allele emerged in a *S. vitulinus* strain around 1957 (Figure 3B).

Regarding the strains that expressed β-lactam resistance, we found that the strains were clustered along the tree. The majority of the resistant *S. sciuri* strains clustered in clade 1 (Figure 3A). In particular, the majority of β-lactam resistant strains that were *mecA* negative clustered in subclade 2 of clade 1 (five strains, asterisk in Figure 3A). Jug17 clustered in subclade 1 of clade 1 (asterisk in Figure 3A). These two subclades emerged in 1962. Our data indicates that alterations leading to β-lactam resistance, either by altering the promoter of *mecA1* or the structure of PBP4, was achieved around the same time in an evolutionary prespective (Figure 3B).

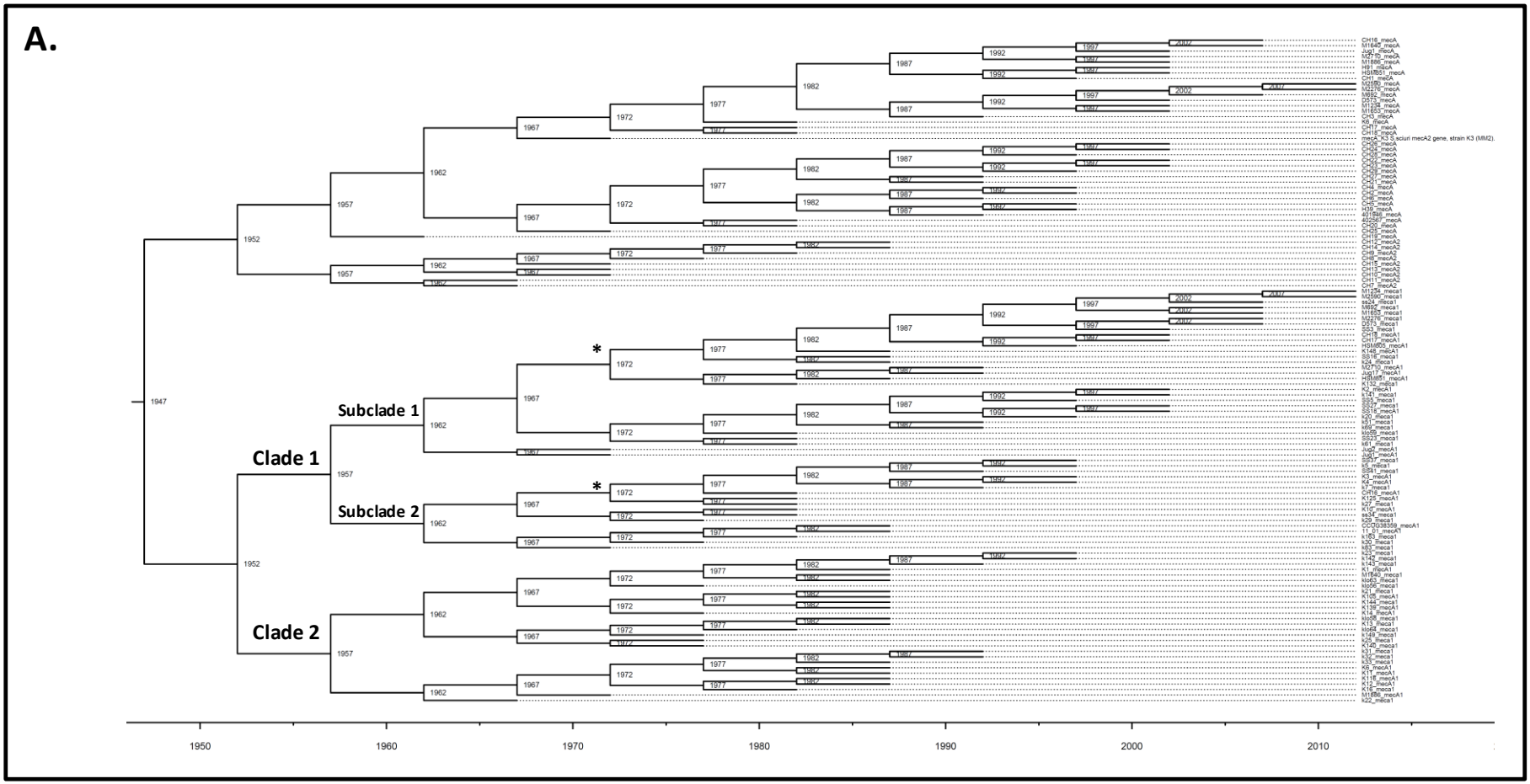


Figure 3A. Evolutionary history of *mecA* homologues alleles. Phylogenetic reconstruction of *mecA* homologues with BEAST.*clades that cluster the great majority of *S. sciuri* β-lactam resistant isolates.

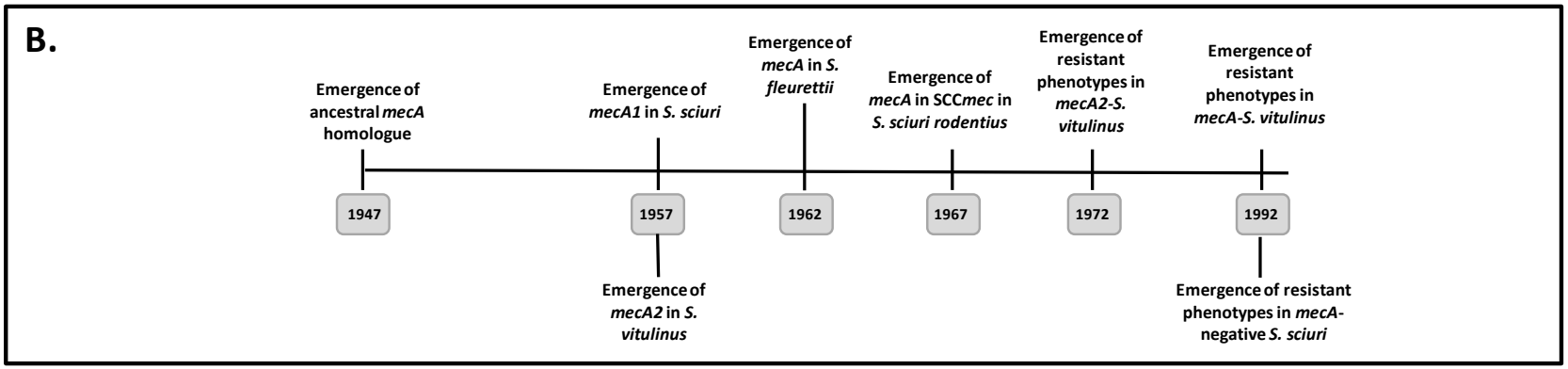


Figure 3B. Evolutionary history of *mecA* homologues alleles. Schematic representation of the emergence of the different *mecA* homologues.

β-lactam resistance was associated mainly to humans and specific phylogenetic clades

To investigate how *mecA* homologues providing β-lactam resistance evolved with phylogeny, we performed a SNP analysis with the predicted core genome of all isolates and reconstructed a phylogenetic tree based on the number of SNP differences (data not shown). Three well-defined phylogenetic groups were identified corresponding to each of the three species; there was an average of 150000 SNPs difference between the predicted core genomes of the three species (data not shown). In addition, we performed a SNP analysis of the predicted core genome of isolates belonging to each species independently, using *S. fleurettii* 402567 as a reference (Supplementary Figures 2A-C). The core genomes of *S. fleurettii* and *S. vitulinus* were more similar with each other (average 3000-9000 SNPs and *S. vitulinus* 4000-7000 SNPs difference, respectively); on the other hand, *S. sciuri* core genomes clustered in five different phylogenetic groups (average of 15000 SNPs difference).

Each strain of *S. fleurettii* belonged to a specific phylogenetic clade (average 3000-9000 SNPs difference) and we found no correlation between the distribution of *mecA* alleles and phylogeny. The susceptible strains CH22 and CH28 belonged to specific phylogenetic clades that differed from the remaining strains 2830-4454 SNPs (Supplementary Figure 2A).

Regarding *S. vitulinus*, we found that *mecA2*-carrying strains were slightly more related to each other (average 4010 SNPs) than to genetic backgrounds that carried *mecA* in the native locus (average 6098 SNPs). There was no correlation between the distribution of *mecA2* alleles*, mecA* alleles carried in the native locus and SCC*mec* and the genetic background of the strains. In addition, the strains carrying SCC*mec* seemed unrelated to each other, suggesting that SCC*mec* was acquired independently by these strains. Moreover, we observed that β-lactam resistant strains belonged to specific phylogenetic clades, unrelated with the remaining strains and with each other (Supplementary Figure 2B).

In our analysis of *S. sciuri* genetic backgrounds, we included type strains, representative of the three subspecies of *S. sciuri*: K1 (*S. sciuri sciuri*), K3 (*S. sciuri*

rodentius) and K11 (*S. sciuri carnaticus*) that belonged to three distinct groups, that we considered to be the different subspecies. Based on the SNP difference, we found that 18 strains belonged to *S. sciuri sciuri* subspecies, 14 strains to *S. sciuri rodentius* subspecies and 11 strains to *S. sciuri carnaticus* subspecies. The remaining strains were considered to belong to two new *S. sciuri* subgroups (three strains in group 1 and 30 strains in group 2), putatively new subspecies (Supplementary Figure 2C).

Moreover, we found that, overall, each *mecA1* allele was specific of each clade and that genetic diversity within each clade was different. In particular, we discovered that the group in which the genetic diversity of *mecA1* was higher was the *S. sciuri* new subspecies group 2 (SID=90.6%; CI:83.8-97.3%), followed by the isolates belonging to *S. sciuri sciuri* group (SID=83.9%; CI:71.2-96.6%). The remaining phylogenetic subgroups had much lower genetic diversities (SID), including 68% in *S. sciuri rodentius* (CI:52.7-83.4%), 57.2% in *S. sciuri carnaticus* (CI:39-75.5%)) and 19.3 % in the ancestral *S. sciuri* subgroup 1 (CI:0.3-38.3%). Interestingly, the phylogenetic groups in which the SID of mecA1 alleles was higher were also the ones that clustered a higher number of isolates collected from humans (*S. sciuri sciuri*, 89%; *S. sciuri* new subgroup 2, 77%).

Regarding the distribution of *S. sciuri* β-lactam resistant isolates, we found that isolates in which resistance was associated to the presence of *mecA* within SCC*mec* were distributed along all phylogenetic clades, with the exception of *S. sciuri carnaticus* clade (Table 1). Moreover, the great majority of isolates were collected from humans (Table 1). In addition, we found that isolates that carried *mecA1* alleles that were able to confer β-lactam resistance most often clustered in a phylogenetic branch belonging to the *S. sciuri rodentius* phylogenetic clade. These isolates were very related with each other; their core genomes had less than 100 SNPs difference. Interestingly, all these isolates carried alterations in the promoter of *mecA1* (Table 1).

Overall, these results suggest that resistant isolates of the phylogenetic *sciuri* group belonged to specific phylogenetic clades. We observed that in *S. vitulinus* and *S. sciuri*, each phylogenetic clade appears to have developed a different mechanism of β-lactam resistance (Table 1). In particular, we found that specific *S. vitulinus* phylogenetic clades acquired SCC*mec*, accumulated differences in the genetic

background or in the promoter of *mecA2*. Specifically, in *S. sciuri*, acquisition of SCC*mec* as well as alterations in the structure of PBP4 were mostly associated with strains belonging to the *S. sciuri* subspecies group 2, and alterations in the promoter of *mecA1* was exclusively observed among *S. sciuri rodentius* strains. While the few *S. vitulinus* strains that were β-lactam resistance were isolated from livestock, we found that most of the β-lactam resistant *S. sciuri* strains were isolated from humans.

DISCUSSION

The mechanism of resistance to antibiotics mediated by *mecA* in *Staphylococcus* is one of the most efficient mechanisms of resistance to antibiotics in bacteria, providing resistance to all members of the large class of β-lactams. Several studies have showed that the *mecA* precursor was a native gene (*mecA1*) not providing resistance in *Staphylococcus sciuri*, the most primitive staphylococcal species (172, 176). However, the evolutionary steps leading to β-lactam resistance have remained unclear. In this study, we showed for the first time that β-lactam resistance emerged several times and by different mechanisms during evolution of the most primitive staphylococcal species, belonging to the *sciuri* group. Moreover, we propose that antibiotic use in humans was the force driving evolution towards a resistant phenotype.

We found that the first evolutionary stages leading to β-lactam resistance occurred in *S. sciuri mecA1*. This included the diversification of *mecA1* sequence through mutation and recombination, which sporadically gave rise to non-synonymous substitutions mainly in the non-binding domain of the *mecA1*-encoded PBP4 that have impact in protein structure. Structural alterations were mainly in the position of residues that are part of the active site groove which we suggest to influence the access of the substrate to the active site Ser401. This could have impact in the activity of the protein when performing its native function in cell wall synthesis. However, our results showed that structural changes could have also been a strategy to decrease susceptibility to β-lactams, since many of structural alterations in the NB domain of the protein had a parallel alteration in the phenotypic resistance to β-lactams.

Overall, susceptibility to β-lactams, observed in the majority of *mecA1* and also in *mecA2* alleles encoded proteins, was associated to a NB domain wherein the residues Ser596 and Thr598 were located outside a β-sheet giving rise to a non-distorted active site that allows acylation by the β-lactam ring. On the other hand, resistance to β-lactams was associated to *mecA1* alleles encoded proteins that contain a distorted active site wherein, like in PBP2a, the access of the substrate to active site is hindered. Although the TP domain has been described to be the crucial domain for

PBPs activity (108), our results point towards a fundamental role of the NB domain for its full performance. The diversification of the NB domain was mainly observed in *S. sciuri* and more rarely observed in *S. fleurettii* and *S. vitulinus*.

Interestingly, we found that besides diversification of *mecA1*, the most primitive group of staphylococcal species appear to have used alternative mechanisms for the generation of β-lactam resistance. In particular, we observed that alterations in the promoter of *mecA* homologues were also frequently associated to changes in phenotypic susceptibility to β-lactams among this group of staphylococcal species. The association of changes in the promoter with an increased *mecA1* expression and a resultant resistant phenotype was a phenomenon previously observed in a few strains of *S. sciuri* (177). In this study we confirmed that these type of events probably occurred in a relatively high frequency in the overall *S. sciuri* population and also in other species of the *S. sciuri* group, like in *S. vitulinus* and *S. fleurettii*. Alterations included mainly deletions around the RBS site and alterations in -10 and -35 regions. This type of alterations occurred more frequently in specific *S. sciuri* subspecies like *S. sciuri rodentius*, suggesting that this subspecies might have been subjected to higher or different antibiotic pressures than other subspecies or that this species is more adapted to express resistance.

Actually, our results also illustrated that like in *S. aureus* (207, 210, 216) another factor that seems to be important for the expression of β-lactam resistance in this primitive group of staphylococcal species is the genetic background. We have found *S. vitulinus* (*mecA* allele 4) and *S. fleurettii* (*mecA* allele 2) isolates carrying the same *mecA* allele that showed different levels of β-lactam resistance. The effect of the genetic background is also well evidenced when the phenotypic resistance to β-lactams was compared in strains of *S. vitulinus* and *S. fleurettii* isolates carrying *mecA* in the native location in which *mecA* regulators are also present: while in *S. vitulinus,* isolates were almost all fully susceptible to β-lactams, *S. fleurettii* were almost all resistant. The low-level of β-lactam resistance when compared to *S. fleurettii* is probably due to the fact that *S. vitulinus* genetic background is not fully adapted to express resistance. Genes involved in general metabolism were previously described to be important for the full expression of resistance in *S. aureus*, suggesting an interplay

between the overall metabolism and β-lactam resistance (216). The same phenomenon probably occurs in *S. fleurettii* and *S. vitulinus*, wherein the metabolism appears to be more favourable to the full expression of resistance in *S. fleurettii*.

The greatest degree of genetic diversity in *S. sciuri mecA1* was observed in isolates of human origin. This result seems to imply that adaptation to humans might have been a driving force to the generation of diversity in the *mecA1* sequence and to the emergence of resistance to β-lactams. The exact human factors that have contributed to *mecA1* diversification are not known. However, we can speculate that the host jump from animals to humans might be itself the event prompting diversity in *mecA1*, as a means of survival. Actually, extensive diversification was not observed in *mecA* homologues from *S. fleurettii* or *S. vitulinus*, which contrarily to *S. sciuri*, were never described in human infection or colonization (170). On the other hand, antibiotic pressure, specifically of penicillin, might have also had a role in *mecA1* diversification, since according to our Bayesian analysis, the time the diversification of *mecA1* begun (1947-1957) coincides with the beginning of the massive use of penicillin in hospitals (106).

Overall our data suggests that the first evolutionary steps leading to β-lactam resistance in *Staphylococcus* occurred in the most primitive staphylococcal species occurred when they jumped into the human host and were subjected to β-lactam pressure. The development of resistance involved different strategies that evolved in a concerted way which included structural diversification of a native PBP, changes in the *mecA* homologues promoter and adaptation of the genetic background of the bacteria. These results highlight the importance of new human colonizing bacterial species as reservoirs of antibiotic resistance genes and emphasize the diverse resources available to bacteria to adapt to new environmental conditions.

ACKNOWLEDGMENTS

The authors acknowledge Dr. Kloos for the unrestricted gift of the great majority of the S. *sciuri* isolates that were used in this study. In addition, we thank Dr. Ivo Sedlácek, Dr. Oto Melter and Dr. José Melo-Cristino for providing some of the S. *sciuri* strains that are part of this study and Dr. Engeline van Duijkeren from the National Institute for Public Health and the Environment (RIVM, Denmark) for providing strain S. fleurettii 402567. We also thank Dr. Isabel Couto for the initial characterization of S. *sciuri* strains. J. Rolo was supported by fellowship SFRH / BD / 72675 / 2010 from Fundação para a Ciência e Tecnologia (FCT), Portugal and grant FEMS-RG-2014-0078 from the Federation of European Microbiology Societies (FEMS). This study was supported by projects PTDC/BIA-EVF/117507/2010, from FCT, and European Society of Clinical Microbiology and Infectious Diseases (ESCMID) Research Grants 2010, awarded to M. Miragaia. Partial support for this study was also provided by a US Public Health Service Award R01-AI457838-15, awarded to A. Tomasz.

SUPPLEMENTARY DATA

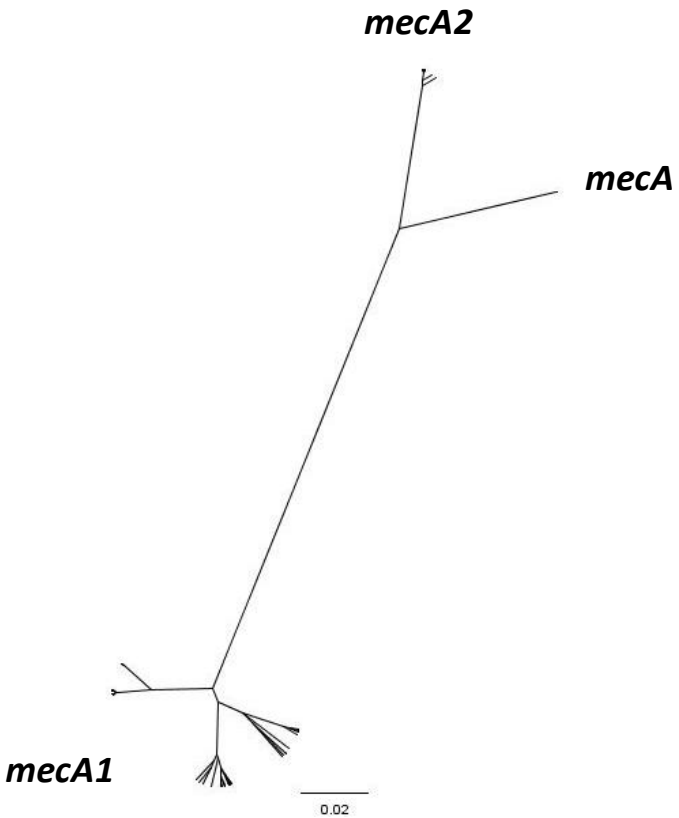
Supplementary Table 1. Epidemiological information of all strains studied. The distribution of the different *mecA* homologue alleles in the population of isolates studied is also shown. Phylogenetic group (*S. sciuri sciuri*, *S. sciuri rodentius*, S, *sciuri carnaticus*, *S. sciuri* new subspecies group 1, *S. sciuri* new subspecies group 2, *S. vitulinus, S. fleurettii*), origin of the strains and oxacillin MIC are also shown. β-lactam resistant strains are highlighted in bold. NB: non-binding domain: TP: transpeptidase domain.

Strain	Phylogenetic group	*mec* homologues allele	NB allele	TP allele	MIC (µg/ml)	Year of isolation	Origin	Geographic location
JUG1	*S. sciuri* new 1	*mecA1* 16/*mecA* 7	*mecA1* 11/ *mecA* 2	*mecA1* 30/ *mecA* 2	>256	2002	Domestic dog	Czech Republic
M1886	*S. sciuri* new 1	*mecA1* 43/*mecA* 7	*mecA1* 18/ *mecA* 14	*mecA1* 7/ *mecA* 2	64	2011	Human	Denmark
K22	*S. sciuri* new 1	*mecA1* 44	19	13	1.5	1992	Morgan horse	USA
CH17	*S. sciuri* new 2	*mecA1* 7/*mecA* 8	*mecA1* 14/ *mecA* 2	*mecA1* 32/ *mecA* 3	>256	2004	Horse	Switzerland
CH18	*S. sciuri* new 2	*mecA1* 7/*mecA* 8	*mecA1* 14/ *mecA* 2	*mecA1* 32/ *mecA* 3	>256	2005	Horse	Switzerland
D573	*S. sciuri* new 2	*mecA1* 6/*mecA* 9	*mecA1* 14/ *mecA* 4	*mecA1* 31/ *mecA* 2	>256	2007	Human	Denmark
HSM805	*S. sciuri* new 2	*mecA1* 7	14	32	0.75	2010	Human	Portugal
HSM851	*S. sciuri* new 2	*mecA1* 4/*mecA* 7	*mecA1* 12/ *mecA* 2	*mecA1* 24/ *mecA* 2	16	2010	Human	Portugal
JUG17	*S. sciuri* new 2	*mecA1* 4	12	24	>256	2002	Human	Yugoslavia
JUG2	*S. sciuri* new 2	*mecA1* 15	10	29	0.75	2002	Domestic dog	Yoguslavia
K132	*S. sciuri* new 2	*mecA1* 5	13	24	0.5	1976	Howler monkey	Panama
K141	*S. sciuri* new 2	*mecA1* 8	15	24	1	1992	California mouse	USA
K148	*S. sciuri* new 2	*mecA1* 3	14	25	0.38	1992	Bottlenose dolphin	USA
K2	*S. sciuri* new 2	*mecA1* 8	15	24	1	1992	Beef tongue	USA
K20	*S. sciuri* new 2	*mecA1* 13	15	24	1	1992	Bottlenose dolphin	USA
K24	*S. sciuri* new 2	*mecA1* 2	14	24	0.75	1992	Morgan horse	USA
K51	*S. sciuri* new 2	*mecA1* 8	15	24	1.5	1971	Human	USA
K61	*S. sciuri* new 2	*mecA1* 14	15	28	1.5	1992	Pilot whale	USA
K69	*S. sciuri* new 2	*mecA1* 9	16	24	1	1986	Human	USA
KLO59	*S. sciuri* new 2	*mecA1* 10	15	24	2	1972	Domestic dog	USA
M1234	*S. sciuri* new 2	*mecA1* 6/*mecA* 9	*mecA1* 14/ *mecA* 4	*mecA1* 31/ *mecA* 2	>256	2009	Human	Denmark
M1653	*S. sciuri* new 2	*mecA1* 6/*mecA* 9	*mecA1* 14/ *mecA* 4	*mecA1* 31/ *mecA* 2	>256	2010	Human	Denmark
M2276	*S. sciuri* new 2	*mecA1* 6/*mecA* 9	*mecA1* 14/ *mecA* 4	*mecA1* 31/ *mecA* 2	>256	2011	Human	Denmark
M2590	*S. sciuri* new 2	*mecA1* 6/*mecA* 9	*mecA1* 14/ *mecA* 4	*mecA1* 31/ *mecA* 2	>256	2012	Human	Denmark
M2710	*S. sciuri* new 2	*mecA1* 4/*mecA* 7	*mecA1* 12/ *mecA* 2	*mecA1* 24/ *mecA* 2	>256	2012	Human	Denmark
M692	*S. sciuri* new 2	*mecA1* 6/*mecA* 9	*mecA1* 14/ *mecA* 4	*mecA1* 31/ *mecA* 2	96	2007	Human	Denmark

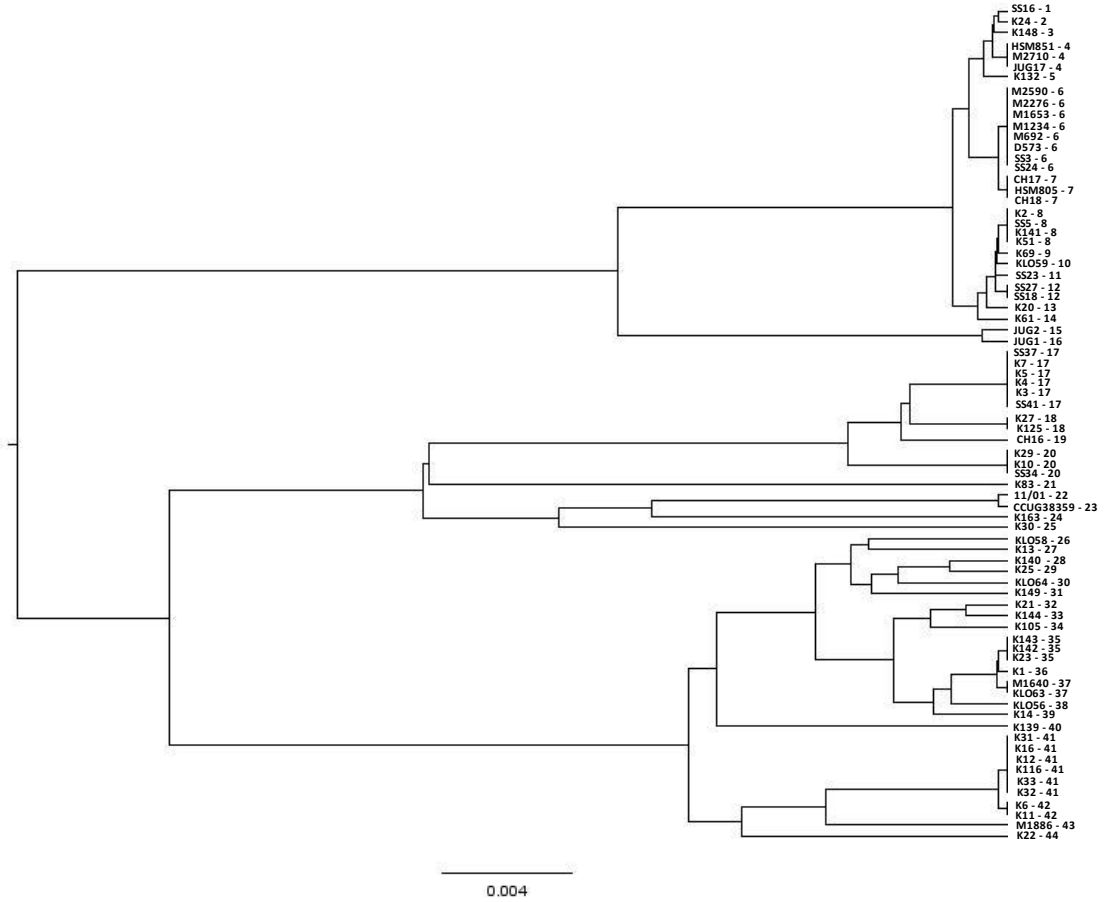
Strain	Phylogenetic group	*mec* homologues allele	NB allele	TP allele	MIC (µg/ml)	Year of isolation	Origin	Geographic location
SS16	*S. sciuri* new 2	*mecA1* 1	14	24	0.75	1996	Human	Portugal
SS18	*S. sciuri* new 2	*mecA1* 12	15	26	0.75	1996	Human	Portugal
SS23	*S. sciuri* new 2	*mecA1* 11	15	27	1	1997	Human	Portugal
SS24	*S. sciuri* new 2	*mecA1* 6	14	31	0.75	1997	Human	Portugal
SS27	*S. sciuri* new 2	*mecA1* 12	15	26	0.75	1998	Human	Portugal
SS3	*S. sciuri* new 2	*mecA1* 6	14	31	1	1996	Human	Portugal
SS5	*S. sciuri* new 2	*mecA1* 8	15	24	1	1996	Human	Portugal
11/01	*S. sciuri carnaticus*	*mecA1* 22	3	21	0.5	2002	Human	Czech Republic
CCUG38359	*S. sciuri carnaticus*	*mecA1* 23	2	21	0.75	2002	Human	Sweden
K11	*S. sciuri carnaticus*	*mecA1* 42	17	15	0.75	1990	Veal leg	USA
K116	*S. sciuri carnaticus*	*mecA1* 41	17	15	0.75	1992	Beef lips	USA
K12	*S. sciuri carnaticus*	*mecA1* 41	17	15	1	1992	Arabian horse	USA
K16	*S. sciuri carnaticus*	*mecA1* 41	17	15	0.75	1992	Jersey calf	USA
K163	*S. sciuri carnaticus*	*mecA1* 24	4	23	0.75	1992	Holstein cow	USA
K30	*S. sciuri carnaticus*	*mecA1* 25	5	22	1.5	1992	Jersey cattle heifer	USA
K31	*S. sciuri carnaticus*	*mecA1* 41	17	15	0.5	1992	Jersey cattle calf	USA
K32	*S. sciuri carnaticus*	*mecA1* 41	17	15	0.75	1992	Jersey cattle calf	USA
K33	*S. sciuri carnaticus*	*mecA1* 41	17	15	0.75	1992	Jersey cattle heifer	USA
CH16	*S. sciuri rodentius*	*mecA1* 19/*mecA* 7	*mecA1* 7/ *mecA* 2	*mecA1* 17/ *mecA* 2	24	2004	Horse	Switzerland
K10	*S. sciuri rodentius*	*mecA1* 20	6	19	0.75	1992	European red squirrel	USA
K125	*S. sciuri rodentius*	*mecA1* 18	8	16	1	1992	Cotton rat	USA
K27	*S. sciuri rodentius*	*mecA1* 18	8	16	0.75	1992	Norway rat	USA
K29	*S. sciuri rodentius*	*mecA1* 20	6	19	0.75	1992	Norway rat	USA
K3	*S. sciuri rodentius*	*mecA1* 17/*mecA* 10	*mecA1* 6/ *mecA* 3	*mecA1* 18/ *mecA* 1	>256	1992	Neonatal ward	Mozambique
K4	*S. sciuri rodentius*	*mecA1* 17	6	18	>256	1992	Human	Mozambique
K5	*S. sciuri rodentius*	*mecA1* 17	6	18	25	1992	Human	Mozambique
K6	*S. sciuri rodentius*	*mecA1* 42/*mecA* 7	*mecA1* 17/ *mecA* 2	*mecA1* 15/ *mecA* 2	>256	1992	Human	Mozambique
K7	*S. sciuri rodentius*	*mecA1* 17	6	18	>256	1992	Human	Mozambique
K83	*S. sciuri rodentius*	*mecA1* 21	9	20	0.75	1992	Human	Czech Republic
SS34	*S. sciuri rodentius*	*mecA1* 20	6	19	0.75	1996	Human	Portugal
SS37	*S. sciuri rodentius*	*mecA1* 17	6	18	25	1996	Human	Portugal
SS41	*S. sciuri rodentius*	*mecA1* 17	6	18	3	1996	Human	Portugal
K1	*S. sciuri sciuri*	*mecA1* 36	25	7	2	1972	Eastern grey squirrel	USA
K105	*S. sciuri sciuri*	*mecA1* 34	20	12	1	1971	Human	USA
K13	*S. sciuri sciuri*	*mecA1* 27	20	5	0.5	1992	Eastern grey squirrel	USA
K139	*S. sciuri sciuri*	*mecA1* 40	21	10	1	1992	Holstein cow	USA
K14	*S. sciuri sciuri*	*mecA1* 39	20	11	0.75	1992	Eastern harvest mouse	USA
K140	*S. sciuri sciuri*	*mecA1* 28	20	4	1	1972	Opossum	USA

Strain	Phylogenetic group	*mec* homologues allele	NB allele	TP allele	MIC (µg/ml)	Year of isolation	Origin	Geographic location
K142	*S. sciuri sciuri*	*mecA1* 35	24	7	1	1992	Horse	USA
K143	*S. sciuri sciuri*	*mecA1* 35	24	7	1	1972	Racoon	USA
K144	*S. sciuri sciuri*	*mecA1* 33	20	9	1	1992	Jersey calf	USA
K149	*S. sciuri sciuri*	*mecA1* 31	1	2	0.75	1972	Eastern grey squirrel	USA
K21	*S. sciuri sciuri*	*mecA1* 32	22	9	0.5	1993	Pilot whale	USA
K23	*S. sciuri sciuri*	*mecA1* 35	24	7	0.75	1992	Red kangaroo	USA
K25	*S. sciuri sciuri*	*mecA1* 29	23	3	0.5	1992	Prairie vole	USA
KLO56	*S. sciuri sciuri*	*mecA1* 38	20	8	0.75	1972	Opossum	USA
KLO58	*S. sciuri sciuri*	*mecA1* 26	20	6	0.75	1972	Squirrel monkey	USA?
KLO63	*S. sciuri sciuri*	*mecA1* 37	24	7	1	1972	Eastern grey squirrel	USA
KLO64	*S. sciuri sciuri*	*mecA1* 30	20	1	2	1972	Southern flying squirrel	USA
M1640	*S. sciuri sciuri*	*mecA1* 37/*mecA* 7	*mecA1* 24/ *mecA* 2	*mecA1* 7/ *mecA* 2	96	2010	Human	Denmark
H39	*S. vitulinus*	*mecA* 4	5	4	0.75	2005	Horse	Denmark
H91	*S. vitulinus*	*mecA* 7	2	2	2	2005	Horse	Denmark
401946	*S. vitulinus*	*mecA* 4	5	4	2	2004	Horse	The Netherlands
CH1	*S. vitulinus*	*mecA* 7	2	2	1	2005	Horse	Switzerland
CH2	*S. vitulinus*	*mecA* 4	5	4	4	2004	Horse	Switzerland
CH3	*S. vitulinus*	*mecA* 7	2	2	0.75	2005	Horse	Switzerland
CH4	*S. vitulinus*	*mecA* 4	5	4	0.75	2004	Horse	Switzerland
CH5	*S. vitulinus*	*mecA* 4	5	4	>256	2005	Horse	Switzerland
CH6	*S. vitulinus*	*mecA* 4	5	4	1.5	2004	Horse	Switzerland
CH7	*S. vitulinus*	*mecA2* 1	1	1	0.75	2004	Horse	Switzerland
CH8	*S. vitulinus*	*mecA2* 2	1	2	0.19	2005	Horse	Switzerland
CH9	*S. vitulinus*	*mecA2* 2	1	2	0.5	2004	Horse	Switzerland
CH10	*S. vitulinus*	*mecA2* 2	1	2	0.5	2004	Horse	Switzerland
CH11	*S. vitulinus*	*mecA2* 1	1	1	0.5	2005	Horse	Switzerland
CH12	*S. vitulinus*	*mecA2* 2	1	2	0.5	2004	Horse	Switzerland
CH13	*S. vitulinus*	*mecA2* 2	1	2	0.1	2004	Horse	Switzerland
CH14	*S. vitulinus*	*mecA2* 2	1	2	0.5	2005	Horse	Switzerland
CH15	*S. vitulinus*	*mecA2* 2	1	2	>256	2004	Horse	Switzerland
402567	*S. fleurettii*	*mecA* 5	6	4	>256	2004	Horse	The Netherlands
CH19	*S. fleurettii*	*mecA* 1	1	6	8	2009	Horse	Switzerland
CH20	*S. fleurettii*	*mecA* 3	5	5	6	2009	Horse	Switzerland
CH21	*S. fleurettii*	*mecA* 2	5	4	4	2010	Horse	Switzerland
CH22	*S. fleurettii*	*mecA* 2	5	4	1	2010	Horse	Switzerland
CH23	*S. fleurettii*	*mecA* 2	5	4	4	2010	Horse	Switzerland
CH24	*S. fleurettii*	*mecA* 2	5	4	>256	2009	Horse	Switzerland
CH25	*S. fleurettii*	*mecA* 6	5	2	4	2010	Horse	Switzerland

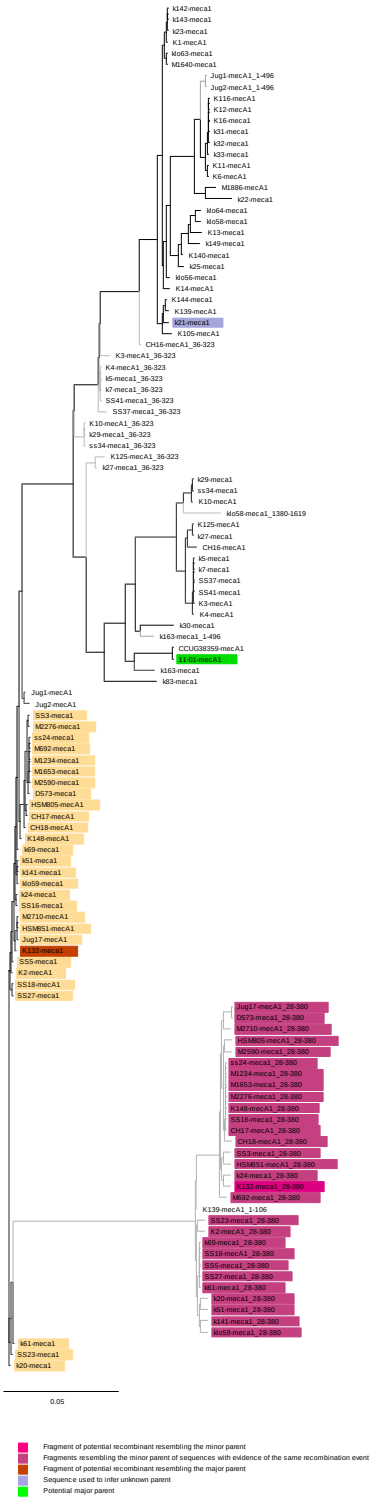
Strain	Phylogenetic group	*mec* homologues allele	NB allele	TP allele	MIC (µg/ml)	Year of isolation	Origin	Geographic location
CH26	*S. fleurettii*	*mecA* 2	5	4	>256	2010	Horse	Switzerland
CH27	*S. fleurettii*	*mecA* 2	4	4	4	2010	Horse	Switzerland
CH28	*S. fleurettii*	*mecA* 2	5	4	1	2010	Horse	Switzerland
CH29	*S. fleurettii*	*mecA* 2	5	4	4	2010	Horse	Switzerland



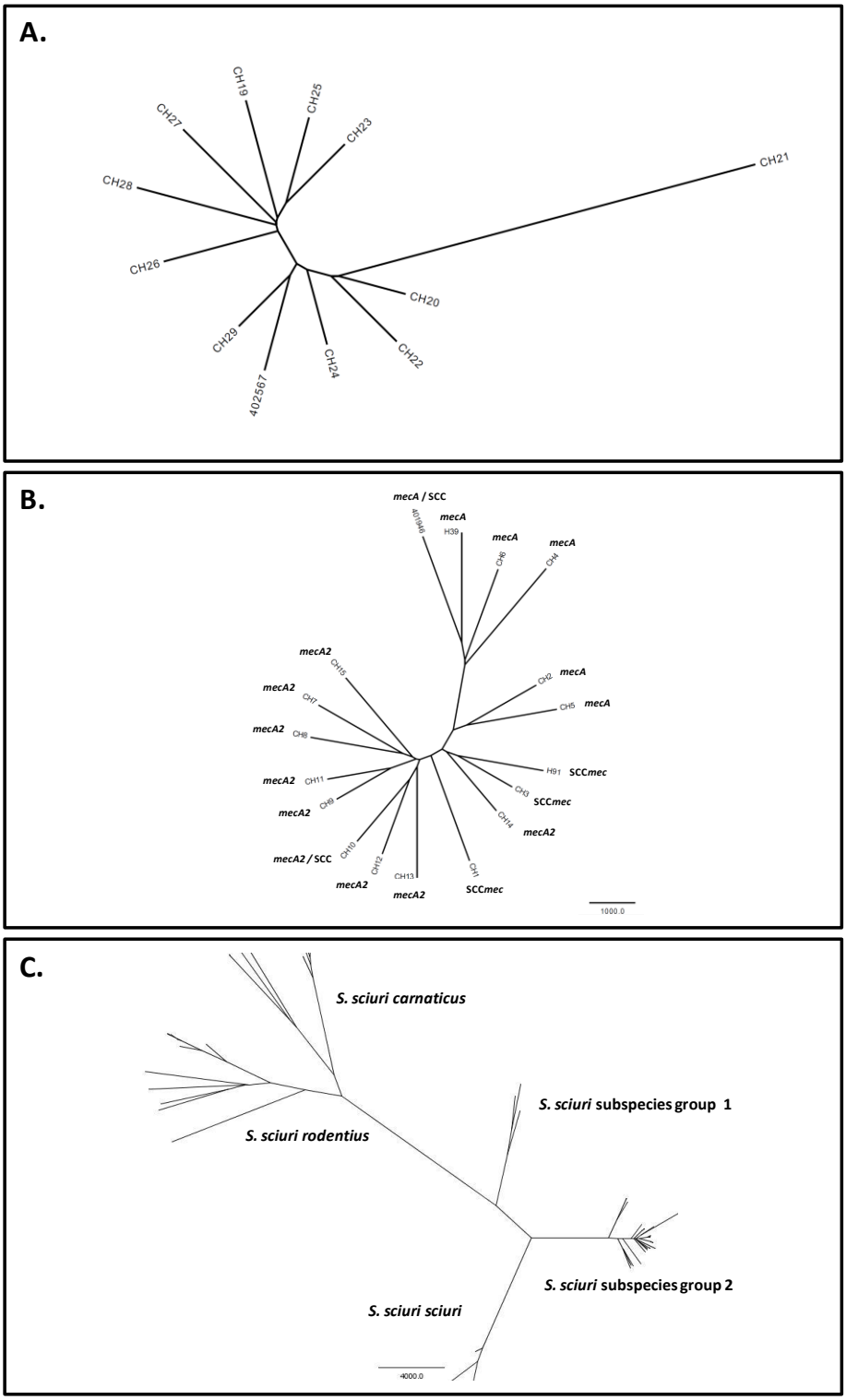
Supplementary Figure 1A. Phylogenetic analysis of the *mecA* homologues nucleotide sequence. The sequences of each *mec* homologue gene was extracted from the de novo assembly contigs and aligned with ClustalW. The tree was performed with UPGMA method, under the Jukes-Cantor substitution model, with a bootstrap of 100 replicates. The unrooted tree is shown.



Supplementary Figure 1B. Phylogenetic analysis of the *mecA1* alleles. nucleotide sequence. The nucleotide sequence of each *mecA1* allele was extracted from the de novo assembly contigs and aligned with ClustalW. The tree was performed with UPGMA method, under the Jukes-Cantor substitution model, with a bootstrap of 100 replicates. The strain ID, followed by the *mecA1* allele number is shown.



Supplementary Figure 1C. Identification of recombination events among *mecA1* alleles. The recombinant parts of *mecA1* alleles are clustered apart from the remaining portion of the allele. In addition, a colour code is applied to identity the putative major parents that were involved in the recombination events.



Supplementary Figure 2. Phylogenetic analysis of the core genome of isolates belonging to the *sciuri* group. Unrooted phylogenetic tree based on the number of SNP differences found among the predicted core genome of the strains. The reference genome used was *S. fleurettii* 402567. **A.** *S. fleurettii.* **B.** *S. vitulinus.* **C.** *S. sciuri.*

Evolutionary origin of the staphylococcal cassette chromosome *mec* (SCC*mec*)

Joana Rolo, Peder Worning, Jesper Boye Nielsen, Rory Bowden, Ons Bouchami, Peter Damborg, Luca Guardabassi ,Vincent Perreten, Alexander Tomasz, Henrik Westh, Hermínia de Lencastre , Maria Miragaia

Manuscript in preparation

All the experimental work and the whole-genome sequencing analysis – with the exception of SNP analysis – was performed by J. Rolo

ABSTRACT

Several lines of evidence indicate that the most primitive staphylococcal species, the *Staphylococcus sciuri* group, were involved in the first stages of evolution of SCC*mec* – the genetic element carrying the β-lactam resistance gene *mecA.* However, many steps are still missing from this evolutionary history. In particular, it is not known how *mecA* was incorporated into the mobile element SCC prior to dissemination among *Staphylococcus aureus* and other pathogenic staphylococcal species.

To gain insights into the possible contribution of several species of the *Staphylococcus sciuri* group to the assembly of SCC*mec*, we sequenced the genomes of 106 isolates, comprising *S. sciuri* (n=76), *Staphylococcus vitulinus* (n=18) and *Staphylococcus fleurettii* (n=12) from animal and human sources, and characterized the native location of *mecA* and the SCC insertion site using a variety of comparative genomic approaches. Moreover, we performed a SNP analysis of the genomes, in order to understand SCC*mec* evolution in relation to phylogeny.

We found that each of three species of the *S. sciuri* group contributed to the evolution of SCC*mec*: *S. vitulinus* and *S. fleurettii* to the assembly of the *mec* complex, and *S. sciuri* most likely provided the mobile element in which *mecA* was later incorporated. We hypothesize that an ancestral SCC*mec* III cassette (an element carried by one of the most epidemic methicillin-resistant *S. aureus* clones), originated in *S. sciuri* possibly by a recombination event in a human host or a human-created environment and later was transferred to *S. aureus*.

IMPORTANCE

Pathogenicity islands like SCC*mec* can alter in critical ways the invasive potential of bacteria. However, steps in the assembly of such mosaic structures are not well understood. In this study we aimed to identify key evolutionary events in the construction of SCC*mec*, an element that had a huge impact on the virulence and resistance potential of *Staphylococcus aureus*. The study described here provides

evidence that the assembly of SCC*mec* occurred in parallel with the evolution of phylogeny and involved several different species within the genera. Our data suggest that primitive prokaryotic species frequent in wild animals may represent gene pools for the construction of pathogenicity islands that have major impact on human health if they are transferred into bacteria capable of colonizing humans.

INTRODUCTION

The most important mechanism of resistance to β-lactam antibiotics is associated with the *mecA* gene which encodes an extra penicillin-binding protein, called PBP2a, that has low affinity to virtually all β-lactam antibiotics (217). The *mecA* gene is carried on a structurally complex mobile genetic element called staphylococcal cassette chromosome *mec* (SCC*mec*) that can also transport determinants of resistance to other antimicrobials, virulence determinants and other genes important for bacterial survival in stress conditions (109). Thus in a single event of genetic acquisition, SCC*mec* can turn susceptible staphylococci into virulent multidrug-resistant pathogens, well adapted to thrive in an infection situation, particularly in the hospital environment. The advantage for bacteria to carry this element is indicated by the wide distribution of SCC*mec* among both nosocomial and community staphylococcal populations (106). Dissemination of SCC*mec* among *Staphylococcus aureus* strains constitutes a real public health threat worldwide (218) due to the associated complications in treatment and disease outcome.

SCC*mec* has a modular structure: it is composed of two essential elements, the *mec* complex – composed of *mecA* and its regulators (*mecR1* and *mecI*) – and the *ccr* complex –containing cassette chromosome recombinase (*ccr*) genes that assure the mobility of the cassette (109, 111). SCC*mec* may also carry insertion sequences (IS), transposons and plasmids, as well as housekeeping genes inside the so-called joining regions (J1-J3). SCC*mec* inserts at a specific site on the chromosome, downstream of *orfX* (109) (recently re-named *rlmH*), which encodes an *rRNA* methyltransferase (115). The insertion of SCC*mec* at *orfX* creates direct (DR) or inverted (IR)repeats, which form the boundaries of the element (109).

Several lines of evidence suggest that SCC*mec* evolution occurred in the most primitive group of *Staphylococcus* species, the *Staphylococcus sciuri* group. Previous studies showed that *mecA1*, a ubiquitous gene in *S. sciuri* with 80% nucleotide identity to *S. aureus mecA*, was the most probable evolutionary precursor of *mecA* (172, 174). Follow-up studies showed that, *mecA1* was able to express resistance and to participate in the cell-wall biosynthesis pathway, just like PBP2a (176, 189), when

introduced into a *S. aureus* genetic background. Moreover, other homologues closer to *mecA* than *mecA1,* along evolutionary lines were also identified in other species of the *S. sciuri* group, namely *mecA2* (90% identity) in *S. vitulinus* and *mecA* (99% identity) in *S. fleurettii*, suggesting a vertical evolution of the gene along phylogeny (173, 178). In contrast to *S. sciuri* and *S. vitulinus*, *mecA* was flanked by the regulators *mecR1* and *mecI* in *S. fleurettii*, leading to the hypothesis that the *mec* complex was first assembled in this species (173). Among the other species of the *S. sciuri* group, *Staphylococcus lentus* has been the least studied, and no information at all is available regarding the distribution of SCC*mec* elements in *Staphylococcus stepanoviccii.*

A recent study investigating the distribution of *ccr* genes among *S. sciuri* showed that the frequency of *ccr* in methicillin-susceptible *S. sciuri* was much higher (35%) than that described for other coagulase-negative staphylococci (CoNS) (219). Moreover, the most frequent *ccr* allotypes corresponded to homologues of *ccrA3B3*, although homologues of all other *ccr* allotypes were also identified (219). This suggests that *S. sciuri* – besides being the original source of the *mecA* determinant – may also have been the donor of the *ccr* complex for the assembly of SCC*mec*.

Studies on SCC*mec* evolution performed so far, suggest an important role of the *S. sciuri* group but those studies were base only on a limited number of isolates and provided scattered evidence on the evolution, diversification and assembly of SCC*mec*. In this study we provide missing links in the evolution of SCC*mec* through the study of a large and diverse collection of isolates belonging to the *S. sciuri* group using whole genome sequencing.

METHODS

Bacterial collection. A collection of 106 staphylococcal isolates, comprising 76 *S. sciuri*, 18 *S. vitulinus* and 12 *S. fleurettii* was assembled. In *S. sciuri*, 29 isolates were obtained from humans, and the remaining 47 were recovered from wild and domestic mammals (Supplementary Table S1). The isolates originated from nine different countries (Czech Republic, Denmark, Portugal, Switzerland, Sweden, former Yugoslavia, Mozambique, Panama and USA) during the period 1972-2012. *S. vitulinus* and *S. fleurettii* isolates were collected from horses and bovine mastitis milk samples, in Denmark, Switzerland and the Netherlands, in 2004, 2005 and 2010. The *S. sciuri* isolates were identified at the species level by 16S RNA ribotyping and API-Staph (Biomerieux, France). *S. fleurettii* and *S. vitulinus* were identified at the species level by sequencing of 16S rRNA or *sodA* and Maldi-tof analysis (Microflex LT, Bruker Daltonics GmbH, Bremen) (178, 220, 221). Species identification was confirmed by phylogenetic analysis of the sequence of the *tuf* gene (222) in the sequencing data produced in this study.

DNA preparation and whole-genome sequencing: *S. sciuri* DNA samples were prepared using a phenol-chlorophorm extraction protocol. *S. fleurettii* and *S. vitulinus* DNA was prepared with DNeasy Blood & Tissue Kit (Qiagen, Limburg, The Netherlands). Sequencing libraries were prepared by sonic fragmentation and adapter ligation and then sequenced on the Illumina HiSeq 2000/2500 platform, producing paired 100 bp reads. The reads were assembled *de novo* using VELVET (33). In addition, strain *S. fleurettii* 402567 was also sequenced using a PacBio RS apparatus and *de novo* assembly was performed using HGAP 3 (https://github.com/PacificBiosciences/Bioinformatics-Training/wiki/HGAP-in-SMRT-Analysis).

Assessment of genetic relatedness between isolates: PacBio reads were combined with Illumina reads obtained for *S. fleurettii* isolate 402567 in CLC Genomics Workbench (Qiagen, Limburg, The Netherlands), using the Genome Finishing module.

Resulting contigs were ordered using *Staphylococcus xylosus* which is the species most closely related to *S. fleurettii* that had a closed genome (NCBI accession number CP007208.1; average nucleotide identity with *S. sciuri*, 78%; *S. vitulinus*, 77.1%; and *S. fleurettii*, 78.5%). Gaps were closed by mapping Illumina data of the remaining *S. fleurettii* strains to the contigs. The resulting closed genome was annotated with RAST (http://rast.nmpdr.org/). In brief, we found that the *S. fleurettii* 402567 genome was 2.58 Mbp in length, with 31.7% G+C content and 2498 coding sequences, comprising 1931 putative genes and 567 pseudogenes. The *S. fleurettii* 402567 closed genome was used as a reference to identify single nucleotide polymorphisms (SNPs) in the draft genomes of each subset of strains belonging to a single species, including *S. fleurettii*, *S. vitulinus* and *S. sciuri*, as well as within the entire collection of samples. Reads were mapped to the reference using Stampy (version 1.0.11) and variants were called using SAMtools (version 0.1.12). Phylogenies were reconstructed using neighbor-joining and drawn in FigTree (http://tree.bio.ed.ac.uk/software/figtree/).

BLAST and phylogenetic analysis. The contigs obtained in the VELVET *de novo* assembly were used to assess the presence of several genes of interest, which were detected using BLAST with a 70% nucleotide identity threshold. Sequences of genes found within SCC*mec* III or published gene sequences of *S. sciuri*, *S. fleurettii* and *S. vitulinus* from NCBI (http://www.ncbi.nlm.nih.gov/) and (173) were used as reference. The nucleotide sequences of specific genes identified by BLAST analysis were extracted from the sequence of the contigs and then aligned using ClustalW. Phylogenetic trees were constructed using neighbor-joining. In addition, the phylogeny of SCC and SCC*mec* elements was produced by aligning the sequence of these elements with Mauve (34).

Manual annotation. The entire contig containing *orfX* was annotated for a group of randomly chosen strains (17 *S. sciuri*, 4 *S. fleurettii* and 5 *S. vitulinus* isolates). The *orfX* was identified by BLAST analysis as described above. The remaining ORFs of the contig were identified using GeneMark-hmm (http://exon.gatech.edu/GeneMark/) and

annotated with BLAST using the NCBI non-redundant database, with a minimum match threshold of 70% across at least 30% of the gene; *ccr* types were assigned as previously suggested (111).

Comparative genomic analysis. The contigs corresponding to SCC and/or SCC*mec* cassettes containing regions of high identity in nucleotide sequence (>80%) and content were compared using WEBACT (http://www.webact.org/WebACT/home).

Statistical analysis. Statistical significance of differences between proportions was evaluated by the Chi-square (χ2) test using a confidence interval of 95%.

Data deposition. The reference genome obtained was deposited in NCBI database and raw reads in ENA.

Rational for the development of the SCC*mec* assembly model. For the construction of the model we considered that the origin of one gene or block of genes would be the species in which the most ancestral form of the gene would be found, the synteny is more conserved and the frequency and genetic diversity is the highest. On the other hand, the order of occurrence of events was recognized by the level of homology with that of genes present in SCC*mec* III as well as by the presence of intermediate forms of a gene or specific structures in a single phylogenetic defined group, like a species or sub-species, which is also observed in their descendants and not observed in their ancestors. Also, when a similar structure would be found in different phylogenetic defined groups this would suggest the occurrence of horizontal gene transfer (HGT) between them.

RESULTS AND DISCUSSION

Studies performed so far, suggest an important role of the *S. sciuri* group in SCC*mec* assembly, but when, where and how the different assembly steps occurred is still elusive. In this study we aim to demonstrate the role of different species within the *S. sciuri* group to the construction of the cassette. In particular we were interested in understanding how the *mecA* homologue present in *S. sciuri* evolved to be part of *mec* complex, how the mobile genetic elements SCC were created and how was *mec* complex incorporated into these elements and became mobile. To address these questions we sequenced the genome of 106 isolates belonging to *S. sciuri*, *S. vitulinus* and *S. fleurettii*, all species belonging to the *S. sciuri* group, collected in different time periods, geographic regions and from both humans and animals and analyzed the *orfX* region and the region in the vicinity of *mecA* homologues.

To identify proofs to the origin and subsequent evolutionary steps of the different pieces within the SCC*mec* mosaic structure we looked for the distribution of structural elements of SCC*mec* type III, the most frequent in species of the *S. sciuri* group, using BLAST analysis. Also, we compared the synteny of these structural elements with those of *S. aureus* SCC*mec* III and examined their relatedness using phylogenetic analysis. Finally, to identify blocks of homology between SCC elements in different species we compared the complete SCC elements using alignment tools and iterative phylogenetic analysis.

Evolution of *mecA* and flanking genes in the native location

The location of the native *mecA* homologues was identified in all isolates by confirming the presence of genes previously shown to be in the vicinity of *mecA1* in *S. sciuri*, *mecA2* in *S. vitulinus* and *mecA* in *S. fleurettii* (173). We found that the native location of *mecA* homologues was approximately 200 kb downstream of *orfX* in the three species analyzed. The location of this PBP gene was the closest to *orfX*, when compared to other native PBP genes in *S. sciuri* and other staphylococcal genomes (http://www.ncbi.nlm.nih.gov/)(25, 27, 52, 126). This location near *orfX* and *oriC* may have been important for integration of the *mec* complex into the SCC elements by

recombination. On the other hand, the integration into a site near *oriC* may have increased the *mecA* gene dosage, providing a potential selective advantage via increased resistance to β-lactams, as previously observed when a strong *mecA1* promotor was formed in *S. sciuri* through the insertion of IS*256* (177).

We also observed that the *mecA* homologues were always flanked by the same genes in the native location, as previously described (173). This included *psm-mec*, a gene implicated in *S. aureus* virulence, and located immediately upstream of *mecA* homologues (137) and *ugpQ*, which is located downstream of *mecA* homologues. Interestingly, these genes are exactly the same as those that are flanking *mecA* in contemporary SCC*mec* types carried by MRSA (138). However, the level of homology of *psm-mec* and *ugpQ* from MRSA with those from the *S. sciuri* group, varied according to the species analyzed. The most similar *psm-mec* and *upgQ* to those of MRSA were found in the vicinity of *S. fleurettii* and *S. vitulinus mecA* (98.55% and 99.87-100%, respectively), followed by those found in the vicinity of *S. vitulinus mecA2* (97%; 86.56%) and in the vicinity of *S. sciuri mecA1* (91-94%; 75-85%) (Table 1).

Although most strains contained at least one copy of a *mecA* homologue in the native location, their identities and frequencies differed in the three species analyzed. The *mecA1* was ubiquitous in *S. sciuri* and *mecA* in *S. fleurettii*. Although ubiquity of *mecA1* was previously described in *S. sciuri* (172), the ubiquity of *mecA* in *S. fleurettii* is in contradiction with previous studies, where a much smaller collection of isolates was analyzed (173). In contrast, in *S. vitulinus* the type of *mecA* homologue found in the native location varied from strain to strain. Half (9/18) of the *S. vitulinus* strains carried *mecA2* (178), but in six *S. vitulinus* strains we identified a *mecA* homologue with 99% identity to that of the *S. aureus mecA*. The remaining three *S. vitulinus* strains had no *mecA* homologue at the native location (Table 1), but instead carried *mecA* near the *orfX*, that was 100% identical to *mecA* of MRSA.

In addition to *mecA1*, we found that 16 out of the total 76 *S. sciuri* strains also carried MRSA *mecA* in the *orfX* region. Moreover, the *mecA* homologue most similar to that of MRSA strains was the *mecA* from *S. sciuri* and *S. vitulinus* located at the *orfX* (100% nucleotide identity), followed by *S. fleurettii* and *S. vitulinus mecA* (99%

nucleotide identity), *S. vitulinus mecA2* (94%) and *S. sciuri mecA1* (80%), all located at the native location.

When *psm-mec*, *upgQ* and *mecA* homologues from the three species were compared to MRSA in a phylogenetic tree (where *S. sciuri* is the most divergent and *S. fleuretti* the most similar to MRSA), the hierarchy observed was comparable to that obtained for other housekeeping genes located far away from the native location, and was similar to the phylogenetic hierarchy obtained when housekeeping genes from the different species were used to construct a phylogenetic tree (data not shown). Therefore, the comparison of *mecA* homologues and its neighboring genes to the ones carried by SCC*mec* reflected the phylogeny of the species in an evolutionary perspective (Table 1).

Altogether, these results suggest that the first steps in the evolution of the β-lactam resistance determinant were the evolution of *S. sciuri mecA1* and its neighboring genes into *mecA2* and neighbouring gene alleles in *S. vitulinus* followed by evolution into *mecA* and neighbouring genes in *S. fleurettii*, which are almost identical to those in MRSA. The finding of *S. vitulinus* strains with similar *mecA* and adjacent genes with the ones carried by *S. fleurettii* suggest that after vertical evolution, *mecA* was probably acquired by *S. vitulinus* from *S. fleurettii* by HGT.

Table 1. Frequency (%) of the genes homologous to genes carried by SCC*mec* elements found in *S. sciuri*, *S. vitulinus* and *S. fleuretti* not containing SCC*mec* at the *orfX*. The nucleotide sequence identity found between these homologues and the ones carried by contemporary SCC*mec* III is also showed.

	S. sciuri		***S. vitulinus***		***S. fleurettii***	
SCC*mec* element	**Frequency (%)**	**Homology (%)**	**Frequency (%)**	**Homology (%)**	**Frequency (%)**	**Homology (%)**
J3 region						
***ugpQ* (J3)**	100	75-85	100	87-100	100	99-100
***maoC* (J3)**	63	100	100	100	100	100
***pre* (J3)**	5	64	60	65	17	63
***polypeptide B* (J3)**	2	63	20	99	17	99
mec complex						
IS431	47	99	80	99	92	99
IS1272	3	63	40	66	25	63
mecA1	100	80	-	-	-	-
mecA2	-	-	50	90	-	-
mecA	-	-	40	99-100	100	97-99
mecR1/mecI	-	-	40	99/92-99	100	99/92
mecR2	23	78	100	99	100	99
J2 region						
***psm-mec* (J2)**	100	91-94	100	97-99	100	98-99
Rhodanese-domain containing protein (J2)	48	77	40	100	100	99
Metallo-βlactamase family protein (J2)	47	78	40	99	100	99
***ccr* complex**						
ccr	38	72-100	13	72-77	8	83
multiple *ccr*	10	-	-	-	-	-
J1 region						
Hypothetical proteins (J1)	38	90-100	11	90-100	17	90-100

The assembly of *mec* complex occurred in a step-wise manner along phylogeny

We found that the *mecA* regulators (*mecI* and *mecR1*) in their native locations were always associated with *mecA* in both *S. fleurettii* and *S. vitulinus*, but were never found near to *mecA1* and *mecA2*. Whereas the native *mecR1* gene carried by *S. fleurettii* and *S. vitulinus* was 99.6% identical to the *mecR1* in SCC*mec*, the nearby *mecI* gene had a lower identity (92.5%), attributable to a 24 nucleotide difference and alternate STOP codon that lengthened the ORF from 345 to 372 bp. We suggest that the ancestral *mecI* gene was longer in length and a deletion of 24 nucleotides occurred during evolution, giving rise to the contemporary *mecI* gene that is found in all SCC*mec* types containing *mec* complex A (111).

We also searched the native location for the presence of the recently described regulatory gene *mecR2*, which encodes an anti-repressor of *mecA*, and is also part of *mec* complex A (135). All *S. vitulinus* and *S. fleurettii* strains carried *mecR2*, 99% identical to MRSA *mecR2*, near to the *mecA* homologue in the native location (*S. vitulinus*, *mecA*/*mecA2*; *S. fleurettii*, *mecA*). In contrast, in *S. sciuri*, we found that only 14 isolates out of 76 (18.42%) carried a *mecR2* homologue in the vicinity of *mecA1* (14 isolates) in the native location. These 14 strains were epidemiologically unrelated, since they were collected from different hosts (including wild animals, production animals and humans), different geographic origins and time periods. The *S. sciuri mecR2* homologue had only 77.56% of nucleotide sequence identity with MRSA *mecR2*. This gene may correspond to the ancestral *mecR2* and may be involved in the regulation of *mecA1* expression in *S. sciuri*.

Another element, which is part of the *mec* complex and is usually located downstream *mecA*, is the IS*431* element. We observed that this element was absent from the close vicinity of *mecA* homologues in the native locus (data not shown). In contrast, IS*431* was found in the close vicinity of *mecA* when it was located in the *orfX* region within a SCC*mec* element, in *S. sciuri* and *S. vitulinus*. In addition, IS*431* was found with high frequency in all genomes, particularly in *S. fleurettii* genomes (Table 1), but not in the vicinity of *mecA* (data not shown).

These observations suggest that the addition of regulators to a basic *mecA* gene cluster may have occurred in a step-wise manner: *mecR2* was the first gene to be added in *S. sciuri* at the native location near *mecA1*. Although only present in a small fraction of the current *S. sciuri* population this gene arrangement seems to have been maintained during evolution of the species and to have become ubiquitous in *S. fleurettii* and *S. vitulinus*. Addition of *mecR1* and *mecI* only occurred later, after the evolution of the ancestral *mecA1* into *mecA* was complete. The IS*431* element was probably only added to *mecA* and regulators, after it was mobilized into SCC*mec* or during mobilization, since it was not observed at the native location.

The *ccr* complex is common in *S. sciuri* and is most similar to *ccrA3B3* of SCC*mec* III

We found that the frequency of *ccr* genes was much higher in *S. sciuri* (55%) than in *S. vitulinus* (28%) and *S. fleurettii* (8%) (Table 1). Furthermore, while only *ccrA* and *ccrB* were identified in *S. fleurettii* and *S. vitulinus*, a more diverse pool of *ccr* genes was observed in *S. sciuri*, including *ccrC*, 3%, and a new *ccr* type (75% nucleotide identity to *ccrC*), 14%, in addition to *ccrA* and *ccrB*. Moreover, *S. sciuri* was the only species where multiple *ccr* genes were detected in some isolates (10%).

The great majority (70%) of the *ccrA* and *ccrB* allotypes identified in *S. sciuri, S. fleurettii* and *S. vitulinus* were similar to *ccrA3*, *ccrA5*, *ccrB3* and *ccrB5* (83-100% nucleotide identity). These *ccr* allotypes are highly related to *ccrA3B3* carried by SCC*mec* III (*ccrA3*, 85% nucleotide identity with *ccrA5*; *ccrB3*, 83% nucleotide identity to *ccrB5*), as previously observed (219). The types of *ccr* complex were diverse and were represented by different combinations. The most frequent *ccr* complex types observed were *ccrA3B5* (27%), *ccrA5B5* (24%) and *ccrA5B3* (13%). The *ccrA3B5* was observed in *S. sciuri* and *S. vitulinus*, while *ccr* types *ccrA5B5* and *ccrA5B3* were observed exclusively in *S. sciuri*. In addition, *ccrA1B1*, *ccrA1B3* and *ccrA1B5* were exclusively observed in *S. sciuri* and with low frequency (2% each). The remaining *ccr* types were new, corresponding to combinations of the new *ccrA* and *ccrB* allotypes identified.

In *S. sciuri* we found the most ancestral alleles of *ccrA* and *ccrB* (Supplementary Table S2), as well as the combination of *ccrA* and *ccrB* genes that were the most closely related to *ccrA3B3* that has disseminated into SCC*mec* types carried by MRSA. In addition, the fact that the frequency and diversity of *ccr* genes and complexes is much higher in *S. sciuri* than in *S. vitulinus* and *S. fleurettii* suggests that *S. sciuri* was the most probable donor of the *ccr* complex to form a primordial SCC*mec*.

On the other hand, the high frequency of *ccr* genes, which have a high degree of similarity with the *ccrA3B3* complex in all species of the *S. sciuri* group supports the hypothesis that SCC elements were transferred within the different species of the *S. sciuri* group and that SCC*mec* III could have been the first cassette to be assembled.

The SCC*mec* J1 region originated in *S. sciuri* and the J2/J3 regions in *S. fleuretti* and *S. vitulinus*

Besides the two essential elements – the *mec* complex and the *ccr* complex – SCC*mec* also contains three joining regions (J1-J3), the organization of which within the element may be represented as *orfX*-J3-*mec*-J2-*ccr*-J1-DR.

The J1 region of SCC*mec* III (linking *ccr* and DR) is composed of six different genes encoding hypothetical proteins. We found that at least one, but often two or more, of these genes was present in higher frequency in *S. sciuri* (38%), than in *S. vitulinus* (11%) and *S. fleurettii* (17%). The similarity of nucleotide sequence of the genes carried by strains belonging to the *sciuri* group and the genes carried by SCC*mec* III was high (90-100%). Particularly in *S. sciuri*, two, three or even six of the genes were found, in the same orientation and synteny as in SCC*mec* III of MRSA. Annotation of the vicinity of these genes in all species showed that they were always located in the *orfX* region, and some of them were located inside SCC elements (Supplementary Table S2). These observations suggest that *S. sciuri* was the most likely donor of the J1 region for the assembly of a primordial SCC*mec*.

The SCC*mec* III J2 genes (linking *mec* and *ccr*) were in general more frequent among *S. fleurettii* and *S. vitulinus* (carrying *mecA*) and less frequent in *S. sciuri*. In

particular, genes encoding a rhodanese-domain-containing protein and a protein of the metallo-β-lactamase family, located in the distant vicinity of *mec* complex in SCC*mec*, were found in all three species analyzed. In *S. fleurettii*, they were ubiquitous (100%), and in *S. vitulinus* they were only present in isolates carrying the *mec* complex (40%) (Table 1). In these species, these genes were located in the vicinity of *mec* complex, in the native location (*S. fleurettii* and six *S. vitulinus* strains) or in the *orfX* region (three *S. vitulinus* strains). However, in *S. sciuri*, these elements were present either in the vicinity of *mecA1* in the native location or in the vicinity of the *mec* complex near *orfX* (Table 1). When in the native location in the three species, these genes were upstream of *psm-mec*, in the same orientation and synteny as the ones carried by SCC*mec*. In addition, the nucleotide identity of the genes found in *S. sciuri* and the ones carried in SCC*mec* was lower, when compared to genes carried by *S. fleuretttii* and *S. vitulinus* (Table 1).

Finally, we observed that genes found in J3, namely genes encoding hypothetical proteins (*pre* and *polypeptide B*, located upstream of the hypervariable region) were more frequent in *S. fleurettii* (*pre*, 17% and *polypeptide B*, 17%) and *S. vitulinus* (*pre*, 60% and *polypeptide B*, 20%) but rare in *S. sciuri* (5-6% and 2%, respectively) (Table 1). We observed that both genes were always present in the same orientation and synteny as the ones carried by SCC*mec*. However, it was not possible to assess their position relative to the position of *mecA* homologues, since *mecA*, *pre* and *polypeptide B* genes were always in different contigs in strains carrying these genes. Similarly *maoC*, a gene that is part of the hypervariable region located in the J3 region was located upstream of the *mecA* homologue in the native location in these three species, with different frequencies *(S. fleurettii* and *S. vitulinus*, 100%; *S. sciuri*, 63%).

These results suggest that the SCC*mec* J3 region probably originated in *S. vitulinus* and *S. fleuretttii.* However, the genes encoding the rhodanese domain-containing protein and the metallo-β-lactamase family protein of the J2 region, which is located in the vicinity of *mec* complex, were probably already present in *S. sciuri*, and evolved in a manner consistent with the phylogeny and were then mobilized together with the *mec* complex and the more distant J3 regions to form an SCC element.

S. sciuri SCC non-_mec_ elements carry housekeeping genes and are similar to MRSA SCC_mec_

SCC structures, bounded by DR or IR, were observed in all three species. However, the frequency and diversity of such structures was higher in *S. sciuri* than in the other two species. A total of eight different structures were found in the 26 isolates analyzed. In Supplementary Table S2 there is a detailed description of the variety of SCC non-*mec* elements found in each of the species.

SCC elements carried by *S. sciuri* had in common the presence of genes that confer resistance to metals and, most unexpectedly, genes that are usually associated with housekeeping functions in different staphylococcal species. In *S. sciuri* strains 11/01 and CCUG38359, a SCC non-*mec* element of 54 Kb ($SCC_{11/01}$) carrying *ccrA5B5* was identified. This SCC was composed of several elements associated with SCC*mec*, such as a type I restriction/modification system; a cadmium-resistance operon; IS*431*; and several genes encoding various hypothetical proteins (Supplementary Table S2). The nucleotide sequence of these genes were most similar (40-100%) to hypothetical proteins present in SCC*mec* III, SCC*mec* IVa, SCC*mec* IX and SCC*mec* XI, and non-typable SCC*mec* cassettes carried by *S. haemolyticus*, *S. xylosus* and *S. pseudintermedius*. Moreover, this element contained ORFs that were similar to genes found in the same region, in *ccr*-negative strains. In particular, in the *ccr*-negative strain K22, genes encoding these exact same proteins (100% nucleotide sequence identity) were present 14 kb downstream *orfX* (Supplementary Table S3). Furthermore, immediately upstream of the DR, this SCC contained genes encoding proteins that are present elsewhere in the core genome of other staphylococcal species. These genes participate in bacterial central metabolism (encoding glycosyltransferases, oxidases, alcohol dehydrogenases, photolyases, general stress proteins).

The remaining SCC non-*mec* elements identified in *S. sciuri* were diverse. Strain K116 carried a 36 kb SCC (SCC_{K116}) that contained a *ccr* gene with 75% nucleotide identity with *ccrC.* Similarly to $SCC_{11/01}$, this SCC element carried the *cad* operon and staphylococcal housekeeping genes. Some of these housekeeping genes were similar to those found inside $SCC_{11/}$01 (e.g. genes encoding oxidases, dehydrogenases,

photolyases) but some were different (genes encoding epimerases). Strain SS27 carried a composite SCC element, formed by two SCCs: a 45 kb SCC and a 15 kb SCC (both carrying *ccrA1B3*, Supplementary Table S2) (SCC-CI_{SS27}). The larger SCC element (SCC_{SS27I}) carried a plasmid with an intact *rep* gene as well as genes conferring resistance to copper, cadmium and arsenic. The smaller SCC carried a quinone reductase, as well as the *cad* operon and housekeeping genes, similarly to those found in SCC_{K116} (Supplementary Table S2). These SCC elements also carried genes found inside SCC*mec* cassettes, such as genes encoding hypothetical proteins (61-98% nucleotide identity) present in several SCC*mec* cassettes (SCC*mec* II, III, IV, V, IX, X and XI). In addition, SCC $_{SS27I}$ also carried a rhodanese-domain containing protein and a metallo-β-lactamase family protein typical of the J2 region of SCC*mec* III (Supplementary Table S2).

In *S. vitulinus* and *S. fleurettii*, the SCC elements found were different from the ones identified in *S. sciuri*. They contained new types of *ccr* and mainly hypothetical proteins, some of which had homology with ORFs that are present in SCC*mec* from MRSA. However, in contrast to what was described for *S. sciuri*, housekeeping genes were not found inside *S. vitulinus* or *S. fleurettii* SCC elements. *S. vitulinus* 401946 carried a 20 kb SCC (SCC_{401946}), with a new *ccr* type (*ccr* type 6, Supplementary Table S2). This SCC also carried genes encoding hypothetical proteins identified in SCC*mec* types V and IX, as well as a type I restriction -modification system. In addition, we also found predicted ORFs that had no significant hits in the BLAST server. *S. vitulinus* CH10 carried an 11 kb SCC (SCC_{CH10}), which contained a new *ccr* type (*ccr* type 7, Supplementary Table S2) and genes encoding hypothetical proteins identified in SCC*mec* types IX and XI, as well as predicted ORFs with no significant hits in the BLAST server. *S. fleurettii* 402567 carried a 18 kb SCC composed of genes encoding hypothetical proteins. Besides the *ccr* genes, which corresponded to new types, the predicted ORFs showed no significant hits in the BLAST server.

Overall, our data showed that SCC structures carried by *S. sciuri* are more similar to the ones seen in contemporary SCC*mec* types, considering both nucleotide sequence identity and synteny. In addition, only *S. sciuri* SCC elements carried housekeeping genes. The frequent inclusion of housekeeping genes inside SCC

elements illustrates the level of genomic re-arrangement that occurs in the *orfX* region of *S. sciuri*.

orfX region of *S. sciuri ccr*-negative strains contains genes that are inside SCC non-*mec* and SCC*mec*

To further explore the possibility that SCC elements may be assembled from housekeeping genes located near *orfX* – we characterized the *orfX* region of 12 strains not carrying any SCC element or *ccr* complex (seven *S. sciuri*, three *S. fleurettii*, and two *S. vitulinus*). We found that this region was different in strains belonging to different species.

Downstream *orfX*, *ccr*-negative *S. sciuri* strains contained mostly genes associated with central metabolism; nevertheless, this region was very diverse with respect to gene content among the strains. This result might indicate that this region of the chromosome frequently undergoes recombination and that genes are being shuffled between strains, due to a high frequency of insertion/excision events. The genes found were not always present or there were insertions of other genes. Examples of the genes carried by these strains are listed in Supplementary Table S3 (genes encoding oxidoreductases, serine proteases, alcohol dehydrogenases, epimerases, isomerases, methyltransferases). A comparison of genes found in this region in *ccr*-negative strains (Supplementary Table S3) and genes found inside SCC and SCC*mec* carried by *S. sciuri* (Supplementary Table S2) revealed that some of these genes (with more than 90% nucleotide sequence identity) are found among both structures, in the same order and synteny: genes encoding glycosyltransferases, epimerases, dehydrogenases and oxidases. In addition, in *ccr*-negative strains, genes encoding resistance to cadmium and arsenic were also identified; and these genes were also identified among SCC structures carried by *S. sciuri*.

In some *ccr*-negative *S. sciuri* strains, structural elements typically found inside SCC*mec* could also be identified despite the absence of *ccr* genes in this region of the chromosome. Strains K10, K31 and KLO63 contained genes encoding hypothetical proteins that have been associated with the J1 region of SCC*mec* types IV, V and IX. In

addition, strain K31 carried a truncated DNA restriction modification system (*hsdRhsdM*) that has been found exclusively in SCC*mec* V.

The fact that housekeeping genes found in *ccr*-negative strains are similar in sequence and synteny to those found in SCC and SCC*mec* elements in *S. sciuri* suggests that a primordial SCC*mec* could have been assembled in this species. Moreover, these observations also suggest that the location of genes near *orfX* was important for their inclusion in a SCC element. Given the fact that the native *mec* location is only 200 Kb apart from *orfX*, it is tempting to speculate that a similar type of phenomenon may explain the insertion of *mec* complex into the SCC element. However, we cannot discard the alternative hypothesis, namely that genes found in *ccr*-negative strains near the *orfX* are remnants of an SCC that was left in this region after imprecise excision of SCC elements.

The *orfX* region of *ccr*-negative *S. vitulinus* strains was also diverse, but not as much as in *S. sciuri* strains. Immediately downstream *orfX*, three ORFs of unknown function were identified among all *S. vitulinus* strains. In addition, genes encoding proteins that probably participate in central metabolism were also identified, such as ATPases, DNA methylases, glycosyltransferases and acetyltransferases. These genes were not identified in all *S. vitulinus* strains. In contrast, the *orfX* region of *S. fleurettii* *ccr*-negative strains was found to be highly conserved, being mainly composed of genes encoding proteins involved in the central metabolism: synthetases and dehydrogenases. In contrast to the results obtained for *S. sciuri*, the SCC non-*mec* elements identified among *S. vitulinus* and *S. fleurettii* strains did not carry any of the genes identified in the *orfX* region of *ccr*-negative strains.

SCC*mec* elements carried by *S. sciuri* are similar but not identical to SCC*mec* III of MRSA

SCC*mec* was identified in *S. sciuri* (16 isolates) and *S. vitulinus* (three isolates), but was absent from all *S. fleurettii* isolates. A total of four different SCC*mec* structures were found. Three of these structures were related to SCC*mec* III ("SCC*mec* III-like

structures A, B and C") and one to SCC*mec* type VII – based on the sequence of the genes, their position and structure.

A comparison of the sequence of SCC*mec* III-like elements found in the *S. sciuri* group to the SCC*mec* III of MRSA with WEBACT (www.webact.org) has revealed that the J2 region, the *mec* complex and the *ccr* complex (approximately 25 kb) were almost exactly the same between the structures in the *S. sciuri* group and the SCC*mec* III of MRSA. The structure more closely related to SCC*mec* III and also the most frequent one was "SCC*mec* III-like A" (Figure 1). Structure A was found in seven *S. sciuri* isolates (K3, M1640, M1886, JUG1, CH16, CH17 and CH18) and three *S. vitulinus* isolates (H91, CH1 and CH3). This structure (63 kb) was very similar to SCC*mec* III, mainly in the J2 region, containing the same genes encoding hypothetical proteins, the Tn*554* transposase and the cadmium resistance operon, besides the *mec* complex A. The J3 region was very small, encompassing only the HVR region located upstream the *mec* complex A, which contrasts with SCC*mec* III J3 region that carries in addition the plasmid pT181 that contains genes conferring resistance to tetracycline (138). The J1 region was the most divergent between "SCC*mec* III-like A" structure and SCC*mec* III. Although some genes encoding hypothetical proteins within this region were common between the two elements, "SCC*mec* III-like A" contained a more "ancient" *ccr* complex (219) (*ccrA3B5* instead of *ccrA3B3*), and also four additional ORFs identified as a glycosyltransferase, a peptidase, a transcriptional regulator and a surface protein, which were absent from MRSA SCC*mec* III.

When we compared "SCC*mec* III-like A" structures with SCC elements identified in our collection we found several regions of homology between the two. In particular, in case of genes that were part of the J1 region of the "SCC*mec* III-like A" element, such as the gene encoding a glycosyltransferase, the same genes were also identified in the J1 region of $SCC_{11/01}$ (Figure 1). Another example is the *ccrA3B5* that is common to "SCC*mec* III–like A" and several *S. sciuri* SCC elements. In Figure 2 there is a tree that resulted from iterative phylogenetic analysis where the level of relatedness between SCC*mec* and SCC elements can be deduced (34) (Figure 2). While the "SCC*mec* III-like A" structure appears to correspond to a primordial form of contemporary SCC*mec* III, the "SCC*mec* VII-like" and the "SCC*mec* III-like B" structures seemed more related to

contemporary SCC*mec* types. Interestingly, according to this analysis, "SCC*mec* III-like A and C" structures, that were present in 63% (12 out of 19) of all strains carrying SCC*mec*, were highly related with SCC carried by *S. sciuri*.

These results further support the hypothesis that "SCC*mec* III-like A" structures could have originated in *S. sciuri*, possibly by recombination of the *mec* complex from *S. fleurettii* or *S. vitulinus* with an SCC carried by this species. Moreover, the evolution into contemporary SCC*mec* III probably occurred afterwards and involved mainly a recombination in the J1 region and acquisition of plasmids in the J3 region.

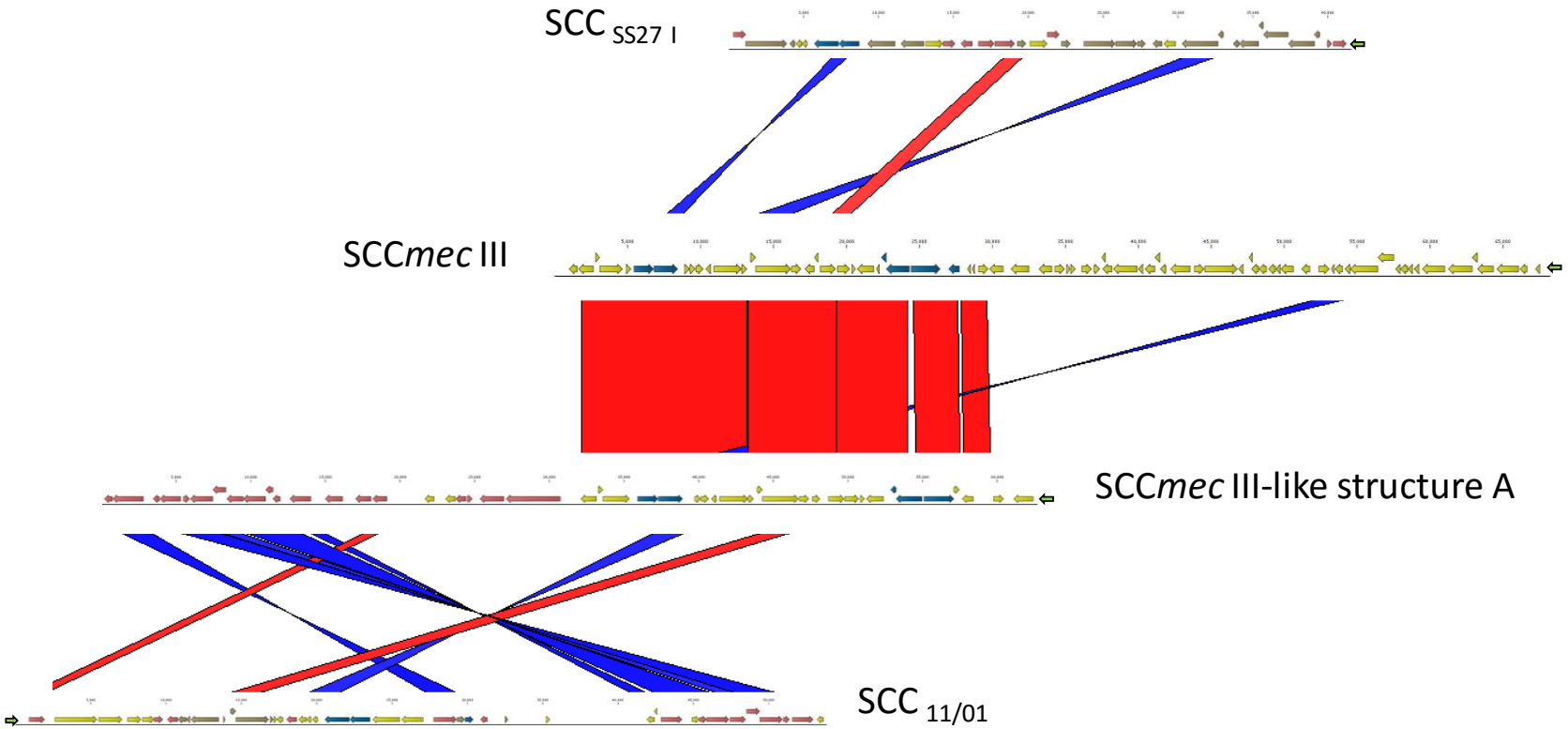


Figure 1. Comparison of regions of homology (highest BLAST score, shown in red) between the most frequent "SCC*mec* III-like structure A" found in *S. sciuri* and *S. vitulinus* with SCC*mec* III carried by MRSA, and "*SCC non-mec carried by S. sciuri"*, using the platform ACT (www.webact.org). A schematic representation of the annotated genes is shown. The core genes are shown in pink while the genes associated with SCC*mec* are shown in yellow. The *ccrAB* and *mecI/mecR1/mecA* genes are highlighted in blue. The *orfX* gene is shown in green. Additional genes which are usually found among staphylococcal mobile genetic elements are depicted in grey (encoding for instance heavy-metal resistance associated genes and plasmid sequences).

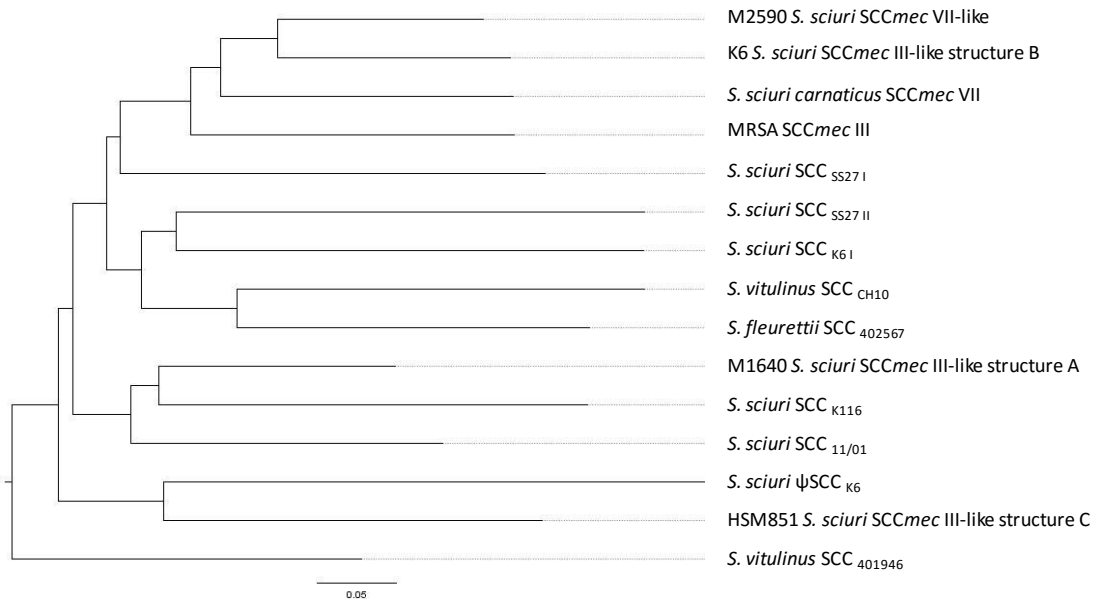


Figure 2. Reconstruction of SCC*mec* and SCC phylogeny with Mauve. *S. sciuri:* 11/01; HSM851 (partial SCC*mec* cassette), K6, K116, M1640, M2590 (partial SCC*mec* cassette), SS27. *S. vitulinus*: CH10, 401946; *S. fleurettii*: 402567.

SCC*mec* elements are identified in several *S. sciuri* genetic backgrounds

The SNP phylogenetic analysis of the genome of all isolates (Supplementary Figure 1A) showed that the *S. sciuri*, *S. vitulinus* and *S. fleurettii* were separated into three distinct clusters and had an average of 150000 SNPs difference between their genomes. On the other hand, the SNP analysis of the genome of isolates belonging to each species independently (Supplementary Figures 1B-D), showed that in *S. fleurettii*, the genomes had a range of 3000-9000 SNPs difference and in *S. vitulinus*, the genomes had a range of 4000-7000 SNPs difference. In *S. sciuri* there was a high diversity of genetic backgrounds, which is in accordance with the fact that this species is the most primitive and most widely distributed in nature. Five distinct phylogenetic groups that differed from each other by more than 15000 SNPs were defined. In this analysis we included the *S. sciuri* subspecies type strains, K1 (*S. sciuri sciuri*), K3 (*S. sciuri rodentius*) and K11 (*S. sciuri carnaticus*). The genomes of these strains clustered separately, and therefore we suggest that these clusters correspond to the three *S. sciuri* subspecies previously described (171). In Supplementary Figure 1B the *S. sciuri*

sciuri is depicted as group 2; *S. sciuri rodentius* as group 4 and *S. sciuri carnaticus* as group 5. However, we found two additional phylogenetic clusters that probably correspond to two additional *S. sciuri* subspecies (groups 1 and 3).

The *S. sciuri "*new subspecies group 3" was the most represented in our collection, including 30 isolates, followed by *S. sciuri sciuri* (n=18); *S. sciuri rodentius* (n=14), *S. sciuri carnaticus* (n=11), and *S. sciuri* "new subspecies group 1" (n=3). The *S. sciuri "*new subspecies group 3" isolates were mainly of human origin; *S. sciuri sciuri* and *S. sciuri carnaticus* isolates were mostly collected from animals (both wild and production animals); and *S. sciuri rodentius* and *S. sciuri* "new subspecies group 1" isolates were evenly collected from humans and animals. Overall, groups 1-3 were the groups in which diversity in terms of SNP differences was lower (20-7000 SNPs). In groups 4 and 5, the diversity was much higher (100-18000 SNPs).

The frequency of SCC*mec* and SCC was different among the five groups identified. Specifically, we found that the overall frequency of SCC*mec* and SCC elements was very high in the "new subspecies group 3" of *S. sciuri* and *S. sciuri rodentius* (69 and 71%, respectively), and low in *S. sciuri sciuri* (16.5%). In contrast, *S. sciuri carnaticus* carried exclusively SCC cassettes in a high frequency (45%) and two out of three of the new *S. sciuri* subspecies 1 isolates carried SCC*mec* elements (67%, 2 out of 3 isolates). Of note, the three subspecies where SCC*mec* was more frequent (groups 1 and 3, and *S. sciuri rodentius*) contained isolates most of which were collected from humans (64-77%). Conversely, the phylogenetic groups that comprised isolates with a lower frequency of SCC*mec* (*S. sciuri sciuri* and *S. sciuri carnaticus*) were mostly collected from animals (82-89%). The difference between these proportions was statistically significant ($p<0.05$).

The high proportion and distribution of SCC*mec* and SCC elements in diverse *S. sciuri* genetic backgrounds as well as the high proportion of SCC*mec* in the most primitive *S. sciuri* cluster suggest that SCC*mec* was probably assembled in *S. sciuri* a long time ago, which may have been before the use of antibiotics in clinical practice. Being *S. sciuri* widely distributed in nature, it is possible that it has been exposed to β-

lactam antibiotics produced by co-colonizing microrganisms in the natural environment, which may have been the driving force for the assembly of SCC*mec*.

In *S. vitulinus*, we did not observe the formation of very defined clusters. However, we found that the genetic backgrounds that carried *mecA2* in the native location were more closely related to each other (average 4010 SNPs) than to genetic backgrounds that carried *mecA* in the native locus (average 6098 SNPs). We propose that an ancestral *mecA2* background might have favored the replacement of the *mecA2* locus by *mec* complex A, which evolved divergently, giving rise to two distinct genetic backgrounds. Regarding the acquisition of SCC elements these appear to have occurred several times during *S. vitulinus* evolution and were independent of genetic background. Actually, the two SCCs found among *S. vitulinus* were acquired both in the genetic background of strains carrying *mecA2* and also in those carrying *mecA*. In contrast, SCC*mec* elements were preferentially acquired by strains of the genetic background carrying *mecA2*.

As to the case of *S. fleurettii*: there was a single genetic background that had a higher number of SNP difference, in comparison to the remaining genetic backgrounds (CH21, 8177 SNPs; remaining genetic backgrounds, average 4955 SNPs). In addition, this strain was collected in the same country (Switzerland) and during the same time period.

Despite a relative high number of isolates included in this study compared to previous studies about the *S. sciuri* group, the greatest limitation of this study is that the interpretation of the results is limited to the collection of isolates that we were able to put together.

CONCLUSIONS

Our results suggest that SCC*mec* was assembled long time ago, before introduction of ß-lactam antibiotics into clinical practice. The assembly occurred through several steps involving at least three species of the *S. sciuri* group (Figure 3)

Our data suggest that the first step in SCC*mec* assembly was the evolution of *mecA1* in its native location together with its flanking genes *psm-mec* and *upgQ.* The second step of evolution was the addition of *mec* regulators. The *mecR2* was first acquired by some *S. sciuri* strains and then maintained through phylogeny in *S. vitulinus* and *S. fleurettii*. The *mecA* has probably first emerged in *S. fleurettii* (BEAST analysis of *mec*A homologues, unpublished data). The addition of *mecI* and *mecR1* occurred probably in this same species through recombination between *mecA* and adjacent regions with the β-lactamase locus, usually associated with plasmids. The amino acid identities among genes present in these two loci and the similar orientation and arrangement of the genes suggests an evolutionary link between *mecA* and *blaZ* (223). In addition, it has been demonstrated that the repressor *blaI* can also regulate the expression of *mecA* (224-226). In subsequent steps of evolution the *mec* complex from *S. fleurettii* may have been then incorporated into the *S. vitulinus* native location, by recombination with the *mecA2* locus. We propose that the last donor of *mec* complex A and neighbouring regions to an assembled SCC element might have been *S. vitulinus*, since this was the only species where *mec* complex and neighbouring genes seemed to have been deleted from the native location.

The evolution of SCC elements occurred in parallel with the evolution of the native *mec* locus. The SCC elements most probably originated in *S. sciuri* and were built from housekeeping genes located in the *orfX* region. The integration of *mec* complex and neighbouring genes from *S. vitulinus* into a SCC element probably occurred in *S. sciuri*, since this is the only species where the same housekeeping genes were found both in SCC and SCC*mec* structures. However, the mechanism that mobilized *mec* complex from *S. vitulinus* to an SCC in *S. sciuri* is not known.

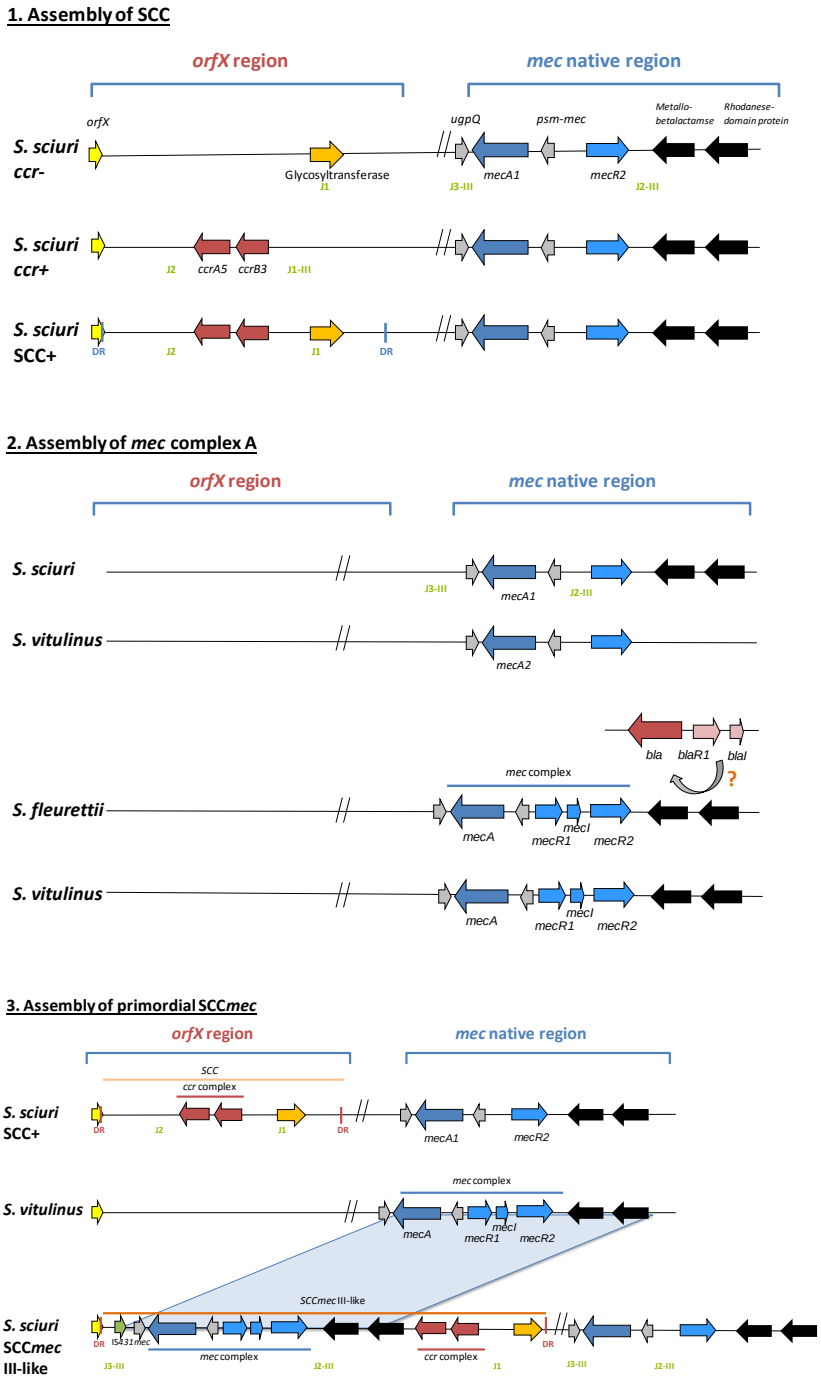


Figure 3. Proposed model for the assembly of the first SCC*mec* structure. The SCC originated in the *S. sciuri* chromosome and insertion/excision events catalyzed by the recombinases encoded by *ccr* were frequent events. On the other hand, *mecA1* and adjacent regions *psm-mec* and *ugpQ* in native locus located 200 kb from *orfX* evolved over phylogeny giving rise to *mecA2* and *mecA* and descendant *psm-mec* and *ugpQ* genes in *S. vitulinus* and *S. fleurettii*, respectively. In *S. fleurettii*, the native locus containing *mecA* might have evolved into a regulatory system like *mec* complex A through recombination with the β-lactamase regulatory locus. This locus would then be incorporated in *S. vitulinus* chromosome by recombination with the native *mecA2* locus. The *mec* complex A, which would have been donated by *S. vitulinus*, would be integrated in an assembled SCC structure carried by *S. sciuri*, giving rise to a primordial SCC*mec* structure related to SCC*mec* III identified in contemporary MRSA clones.

In *S. sciuri* an ancestral structure of SCC*mec* III that we refer to as *"*SCC*mec* III-like A", was the most frequent SCC*mec* structure found. Our results suggest that: 1) "SCC*mec* III-like A" was the first SCC*mec* cassette, that probably emerged in *S. sciuri* and 2) SCC*mec* III may have been the first contemporary SCC*mec* type that emerged. In fact, a recent study that focused on the evolution of the pandemic MRSA clone ST239-III, carrying SCC*mec* III, has suggested that this clone has emerged in the 1960s, in Europe (31). Therefore, one can speculate that SCC*mec* III was probably already circulating among the staphylococcal population around that time or even before.

Our findings highlight the role of primitive staphylococcal species in the origin of complex pathogenicity islands such as SCC*mec* that once acquired can increase in a dramatic way the pathogenic and resistance potential of human colonizing bacteria such as *S. aureus*.

ACKNOWLEDGMENTS

The authors acknowledge Isabel Couto for the initial characterization of the *S. sciuri* isolates. In addition, the authors thank Ivo Sedlácek (Masaryk University, Czech Republic), Oto Melter (Charles University in Prague and Motol University Hospital, Czech Republic) and José Melo-Cristino (Hospital de Santa Maria, Portugal) for providing some of the *S. sciuri* isolates used in this study and Engeline van Duijkeren (National Institute for Public Health and the Environment, RIVM, Denmark) for providing strain *S. fleurettii* 402567. J. Rolo was supported by fellowship SFRH / BD / 72675 / 2010 from Fundação para a Ciência e Tecnologia (FCT), Portugal and grant FEMS-RG-2014-0078 from the Federation of European Microbiology Societies (FEMS). R. Bowden was a member of the UKCRC Modernising Medical Microbiology Consortium supported by UK Medical Research Council (Grant G0800778) and the Wellcome Trust (Grant 087646/Z/ 08/Z). This study was supported by projects PTDC/BIA-EVF/117507/2010, from FCT, and European Society of Clinical Microbiology and Infectious Diseases (ESCMID) Research Grants 2010, awarded to M. Miragaia. Partial support for this study was also provided by a US Public Health Service Award R01-AI457838-15, awarded to A. Tomasz. DNA sequencing was supported in part by core funding to the Wellcome Trust Centre for Human Genetics, provided by the Wellcome Trust (090532/Z/09/Z).

SUPPLEMENTARY DATA

Supplementary Table S1. Epidemiological data of all isolates characterized in this study by whole-genome sequencing. Phylogenetic group 1: ancestral *S. sciuri* background; phylogenetic group 2: S. *sciuri sciuri*; phylogenetic group 3: *S. sciuri* new subspecies; phylogenetic group 4: *S. sciuri rodentius*; phylogenetic group 5: *S. sciuri carnaticum.*

Strain	Species (*S. sciuri* phylogenetic group)	MIC (µg/ml)	Year of isolation	Origin	Geographic location	Source or Reference
402567	*S. fleurettii*	>256	2004	Horse	The Netherlands	E. van Duijkeren
CH19	*S. fleurettii*	8	2010	Bovine mastitis milk	Switzerland	(221)
CH20	*S. fleurettii*	6	2010	Bovine mastitis milk	Switzerland	(221)
CH21	*S. fleurettii*	4	2010	Bovine mastitis milk	Switzerland	(221)
CH22	*S. fleurettii*	1	2010	Bovine mastitis milk	Switzerland	(221)
CH23	*S. fleurettii*	4	2010	Bovine mastitis milk	Switzerland	(221)
CH24	*S. fleurettii*	>256	2010	Bovine mastitis milk	Switzerland	(221)
CH25	*S. fleurettii*	4	2010	Bovine mastitis milk	Switzerland	(221)
CH26	*S. fleurettii*	>256	2010	Bovine mastitis milk	Switzerland	(221)
CH27	*S. fleurettii*	2	2010	Bovine mastitis milk	Switzerland	(221)
CH28	*S. fleurettii*	1	2010	Bovine mastitis milk	Switzerland	(221)
CH29	*S. fleurettii*	4	2010	Bovine mastitis milk	Switzerland	(221)
JUG1	*S. sciuri* (1)	>256	2002	Domestic dog	Czech Republic	I. Sedlácek
K22	*S. sciuri* (1)	1.5	1992	Morgan Horse	USA	(172)
M1886	*S. sciuri* (1)	64	2011	Human	Denmark	Hvidovre Hospital
K1	*S. sciuri* (2)	0.75	1972	Eastern grey squirrel	USA	(172)
K105	*S. sciuri* (2)	1	1971	Human	USA	(172)
K13	*S. sciuri* (2)	0.5	1992	Eastern grey squirrel	USA	(172)
K139	*S. sciuri* (2)	1	1992	Holstein cow	USA	(172)
K14	*S. sciuri* (2)	0.75	1992	Eastern harvest mouse	USA	(172)
K140	*S. sciuri* (2)	1	1972	Opossum	USA	(172)
K142	*S. sciuri* (2)	1	1992	Horse	USA	(172)
K143	*S. sciuri* (2)	1	1972	Racoon	USA	(172)
K144	*S. sciuri* (2)	1	1992	Jersey calf	USA	(172)
K149	*S. sciuri* (2)	0.75	1972	Eastern grey squirrel	USA	(172)
K21	*S. sciuri* (2)	0.5	1993	Pilot whale	USA	(172)
K23	*S. sciuri* (2)	0.75	1992	Red kangaroo	USA	(172)
K25	*S. sciuri* (2)	0.5	1992	Prairie vole	USA	(172)
KLO56	*S. sciuri* (2)	0.75	1972	Opossum	USA	(172)
KLO58	*S. sciuri* (2)	0.75	1972	Squirrel monkey	USA	(172)
KLO63	*S. sciuri* (2)	1	1972	Eastern grey squirrel	USA	(172)
KLO64	*S. sciuri* (2)	2	1972	Southern flying squirrel	USA	(172)
CH17	*S. sciuri* (3)	>256	2004	Horse	Switzerland	(178)
CH18	*S. sciuri* (3)	>256	2005	Horse	Switzerland	(178)
D573	*S. sciuri* (3)	>256	2007	Human	Denmark	Hvidovre Hospital
HSM805	*S. sciuri* (3)	0.75	2010	Human	Portugal	J. Melo-Cristino
HSM851	*S. sciuri* (3)	16	2010	Human	Portugal	J. Melo-Cristino
JUG17	*S. sciuri* (3)	>256	2002	Human	Yugoslavia	I. Sedlácek
JUG2	*S. sciuri* (3)	0.75	2002	Domestic dog	Yoguslavia	I. Sedlácek
K132	*S. sciuri* (3)	0.5	1976	Howler monkey	Panama	(172)
K141	*S. sciuri* (3)	1	1992	California mouse	USA	(172)
K148	*S. sciuri* (3)	0.38	1992	Bottlenose dolphin	USA	(172)
K2	*S. sciuri* (3)	1	1992	Beef tongue	USA	(172)
K20	*S. sciuri* (3)	1	1992	Bottlenose dolphin	USA	(172)

Strain	Species (*S. sciuri* phylogenetic group)	MIC (μg/ml)	Year of isolation	Origin	Geographic location	Source or Reference
K24	*S. sciuri* (3)	0.75	1992	Morgan Horse	USA	(172)
K51	*S. sciuri* (3)	1.5	1971	Human	USA	(172)
K61	*S. sciuri* (3)	1.5	1992	Pilot whale	USA	(172)
K69	*S. sciuri* (3)	1	1986	Human	USA	(172)
KLO59	*S. sciuri* (3)	2	1972	Domestic dog	USA	(172)
M1234	*S. sciuri* (3)	>256	2009	Human	Denmark	Hvidovre Hospital
M1640	*S. sciuri* (3)	96	2010	Human	Denmark	Hvidovre Hospital
M1653	*S. sciuri* (3)	>256	2010	Human	Denmark	Hvidovre Hospital
M2276	*S. sciuri* (3)	>256	2011	Human	Denmark	Hvidovre Hospital
M2590	*S. sciuri* (3)	>256	2012	Human	Denmark	Hvidovre Hospital
M2710	*S. sciuri* (3)	>256	2012	Human	Denmark	Hvidovre Hospital
M692	*S. sciuri* (3)	96	2007	Human	Denmark	Hvidovre Hospital
SS16	*S. sciuri* (3)	0.75	1996	Human	Portugal	(202)
SS18	*S. sciuri* (3)	0.75	1996	Human	Portugal	(202)
SS23	*S. sciuri* (3)	1	1997	Human	Portugal	(202)
SS24	*S. sciuri* (3)	0.75	1997	Human	Portugal	(202)
SS27	*S. sciuri* (3)	0.75	1998	Human	Portugal	(202)
SS3	*S. sciuri* (3)	1	1996	Human	Portugal	(202)
SS5	*S. sciuri* (3)	1	1996	Human	Portugal	(202)
K3	*S. sciuri* (4)	>256	1992	Neonatal ward	Mozambique	(172)
CH16	*S. sciuri* (4)	24	2004	Horse	Switzerland	(178)
K10	*S. sciuri* (4)	0.75	1992	European red squirrel	USA	(172)
K125	*S. sciuri* (4)	1	1992	Cotton rat	USA	(172)
K27	*S. sciuri* (4)	0.75	1992	Norway rat	USA	(172)
K29	*S. sciuri* (4)	0.75	1992	Norway rat	USA	(172)
K4	*S. sciuri* (4)	>256	1992	Human	Mozambique	(172)
K5	*S. sciuri* (4)	2	1992	Human	Mozambique	(172)
K6	*S. sciuri* (4)	>256	1992	Human	Mozambique	(172)
K7	*S. sciuri* (4)	>256	1992	Human	Mozambique	(172)
K83	*S. sciuri* (4)	0.75	1992	Human	Czech Republic	(172)
SS34	*S. sciuri* (4)	0.75	1996	Human	Portugal	(202)
SS37	*S. sciuri* (4)	25	1996	Human	Portugal	(202)
SS41	*S. sciuri* (4)	3	1996	Human	Portugal	(202)
11/01	*S. sciuri* (5)	0.5	2002	Human	Czech Republic	O. Melter
CCUG38359	*S. sciuri* (5)	0.75	2002	Human	Czech Republic	O. Melter
K11	*S. sciuri* (5)	0.75	1990	Veal leg	USA	(172)
K116	*S. sciuri* (5)	0.75	1992	Beef lips	USA	(172)
K12	*S. sciuri* (5)	1	1992	Arabian Horse	USA	(172)
K16	*S. sciuri* (5)	0.75	1992	Jersey calf	USA	(172)
K163	*S. sciuri* (5)	0.75	1992	Holstein cow	USA	(172)
K30	*S. sciuri* (5)	1.5	1992	Jersey cattle heifer	USA	(172)
K31	*S. sciuri* (5)	0.5	1992	Jersey cattle calf	USA	(172)
K32	*S. sciuri* (5)	0.75	1992	Jersey cattle calf	USA	(172)
K33	*S. sciuri* (5)	0.75	1992	Jersey cattle heifer	USA	(172)
H39	*S. vitulinus*	0.75	2005	Horse	Denmark	L. Guardabassi P. Damborg
H91	*S. vitulinus*	2	2005	Horse	Denmark	L. Guardabassi P. Damborg
401946	*S. vitulinus*	2	2004	Horse	The Netherlands	E. van Duijkeren
CH1	*S. vitulinus*	1	2005	Horse	Switzerland	(178)
CH2	*S. vitulinus*	4	2004	Horse	Switzerland	(178)

Strain	Species (*S. sciuri* phylogenetic group)	MIC (µg/ml)	Year of isolation	Origin	Geographic location	Source or Reference
CH3	*S. vitulinus*	0.75	2005	Horse	Switzerland	(178)
CH4	*S. vitulinus*	0.75	2004	Horse	Switzerland	(178)
CH5	*S. vitulinus*	>256	2005	Horse	Switzerland	(178)
CH6	*S. vitulinus*	1.5	2004	Horse	Switzerland	(178)
CH7	*S. vitulinus*	0.75	2004	Horse	Switzerland	(178)
CH8	*S. vitulinus*	0.19	2005	Horse	Switzerland	(178)
CH9	*S. vitulinus*	0.5	2004	Horse	Switzerland	(178)
CH10	*S. vitulinus*	0.5	2004	Horse	Switzerland	(178)
CH11	*S. vitulinus*	0.5	2005	Horse	Switzerland	(178)
CH12	*S. vitulinus*	0.5	2004	Horse	Switzerland	(178)
CH13	*S. vitulinus*	0.1	2004	Horse	Switzerland	(178)
CH14	*S. vitulinus*	0.5	2005	Horse	Switzerland	(178)
CH15	*S. vitulinus*	>256	2004	Horse	Switzerland	(178)

Supplementary Table S2. Content of SCC and SCC*mec* elements found among *S. sciuri*, *S. vitulinus* and *S. fleurettii*. Only the elements flanked by DR and/or IR repeats are listed on the table.

Strain (species)	Element (bp)	Direct repeats (DR) or inverse repeats (IR)	Gene/protein*	Position in contig
11/01 (*S. sciuri*)	SCC $_{11/01}$ (53848)	GGGGAAGCGTATCATAAATGATGCGGTTTTTT	Threonine dehydrogenase	47225-48262
			Type I restriction modification enzyme	48946-51738
			Type I restriction modification methyltransferase	51850-53406
			Type I restriction modification S subunit	53755-54669
			Hypothetical protein 1: 47.6% KF234240 (SCC*mec* IVa – J1)	54754-55527
			ORF	55532-56110
			Transcriptional regulator	56407-57099
			Ferroxidase	57116-57667
			Copper chaperone	57669-57905
			Cadmium-transporting ATPase	57969-59819
			ORF	60127-60252
			cadC	60591-60959
			Cadmium-transporting ATPase	60952-63135
			cadD	63233-63460
			Hypothetical protein 2: 96.62% CP006838 (SCC*mec* III – J2)	63478-63684
			Hypothetical protein 3: 66.9% CP006838 (SCC*mec* III – J2)	63686-64090
			repB	64320-65000
			Hypothetical protein 4: 76.92% HF569115 (SCC*mec* IVa+*ccrAB1* – J3)	65126-65629
			Hypothetical protein 5: 68.69% AB505628 (SCC*mec* IX – J3)	65645-65953
			Hypothetical protein 6: 90.69% AB505628 (SCC*mec* IX – J3)	66038-66391
			ccrB5	66861-68489
			ccrA5	68510-69859
			Hypothetical protein 7: 78.56% AB705452 (SCC*mec* IX – J2)	70038-71828
			Hypothetical protein 8: 93.36% AB505628 (SCC*mec* IX – J2)	71954-73384
			ATPase	74084-75607
			copaA – truncrated	75629-76147
			Putative lipoprotein	76165-76710

Strain (species)	Element (bp)	Direct repeats (DR) or inverse repeats (IR)	Gene/protein*	Position in contig
			ORF	77190-77651
			arsC	78815-79021
			Metallo-βlactamase family protein-truncrated	81553-81814
			IS257 transposase	88244-88759
			IS431 transposase	88719-88919
			Acyltransferase precursor	91722-92195
			Oxidase	92196-93692
			Glycosyltransferase	93753-94796
			Dehydrosqualene synthase	94852-95733
			Dehydrosqualene desaturase	95730-97235
			General stress protein 26	97284-97709
			Deoxyribodipyrimidine photolyase	97893-99275
			Hypothetical protein 10: 58.69% HE993884 (SCC*mec mecC S. xylosus* - J1)	99525-99965
K6 (*S. sciuri*)	SCC_{K6} (15937)	ATGATGCGGTTTTTT	Hypothetical protein 1: 78.81% CP006630 (SCC*mec* V – J1)	152605-153912
			Hypothetical protein 2: 90.29% CP006838 (SCC*mec* III – J1)	154412-155461
			ORF	155593-155883
			Putative helicase	155883-157676
			ccrA5	157854-159206
			ccrB new type 2: 72% *ccrB1*	159223-160854
			Hypothetical protein 3: 95.19% CP006838 (SCC*mec* III – J2)	161749-162060
			Hypothetical protein 4: 99.8% AB505628 (SCC*mec* IX – J3)	162081-162584
			Plasmid replication initiation protein	162786-163745
			DNA modification methylase	163595-167209
			Conserved domain protein	167299-168210
K6 (*S. sciuri*)	ΨSCC_{K6} (10930)	AAAACCGCATCATTT	Permease	168506-168886
			ORF	168962-169546
			Mobile element protein	170153-171040
			TPR domain protein in aerotolerance operon	171224-173869
			ORF	174071-174214

Strain (species)	Element (bp)	Direct repeats (DR) or inverse repeats (IR)	Gene/protein*	Position in contig
			Hypothetical protein 1: 54.97% CP006838 (SCC*mec* III – J3)	174287-174838
			Hypothetical protein 2: 80.65% HE980450 (SCC*mec* III – J3)	174842-175822
			ORF	176033-176935
			ORF	177041-177358
			Hypothetical protein 3: 65.17% AB373032 (SCC*mec* 5C1 – J1)	177847-179250
K6 (*S. sciuri*)	SCC*mec* III-like structure B (43815)	GATGCGGTTTTTT	Arsenic resistance operon repressor	179490-179804
			ORF	179758-180168
			Arsenic efflux pump	180722-181144
			Arsenate reductase	181162-181563
			Transcriptional regulator	181689-181991
			chrA	182142-183164
			Polysulfide binding protein	183357-184691
			Disulfide bond regulator	184722-185789
			Transcriptional regulator	185927-186187
			Putative transmembrane protein	186208-186942
			Abortive phage resistance protein	187360-188163
			Abi-alpha protein	188176-189339
			ORF	189449-190987
			ORF	191190-191426
			ORF	191428-193215
			Hypothetical protein 1: 72.76% AB705452 (SCC*mec* IX – J2)	193305-193583
			ccrA3	193772-195118
			ccrB5	195139-196767
			Hypothetical protein 2: 100% JQ412578 (SCC*mec* II – J2)	197238-197588
			Hypothetical protein 3: 99.68% JQ412578 (SCC*mec* II – J2)	197673-197984
			Hypothetical protein 4: 100% JQ412578 (SCC*mec* II – J2)	198003-198524
			Hypothetical protein 5: 99.52% CP006838 (SCC*mec* III – J2)	198526-198732
			DNA repair protein RadC	198725-199048
			Mobile element protein	199243-201135

Strain (species)	Element (bp)	Direct repeats (DR) or inverse repeats (IR)	Gene/protein*	Position in contig
			Mobile element protein	201142-201303
			Mobile element protein	201298-202620
			Mercuric ion reductase	202862-204502
			Mercuric resistance operon regulatory protein	204527-204919
			Cadmium efflux system accessory protein	205631-205996
			Cadmium transporting ATPase	205989-208403
			Cadmium resistance protein	208484-209101
			Hypothetical protein 1: 60.32% CP006838 (SCC*mec* III – J2)	209369-209998
			Hypothetical protein 2: 100% CP006838 (SCC*mec* III – J2)	210013-210273
			Disulfide bond regulator	210409-211473
			Zn-dependent hydroacylglutathione hydrolase	211503-212444
			Polysulfide binding protein	212535-212840
			Transcriptional regulator	212954-214102
			mecI	214575-214946
			mecR1	214946-216703
			mecA	216803-218809
			maoC	218855-219283
			ugpQ	219380-220123
			hmg-coA	221040-221207
			Mobile element protein	221465-222139
			Hypothetical protein 3: 100% FN433596 (SCC*mec* III – J3)	222271-222513
M1640 (*S. sciuri*)	SCC*mec* III-like structure A (62689)	TTAATGATGCGGTTTTT	CadD	41887-42516
			Hypothetical protein 1: 96.4% AB505630 (SCC*mec* X-J1)	43277-43942
			putative peptidase	43964-44647
			putative transcriptional regulator	44725-45072
			putative glycosyltransferase	45581-47197
			putative surface protein	47339-50968
			Hypothetical protein 2: 98.48% CP006838 (SCC*mec* III-J1)	52344-53393
			Hypothetical protein 3: 90.57% CP006838 (SCC*mec* III-J1)	53497-53814

Strain (species)	Element (bp)	Direct repeats (DR) or inverse repeats (IR)	Gene/protein*	Position in contig
			Hypothetical protein 4: 100% CP006838 (SCC*mec* III-J1)	53814-55601
			ORF	55628-55966
			ccrA3	56141-57505
			ccrB5	57526-59154
			ORF	59625-59975
			Hypothetical protein 5: 78.79% CP006838 (SCC*mec* III-J2)	59976-60371
			Hypothetical protein 6: 100% CP006838 (SCC*mec* III-J2)	60390-60911
			Hypothetical protein 7: 43.85% CP006838 (SCC*mec* III-J2)	61084-61428
			Tnp	61630-63522
			tnp for Tn554	63529-63906
			cadC	64144-64509
			cadA	64493-66916
			CadD	66955-67614
			Hypothetical protein 8: 38.18% CP006838 (SCC*mec* III-J2)	67832-68353
			Rhodanese-domain containing protein	68922-69986
			B-lactamase domain containing protein	70019-70957
			Hypothetical protein 8: 100% CP006838 (SCC*mec* III-J1)	71048-71353
			mecR2	71467-72615
			mecI	73088-73459
			mecR1	73459-75198
			mecA	75313-77322
			Hypothetical protein 9: 42.40% AB872255 (SCC*mec* IV-J3)	77315-77698
			ugpQ	77893-78636
			Tnp	79978-80652
			Hypothetical protein 10: 96.32% AB872255 (SCC*mec* IV-J3)	81338-82642
K116 (*S. sciuri*)	SCC_{K116} (35661)	AGTTTTATTTGTGATATGCTT	Photolyase-truncrated	17118-17730
			ABC transporter permease	18548-19597
			ORF	19609-20224
			General stress protein 26	21135-21428
			Dehydrosqualene desaturase	21480-22985

Strain (species)	Element (bp)	Direct repeats (DR) or inverse repeats (IR)	Gene/protein*	Position in contig
			Dehydrosqualene synthase	22982-23863
			Glycosyltransferase	23919-24962
			Oxidase	25023-26519
			Acyltransferase precursor	26520-26993
			ORF	26998-27498
			ORF	28140-29534
			adhC/alcohol dehydrogenase	30512-31663
			ORF	31845-32360
			Threonine dehydrogenase	32544-33581
			chrA	33690-34852
			Tnp	36633-37496
			Tnp	37460-38011
			cadX	38208-38630
			cadD	38648-39265
			NAD dependent epimerase/dehydrogenase	40058-40495
			NAD dependent epimerase	40479-40724
			Hypothetical protein 1: 91.19% CP006838 (SCC*mec* III-J2)	41545-42057
			Hypothetical protein 2: 98.08% CP006838 (SCC*mec* III-J2)	42077-42388
			Hypothetical protein 3: 89.74% JQ412578 (SCC*mec* II-J2)	42473-42823
			75% *ccrC*	42928-44604
			Hypothetical protein 4: 79.68% AB781449 (SCC*mec* V-J2)	44830-46473
			Hypothetical protein 5: 73.17% AB781449 (SCC*mec* V-J2)	46473-46841
SS27 (SCC-CI $_{SS27}$) (*S. sciuri*)	SCC $_{SS27\ I}$ (41603)	TATCATAAATGATGCGGTTTTTT	Type II methyltransferase	907-1737
			ORF	1724-4489
			ORF	4673-5038
			Hypothetical protein 1: 77.02% CP006838 (SCC*mec* III-J2)	5097-5492
			Hypothetical protein 2: 89.95% AB505630 (SCC*mec* X-J2)	5493-5846
			ccrB3	6318-7946
			ccrA1	7967-9319

Strain (species)	Element (bp)	Direct repeats (DR) or inverse repeats (IR)	Gene/protein*	Position in contig
			ORF	9875-11713
			ORF	12103-13641
			rep	13751-14914
			ORF	14928-15722
			Transmembrane protein	16119-16874
			ORF	16147-16542
			Rhodanese-domain containing protein	17272-18339
			Metallo-β lactamase family protein	18368-19702
			DNA-invertase	19873-20454
			Tnp	20710-21879
			ATP-binding domain	21872-22687
			sin recombinase	22821-23426
			ORF	23487-23831
			Copper ATPase	24325-26430
			Multicopper oxidase	26445-27878
			Lipoprotein ACME	27898-28440
			cadD	28933-29547
			Tnp	29658-30461
			cadA	30877-33294
			cadC	33287-33655
			Arsenate reductase	34295-34690
			Arsenic efflux pump protein	34708-36000
			ArsR	36000-36314
			Putative dehydrogenase	36311-37975
			Arsenical pump-driving ATPase	37975-39738
			Trans regulator of arsenic operon	39704-40093
			Arsenical resistance repressor	40572-40892
			Putative membrane protein	40975-41862
SS27 (SCC-CI $_{SS27}$) (*S. sciuri*)	SCC $_{SS27\ II}$ (15334)	AAAACCGCATCACTATATGATAAGC	ORF	42715-44439

Strain (species)	Element (bp)	Direct repeats (DR) or inverse repeats (IR)	Gene/protein*	Position in contig
			ORF	44510-45121
			ORF	45150-46352
			Hypothetical protein 1: 73.14% AB505628 (SCC*mec* IX-J2)	46433-46954
			Hypothetical protein 2: 97.47% AB505628 (SCC*mec* IX-J2)	46972-47367
			Hypothetical protein 3: 94.07% CP002120 (SCC*mec* II-J2)	47368-47721
			ccrB3	48187-49818
			ccrA1	49839-51188
			Hypothetical protein 4: 76.9% AB705452 (SCC*mec* IX-J1)	51380-51721
			Hypothetical protein 5: 73.14% LK02544 (SCC*mec* XI-J1)	51724-53538
			Hypothetical protein 6: 61.81% AB097677 (SCC*mec* IV-J2)	53949-55019
			Hypothetical protein 7: 79.28% AB097677 (SCC*mec* IV-J2)	55019-56548
			ORF	56548-57501
402567 (*S. fleurettii*)	SCC_{402567} (17929)	AAAAAACCGACTCATTTATGT	ORF	4731-6800
			ORF	6801-8126
			ORF	7420-7740
			ORF	8263-10338
			ORF	10968-12395
			Hypothetical protein 1: 75% AB705452(SCC*mec* IX-J2)	12521-14311
			ORF	12976-13287
			ccrA new type 2: 77% *ccrA1*	14793-16160
			ccrB5	16180-17808
			Hypothetical protein 2: 96% CP005288 (SCC*mec* III-J2)	18279-18629
			Hypothetical protein 3: 77.53% CP005288 (SCC*mec* III-J2)	18630-19025
			ORF	19041-19547
			ORF	20460-21422
			ORF	21454-22404
401946 (*S. vitulinus*)	SCC_{401946} (32443)	AAAAAACCGCATCATT	ORF	446-1519
			Type I restriction/modification system	4045-5676
			ORF	5661-6221
			ORF	6230-7261

Strain (species)	Element (bp)	Direct repeats (DR) or inverse repeats (IR)	Gene/protein*	Position in contig
			ORF	7194-8234
			ORF	8235-8795
			Hypothetical protein 1: 51.26% HF569109 (SCC*mec* IV-J3)	8780-10462
			Type I restriction/modification system DNA methyltransferase subunit M	15060-16514
			Type I restriction/modification system DNA methyltransferase subunit S	16504-17763
			hsdR	17741-20869
			ORF	21448-21954
			ORF	21970-22287
			ccrB new type 1: 76% *ccrB1*	23189-24829
			ccrA new type 1: 82% *ccrA1*	24838-26187
			Hypothetical protein 2: 75.86% AB063173 (SCC*mec* IX-J1)	26690-28480
			Hypothetical protein 3: 81.89% KM369884 (SCC*mec* V-J1)	31355-32257
CH10 (*S. vitulinus*)	SCC_{CH10} (11684)	ATTATTAGATTTTTT	ORF	243-2318
			Hypothetical protein 1: 97.20% AB705452 (SCC*mec* IX-J2)	2949-4376
			Hypothetical protein 2: 95.20% AB705453 (SCC*mec* IX-J1)	4502-6292
			ccrA new type 2: 82.74% *ccrA3*	6465-7832
			ccrB new type 4: 81% *ccrB4*	7852-9486
			Hypothetical protein 3: 89.97% AB705452 (SCC*mec* IX-J1)	9960-10298
			Hypothetical protein 4: 83.33% AB705452 (SCC*mec* IX-J1)	10394-10705
			Hypothetical protein 5: 56.61% AB705452 (SCC*mec* IX-J1)	10721-11233
			ORF	11615-12034
			ORF	12485-13579
			ORF	13621-15297

* when the gene is not identified, only the most similar protein present in the NCBI database is shown.

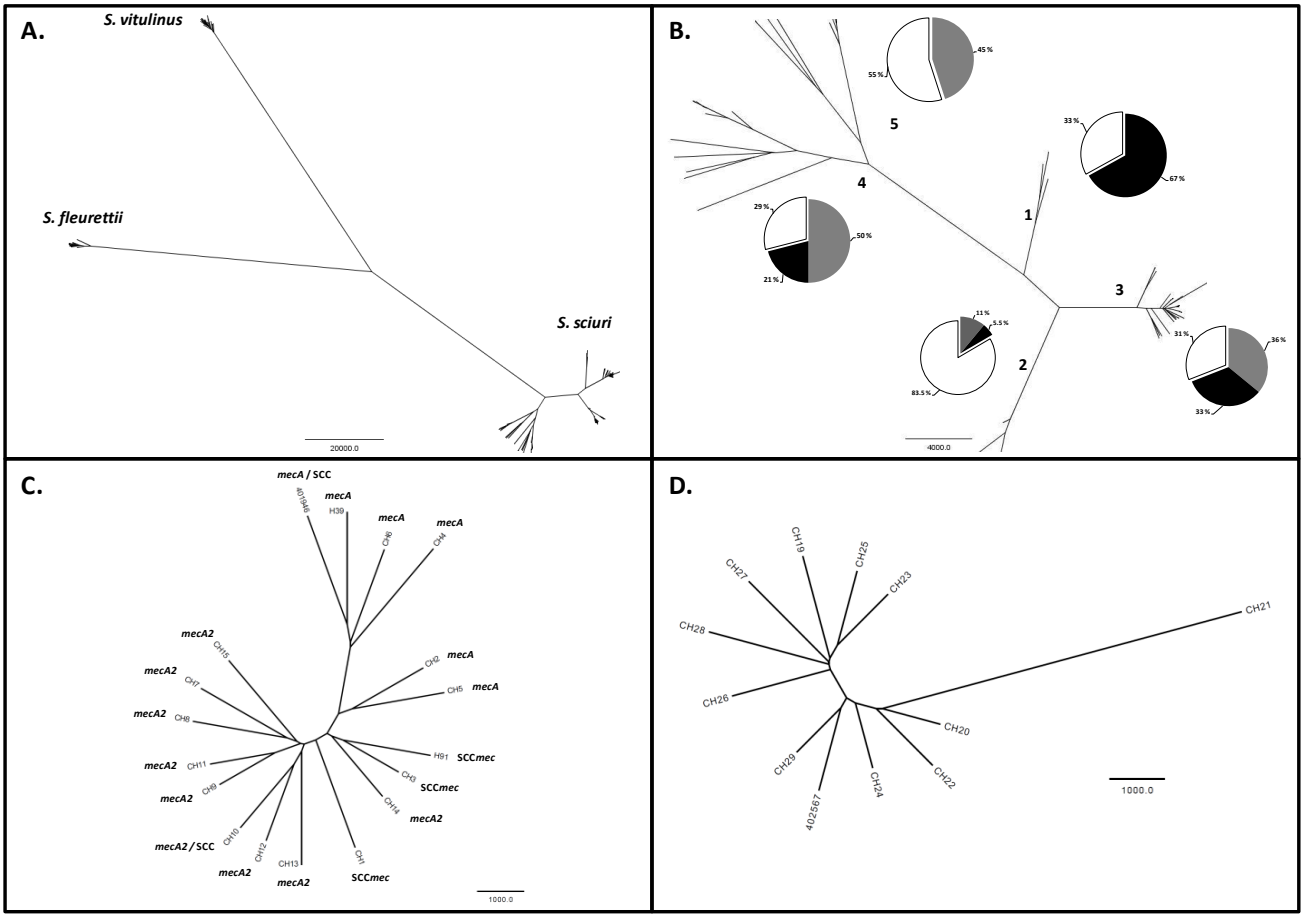
Supplementary Table S3. Predicted ORFs of the *orfX*-containing contig of *ccr*-negative strain K22. Nucleotide sequences of the predicted ORFs were compared with the ones available at http://www.ncbi.nlm.nih.gov/.

Position in contig	Gene *	Species *
10..846	Similar to MCCL_1950/*parB*	*M. caseolyticus*
1415..2521	Similar to A284_11420/cysthathione gamma synthase	*S. warneri*
2518..3690	Similar to SERP0036/trans-sulfuration enzyme	*S. epidermidis*
3647..5485	Similar to SSP2415/methyltransferase	*S. saprophyticus*
5482..7728	Similar to SAMSHR1132_03290/methyltransferase	*S. aureus*
7745..8488	Similar to A284_11440/metal dependent hydrolase	*S. warneri*
8602..10056	Similar to SSP2440/*cstA*	*Bacillus sp."*
10208..11071	Similar to MCCL_1952/ParB protein homologue	*M. caseolyticus*
11088..11807	*gidB*	*M. caseolyticus*
11809..13683	*gidA*	*S. pasteuri*
13699..15078	Similar to SAEMRSA15_26100/tRNA modification GTPase	*S. aureus*
15192..15989	Similar to MCCL_1956/single stranded DNA binding protein	*M. caseolyticus*
16019..16369	Similar to rnpA/ribonuclease protein component	*S. epidermidis*
16461..16598	Similar to SPSINT_0001/ribosomal protein	*S. pseudointermedius*
17150..18490	*dnaA*	*S. haemolyticus*
18657..19790	*dnaN*	*S. carnosus*
20152..21051	Similar to MCCL_0003/Na efflux pump	*M. caseolyticus*
24025..24246	Similar to X998_0003/S4 domain protein	*S. aureus*
24250..25365	*recF*	*S haemolyticus*
25388..27310	*gyrB*	*S. warneri*
27344..29944	Similar to A284_00030/DNA gyrase	*S. warneri*
30018..30614	Similar to SH0007/hyp core protein	*S. haemolyticus*
31050..31487	*hutP*	*Bacillus sp."*
31582..33066	Similar to STP1_1390/histidina ammonia lyase	*S. pasteuri*
33094..34404	Similar to SCA_1829/hyp core protein	*S. carnosus*
34422..36089	Similar to *hutU*/urocanate hydratase"	*S. carnosus*
36082..37323	Similar to STP1_0815/imidazolepropionase	*S. pasteuri*
37631..38908	Similar to SAMSHR1132_00090/seryl tRNA synthase	*S. aureus*
39033..39548	Similar to CH52_05575/hyp core protein	*S. aureus*
39824..40468	Similar to SAMSHR1132_00100/membrane protein	*S. aureus*
41027..41998	Similar to SERP2541/homoserine-o-acetyltransferase	*S. epidermidis*
43012..44976	Similar to STP1_1384/DHHA1 domain containing protein	*S. pasteuri*
45513..46922	Similar to CH52_05370/DNA helicase	*S. aureus*
47209..48495	*purA*	*S. aureus*
49483..50367	Similar to MCCL_0020/hyp core protein	*M. caseolyticus*

Position in contig	Gene *	Species *
50808..51509	Two-component response regulator	*S. sciuri*
51519..53360	Two-component sensor kinase	*S. sciuri*
53335..54675	*yycH*	*S. sciuri*
54675..55466	hypothetical protein SA_21	*S. sciuri*
55498..56289	Zn-dependent hydrolase	*S. sciuri*
56584..5706	*orfX*	*S. sciuri*
57138..58442	Similar to SAKOR_01992/potassium uptake protein	*S. aureus*
58611..58955	*arsC*	*S. sciuri*
59027..60316	*arsB*	*S. sciuri*
60845..63874	putative phage infection protein	*S. sciuri*
64173..65006	Similar to SERP2401/serine protease	*S. epidermidis*
70703..72076	Similar to deoxyribodipyrimidine photolyase"	*S. xylosus*
72122..72541	hyp core protein *S. xylosus*	*S. xylosus*
72701..73741	putative alcohol dehydrogenase	*S. sciuri*
74151..75131	*dus*/tRNA synthase	*S. aureus*
75352..76749	Similar to SCA_2443/hyp core protein	*S. carnosus*
77115..78032	Similar to SAMSHR1132_20200/putative exported protein	*S. aureus*
78488..79813	Similar to SPSINT_2422/hyp core protein	*S. pseudointermedius*
80071..83217	Similar to SSP0074/hyp core protein	*S. saprophyticus*
84623..86254	*xylB*/xylulokinase	*S. carnosus*
86223..86918	Similar to M7W_2352/epimerase	*E. faecium*
86932..88353	Similar to araA2/arabinose isomerase	*E. faecium*
88409..89815	Similar to T256_00940/sugar transporter	*Pediococcus sp.*
89882..90736	*blaZ*	*S. aureus*
90874..91191	Similar to SCA_2438/hyp core protein	*S. carnosus*
91286..92236	Similar to SSP1638/putative lipase	*S. saprophyticus*
92431..93354	Similar to SCA_2436/hyp core protein	*S. carnosus*
93356..94051	Similar to SCA_2435/hyp core protein	*S. aureus*
94280..95605	Similar to SSP0521/hyp core protein	*S. saprophyticus*
95602..96618	Similar to SSP0520/putative oxidoreductase	*S. saprophyticus*
96960..97487	Similar to SAMSHR1132_22300/acetyltransferase	*S. aureus*
97503..98294	Similar to lin2443/ABC transporter	*Listeria sp.*
98313..99026	Similar to LMOSLCC2376_2243/ABC transporter/permease	*Listeria sp.*
99038..99742	Similar to AX10_05710/cystheine ABC transporter/permease	*Listeria sp.*
99743..100525	Similar to *yecC*/ATP transporter permease ATP binding protein	*Listeria sp.*
100720..101481	Similar to SSP2181/transcriptional regulator	*S. saprophyticus*
101546..102784	Similar to SSP2180/alpha-ketoglutarate permease	*S. saprophyticus*
102808..103878	Similar to A284_10560/dehydrogenase	*S. warneri*
104040..105671	Similar to A284_10555/ribulokinase	*S. warneri*
105831..106085	Similar to STP1_1355/prevent host death protein	*S. pasteuri*
106900..107373	Similar to MC28_F090/acetyltransferase	*Bacillus sp*
112374..113495	FAD-dependent oxidoreductase	*S. haemolyticus*

Position in contig	Gene *	Species *
113489..114829	*feoB*+putative membrane protein	*S. haemolyticus*
115380..116267	SERP2386/colabamin synthesis protein	*S. epidermidis*
116528..117478	SPSINT_2071/uroporphyrinogen methyltransferase	*S. pseudointermedius*
117469..117786	*nirD*	*S. pseudointermedius*
117789..120194	SPSINT_2073/nitrite reductase	*S. pseudointermedius*
121609..123039	Similar to Pcryo_0766/dicarboxylate anaerobic carried	*Psychrobacter sp.*
124183..126057	Similar to Cbei_0699/phosphotransferase	*Clostridium sp.*
126054..127472	Similar to Cbei_0700/glycoside hydrolase	*Clostridium sp.*
127687..128667	Similar to STP1_1330/alpha-β family hydrolase	*S. pasteuri*
128682..129251	Similar to STP1_1331/acetyltransferase	*S. pasteuri*
129372..130241	Similar to STP1_1332/LysR transcriptional regulator	*S. pasteuri*
130979..132601	Similar to SCA_2062/hyp core protein	*S. carnosus*
132890..134380	Similar to SCA_2061/allontoin permease	*S. carnosus*
134370..135731	Allantoinase	*S. xylosus*

* nucleotide sequences with the highest BLAST score; hyp: hypothetical



Supplementary Figure 1. Phylogenetic tree reflecting the number of SNP differences among the genomes of the isolates studied. *S. fleurettii* 402567 was used as a reference. **A.** All genomes (*S. sciuri, S. fleurettii, S. vitulinus*). **B.** *S. sciuri* genomes. Phylogenetic groups were defined according with the number of single-nucleotide polymorphisms (SNPs) differences. Phylogenetic group 1: new subspecies group 1; phylogenetic group 2: *S. sciuri sciuri*; phylogenetic group 3: new subspecies group 3; phylogenetic group 4: *S. sciuri rodentius*; phylogenetic group 5: *S. sciuri carnaticus*. The distribution of SCC*mec* and SCC elements among the isolates belonging to each phylogenetic group (1-5) is also shown; black: SCC*mec*; grey: SCC; white: no carriage of SCC*mec*/SCC. **C.** *S. vitulinus* genomes. The distribution of *mec* homologues, SCC and SCC*mec* elements is shown**. D.** *S. fleurettii* genomes.

PART II

Impact of SCC*mec* in the evolution and adaptation of *S. epidermidis* to the hospital environment

Evolution of the skin commensal *Staphylococcus epidermidis* in the hospital environment: insights from an early collection

Joana Rolo, Rita Sobral, Jesper Boye Nielsen, Peder Worning, Rory Bowden, Samuel Sheppard, Guillaume Méric, Robert Skov, Henrik Westh, Hermínia de Lencastre and Maria Miragaia

Manuscript in preparation

All the experimental work and the whole-genome sequencing analysis -with the exception of SNP analysis-was performed by J. Rolo

ABSTRACT

The emergence of the human skin colonizer *Staphylococcus epidermidis* as a pathogen related to medical devices-associated infections occurred relatively recently. However, how the pressure of the nosocomial setting drove *S. epidermidis* evolution, remains to be clarified. To identify the genomic events associated with the recent evolutionary history of *S. epidermidis* we compared by whole genome sequencing early and contemporary nosocomial *S. epidermidis* strains collected in Denmark.

Early and contemporary isolates had a conserved genetic background, related to clonal complex 2 (67% and 79%, respectively). Early isolates carried antibiotic resistance genes (*aaD*, 9.5%; *ermA*, 9.5%; *fusB*, 9.5%, *tet(K)*, 14.3%; *cat*, 14.3%; *blaZ*, 67%, *norA*, 100%, *fosA*, 100%); a single isolate carried *mecA*, disrupted by IS*431*, within a staphylococcal cassette chromosome *mec* (SCC*mec*) IV-like element. Biofilm-associated genes (*ica* 5%; *aap* 38%), the arginine catabolic mobile element (ACME I and II, 38% each) and insertion sequences (IS) were common and a high frequency of prophages (95%) was found. In contrast, in contemporary strains, resistance determinants (including SCC*mec* IV, 57%; *tet(L)*, 5%; *lnu(B),* 5%; *ermC*, 14%; *aac(6')-aph(2'')*, 19%; *fusB*, 43%; *blaZ*, 100%, *norA*, 100%, *fosA*, 100%), biofilm-associated genes (*bap*, 7%; *sdrF*, 14%; *ica*, 21%; *aap*, 64%) and ACME-I (57%) were more represented and additional IS were identified. However, the frequency of prophages was lower (50%).

We propose that the continued exposure of *S. epidermidis* to hospital environment led to accumulation of genes associated to antibiotic resistance, colonization, biofilm formation, genome plasticity and to phage loss what might have contributed to its success as a pathogen.

INTRODUCTION

Staphylococcus epidermidis is the main colonizer of human skin, playing an important role in the protection against pathogenic bacteria through the stimulation of immune responses (60). For this reason for many years *S. epidermidis* was regarded as a harmless commensal. Coagulase-negative staphylococci (CoNS) like *S. epidermidis* were first recognized as the probable cause of a septicemia in 1958 (227). But it was only when the first death due to a CoNS infection was reported in 1964 that these bacteria became recognized as potential human pathogens (47). However, until the 1980s, CoNS were frequently regarded as contaminants of clinical products and were rarely identified at species level. The remarkable progresses made in the last decades in staphylococcal taxonomy and the development of phenotypic and genotypic identification methods contributed to the recognition of *S. epidermidis* as a pathogen. Moreover, other factors like the increase of the use of intravascular catheters and medical devices as well as the rise in the number of immunocompromized patients inside the hospital associated to a higher life expectancy, prolongation of hospitalization periods and use of immunosuppressive therapies and antibiotics, contributed greatly for the growing incidence and clinical significance of *S. epidermidis* as an infectious agent.

Today *S. epidermidis* is recognized as one of the most frequent pathogen related to medical device-associated infections worldwide (2). Most of the infections are believed to be endogenous, wherein bacteria from patient's own flora gain access to bloodstream through the impairment of skin barrier occurring during medical procedures, like inoculation of needles, insertion of catheters or implantation of foreign bodies. The success of *S. epidermidis* as a pathogen is dependent mainly on its ability to adhere to host or foreign body surfaces and form biofilms (51), which is known to be achieved by numerous adhesins, proteins, DNA and the polyssacharide intercellular adhesion (PIA) (51). Another factor that is highly frequent in *S. epidermidis* (100, 101) and was described to be important for colonization and dissemination of community-associated *S. aureus* (98) is the arginine catabolic mobile element (ACME), but the associated mechanism is still elusive.

The molecular epidemiology of *S. epidermidis* has been well-studied using different molecular typing techniques. The application of multilocus sequence typing to a diverse collection of isolates in terms of clinical and geographic origin showed that there was a high genetic diversity (www.mlst.net) (17, 198). Nonetheless, it was observed that a single clonal complex, clonal complex 2 (CC2), comprised the great majority of isolates independently from its origin, including hospital and community (17, 198, 228) and that a particular sequence type, ST2, was preferentially associated to disease (77). More recently, the analysis of *S. epidermidis* population structure by whole genome sequencing showed that *S. epidermidis* was composed of two well-separated clusters, one containing CC2 isolates and the other containing minor clonal complexes and sporadic isolates (42). However, the factors associated to the increased epidemicity and pathogenicity of this clonal lineage are unknown.

Another key feature of nosocomial *S. epidermidis* is its resistance to a multitude of antimicrobials (2), of which the most epidemiologically relevant are β-lactams. Resistance to β-lactams in staphylococci is encoded by the *mecA* gene, which is carried in a mobile genetic element called the staphylococcal cassette chromosome *mec* (SCC*mec*)(109). The *mecA* gene encodes an additional penicillin-binding protein, that has low affinity to β-lactams (7). By acquiring SCC*mec*, β-lactam susceptible strains become resistant and MRS (methicillin resistance staphylococci) have emerged. Since its finding in a *Staphylococcus aureus* isolate in 1961, in the UK (69), eleven different types of SCC*mec* have been described in this species (111-114) and many more appear to be present in coagulase-negative staphylococci (2). In *S. epidermidis*, SCC*mec* IV is the most frequent type (124), but other types have been described. Although in the community environment strains colonizing the anterior nares were found to be mostly susceptible to β-lactams, contact with hospital environment was shown to promote acquisition and amplification of SCC*mec* in *S. epidermidis* (198). Nonetheless, it is not known when in *S. epidermidis* history was SCC*mec* acquired.

In this study, we aimed to identify the genomic events associated to the emergence and establishment of *S. epidermidis* as a nosocomial pathogen. To accomplish this we studied *S. epidermidis* evolutionary history by comparing by whole

genome sequencing early and contemporary nosocomial *S. epidermidis* strains collected in Denmark.

METHODS

Bacterial collection: a collection of 21 *S. epidermidis* isolates collected from nasal swabs in hospitalized patients in Denmark, in 1965, was assembled. Each sample was obtained from a single patient. The samples were stored at the Statens Serum Institut, Copenhagen, Denmark, until their inclusion in this particular study. In addition, a collection of 14 *S. epidermidis* isolates, collected in Denmark in five hospitals in Greater Copenhagen between 1997 and 1998 was included for comparison. The great majority of these isolates (11 out of 14) have been obtained from blood cultures, and correspond to infection isolates. Two of the remaining isolates were considered to be colonization isolates (one of these was collected from a wound and for the other no information was available regarding the clinical product). Finally, for one isolate there was no information available regarding its origin. All isolates were obtained from inpatients. These isolates have been previously characterized by molecular typing techniques (17, 79) and their genome sequence has been previously published (42).

Antimicrobial susceptibility testing: susceptibility to two β-lactams (oxacillin and cefoxitin) was assessed by eTest (Oxoid, Basingstoke, United Kingdom) for all the isolates. In addition, cefoxitin (Sigma, St. Louis, United States of America) susceptibility population analysis profiles were performed for the early isolates, as described before (211). Isolates were considered susceptible for oxacillin MIC values <0.25 μg/ml or for cefoxitin MIC values <4 μg/ml.

DNA preparation and whole genome sequencing: DNA was extracted for *S. epidermidis* isolates collected in 1965 with a phenol/chlorophorm extraction protocol, as described in (25). Sequencing libraries were prepared by sonic fragmentation and adapter ligation and then sequenced on the Illumina HiSeq 2000/2500 platform, producing paired 100 bp reads. The DNA of the 14 contemporary *S. epidermidis* isolates have been prepared and sequenced as described in (42).

SNP analysis: the raw reads of all strains collected in 1965 were mapped against the reference *S. epidermidis* strain ATCC12228. SNP analysis was performed using Stampy (version 1.0.11) where reads were mapped to the reference genome. SNP calling was performed using SAMtools (version 0.1.12). Phylogenies were reconstructed using Neighbor Joining (NJ) and drawn in FigTree (http://tree.bio.ed.ac.uk/software/figtree/).

Comparative genome analysis: the core genome of all strains was predicted with Mugsy (229), using no reference strain. The resulting alignment file was used to perform a NJ tree based on the similarity of the homology blocks found with Mauve (34).

Assembly *de novo* and MLST: the raw reads of each strain were assembled into contigs using VELVET (33). To further study the clonal relatedness between the strains, a multilocus sequence type was defined for each strain by extracting the corresponding allele sequence from the resulting contigs according to (17). In addition, eBURST (230) was used to estimate the clonal relatedness of the isolates using the allelic profiles of each sequence types.

Manual annotation: the entire contig containing *orfX* was annotated. The *orfX* was identified by BLAST analysis of the *orfX* sequence carried by the reference *S. epidermidis* strain RP62A (NCBI accession number NC_002976.3). The remaining ORFs of the contig were identified using GeneMark-hmm (http://exon.gatech.edu/GeneMark/) and annotated by BLAST. The sequences of the predicted ORFs were compared with the nucleotide sequences available at http://www.ncbi.nlm.nih.gov/. BLAST was performed with a minimum of 70% of nucleotide identity. Only scores comprising at least 30% of the size of the gene were considered.

***ccr* allotypes**: *ccr* types were assigned when the nucleotide sequence shared more than 85% identity with known *ccr* types, as previously suggested (111).

Identification of ψSCC*mec*: the contigs containing SCC*mec* elements of the early strain DNK22 were aligned with SCC*mec* IV (NCBI accession number AB063172) using the Microbial Genome Finishing Module of the CLC Genomics Workbench (Qiagen, Venlo, The Netherlands). The sequence of the IS*431* disrupting *mecA* was confirmed by sequencing a DNA fragment obtained by long-range PCR between *orfX* and *mecRI* (13 kb). The primers used for primer walking are listed on Supplementary Table 1. Manual annotation of the resulting contig containing the ψSCC*mec* structure was performed by BLAST analysis of the predicted ORFs with the nucleotide sequences available at http://www.ncbi.nlm.nih.gov/. WebACT was used to compare the sequence of ψSCC*mec* and SCC*mec* IV (http://www.webact.org/).

SCC/SCC*mec* nucleotide sequence comparison: the homology blocks carried by the SCC and SCC*mec* elements identified delimited by direct and/or inverted repeats were identified and aligned with Mauve (34). The most frequently found SCC*mec* types in *S. epidermidis*, SCC*mec* types I (NCBI accession number AB033763.2), II (NCBI accession number D86934), III (NCBI accession number AB037671), IV (NCBI accession number AB063172.2), V (NCBI accession number AB121219) and VI (NCBI accession number AF411935.3) were also used in the alignment.

Assessment of the carriage of antibiotic resistance genes: the online version of ResFinder (231)(https://cge.cbs.dtu.dk//services/ResFinder/) was used to search for the presence of antibiotic resistance genes in the contigs obtained with VELVET for all strains.

Assessment of the carriage of virulence genes: virulence genes were identified in the contigs obtained with VELVET by BLAST analysis of genes involved in biofilm formation, as described in (51). In addition, the two operons that are part of ACME-I (98), *arc* and *opp*, were searched by BLAST analysis.

Determination of the mobilome: the mobilome was considered to include insertion sequences, plasmids and intact prophages. Several online based servers were used to search for the presence of these elements for all strains, using the contigs obtained with VELVET; specifically, insertion sequences were searched with IS finder (232) (https://www-is.biotoul.fr/), *rep* genes with Plasmid finder (233) (https://cge.cbs.dtu.dk/services/PlasmidFinder/) and intact prophages with PHAST (234) (http://phast.wishartlab.com/).

Statistical analysis: Statistical significance of differences between proportions was evaluated by the Chi-square ($\chi 2$) test using a confidence interval of 95%.

Data deposition: the raw data was deposited in ENA.

RESULTS

S. epidermidis of the early-antibiotic era belong to 16 different clonal types

The earliest *S. epidermidis* isolates characterized so far were isolated in the 1970s (110). In this study we analyzed 21 *S. epidermidis* isolates collected in Denmark five years earlier, in 1965, which coincides with the beginning of the large-scale use of antibiotics in clinical practice.

The genomes of these 21 early *S. epidermidis* isolates were sequenced and single-nucleotide polymorphism (SNP) analysis of the genomes using ATCC12228 as a reference strain was performed. Considering a threshold value of 50000 SNPs, the isolates were included into three different clusters (Figure 1). Inside each cluster, the highest SNP difference was 5000-12000 SNPs. This huge degree of genetic diversity in the SNP tree was concordant with the MLST results: the MLST alleles of the seven housekeeping genes that compose the *S. epidermidis* MLST scheme (*arcC, aroE, gtr, mutS, pyrR*, *tpi* and *yqil*) were extracted and the 21 isolates were found to belong to 16 different sequence types (STs). Moreover, the majority was not present in the MLST database, and new STs were assigned (STs 409, 410, 411, 412, 413, 414, 417, 500, 501 and 502).

On the other hand, we found that isolates with the same ST (ST5, DNK3 and DNK8; ST6, DNK2 and DNK20; ST8, DNK11 and DNK12; ST225, DNK15 and DNK18; ST500, DNK1 and DNK13) clustered together in the SNP tree, confirming that they are related to each other. In addition, the core genome of these isolates with the same ST showed a low number of SNP differences (ST5, 103 SNPs; ST8, 6 SNPs; ST225, 150 SNPs; ST500, 76 SNPs), except for ST6 isolates with 5191 SNPs between their core genomes, indicating that this particular ST includes more distantly related isolates.

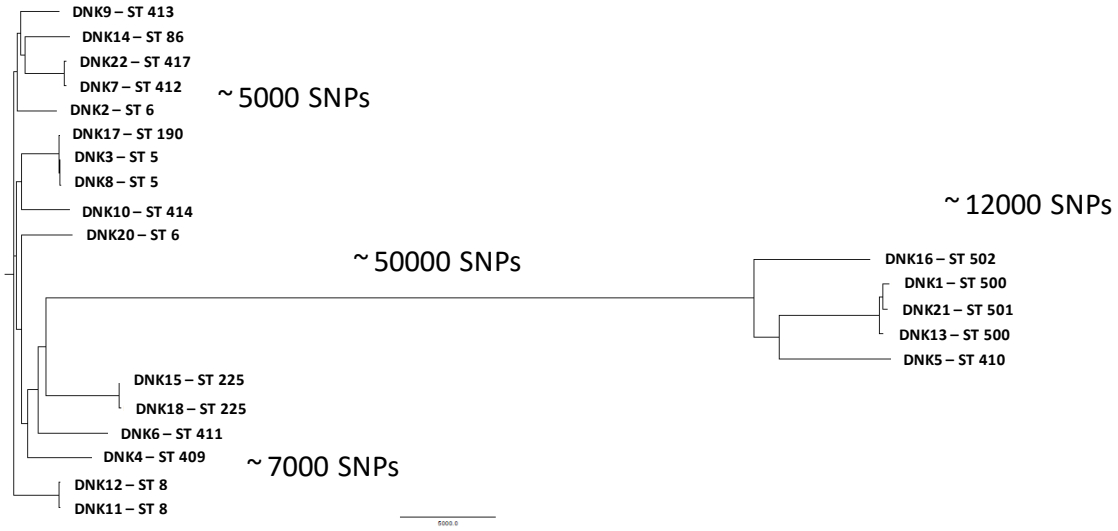


Figure 1. SNP tree of the predicted core genome of *S. epidermidis* isolates collected in Denmark in 1965. The tree reflects the number of SNP differences among the isolates. The average number of SNPs between the genomes of each cluster is shown. The reference strain used was *S. epidermidis* ATCC12228. Sequence types (STs) of each isolate is also shown.

The main contemporary *S. epidermidis* clonal lineage, CC2, was also prevalent in the early-antibiotic era

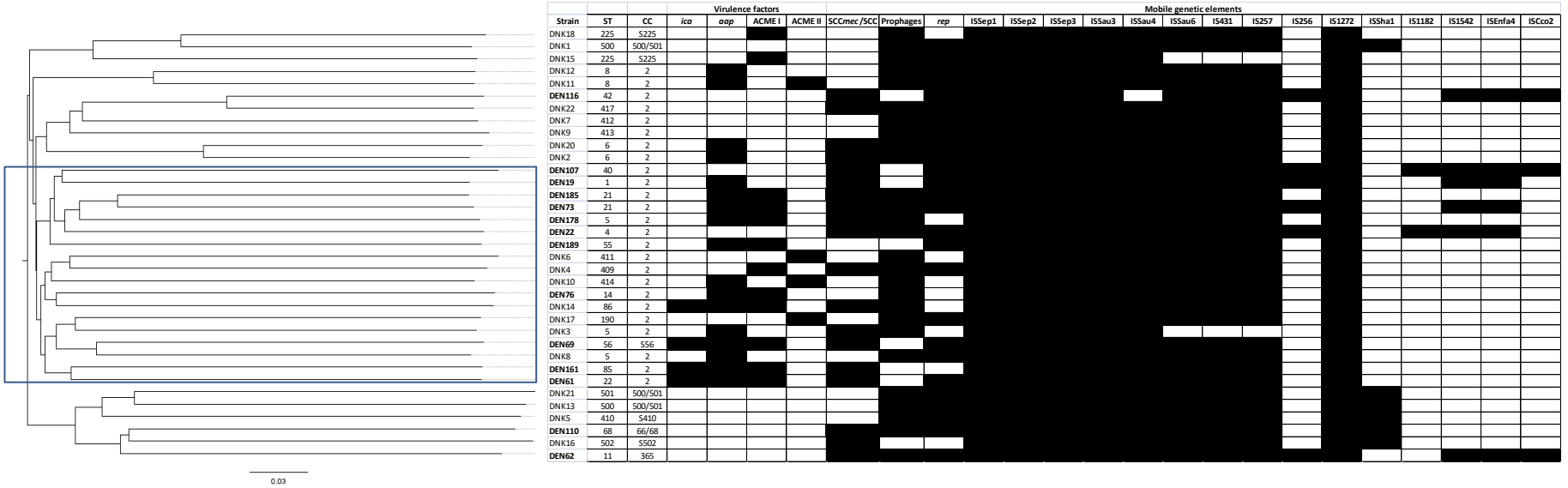
Most of the early isolates belonged to new sequence types, but although new, the STs identified were related with contemporary hospital-associated isolates, including ST2 (STs 5, 6, 8, 86, 190, 225). Interestingly, no isolate of the 1965 collection belonged to the presently most successful hospital-associated ST2. The relatedness of early and contemporary isolates was further evidenced when we compared the sequence types found among the early isolates to all sequence types available at www.mlst.net, through eBURST analysis (data not shown). Actually, the majority of the sequence types found among early isolates belonged to CC2 (ST5, -6, -8, -86, -190, -409, -411, -412, -413, -414 and -417). In addition, we found that ST5, a single-locus variant of ST2, was the only ST shared by early and contemporary isolates (17). This observation corroborates a previous suggestion that ST5 might correspond to the founder of CC2 and be the ancestral of the highly successful hospital-associated clonal type, ST2 (198).

To further understand the level of relatedness between early and contemporary strains, we predicted their core genome, by aligning the contigs obtained by VELVET with Mugsy, without a reference genome (Figure 2) and estimated the homology between them using Mauve. Our analysis showed that in fact the average degree of homology between the core genomes of early and contemporary isolates was extremely high, reaching 92.7%. When we considered only early isolates, the homology of the core genome of these isolates with the core of the entire collection (early and contemporary genomes) was lower, 89.4%, which reflects their primordial genome. Accordingly, if we consider the homology of the core of contemporary isolates with that of the entire collection, it increases to 97.5%.

Early *S. epidermidis* isolates had a susceptible heterogeneous profile to β-lactams

Contemporary hospital-associated *S. epidermidis* have been described to be highly resistant to β-lactams (2), associated with a high frequency of carriage of SCC*mec*. Nevertheless, the susceptibility to β-lactams in early *S. epidermidis* isolates has never been studied. We have determined the susceptibility of the early *S. epidermidis* isolates to two β-lactams, oxacillin and cefoxitin (Supplementary Figure 1). All isolates were highly susceptible to both β-lactams by eTest: the MIC to oxacillin ranged from 0.023 to 0.19 μg/ml and the MIC to cefoxitin, from 0.38 to 1.5 μg/ml. To determine in detail the resistance profile, cefoxitin population analysis profiles were performed and the great majority (18 out 21) strains showed a slightly hetereogeneous susceptibility profile, indicating that a small proportion of the population was able to grow at higher concentrations of cefoxitin (0.75-3 μg/ml). As expected, the contemporary isolates showed low susceptibility profiles to oxacillin as determined by disk diffusion (published before in (17, 79)).

A.



			Virulence factors				Mobile genetic elements															
Strain	ST	CC	*ica*	*aap*	ACME I	ACME II	SCC*mec*/SCC	Prophages	*rep*	ISSep1	ISSep2	ISSep3	ISSau3	ISSau4	ISSau6	IS431	IS257	IS256	IS1272	ISSha1	IS1182	IS1542	ISEnfa4	ISCco2
DNK18	225	S225			■			■		■	■	■	■	■			■		■					
DNK1	500	500/501			■			■		■	■	■	■	■			■		■	■				
DNK15	225	S225			■			■		■	■	■	■	■			■		■					
DNK12	8	2		■				■		■	■	■	■	■	■	■	■		■					
DNK11	8	2		■		■		■		■	■	■	■	■	■	■	■		■					
DEN116	42	2					■			■	■	■	■		■	■	■		■			■	■	■
DNK22	417	2					■	■		■	■	■	■	■	■	■	■		■					
DNK7	412	2						■		■	■	■	■	■	■	■	■		■					
DNK9	413	2						■		■	■	■	■	■	■	■	■		■					
DNK20	6	2		■			■	■		■	■	■	■	■	■	■	■		■					
DNK2	6	2					■	■		■	■	■	■	■	■	■	■		■					
DEN107	40	2					■			■	■	■	■	■	■	■	■	■	■			■	■	■
DEN19	1	2		■			■			■	■	■	■	■	■	■	■	■	■			■		
DEN185	21	2		■	■		■	■		■	■	■	■	■	■	■	■	■	■					
DEN73	21	2		■	■		■	■		■	■	■	■	■	■	■	■		■			■	■	
DEN178	5	2					■	■	■	■	■	■	■	■	■	■	■	■	■					
DEN22	4	2					■	■		■	■	■	■	■	■	■	■		■			■	■	■
DEN189	55	2		■	■		■			■	■	■	■	■	■	■	■	■	■					
DNK6	411	2				■		■	■	■	■	■	■	■	■	■	■		■					
DNK4	409	2			■		■	■		■	■	■	■	■	■	■	■		■					
DNK10	414	2		■		■	■	■	■	■	■	■	■	■	■	■	■		■					
DEN76	14	2		■	■		■	■	■	■	■	■	■	■	■	■	■		■					
DNK14	86	2	■	■	■		■	■	■	■	■	■	■	■	■	■	■		■					
DNK17	190	2		■		■	■	■		■	■	■	■	■	■	■	■		■					
DNK3	5	2		■	■		■	■	■	■	■	■	■	■	■	■	■		■					
DEN69	56	S56	■	■	■		■		■	■	■	■	■	■	■	■	■		■					
DNK8	5	2		■	■		■	■		■	■	■	■	■	■	■	■		■					
DEN161	85	2	■	■	■		■		■	■	■	■	■	■	■	■	■		■					
DEN61	22	2	■	■	■		■	■	■	■	■	■	■	■	■	■	■		■					
DNK21	501	500/501						■		■	■	■	■	■	■	■	■		■	■				
DNK13	500	500/501						■		■	■	■	■	■	■	■	■		■	■				
DNK5	410	S410						■		■	■	■	■	■	■	■	■		■	■				
DEN110	68	66/68					■	■		■	■	■	■	■	■	■	■		■	■				
DNK16	502	S502					■		■	■	■	■	■	■	■	■	■		■	■				
DEN62	11	365					■	■	■	■	■	■	■	■	■	■	■	■	■			■	■	■

Figure 2A. Comparison between the genetic backgrounds of *S. epidermidis* isolates collected in Denmark in 1965 and 1997/1998. **A.** Phylogenetic reconstruction of the core genomes of *S. epidermidis* isolates as defined by Mugsy. The distance as predicted by Mauve is shown. In addition, the frequency of virulence genes and mobile genetic elements are highlighted for each strain in black. The strain ID of contemporary isolates is highlighted in bold. In addition, the cluster englobing a higher number of virulent strains is highlighted in blue.

B.

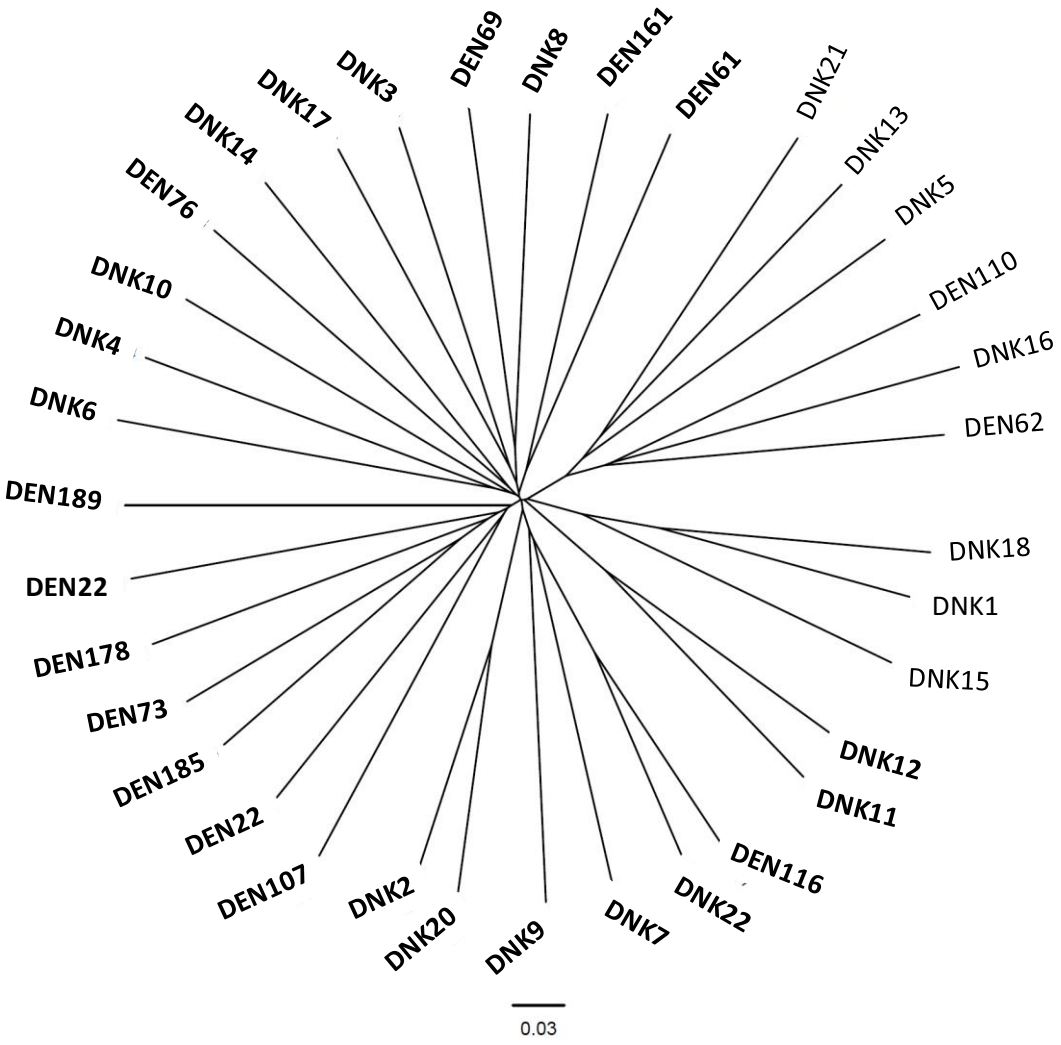


Figure 2B. Comparison between the genetic backgrounds of *S. epidermidis* isolates collected in Denmark in 1965 and 1997/1998. Radial representation of the tree showed in A. Isolates belonging to clonal complex 2 are highlighted in bold. DNK: early strains; DEN: contemporary strains.

SCC and ψSCC*mec* elements related with SCC*mec* were present in early *S. epidermidis*

The most frequent SCC*mec* type in *S. epidermidis* is SCC*mec* type IV, suggesting this species could have been the origin of this SCC*mec* element, but the contribution of *S. epidermidis* to SCC*mec* evolution is not known. On the other hand the acquisition of SCC*mec* was previously shown to be promoted in the hospital environment, suggesting this element could help *S. epidermidis* to adapt to this setting, but is still not clear

when was SCC*mec* first acquired by this species. To further explore the possible associations between *S. epidermidis* and SCC*mec*, we searched for the presence of SCC*mec* elements in the genome of early *S. epidermidis* isolates. The vicinity of *orfX* was annotated for 16 isolates collected in 1965. In the remaining five isolates, the elements were scattered among several small contigs or the *orfX* was located at the end of a long contig, which hindered the analysis.

The annotation of the *orfX* region of *S. epidermidis* strains collected in the pre-antibiotic era revealed a very complex and diverse genetic region (Supplementary Table 1, Figure 3A). Out of the 16 strains that were annotated, seven (DNK2, DNK3, DNK4, DNK14, DNK16, DNK20 and DNK22) (44%) had SCC and ψSCC elements flanked by direct and/or inverse repeats (DR/IR) and in two isolates, the elements were inserted in the chromosome in tandem (DNK3 SCC/DNK3 ψSCC; DNK20 ψSCC 1/ DNK20 ψSCC 2).

The comparison of the sequence of SCC and ψSCC found among early *S. epidermidis* isolates with SCC*mec* cassettes most often carried by contemporary *S. epidermidis* isolates (Figure 3C) showed that one strain, DNK22, carried a ψSCC*mec* that was closely related with SCC*mec* IV and SCC*mec* I. The regions of homology corresponded to *mec* complex and the J3 region (100% nucleotide identity of the elements carried in the same orientation and synteny as the ones found in SCC*mec* IV). The remaining ψSCC carried by early *S. epidermidis* isolates were less related with contemporary SCC*mec* cassettes, despite the finding of genes encoding SCC*mec*-associated hypothetical proteins inside these elements (Supplementary Table 2). On the other hand, the SCC carried by early *S. epidermidis* were more related with SCC*mec* II, having in common *ccrAB2* and related J1 regions, composed by genes encoding hypothetical proteins that were found to be carried in SCC*mec* II and SCC*mec* II composite islands identified in several staphylococcal species (Figure 3C and Supplementary Table 2).

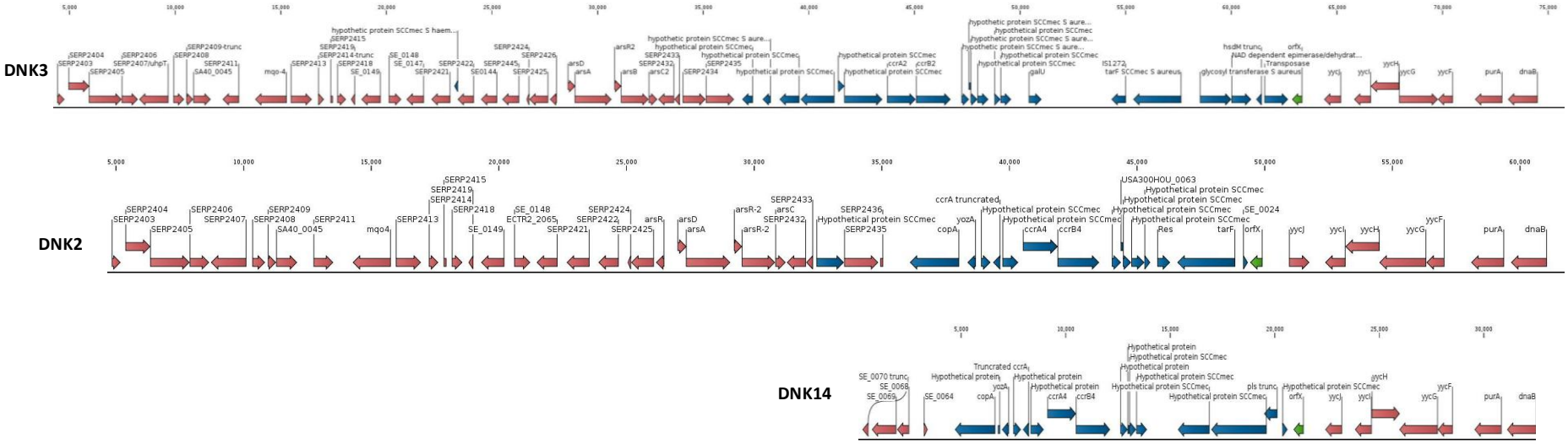


Figure 3A. Examples of SCC elements carried by ancestral *S. epidermidis* strains. Green arrows: *orfX*; blue arrows: elements related with SCC*mec* structures; pink arrows: elements related with housekeeping genes.

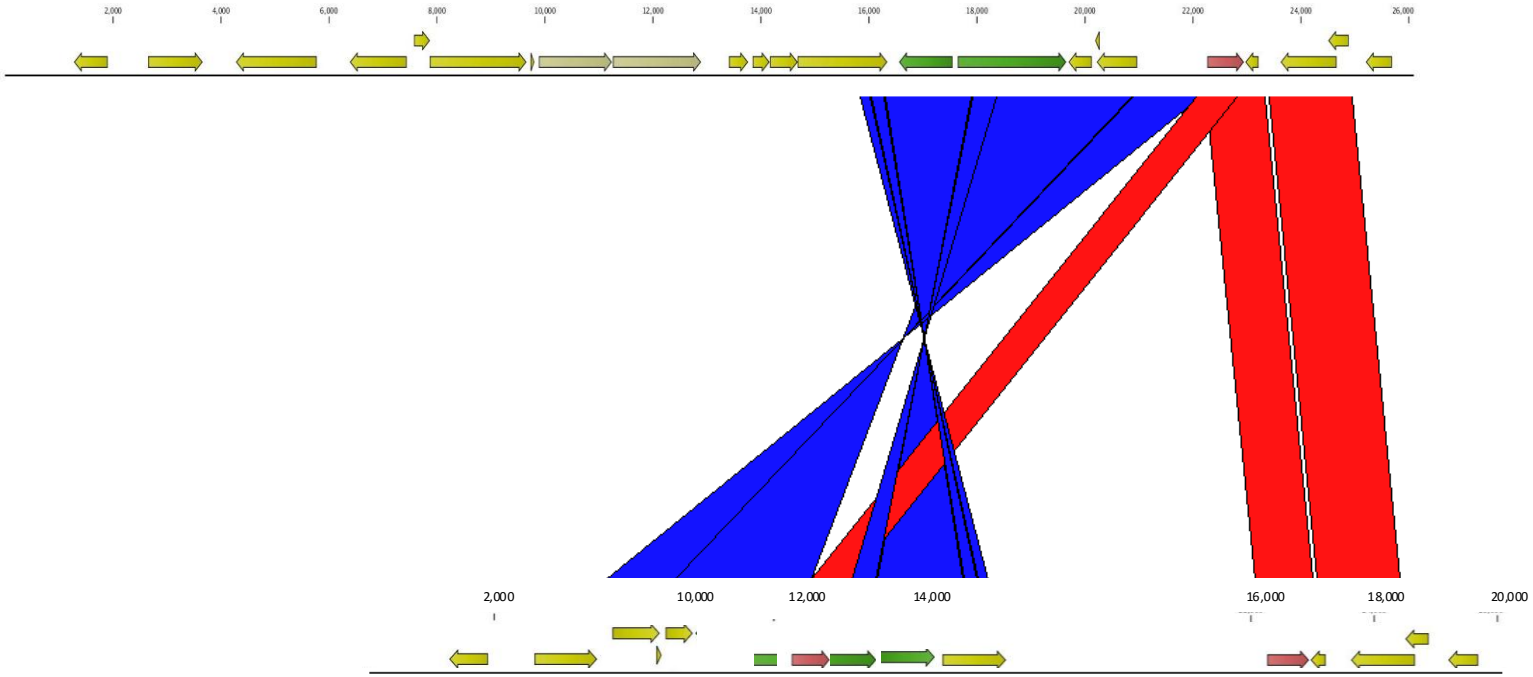


Figure 3B. Comparison of SCC*mec* IV and the ψSCC*mec* IV-like carried by DNK22 strain, with WEBACT. Red blocks: regions with high homology; blue blocks: regions with high homology and inverse sinteny. Green arrows: *mec* complex; grey arrows: *ccr* complex; pink arrows: IS*431*; light blue arrows: *orfX*. The remaining ORFs of each cassette are shown in yellow.

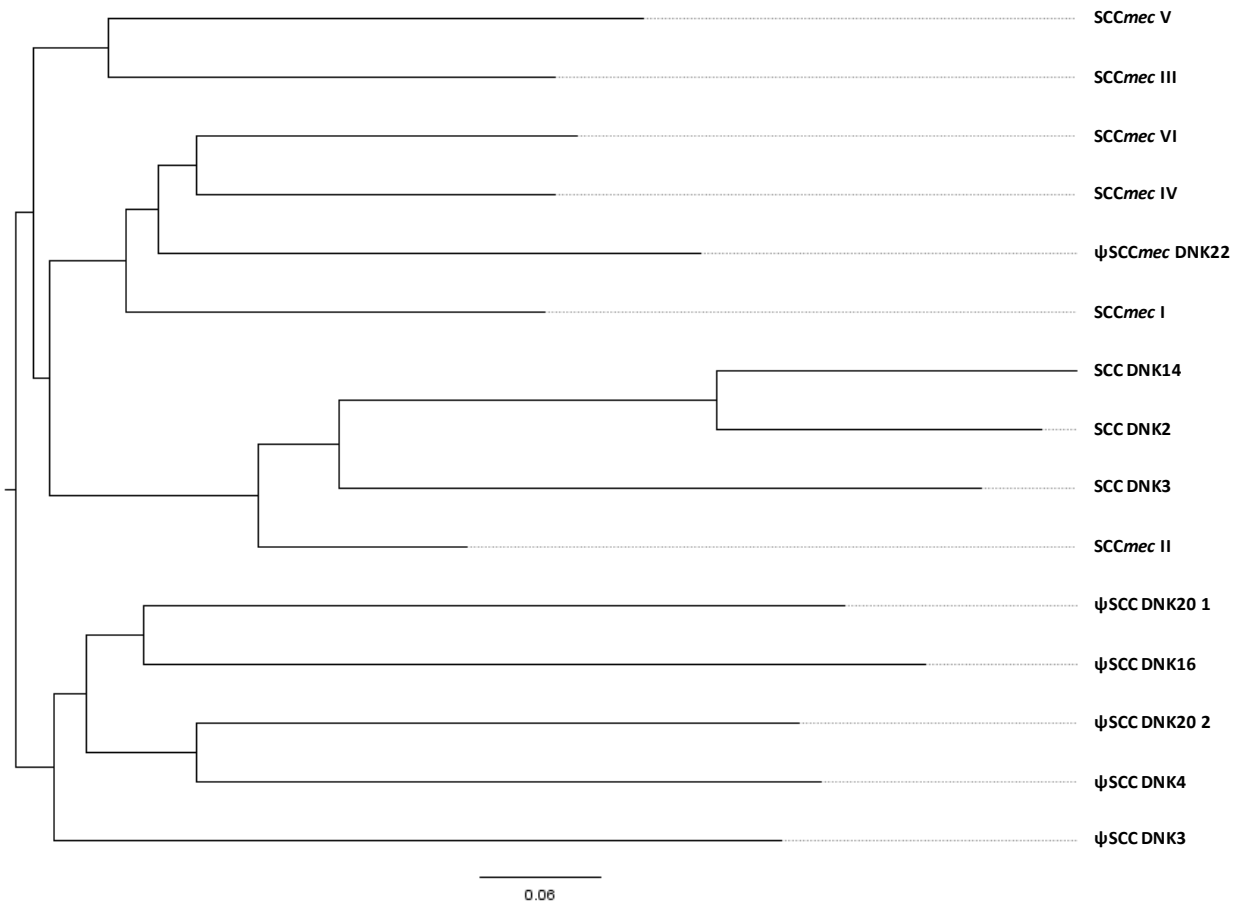


Figure 3C. Phylogeny reconstruction of the first SCC*mec* cassettes described until the early 2000s and the SCC cassettes carried by ancestral *S. epidermidis* strains.

From the nine isolates that had no DR/IR in the vicinity of the *orfX*, seven isolates contained several different SCC*mec*-associated structural elements, such as *ccr* and J1/J2/J3 genes encoding hypothetical proteins (data not shown). The genes encoding hypothetical proteins identified have been found mainly in contemporary SCC*mec* II, III, IV and V (Table 1). Moreover, some SCC*mec*-associated genes encoding hypothetical proteins were found in elements that belonged to distinct genetic backgrounds in the same chromosomal region, suggesting that they could be part of the core genomes of early isolates. The remaining two isolates did not carry any SCC*mec* structural elements in the vicinity of *orfX*. Noteworthy, as many as 29% of the early strains carried *ccr* genes, including those within a SCC or within *orfX*, outside any delimited element. The *ccrAB4* and related alotypes were the most frequent *ccr* types found among early *S. epidermidis* isolates (19%, four out of 21, DNK2, DNK11, DNK12 and DNK14), followed by *ccrAB2* (DNK3) and *ccrC* (DNK7) (5%, one isolate out of 21 each).

The great majority of the contemporary isolates included in this study (57%, eight out of 14) carried SCC*mec* IV in the *orfX* region. Regarding the remaining contemporary isolates, four isolates carried SCC*mec* (SCC*mec* I, III, V and a combination of *ccrAB2* with *mec* complex C2) and two isolates (MSSE) did not carry any SCC in their *orfX* region. In addition, the *orfX* region of MSSE did not carry any other structural SCC*mec* elements.

Overall, we found a high number of SCC*mec* IV structural elements among early *S. epidermidis* (Table 1); specifically, genes encoding hypothetical proteins carried in the J regions of the element. The nucleotide sequence of the genes found had 76-100% of nucleotide identity with the genes carried in SCC*mec* IV, which suggests that early strains might have played a role in the assembly of SCC*mec* IV or of a related intermediate structure.

Table 1. Distribution of SCC*mec* IV (reference sequence, NCBI accession number AB063172.2) structural elements among early *S. epidermidis* genomes.

Genes	% Nucleotide identity	Element
Hypothetical proteins (J3)	97.22-100% hypothetical protein KF234240 (*S. aureus* SCC*mec* IVa-ACME, upstream mecA)	*orfX* vicinity
Hypothetical proteins (J3)	100% hypothetical protein AB063172 (SCC*mec* IV J3 reference sequence)	ψSCCmec
Hypothetical proteins (J3)	87.46% hypothetical protein (BA000017 *S. aureus* SCC*mec* J3 IV) 75.77-81.83% hypothetical protein (HF569109 *S. aureus* J3 SCC*mec* IV) 97.78-100% hypothetical protein J3 SCC*mec* IVa *S. aureus USA300* (KF234240)	ψSCC
IS431	100% SCCmec IV reference sequence	ψSCC*mec*
mecA	99% SCCmec IV reference sequence	ψSCC*mec*
mecR1	99% SCCmec IV reference sequence	ψSCC*mec*
IS1272	100% SCCmec IV reference sequence	ψSCC; ψSCC*mec*
Hypothetical proteins (J2)	100% hypothetical protein J2 SCC*mec* (SCC*mec* IVa+ACME composite island – *S. aureus* KF175393) 94.12% hypothetical protein J2 SCC*mec* (SCC*mec* IVh – *S. aureus* HE681097)	SCC
Hypothetical proteins (J2)	100% hypothetical protein AB063172 (SCC*mec* IV J2 reference sequence)	ψSCC*mec*
ccrA2	98.2% *ccrA2* SCC*mec* IV reference sequence	SCC
Hypothetical proteins (J1)	98.44-100% hypothetical protein J1 SCC*mec* IVa/ACME *S. aureus* USA300 (KF175393)	ψSCC
Hypothetical proteins (J1)	85.8-100% hypothetical protein J1 SCC*mec* (SCC*mec* IVa+ACME composite island – *S. aureus* FR753166)	SCC

An interrupted *mecA* gene was carried by a single early isolate

The single early isolate that carried *mecA* gene, in the vicinity of the *orfX*, was DNK22. To our knowledge, this is the earliest description of a *S. epidermidis* carrying *mecA*. DNK22 strain carried a ψSCC*mec* element, flanked by DR/IR repeats but without *ccr* genes. In spite of the presence of the *mecA* gene in the chromosome of this strain, DNK22 was fully susceptible to the two β-lactam antibiotics tested. The *mecA* gene sequence was very conserved, showing 99% nucleotide identity with the sequence carried by the reference MRSE strain RP62A. However, DNK22 *mecA* nucleotide sequence was disrupted at position 727 with a copy of IS*431*. IS*431* was flanked by direct 8 bp repeats. In the vicinity of *mecA*, *mecRI* was identified upstream; a second copy of IS*431* was carried downstream *mecA*. In addition, we could identify a copy of *mecR1* disrupted by IS*1272*, downstream *mecA* (Supplementary Table 2). Therefore, DNK22 ψSCC*mec* carried *mec* complex B. BLAST analysis of the remaining ORFs of the

element with the nucleotide sequences available at http://www.ncbi.nlm.nih.gov/ has revealed several hits with genes carried by SCC*mec* IV. We performed a comparison of SCC*mec* IV with DNK22 ψSCC*mec* with WEBACT and we found that these elements were closely related (Figure 3B). Therefore, these results suggest that a SCC*mec* IV-like element was already carried by early *S. epidermidis* isolates, collected in the beginning of the antibiotic era.

Early hospital-associated *S. epidermidis* carried genes associated with antibiotic and heavy metal resistance

Contemporary hospital-associated *S. epidermidis* are resistant to several classes of antimicrobial agents. To determine if early hospital-associated isolates already carried genetic determinants of resistance, searches were performed using web-based software and BLAST analysis of the assembled *de novo* contigs (Table 2). Interestingly, most of the antibiotic resistance determinants carried by contemporary isolates were already present in early *S. epidermidis* isolates (Table 2). In fact, early isolates carried genes associated to resistance to several different antibiotics commonly used in clinical practice in the 1960s; some were highly prevalent (*blaZ*, 67%, *fosA*, 90%; *norA*, 100%) and others were found in a low frequency (*aaD*, 9.5%; *ermA*, 9.5%, *fusB*, 9.5%; *cat*, 14.3%, *tet(K)*, 14.3%).

A higher frequency of genetic determinants involved in antibiotic resistance was found among the contemporary isolates; specifically, *blaZ* (100%) and *fusB* (43%). Moreover, additional resistance determinants were identified exclusively among the contemporary isolates: *aac(6')-aph(2'')* (19%), *ermC* (14%), *tet (L)* (5%) and *lnu(B)* (5%). The resistance determinants *cat* (14.2%) and *tet(K)* (14.2%) were also carried by the contemporary isolates; in addition, *norA* and *fosA* were also found in the genomes of all contemporary isolates (100%).

In addition, the accumulation of such elements in a single strain, leading to multi-resistance (resistance to four or more antimicrobial agents), was more frequently found among contemporary isolates (Table 2) (11 of 14, 79%) than early

isolates (7 of 21, 33%). Moreover, at least in two multiresistant early strains, DNK11 and DNK12, the majority of the antimicrobial resistance determinants, namely *aadD*, *cat* and *tet (K)*, were identified in the vicinity of a *rep* gene, indicating that these two particular strains might be carrying a multi-resistance plasmid. However, we did not find any plasmid with the same genes deposited in the NCBI database.

In addition to antibiotic resistance genes, we searched for genes conferring resistance to heavy metals. We found that at least four early *S. epidermidis* isolates (DNK2, DNK5, DNK14 and DNK22) carried heavy metals resistance genes, such as *copA* (conferring resistance to copper) and *yozA* (conferring resistance to arsenic). In early strains, these elements were found inside SCC cassettes, flanked by direct repeats, being putatively mobile (Supplementary Table 1), but they were not identified among any of the contemporary *S. epidermidis* genomes.

Early hospital-associated *S. epidermidis* carried genes associated with virulence

The search for the presence of virulence genes was focused on the genes that were previously described to be associated to biofilm formation, namely, *ica*, *aap*, *bap*, *bhp, sdrG, sdrH* and *sdrF*. In addition, we searched for the autolysin *atlE*, the capsule operon *cap*, and the proteases *geh-1, geh-2, sepA, sspB* and *sspC*, that have been described to also influence biofilm formation in *S. epidermidis* (51).

The *atlE*, *sdrG*, *sdrH, geh-1* and *geh-2* were present in all 35 isolates analyzed, including early and contemporary isolates. The *cap* operon was also almost ubiquitous in our collection, since a single early isolate, DNK5, did not carry this element. In addition, *sspB* and *sspC* were absent from the genomes of all isolates. Regarding the remaining genes, we found that *bap* and *sdrF* were exclusively carried by contemporary isolates (7%, one isolate out of 14; 2 out of 14 isolates, respectively). *bhp* was equally abundant among both collections (early, 28.5%, 6 out of 21 isolates; contemporary, 21.4%, three out of 14 isolates). The *sepA* was ubiquitous among early isolates, while it was only carried by 10 out of 14 contemporary isolates (71.4%). The *ica* operon (all five genes, *icaR*, *icaA*, *icaB*, *icaC* and *icaD*) was carried by a single early

S. epidermidis isolate, DNK14 (5%) (Table 3), but its frequency was higher in contemporary isolates (21%, three out of 14 isolates). Also an increase in frequency was observed for *aap* that raised from 38% (eight out of 21 isolates) in early isolates to 64% (nine out of 14 isolates) in contemporary isolates. ACME-I is composed by two operons, *arc* and *opp-3*, while ACME-II is composed by the *arc* operon alone. Among the early isolates, 38% (eight isolates out of 21) carried ACME and half of these strains carried ACME-I while the other half carried ACME-II. Interestingly, the frequency of the ACME I in contemporary isolates was found to be much higher (57%, eight out of 14 isolates) and ACME II was not found at all among this population.

Statistical significance between the different proportions observed regarding the distribution of virulence factors was tested with the Chi-square test. We found that the only significant differences ($p<0.05$) were observed regarding the higher frequency of *sdrF* among contemporary isolates and the higher frequency of *sepA* among the early isolates.

To understand how virulence factors might be associated to evolution of *S. epidermidis*, we analysed the distribution of these virulence factors among the different genetic backgrounds (Figure 2B). Interestingly, we found that the genetic background of contemporary isolates, that carried biofilm-associated genes and ACME, clustered more frequently with the genetic background of early isolates carrying also these elements (highlighted in Figure 2B in blue). This result might indicate that the genetic backgrounds that later adapted to the hospital environment were already present in the 60s. Moreover, this adaptation might be related with the maintenance and accumulation of virulence factors in the genetic background of the strains.

Table 2. Genetic background of the *S. epidermidis* isolates collected in Denmark in 1965 and 1997/1998. The presence of antibiotic resistance genes and virulence genes is also shown.

Strain	Year of isolation	Colonization/ infection	ST	Antibiotic resistance genes	Predicted antibiotic resistance profile	SCC*mec* type	Virulence genes*
DNK1	1965	Colonization	500	*blaZ/norA/fosA*	$PEN^{R}FQL^{R}FOS^{R}$	-	*sepA/bhp*
DNK2	1965	Colonization	6	*norA/fosA*	$FQL^{R}FOS^{R}$	-	*aap/sepA*
DNK3	1965	Colonization	5	*blaZ/norA/fosA*	$PEN^{R}FQL^{R}FOS^{R}$	-	*aap/sepA*
DNK4	1965	Colonization	409	*norA/fosA*	$FQL^{R}FOS^{R}$	-	ACME I*/sepA/bhp*
DNK5	1965	Colonization	410	*str/blaZ/norA/fosB*	$AMN^{R}PEN^{R}FQL^{R}FOS^{R}$	-	*sepA/bhp*
DNK6	1965	Colonization	411	*norA/fosA*	$FQL^{R}FOS^{R}$	-	ACME II*/sepA*
DNK7	1965	Colonization	412	*blaZ/norA/fosA*	$PEN^{R}FQL^{R}FOS^{R}$	-	*sepA*
DNK8	1965	Colonization	5	*blaZ/norA/fosA/tetK/tetM*	$PEN^{R}FQL^{R}FOS^{R}TET^{R}$	-	*aap/sepA*
DNK9	1965	Colonization	413	*blaZ/norA/fosA/fusB/cat*	$PEN^{R}FQL^{R}FOS^{R}FUS^{R}CHL^{R}$	-	*sepA*
DNK10	1965	Colonization	414	*blaZ/norA/fosA*	$PEN^{R}FQL^{R}FOS^{R}$	-	*aap*/ACME II*/sepA*
DNK11	1965	Colonization	8	*aadD/spc/blaZ/norA/fosA/ermA/cat/tet(K)*	$AMN^{R}SPE^{R}PEN^{R}FQL^{R}FOS^{R}TET^{R}CHL^{R}MLS^{R}$	-	*aap*/ACME II*/sepA*
DNK12	1965	Colonization	8	*aadD/spc/blaZ/norA/fosA/ermA/cat/tet(K)*	$AMN^{R}SPE^{R}PEN^{R}FQL^{R}FOS^{R}TET^{R}CHL^{R}MLS^{R}$	-	*aap/sepA*
DNK13	1965	Colonization	500	*blaZ/norA/fosA*	$PEN^{R}FQL^{R}FOS^{R}$	-	*sepA/bhp*
DNK14	1965	Colonization	86	*norA/fosA*	$FQL^{R}FOS^{R}$	-	*ica/aap*/ACME I*/sepA*
DNK15	1965	Colonization	225	*norA/fosA*	$FQL^{R}FOS^{R}$	-	ACME I*/sepA/bhp*
DNK16	1965	Colonization	502	*norA/fosA/fusB*	$FQL^{R}FOS^{R}FUS^{R}$	-	*sepA*
DNK17	1965	Colonization	190	*str/blaZ/norA/fosB*	$AMN^{R}PEN^{R}FQL^{R}FOS^{R}$	-	ACME II*/sepA*
DNK18	1965	Colonization	225	*norA/fosA*	$FQL^{R}FOS^{R}$	-	ACME I*/sepA/bhp*
DNK20	1965	Colonization	6	*blaZ/norA/fosA*	$PEN^{R}FQL^{R}FOS^{R}$	-	*aap/sepA*
DNK21	1965	Colonization	501	*blaZ/norA/fosA*	$PEN^{R}FQL^{R}FOS^{R}$	-	*sepA*
DNK22	1965	Colonization	417	*ΔmecA/blaZ/norA/fosA*	$PEN^{R}FQL^{R}FOS^{R}$	ψSCC*mec*	-
DEN19	1997	Infection	1	*aac (6')-aph (2'')/blaZ/mecA/norA/fosA*	$AMN^{R}OXA^{R}PEN^{R}FQL^{R}FOS^{R}$	IV	*aap/sepA*
DEN22	1997	Infection	4	*blaZ/mecA/norA/fosA/fusB*	$OXA^{R}PEN^{R}FQL^{R}FOS^{R}FUS^{R}$	IV	*sepA*
DEN61	1997	Colonization	22	*blaZ/mecA/norA/fosA/fusB/ermC*	$OXA^{R}PEN^{R}FQL^{R}FOS^{R}FUS^{R}MLS^{R}$	III	*ica/aap*/ACME I
DEN62	1997	Infection	11	*aac (6')-aph (2'')/blaZ/mecA/norA/fosA/fusB*	$AMN^{R}OXA^{R}PEN^{R}FQL^{R}FOS^{R}FUS^{R}$	IV	-

Strain	Year of isolation	Colonization/ infection	ST	Antibiotic resistance genes	Antibiotic resistance profile	SCC*mec* type	Virulence genes
DEN69	1997	Infection	56	*blaZ/mecA/norA/fosA/fusB/cat/tetL*	$OXA^{R}PEN^{R}FQL^{R}FOS^{R}FUS^{R}CHL^{R}TET^{R}$	V	*ica*/*aap*/ACME I/*sepA*
DEN73	1997	Infection	21	*blaZ/mecA/norA/fosA*	$OXA^{R}PEN^{R}FQL^{R}FOS^{R}$	C/2	*aap*/ACME I/*bhp*/*sepA*
DEN76	1997	Infection	14	*blaZ/norA/fosA*	$PEN^{R}FQL^{R}FOS^{R}$	-	*aap*/ACME I/*sepA*
DEN107	1998	Infection	40	*aac (6')-aph (2'')/blaZ/mecA/norA/fosA/fusB/cat/tetK*	$AMN^{R}OXA^{R}PEN^{R}FQL^{R}FOS^{R}FUS^{R}CHL^{R}TET^{R}$	IV	*sdrF/sepA*
DEN110	1998	Colonization	68	*aadD/blaZ/mecA/norA/fosA/fusB/ermC*	$AMN^{R}OXA^{R}PEN^{R}FQL^{R}FOS^{R}FUS^{R}MLS^{R}$	IV	*bap/sepA*
DEN116	1998	Infection	42	*aac (6')-aph (2'')/blaZ/mecA/norA/fosA/ermC/cat*	$AMN^{R}OXA^{R}PEN^{R}FQL^{R}FOS^{R}MLS^{R}CHL^{R}$	I	*sepA*
DEN161	1998	Infection	85	*blaZ/mecA/norA/fosA/fusB*	$OXA^{R}PEN^{R}FQL^{R}FOS^{R}FUS^{R}$	IV	*ica*/*aap*/ACME I
DEN178	1998	Infection	5	*blaZ/mecA/norA/fosA*	$OXA^{R}PEN^{R}FQL^{R}FOS^{R}$	IV	*aap*/ACME I/*bhp*/*sepA*
DEN185	1998	Infection	21	*blaZ/mecA/norA/fosA/fusB/tetK*	$OXA^{R}PEN^{R}FQL^{R}FOS^{R}FUS^{R}TET^{R}$	IV	*aap*/ACME I/*bhp*/*sepA*
DEN189	1998	NA	55	*blaZ/norA/fosA/fusB/lnuA*	$PEN^{R}FQL^{R}FOS^{R}FUS^{R}MLS^{R}$	-	*aap*/ACME I/*sdrF*/*sepA*

***** all isolates carried in addition *atlE*, *sdrG*, *sdrH, geh-1, geh-2* and *cap* (with the exception of DNK5 that did not carry *cap*).

PEN: penicillin; OXA:oxacillin; FQL: fluoroquinolones; FOS: fosfomycin; FUS: fusidic acid; MLS: macrolides; AMN: aminoglycosides; CHL: cloranphenicol; TET: tetracycline; SPE: spectinomycin

Distinctive mobilomes of early and contemporary *S. epidermidis* isolates

The frequency of mobile genetic elements, namely plasmids, prophages and insertion sequences (IS), was compared for early and contemporary *S. epidermidis* isolates. The proportion of isolates carrying plasmid sequences, as indicated by the presence of *rep* genes, was similar in both collections (71%, 15 out of 21 early isolates; 79%, 11 out of 14 contemporary isolates). On the other hand, the presence of intact prophages was markedly different between both collections. While for the early isolates, intact prophages were carried by almost all isolates (95%, 20 out of 21), in the contemporary isolates, only half (7 out of 14) carried these elements (Figure 2B). By analyzing the predicted size and G+C content of the prophages, we found that the pool of prophages was very diverse (29 different phages in 20 early strains and seven different phages carried by seven contemporary strains). In fact, we were able to identify the same phage in only a pair of early strains (DNK11 and DNK12) and contemporary strains (DEN73 and DEN185).

In addition, we looked at the highest BLAST hit of the proteins encoded by the majority of the predicted phage genes. Based on this analysis, we found that the majority of the prophages that seemed to carry the same or related phage genes were restricted to early or contemporary genomes. However, we found some exceptions. Early strains DNK15 and DNK18 carried two related phages that were also identified in the contemporary strain DEN110 (similar to NCBI database entry NC_022758); and early strain DNK18 carried a prophage related with the one carried by contemporary strains DEN22, DEN73 and DEN178 (similar to NCBI database entry NC_008723). These prophages contained solely genes associated to the phage cycle.

The presence of different IS sequences was searched using the IS finder database. All strains carried a core set of IS elements in their genome, composed by IS*Sep1*, IS*Sep2*, IS*Sep3*, IS*Sau3* and IS*1272* (Figure 2B). IS*Sau4* was also almost ubiquitous (a single isolate out of the 35 studied did not carry this element). The high frequency of IS*1272*, a *S. haemolyticus* IS element and IS*Sau3*/IS*Sau4*, *S. aureus* IS elements, in early *S. epidermidis* isolates, indicates high frequency of genetic exchange between these species and raises the question on the source of these elements.

Interestingly, IS*431* and related IS sequences (IS*257* and IS*Sau6*) were also almost ubiquitous in our collection. Only two early isolates did not carry these elements (DNK1 and DNK3).

By comparing the distribution of different IS elements in both collections, we found that contemporary isolates have a higher number of different IS elements in their genomes than the early isolates. As many as 14 different IS elements in a single isolate were identified in the contemporary isolates, while in early isolates the maximum number of different IS elements per isolate was ten. In fact, specific IS elements were only identified in contemporary isolates, such as *S. aureus* IS*1182*, *Enterococcus faecium* IS*1542*, *Enterococcus faecalis* IS*Enfa4* and *Campylobacter coli* IS*CCo2*. Interestingly, all of these bacteria from which IS were putatively originated are nosocomial pathogens, like *S. epidermidis*. The accumulation of these IS elements in the genome of contemporary *S. epidermidis* isolates might be related with their interaction with other nosocomial pathogens at the hospital environment.

We analysed the number of copies of IS elements in a proportion of the early and contemporary strains (six out of the 35 strains) and we did not find an overall alteration of the number of copies of each IS element per genome (each IS is found on average in single copy, two or three copies). The IS elements were we found a great variation were IS*431* and IS*1272*, that ranged from none to until four copies per genome, but we did not find a correlation between number of copies and year of isolation of the strain.

Of particular interest was the finding of IS*256* exclusively among contemporary genomes (six out of 14 strains, 43%). This IS element has been associated with modulation of biofilm formation and resistance to aminoglycosides, which are particularly relevant among strains associated with infection in the hospital environment, such as the ones collected in the post-antibiotic era (78, 235).

DISCUSSION

Staphylococcus epidermidis, a harmless commensal, has established recently as a human pathogen associated to medical-devices associated infections (2, 51). In recent years, several studies have addressed its pathogenic potential (52, 84, 100, 236) and epidemiology (17, 228, 237). A single genetic background, the clonal complex 2 (CC2), was described to be predominant in colonization and infection in the community and hospital (17, 198, 228). However, how nosocomial setting environmental pressures contributed to *S. epidermidis* evolution towards a more pathogenic lifestyle, remained to be clarified.

To address this question we compared the genomes of nosocomial strains collected in the same geographic region in the beginning of the 1960s, wherein *S. epidermidis* was still rarely identified as etiological agent of infection, with strains collected in the end of the 1990s, wherein *S. epidermidis* was already a recognized infectious agent.

The characterization of the genetic background of early and contemporary *S. epidermidis* isolates by both MLST and phylogenetic analysis showed that isolates collected almost 40 years apart were surprisingly highly related regarding their core genome, belonging almost all to the so-called CC2. This homogeny might result from the fact that recombination between similar *S. epidermidis* isolates belonging to CC2 is favored over recombination with different strains, as previously suggested (42). Notwithstanding, some differences between the two collections were observed. One of the most striking observations was the absence from the early collection of ST2, the ST most frequently associated to infections nowadays. Although our study is the first to describe the population structure of *S. epidermidis* in the 1960s, a previous study analyzed isolates collected in the 1970s, in Canada, by multilocus sequence typing (110). In this former study, a different typing scheme was used, but one of the main sequence types carried by 1973/76 isolates identified was that correspondent to ST2 (former ST27), suggesting that ST2 might have emerged between 1965 and 1973.

Actually, although belonging to CC2, the great majority of sequence types identified in early isolates were new sequence types, which might indicate that they

correspond to ancestral genetic backgrounds that no longer exist in circulation. Only one sequence type, ST5, was found both in the early and in the contemporary collection, which was a single-locus variant of ST2. Although ST2 was initially considered as the founder of CC2, when MLST analysis was extended from the hospital settings to the community, ST5 has been suggested to be the founder of CC2 instead (198). The fact that ST5 was already present in the early isolates and ST2 was not further supports this hypothesis.

In contrast with the conservation observed in the core genome, we observed a high number of differences in the mobilome of early and contemporary isolates. In particular we found a much higher frequency of intact prophages in the early population. Excision and loss of prophages have been described to be induced by hospital-associated stresses such as contact with antibiotics, like it was described for *Enterococcus faecalis*, when exposed to fluoroquinolones (238). Also, loss of prophages has been associated to increased virulence features, like increase in colonization capacity and biofilm formation. Actually, recent studies have highlighted the role of prophages in the modulation/inhibition of biofilm formation in *S. epidermidis* (239-241). Since biofilm formation constitutes the main virulence factor of *S. epidermidis*, it is tempting to speculate that the loss of intact prophages, possibly promoted by antibiotics, might have resulted in the emergence of strains with increased biofilm formation ability in contemporary isolates.

Another mobile genetic element that showed to be differentially distributed between the two collections were insertion sequences, that varied overtime not only in diversity but also in the type of IS carried. In particular, we observed that the number of different IS elements in the genome of early isolates was lower than in the contemporary isolates and that some IS elements, like IS*256*, IS*1182*, IS*1542*, IS*Enfa4* and IS*CCo2* were exclusively present in contemporary strains. IS elements have been previously described to be related to an increased ability to acquire and maintain antibiotic resistance genes and can be part of transposons, carrying antibiotic and heavy metal resistance genes into mosaic multi-resistance plasmids (242). The accumulation of antibiotic resistance genes that we observed to have occurred overtime in this study may have been facilitated by the expansion of IS during

evolution of *S. epidermidis*. Another known possible activity of IS elements is their ability to control gene expression through insertion into promotors or through gene disruption. In particular, in *S. epidermidis* it has been demonstrated that IS*256* could modulate the biofilm formation (78, 86, 235). The acquisition of IS*256* by contemporary strains might have enabled a refined tuning and regulation of biofilm formation, which is an added value in a medical device related infection situation, in the hospital environment. On the other hand, previous comparative genomic studies have previously documented the loss of prophages and accumulation of IS elements during adaptation of *Bordetella* genera to the human host (243). As suggested by Parkhill and co-authors these events might occur together with loss of coding sequences and the consequent loss of regulatory functions that can result in alterations in the regulation of virulence.

We have also found that the distribution of different SCC elements, like ACME and SCC carrying heavy-metal resistance genes was remarkably different between the early and contemporary populations. Our data showed that in particular ACME I was already part of *S. epidermidis* chromosome in the 1960s, however it only became widely disseminated in the contemporary collection. This finding is in accordance with previous studies wherein ACME I was found to be the most frequent among a representative population of nosocomial *S. epidermidis* from different countries (100, 101). The increased prevalence of this element in contemporary strains might provide a higher capacity of dissemination to *S. epidermidis*, like was previously shown to occur in *S. aureus* (98), a characteristic that is highly advantageous for a nosocomial bacteria. On the other hand, SCC carrying heavy metal resistance genes were exclusively found in early isolates. The existence of these genes probably resulted from prolonged contact with a heavy-metal rich environments created by humans. In the context of human skin and hospital, the frequent use of antiseptics and desinfectants might have been the environmental pressure driving the maintenance of heavy-metal resistance genes. The existence of heavy metal resistance genes in early populations of other bacteria, like *S. aureus* was previously reported (244) and their loss in the contemporary isolates might result from changes in hygiene practices in the hospital environment (245).

As expected and as previously described (2), the contemporary nosocomial *S. epidermidis* isolates were found to carry genes conferring resistance to multiple antibiotics and have accumulated a higher number of antibiotic resistance determinants; however, most of these genes were already found among early isolates. In particular, early *S. epidermidis* isolates carried resistance determinants to penicillin, erythromycin, tetracycline, fusidic acid and spectinomycin which were antibiotics already used in clinical practice by 1965 (246-248). This same antimicrobial resistance pattern was previously found among early MRSA and MSSA isolates collected in Denmark in a similar time period (1960s) (249). Our genomic analysis revealed that some of these determinants were likely located in a plasmid. We hypothesize that the dissemination of this plasmid among *S. epidermidis* and *S. aureus* might have been promoted by contact with antibiotics in the hospital environment and provide a means a survival and persistence in this setting.

We found that *fosA*, encoding resistance to fosfomycin and *norA*, encoding resistance to fluoroquinolones were ubiquitously found among our collection of *S. epidermidis*; none of these genes were found in the close vicinity of a *rep* gene, which is an indication that they are part of *S. epidermidis* core genome. *fosA* encodes a glutathione S-transferase that inactivates fosfomycin and it has been described to be plasmid-encoded (250). Fosfomycin was introduced in the clinical practice in 1969 (251) and resistance to this antibiotic has been reported to be around 20-30% among coagulase-negative staphylococci (2). However, studies on the distribution of *fosA* gene among these fosfomycing-resistant isolates have not been performed. *norA* encodes an efflux pump and it usually chromosomally-encoded (252). Fluoroquinolones such as ciprofloxacin were introduced in the clinical practice later, in the late 1980s. Resistance to ciprofoloxacin has emerged in *S. aureus* soon after its introduction in the clinical practice (253) and among coagulase-negative staphylococci, resistance to this antibiotic has been described to be high (60-70%)(74). Taking together the high rates of resistance to fluoroquinolones, as well as to fosfomycin, among hospital-associated coagulase-negative staphylococci and the ubiquitous nature of *fosA* and *norA* among *S. epidermidis* collected in a period of 40 years, led us

to hypothesize if this species could be a reservoir for these genes and resistance to these particular antibiotics.

In our collection of early isolates we found that all isolates were fully susceptible to oxacillin and cefoxitin. However, the majority of strains contained a heterogeneous profile of susceptibility to cefoxitin, having subpopulation of cells able to grow to higher concentrations (up to 3 μg/ml). In spite of all early isolates in our collection being susceptible to β-lactams, we found that one isolate carried *mecA* that was disrupted by a copy of IS*431*, which probably explains the oxacillin susceptibility profile of the strain. In addition, IS*431* was flanked by 8 bp repeats, an indication that this IS was still mobilizable (254). The acquisition of *mecA* has been previously described to be dependent on the adaptation of the strain genetic background (255). The interruption of *mecA* by IS*431* might have been a strategy of early *S. epidermidis* to accommodate *mecA* in its chromosome. Alternatively, it could function as a switch on-off mechanism that enables *mecA* activation, when necessary. Annotation of the SCC element carrying this interrupted *mecA* copy has revealed that this was a ψSCC*mec*, highly similar to contemporary SCC*mec* IV, but missing *ccr*-encoded recombinases. SCC*mec* IV has been found in *S. aureus* isolates in the early 1980s only (123, 138), but it is the most frequent cassette carried by contemporary nosocomial *S. epidermidis* (79), and has been identified among isolates collected in 1973 (110). In addition, SCC*mec* IV carries *ccrAB2*, which is the *ccr* type most frequently carried among methicillin-susceptible *S. epidermidis* isolates (198)(Rolo and Miragaia, unpublished data). Moreover, in this current study we found that early MSSE in the *orfX* region and outside any SCC also carried genes encoding hypothetical proteins that are usually part of SCC*mec* III, IV and V. Interestingly, according to our own studies, these are the SCC*mec* types that are most frequently found among contemporary MRSE (79). Overall, our results suggest that the SCC*mec* IV element might have been first assembled in *S. epidermidis*, possibly through acquisition of an element containing *mec* complex B into an already formed SCC carrying *ccrAB2* and SCC*mec* IV genes encoding hypothetical proteins. The emergence of SCC*mec* IV might have occurred by recombination, which has been showed to occur very frequently in this species (42).

In this study, we found that *S. epidermidis* have suffered a high number of evolutionary events since they entered the hospital environment until they became important major nosocomial pathogen. These included, the selection of strains belonging to CC2, the acquisition of genes related to biofilm-formation and antibiotic resistance, the expansion of IS and the deletion of prophages from the chromosome. On the other hand we provided evidence that point towards a key role of *S. epidermidis* in the assembly of the first SCC*mec* type IV that later became widely disseminated in the contemporary population.

ACKNOWLEDGMENTS

J. Rolo was supported by fellowship SFRH / BD / 72675 / 2010 from Fundação para a Ciência e Tecnologia (FCT), Portugal. This study was partially supported by projects ESCMID research grant 2010, from European Society of Clinical Microbiology and Infectious Diseases, and project ref. PTDC/BIA-EVF/117507/2010 and R&D Unit ref. UID/CBQ/04612/2013 from Fundação para a Ciência e Tecnologia (FCT).

SUPPLEMENTARY DATA

Supplementary Table 1. List of primers used for sequencing a DNA fragment of the early *S. epidermidis* strain DNK22, obtained between *orfX* and *mecRI*. The fragment was obtained by long-range PCR and the sequencing was performed by primer walking.

Primer sequence	Target gene	Reference
orfXF1: GAAAAATATTGGAAGCAAGC	*orfX*	(124)
mecR4: GTCGTTCATTAAGATATGACG	*mecRI*	(125)
IS2: TGAGGTTATTCAGATATTTCGATGT	IS431	(127)
IS431 P1: AAGGAGTCTTCTGTATGAAC	IS431	unpublished. Oliveira.
IS431 P4: CAGGTCTCTTCAGATCTACG	IS431	(160)
promF1: GATAACACCTTCTACACCTCC	Promoter region of *mecA*	This paper
mecA P4: TCCAGATTACAACTTCACCAGG	*mecA*	(160)
mecA P7: CCACTTCATATCTTGTAACG	*mecA*	(160)
GLR2: CGAAGGTATCATCTTGTACCC	*mecA*	This paper
mecA GL F: TATGAGATAGGCATCGTTCC	*mecA*	This paper
mecA GL R: TTACCAATAACTGCATCATC	*mecA*	This paper
termR1: GCAACCATCGTTACGGATTGC	*mecA*	This paper
termF1: GGAATGGCTAGCTACAATGCC	*mecA*	This paper

Supplementary Table 2: Genes carried by SCC and ψSCC elements found in the *orfX* region of early *S. epidermidis* isolates.

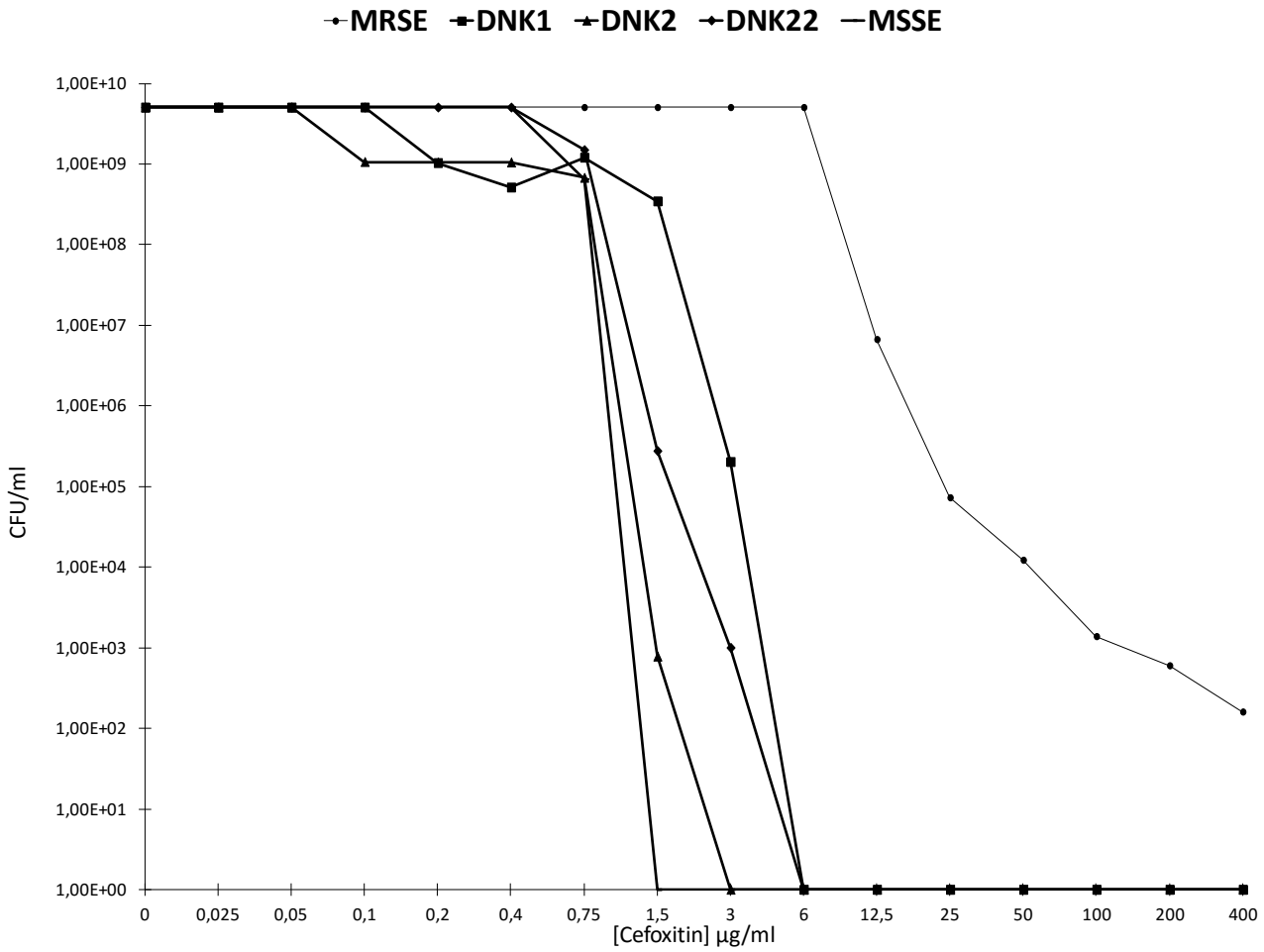
Element (size, bp)	Direct/inverted repeats	Region (size, bp)	Genes
SCC DNK3 (13546)	CTTGCATAGTATTGTATA	J1 (7058 bp)	36852-37349: Hypothetical protein 1: 100% hypothetical protein J1 SCC*mec* (SCC*mec* IVa+ACME composite island – *S. aureus* FR753166) 37818-38195: Hypothetical protein 2: 100% hypothetical protein J1 SCC*mec* (SCC*mec* IVa+ACME composite island – *S. aureus* FR753166) 38620-39552: Hypothetical protein 3: 100% hypothetical protein J1 SCC*mec* (SCC*mec* IVa+*ccrAB1* SCC – *S. aureus* HF569115) 39624-41207: Hypothetical protein 4: 85.8% hypothetical protein J1 SCC*mec* (SCC*mec* IVa+*ccrAB1* SCC – *S. aureus* HF569115) 41379-41675: Hypothetical protein 5: 100% hypothetical protein J1 SCC*mec* (SCC*mec* IVa+*ccrAB1* SCC – *S. aureus* HF569115) 41676-43468: Hypothetical protein 6: 100% hypothetical protein J1 SCC*mec* (SCC*mec* IVa+*ccrAB1* SCC – *S. aureus* HF569115)
		ccr	43702-45051: *ccrA2* 45073-46701: *ccrB2*
		J2 (3489 bp)	47223-47573: Hypothetical protein 1: 100% hypothetical protein J2 SCC*mec* (SCC*mec* IVa+ACME composite island – *S. aureus* KF175393) 47566-47658: Hypothetical protein 2: 100% hypothetical protein J2 SCC*mec* (SCC*mec* IVa+ACME composite island – *S. aureus* KF175393) 47660-47971: Hypothetical protein 3: 100% hypothetical protein J2 SCC*mec* (SCC*mec* IVa+ACME composite island – *S. aureus* KF234240) 47983-48489: Hypothetical protein 4: 94.12% hypothetical protein J2 SCC*mec* (SCC*mec* IVh – *S. aureus* HE681097) 48787-49065: Hypothetical protein 5: 66.67% hypothetical protein J3 SCC*mec* (SCC*mec* III – *S. aureus* CP006838) 49072-49569: Hypothetical protein 6: 60.68% hypothetical protein J3 SCC*mec* (SCC*mec* III – *S. aureus* CP006838)
ψSCC DNK3 (12690)	CTTCATACGATATACATA	-	50414-51009: *galU* (UTP-glucose-1-phosphate uridylyltransferase) 54334-55007: IS1272 55364-57631: *tarF S. aureus* (teichoic acid biosynthesis protein F) 58516-59997: glycosyl transferase 60010-60936: NAD dependent epimerase/dehydratase *S pseudointermedius* 61187-61438: *hsdM* truncrated 61572-62678: transposase
SCC DNK2 (14363)	AAAAACCGCCTCACTTACGATA	J1 (5412)	38029-36104: *copA* (copper transporting ATPase) 38685-38359: *yozA* (arsenical resistance operon repressor)

Element (size, bp)	Direct/inverted repeats	Region (size, bp)	Genes
			39251-38892: Hypothetical protein 1: 100% hypothetical protein J2 SCC266 – SCC266+SCC*mec*II composite island *S. aureus* AB774374) 39638-39363: truncrated *ccrA*: 78.3% *ccrA1* 40339-39725: Hypothetical protein 2: 100% hypothetical protein J2 SCC266 – SCC266+SCC*mec*II composite island *S. aureus* AB774374)
		ccr	40526-41887: *ccrA4* 41884-43512: New type *ccrB* (77% nucleotide identity with *ccrB4*)
		J2 (5936)	44363-44013: Hypothetical protein 3: 100% hypothetical protein J2 SCC266 – SCC266+SCC*mec*II composite island *S. aureus* AB774374) 44448-44356: 98.92% USA300HOU_0063 (hypothetical protein region between SCC*mec* IVa and ACME) 44761-44450: Hypothetical protein 4: 98.4% hypothetical protein J3 SCC4610 – SCC4610+SCC*mec*V composite island *S. aureus* AB773816) 45279-44776: Hypothetical protein 5: 76.59% hypothetical protein J3 SCC4610 – SCC4610+SCC*mec*V composite island *S. aureus* AB773816) 45514-45293: Hypothetical protein 6: 80.56% hypothetical protein J3 SCC*mec* –*mec* complex A+*ccrAB1 S. aureus* HF569114) 46290-45793: Res (spermidine acetyltransferase) 48846-46579: *tarF* (teichoic acid biosynthesis protein F) 49155-49337: SE_0024 (hypothetical protein ATCC12228, downstream *orfX*)
SCC DNK14 (17089)	CTAATCTAATCAAACAAAAA	J1 (4422)	4703-6628: *copA* 6757-6855: Hypothetical protein 1: 100% hypothetical protein J1 HE858191 (*S. aureus* composite island SCC*mec*IIE-SCCM1) 6959-7284: *yozA* 7491-7850: Hypothetical protein 2: 99.72% hypothetical protein J1 AB774374 (*S. aureus* composite island SCC266 upstream *yozA*) 7962-8237: truncrated *ccrA*: 78.3% *ccrA1* 8324-8938: Hypothetical protein 3: 100% hypothetical protein J1 AB774374 (*S. aureus* composite island SCC266 upstream *yozA*)
		ccr	9125-10486: *ccrA4* 10483-12111: New type *ccrB* (77% nucleotide identity with *ccrB4*)
		J2 (8489)	12614-12955: Hypothetical protein 4: 95.16% hypothetical protein JQ746621 (*S. aureus* composite island SCC*mec*V-SCC*ccrAB2* upstream *ccrAB2*) 12957-13046: Hypothetical protein 5: 37.5% hypothetical protein J1 AB121219 (*S. aureus* WIS SCC*mec* V downstream *ccrC*) 13048-13362: Hypothetical protein 6: 99.72% hypothetical protein HF569114 (*S. aureus* hypothetical protein J3 SCC*mec* –*mec* complex A+*ccrAB1* upstream *ccrB1*) 13377-13880: Hypothetical protein 7: 90.67% hypothetical protein KF049201 (SCC+SCC*mec*+SCC CRISPR composite island *S. capitis*, downstream 2nd copy of *ccrC*)

Element (size, bp)	Direct/inverted repeats	Region (size, bp)	Genes
			15373-16881: Hypothetical protein 8: 68.05% hypothetical protein SH0041 AP006716 (*S. haemolyticus* SCC*mec* V upstream *ccrC*) 16970-19609: Hypothetical protein 9: 41.81% hypothetical protein SH0040 AP006716 (*S. haemolyticus* SCC*mec* V upstream *ccrC*) 20367-20600: Hypothetical protein 10: 97.86% hypothetical protein JQ746621 (*S. aureus* composite island SCC*mec*V-SCC*ccrAB2* upstream *pls*)
ψSCC*mec* DNK22 (20521)	AAAACCGCATCATTTA	J2 (10284)	137314-138786: multicopper oxidase 138762-140963: *copA* 143509-143952: ORF
		mec complex (7495)	148241-148855: IS1272 147032:148006: *mecRI* 144128-1469332: Δ*mecA* 145416-146206: IS431 14654-144082: Hypothetical protein 2: 100% hypothetical protein AB063172 (SCC*mec* IV J2 reference sequence) 144148-144228: Hypothetical protein 3: 100% hypothetical protein AB063172 (SCC*mec* IV J2 reference sequence) 144179-144922: Hypothetical protein 4: 100% hypothetical protein AB063172 (SCC*mec* IV J2 reference sequence) 153941-154527: IS431
		J3 (2742)	154647-154798: Hypothetical protein 5: 100% hypothetical protein AB063172 (SCC*mec* IV J3 reference sequence) 155556-155505: Hypothetical protein 6: 100% hypothetical protein AB063172 (SCC*mec* IV J3 reference sequence) 156182-156468: Hypothetical protein 7: 100% hypothetical protein AB063172 (SCC*mec* IV J3 reference sequence) 156878-157269: Hypothetical protein 8: 100% hypothetical protein AB063172 (SCC*mec* IV J3 reference sequence)
ψSCC DNK4 (27016)	GAAGTTGAAAATAAAAT	-	12194-14352: Hypothetical protein 1-truncrated: 65.70% hypothetical protein J3 AB121219 (SCC*mec* V – WIS) 14739-14906: Hypothetical protein 2: 90.48% hypothetical protein KF049201 (*S. capitis SCCmec*-SCCCRISPR composite element) 15546-15758: Hypothetical protein 3-truncrated: 64% hypothetical protein J1 AB12767 (S. aureus SCC*mec* V) 15841-17265: Hypothetical protein 4: 99.72% hypothetical protein SH0069 (AP006716 S. haemolyticus SCC*mec* V) 17333-17506: Hypothetical protein 5: 98.85% hypothetical protein SA957_0061 (CP003603 *S. aureus* SCC*mec* V) 17986-18300: Transposase SA40_0044

Element (size, bp)	Direct/inverted repeats	Region (size, bp)	Genes
			18312-19127: Transposase SA40_0045 19227-20156: *arc* 20176-21174: *arcB* 21212-21901: *arcR* 21943-23364: *arcD* 23450-24685: *arcA* 24954-25400: *argR* 26210-26569: *hsdR*-truncrated 26807-27499: Hypothetical protein ACME 27514-29049: *opp-3A* 29052-30008: *opp-3B* 30008-30775: *opp-3C* 30742-31509: *opp-3D* 31502-32137: *opp-3E* 32776-35568: *hsdR* 35660-36241: *hsdS* 36242-37798: *hsdM* 37791-39035: Type restriction/modification protein-truncrated
ψSCC DNK16 (12354)	CAACTTATTTTTTAGTTTTATTTGTGAT	-	20251-21228: Hypothetical protein 1: 87.46% hypothetical protein (BA000017 *S. aureus* SCC*mec* IV) 21270-22020: Hypothetical protein 2: 75.77% hypothetical protein (HF569109 *S. aureus* SCC*mec* IV) 22054-23127: Hypothetical protein 3: 81.83% hypothetical protein (HF569109 *S. aureus* SCC*mec* IV) 23265-23939: *tnp* 24066-24281: putative reductase 24322-25008: protein plasmid *S. simulans* 25550-28231: SERP0245 28255-29199: SERP0246 29549-30118: SERP0247 30637-31002: tnp truncrated 31103-31921: Hypothetical protein 4-truncrated: 98.05% hypothetical protein (KF234240 *S. aureus* SCC*mec* IV) 31958-32284: Hypothetical protein 5: 100% hypothetical protein (KF234240 *S. aureus* SCC*mec* IV)
ψSCC DNK20 1 (15049)	ACCATATGTTTTTTAGTTTTATTTGTGATACGCTTCGCCT	-	15683-17239: *arcD* 17365-18312: SE_0102 18313-19311: *arcB* 19349-20044: *arcR*

Element (size, bp)	Direct/inverted repeats	Region (size, bp)	Genes
			20780-21232: *argR* (ACME) 22164-22262: Hypothetical protein 1: 98.99% SH0027 J3 SCC*mec* V (*S. haemolyticus* AP006716) 22524-23495: Hypothetical protein 2: 98.99% SH0026 J3 SCC*mec* V (*S. haemolyticus* AP006716) 24199-25308: Hypothetical protein 3: 58.64% hypothetical protein SCC*mec* V (upstream 2nd copy of *ccrC S. aureus* GQ902038) 28066-28191: Hypothetical protein *S. aureus* plasmid 28348-28707: Hypothetical protein 4: 97.78% protein J3 SCC*mec* IVa *S. aureus USA300* (KF234240) 28946-29719: Hypothetical protein 5: 100% protein J3 SCC*mec* IVa *S. aureus* USA300 (KF234240)
ψSCC DNK20 2 (6234)	TTTTAGTTTTATTTGTGATACGCTTC	-	31453-32556: *hsdM* 32581-33255: *tnp* 33569-33898: Hypothetical protein 1: 99.09% protein J1 SCC*mec* IVa/ACME *S. aureus* USA300 (KF175393) 33990-34178: Hypothetical protein 2: 98.44% protein J1 SCC*mec* IVa/ACME *S. aureus* USA300 (KF175393) 34275-35240: Hypothetical protein 3: 100% protein J1 SCC*mec* IVa/ACME *S. aureus* USA300 (KF175393) 35734-35904: Hypothetical protein 4: 100% protein J1 SCC*mec* IVa/ACME *S. aureus* USA300 (KF175393)



Supplementary Figure 1. Cefoxitin susceptibility population analysis profiles of representative early *S. epidermidis* strains. MRSE RP62A and MSSE ATCC12228 were used as control strains.

Strategies of adaptation of *Staphylococcus epidermidis* to hospital and community: amplification and diversification of SCC*mec*

Joana Rolo, Hermínia de Lencastre, and Maria Miragaia

Journal of Antimicrobial Chemotherapy. 2012 Jun;67(6):1333-41

All the experimental work was performed by J. Rolo.

ABSTRACT

Objectives. *Staphylococcus epidermidis* is a harmless commensal but, it can become a human pathogen, mainly in the hospital environment. In order to clarify strategies used by these bacteria to adapt to hospital environment, we compared the population structure and SCC*mec* contents of *S. epidermidis* from the community and hospital.

Methods. *S. epidermidis* were collected from nasal swabs of both healthy military draftees (192 isolates) and patients (94 isolates) recovered in the same time period and geographic region. *S. epidermidis* were characterized by pulsed-field gel electrophoresis (PFGE), multilocus sequence typing (MLST) and staphylococcal chromosomal cassette *mec* (SCC*mec*) typing.

Results. Clonal complex 5 (CC5) was predominant in hospital (100%) and community (58%), but some clonal types were specific of each environment and others were found in both (C/H clones). Methicillin-resistant *S. epidermidis* (MRSE) colonization rate in the community was very low (7%) when compared to the hospital (30%, $p<0.05$). Community-associated MRSE (CA-MRSE) carried mostly SCC*mec* IV and V (Simpson Index of Diversity, SID = 57.52; 95% CI=[38.35-76.69]), whereas hospital-associated MRSE (HA-MRSE) carried as many as 18 SCC*mec* structures (SID=82.67; 95% CI=[77.38-87.96]). Isolates of the same PFGE type had a much higher number of different SCC*mec* types when collected in the hospital than in the community.

Conclusions. Our data suggests that the *S. epidermidis* population is composed of hospital-associated clonal types, community-associated clonal types and types that are able to survive in both environments. Moreover, adaptation to the hospital environment in *S. epidermidis* appears to promote an increase in the frequency and diversification of SCC*mec*.

INTRODUCTION

Staphylococcus epidermidis is one of the main colonizers of the human skin, but can turn into a pathogen, if the cutaneous barrier is broken or the host is compromised (51). *S. epidermidis* is one of the most important pathogens in the hospital environment, being responsible for 40-90% of infections associated to indwelling devices (5).

The success of *S. epidermidis* as a pathogen is mainly linked to its capacity to form biofilm, a multistep process involving several genetic determinants (*atlE*, *aap* and *ica* operon) (51). In addition, *S. epidermidis* has the ability to accumulate multiple antibiotic resistance determinants (5, 51). In particular, the frequency of methicillin-resistant *S. epidermidis* (MRSE) can reach 80% in hospitals worldwide,(256) although is considerably lower (30-40%) in countries like Denmark and Iceland where methicillin-resistant *Staphylococcus aureus* (MRSA) frequency is below 1% (79).

Methicillin resistance is conferred by the *mecA* gene, which is carried in a family of mobile genetic elements called SCC (staphylococcal chromosome cassette)(109). In *S. aureus* SCCs insert in a unique specific chromosomal site (*orfX*), have characteristic inverted and direct repeats and contain chromosome cassette recombinases (*ccr*) that are responsible for SCC mobility. SCC*mec* is composed of the *mec* complex containing the *mecA* and its regulators, and the *ccr* complex, containing one (*ccrC*) (193) or two recombinases (*ccrAB*) (109). Up until now eleven major types of SCC*mec* (I-XI) (111, 113, 114, 257) and eight subtypes of SCC*mec* IV (111, 165) have been described in *S. aureus*. The few data available suggest that in *S. epidermidis*, SCC*mec* is also inserted in the *orfX* (52, 124) and has a structure similar to that described for *S. aureus* (52)' (110). However several studies have demonstrated that there is a large pool of uncharacterized SCC*mec* types among *S. epidermidis* and other coagulase-negative staphylococci (13, 258). Besides SCC*mec*, other SCC elements have been described in staphylococci that transport genes important for survival and virulence (99, 100, 137).

The population structure of *S. epidermidis* in the hospital environment (1996-2001) was shown to be composed by a major and highly diverse genetic lineage disseminated worldwide (clonal complex 2, CC2) and several minor CCs (17). In this

population, SCC*mec* IV (13, 110) was found to be the most frequent type, but other SCC*mec* types (13, 258) and a high number of novel cassettes have been identified as well (13, 258). Nosocomial *S. epidermidis* were described to carry several *ccr* alleles, that were suggested to correspond to the acquisition of multiple SCC *in tandem.* (124) It was also shown that these isolates had a higher estimated rate of recombination/mutation (2.5:1) and a higher frequency of SCC*mec* acquisition than *S. aureus* (17). Additionally, nosocomial isolates, particularly those belonging to CC2 were found to be associated to the presence of the *ica* operon (77) and the arginine catabolic mobile genetic element (ACME I.02),(100) considered a virulence factor in the community-associated *S. aureus* (CA-MRSA) strain USA300.

Much less is known on the epidemiology of community-associated *S. epidermidis* (CA-SE). The few studies available describe frequencies of nasal colonization with CA-MRSE of 20% among children (80) and military personnel (82). The molecular characterization of CA-SE, showed a high genetic diversity as illustrated by the high number of types of PFGE (pulsed-field gel electrophoresis) found; however dissemination of *S. epidermidis* epidemic strains was also observed to occur in this setting (82). Regarding SCC*mec* distribution, the few studies available showed that SCC*mec* type IVa was the most prevalent among CA-MRSE, but other types were also found (80, 236). However, the population structure of *S. epidermidis* in the community and hospital and the frequency of antibiotic resistant genes have never been compared before.

In the study described here, we addressed this question by comparing *S. epidermidis* from the hospital and the community, collected from the same ecological niche, same time period and geographic location, in terms of genetic background and contents of SCC*mec*.

METHODS

Study population. A total of 1,483 Air Force draftees from different regions in Portugal and attending Centro de Formação da Ota (Lisbon, Portugal) were swabbed in the anterior nares. This study was conducted in four consecutive years (1996-1999). Each draftee filled a questionnaire assessing demographic data and the presence of risk factors for carriage of antibiotic resistant staphylococci (namely, recent antibiotic consumption, reason for antibiotics prescription, and contact with animals) and recent contact with the hospital (specifically, recent emergency department attendance and previous hospitalization or surgery) (see supplemental data S1). A group of 1,160 draftees did not take any antibiotics and had no contact with the hospital in the three months prior to sampling and were considered as being "healthy draftees".

In addition, 253 patients attending the Medicine (160 patients) and Orthopedics services (93 patients) of Hospital da Força Aérea (Lisbon, Portugal) were swabbed in the anterior nares during two sampling periods (2000-2001). This hospital is a private hospital with only 86 beds and presents particular characteristics: it does not contain an emergency department, the surgeries are all scheduled, the average period of internment is of 11 days and the prevalence of MRSA is extremely low: 3% in Medicine and 0.9% in Orthopedics (I. Santos-Sanches, FCT/UNL, personal communication).

Ethical statement. The nasal screening of patients from Hospital da Força Aérea, (Lisbon, Portugal) was performed with approval from Local Medical Ethical Committee and the screening of draftees attending Centro de Formação da Ota (Lisbon, Portugal) was performed with written informed consent. Patient records were de-identified and analyzed anonymously and the strains, not human subjects, were studied.

Bacteria isolation. The swabs obtained from the 1,160 "healthy draftees" were streaked onto Mannitol Salt Agar (MSA, Difco, BBL, Becton Dickinson, Franklin Lakes, New Jersey, USA) and incubated 24h at 37ºC. Bacterial isolates were further tested for

coagulase production using the Staphytec Plus assay (Oxoid, Cambridge, United Kingdom). A total of 736 healthy draftees were colonized with coagulase-negative staphylococci (CoNS) of which 199 were selected for further study to include the highest diversity in terms of geographic origin, gender and smoking habits. A total of 170/199 draftees were colonized with *S. epidermidis*, from whom 192 *S. epidermidis* isolates were collected (some draftees were colonized with more than one strain). These 192 *S. epidermidis* isolates were considered as having a community origin (CA-SE) and were included in this study.

Additionally, *Staphylococcus* were isolated from the nasal swabs obtained from 253 patients by growth on mannitol salt agar as previously described (259). The presence of the coagulase enzyme was assessed for all staphylococcal isolates by the Staphytec Plus assay (Oxoid, Cambridge, United Kingdom). A total of 94 hospital patients out of 253 were colonized with *S. epidermidis*. These isolates were considered as being associated to the hospital (HA-SE).

Species identification. *S. epidermidis* strains were identified by internal transcribed spacer PCR (ITS-PCR) (260).

***mecA* detection.** The presence of the *mecA* gene was detected by PCR amplification for all isolates (124). *S. epidermidis* isolates carrying the *mecA* were considered as methicillin-resistant *S. epidermidis* (MRSE) and those lacking the *mecA* were considered methicillin-susceptible *S. epidermidis* (MSSE), regardless of the oxacillin susceptibility results obtained.

***S. epidermidis* nasal colonization rate in the community.** *S. epidermidis* colonization rate in the community analyzed in this study was estimated by calculating: (i) the rate between the number of draftees colonized with *S. epidermidis* and the total number of draftees selected for study – colonized with CoNS (170/199=0.854) (85.4%); (ii) the number of draftees colonized with *S. epidermidis* in the total of healthy draftees

colonized with CoNS (736), considering that the rate is the same as obtained in (i) (0.854x736=628 draftees). The rate between the number obtained in (ii) and the total of healthy draftees screened (1,160) was considered the CA-SE colonization rate (628/1160x100=54.1%). The same approach was applied to estimate CA-MRSE nasal colonization rate (see sections Study population and Bacteria isolation above).

Antimicrobial susceptibility testing. Antimicrobial susceptibility testing was performed to penicillin, oxacillin, erythromycin, clindamycin, vancomycin, tetracycline and ciprofloxacin (Oxoid, Cambridge, United Kingdom) according to the guidelines of the Clinical and Laboratory Standards Institute (CLSI) (194).

DNA preparation. Agarose disks for PFGE and DNA for PCR were prepared as described (79, 124, 201).

PFGE. The SmaI DNA restriction fragments were separated by PFGE (201) and resulting patterns were analyzed using the BioNumerics software (version 4.61 of Applied Maths, Saint-Martens-Latem, Belgium) with previously optimized settings for *S. epidermidis* (13).

MLST. Multilocus sequence typing (MLST) was performed for one isolate of each PFGE type in the case of CA-SE isolates and for one isolate of each of the most representative PFGE types (containing more than four isolates) in the case of HA-SE. The MLST data was analyzed using the goeBURST algorithm (http://goeBURST.phylowiz.net).

SCC*mec* typing. The SCC*mec* type was determined by the combination of the class of *mec* complex and the type of *ccr* complex as previously suggested (111). In the

detection of *ccr* genes, the following strains were used as positive controls: *S. aureus* COL (*ccrAB1*),(261) *S. epidermidis* RP62A (*ccrAB2*),(52) *S. aureus* ANS46 (*ccrAB3*),(261) *S. epidermidis* ATCC12228 (*ccrAB4*),(99) and *S. aureus* WIS (*ccrC*) (127). To determine the class of *mec* complex, the strains *S. aureus* N315 (*mec complex* A)(193), COL (*mec complex* B) (125) and WIS (*mec* complex C) (127) were used as controls.

The subtype of SCC*mec* IV was determined by multiplex PCR as described by Milheiriço *et* al (165). SCC*mec* was considered non-typable when either *mec* complex or *ccr* complex, or both, were non-typable by the methods used or when the isolate carried more than one *ccr* type. SCC*mec* was considered to be new if a new combination of *mec* complex and *ccr* complex was found.

Clonal type definition. *S. epidermidis* clonal types were defined by the association of PFGE type and SCC*mec* type for MRSE as previously proposed (13) and by PFGE type alone for MSSE (13).

Statistical analysis. The degree of genetic diversity was assessed by the Simpson's index of diversity, using a confidence interval of 95% (215). In this analysis, each PFGE type or SCC*mec* subtype was considered a "type or a species". Statistical significance of differences between proportions was evaluated by the Chi-square (χ^2) test using a confidence interval of 95%.

RESULTS

Frequency of nasal colonization of *S. epidermidis* and methicillin-resistant *S. epidermidis* in the community

Out of 199 draftees selected, 170 were colonized with at least one *S. epidermidis* isolate (n=192 isolates), which corresponds to a colonization rate of *S. epidermidis* among healthy people of 54%. However, we found a low number of MRSE (18 isolates carrying the *mecA*) in the community, which corresponds to a colonization rate of 7%. Moreover, we observed that the rates of resistance to non-β-lactam antibiotics among this population were also low: 33% were resistant to erythromycin, 18% to clindamycin, and 9% to tetracycline. In addition, only 19% (36 out of 192 isolates) of CA-SE isolates, both MRSE (17 isolates) and MSSE (19 isolates) were resistant to three or more classes of antimicrobial agents and as many as 17% of the isolates were susceptible to all antimicrobials.. Resistance to penicillin was high and reached 72%.

In contrast, in the hospital we found a much higher number of MRSE (75 out of 94 isolates), which corresponds to a nasal colonization rate of 30% ($p<0.05$). This rate is comparable to the ones obtained in countries with a low frequency of MRSA (79).

Population structure of *S. epidermidis* isolated in the community

A total of 50 PFGE types were identified among the 192 CA-SE isolates studied, which corresponds to a high level of genetic diversity (Simpson's Index of Diversity, SID =94.71%, 95% CI=[92.94-96.48]). A major PFGE type (PFGE type 10) comprised 17% of the isolates, 45% belonged to 11 minor PFGE types and 38% were sporadic (43 different types).

A total of 53 CA-SE isolates, were analyzed by MLST and forty different STs were found. ST184 was the most prevalent ST (five isolates), followed by ST59 (three isolates) and ST402 (two isolates). The remaining STs were detected in single isolates only, with the great majority of the isolates (72%, 38) being new.

The application of the algorithm goeBURST to MLST data obtained in this study and the data available online (www.mlst.net), allowed the identification of a change in the ancestor of the major clonal complex from ST2 to ST5. This occurred as a result of the increase in the number of isolates belonging to ST5 and of SLVs of ST5 in the S. epidermidis MLST database. Consequently, the previously clonal complex 2 (CC2) is now called CC5. This CC5 is now composed of 27 subgroup founders, including ST2 (see Figure 1). 31 out of the 53 isolates (58%) analyzed in this study belonged to the major clonal lineage (CC5) (see Figure 1). In addition, two isolates belonged to CC171, one isolate was related with CC19 and another one with CC212. The remaining 18 isolates (34%) were singletons.

In comparison to CA-SE, HA-SE isolates were more clonal. Twenty PFGE types were determined among the 94 nosocomial isolates analyzed (SID=81.79, 95%CI=[76.57-87.01]) and 70% of the isolates belonged to three major PFGE types: 10 (33%), 12 (25%) and 11 (12%). The remaining 17 PFGE types corresponded to less than 5%.The analysis of 16 representative isolates (one of each major PFGE type) by MLST showed that they all belonged to previously described STs and to the major CC, CC5.

When we compared the population structures of CA- and HA-SE we found some PFGE types that were hospital specific (class H: eight different types) or community specific (class C: 37 different types). Noteworthy, we also found that the most prevalent genetic lineage in both environments was CC5 (see Table 1) and that isolates within this lineage belonging to specific STs and PFGE types were frequently sampled in the hospital and community (class C/H).

The clones identified as C/H were more frequently found in the hospital (85%) than in the community (35%) and had particular genetic features that may be related to their capacity to survive in both environments. Specifically, all the isolates (except one) belonging to C/H clones, were from CC5 and contained a high number of different SCC*mec* types when collected in the hospital environment (see PFGE types 10 and 12 in Table 1). In addition, they carried ACME with ClaI-*arcCB* type 4 or 6 (43%), (similar to the ACME I from USA300) and the *ica* genes (50%) (data not shown).

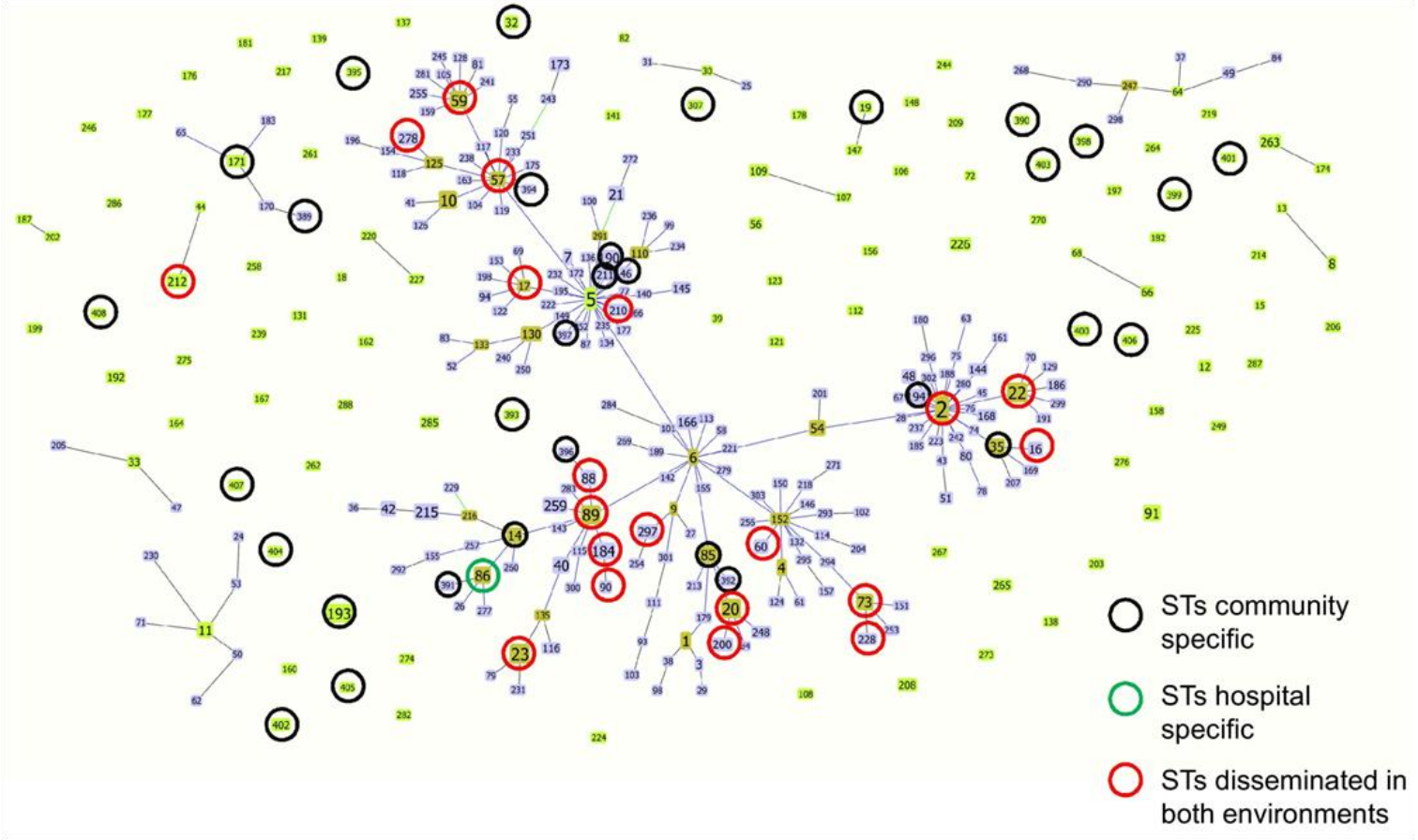


Figure 1. Analysis of MLST data with go-EBURST. The most prevalent STs are represented by a bigger font. Light green STs indicate probable ancestors (group founders) and dark green STs constitute subgroup founders. Blue STs correspond to STs that share the same background (clonal clomplex). Dark circles indicate STs that were found exclusively in the community. Green circles indicate STs exclusively associated with the hospital setting. Red circles indicate STs that were found in both environments.

Table 1. Molecular characterization of *S. epidermidis* isolates that belong to PFGE types identified both in community and hospital. STs: sequence types; CCs: clonal complexes.

PFGE pattern (no of isolates)	STs	CCs	SCC*mec* types (no of isolates)	
			Community	Hospital
10 (66)	16, 22 2, 17, 60, 210, 212	5 5, 212	2&5A (1) MSSE (34)	2&5A, 4A, 5A, NTA, 5B, 2&5C1, 3&5C1, II, IVa, IVc, IVg, 2&5NT (25) MSSE (6)
12 (32)	20, 22, 194 73, 228	5 5	IVg (1) MSSE (8)	2&4&5A, 5A, NTA, 2NT, II, IVg (21) MSSE (2)
16 (15)	184, 90	5	MSSE (13)	MSSE (2)
11 (13)	23	5	MSSE (2)	IVa, IVc, IVd, IVg, IVnt (11)
9 (12)	59 88	5 5	IVa, IVnt (6) MSSE (2)	5A, IVa (3) MSSE (1)
24 (7)	57	5	MSSE (6)	IVa (1)
13 (5)	20 -	5 -	5D (1) MSSE (3)	IVa (1) -
30 (5)	- 278	- 5	- MSSE (2)	3&5A (2) MSSE (1)
8 (4)	297 -	5 -	IVa (1) MSSE (2)	IVa (1) -
31 (3)	89	5	MSSE (2)	MSSE (1)
4 (2)	57	5	MSSE (1)	MSSE (1)
32 (2)	200	5	MSSE (1)	MSSE (1)

Genetic diversity of SCC*mec* among CA-MRSE

From the 18 CA-MRSE isolates, 16 carried either SCC*mec* IV (11 isolates) or SCC*mec* V (five isolates), one isolate carried a new combination of *mec* complex and *ccr* type (5D) and one was non-typeable (4&5A) (see Table 2). From the 11 isolates carrying SCC*mec* type IV, six isolates carried subtype IVa, and two carried subtype IVc. Subtypes IVg, IVh and a non-subtypable variant were identified in single isolates each.

Overall the genetic diversity of SCC*mec*, considering the SCC*mec* subtypes, in the community environment was relatively low (SID = 57.52; 95%CI=[38.35-76.69]).

Table 2. Comparison of community-associated *S. epidermidis* (CA-SE) and hospital-associated *S. epidermidis* (HA-SE).

Feature	**CA-SE**	**HA-SE**	**Statistical validation**
MRSE nasal colonization	7%	30%	p<0,05
Frequency of C/H clones	35%	85%	p<0,05
Multiple ccr	7%	17%	p <0,05
SCC*mec* types	IV, V, 5D, NT1	IV, II, VIII, 4A, 5A, 5B, 5D, VIII, NT2-NT8	-
Diversity in SCC*mec* types[a] (SID)	57%, 95%CI= (38,35-76,69)	83%, 95%CI=(77,38-87,96)	95%CI do not overlap
Diversity in genetic backgrounds (SID)	95%, 95%CI =(92,94-96,48)	82%, 95%CI=(76,57-87,01)	95%CI do not overlap

[a]considering SCC*mec* subtypes

A much higher genetic diversity in SCC*mec* was observed among hospital isolates (SID=82.67; 95%CI=[77.38-87.96]). Of the 75 HA-MRSE isolates studied, 25 (33%) carried SCC*mec* IV and 23 (31%) carried non-typable structures (NT) (see Table 2). In addition, 12 of the isolates (16%) carried new associations between the class of *mec* complex and the *ccr* type (see Table 2). Of the remaining isolates, 13 (17%) carried SCC*mec* II and two (3%) SCC*mec* VIII. The most prevalent subtype of SCC*mec* IV was subtype IVa (11 isolates), followed by SCC*mec* IVc (six isolates), SCC*mec* IVg (three isolates) and SCC*mec* IVd (one isolate). We also identified four isolates with SCC*mec* IV not subtypable by the methods used.

Evidence for the role of hospital environment on SCC*mec* amplification and diversification

We observed that SCC*mec* frequency and genetic diversity in the hospital was significantly higher than in the community (see Table 2). In order to understand in detail the impact of hospital environment on SCC*mec* diversification and amplification we compared the SCC*mec* content of isolates with the same PFGE type collected either in the community or in the hospital (C/H). Interestingly we observed that in several cases CA- and HA-SE with same PFGE type differed significantly in the contents of SCC*mec*. In particular, isolates belonging to PFGE type 10 collected in the community environment carried the SCC*mec* type with *mec* complex class A associated to *ccrC* and *ccrAB4* only, whereas isolates with PFGE 10 collected in the hospital carried as many as 12 different SCC*mec* structures (see Table 1). Similarly, isolates of PFGE type 12 that originated in the community carried only SCC*mec* type IVg, whereas isolates from the hospital carried six different SCC*mec* structures. The same type of results was obtained when we compared CA-MSSE and HA-MRSE with exactly the same PFGE type. For example, isolates with PFGE type 11 collected in the community were all MSSE, whereas isolates collected in the hospital were all MRSE and carried five different SCC*mec* IV subtypes (see Table 1). Similar observations were made for isolates belonging to PFGE types 24 and 30. Overall we observed a positive correlation between MRSE sample size and the number of different SCC*mec* types in the hospital environment and the same was not observed in the community (see Table 1). Moreover, we observed that multiple *ccr*s were present in a higher frequency among hospital isolates when compared to community isolates (HA: 17%, CA: 7%, $p<0,05$). Overall, the results suggest that this may represent new acquisitions of SCC*mec*/SCC, but the selection of isolates followed by recombination at the SCC*mec* level occurring in the hospital environment cannot be disregarded.

Interestingly, SCC*mec* types II, III, IVd, and VIII, as well as new (5A, 5B) and non-typeable (2&4&5A; 3&5A; 2&5C1; 3&5C1; 2&5A; NTA; NTNT; 2New; 2&5New) SCC*mec* structures were detected among hospital-associated isolates only, suggesting that these SCC*mec* types most probably were acquired and/or assembled in the hospital environment.

The two sets of isolates analyzed in this study are not exactly contemporary (CA isolates were collected between 1996-1999 and HA isolates were collected in 2000-2001). In order to exclude the hypothesis that time could be a factor influencing the results obtained, we analyzed SCC*mec* diversity in isolates from community and hospital in two-year time blocks. The results obtained from block to block were comparable to the entire time period suggesting that time should not be a factor contributing to the differences observed between the two environments.

Overall, the results obtained suggest that hospital environment promotes SCC*mec* diversification and amplification - either by SCC*mec* acquisition or by selection of MRSE strains.

DISCUSSION

In the present work we compared the population structures and the frequency and diversity of SCC*mec* in *S. epidermidis* collected in two different environments from the same geographic region and comparable time periods. The molecular characterization of such collections showed that *S. epidermidis* strategies to adapt to hospital and community environments involved the divergent adaptation or selection of specific genetic backgrounds and SCC*mec* elements.

We found a low nasal colonization rate with MRSE (7%) in healthy Portuguese draftees. Other studies concerning different healthy populations have found a higher MRSE colonization rate, like among military personnel (20%) (82) and children attending day care centers (80). However, these two populations present risk behaviors for MRSE dissemination and colonization like frequent physical contact, and higher antibiotic consumption, which are not observed in the population under study here that is constituted by unrelated healthy young individuals.

In addition, we also found a low frequency of multiresistance to antibiotics (7%) among isolates originated in the community. This scenario contrasts sharply to what was observed in the hospital environment where MRSE colonization rates obtained were higher (30%, $p<0.05$) and multiresistance is frequent (51, 79). As a whole these results suggest that CA-SE are probably not functioning as the primary reservoirs of SCC*mec* and other antibiotic resistance determinants to other more pathogenic species like *S. aureus.* Nevertheless, we cannot exclude the hypothesis that certain CA-MRSE when introduced into the hospital may become amplified due to selective pressure and become reservoirs of SCC*mec* for *S. aureus*.

The difference observed in the frequencies of SCC*mec* and multiple *ccr* in the community and hospital, suggests in addition that specific physiological conditions during infection and stresses imposed by the hospital environment can promote SCC*mec* excision/acquisition and dissemination in the *S. epidermidis* hospital population. This hypothesis is further sustained in our study, by the existence of isolates with the same PFGE type that either lack or contain several different SCC*mec* types, depending on whether they were isolated in the community or hospital.

Additionally, the discovery by others of *S. epidermidis* subpopulations with spontaneously deletion of *mecA locus* in isolates from persistent infection (262) and the finding that β-lactams and vancomycin upregulate *ccrA* expression, (263) further supports the hypothesis that the hospital environment may promote SCC*mec* excision and transfer. Likewise, it is possible that SCC*mec* transfer is promoted during biofilm formation, which is the most important virulence propriety of *S. epidermidis*, as was previously proved for other mobile genetic elements (264). However, we cannot exclude the hypothesis that the higher frequency of SCC*mec* observed in the hospital result from the occurrence of selection of MRSE strains originated either in the community or in the hospital.

Besides promoting SCC*mec* acquisition, the hospital environment appears also to contribute to generate genetic diversity in the SCC*mec* elements. This hypothesis is supported by the very high number of SCC*mec* structures with new combinations of classes of *mec* complex and *ccr* types and non-typable SCC*mec* found in nosocomial isolates belonging to a single PFGE type (PFGE type 10 or 12), when compared to isolates of the same PFGE type originated in the community. Previous studies analyzing HA-MRSE also described a high number of new SCC*mec* structures (124, 258). However, only in this study by comparing hospital and community isolates it was possible to envision that the genetic diversity previously observed in SCC*mec* should be mostly created by factors associated to hospital environment. One of the factors that probably contribute to this diversity is the recombination between SCC elements that might occur in the same strain upon multiple SCC acquisitions. Also the increased expression of *ccr* genes, after antimicrobial exposure, as previously observed (263) may increase the opportunities for recombination between several excised SCC elements. Moreover, we should not disregard the fact that in the hospital a large reservoir of SCC*mec* types exist in other coagulase-negative species, what may be contributing to the overall genetic diversity observed in *S. epidermidis*. The high number of different SCC*mec* types present in *S. epidermidis* together with the ones present in other CoNS build up a large reservoir of new SCC*mec* types for *S. aureus*. The acquisition of an additional SCC*mec* type (type IV) by *S. aureus* in the beginning of

the 1990s could have resulted from the acquisition of this element from this highly diverse pool of SCC*mec*.

Although the genetic diversity appears to be higher among hospital isolates than among community isolates, we cannot rule out the possibility that if a higher number of MRSE would be collected in the community, a higher diversity in terms of SCC*mec* would also be found. The analysis of SCC*mec* structure under stress conditions promoted in the hospital environment, like sub-inhibitory concentrations of antibiotics, would probably contribute to the clarification of this question

The comparison of *S. epidermidis* in the hospital and community, showed that both populations had a high genetic diversity, but the community population was more diverse than hospital population, meaning that certain *S. epidermidis* clones are probably more fit to hospital-associated stresses and spread easily in those surroundings. In spite of the high number of different PFGE types and STs in the two collections analyzed, the great majority of the isolates from both community and hospital belonged to a single clonal complex - CC5 (former CC2). This clonal complex has been previously described as the most prevalent clonal complex in the nosocomial population of *S. epidermidis* (17) being characterized by a large genetic diversity, an increased recombination/mutation rate and a high number of acquisitions of SCC*mec* elements (17). The fact that this lineage was identified in this study also as predominant in the community, suggests it is well adapted to the host and that it has also the capacity to adapt to environments with distinct characteristics. In spite of belonging to CC5, CA-SE and HA-SE isolates presented specific PFGE types and STs according to their origin, suggesting divergent evolution.

Moreover, we also identified among CC5, clones with the capacity to survive in both environments. These clones appear to adapt to different environments by modulating the acquisition of SCC*mec* and their level of genetic diversity. Interestingly, the great majority of the isolates belonging to these clones carried SCC*mec* type IV, which probably confer advantages and have no fitness cost in either environment. The fact that these isolates can survive in community and hospital, provides them with a higher capacity of dissemination and accumulation of relevant genetic traits for

survival in both settings. Indeed, it is possible that these clones are the ones responsible for shuffling genetic traits between the community and hospital, namely ACME and SCC*mec* IV. These clones appear to be, by some means, similar to USA300 and EMRSA-15 that carry SCC*mec* IV and gained the ability to survive in both community and hospital environments (265-267).

The collection analyzed in this study is not contemporary, but otherwise reflects the *S. epidermidis* epidemiology in a certain period of time (10-15 years ago) in a specific location. Since the time this study was performed some alterations in *S. aureus* epidemiology occurred, namely the emergence of CA-MRSA as an epidemic, and additional changes might have also occurred in the epidemiology of *S. epidermidi*s. Nonetheless, studies wherein more recent *S. epidermidis* isolates were analyzed continue to report ST2 as the most frequent in hospitals in several different geographic location and SCC*mec* as very diverse in the hospital environment (268-270). These observations suggest that the conclusions drawn for the isolates analyzed in our study are probably still true in the current reality.

Our data demonstrated for the first time the role of hospital environment in the selection of some genetic backgrounds and diversification and acquisition and/or selection of SCC*mec* in *S. epidermidis*. Moreover, it enabled us to identify a class of clones that is able to move between hospital and community. These features of *S. epidermidis* epidemiology will be critical to take into consideration in any infection control program directed to *S. epidermidis* or evolutionary studies regarding SCC*mec*.

ACKNOWLEDGMENTS

We thank all the participants of projects “Disseminação de genes de resistência a antibióticos em populações sadias (crianças, mancebos e recrutas)”, JNICT, PECS/C/SAU/145/95 1997-1998 (Ilda Santos-Sanches, Rosario Mato, Raquel Sá-Leão, Sónia Nunes and Rute Calvão) and “Infection and colonization of multiresistant staphylococci: epidemiology, control and molecular characterization”, Fundação Calouste Gulbenkian, project FCG/ITQB/HFA 1999-2001 (Ilda Santos-Sanches, Rosario Mato, Nuno Augusto, Rui Pimentel and Gabriel Olim) that participated in studies that led to the isolation of the *Staphylococcus epidermidis* strains characterized here. We also thank Isabel Couto and Sérgio Fernandes for the initial characterization of part of the *S. epidermidis* isolates that were further studied here. We thank K. Hiramatsu, T. Ito, D. C. Coleman, R. Daum, K. T. Park and W. B. Grubb for having kindly provided the prototype strains used as controls in this study.

FUNDING

J. Rolo was supported by fellowships 045/BI-BTI/2008 from project “A comprehensive dissection of pneumococcal-host interactions” (Project PNEUMOPATH -Health-F3-2008-222983 from the European Commission), 007/BI/2009 from project “CONtrol of COmmunity-acquired MRSA: Rationale and Development of counteractions” (Project CONCORD–HEALTH-F3-2008-222718 from the European Commission) and SFRH/BD/ 72675/2010 from Fundação para a Ciência e Tecnologia (FCT), Portugal. Support for this work was provided by project P-99911 from Fundação Calouste Gulbenkian and CONCORD-HEALTH-F3-2008/Project Number 222718/European Commission. This work was also supported by Fundação para a Ciência e a Tecnologia through grant #PEst-OE/EQB/LA0004/2011.

TRANSPARENCY DECLARATION

None to declare.

SUPPLEMENTARY DATA

PROJECT: "Dissemination of antibiotic resistance determinants among healthy populations"

The bacteria that cause infections are becoming more resistant to antimicrobial agents, what hinders its control and treatment. For that reason studies aiming to better understand these infectious agents and their capacity of dissemination are fundamental.

The nasopharynx, skin and intestinal flora are constituted by a huge number of microorganisms of different species that can contain several different antibiotic resistance determinants.

The study we are performing aims to assess the frequency of resistance to antimicrobial agents among bacteria colonizing healthy populations. In order to accomplish this objective your participation is indispensable.

We appeal to your collaboration in this study by allowing the collection of samples from the nasal cavity, nasopharynx and skin.

The sample collection does not involve any risk or danger to your health.

In case you agree to collaborate in this study, we ask you to answer to the following questions. The filled questionnaire should be delivered during sample collection.

QUESTIONNAIRE (1996)

1. Are you taking any antibiotic at this moment? Yes ☐ No ☐

   If yes, please indicate the antibiotic name_____

2. Did you take any antibiotic in the last month? Yes ☐ No ☐

   If yes, please indicate the antibiotic name_____

3. If you took any antibiotic, please indicate the reasons why you did it_____

   _____

4. Have you attended any urgency service at the hospital in the last 3 months?

   Yes ☐ No ☐

   If yes please indicate the name of the hospital_____

5. Have you recently been interned in a hospital? Yes ☐ No ☐

   If yes please indicate the name of the hospital_____

6. Do you smoke? Yes ☐ No ☐

Please fill in the following information

Birthday: ___ /___ /____

Residence District:_____

Academic qualifications:_____

Do not write here

SAMPLE CODE: _____ Collection date: ___ /___ /____

QUESTIONNAIRE (1997-1998)

1. Are you taking any antibiotic at this moment? Yes ☐ No ☐

   If yes, please indicate the antibiotic name_____

2. Did you take any antibiotic in the last month? Yes ☐ No ☐

   If yes, please indicate the antibiotic name_____

3. If you took any antibiotic, please indicate the reasons why you did it_____

   _____

4. Have you attended any urgency service at the hospital in the last 3 months?

   Yes ☐ No ☐

   If yes please indicate the name of the hospital_____

5. Have you recently been interned in a hospital? Yes ☐ No ☐

   If yes please indicate the name of the hospital_____

6. Do you smoke? Yes ☐ No ☐

7. At your residence are you usually in contact with animals?

   Yes ☐ No ☐

   If yes, please underline the animals that you have contact with:
   horses, donkey, cows, sheeps, pigs, chickens, ducks, rabbits, dogs, cats, or others.

Please fill in the following information

Birthday: ___ /___ /____

Residence District:_____

Academic qualifications:_____

Do not write here

SAMPLE CODE: _____ Collection date: ___ /___ /____

QUESTIONNAIRE (1999)

1. Please indicate the date of your inscription to CFMTA-OTA ___ / ___ /___

2. Are you taking any antibiotic at this moment? Yes ☐ No ☐

   If yes, please indicate the antibiotic name_____

   If yes, please indicate the reason why you did it _____

3. Did you take any antibiotic in the last month? Yes ☐ No ☐

   If yes, please indicate the antibiotic name_____

   If yes, please indicate the reason why you did it_____

4. Have you attended any urgency service at the hospital in the last 3 months?

   Yes ☐ No ☐

   If yes please indicate the name of the hospital_____

5. Have you recently been interned in a hospital? Yes ☐ No ☐

   If yes, please indicate the name of the hospital_____
   If yes, please indicate the reason why_____

6. Do you smoke? Yes ☐ No ☐

7. At your residence are you usually in contact with animals?
   Yes ☐ No ☐

   If yes, please underline the animals that you have contact with:
   horses, donkey, cows, sheeps, pigs, chickens, ducks, rabbits, dogs, cats, or others.

Please fill in the following information

Birthday: ___ /___ /____ Gender: Female ☐ Male ☐

Residence District:_____

Academic qualifications:_____

Do not write here

SAMPLE CODE: _____ Collection date: ___ /___ /____

Chapter VII

Concluding Remarks

The SCC*mec* element, carrying the β-lactam resistance determinant *mecA*, is one of the most important and well-characterized bacterial mobile genetic elements. However, its origin and the mechanisms involved in SCC assembly and diversification are not well understood. Previous studies have revealed that distinct species of *Staphylococcus* were involved in the evolution of SCC*mec*, but these included a few number of isolates and traditional typing methods, that have been precluding the full understanding of the contribution of each species for this process. In this Thesis we aimed to approach this issue at the population level using whole genome sequencing approaches.

Evidence gathered prior to this Thesis indicated that the ancestral form of the *mecA* gene corresponds to the *Staphylococcus sciuri mecA1* gene, which encodes the native penicillin-binding protein PBP4 (172, 176). In **Chapters II-IV**, we explored the role of three species of the phylogenetic *sciuri* group, in the evolution of this housekeeping gene into the resistance determinant *mecA* and the steps that lead to its integration into a mobile genetic element to give rise to SCC*mec*.

Since its original assembly, SCC*mec* was disseminated among the remaining staphylococcal species and has diversified into at least eleven different structural types (111-113). Coagulase-negative staphylococci (CoNS) appear to have been crucial in this diversification process. In particular, *Staphylococcus epidermidis* displays a highly diverse pool of SCC*mec* types and presents increased ability to acquire SCC*mec* (17), which suggests it could act as an assembly platform of new SCC*mec* types for the remaining population of staphylococci. The resources and mechanisms driving the diversification of SCC*mec* in this species were yet to be described. In **Chapters V-VI**, we explored the specific role of *S. epidermidis* in the origin and evolution of SCC*mec*.

The gene *mecA* evolved from a native to a resistant determinant multiple times over phylogeny

The studies included in **Chapter II** described, for the first time, the native genetic location of *mecA* homologues in the most primitive species of *Staphylococcus*, *S. sciuri.* DNA hybridization assays with a *mecA* probe showed that, in *S. sciuri*, *mecA1*

was located near the SCC*mec* integration site, the *orfX*. These results were further detailed in **Chapter IV**, through a whole-genome sequencing approach that allowed to determine that the *mecA* homologues *mecA1* in *S. sciuri*, *mecA2*/*mecA* in *S. vitulinus* (with the exception of three *mecA* positive *S. vitulinus* strains) and *mecA* in *S. fleurettii* were all located 200 kb downstream *orfX*. This genetic location was the closest to the *orfX* observed for a gene encoding a penicillin-binding protein among sequenced staphylococcal genomes (http://www.ncbi.nlm.nih.gov/) and has probably been crucial in the incorporation of *mecA* into SCC*mec*. Interestingly, and as already described (173), we found that the close vicinity of *mecA* homologues carried by this group of phylogenetically related species was very similar, evidencing its vertical inheritance from a common ancestor. Phylogenetic analysis of all *mecA* homologues identified *mecA1*, a native housekeeping gene of *S. sciuri*, as the most ancient form suggesting it was the precursor of *mecA*; and *mecA2* from *S. vitulinus* as an intermediary form between *mecA1* and *mecA* in *S. fleurettii* and *S. aureus*.

On the other hand, in **Chapter III,** phenotypic data gathered for strains within the *sciuri* group of species showed that during the process of evolution within the native location, the *mecA* homologues evolved from a susceptible towards a resistant phenotype. Actually, whereas most of *S. sciuri* strains carrying only *mecA*1 and *S. vitulinus* strains carrying only *mecA2* were susceptible to β-lactams, all the *S. fleurettii* strains were resistant. However, there were some exceptions, namely we found that 10% of *S. sciuri* and 17% of *S. vitulinus* strains not carrying *mecA* were resistant to β-lactams. Investigation of the putative genetic mechanisms behind the expression of β-lactam resistance in these strains showed that some strains carried mutations in the *mecA1* promoter, in *S. sciuri,* and in the *mecA2* promoter, in *S. vitulinus*. This phenomenon has been previously described in clinical *S. sciuri* and in *S. sciuri* submitted to a stepwise exposure to increasing concentrations of methicillin, wherein the accumulation of SNPs in the promoter of *mecA1* lead to overexpression of *mecA1*-encoded PBP4 and consequently to a β-lactam resistance phenotype (177).

In another group of oxacillin resistant strains we identified the accumulation of non-synonymous SNPs in the non-binding domain of the native penicillin-binding 4 (PBP4) of *S. sciuri*. Unlike mutations in the promoter, alterations in the non-binding

domain have never been described to be associated to a resistant phenotype. In fact, the active site of *mecA*-encoded PBP2a is located in the other protein domain, the transpeptidase domain (108) and the β-lactam resistance phenotype is achieved by a torsion of the protein backbone of the regions surrounding the active site groove that results in a decreased acylation rate by β-lactam antibiotics (108). However, through the application of *in silico* structural prediction tools to the aminoacid sequence of the proteins encoded by the *mecA1* alleles associated to β-lactam resistance, we found that aminoacid alterations on the non-binding domain could lead to an overall altered folding of PBP4 that result in a "closed groove" near the active site of the protein, which may protect the active residues from the antibiotic action. We hypothesize that these alleles could originate a PBP4 form that would be resistant to acylation by β-lactams. To our knowledge, this is the first description of alterations in the non-binding domain of a *mecA*-encoded PBP that may lead to a resistant phenotype. Nonetheless, further studies would be needed to confirm this hypothesis. Site-directed mutagenesis of *mecA1* alleles and assessment of β-lactam resistance of mutants carrying these alleles would be a suitable strategy.

Besides the diversification of the promoter and the accumulation of SNPs in the non-binding domain of *mecA* homologues, our results showed that alterations in the genetic background during speciation could also have led to the development of β-lactam resistance. Evidence for that is the fact that strains of *S. sciuri*/*S. fleurettii* and *S. vitulinus* carrying virtually identical regulators, promoters and *mecA* alleles, were either heterogeneously resistant or susceptible to β-lactams. Examples of other genetic determinants that were described to influence the expression of *mecA* and the resistant phenotype are the *blaRI-blaI* regulators of the β-lactamase (*blaZ*) (224-226) and the auxiliary genes (207). However, the susceptibility of *S. vitulinus* isolates cannot be explained by the presence of *blaZ* locus, because no link was found between the presence/absence of β-lactamase and the decreased β-lactam resistance profile of isolates belonging to this species. Also, the contribution of auxiliary genes to this phenotype is difficult to assess, since the mechanisms that link auxiliary genes to the resistance/susceptible phenotype are still poorly understood.

Our data implies that the selective pressure driving the evolution of *mecA* was the exposure of species of the *S. sciuri* group to β-lactam antibiotics as shown by the BEAST analysis. Evidence for that come from the estimation that the great majority of *mecA1* alleles have emerged around the time that penicillin was introduced in the clinical practice in humans (1940s). In addition, the highest degree of genetic diversity was observed among isolates collected from humans, which suggests that the selective pressure imposed by antibiotic use in humans might have triggered the diversification of *S. sciuri mecA1*. Although *S. sciuri* is not a frequent human colonizer or infecting agent, interaction of these bacteria with antibiotics could have occurred through direct or indirect contact with contaminated hospital effluents upon its use. BEAST analysis also have estimated that emergence of *mecA* have occurred in the late 1950s, a time wherein penicillin became to be largely used as animal feed additives in subtherapeutic concentrations (271).

Our results suggest that the use of penicillin in clinical practice and as animal feed additives was probably the driving force leading to an increase in diversity of *mecA1*, which gave rise to the emergence of new *mecA* homologues with decreased susceptibility to β-lactams, including *mecA*. This is in accordance with the view that *mecA* emerged before the introduction of methicillin in the clinical practice in 1960s.

The assembly of *mec* complex involved several steps and different species belonging to the *sciuri* group

All *mecA* homologues in the native location presented the same close vicinity, being flanked downstream by *ugpQ* and upstream by *mecR2 and psm-mec* (135). This same close vicinity was also observed flanking *mecA* in contemporary SCC*mec* types. Furthermore, the phylogenetic tree constructed from the alignment of the nucleotide sequence of *ugpQ*, *psm-mec* and *mecR2* genes, showed an identical clustering and hierarchy to that of the phylogenetic tree constructed for *mecA* homologues wherein genes of *S. sciuri* were the most ancestral and those of *S. fleurettii*, the most recent. This type of results had been already observed, for other genes located in the vicinity of *mecA* homologues, like *mvaS* (173), suggesting that, not only *mecA* homologues, but

also their close vicinity, which together are part of the *mec* complex, have evolved across phylogeny in the native location from *S. sciuri* to *S. fleurettii*.

In contrast to *mecR2*, the other two *mec* regulators, *mecI* and *mecR1*, were only associated with *mecA* in the native location and never with *mecA1* and *mecA2*. In addition, the degree of homology between *mecI* /*mecR1* from strains belonging to the *sciuri* group with those found in SCC*mec* was comparably high (>90% nucleotide identity). These results indicate that *mecR2* was the most ancient regulatory gene and that *mecI* and *mecR1* were incorporated later in evolution, in the chromosome of *S. vitulinus* or *S. fleurettii* after the evolution of *mecA1* into *mecA*.

The BEAST analysis (**Chapter III**) revealed that *mecA* has first emerged in *S. fleurettii* and only afterwards in *S. vitulinus*. We hypothesize that recombination between *mecA* and adjacent regions carried by *S. fleurettii* in the native locus and the β-lactamase locus, *blaZ-blaR1-blaI*, usually carried by plasmids, might have occurred. Actually, the nucleotide identities present in these two loci and the similar orientation and arrangement of the genes have been previously suggested to reflect an evolutionary link between these two loci (223). In addition, a functional link has also been demonstrated since the repressor *blaI* can regulate the expression of *mecA* (224-226). However, this hypothesis has yet to be confirmed, since no intermediate structures between the two loci were found in our collection.

Finally, the last step of construction of *mec* complex would have been the introduction of a copy of IS*431* upstream of the HVR region. Although in our study IS*431* was not found in the vicinity of *mec* complex A in the native locus of *S. fleurettii* or *S. vitulinus*, a copy of IS*431* was found in the vicinity of *mecA* complex in the *orfX* of *S. vitulinus*. This result suggests that IS*431* integration into the *upgQ-mecA-mecR1-mecI-mecR2-psm-mec* structure, occurred either in the *orfX* region after mobilization of *mec* complex or during its mobilization. The *upgQ-mecA-mecR1-mecI-mecR2-psm-mec* structure already formed in *S. fleurettii* was then probably transferred to some strains of *S. vitulinus* through recombination with the native *mecA2*, by an unknown mechanism.

Our results also showed that during this mobilization, *mec* complex carried additionally other regions. These regions, that were located upstream and downstream the *mec* complex correspond to SCC*mec* J2 and J3 regions and were also found in the native location of *S. sciuri*, *S. vitulinus* and *S. fleurettii*. Like other genes in the native location, these genes were more frequent and had higher homology with SCC*mec* genes in *S. fleurettii* and *S. vitulinus.* Altogether our study showed that *mec* complex and the J2 and J3 regions evolved over phylogeny and were transferred to a SCC cassette from *S. fleurettii* or *S. vitulinus*.

Apart from the fact that *mecA* homologues native location is near the *orfX*, from our data is not obvious how and in which species *mec* complex has been mobilized into a SCC element. Our results showed that this event might include the deletion of *mec* complex from the native location in both *S. fleurettii* and *S. vitulinus*. Actually, some *S. vitulinus* strains did not carry any form of *mecA* homologues in the native location and although, in our collection, the *mec* complex A was ubiquitous among *S. fleurettii*, there have been reports of *S. fleurettii* strains lacking this locus (173). We suggest that after its assembly, *mec* complex was mobilized from *S. fleurettii* into *S. vitulinus* and recombined with *mecA2* locus to give rise to *mecA* carrying strains in the native locus. In a second step, the *mec* complex from *S. vitulinus* was excised from the native location and transported to *S. sciuri* (see below).

The mobilization of the *mec* complex A between the different species could have occurred in transposons as previously observed in *S. haemolyticus* (154) and in the related species *M. caseolyticus* (120) or by transduction as previously shown (116-118). Actually, by using the genome-search tool PHAST (234), we were able to find a high frequency of phage DNA in the genomes of *S. sciuri*, *S. vitulinus* and *S. fleuretttii* (84%, 77% and 75%, respectively).

The mobile genetic elements SCC originated in *S. sciuri*

The next step in the assembly of SCC*mec* was the integration of the *mec* complex into a SCC element. Our results showed that assembly of SCC elements occurred in parallel with the assembly of *mec* complex and also involved several

species of the *sciuri* group. Moreover, our data provided evidence that indicate that the primordial SCC*mec*, and thus the integration of *mec* complex into a SCC element was accomplished in *S. sciuri* (see **Chapter II** and **IV)**.

We identified for the first time a high frequency, in methicillin-susceptible *S. sciuri*, of *ccr* genes, responsible for the mobilization of SCC elements **(Chapter II)**. The *ccr* genes were widely distributed among the *S. sciuri* population, in different genetic backgrounds (illustrated by the different PFGE types found among *ccr*-carrying strains), and in strains from different geographic regions and periods of isolation. Furthermore, contrarily to other coagulase-negative staphylococcal species – especially the ones that are human-associated (124, 131, 167, 198)– we found a high genetic diversity of *ccrA* and *ccrB* alleles, as well as several different combinations of the allotypes found. Noteworthy, all the *ccrA* and *ccrB* allotypes related to those carried by contemporary SCC*mec* cassettes identified to date, were found in *S. sciuri*. In particular, *ccrA* and *ccrB* alleles related with *ccrAB3*, occurred in a high frequency. This is in accordance with previous studies wherein SCC*mec* type III, carrying *ccrA3B3*, was shown to be the most prevalent among isolates belonging to the *sciuri* group (170, 180, 202).

In **Chapter IV**, through whole-genome sequencing, the *ccr* genes were confirmed to be highly frequent and diverse among *S. sciuri*, but rare among related species (*S. vitulinus* and *S. fleurettii*). Furthermore, as opposed to *S. sciuri*, most of the *ccr* allotypes carried by *S. vitulinus* and *S. fleurettii* were not related with the ones carried by contemporary SCC*mec* cassettes. Moreover, the phylogenetic analysis showed that, in general, the *ccr* allotypes of *S. sciuri* are the most ancestral among the three species. On the other hand, *ccr* allotypes phylogenetically distantly related with those found in staphylococci were previously identified in *M. caseolyticus* near the *orfX* within a SCC element (120, 146). Altogether these findings suggest that *M. caseolyticus* recombinases could be the ancestral forms of recombinases found in *S. sciuri* that in turn was the first staphylococcal species to acquire *ccr*. Nonetheless, further studies are needed to assess the functionality of the *ccr* genes carried by *S. sciuri*, in particular if a circular intermediate would be formed upon excision, as has previously been described for *S. aureus* (109) and *M. caseolyticus* (120).

The whole genome sequencing data also showed that besides having a high frequency of *ccr*, *S. sciuri* also contains a high number of SCC elements, in which these *ccr* genes are delimited by DR/IR junctions. Like what was described for *S. aureus*, these *S. sciuri* SCCs carried different genetic determinants associated with colonization and survival in the host, such as permeases, dehydrogenates, oxidases and genes encoding metal resistance and detoxification enzymes. In particular we found that some of the genes typically found in the J1 region of SCC*mec* were found within SCC elements. Interestingly, some of these J1 genes in several *S. sciuri* strains were found outside any element in the *orfX* region, or elsewhere in the chromosome. These observations suggest that SCC elements content in *S. sciuri* is built from the incorporation of *S. sciuri* housekeeping genes into the *orfX* region and SCC elements, probably through recombination events. Although SCC elements could be also found in *S. vitulinus* and *S. fleurettii*, they were rare; also, the few SCC that existed were not composed of genes found in the *orfX* outside SCC elements or in other parts of the chromosome.

Taken together, our results suggest that *S. sciuri* have provided two building blocks for the assembly of SCC*mec*, the *ccr* genes, encoding the recombinases that are responsible for the mobility of SCC*mec* and the J1 region. Also *S. sciuri* appear to function as sources of new SCC elements. We believe that at least one of these *S. sciuri* SCC elements was the recipient of the *mec* complex previously assembled as described before, to give rise to the primordial SCC*mec*.

The primordial SCC*mec* element was assembled in *S. sciuri*

In **Chapter IV**, we describe the results that support the hypothesis that an ancestral SCC*mec* type was assembled in *S. sciuri*, with the contribution of the related species *S. vitulinus* and *S. fleurettii*.

The main evidence for that is the fact that we found in *S. sciuri* a SCC*mec*-like structure (SCC*mec* III-like structure A) that is similar (but not identical) in content and synteny to SCC*mec* type III, but that contains ancestral forms of the SCC*mec* III genes. This included mainly a *ccrAB* type related with *ccrAB3* and genes that are part of the

SCC*mec* III J1, J2 and the J3 regions, as well as *mec* complex A. Moreover, we found that this structure is a mosaic composed of part of one SCC carried by several *S. sciuri* in our collection (specifically the *ccrAB* genes and J1 region) and part of a 25 kb fragment encompassing the J2 region, the *mec* complex A and part of the J3 region, which are very similar to SCC*mec* type III and to *mec* complex vicinity in the native location. We suggest that SCC*mec* III-like structure A corresponds to an ancestral SCC*mec* that was assembled in *S. sciuri* through the incorporation of *mec* complex and adjacent J2 and J3 regions from *S. fleurettii*/*S. vitulinus* into a resident SCC element.

SCC*mec* III could have been though the first SCC*mec* type to emerge. Although it has been firstly identified in a *S. aureus* strain recovered in New Zealand in 1985 (138), SCC*mec* III could already exist among CoNS isolates. SCC*mec* III is one of the most widely distributed SCC*mec* types, found among several staphylococcal species (180, 181). It is also frequently carried by the most relevant human pathogens *S. aureus* (106) and *S. epidermidis* (124). Due to its high frequency and wide distribution among staphylococcal species, it is plausible that SCC*mec* III is ancient.

Based in all our findings, we propose the following model for the assembly of SCC*mec* (Figure 3 of **Chapter IV**): (i) SCC elements emerged in *S. sciuri* by incorporation of *ccr* and core genes within DR/IR, probably through recombination; (ii) *mecA* homologues and genes in their vicinity (J2 and J3 genes) evolved over phylogeny; (iii) *mecA* homologues evolved towards a resistant phenotype from *S. sciuri mecA1* to *S. fleurettii mecA* and the main driving force was the use of penicillin in humans and animals; resistance to β-lactams emerged several times during evolution, through accumulation of mutation in the *mecA* homologues promoters and non-binding domain and culminated in the emergence of *mecA*; (iv) a parallel adaptation of staphylococcal genetic background to the expression of resistance took place; (v) the *mec* complex A has emerged in *S. fleurettii* through recombination with the *bla* operon; (vi) *mec* complex A was excised from *S. fleurettii* and transferred to *S. vitulinu*s where it recombined with *mecA2* in the native region; (vii) *mec* complex and adjacent regions were excised from *S. vitulinus* by the action of transposases, as the one carried by IS*431*, and transferred to *S. sciuri*, where it recombined with a resident SCC; (viii) SCC*mec* III disseminated among the *sciuri* group.

***S. epidermidis* has contributed to the assembly of SCC*mec* IV**

The results described above provided evidence for the emergence of SCC*mec* type III, but failed to provide an explanation for the origin and evolutionary steps leading to the assembly of the remaining SCC*mec* types.

Our analysis showed that *S. sciuri* contained ancestral forms of all the *ccr* allotypes described to date, suggesting it could be their origin. On the other hand, it has been previously observed that different CoNS are enriched in particular *ccr* allotypes (124, 131, 167, 198), irrespective of the fact that strains are resistant or susceptible to methicillin, indicating that SCC elements carried by methicillin susceptible (MS)-CoNS carry the same *ccr* allotypes than SCC*mec* carried by methicillin resistant (MR)-CoNS of the same species. It is though tempting to speculate that during staphylococcal speciation each *ccr* allotype was fixed in each species. And that SCC*mec* types emerged by a mechanism similar to that described above for SCC*mec* III, wherein *mec* complex was integrated into resident SCC carrying different *ccr* allotypes in each species.

To try to prove this hypothesis we decided to study early CoNS isolates that could carry primordial forms of other types of SCC*mec*. We focused in the study of *S. epidermidis*, a species enriched in *ccrAB2* in which the most frequent SCC*mec* is SCC*mec* IV (17, 110, 237, 258). The earliest report of a MRSE strain corresponds to a strain isolated in 1973, in Canada (110). However, reports of *S. epidermidis* strains with decreased susceptibility to methicillin date back to 1960s, in Denmark (70). In **Chapter V**, we provide the results obtained from the analysis of the *orfX* region in a collection of *S. epidermidis* isolated in the 1960s, in Danish hospitals.

The analysis of 21 *S. epidermidis* isolates collected in 1965, from nasal swabs obtained from hospitalized patients, by whole-genome sequencing, showed that a single strain (1/22, 4.5%) (DNK22) carried the *mecA* copy. This strain carried a ψSCC*mec*-IV-like element, that did not contain *ccr* genes, but carried *mec* complex B and regions of homology mainly with SCC*mec* IV. The ψSCC*mec* IV-like element identified among early *S. epidermidis* had a singularity: the *mecA* gene was interrupted by a copy of IS*431*. Since a second copy of IS*431* was identified downstream of *mecA*,

we hypothesize that the mobilization of *mec* complex B to the *orfX* vicinity has occurred by the action of the transposases present in IS*431*. An inaccurate insertion event would have placed IS*431* in the coding frame of *mecA*. Alternatively, the interruption of *mecA* by IS*431* may have been a strategy to accommodate *mecA* in the chromosome. In strain DNK22, the IS*431* interrupting *mecA* was flanked by 8 bp repeats, indicating that its integration was recent and that the element is probably still mobilizable (254). The ψSCC*mec* element observed could be though an intermediate structure between a SCC carried by *S. epidermidis* and SCC*mec* IV. In fact, among our collection of isolates from 1965, we identified a high frequency of SCC carrying SCC*mec* IV-associated genes encoding hypothetical proteins. Therefore, it is possible that the ψSCC*mec* IV-like element emerged from the integration of *mec* complex B and adjacent regions in an already assembled ψSCC.

This observation, together with the low frequency of *mecA* in our collection of early *S. epidermidis*, suggests that the introduction of *mec* complex B in the *orfX* vicinity of *S. epidermidis* from 1965 was a rare event, probably because the genetic background was not yet adapted to allow the expression of a novel penicillin-binding protein. The interruption of *mecA* by IS*431* could have been a strategy to provide time for adaptive mutations to occur in the genetic background. Finally, the interruption of *mecA* by IS*431* could be a strategy to regulate the expression of *mecA*; IS*431* mobilization could function as a switch to turn on the expression of β-lactam resistance only when necessary. However, we were not able to excise IS*431* from *mecA*, when we incubated the strain with subinhibitory concentrations of penicillin and oxacillin or high salt concentrations. This could be due to the fact that the genetic background of DNK22 was not yet adapted to cope with the expression of *mecA* or because the conditions tested were not the most appropriate to induce the excision of IS*431*.

The evolution of *S. epidermidis* and SCC*mec* was promoted by contact with the hospital environment

S. epidermidis is the main commensal of the skin flora in humans, and most of its life time lives in the absence of antibiotic pressure. However, in the hospital environment, *S. epidermidis* can become a pathogen, mainly due to its capacity to accumulate antibiotic resistance genes and form biofilms. In spite of this duality, the majority of *S. epidermidis* studied are nosocomial strains. In fact, not much was known regarding the epidemiology of strains isolated in the community. Moreover, the impact in *S. epidermidis* evolution of the contact with environmental pressures imposed inside hospitals, such as antibiotics or the human immune system during infection, was also not fully understood. These issues were explored in **Chapter VI** of this Thesis.

The comparison of the population structure of contemporary *S. epidermidis* isolates from the community and hospital obtained in the same time period and in the same geographical area showed that the genetic backgrounds of community-associated *S. epidermidis* were more diverse as illustrated by the higher number of PFGE types and STs. However, eBURST analysis indicated that the majority of *S. epidermidis* strains, irrespective of being isolated in the community or the hospital belonged to the same lineage, clonal complex 2 (CC2). We also found that although belonging to the same lineage, the frequency of MRSE strains in hospital was much higher than in the community, suggesting that hospital contact promoted either the acquisition and/or amplification of SCC*mec* or spread of MRSE. On the other hand when we compared closely related strains (with the same PFGE type) from the two settings we found that the genetic diversity of SCC*mec* elements within hospital isolates was much higher than among community isolates, which carried almost exclusively SCC*mec* IV. This suggests that in addition to promoting the acquisition of SCC*mec*, hospital contact also appears to promote its diversification. These results are in accordance with previous studies that show that CC2 strains have an increased ability to acquire SCC*mec* and recombine (17, 42). Furthermore, the expression of *ccr* has been described to be upregulated by subinhibitory concentrations of antibiotics (263), a condition associated with the hospital environment. The increased expression of *ccr* genes in the hospital environment, may promote not only SCC*mec*

excision/insertion events, but also recombination between different SCC elements, leading to SCC*mec* diversification.

Another approach that we used to study the impact of hospital environment on *S. epidermidis* evolution was to compare early and contemporary strains. Since the 1960s until the present, the use of medical devices has sharply increased, the hospitalization periods have become longer and exposure to antibiotics more frequent and extended. Due to these dramatic changes in the nosocomial setting, early strains (collected in the 1960s) could be considered as strains that have had less contact with the hospital and associated stresses than contemporary strains.

In **Chapter V**, we compared the population structure of early (1965) and contemporary (1990s) hospital-associated *S. epidermidis* isolates collected in Denmark. Several differences were observed both in population structure and in the content of virulence genes and mobile genetic elements between these two types of isolates.

We observed that no isolate of sequence type ST2 was present among the early *S. epidermidis*, but several sequence types that belonged to CC2, were found. The single ST that was shared between both collections was ST5, a single-locus variant of ST2 that contains, as ST2, a large number of single-locus variants (www.mlst.net). When the predicted core genomes of early and contemporary isolates was compared in a phylogenetic tree, we found that their genetic background was not very different. Both early strains belonging to ST5 and virulent contemporary isolates (carrying antibiotic resistance determinants, virulence factors and associated to infection) belonging to CC2, clustered together, suggesting that the genetic background that allowed a full adaptation to the hospital environment was already present by 1965. Due to the fact that ST5 appears to have emerged first than ST2, and due to its relatedness with ST2 and isolates within CC2, we suggest that ST5 is the founder of clonal lineage CC2, instead of ST2.

Although, the core genomes of isolates collected in Danish hospitals over a period of 30 years did not change considerably, striking differences were found between these collections, when we analyzed their mobilome. Specifically, we found a

significant higher frequency of intact prophages in the early collection, and a higher frequency of insertion sequences (IS) and SCC*mec* in the contemporary collection.

The high frequency of prophages among *S. epidermidis* genomes collected in 1965 might have provided to these strains a mechanism of genetic transfer that allowed for the accumulation of the antibiotic resistance determinants, heavy metal resistance genes and SCCs in these genomes. On the other hand, the decrease in frequency of intact prophages in contemporary *S. epidermidis* could be a strategy to maintain and regulate the expression of biofilm, which is the most well recognized *S. epidermidis* virulence factor. In fact, the negative impact of prophages in the biofilm formation ability of *S. epidermidis* is well-documented (239, 241).

Besides having a higher frequency, we observed that the diversity of IS was also slightly higher in contemporary than in early genomes. Namely, contemporary isolates could carry as many as 14 different IS elements while early isolates would carry only as many as 10 different IS. The higher accumulation of IS in the genome of contemporary isolates might be related with the incorporation of antibiotic resistance determinants and other genes important for survival in the hospital environment, since IS are often transported as transposons, which typically carry antibiotic resistant genes (272). Also they could have had impact in the expression of virulence factors, such as biofilm formation, through inactivation of *ica* genes or their regulators, as previously reported (90, 235, 243).

On the other hand, the contemporary Danish isolates had a higher frequency of SCC*mec* than the isolates collected in 1965. Most of the contemporary isolates carried SCC*mec* IV (67%), but other types were also found (SCC*mec* type I, III, V and a combination of *mec* complex C2 with *ccrAB2*). Therefore, the introduction of SCC*mec*, particularly SCC*mec* IV, in *S. epidermidis* genetic backgrounds must have been a crucial event in the adaptation of *S. epidermidis* to the hospital environment. Studies with MRSA have described a link between the presence of SCC*mec* and the ability to produce biofilm, namely hospital-associated (HA) strains carrying SCC*mec* II and III, produced more biofilm than community-associated strains carrying SCC*mec* IV (123). Further studies have shown that the increased ability of HA-MRSA to form biofilm was

associated with the presence of *psm-mec*, a cytolysin carried by SCC*mec* II and III (273). Such studies have not been performed in *S. epidermidis*, but it would be interesting to verify if like in MRSA, MRSE carrying these SCC*mec* types have increased biofilm formation.

In addition we observed that the *S. epidermidis* collected in 1965 showed a high frequency of carriage of both types of ACME: ACME-I, composed by *arc* and *opp*3 operons and ACME-II, which is composed by the *arc* operon alone. In contrast, in contemporary Danish isolates, exclusively ACME-I was found in high frequency. This is in accordance with previous studies that have reported the dissemination of ACME-I in contemporary *S. epidermidis* (100, 101). In a *S. aureus* rabbit model of bacteremia, it has been demonstrated that ACME-I can potentiate colonization and dissemination of the strains carrying it (98). Similarly, the selection of ACME-I in contemporary strains could have provided *S. epidermidis* with enhanced epidemic features, which are convenient characteristics to harbor in a nosocomial environment.

Besides varying in the content of mobile genetic elements early and contemporary *S. epidermidis* varied in the distribution of genes associated with biofilm formation. In particular the contemporary isolates had a higher frequency of the genes *ica* and *aap* and carried in addition *bap* and *sdrF*, which were not found among the early population. On the other hand, in the early collection, the metalloprotease SepA was widely spread, in contrast with contemporary isolates. Biofilm formation is one of the key features of *S. epidermidis* that is directly responsible for its success in the hospital environment nowadays (51, 71, 89, 91, 92). Our results showed that the accumulation of genes involved in biofilm formation was part of the strategy of adaptation of *S. epidermidis* to the hospital environment. SepA is described to promote biofilm formation by mediating the release of extracellular DNA through upregulation of the activity of the autolysin AtlE (274). These results suggest that early *S. epidermidis* had already the potential to form biofilm, however its structure and composition was certainly different from the biofilms formed by contemporary isolates. This may have had important implications in the resistance of the biofilms to antibiotics and immune system. Moreover, the establishment of *ica* operon in contemporary *S. epidermidis* may have complicated the infections caused by this

bacteria, due to the known immunogenic activity of *ica*-encoded PIA (60). The comparison of pathogenesis of early and contemporary strains belonging to ST5 in a mouse catheter model could be used to prove this hypothesis.

We found a wide distribution of antibiotic resistance genes among the nosocomial *S. epidermidis* genomes. Interestingly, we found that most of these genes were already present among early isolates. In particular, early *S. epidermidis* isolates carried resistance determinants to antibiotics already used in clinical practice by 1965 (246-248) (penicillin, erythromycin, tetracycline, fusidic acid and spectinomycin). This antimicrobial resistance pattern was previously found among early MRSA and MSSA isolates collected in Denmark in a similar time period (1960s) (249). The same kind of observation was obtained regarding the distribution of SCC elements carrying heavy metal resistance genes, which were carried exclusively by the early isolates. Likewise, genes conferring heavy metal resistance have been observed in early populations of *S. aureus* (244). Therefore, we hypothesize that acquiring resistance to these particular antibiotics and heavy metals was beneficial for the population of nosocomial staphylococci at that time.

Taking into account the results described in **Chapters V and VI**, we suggest the following model of evolution and adaptation of hospital-associated *S. epidermidis* (Figure 1). The clonal types of early *S. epidermidis* were different from those of contemporary isolates, but the ancestral ST5 was already present and they virtually all belonged to the same genetic lineage. *blaZ* was very frequent among the early collection, so we suggest that this gene was obtained by plasmid acquisition soon after the introduction of penicillin in the clinical practice, as described for *S. aureus* (68, 275).

Through phage acquisition and dissemination, the early hospital-associated *S. epidermidis* acquired genes important to survive in the increasingly harsh hospital environment: antibiotic resistance determinants, SCC*mec* IV and other elements such as ACME and heavy metal resistance genes. As the development of medical invasive techniques increased, *S. epidermidis* genetic lineages carrying biofilm-associated genes disseminated, and most likely the highly invasive and well-adapted ST2 increased its

frequency in the population structure of *S. epidermidis*. In addition, IS elements that would regulate the expression of the biofilm phenotype and possibly other genes, were acquired and maintained in the genome. Finally, multiple acquisition of SCC*mec* led to diversification of this structure.

.

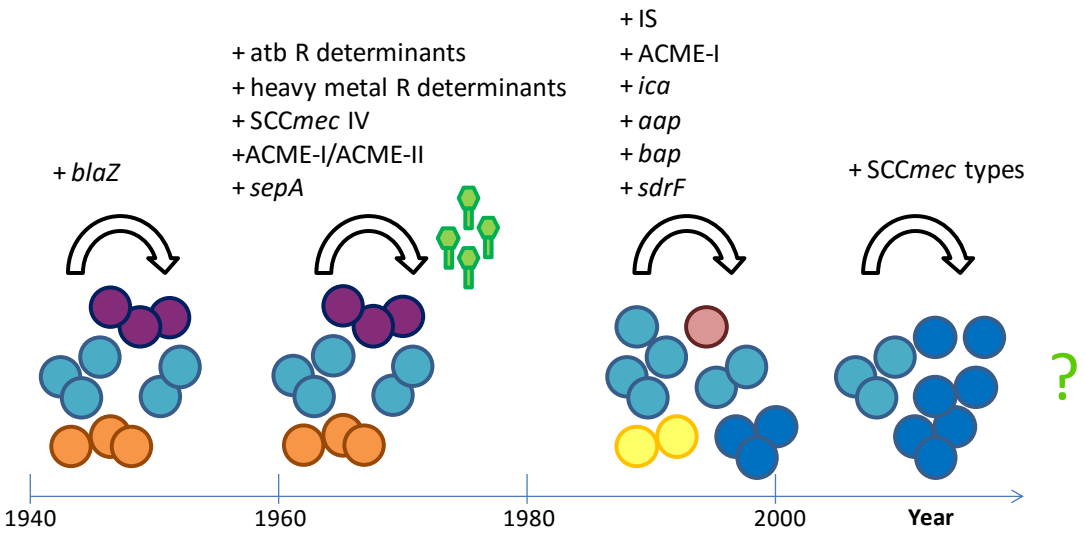


Figure 1. Model for the evolution of hospital-associated *S. epidermidis* genetic backgrounds. The early *S. epidermidis* population structure was composed by ancestral sequence types, namely ST5 (light blue circles) and others (remaining circles). The acquisition and dissemination of phages in this early population possibly allowed the acquisition of antibiotic and heavy metal resistance determinants, SCC*mec* IV and ACME. Among the highly genetic diverse population structure of *S. epidermidis*, ST2 (dark blue circles) frequency would increase, since expansion of clones carrying biofilm-associated genes and IS elements would be favored. Finally, optimized adaptation to the hospital would be achieved through SCC*mec* acquisition and diversification.

In conclusion, in this Thesis we clarified the role of *S. sciuri* as the origin of the two building blocks of the staphylococcal mobile genetic element SCC*mec* and provided a model for the assembly of the first SCC*mec* element. The origin of SCC*mec* was revealed for the first time and might help to understand the mechanisms used by *Staphylococcus* to generate an efficient and mobile mechanism of resistance to an antimicrobial. On the other hand, we shed light on the contribution of *S. epidermidis* for the emergence of SCC*mec* IV and found that *S. epidermidis* adaptation to the hospital environment involved the accumulation and loss of specific mobile genetic

elements and genes involved in antibiotic resistance, colonization and biofilm formation.

References

1. **Ogston A.** 1882. *Micrococcus* Poisoning. J Anat Physiol **17:**24-58.
2. **Becker K, Heilmann C, Peters G.** 2014. Coagulase-negative staphylococci. Clin Microbiol Rev **27:**870-926.
3. **Euzéby J.** 2010. List of new names and new combinations previously effectively, but not validly, published. Int J Syst Evol Microbiol **60:**1009-1010.
4. **Anonymous.** 2015. List of Prokaryotic Names with Standing in Nomenclature. http://www.bacterio.net/staphylococcus.html. Accessed
5. **Huebner J, Goldmann DA.** 1999. Coagulase-negative staphylococci: role as pathogens. Annu Rev Med **50:**223-236.
6. **Bannerman TL, Peacock, S.J.** 2007. *Staphylococcus*, *Micrococcus*, and Other Catalase-Positive Cocci, p 390-411. *In* Murray PR (ed), Manual of Clinical Microbiology, 9th ed, vol 1. ASM Press, Washington, D.C.
7. **Tomasz A.** 2006. Staphylococcal Cell Wall, p 443-455. *In* Fischetti VA, *et al* (ed), Gram-Positive Pathogens, 2nd ed. ASM Press, Washington D.C.
8. **Lamers RP, Muthukrishnan G, Castoe TA, Tafur S, Cole AM, Parkinson CL.** 2012. Phylogenetic relationships among *Staphylococcus* species and refinement of cluster groups based on multilocus data. BMC Evol Biol **12:**171.
9. **Riley LW.** 2004. Laboratory Methods used for Strains Typing of Pathogens: Conventional and Molecular Techniques, p 29-62. *In* Riley LW (ed), Molecular Epidemiology of Infectious Diseases: Principles and Practices, 1st ed. ASM Press, Washington D.C.
10. **van Belkum A, van Leeuwen W, Kaufmann ME, Cookson B, Forey F, Etienne J, Goering R, Tenover F, Steward C, O'Brien F, Grubb W, Tassios P, Legakis N, Morvan A, El Solh N, de Ryck R, Struelens M, Salmenlinna S, Vuopio-Varkila J, Kooistra M, Talens A, Witte W, Verbrugh H.** 1998. Assessment of resolution and intercenter reproducibility of results of genotyping *Staphylococcus aureus* by pulsed-field gel electrophoresis of SmaI macrorestriction fragments: a multicenter study. J Clin Microbiol **36:**1653-1659.
11. **Struelens MJ.** 1996. Consensus guidelines for appropriate use and evaluation of microbial epidemiologic typing systems. Clin Microbiol Infect **2:**2-11.
12. **Tenover FC, Arbeit RD, Goering RV, Mickelsen PA, Murray BE, Persing DH, Swaminathan B.** 1995. Interpreting chromosomal DNA restriction patterns produced by pulsed-field gel electrophoresis: criteria for bacterial strain typing. J Clin Microbiol **33:**2233-2239.
13. **Miragaia M, Carrico JA, Thomas JC, Couto I, Enright MC, de Lencastre H.** 2008. Comparison of molecular typing methods for characterization of *Staphylococcus epidermidis*: proposal for clone definition. J Clin Microbiol **46:**118-129.
14. **Hanage WP, Fraser, C., Connor, T. R., Spratt, B. G.** 2008. Multilocus Models of Bacterial Population Genetics, p 93-104. *In* Baquero F, Nombela, C., Cassel, G. H., Gutiérrez-Fuentes, J. A. (ed), Evolutionary Biology of Bacterial and Fungal Pathogens, 1st ed. ASM Press, Washington D.C.
15. **Feil EJ, Cooper JE, Grundmann H, Robinson DA, Enright MC, Berendt T, Peacock SJ, Smith JM, Murphy M, Spratt BG, Moore CE, Day NP.** 2003. How clonal is *Staphylococcus aureus*? J Bacteriol **185:**3307-3316.
16. **Thomas JC, Vargas MR, Miragaia M, Peacock SJ, Archer GL, Enright MC.** 2007. Improved multilocus sequence typing scheme for *Staphylococcus epidermidis*. J Clin Microbiol **45:**616-619.
17. **Miragaia M, Thomas JC, Couto I, Enright MC, de Lencastre H.** 2007. Inferring a population structure for *Staphylococcus epidermidis* from multilocus sequence typing data. J Bacteriol **189:**2540-2552.
18. **Sanger F, Coulson AR.** 1975. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. J Mol Biol **94:**441-448.

19. **Morozova O, Marra MA.** 2008. Applications of next-generation sequencing technologies in functional genomics. Genomics **92:**255-264.
20. **Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer ML, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, et al.** 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature **437:**376-380.
21. **Ronaghi M, Karamohamed S, Pettersson B, Uhlén M, Nyrén P.** 1996. Real-time DNA sequencing using detection of pyrophosphate release. Anal Biochem **242:**84-89.
22. **Didelot X, Bowden R, Wilson DJ, Peto TE, Crook DW.** 2012. Transforming clinical microbiology with bacterial genome sequencing. Nat Rev Genet **13:**601-612.
23. **Risse J, Thomson M, Patrick S, Blakely G, Koutsovoulos G, Blaxter M, Watson M.** 2015. A single chromosome assembly of *Bacteroides fragilis* strain BE1 from Illumina and MinION nanopore sequencing data. Gigascience **4:**60.
24. **Loman NJ, Quick J, Simpson JT.** 2015. A complete bacterial genome assembled de novo using only nanopore sequencing data. Nat Methods **12:**733-735.
25. **Harris SR, Feil EJ, Holden MT, Quail MA, Nickerson EK, Chantratita N, Gardete S, Tavares A, Day N, Lindsay JA, Edgeworth JD, de Lencastre H, Parkhill J, Peacock SJ, Bentley SD.** 2010. Evolution of MRSA during hospital transmission and intercontinental spread. Science **327:**469-474.
26. **Harris SR, Cartwright EJ, Török ME, Holden MT, Brown NM, Ogilvy-Stuart AL, Ellington MJ, Quail MA, Bentley SD, Parkhill J, Peacock SJ.** 2013. Whole-genome sequencing for analysis of an outbreak of meticillin-resistant *Staphylococcus aureus*: a descriptive study. Lancet Infect Dis **13:**130-136.
27. **Holden MT, Feil EJ, Lindsay JA, Peacock SJ, Day NP, Enright MC, Foster TJ, Moore CE, Hurst L, Atkin R, Barron A, Bason N, Bentley SD, Chillingworth C, Chillingworth T, Churcher C, Clark L, Corton C, Cronin A, Doggett J, Dowd L, Feltwell T, Hance Z, Harris B, Hauser H, Holroyd S, Jagels K, James KD, Lennard N, Line A, Mayes R, Moule S, Mungall K, Ormond D, Quail MA, Rabbinowitsch E, Rutherford K, Sanders M, Sharp S, Simmonds M, Stevens K, Whitehead S, Barrell BG, Spratt BG, Parkhill J.** 2004. Complete genomes of two clinical *Staphylococcus aureus strains*: evidence for the rapid evolution of virulence and drug resistance. Proc Natl Acad Sci U S A **101:**9786-9791.
28. **Köser CU, Holden MT, Ellington MJ, Cartwright EJ, Brown NM, Ogilvy-Stuart AL, Hsu LY, Chewapreecha C, Croucher NJ, Harris SR, Sanders M, Enright MC, Dougan G, Bentley SD, Parkhill J, Fraser LJ, Betley JR, Schulz-Trieglaff OB, Smith GP, Peacock SJ.** 2012. Rapid whole-genome sequencing for investigation of a neonatal MRSA outbreak. N Engl J Med **366:**2267-2275.
29. **Lindsay JA.** 2014. Evolution of *Staphylococcus aureus* and MRSA during outbreaks. Infect Genet Evol **21:**548-553.
30. **Drummond AJ, Suchard MA, Xie D, Rambaut A.** 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol Biol Evol **29:**1969-1973.
31. **Castillo-Ramírez S, Corander J, Marttinen P, Aldeljawi M, Hanage WP, Westh H, Boye K, Gulay Z, Bentley SD, Parkhill J, Holden MT, Feil EJ.** 2012. Phylogeographic variation in recombination rates within a global clone of methicillin-resistant *Staphylococcus aureus*. Genome Biol **13:**R126.
32. **Hsu LY, Harris SR, Chlebowicz MA, Lindsay JA, Koh TH, Krishnan P, Tan TY, Hon PY, Grubb WB, Bentley SD, Parkhill J, Peacock SJ, Holden MT.** 2015. Evolutionary

dynamics of methicillin-resistant *Staphylococcus aureus* within a healthcare system. Genome Biol **16:**81.

33. **Zerbino DR, Birney E.** 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res **18:**821-829.
34. **Darling AC, Mau B, Blattner FR, Perna NT.** 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res **14:**1394-1403.
35. **Bashir A, Klammer AA, Robins WP, Chin CS, Webster D, Paxinos E, Hsu D, Ashby M, Wang S, Peluso P, Sebra R, Sorenson J, Bullard J, Yen J, Valdovino M, Mollova E, Luong K, Lin S, LaMay B, Joshi A, Rowe L, Frace M, Tarr CL, Turnsek M, Davis BM, Kasarskis A, Mekalanos JJ, Waldor MK, Schadt EE.** 2012. A hybrid approach for the automated finishing of bacterial genomes. Nat Biotechnol **30:**701-707.
36. **Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA.** 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol **19:**455-477.
37. **Myers EW, Sutton GG, Delcher AL, Dew IM, Fasulo DP, Flanigan MJ, Kravitz SA, Mobarry CM, Reinert KH, Remington KA, Anson EL, Bolanos RA, Chou HH, Jordan CM, Halpern AL, Lonardi S, Beasley EM, Brandon RC, Chen L, Dunn PJ, Lai Z, Liang Y, Nusskern DR, Zhan M, Zhang Q, Zheng X, Rubin GM, Adams MD, Venter JC.** 2000. A whole-genome assembly of *Drosophila*. Science **287:**2196-2204.
38. **Jolley KA, Maiden MC.** 2010. BIGSdb: Scalable analysis of bacterial genome variation at the population level. BMC Bioinformatics **11:**595.
39. **Leopold SR, Goering RV, Witten A, Harmsen D, Mellmann A.** 2014. Bacterial whole-genome sequencing revisited: portable, scalable, and standardized analysis for typing and detection of virulence and antibiotic resistance genes. J Clin Microbiol **52:**2365-2370.
40. **Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG.** 2007. Clustal W and Clustal X version 2.0. Bioinformatics **23:**2947-2948.
41. **Edgar RC.** 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res **32:**1792-1797.
42. **Méric G, Miragaia M, de Been M, Yahara K, Pascoe B, Mageiros L, Mikhail J, Harris LG, Wilkinson TS, Rolo J, Lamble S, Bray JE, Jolley KA, Hanage WP, Bowden R, Maiden MC, Mack D, de Lencastre H, Feil EJ, Corander J, Sheppard SK.** 2015. Ecological Overlap and Horizontal Gene Transfer in *Staphylococcus aureus* and *Staphylococcus epidermidis*. Genome Biol Evol **7:**1313-1328.
43. **Bradley P, Gordon NC, Walker TM, Dunn L, Heys S, Huang B, Earle S, Pankhurst LJ, Anson L, de Cesare M, Piazza P, Votintseva AA, Golubchik T, Wilson DJ, Wyllie DH, Diel R, Niemann S, Feuerriegel S, Kohl TA, Ismail N, Omar SV, Smith EG, Buck D, McVean G, Walker AS, Peto TE, Crook DW, Iqbal Z.** 2015. Rapid antibiotic-resistance predictions from genome sequence data for *Staphylococcus aureus* and *Mycobacterium tuberculosis*. Nat Commun **6:**10063.
44. **Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B.** 2006. AUGUSTUS: ab initio prediction of alternative transcripts. Nucleic Acids Res **34:**W435-439.
45. **Ekblom R, Wolf JB.** 2014. A field guide to whole-genome sequencing, assembly and annotation. Evol Appl **7:**1026-1042.
46. **Capone KA, Dowd SE, Stamatas GN, Nikolovski J.** 2011. Diversity of the human skin microbiome early in life. J Invest Dermatol **131:**2026-2032.
47. **Kloos WE, Bannerman TL.** 1994. Update on clinical significance of coagulase-negative staphylococci. Clin Microbiol Rev **7:**117-140.
48. **Weiss S, Kadlec K, Fessler AT, Schwarz S.** 2013. Identification and characterization of methicillin-resistant *Staphylococcus aureus*, *Staphylococcus epidermidis*,

Staphylococcus haemolyticus and *Staphylococcus pettenkoferi* from a small animal clinic. Vet Microbiol **167:**680-685.

49. **McManus BA, Coleman DC, Deasy EC, Brennan GI, O' Connell B, Monecke S, Ehricht R, Leggett B, Leonard N, Shore AC.** 2015. Comparative Genotypes, Staphylococcal Cassette Chromosome *mec* (SCC*mec*) Genes and Antimicrobial Resistance amongst *Staphylococcus epidermidis* and *Staphylococcus haemolyticus* Isolates from Infections in Humans and Companion Animals. PLoS One **10:**e0138079.
50. **Vanderhaeghen W, Vandendriessche S, Crombe F, Dispas M, Denis O, Hermans K, Haesebrouck F, Butaye P.** 2012. Species and staphylococcal cassette chromosome *mec* (SCC*mec*) diversity among methicillin-resistant non-*Staphylococcus aureus* staphylococci isolated from pigs. Vet Microbiol.
51. **Otto M.** 2009. *Staphylococcus epidermidis* - the 'accidental' pathogen. Nat Rev Microbiol **7:**555-567.
52. **Gill SR, Fouts DE, Archer GL, Mongodin EF, Deboy RT, Ravel J, Paulsen IT, Kolonay JF, Brinkac L, Beanan M, Dodson RJ, Daugherty SC, Madupu R, Angiuoli SV, Durkin AS, Haft DH, Vamathevan J, Khouri H, Utterback T, Lee C, Dimitrov G, Jiang L, Qin H, Weidman J, Tran K, Kang K, Hance IR, Nelson KE, Fraser CM.** 2005. Insights on evolution of virulence and resistance from the complete genome analysis of an early methicillin-resistant *Staphylococcus aureus* strain and a biofilm-producing methicillin-resistant *Staphylococcus epidermidis* strain. J Bacteriol **187:**2426-2438.
53. **Iwase T, Uehara Y, Shinji H, Tajima A, Seo H, Takada K, Agata T, Mizunoe Y.** 2010. Staphylococcus epidermidis Esp inhibits *Staphylococcus aureus* biofilm formation and nasal colonization. Nature **465:**346-349.
54. **Cogen AL, Yamasaki K, Sanchez KM, Dorschner RA, Lai Y, MacLeod DT, Torpey JW, Otto M, Nizet V, Kim JE, Gallo RL.** 2010. Selective antimicrobial action is provided by phenol-soluble modulins derived from *Staphylococcus epidermidis*, a normal resident of the skin. J Invest Dermatol **130:**192-200.
55. **Cogen AL, Yamasaki K, Muto J, Sanchez KM, Crotty Alexander L, Tanios J, Lai Y, Kim JE, Nizet V, Gallo RL.** 2010. *Staphylococcus epidermidis* antimicrobial delta-toxin (phenol-soluble modulin-gamma) cooperates with host antimicrobial peptides to kill group A *Streptococcus*. PLoS One **5:**e8557.
56. **Lina G, Boutite F, Tristan A, Bes M, Etienne J, Vandenesch F.** 2003. Bacterial competition for human nasal cavity colonization: role of Staphylococcal *agr* alleles. Appl Environ Microbiol **69:**18-23.
57. **Jack RW, Tagg JR, Ray B.** 1995. Bacteriocins of gram-positive bacteria. Microbiol Rev **59:**171-200.
58. **Sandiford S, Upton M.** 2012. Identification, characterization, and recombinant expression of epidermicin NI01, a novel unmodified bacteriocin produced by *Staphylococcus epidermidis* that displays potent activity against Staphylococci. Antimicrob Agents Chemother **56:**1539-1547.
59. **Bennallack PR, Burt SR, Heder MJ, Robison RA, Griffitts JS.** 2014. Characterization of a novel plasmid-borne thiopeptide gene cluster in *Staphylococcus epidermidis* strain 115. J Bacteriol **196:**4344-4350.
60. **Stevens NT, Sadovskaya I, Jabbouri S, Sattar T, O'Gara JP, Humphreys H, Greene CM.** 2009. *Staphylococcus epidermidis* polysaccharide intercellular adhesin induces IL-8 expression in human astrocytes via a mechanism involving TLR2. Cell Microbiol **11:**421-432.
61. **Dong Y, Speer CP.** 2014. The role of *Staphylococcus epidermidis* in neonatal sepsis: guarding angel or pathogenic devil? Int J Med Microbiol **304:**513-520.
62. **Milisavljevic V, Wu F, Cimmotti J, Haas J, Della-Latta P, Larson E, Saiman L.** 2005. Genetic relatedness of *Staphylococcus epidermidis* from infected infants and staff in the neonatal intensive care unit. Am J Infect Control **33:**341-347.

63. **Tammelin A, Domicel P, Hambraeus A, Stahle E.** 2000. Dispersal of methicillin-resistant *Staphylococcus epidermidis* by staff in an operating suite for thoracic and cardiovascular surgery: relation to skin carriage and clothing. J Hosp Infect **44:**119-126.
64. **Brook I.** 2002. Bacterial contamination of saline nasal spray/drop solution in patients with respiratory tract infection. Am J Infect Control **30:**246-247.
65. **Hedin G.** 1993. *Staphylococcus epidermidis*--hospital epidemiology and the detection of methicillin resistance. Scand J Infect Dis Suppl **90:**1-59.
66. **Kanai H, Sato H, Takei Y.** 2014. Community-acquired methicillin-resistant *Staphylococcus epidermidis* pyelonephritis in a child: a case report. J Med Case Rep **8:**415.
67. **Kitterer D, Latus J, Pöhlmann C, Alscher MD, Kimmel M.** 2015. Microbiological Surveillance of Peritoneal Dialysis Associated Peritonitis: Antimicrobial Susceptibility Profiles of a Referral Center in GERMANY over 32 Years. PLoS One **10:**e0135969.
68. **Griffith GC, Levinson DC.** 1949. Subacute bacterial endocarditis; a report on 57 patients treated with massive doses of penicillin. Calif Med **71:**403-408.
69. **Jevons MP.** 1961. 'Celbenin' - Resistant Staphylococci. British Medical Journal. **1**: 124-125.
70. **Kjellander JO, Klein JO, Finland M.** 1963. In vitro activity of penicillins against *Staphylococcus albus*. Proc Soc Exp Biol Med **113:**1023-1031.
71. **Cherifi S, Byl B, Deplano A, Nonhoff C, Denis O, Hallin M.** 2013. Comparative epidemiology of *Staphylococcus epidermidis* isolates from patients with catheter-related bacteremia and from healthy volunteers. J Clin Microbiol **51:**1541-1547.
72. **Gordon RJ, Miragaia M, Weinberg AD, Lee CJ, Rolo J, Giacalone JC, Slaughter MS, Pappas P, Naka Y, Tector AJ, de Lencastre H, Lowy FD.** 2012. *Staphylococcus epidermidis* colonization is highly clonal across US cardiac centers. J Infect Dis. **205 (9)**:1391-8.
73. **Rogers KL, Fey PD, Rupp ME.** 2009. Coagulase-negative staphylococcal infections. Infect Dis Clin North Am **23:**73-98.
74. **May L, Klein EY, Rothman RE, Laxminarayan R.** 2014. Trends in antibiotic resistance in coagulase-negative staphylococci in the United States, 1999 to 2012. Antimicrob Agents Chemother **58:**1404-1409.
75. **Schwalbe RS, Stapleton JT, Gilligan PH.** 1987. Emergence of vancomycin resistance in coagulase-negative staphylococci. N Engl J Med **316:**927-931.
76. **Bonilla H, Huband MD, Seidel J, Schmidt H, Lescoe M, McCurdy SP, Lemmon MM, Brennan LA, Tait-Kamradt A, Puzniak L, Quinn JP.** 2010. Multicity outbreak of linezolid-resistant *Staphylococcus epidermidis* associated with clonal spread of a cfr-containing strain. Clin Infect Dis **51:**796-800.
77. **Kozitskaya S, Olson ME, Fey PD, Witte W, Ohlsen K, Ziebuhr W.** 2005. Clonal analysis of *Staphylococcus epidermidis* isolates carrying or lacking biofilm-mediating genes by multilocus sequence typing. J Clin Microbiol **43:**4751-4757.
78. **Kozitskaya S, Cho SH, Dietrich K, Marre R, Naber K, Ziebuhr W.** 2004. The bacterial insertion sequence element IS*256* occurs preferentially in nosocomial *Staphylococcus epidermidis* isolates: association with biofilm formation and resistance to aminoglycosides. Infect Immun **72:**1210-1215.
79. **Miragaia M, Couto I, Pereira SF, Kristinsson KG, Westh H, Jarlov JO, Carrico J, Almeida J, Santos-Sanches I, de Lencastre H.** 2002. Molecular characterization of methicillin-resistant *Staphylococcus epidermidis* clones: evidence of geographic dissemination. J Clin Microbiol **40:**430-438.
80. **Jamaluddin TZ, Kuwahara-Arai K, Hisata K, Terasawa M, Cui L, Baba T, Sotozono C, Kinoshita S, Ito T, Hiramatsu K.** 2008. Extreme genetic diversity of methicillin-resistant *Staphylococcus epidermidis* strains disseminated among healthy Japanese children. J Clin Microbiol **46:**3778-3783.

81. **Widerström M, Wiström J, Ek E, Edebro H, Monsen T.** 2011. Near absence of methicillin-resistance and pronounced genetic diversity among *Staphylococcus epidermidis* isolated from healthy persons in northern Sweden. APMIS **119:**505-512.
82. **Silva FR, Mattos EM, Coimbra MV, Ferreira-Carvalho BT, Figueiredo AM.** 2001. Isolation and molecular characterization of methicillin-resistant coagulase-negative staphylococci from nasal flora of healthy humans at three community institutions in Rio de Janeiro City. Epidemiol Infect **127:**57-62.
83. **O'Gara JP, Humphreys H.** 2001. *Staphylococcus epidermidis* biofilms: importance and implications. J Med Microbiol **50:**582-587.
84. **O'Gara JP.** 2007. *ica* and beyond: biofilm mechanisms and regulation in *Staphylococcus epidermidis* and *Staphylococcus aureus*. FEMS Microbiol Lett **270:**179-188.
85. **Wang R, Khan BA, Cheung GY, Bach TH, Jameson-Lee M, Kong KF, Queck SY, Otto M.** 2011. *Staphylococcus epidermidis* surfactant peptides promote biofilm maturation and dissemination of biofilm-associated infection in mice. J Clin Invest **121:**238-248.
86. **Ziebuhr W, Krimmer V, Rachid S, Lossner I, Gotz F, Hacker J.** 1999. A novel mechanism of phase variation of virulence in *Staphylococcus epidermidis*: evidence for control of the polysaccharide intercellular adhesin synthesis by alternating insertion and excision of the insertion sequence element IS*256*. Mol Microbiol **32:**345-356.
87. **Nuryastuti T, van der Mei HC, Busscher HJ, Kuijer R, Aman AT, Krom BP.** 2008. *recA* mediated spontaneous deletions of the *icaADBC* operon of clinical *Staphylococcus epidermidis* isolates: a new mechanism of phenotypic variations. Antonie Van Leeuwenhoek **94:**317-328.
88. **Arciola CR, Baldassarri L, Montanaro L.** 2002. In catheter infections by *Staphylococcus epidermidis* the intercellular adhesion (*ica*) locus is a molecular marker of the virulent slime-producing strains. J Biomed Mater Res **59:**557-562.
89. **Arciola CR, Campoccia D, Gamberini S, Donati ME, Baldassarri L, Montanaro L.** 2003. Occurrence of *ica* genes for slime synthesis in a collection of *Staphylococcus epidermidis* strains from orthopedic prosthesis infections. Acta Orthop Scand **74:**617-621.
90. **Koskela A, Nilsdotter-Augustinsson A, Persson L, Soderquist B.** 2009. Prevalence of the *ica* operon and insertion sequence IS*256* among *Staphylococcus epidermidis* prosthetic joint infection isolates. Eur J Clin Microbiol Infect Dis **28:**655-660.
91. **Ninin E, Caroff N, Espaze E, Maraillac J, Lepelletier D, Milpied N, Richet H.** 2006. Assessment of *ica* operon carriage and biofilm production in *Staphylococcus epidermidis* isolates causing bacteraemia in bone marrow transplant recipients. Clin Microbiol Infect **12:**446-452.
92. **Ziebuhr W, Hennig S, Eckart M, Kranzler H, Batzilla C, Kozitskaya S.** 2006. Nosocomial infections by *Staphylococcus epidermidis*: how a commensal bacterium turns into a pathogen. Int J Antimicrob Agents **28 Suppl 1:**S14-20.
93. **Cheung GY, Joo HS, Chatterjee SS, Otto M.** 2014. Phenol-soluble modulins--critical determinants of staphylococcal virulence. FEMS Microbiol Rev **38:**698-719.
94. **Mehlin C, Headley CM, Klebanoff SJ.** 1999. An inflammatory polypeptide complex from *Staphylococcus epidermidis*: isolation and characterization. J Exp Med **189:**907-918.
95. **Wang R, Braughton KR, Kretschmer D, Bach TH, Queck SY, Li M, Kennedy AD, Dorward DW, Klebanoff SJ, Peschel A, DeLeo FR, Otto M.** 2007. Identification of novel cytolytic peptides as key virulence determinants for community-associated MRSA. Nat Med **13:**1510-1514.
96. **Cheung GY, Rigby K, Wang R, Queck SY, Braughton KR, Whitney AR, Teintze M, DeLeo FR, Otto M.** *Staphylococcus epidermidis* strategies to avoid killing by human neutrophils. PLoS Pathog **6**.

97. **Diep BA, Gill SR, Chang RF, Phan TH, Chen JH, Davidson MG, Lin F, Lin J, Carleton HA, Mongodin EF, Sensabaugh GF, Perdreau-Remington F.** 2006. Complete genome sequence of USA300, an epidemic clone of community-acquired meticillin-resistant *Staphylococcus aureus*. Lancet **367:**731-739.
98. **Diep BA, Stone GG, Basuino L, Graber CJ, Miller A, des Etages SA, Jones A, Palazzolo-Ballance AM, Perdreau-Remington F, Sensabaugh GF, DeLeo FR, Chambers HF.** 2008. The arginine catabolic mobile element and staphylococcal chromosomal cassette *mec* linkage: convergence of virulence and resistance in the USA300 clone of methicillin-resistant *Staphylococcus aureus*. J Infect Dis **197:**1523-1530.
99. **Zhang YQ, Ren SX, Li HL, Wang YX, Fu G, Yang J, Qin ZQ, Miao YG, Wang WY, Chen RS, Shen Y, Chen Z, Yuan ZH, Zhao GP, Qu D, Danchin A, Wen YM.** 2003. Genome-based analysis of virulence genes in a non-biofilm-forming *Staphylococcus epidermidis* strain (ATCC 12228). Mol Microbiol **49:**1577-1593.
100. **Miragaia M, de Lencastre H, Perdreau-Remington F, Chambers HF, Higashi J, Sullam PM, Lin J, Wong KI, King KA, Otto M, Sensabaugh GF, Diep BA.** 2009. Genetic diversity of arginine catabolic mobile element in *Staphylococcus epidermidis*. PLoS One **4:**e7722.
101. **Barbier F, Lebeaux D, Hernandez D, Delannoy AS, Caro V, François P, Schrenzel J, Ruppé E, Gaillard K, Wolff M, Brisse S, Andremont A, Ruimy R.** 2011. High prevalence of the arginine catabolic mobile element in carriage isolates of methicillin-resistant *Staphylococcus epidermidis*. J Antimicrob Chemother **66:**29-36.
102. **Lindgren JK, Thomas VC, Olson ME, Chaudhari SS, Nuxoll AS, Schaeffer CR, Lindgren KE, Jones J, Zimmerman MC, Dunman PM, Bayles KW, Fey PD.** 2014. Arginine deiminase in *Staphylococcus epidermidis* functions to augment biofilm maturation through pH homeostasis. J Bacteriol **196:**2277-2289.
103. **Thurlow LR, Joshi GS, Clark JR, Spontak JS, Neely CJ, Maile R, Richardson AR.** 2013. Functional modularity of the arginine catabolic mobile element contributes to the success of USA300 methicillin-resistant *Staphylococcus aureus*. Cell Host Microbe **13:**100-107.
104. **Kocianova S, Vuong C, Yao Y, Voyich JM, Fischer ER, DeLeo FR, Otto M.** 2005. Key role of poly-gamma-DL-glutamic acid in immune evasion and virulence of *Staphylococcus epidermidis*. J Clin Invest **115:**688-694.
105. **Vuong C, Voyich JM, Fischer ER, Braughton KR, Whitney AR, DeLeo FR, Otto M.** 2004. Polysaccharide intercellular adhesin (PIA) protects *Staphylococcus epidermidis* against major components of the human innate immune system. Cell Microbiol **6:**269-275.
106. **Chambers HF, Deleo FR.** 2009. Waves of resistance: *Staphylococcus aureus* in the antibiotic era. Nat Rev Microbiol **7:**629-641.
107. **Pinho MG, de Lencastre H, Tomasz A.** 2001. An acquired and a native penicillin-binding protein cooperate in building the cell wall of drug-resistant staphylococci. Proc Natl Acad Sci U S A **98:**10886-10891.
108. **Lim D, Strynadka NC.** 2002. Structural basis for the beta lactam resistance of PBP2a from methicillin-resistant *Staphylococcus aureus*. Nat Struct Biol **9:**870-876.
109. **Katayama Y, Ito T, Hiramatsu K.** 2000. A new class of genetic element, staphylococcus cassette chromosome *mec*, encodes methicillin resistance in *Staphylococcus aureus*. Antimicrob Agents Chemother **44:**1549-1555.
110. **Wisplinghoff H, Rosato AE, Enright MC, Noto M, Craig W, Archer GL.** 2003. Related clones containing SCC*mec* type IV predominate among clinically significant *Staphylococcus epidermidis* isolates. Antimicrob Agents Chemother **47:**3574-3579.
111. **IWG-SCC.** 2009. Classification of staphylococcal cassette chromosome *mec* (SCC*mec*): guidelines for reporting novel SCCmec elements. Antimicrob Agents Chemother **53:**4961-4967.
112. **Li S, Skov RL, Han X, Larsen AR, Larsen J, Sorum M, Wulf M, Voss A, Hiramatsu K, Ito T.** 2011. Novel Types of Staphylococcal Cassette Chromosome *mec* Elements Identified

in Clonal Complex 398 Methicillin-Resistant *Staphylococcus aureus* Strains. Antimicrob Agents Chemother **55:**3046-3050.

113. **Garcia-Alvarez L, Holden MT, Lindsay H, Webb CR, Brown DF, Curran MD, Walpole E, Brooks K, Pickard DJ, Teale C, Parkhill J, Bentley SD, Edwards GF, Girvan EK, Kearns AM, Pichon B, Hill RL, Larsen AR, Skov RL, Peacock SJ, Maskell DJ, Holmes MA.** Meticillin-resistant *Staphylococcus aureus* with a novel *mecA* homologue in human and bovine populations in the UK and Denmark: a descriptive study. Lancet Infect Dis **11 (8)**: 595–603.
114. **Shore AC, Deasy EC, Slickers P, Brennan G, O'Connell B, Monecke S, Ehricht R, Coleman DC.** Detection of staphylococcal cassette chromosome *mec* type XI carrying highly divergent *mecA, mecI, mecR1, blaZ,* and *ccr* genes in human clinical isolates of clonal complex 130 methicillin-resistant *Staphylococcus aureus*. Antimicrob Agents Chemother **55:**3765-3773.
115. **Boundy S, Safo MK, Wang L, Musayev FN, O'Farrell HC, Rife JP, Archer GL.** 2013. Characterization of the *Staphylococcus aureus* rRNA methyltransferase encoded by *orfX*, the gene containing the staphylococcal chromosome Cassette *mec* (SCC*mec*) insertion site. J Biol Chem **288:**132-140.
116. **Cohen S, Sweeney HM.** 1970. Transduction of Methicillin Resistance in *Staphylococcus aureus* Dependent on an Unusual Specificity of the Recipient Strain. J Bacteriol **104:**1158-1167.
117. **Mašlaňová I, Doškař J, Varga M, Kuntová L, Mužík J, Malúšková D, Růžičková V, Pantůček R.** 2013. Bacteriophages of *Staphylococcus aureus* efficiently package various bacterial genes and mobile genetic elements including SCC*mec* with different frequencies. Environ Microbiol Rep **5:**66-73.
118. **Chlebowicz MA, Mašlaňová I, Kuntová L, Grundmann H, Pantůček R, Doškař J, van Dijl JM, Buist G.** 2014. The Staphylococcal Cassette Chromosome *mec* type V from *Staphylococcus aureus* ST398 is packaged into bacteriophage capsids. Int J Med Microbiol **304:**764-774.
119. **Morikawa K, Takemura AJ, Inose Y, Tsai M, Nguyen Thi lT, Ohta T, Msadek T.** 2012. Expression of a cryptic secondary sigma factor gene unveils natural competence for DNA transformation in *Staphylococcus aureus*. PLoS Pathog **8:**e1003003.
120. **Tsubakishita S, Kuwahara-Arai K, Baba T, Hiramatsu K.** 2010. Staphylococcal cassette chromosome *mec*-like element in *Macrococcus caseolyticus*. Antimicrob Agents Chemother **54:**1469-1475.
121. **de Lencastre H, Chung M, Westh H.** 2000. Archaic strains of methicillin-resistant *Staphylococcus aureus*: molecular and microbiological properties of isolates from the 1960s in Denmark. Microb Drug Resist **6:**1-10.
122. **Zhang L, Thomas JC, Miragaia M, Bouchami O, Chaves F, d'Azevedo PA, Aanensen DM, de Lencastre H, Gray BM, Robinson DA.** 2013. Multilocus sequence typing and further genetic characterization of the enigmatic pathogen, *Staphylococcus hominis*. PLoS One **8:**e66496.
123. **Deurenberg RH, Stobberingh EE.** 2009. The molecular evolution of hospital- and community-associated methicillin-resistant *Staphylococcus aureus*. Curr Mol Med **9:**100-115.
124. **Miragaia M, Couto I, de Lencastre H.** 2005. Genetic diversity among methicillin-resistant *Staphylococcus epidermidis* (MRSE). Microb Drug Resist **11:**83-93.
125. **Okuma K, Iwakawa K, Turnidge JD, Grubb WB, Bell JM, O'Brien FG, Coombs GW, Pearman JW, Tenover FC, Kapi M, Tiensasitorn C, Ito T, Hiramatsu K.** 2002. Dissemination of new methicillin-resistant *Staphylococcus aureus* clones in the community. J Clin Microbiol **40:**4289-4294.
126. **Cavanagh JP, Hjerde E, Holden MT, Kahlke T, Klingenberg C, Flægstad T, Parkhill J, Bentley SD, Sollid JU.** 2014. Whole-genome sequencing reveals clonal expansion of

multiresistant *Staphylococcus haemolyticus* in European hospitals. J Antimicrob Chemother **69:**2920-2927.

127. **Ito T, Ma XX, Takeuchi F, Okuma K, Yuzawa H, Hiramatsu K.** 2004. Novel type V staphylococcal cassette chromosome *mec* driven by a novel cassette chromosome recombinase, *ccrC*. Antimicrob Agents Chemother **48:**2637-2651.
128. **Oliveira DC, Milheirico C, de Lencastre H.** 2006. Redefining a structural variant of staphylococcal cassette chromosome *mec*, SCC*mec* type VI. Antimicrob Agents Chemother **50:**3457-3459.
129. **Takano T, Higuchi W, Otsuka T, Baranovich T, Enany S, Saito K, Isobe H, Dohmae S, Ozaki K, Takano M, Iwao Y, Shibuya M, Okubo T, Yabe S, Shi D, Reva I, Teng LJ, Yamamoto T.** 2008. Novel characteristics of community-acquired methicillin-resistant *Staphylococcus aureus* strains belonging to multilocus sequence type 59 in Taiwan. Antimicrob Agents Chemother **52:**837-845.
130. **Zhang K, McClure JA, Elsayed S, Conly JM.** 2009. Novel staphylococcal cassette chromosome *mec* type, tentatively designated type VIII, harboring class A *mec* and type 4 *ccr* gene complexes in a Canadian epidemic strain of methicillin-resistant *Staphylococcus aureus*. Antimicrob Agents Chemother **53:**531-540.
131. **Bouchami O, Ben Hassen A, de Lencastre H, Miragaia M.** 2011. Molecular epidemiology of methicillin-resistant *Staphylococcus hominis* (MRSHo): low clonality and reservoirs of SCC*mec* structural elements. PLoS One **6:**e21940.
132. **Harrison EM, Paterson GK, Holden MT, Morgan FJ, Larsen AR, Petersen A, Leroy S, De Vliegher S, Perreten V, Fox LK, Lam TJ, Sampimon OC, Zadoks RN, Peacock SJ, Parkhill J, Holmes MA.** 2013. A *Staphylococcus xylosus* isolate with a new *mecC* allotype. Antimicrob Agents Chemother **57:**1524-1528.
133. **Harrion EM, Paterson GK, Holden MT, Ba X, Rolo J, Morgan FJ, Pichon B, Kearns A, Zadocks RN, Peacock SJ, Parkhill J, Holmes MA.** 2013. A novel hybrid SCC*mec-mecC* region in *Staphylococcus sciuri*. J Antimicrob Chemother **69 (4)**:911-8.
134. **Ryffel C, Kayser FH, Berger-Bächi B.** 1992. Correlation between regulation of *mecA* transcription and expression of methicillin resistance in staphylococci. Antimicrob Agents Chemother **36:**25-31.
135. **Arêde P, Milheiriço C, de Lencastre H, Oliveira DC.** 2012. The anti-repressor MecR2 promotes the proteolysis of the *mecA* repressor and enables optimal expression of β-lactam resistance in MRSA. PLoS Pathog **8:**e1002816.
136. **Ryffel C, Bucher R, Kayser FH, Berger-Bächi B.** 1991. The *Staphylococcus aureus mec* determinant comprises an unusual cluster of direct repeats and codes for a gene product similar to the *Escherichia coli* sn-glycerophosphoryl diester phosphodiesterase. J Bacteriol **173:**7416-7422.
137. **Queck SY, Khan BA, Wang R, Bach TH, Kretschmer D, Chen L, Kreiswirth BN, Peschel A, Deleo FR, Otto M.** 2009. Mobile genetic element-encoded cytolysin connects virulence to methicillin resistance in MRSA. PLoS Pathog **5:**e1000533.
138. **Ito T, Katayama Y, Asada K, Mori N, Tsutsumimoto K, Tiensasitorn C, Hiramatsu K.** 2001. Structural comparison of three types of staphylococcal cassette chromosome *mec* integrated in the chromosome in methicillin-resistant *Staphylococcus aureus*. Antimicrob Agents Chemother **45:**1323-1336.
139. **Oliveira DC, Wu SW, de Lencastre H.** 2000. Genetic organization of the downstream region of the *mecA* element in methicillin-resistant *Staphylococcus aureus* isolates carrying different polymorphisms of this region. Antimicrob Agents Chemother **44:**1906-1910.
140. **Katayama Y, Ito T, Hiramatsu K.** 2001. Genetic organization of the chromosome region surrounding *mecA* in clinical staphylococcal strains: role of IS*431*-mediated *mecI* deletion in expression of resistance in *mecA*-carrying, low-level methicillin-resistant *Staphylococcus haemolyticus*. Antimicrob Agents Chemother **45:**1955-1963.

141. **Ma XX, Ito T, Tiensasitorn C, Jamklang M, Chongtrakool P, Boyle-Vavra S, Daum RS, Hiramatsu K.** 2002. Novel type of staphylococcal cassette chromosome *mec* identified in community-acquired methicillin-resistant *Staphylococcus aureus* strains. Antimicrob Agents Chemother **46:**1147-1152.
142. **Oliveira DC, Milheirico C, Vinga S, de Lencastre H.** 2006. Assessment of allelic variation in the *ccrAB* locus in methicillin-resistant *Staphylococcus aureus* clones. J Antimicrob Chemother **58:**23-30.
143. **Wang L, Archer GL.** 2010. Roles of CcrA and CcrB in excision and integration of staphylococcal cassette chromosome *mec*, a *Staphylococcus aureus* genomic island. J Bacteriol **192:**3204-3212.
144. **Wang L, Safo M, Archer GL.** 2012. Characterization of DNA sequences required for the CcrAB-mediated integration of staphylococcal cassette chromosome *mec*, a *Staphylococcus aureus* genomic island. J Bacteriol **194:**486-498.
145. **Bjørkeng EK, Tessema GT, Lundblad EW, Butaye P, Willems R, Sollid JE, Sundsfjord A, Hegstad K.** 2010. *ccrABEnt* serine recombinase genes are widely distributed in the *Enterococcus faecium* and *Enterococcus casseliflavus* species groups and are expressed in *E. faecium*. Microbiology **156:**3624-3634.
146. **Gómez-Sanz E, Schwendener S, Thomann A, Gobeli Brawand S, Perreten V.** 2015. First Staphylococcal Cassette Chromosome *mec* Containing a *mecB*-Carrying Gene Complex Independent of Transposon Tn*6045* in a *Macrococcus caseolyticu*s Isolate from a Canine Infection. Antimicrob Agents Chemother **59:**4577-4583.
147. **Chongtrakool P, Ito T, Ma XX, Kondo Y, Trakulsomboon S, Tiensasitorn C, Jamklang M, Chavalit T, Song JH, Hiramatsu K.** 2006. Staphylococcal cassette chromosome *mec* (SCC*mec*) typing of methicillin-resistant *Staphylococcus aureus* strains isolated in 11 Asian countries: a proposal for a new nomenclature for SCC*mec* elements. Antimicrob Agents Chemother **50:**1001-1012.
148. **Luong TT, Ouyang S, Bush K, Lee CY.** 2002. Type 1 capsule genes of *Staphylococcus aureus* are carried in a staphylococcal cassette chromosome genetic element. J Bacteriol **184:**3623-3629.
149. **Mongkolrattanothai K, Boyle S, Murphy TV, Daum RS.** 2004. Novel non-*mecA*-containing staphylococcal chromosomal cassette composite island containing *pbp4* and *tagF* genes in a commensal staphylococcal species: a possible reservoir for antibiotic resistance islands in *Staphylococcus aureus*. Antimicrob Agents Chemother **48:**1823-1836.
150. **Katayama Y, Takeuchi F, Ito T, Ma XX, Ui-Mizutani Y, Kobayashi I, Hiramatsu K.** 2003. Identification in methicillin-susceptible *Staphylococcus hominis* of an active primordial mobile genetic element for the staphylococcal cassette chromosome *mec* of methicillin-resistant *Staphylococcus aureus*. J Bacteriol **185:**2711-2722.
151. **Takeuchi F, Watanabe S, Baba T, Yuzawa H, Ito T, Morimoto Y, Kuroda M, Cui L, Takahashi M, Ankai A, Baba S, Fukui S, Lee JC, Hiramatsu K.** 2005. Whole-genome sequencing of *Staphylococcus haemolyticus* uncovers the extreme plasticity of its genome and the evolution of human-colonizing staphylococcal species. J Bacteriol **187:**7292-7308.
152. **Kuroda M, Yamashita A, Hirakawa H, Kumano M, Morikawa K, Higashide M, Maruyama A, Inose Y, Matoba K, Toh H, Kuhara S, Hattori M, Ohta T.** 2005. Whole genome sequence of *Staphylococcus saprophyticus* reveals the pathogenesis of uncomplicated urinary tract infection. Pros Natl Acad Sci U S A **102:**13272-13277.
153. **Yu D, Pi B, Chen Y, Wang Y, Ruan Z, Otto M, Yu Y.** 2014. Characterization of the staphylococcal cassette chromosome composite island of *Staphylococcus haemolyticus* SH32, a methicillin-resistant clinical isolate from China. PLoS One **9:**e87346.

154. **Zong Z.** 2013. Characterization of a complex context containing *mecA* but lacking genes encoding cassette chromosome recombinases in *Staphylococcus haemolyticus*. BMC Microbiol **13:**64.
155. **Xue H, Wu Z, Li L, Li F, Wang Y, Zhao X.** 2015. Co-existence of Heavy Metal and Antibiotic Resistance within a Novel Composite Staphylococcal Cassette Chromosome in a *Staphylococcus Haemolyticus* from Bovine Mastitis Milk. Antimicrob Agents Chemother **59 (9)**:5788-92.
156. **Berglund C, Ito T, Ikeda M, Ma XX, Soderquist B, Hiramatsu K.** 2008. Novel type of staphylococcal cassette chromosome *mec* in a methicillin-resistant *Staphylococcus aureus* strain isolated in Sweden. Antimicrob Agents Chemother **52:**3512-3516.
157. **Lindsay JA.** 2010. Genomic variation and evolution of *Staphylococcus aureus*. Int J Med Microbiol **300:**98-103.
158. **Fitzgerald JR.** 2012. Livestock-associated *Staphylococcus aureus*: origin, evolution and public health threat. Trends Microbiol **20:**192-198.
159. **Enright MC, Robinson DA, Randle G, Feil EJ, Grundmann H, Spratt BG.** 2002. The evolutionary history of methicillin-resistant *Staphylococcus aureus* (MRSA). Proc Natl Acad Sci U S A **99:**7687-7692.
160. **Oliveira DC, de Lencastre H.** 2002. Multiplex PCR strategy for rapid identification of structural types and variants of the *mec* element in methicillin-resistant *Staphylococcus aureus*. Antimicrob Agents Chemother **46:**2155-2161.
161. **Zhang K, McClure JA, Elsayed S, Louie T, Conly JM.** 2005. Novel multiplex PCR assay for characterization and concomitant subtyping of staphylococcal cassette chromosome *mec* types I to V in methicillin-resistant *Staphylococcus aureus*. J Clin Microbiol **43:**5026-5033.
162. **Milheiriço C, Oliveira DC, de Lencastre H.** 2007. Update to the multiplex PCR strategy for assignment of *mec* element types in *Staphylococcus aureus*. Antimicrob Agents Chemother **51:**3374-3377.
163. **Kondo Y, Ito T, Ma XX, Watanabe S, Kreiswirth BN, Etienne J, Hiramatsu K.** 2007. Combination of multiplex PCRs for staphylococcal cassette chromosome *mec* type assignment: rapid identification system for *mec*, *ccr*, and major differences in junkyard regions. Antimicrob Agents Chemother **51:**264-274.
164. **Boye K, Bartels MD, Andersen IS, Moller JA, Westh H.** 2007. A new multiplex PCR for easy screening of methicillin-resistant *Staphylococcus aureus* SCC*mec* types I-V. Clin Microbiol Infect **13:**725-727.
165. **Milheiriço C, Oliveira DC, de Lencastre H.** 2007. Multiplex PCR strategy for subtyping the staphylococcal cassette chromosome *mec* type IV in methicillin-resistant *Staphylococcus aureus*: 'SCC*mec* IV multiplex'. J Antimicrob Chemother **60:**42-48.
166. **Oliveira DC, Santos M, Milheirico C, Carrico JA, Vinga S, Oliveira AL, de Lencastre H.** 2008. CcrB typing tool: an online resource for staphylococci *ccrB* sequence typing. J Antimicrob Chemother **61:**959-960.
167. **Bouchami O, Ben Hassen A, de Lencastre H, Miragaia M.** 2012. High prevalence of *mec* complex C and *ccrC* is independent of SCC*mec* type V in *Staphylococcus haemolyticus*. Eur J Clin Microbiol Infect Dis **31:**605-614.
168. **Zong Z, Lü X.** 2010. Characterization of a new SCC*mec* element in *Staphylococcus cohnii*. PLoS One **5:**e14016.
169. **Ito T, Hiramatsu K, Tomasz A, de Lencastre H, Perreten V, Holden MT, Coleman DC, Goering R, Giffard PM, Skov RL, Zhang K, Westh H, O'Brien F, Tenover FC, Oliveira DC, Boyle-Vavra S, Laurent F, Kearns AM, Kreiswirth B, Ko KS, Grundmann H, Sollid JE, John JF, Daum R, Soderquist B, Buist G, (IWG-SCC) IWGotCoSCCE.** 2012. Guidelines for reporting novel *mecA* gene homologues. Antimicrob Agents Chemother **56:**4997-4999.

170. **Nemeghaire S, Argudín MA, Feßler AT, Hauschild T, Schwarz S, Butaye P.** 2014. The ecological importance of the *Staphylococcus sciuri* species group as a reservoir for resistance and virulence genes. Vet Microbiol **171:**342-356.
171. **Kloos WE, Ballard DN, Webster JA, Hubner RJ, Tomasz A, Couto I, Sloan GL, Dehart HP, Fiedler F, Schubert K, de Lencastre H, Sanches IS, Heath HE, Leblanc PA, Ljungh A.** 1997. Ribotype delineation and description of *Staphylococcus sciuri* subspecies and their potential as reservoirs of methicillin resistance and staphylolytic enzyme genes. Int J Syst Bacteriol **47:**313-323.
172. **Couto I, de Lencastre H, Severina E, Kloos W, Webster JA, Hubner RJ, Sanches IS, Tomasz A.** 1996. Ubiquitous presence of a *mecA* homologue in natural isolates of *Staphylococcus sciuri*. Microb Drug Resist **2:**377-391.
173. **Tsubakishita S, Kuwahara-Arai K, Sasaki T, Hiramatsu K.** 2010. Origin and molecular evolution of the determinant of methicillin resistance in staphylococci. Antimicrob Agents Chemother **54:**4352-4359.
174. **Wu S, Piscitelli C, de Lencastre H, Tomasz A.** 1996. Tracking the evolutionary origin of the methicillin resistance gene: cloning and sequencing of a homologue of *mecA* from a methicillin susceptible strain of *Staphylococcus sciuri*. Microb Drug Resist **2:**435-441.
175. **Zhou Y, Antignac A, Wu SW, Tomasz A.** 2008. Penicillin-binding proteins and cell wall composition in beta-lactam-sensitive and -resistant strains of *Staphylococcus sciuri*. J Bacteriol **190:**508-514.
176. **Antignac A, Tomasz A.** 2009. Reconstruction of the phenotypes of methicillin-resistant *Staphylococcus aureus* by replacement of the staphylococcal cassette chromosome *mec* with a plasmid-borne copy of Staphylococcus sciuri *pbpD* gene. Antimicrob Agents Chemother **53:**435-441.
177. **Couto I, Wu SW, Tomasz A, de Lencastre H.** 2003. Development of methicillin resistance in clinical isolates of *Staphylococcus sciuri* by transcriptional activation of the *mecA* homologue native to the species. J Bacteriol **185:**645-653.
178. **Schnellmann C, Gerber V, Rossano A, Jaquier V, Panchaud Y, Doherr MG, Thomann A, Straub R, Perreten V.** 2006. Presence of new *mecA* and *mph(C)* variants conferring antibiotic resistance in *Staphylococcu*s spp. isolated from the skin of horses before and after clinic admission. J Clin Microbiol **44:**4444-4454.
179. **Vanderhaeghen W, Vandendriessche S, Crombe F, Nemeghaire S, Dispas M, Denis O, Hermans K, Haesebrouck F, Butaye P.** 2012. Characterization of methicillin-resistant non-*Staphylococcus aureus* staphylococci carriage isolates from different bovine populations. J Antimicrob Chemother.
180. **Zhang Y, Agidi S, LeJeune JT.** 2009. Diversity of staphylococcal cassette chromosome in coagulase-negative staphylococci from animal sources. J Appl Microbiol **107:**1375-1383.
181. **Vanderhaeghen W, Vandendriessche S, Crombé F, Nemeghaire S, Dispas M, Denis O, Hermans K, Haesebrouck F, Butaye P.** 2013. Characterization of methicillin-resistant non-*Staphylococcus aureus* staphylococci carriage isolates from different bovine populations. J Antimicrob Chemother **68:**300-307.
182. **Kloos WE, Ballard DN, George CG, Webster JA, Hubner RJ, Ludwig W, Schleifer KH, Fiedler F, Schubert K.** 1998. Delimiting the genus *Staphylococcus* through description of *Macrococcus caseolyticu*s gen. nov., comb. nov. and *Macrococcus equipercicus* sp. nov., and *Macrococcus bovicus* sp. no. and *Macrococcus carouselicus* sp. nov. Int J Syst Bacteriol **48 Pt 3:**859-877.
183. **Baba T, Kuwahara-Arai K, Uchiyama I, Takeuchi F, Ito T, Hiramatsu K.** 2009. Complete genome sequence of *Macrococcus caseolyticus* strain JCSCS5402, [corrected] reflecting the ancestral genome of the human-pathogenic staphylococci. J Bacteriol **191:**1180-1190.

184. **Paterson GK, Harrison EM, Holmes MA.** 2014. The emergence of *mecC* methicillin-resistant *Staphylococcus aureus*. Trends Microbiol **22:**42-47.
185. **Schaumburg F, Köck R, Mellmann A, Richter L, Hasenberg F, Kriegeskorte A, Friedrich AW, Gatermann S, Peters G, von Eiff C, Becker K, group s.** 2012. Population dynamics among methicillin-resistant *Staphylococcus aureus* isolates in Germany during a 6-year period. J Clin Microbiol **50:**3186-3192.
186. **Petersen A, Stegger M, Heltberg O, Christensen J, Zeuthen A, Knudsen LK, Urth T, Sorum M, Schouls L, Larsen J, Skov R, Larsen AR.** 2013. Epidemiology of methicillin-resistant *Staphylococcus aureus* carrying the novel *mecC* gene in Denmark corroborates a zoonotic reservoir with transmission to humans. Clin Microbiol Infect **19:**E16-22.
187. **Kim C, Milheiriço C, Gardete S, Holmes MA, Holden MT, de Lencastre H, Tomasz A.** 2012. Properties of a novel PBP2A protein homolog from *Staphylococcus aureus* strain LGA251 and its contribution to the β-lactam-resistant phenotype. J Biol Chem **287:**36854-36863.
188. **Loncaric I, Kübber-Heiss A, Posautz A, Stalder GL, Hoffmann D, Rosengarten R, Walzer C.** 2013. Characterization of methicillin-resistant *Staphylococcus* spp. carrying the *mecC* gene, isolated from wildlife. J Antimicrob Chemother **68:**2222-2225.
189. **Wu SW, de Lencastre H, Tomasz A.** 2001. Recruitment of the *mecA* gene homologue of *Staphylococcus sciuri* into a resistance determinant and expression of the resistant phenotype in *Staphylococcus aureus*. J Bacteriol **183:**2417-2424.
190. **Boundy S, Safo MK, Wang L, Musayev FN, O'Farrell HC, Rife JP, Archer GL.** Characterization of the *Staphylococcus aureus rRNA* methyltransferase encoded by *orfX*, the gene containing the staphylococcal chromosome Cassette *mec* (SCC*mec*) insertion site. J Biol Chem **288:**132-140.
191. **Juuti K, Ibrahem S, Virolainen-Julkunen A, Vuopio-Varkila J, Kuusela P.** 2005. The *pls* gene found in methicillin-resistant *Staphylococcus aureus* strains is common in clinical isolates of *Staphylococcus sciuri*. J Clin Microbiol **43:**1415-1419.
192. **Wu S, de Lencastre H, Tomasz A.** 1998. Genetic organization of the *mecA* region in methicillin-susceptible and methicillin-resistant strains of *Staphylococcus sciuri*. J Bacteriol **180:**236-242.
193. **Ito T, Katayama Y, Hiramatsu K.** 1999. Cloning and nucleotide sequence determination of the entire *mec* DNA of pre-methicillin-resistant *Staphylococcus aureus* N315. Antimicrob Agents Chemother **43:**1449-1458.
194. **CLSI.** 2009. Performance standards for antimicrobial disk susceptibility tests; approved standard, 10th edn. Wyane, PA: CLSI publication M02-A10. Clinical and Laboratory Standards Institute, 2009.
195. **Chung M, de Lencastre H, Matthews P, Tomasz A, Adamsson I, Aires de Sousa M, Camou T, Cocuzza C, Corso A, Couto I, Dominguez A, Gniadkowski M, Goering R, Gomes A, Kikuchi K, Marchese A, Mato R, Melter O, Oliveira D, Palacio R, Sa-Leao R, Santos Sanches I, Song JH, Tassios PT, Villari P.** 2000. Molecular typing of methicillin-resistant *Staphylococcus aureus* by pulsed-field gel electrophoresis: comparison of results obtained in a multilaboratory effort using identical protocols and MRSA strains. Microb Drug Resist **6:**189-198.
196. **Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S.** MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol **28:**2731-2739.
197. **Mekni MA, Bouchami O, Achour W, Ben Hassen A.** Strong biofilm production but not adhesion virulence factors can discriminate between invasive and commensal *Staphylococcus epidermidis* strains. APMIS **120:**605-611.

198. **Rolo J, de Lencastre H, Miragaia M.** 2012. Strategies of adaptation of *Staphylococcus epidermidis* to hospital and community: amplification and diversification of SCC*mec*. J Antimicrob Chemother **67:**1333-1341.
199. **Rolo J, Miragaia M, Turlej-Rogacka A, Empel J, Bouchami O, Faria NA, Tavares A, Hryniewicz W, Fluit AC, de Lencastre H.** High genetic diversity among community-associated *Staphylococcus aureus* in Europe: results from a multicenter study. PLoS One **7:**e34768.
200. **Dyke KG, Jevons MP, Parker MT.** 1966. Penicillinase production and intrinsic resistance to penicillins in Staphylococcus aures. Lancet **1:**835-838.
201. **Chung M, de Lencastre H, Matthews P, Tomasz A, Adamsson I, Aires de Sousa M, Camou T, Cocuzza C, Corso A, Couto I, Dominguez A, Gniadkowski M, Goering R, Gomes A, Kikuchi K, Marchese A, Mato R, Melter O, Oliveira D, Palacio R, Sa-Leao R, Santos Sanches I, Song JH, Tassios PT, Villari P.** 2000. Molecular typing of methicillin-resistant *Staphylococcus aureus* by pulsed-field gel electrophoresis: comparison of results obtained in a multilaboratory effort using identical protocols and MRSA strains. Microb Drug Resist **6:**189-198.
202. **Couto I, Sanches IS, Sá-Leão R, de Lencastre H.** 2000. Molecular characterization of *Staphylococcus sciuri* strains isolated from humans. J Clin Microbiol **38:**1136-1143.
203. **de Lencastre H, Oliveira D, Tomasz A.** 2007. Antibiotic resistant *Staphylococcus aureus*: a paradigm of adaptive power. Curr Opin Microbiol **10:**428-435.
204. **Hiramatsu K, Ito T, Tsubakishita S, Sasaki T, Takeuchi F, Morimoto Y, Katayama Y, Matsuo M, Kuwahara-Arai K, Hishinuma T, Baba T.** 2013. Genomic Basis for Methicillin Resistance in *Staphylococcus aureus*. Infect Chemother **45:**117-136.
205. **Fuda C, Suvorov M, Vakulenko SB, Mobashery S.** 2004. The basis for resistance to beta-lactam antibiotics by penicillin-binding protein 2a of methicillin-resistant *Staphylococcus aureus*. J Biol Chem **279:**40802-40806.
206. **Jevons MP, Coe AW, Parker MT.** 1963. Methicillin resistance in staphylococci. Lancet **1:**904-907.
207. **de Lencastre H, Tomasz A.** 1994. Reassessment of the number of auxiliary genes essential for expression of high-level methicillin resistance in *Staphylococcus aureus*. Antimicrob Agents Chemother **38:**2590-2598.
208. **Mwangi MM, Kim C, Chung M, Tsai J, Vijayadamodar G, Benitez M, Jarvie TP, Du L, Tomasz A.** 2013. Whole-genome sequencing reveals a link between β-lactam resistance and synthetases of the alarmone (p)ppGpp in *Staphylococcus aureus*. Microb Drug Resist **19:**153-159.
209. **Kim C, Mwangi M, Chung M, Milheiriço C, Milheirço C, de Lencastre H, Tomasz A.** 2013. The mechanism of heterogeneous beta-lactam resistance in MRSA: key role of the stringent stress response. PLoS One **8:**e82814.
210. **Dordel J, Kim C, Chung M, Pardos de la Gándara M, Holden MT, Parkhill J, de Lencastre H, Bentley SD, Tomasz A.** 2014. Novel determinants of antibiotic resistance: identification of mutated loci in highly methicillin-resistant subpopulations of methicillin-resistant *Staphylococcus aureus*. MBio **5:**e01000.
211. **de Lencastre H, Sa Figueiredo AM, Urban C, Rahal J, Tomasz A.** 1991. Multiple mechanisms of methicillin resistance and improved methods for detection in clinical isolates of *Staphylococcus aureus*. Antimicrob Agents Chemother **35:**632-639.
212. **Gray RR, Tatem AJ, Johnson JA, Alekseyenko AV, Pybus OG, Suchard MA, Salemi M.** 2011. Testing spatiotemporal hypothesis of bacterial evolution using methicillin-resistant *Staphylococcus aureus* ST239 genome-wide data within a bayesian framework. Mol Biol Evol **28:**1593-1603.
213. **Martin D, Rybicki E.** 2000. RDP: detection of recombination amongst aligned sequences. Bioinformatics **16:**562-563.

214. **Martí-Renom MA, Stuart AC, Fiser A, Sánchez R, Melo F, Sali A.** 2000. Comparative protein structure modeling of genes and genomes. Annu Rev Biophys Biomol Struct **29:**291-325.
215. **Simpson EH.** 1949. Measurement of species diversity. Nature **163:**688.
216. **de Lencastre H, de Jonge BL, Matthews PR, Tomasz A.** 1994. Molecular aspects of methicillin resistance in *Staphylococcus aureus*. J Antimicrob Chemother **33:**7-24.
217. **Hartman BJ, Tomasz A.** 1984. Low-affinity penicillin-binding protein associated with beta-lactam resistance in *Staphylococcus aureus*. J Bacteriol **158:**513-516.
218. **CDC.** 2013. Antibiotic Resistance Threats in the United States, 2013. U.S.Centers for Disease Control and Prevention.
219. **Rolo J, de Lencastre H, Miragaia M.** 2014. High frequency and diversity of cassette chromosome recombinases (*ccr*) in methicillin-susceptible *Staphylococcus sciuri*. J Antimicrob Chemother **69 (6)**:1461-9.
220. **Poyart C, Quesne G, Boumaila C, Trieu-Cuot P.** 2001. Rapid and accurate species-level identification of coagulase-negative staphylococci by using the *sodA* gene as a target. J Clin Microbiol **39:**4296-4301.
221. **Frey Y, Rodriguez JP, Thomann A, Schwendener S, Perreten V.** 2013. Genetic characterization of antimicrobial resistance in coagulase-negative staphylococci from bovine mastitis milk. J Dairy Sci **96:**2247-2257.
222. **Martineau F, Picard FJ, Ke D, Paradis S, Roy PH, Ouellette M, Bergeron MG.** 2001. Development of a PCR assay for identification of staphylococci at genus and species levels. J Clin Microbiol **39:**2541-2547.
223. **Song MD, Wachi M, Doi M, Ishino F, Matsuhashi M.** 1987. Evolution of an inducible penicillin-target protein in methicillin-resistant *Staphylococcus aureus* by gene fusion. FEBS Lett **221:**167-171.
224. **Sabat AJ, Pournaras S, Akkerboom V, Tsakris A, Grundmann H, Friedrich AW.** 2015. Whole-genome analysis of an oxacillin-susceptible CC80 *mecA*-positive *Staphylococcus aureus* clinical isolate: insights into the mechanisms of cryptic methicillin resistance. J Antimicrob Chemother.
225. **Cohen S, Gibson CJ, Sweeney HM.** 1972. Phenotypic suppression of methicillin resistance in *Staphylococcus aureus* by mutant noninducible penicillinase plasmids. J Bacteriol **112:**682-689.
226. **Hackbarth CJ, Chambers HF.** 1993. *blaI* and *blaR1* regulate beta-lactamase and PBP 2a production in methicillin-resistant *Staphylococcus aureus.* Antimicrob Agents Chemother **37:**1144-1149.
227. **Smith IM, Beals PD, Kingsbury KR, Hasenclever HF.** 1958. Observations on *Staphylococcus albus* septicemia in mice and men. AMA Arch Intern Med **102:**375-388.
228. **Argudín MA, Vanderhaeghen W, Vandendriessche S, Vandecandelaere I, André FX, Denis O, Coenye T, Butaye P.** 2015. Antimicrobial resistance and population structure of *Staphylococcus epidermidis* recovered from animals and humans. Vet Microbiol **178:**105-113.
229. **Angiuoli SV, Salzberg SL.** 2011. Mugsy: fast multiple alignment of closely related whole genomes. Bioinformatics **27:**334-342.
230. **Feil EJ, Li BC, Aanensen DM, Hanage WP, Spratt BG.** 2004. eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. J Bacteriol **186:**1518-1530.
231. **Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, Aarestrup FM, Larsen MV.** 2012. Identification of acquired antimicrobial resistance genes. J Antimicrob Chemother **67:**2640-2644.
232. **Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M.** 2006. ISfinder: the reference centre for bacterial insertion sequences. Nucleic Acids Res **34:**D32-36.

233. **Carattoli A, Zankari E, García-Fernández A, Voldby Larsen M, Lund O, Villa L, Møller Aarestrup F, Hasman H.** 2014. In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. Antimicrob Agents Chemother **58:**3895-3903.
234. **Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS.** 2011. PHAST: a fast phage search tool. Nucleic Acids Res **39:**W347-352.
235. **Conlon KM, Humphreys H, O'Gara JP.** 2004. Inactivations of *rsbU* and *sarA* by IS*256* represent novel mechanisms of biofilm phenotypic variation in *Staphylococcus epidermidis*. J Bacteriol **186:**6208-6219.
236. **Barbier F, Ruppe E, Hernandez D, Lebeaux D, Francois P, Felix B, Desprez A, Maiga A, Woerther PL, Gaillard K, Jeanrot C, Wolff M, Schrenzel J, Andremont A, Ruimy R.** 2010. Methicillin-resistant coagulase-negative staphylococci in the community: high homology of SCC*mec* IVa between *Staphylococcus epidermidis* and major clones of methicillin-resistant *Staphylococcus aureus*. J Infect Dis **202:**270-281.
237. **Ruppe E, Barbier F, Mesli Y, Maiga A, Cojocaru R, Benkhalfat M, Benchouk S, Hassaine H, Maiga I, Diallo A, Koumare AK, Ouattara K, Soumare S, Dufourcq JB, Nareth C, Sarthou JL, Andremont A, Ruimy R.** 2009. Diversity of staphylococcal cassette chromosome *mec* structures in methicillin-resistant *Staphylococcus epidermidis* and *Staphylococcus haemolyticus* strains among outpatients from four countries. Antimicrob Agents Chemother **53:**442-449.
238. **Matos RC, Lapaque N, Rigottier-Gois L, Debarbieux L, Meylheuc T, Gonzalez-Zorn B, Repoila F, Lopes MeF, Serror P.** 2013. *Enterococcus faecalis* prophage dynamics and contributions to pathogenic traits. PLoS Genet **9:**e1003539.
239. **Gutiérrez D, Martínez B, Rodríguez A, García P.** 2012. Genomic characterization of two *Staphylococcus epidermidis* bacteriophages with anti-biofilm potential. BMC Genomics **13:**228.
240. **Gutiérrez D, Ruas-Madiedo P, Martínez B, Rodríguez A, García P.** 2014. Effective removal of staphylococcal biofilms by the endolysin LysH5. PLoS One **9:**e107307.
241. **Gutiérrez D, Vandenheuvel D, Martínez B, Rodríguez A, Lavigne R, García P.** 2015. Two Phages, phiIPLA-RODI and phiIPLA-C1C, Lyse Mono- and Dual-Species Staphylococcal Biofilms. Appl Environ Microbiol **81:**3336-3348.
242. **Weiß S, Kadlec K, Feßler AT, Schwarz S.** 2014. Complete sequence of a multiresistance plasmid from a methicillin-resistant *Staphylococcus epidermidis* ST5 isolated in a small animal clinic. J Antimicrob Chemother **69:**847-859.
243. **Parkhill J, Sebaihia M, Preston A, Murphy LD, Thomson N, Harris DE, Holden MT, Churcher CM, Bentley SD, Mungall KL, Cerdeño-Tárraga AM, Temple L, James K, Harris B, Quail MA, Achtman M, Atkin R, Baker S, Basham D, Bason N, Cherevach I, Chillingworth T, Collins M, Cronin A, Davis P, Doggett J, Feltwell T, Goble A, Hamlin N, Hauser H, Holroyd S, Jagels K, Leather S, Moule S, Norberczak H, O'Neil S, Ormond D, Price C, Rabbinowitsch E, Rutter S, Sanders M, Saunders D, Seeger K, Sharp S, Simmonds M, Skelton J, Squares R, Squares S, Stevens K, Unwin L, et al.** 2003. Comparative analysis of the genome sequences of *Bordetella pertussis, Bordetella parapertussis* and *Bordetella bronchiseptica*. Nat Genet **35:**32-40.
244. **Novick RP, Roth C.** 1968. Plasmid-linked resistance to inorganic salts in *Staphylococcus aureus*. J Bacteriol **95:**1335-1342.
245. **Slifierz MJ, Friendship RM, Weese JS.** 2015. Methicillin-resistant *Staphylococcus aureus* in commercial swine herds is associated with disinfectant and zinc usage. Appl Environ Microbiol **81:**2690-2695.
246. **Rosendal K, Bülow P, Bentzon MW, Eriksen KR.** 1976. *Staphylococcus aureus* strains isolated in Danish hospitals from January 1st, 1966, to December 31st, 1974. Acta Pathol Microbiol Scand B **84B:**359-368.

247. **Eriksen KR, Dons N, Erichsen I.** 1960. Erythromycin-resistant Staphylococci. Incidence and mode of development. Ugeskr Laeger **122:**1635-1642.
248. **Harvey CL, Knight SG, Sih CJ.** 1966. On the mode of action of fusidic acid. Biochemistry **5:**3320-3327.
249. **Crisóstomo MI, Westh H, Tomasz A, Chung M, Oliveira DC, de Lencastre H.** 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. Proc Natl Acad Sci U S A **98:**9865-9870.
250. **Llaneza J, Villar CJ, Salas JA, Suarez JE, Mendoza MC, Hardisson C.** 1985. Plasmid-mediated fosfomycin resistance is due to enzymatic modification of the antibiotic. Antimicrob Agents Chemother **28:**163-164.
251. **Raz R.** 2012. Fosfomycin: an old--new antibiotic. Clin Microbiol Infect **18:**4-7.
252. **Kaatz GW, Seo SM.** 1995. Inducible NorA-mediated multidrug resistance in *Staphylococcus aureus*. Antimicrob Agents Chemother **39:**2650-2655.
253. **Blumberg HM, Rimland D, Carroll DJ, Terry P, Wachsmuth IK.** 1991. Rapid development of ciprofloxacin resistance in methicillin-susceptible and -resistant *Staphylococcus aureus*. J Infect Dis **163:**1279-1285.
254. **Shearer JE, Wireman J, Hostetler J, Forberger H, Borman J, Gill J, Sanchez S, Mankin A, Lamarre J, Lindsay JA, Bayles K, Nicholson A, O'Brien F, Jensen SO, Firth N, Skurray RA, Summers AO.** 2011. Major families of multiresistant plasmids from geographically and epidemiologically diverse staphylococci. G3 (Bethesda) **1:**581-591.
255. **Katayama Y, Zhang HZ, Hong D, Chambers HF.** 2003. Jumping the barrier to beta-lactam resistance in *Staphylococcus aureus*. J Bacteriol **185:**5465-5472.
256. **Diekema DJ, Pfaller MA, Schmitz FJ, Smayevsky J, Bell J, Jones RN, Beach M.** 2001. Survey of infections due to *Staphylococcus* species: frequency of occurrence and antimicrobial susceptibility of isolates collected in the United States, Canada, Latin America, Europe, and the Western Pacific region for the SENTRY Antimicrobial Surveillance Program, 1997-1999. Clin Infect Dis **32 Suppl 2:**S114-132.
257. **Li S, Skov RL, Han X, Larsen AR, Larsen J, Sorum M, Wulf M, Voss A, Hiramatsu K, Ito T.** Novel Types of Staphylococcal Cassette Chromosome *mec* Elements Identified in Clonal Complex 398 Methicillin-Resistant *Staphylococcus aureus* Strains. Antimicrob Agents Chemother **55:**3046-3050.
258. **Hanssen AM, Sollid JU.** 2007. Multiple staphylococcal cassette chromosomes and allelic variants of cassette chromosome recombinases in *Staphylococcus aureus* and coagulase-negative staphylococci from Norway. Antimicrob Agents Chemother **51:**1671-1677.
259. **Kampf G, Lecke C, Cimbal AK, Weist K, Ruden H.** 1998. Evaluation of mannitol salt agar for detection of oxacillin resistance in *Staphylococcus aureus* by disk diffusion and agar screening. J Clin Microbiol **36:**2254-2257.
260. **Couto I, Pereira S, Miragaia M, Sanches IS, de Lencastre H.** 2001. Identification of clinical staphylococcal isolates from humans by internal transcribed spacer PCR. J Clin Microbiol **39:**3099-3103.
261. **Oliveira DC, Tomasz A, de Lencastre H.** 2001. The evolution of pandemic clones of methicillin-resistant *Staphylococcus aureus*: identification of two ancestral genetic backgrounds and the associated mec elements. Microb Drug Resist **7:**349-361.
262. **Weisser M, Schoenfelder SM, Orasch C, Arber C, Gratwohl A, Frei R, Eckart M, Fluckiger U, Ziebuhr W.** 2010. Hypervariability of biofilm formation and oxacillin resistance in a *Staphylococcus epidermidis* strain causing persistent severe infection in an immunocompromised patient. J Clin Microbiol **48:**2407-2412.
263. **Higgins PG, Rosato AE, Seifert H, Archer GL, Wisplinghoff H.** 2009. Differential expression of *ccrA* in methicillin-resistant *Staphylococcus aureus* strains carrying

staphylococcal cassette chromosome *mec* type II and IVa elements. Antimicrob Agents Chemother **53:**4556-4558.

264. **Sorensen SJ, Bailey M, Hansen LH, Kroer N, Wuertz S.** 2005. Studying plasmid horizontal transfer in situ: a critical review. Nat Rev Microbiol **3:**700-710.
265. **O'Brien FG, Pearman JW, Gracey M, Riley TV, Grubb WB.** 1999. Community strain of methicillin-resistant *Staphylococcus aureus i*nvolved in a hospital outbreak. J Clin Microbiol **37:**2858-2862.
266. **Seybold U, Kourbatova EV, Johnson JG, Halvosa SJ, Wang YF, King MD, Ray SM, Blumberg HM.** 2006. Emergence of community-associated methicillin-resistant *Staphylococcus aureus* USA300 genotype as a major cause of health care-associated blood stream infections. Clin Infect Dis **42:**647-656.
267. **O'Brien FG, Lim TT, Chong FN, Coombs GW, Enright MC, Robinson DA, Monk A, Said-Salim B, Kreiswirth BN, Grubb WB.** 2004. Diversity among community isolates of methicillin-resistant *Staphylococcus aureus* in Australia. J Clin Microbiol **42:**3185-3190.
268. **Widerstrom M, Monsen T, Karlsson C, Edebro H, Johansson A, Wistrom J.** 2009. Clonality among multidrug-resistant hospital-associated *Staphylococcus epidermidis* in northern Europe. Scand J Infect Dis **41:**642-649.
269. **de Almeida LM, Lincopan N, de Araujo MR, Mamizuka EM.** Dissemination of the linezolid-resistant *Staphylococcus epidermidis* clone ST2 exhibiting the G2576T mutation in the 23S *rRNA* gene in a tertiary-care hospital, Brazil. J Antimicrob Chemother.
270. **Li M, Wang X, Gao Q, Lu Y.** 2009. Molecular characterization of *Staphylococcus epidermidis* strains isolated from a teaching hospital in Shanghai, China. J Med Microbiol **58:**456-461.
271. **Zimmerman DR.** 1986. Role of subtherapeutical levels of antimicrobials in pig production. Journal of Animal Science **62:**6-16.
272. **Malachowa N, DeLeo FR.** Mobile genetic elements of *Staphylococcus aureus*. Cell Mol Life Sci **67:**3057-3071.
273. **Kaito C, Saito Y, Nagano G, Ikuo M, Omae Y, Hanada Y, Han X, Kuwahara-Arai K, Hishinuma T, Baba T, Ito T, Hiramatsu K, Sekimizu K.** 2011. Transcription and translation products of the cytolysin gene *psm-mec* on the mobile genetic element SCC*mec* regulate *Staphylococcus aureus* virulence. PLoS Pathog **7:**e1001267.
274. **Christner M, Heinze C, Busch M, Franke G, Hentschke M, Bayard Dühring S, Büttner H, Kotasinska M, Wischnewski V, Kroll G, Buck F, Molin S, Otto M, Rohde H.** 2012. *sarA* negatively regulates *Staphylococcus epidermidis* biofilm formation by modulating expression of 1 MDa extracellular matrix binding protein and autolysis-dependent release of eDNA. Mol Microbiol **86:**394-410.
275. **McDougal LK, Thornsberry C.** 1986. The role of beta-lactamase in staphylococcal resistance to penicillinase-resistant penicillins and cephalosporins. J Clin Microbiol **23:**832-839.