SCCmecFinder, a Web-Based Tool for Typing of Staphylococcal Cassette Chromosome mec in Staphylococcus aureus Using Whole-Genome Sequence Data

Kaya, Hülya; Hasman, Henrik; Larsen, Jesper; Stegger, Marc; Johannesen, Thor Bech; Allesøe, Rosa Lundbye; Lemvigh, Camilla Koldbæk; Aarestrup, Frank Møller; Lund, Ole; Larsen, Anders Rhod

Published in:
mSphere

Link to article, DOI:
10.1128/mSphere.00612-17

Publication date:
2018

Document Version
Publisher's PDF, also known as Version of record

Link back to DTU Orbit

Citation (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
SCCmecFinder, a Web-Based Tool for Typing of Staphylococcal Cassette Chromosome mec in Staphylococcus aureus Using Whole-Genome Sequence Data

Hülya Kaya, Henrik Hasman, Jesper Larsen, Marc Stegger, Thor Bech Johannesen, Rosa Lundbye Allesøe, Camilla Koldbaek Lemvigh, Frank Møller Aarestrup, Ole Lund, Anders Rhod Larsen

Department of Bacteria, Parasites and Fungi, Statens Serum Institut, Copenhagen, Denmark
Research Group for Genomic Epidemiology, National Food Institute, Technical University of Denmark, Kgs Lyngby, Denmark
Center for Biological Sequence Analysis, Department of Bioinformatics, Technical University of Denmark, Kgs Lyngby, Denmark

ABSTRACT Typing of methicillin-resistant Staphylococcus aureus (MRSA) is important in infection control and surveillance. The current nomenclature of MRSA includes the genetic background of the S. aureus strain determined by multilocus sequence typing (MLST) or equivalent methods like spa typing and typing of the mobile genetic element staphylococcal cassette chromosome mec (SCCmec), which carries the mecA or mecC gene. Whereas MLST and spa typing are relatively simple, typing of SCCmec is less trivial because of its heterogeneity. Whole-genome sequencing (WGS) provides the essential data for typing of the genetic background and SCCmec, but so far, no bioinformatic tools for SCCmec typing have been available. Here, we report the development and evaluation of SCCmecFinder for characterization of the SCCmec element from S. aureus WGS data. SCCmecFinder is able to identify all SCCmec element types, designated I to XIII, with subtyping of SCCmec types IV (2B) and V (5C2). SCCmec elements are characterized by two different gene prediction approaches to achieve correct annotation, a Basic Local Alignment Search Tool (BLAST)-based approach and a k-mer-based approach. Evaluation of SCCmecFinder by using a diverse collection of clinical isolates (n = 93) showed a high typeability level of 96.7%, which increased to 98.9% upon modification of the default settings. In conclusion, SCCmecFinder can be an alternative to more laborious SCCmec typing methods and is freely available at https://cge.cbs.dtu.dk/services/SCCmecFinder.

IMPORTANCE SCCmec in MRSA is acknowledged to be of importance not only because it contains the mecA or mecC gene but also for staphylococcal adaptation to different environments, e.g., in hospitals, the community, and livestock. Typing of SCCmec by PCR techniques has, because of its heterogeneity, been challenging, and whole-genome sequencing has only partially solved this since no good bioinformatic tools have been available. In this article, we describe the development of a new bioinformatic tool, SCCmecFinder, that includes most of the needs for infection control professionals and researchers regarding the interpretation of SCCmec elements. The software detects all of the SCCmec elements accepted by the International Working Group on the Classification of Staphylococcal Cassette Chromosome Elements, and users will be prompted if diverging and potential new elements are uploaded. Furthermore, SCCmecFinder will be curated and updated as new elements are found and it is easy to use and freely accessible.

KEYWORDS MRSA, SCCmec, bioinformatics, mecA, mecC, typing
The presence of methicillin-resistant Staphylococcus aureus (MRSA) imposes a significant burden on the public health care system, where accurate molecular typing is essential for infection control and surveillance of MRSA. The current standard MRSA nomenclature includes identification of the chromosomal background, annotated by the multilocus sequence type (ST) or clonal complex (CC) and the type of staphylococcal cassette chromosome mec (SCCmec) element (indicated by roman numerals I to XIII), sometimes with the addition of a letter indicating the SCCmec subtype (e.g., ST80-IVc) (1).

The key components of SCCmec are the mec and ccr gene complexes, which contain genes responsible for methicillin resistance and the mobility of the SCCmec element, respectively (2). These two gene complexes are surrounded by three highly variable joining regions (J1 to J3), which contain nonessential components of the SCCmec element but may harbor additional resistance genes (3). The currently known SCCmec elements of S. aureus all integrate at the same unique site in an open reading frame called orfX, and all have the same backbone structure with the general organization orfX-J3-mec-J2-ccr-J1, with the exception of SCCmec types VIII (4A) and IX (1C2), which have a similar organization, namely, orfX-J3-ccr-J2-mec-J1, as shown in Fig. 1, which was previously published and kindly allowed to be used here by Hiramatsu et al. (4). The combination of the mec and ccr gene complexes determines the SCCmec type, while the J1 region is used for subtyping (2). Historically, subtyping of the SCCmec element is only conducted for SCCmec types IV (2B) and V (5C2 and 5C2&5).

Genetic rearrangement of the SCCmec element can, however, result in novel elements, variants of existing SCCmec elements, and composite elements, hence complicating the nomenclature of SCCmec elements.

In 2009, the International Working Group on the Classification of Staphylococcal Cassette Chromosome Elements (IWG-SCC) established a consensus uniform nomenclature for the SCCmec element that defined the requirements for the annotation of new SCCmec elements and provided a homepage for keeping track of approved SCCmec elements and information about them. Currently, SCCmec types I to XI are listed on the homepage (http://www.sccmec.org), while SCCmec types XII (9C2) (5) and XIII (9A) have been reported and approved by the IWG-SCC but yet not listed on the homepage.

The SCCmec typing schemes are based on several multiplex PCR (M-PCR) protocols that do not encompass the plasticity of SCCmec elements, leading to an increasing proportion of nontypeable (NT) SCCmec elements (6–8). Recent advances in whole-genome sequencing (WGS) have made data essential for SCCmec typing available, but no bioinformatic tools for in silico typing of the SCCmec element have been available. Here, we introduce SCCmecFinder, an in silico web-based bioinformatic tool for the identification and typing of SCCmec elements listed by the IWG-SCC from WGS data on S. aureus isolates. Together with other existing tools for multilocus sequence typing (MLST) and spa typing, SCCmecFinder enables high-throughput WGS-based typing of MRSA isolates. The tool is freely available at the Center for Genomic Epidemiology (CGE) homepage, http://genomicepidemiology.org/.

RESULTS AND DISCUSSION

SCCmec typing by SCCmecFinder. The initial evaluation with the type I to XI reference SCCmec elements, the composite elements SCCmec types IV (2B&5) and V (5C2&5), subtypes of SCCmec types IV (2B) and V (5C2), and the two new elements SCCmec types XII (9C2) and XIII (9A) revealed concordant results for all 36 SCCmec elements. When both SCCmec typing methods (SCCmecFinder and the M-PCR-based typing) were applied to the 93 clinical isolates (Table 1), an initial concordance of 96.7% was observed (Table 1). SCCmecFinder correctly differentiated methicillin-susceptible S. aureus (MSSA) isolates from MRSA isolates in all cases. Of the 50 MSSA isolates, 3 were predicted to harbor an SCC-like element having only the type 1 ccr gene complex by SCCmecFinder. M-PCR-based typing confirmed the presence of a type 1 ccr gene complex and the absence of the mec gene and gene complex. All three isolates...
belonged to ST1/t127, which has previously been associated with the SCC-like element SCC\textit{fus}, which contains the fusidic acid resistance gene \textit{fusC} (9). The presence of the \textit{fusC} gene in the three isolates was confirmed, indicating that the isolates harbor the SCC\textit{fus} element. SCC\textit{mec}Finder correctly predicted the SCC\textit{mec} elements in 39 of the 43

\textbf{FIG 1} Backbone structure organization of SCC\textit{mec} elements in \textit{S. aureus}. This figure was published previously (4) and is shown here with the kind permission of Keiichi Hiramatsu.
MRSA isolates, with an average \((k\)-mer\) template coverage 83.83\%, ranging from 44.12 to 99.79\% for the 39 SCC
\emph{mec} elements.

SCC\emph{mec}Finder and the M-PCR-based typing gave NT results for four isolates but without overlaps. Of the four isolates NT by SCC\emph{mec}Finder, two could be recovered by a reassembly of the raw reads using SPAdes instead of the default assembler, e.g., Velvet, and one SCC\emph{mec} element was recovered by modifying the database to the extended database, while the last isolate could not be recovered (Table 2). Upon modification of the thresholds, the percentage of concordance increased to 98.9\%.

Publically available \emph{S. aureus} genomes from the NCBI RefSeq database \((n = 6,852)\) were included to evaluate the performance of SCC\emph{mec}Finder on genomes with different sequencing qualities from different platforms and sources to catch possible systematic errors by SCC\emph{mec}Finder and identify the most prevalent warning messages. SCC\emph{mec}Finder identified 6,021 genomes as MRSA and 831 as MSSA (Table 3). Less than 8\% of the predictions were accompanied by a warning. The predominant warning was

### TABLE 1

<table>
<thead>
<tr>
<th>Strain</th>
<th>ST/spa type</th>
<th>SCC\emph{mec} typing result</th>
</tr>
</thead>
<tbody>
<tr>
<td>75013</td>
<td>ST30/t019</td>
<td>NT</td>
</tr>
<tr>
<td>75015</td>
<td>ST22/t032</td>
<td>IV (2B)</td>
</tr>
<tr>
<td>75019</td>
<td>ST5/t002</td>
<td>IV (2B)</td>
</tr>
<tr>
<td>75020</td>
<td>ST22/t014</td>
<td>NT</td>
</tr>
<tr>
<td>75024</td>
<td>ST152/t355</td>
<td>V (5C2)</td>
</tr>
<tr>
<td>75049</td>
<td>ST398/t034</td>
<td>NT</td>
</tr>
<tr>
<td>75050</td>
<td>ST30/t019</td>
<td>IV (2B)</td>
</tr>
<tr>
<td>75052</td>
<td>ST8/t1216</td>
<td>IV (2B)</td>
</tr>
<tr>
<td>75054</td>
<td>ST398/t034</td>
<td>V (5C2)</td>
</tr>
<tr>
<td>75055</td>
<td>ST30/t1752</td>
<td>IV (2B)</td>
</tr>
<tr>
<td>75056</td>
<td>ST8/t008</td>
<td>IV (2B)</td>
</tr>
<tr>
<td>75058</td>
<td>ST6/t4403</td>
<td>NT</td>
</tr>
<tr>
<td>75061</td>
<td>ST30/t019</td>
<td>IV (2B)</td>
</tr>
<tr>
<td>75083</td>
<td>ST22/t3638</td>
<td>IV (2B)</td>
</tr>
<tr>
<td>75098</td>
<td>ST5/t1062</td>
<td>IV (2B)</td>
</tr>
<tr>
<td>75100</td>
<td>ST8/t008</td>
<td>IV (2B)</td>
</tr>
<tr>
<td>75101</td>
<td>ST8/t2171</td>
<td>IV (2B)</td>
</tr>
<tr>
<td>75103</td>
<td>ST30/t012</td>
<td>IV (2B)</td>
</tr>
<tr>
<td>75104</td>
<td>ST5/t5550</td>
<td>I (1B)</td>
</tr>
<tr>
<td>75108</td>
<td>ST45/t026</td>
<td>IV (2B)</td>
</tr>
<tr>
<td>75125</td>
<td>ST5/t002</td>
<td>IV (2B)</td>
</tr>
<tr>
<td>75126</td>
<td>ST45/t362</td>
<td>IV (2B)</td>
</tr>
<tr>
<td>75127</td>
<td>ST45/t362</td>
<td>IV (2B)</td>
</tr>
<tr>
<td>75129</td>
<td>ST22/t022</td>
<td>IV (2B)</td>
</tr>
<tr>
<td>75146</td>
<td>ST398/t034</td>
<td>V (5C2)</td>
</tr>
<tr>
<td>75147</td>
<td>ST1/t127</td>
<td>IV (2B)</td>
</tr>
<tr>
<td>75149</td>
<td>ST22/t223</td>
<td>IV (2B)</td>
</tr>
<tr>
<td>75150</td>
<td>ST45/t026</td>
<td>IV (2B)</td>
</tr>
<tr>
<td>75178</td>
<td>ST772/t657</td>
<td>V (5C2)</td>
</tr>
<tr>
<td>75180</td>
<td>ST5/t002</td>
<td>IV (2B)</td>
</tr>
<tr>
<td>75181</td>
<td>ST30/t019</td>
<td>IV (2B)</td>
</tr>
<tr>
<td>75182</td>
<td>ST5/t002</td>
<td>IV (2B)</td>
</tr>
<tr>
<td>75201</td>
<td>ST30/unknown</td>
<td>IV (2B)</td>
</tr>
<tr>
<td>75203</td>
<td>ST22/t032</td>
<td>IV (2B)</td>
</tr>
<tr>
<td>75205</td>
<td>ST5/t045</td>
<td>II (2A)</td>
</tr>
<tr>
<td>75208</td>
<td>ST30/t019</td>
<td>IV (2B)</td>
</tr>
<tr>
<td>75209</td>
<td>ST8/t690</td>
<td>IV (2B)</td>
</tr>
<tr>
<td>75211</td>
<td>ST22/t379</td>
<td>IV (2B)</td>
</tr>
<tr>
<td>75212</td>
<td>ST5/t010</td>
<td>I (1B)</td>
</tr>
<tr>
<td>75316</td>
<td>ST22/t022</td>
<td>IV (2B&amp;5)</td>
</tr>
<tr>
<td>75319</td>
<td>ST8/t008</td>
<td>IV (2B)</td>
</tr>
<tr>
<td>75336</td>
<td>ST152/t355</td>
<td>V (5C2)</td>
</tr>
<tr>
<td>75339</td>
<td>ST398/t011</td>
<td>V (5C2)</td>
</tr>
</tbody>
</table>

*The PCR-based SCC\emph{mec} typing protocol described by Kondo et al. (6).*
the detection of multiple (more than one) ccr gene complexes and does not count for SCCmec types V (5C2&5) and IV (2B&5), as these are detectable by SCCmecFinder. Detection of multiple gene complexes complicates the Basic Local Alignment Search Tool (BLAST)-based prediction, as the method is unable to resolve which gene complexes make up the SCCmec element and which may be part of an SCC-like element, and hence, in such cases, SCCmecFinder reports the detection of multiple gene complexes to alert the user to make a final decision. Failure of the BLAST-based approach to give a definitive SCCmec prediction was also among the most frequently observed warnings, often because of stringent threshold levels or insufficient assembly

<table>
<thead>
<tr>
<th>Strain</th>
<th>ST/spa type</th>
<th>PCR-based protocola</th>
<th>SCCmecFinder</th>
<th>Recovery method</th>
<th>Final SCCmec type</th>
</tr>
</thead>
<tbody>
<tr>
<td>75024</td>
<td>ST152/t355</td>
<td>V (SC2)</td>
<td>NT</td>
<td>Change from Velvet to SPAdes assembler</td>
<td>V (SC2&amp;5)</td>
</tr>
<tr>
<td>75149</td>
<td>ST22/t223</td>
<td>IV (2B)</td>
<td>NT</td>
<td>Change from Velvet to SPAdes assembler</td>
<td>IV (2B)</td>
</tr>
<tr>
<td>75178</td>
<td>ST772/t657</td>
<td>V (SC2)</td>
<td>V (SC2) or V (SC2&amp;5)</td>
<td>Change from reference to extended database</td>
<td>V (SC2)</td>
</tr>
<tr>
<td>75180</td>
<td>ST5/t002</td>
<td>IV (2B)</td>
<td>IV (2B) or II (2A)</td>
<td>NT</td>
<td></td>
</tr>
</tbody>
</table>

The PCR-based SCCmec typing protocol described by Kondo et al. (6).

<table>
<thead>
<tr>
<th>S. aureus group and SCCmec typing result</th>
<th>No. of genomes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Default thresholdsa</td>
</tr>
<tr>
<td>MRSA</td>
<td>201</td>
</tr>
<tr>
<td>No alerts</td>
<td>2,241</td>
</tr>
<tr>
<td>SCCmec type I (1B)</td>
<td>136</td>
</tr>
<tr>
<td>SCCmec type II (2A)</td>
<td>2,616</td>
</tr>
<tr>
<td>SCCmec type IV (2B)</td>
<td>20</td>
</tr>
<tr>
<td>SCCmec type IV (2B&amp;5)</td>
<td>0</td>
</tr>
<tr>
<td>SCCmec type V (SC2)</td>
<td>341</td>
</tr>
<tr>
<td>SCCmec type VI (4B)</td>
<td>3</td>
</tr>
<tr>
<td>SCCmec type VII (5C1)</td>
<td>0</td>
</tr>
<tr>
<td>SCCmec type VIII (4A)</td>
<td>5</td>
</tr>
<tr>
<td>SCCmec type IX (1C2)</td>
<td>0</td>
</tr>
<tr>
<td>SCCmec type X (7C1)</td>
<td>0</td>
</tr>
<tr>
<td>SCCmec type XI (8E)</td>
<td>8</td>
</tr>
</tbody>
</table>

Alerts

Multiple complexes

- ccr type 1, ccr type 2, mec class B
- ccr type 1, ccr type 4, ccr type 5, mec class C2
- ccr type 2, ccr type 4, mec class A
- ccr type 2, ccr type 4, mec class B
- ccr type 2, ccr type 4, mec class 5, mec class A

Contradictory prediction at:

- Type level: 76
- Subtype level: 76
- Failure of BLAST: 73
- Failure of k-mer: 27
- SCC-like: 2

MSSA

- No alerts, no SCCmec element: 773

Alerts

- Failure of BLAST: 35
- SCC-like: 23

In all, 6,021 MRSA and 831 MSSA isolates were tested.

The modifications included lowering the thresholds to >50% nucleotide identity and >50% template coverage and changing from the reference database to the extended database. In all, 6,022 MRSA and 830 MSSA isolates were tested.

A total of 147/168 strains causing alerts belonged to the same ST (ST5), indicating a single NT element.
of contigs. SCCmecFinder always detected SCCmec types VII (5C1) and X (7C1) by only one of the approaches, which could be indicative of a systematic error by SCCmecFinder. However, further investigation revealed no systematic errors.

**SCCmec subtype prediction by SCCmecFinder.** The subtyping predictions by SCCmecFinder of SCCmec types IV (2B) and V (SC2) were initially evaluated by using the reference elements listed by the IWG-SCC. The evaluation showed perfect concordance. The subtyping predictions were further evaluated by comparing the findings of SCCmecFinder with those of PCR-based SCCmec typing (Table 4). Sorting on the basis of the ST type, spa type, and subtype predicted by SCCmecFinder revealed a clustering of the isolates. Comparing the clustering results with the expected subtype prediction based on the multilocus ST, the spa type, a literature search, and PCR-based subtyping revealed a high concordance of the subtyping predictions, indicating that the SCCmec (sub)types are relatively stable within the ST/spa type clone (Table 4).

**Database curation and identification of novel (sub)types.** SCCmecFinder is designed to detect all of the reference SCCmec types listed by the IWG-SCC, and ideally, novel SCCmec (sub)types will be added to the databases by the curator after agreement with the IWG-SCC group upon the detection of a novel SCCmec (sub)type. Likewise, we suggest that notification of the IWG-SCC should be reported in the article describing the element.

### TABLE 4 Comparison of the subtype predictions of SCCmec types IV (2B) and V (SC2) from the clinical data set by the three SCCmec typing methods

<table>
<thead>
<tr>
<th>Strain</th>
<th>ST/spa type</th>
<th>PCR-based protocol&lt;sup&gt;a&lt;/sup&gt;</th>
<th>SCCmecFinder</th>
</tr>
</thead>
<tbody>
<tr>
<td>75178</td>
<td>ST772/t657</td>
<td>NT</td>
<td>Va (SC2)</td>
</tr>
<tr>
<td>75024</td>
<td>ST152/t355</td>
<td>NT</td>
<td>Vnt (SC2&amp;5)</td>
</tr>
<tr>
<td>75336</td>
<td>ST152/t355</td>
<td>NT</td>
<td>Vc (SC2&amp;5)</td>
</tr>
<tr>
<td>75054</td>
<td>ST398/t034</td>
<td>NT</td>
<td>Vc (SC2&amp;5)</td>
</tr>
<tr>
<td>75339</td>
<td>ST398/t011</td>
<td>NT</td>
<td>Vc (SC2&amp;5)</td>
</tr>
<tr>
<td>75049</td>
<td>ST398/t034</td>
<td>NT</td>
<td>Vc (SC2&amp;5)</td>
</tr>
<tr>
<td>75013</td>
<td>ST30/t019</td>
<td>IVC (2B)</td>
<td>IVC (2B)</td>
</tr>
<tr>
<td>75050</td>
<td>ST30/t019</td>
<td>IVC (2B)</td>
<td>IVC (2B)</td>
</tr>
<tr>
<td>75055</td>
<td>ST30/t1752</td>
<td>IVC (2B)</td>
<td>IVC (2B)</td>
</tr>
<tr>
<td>75061</td>
<td>ST30/t019</td>
<td>IVC (2B)</td>
<td>IVC (2B)</td>
</tr>
<tr>
<td>75181</td>
<td>ST5/t002</td>
<td>IVC (2B)</td>
<td>IVC (2B)</td>
</tr>
<tr>
<td>75208</td>
<td>ST30/t019</td>
<td>IVC (2B)</td>
<td>IVC (2B)</td>
</tr>
<tr>
<td>75108</td>
<td>ST45/t026</td>
<td>IVa (2B)</td>
<td>IVa (2B)</td>
</tr>
<tr>
<td>75190</td>
<td>ST45/t126</td>
<td>NT</td>
<td>IVg (2B)</td>
</tr>
<tr>
<td>75015</td>
<td>ST22/t032</td>
<td>ST</td>
<td>IVnt (2B)&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>75129</td>
<td>ST22/t022</td>
<td>NT</td>
<td>IVnt (2B)&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>75203</td>
<td>ST22/t032</td>
<td>NT</td>
<td>IVnt (2B)&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>75211</td>
<td>ST22/t379</td>
<td>NT</td>
<td>IVnt (2B)&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>75098</td>
<td>ST5/t1062</td>
<td>IVC (2B)</td>
<td>IVC (2B)</td>
</tr>
<tr>
<td>75180</td>
<td>ST5/t002</td>
<td>IVC (2B)</td>
<td>IVC (2B)</td>
</tr>
<tr>
<td>75019</td>
<td>ST5/t002</td>
<td>IVC (2B)</td>
<td>IVC (2B)</td>
</tr>
<tr>
<td>75125</td>
<td>ST5/t002</td>
<td>IVC (2B)</td>
<td>IVC (2B)</td>
</tr>
<tr>
<td>75182</td>
<td>ST5/t002</td>
<td>IVC (2B)</td>
<td>IVC (2B)</td>
</tr>
<tr>
<td>75201</td>
<td>ST30/t019</td>
<td>IVC (2B)</td>
<td>IVC (2B)</td>
</tr>
<tr>
<td>75103</td>
<td>ST30/t012</td>
<td>IVC (2B)</td>
<td>IVC (2B)</td>
</tr>
<tr>
<td>75319</td>
<td>ST8/t008</td>
<td>IVA (2B)</td>
<td>IVA (2B)</td>
</tr>
<tr>
<td>75056</td>
<td>ST8/t008</td>
<td>IVC (2B)</td>
<td>IVC (2B)</td>
</tr>
<tr>
<td>75100</td>
<td>ST8/t008</td>
<td>NT</td>
<td>IVnt (2B)&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>75147</td>
<td>ST1/t127</td>
<td>IVA (2B)</td>
<td>IVA (2B)</td>
</tr>
<tr>
<td>75058</td>
<td>ST6/t4403</td>
<td>IVA (2B)</td>
<td>IVA (2B)</td>
</tr>
<tr>
<td>75209</td>
<td>ST8/t690</td>
<td>IVA (2B)</td>
<td>IVA (2B)</td>
</tr>
</tbody>
</table>

<sup>a</sup>The PCR-based SCCmec typing protocol described by Kondo et al. (6).

<sup>b</sup>Subtype either IVh (2B) or IVj (2B).

<sup>c</sup>Contradictory subtype prediction.
The combination of two different approaches for the prediction of the SCCmec element in SCCmecFinder creates an opportunity for the user to assess whether the SCCmec element in question is a variant of an existing element or a novel element, as SCCmecFinder provides data on the detected target genes and the k-mer template coverage. An example is *S. aureus* JCSC7481 (GenBank accession no. AB774378), which the authors have stated is a variant of SCCmec type V (SC2). Running the SCCmec element of *S. aureus* JCSC7481 through SCCmecFinder by using the reference database, an alert about contradictory predictions was produced. The k-mer template coverage was 42.4%, indicating that the element differed from the reference elements and in concordance with the conclusion of the authors that it was a variant of SCCmec type V (SC2). However, the element of *S. aureus* JCSC7481 has, to our knowledge, not yet been assigned by the IWG-SCC, and hence, the SCCmec element has not yet been added to the reference database (but is currently found in the extended database). Of importance; researchers are encouraged to consult members of the IWG-SCC and the curator of the SCCmecFinder database to keep the number of NT SCCmec elements low.

**Conclusion.** In conclusion, SCCmecFinder ([https://cge.cbs.dtu.dk/services/SCCmecFinder](https://cge.cbs.dtu.dk/services/SCCmecFinder)) is a validated tool for rapid whole-genome-based SCCmec typing of MRSA that, together with the existing tools for the *in silico* prediction of the multilocus ST ([https://cge.cbs.dtu.dk/services/MLST](https://cge.cbs.dtu.dk/services/MLST)) and spa type ([https://cge.cbs.dtu.dk/services/spatyper](https://cge.cbs.dtu.dk/services/spatyper)), can provide a standardized nomenclature of *S. aureus* genomes. It is our intention that the SCCmecFinder database will be curated in a way that fulfills future needs to encompass the changing epidemiology of MRSA accompanied by increasing genic diversity.

**MATERIALS AND METHODS**

**Construction of SCCmecFinder.** SCCmecFinder combines two existing and validated gene prediction algorithms (10, 11), a BLAST-based approach and a k-mer-based approach. By utilizing the BLAST approach (BLASTn algorithm) with default thresholds set at ≥90% nucleotide identity and a minimum length of 60%, SCCmecFinder detects the presence of selected target genes, identifies the mec and ccr gene complex, and types the SCCmec element, whereas by utilizing the k-mer approach, SCCmecFinder types the element on the basis of homology to reference SCCmec elements by identifying the number of co-occurring DNA sequences with the length k, known as k-mers, between the query genome and the entries in the curated database, of which only hits with ≥50% k-mer template coverage are considered (Fig. 2). Less stringent thresholds for the BLAST-based approach can be set, but less stringent thresholds could lead to outputs that require more downstream analysis and might give misleading annotations. The software can be found at [https://bitbucket.org/genomicepidemiology/sccmecfinder.git](https://bitbucket.org/genomicepidemiology/sccmecfinder.git). Prior for choosing the thresholds for each algorithm, a parameterization study was conducted.

SCCmecFinder compares the predicted SCCmec types from the BLAST- and the k-mer-based predictions and checks for contradictory results to achieve correct SCCmec annotation. Identification of SCCmec type IV (2B) or V (SC2) results in subtyping of the element. Subtyping is performed as with the typing of the SCCmec element, on the basis of a BLAST search for selected subtyping target genes from the J1 region and the k-mer-based approach. On the basis of the comparison, SCCmecFinder outputs one or more of the five possible output types, e.g., notification of a positive/negative prediction, a contradictory SCCmec type/subtype, a missing/additional gene(s), the detection of multiple gene complexes, or the presence of a possible SCC-like element. If one of the prediction approaches fails, the output will be negative but will contain information about the possible presence of an SCCmec element to guide the user to perform further analysis (Fig. 2).

**Sequence databases for SCCmecFinder.** Two databases were developed to support the two different approaches. For the BLAST-based approach, a collection of 37 target gene sequences (Table S1 and S2) and for the k-mer-based approach, a collection of 36 archetypal reference SCCmec were selected to represent sequences of known types available at/according to the IWG-SCC (Table S3). Additionally, an extended version of the database for the k-mer-based approach, containing all of the archetypal reference SCCmec elements and variants of SCCmec type V (SC2 and SC2&5) from *S. aureus* that have not yet been validated or assigned a subtype by the IWG-SCC, was also generated (Table S4). Each entry in all three databases is accompanied by an NCBI GenBank accession number, thus allowing for the correlation of an identified gene(s)/element(s) in the query genome with an annotated gene(s)/element(s) in the archetypal reference SCCmec elements. The databases for SCCmecFinder can be found at [https://bitbucket.org/genomicepidemiology/sccmecfinder_db.git](https://bitbucket.org/genomicepidemiology/sccmecfinder_db.git).

**User interface of SCCmecFinder.** SCCmecFinder builds on the common design of the CGE user interface. The upload page consists of four dropdown menus by which the type of sequencing platform can be chosen and default settings for the analysis can be changed. The output page consists of three sections, with the first reporting the presence/absence of the *mecaA/mecC* gene. The second section is the actual output of the SCCmec prediction, in which the mec and ccr gene complexes, SCCmec
type, and possible subtype from both approaches are reported. The third and last section is reporting of the genes detected by the BLAST-based approach and the top five hits from the k-mer-based approach.

SCCmecFinder can be done by using either preassembled draft genomes or raw reads from the 454, Ion Torrent, or Illumina sequencing platform. Prior to the BLAST-based prediction, 454 and Ion Torrent raw reads are de novo assembled by using Newbler, while Illumina raw reads are de novo assembled by using Velvet (10).

The integrity of SCCmecFinder was initially evaluated by submitting reference genomes and the archetypal SCCmec elements included in the k-mer reference database by using the default thresholds. After the initial validation process, draft genomes of 93 clinical isolates were submitted to SCCmecFinder without further processing. The isolates covered a diverse collection of 43 MRSA and 50 MSSA strains representing 13 different STs of predominant MRSA and MSSA clones in Denmark. The genomes were obtained after genomic DNA extraction (DNeasy Blood and Tissue kit; Qiagen, Copenhagen, Denmark), and fragment libraries were constructed by using an in-house library protocol (12), followed by 100-bp paired-end sequencing (HiSeq; Illumina) in accordance with the manufacturer’s instructions. De novo assembly was performed with the modified Velvet algorithm integrated into the web tool Assembler v1.2 (available at https://cge.cbs.dtu.dk/services/Assembler/). The in silico predictions by SCCmecFinder were evaluated by comparing the results with those predicted by the PCR-based protocol of Kondo et al. (6), M-PCRs 1 and 2. Subtyping were performed by M-PCR 3 as described by Kondo et al. (6). Composite elements were detected by ensuring the presence of both ccr gene complexes in SCCmec types IV (2B&5) and V (5C2&5), and an in-house protocol was used for the typing of ccrC1 gene alleles 2 and 8, respectively, for the verification of SCCmec type V (SC2&5) (13).

Additional genomes. Finally, 6,852 publicly available S. aureus draft genomes from the NCBI Reference Sequence Database (RefSeq [https://ncbi.nlm.nih.gov/refseq/], accessed 16 September 2016) were downloaded to validate the SCCmecFinder outputs for any systematic errors. The assemblies were subjected to rough quality filtering on the basis of being S. aureus (determined on the basis of MLST and single nucleotide polymorphism and phylogenetic analyses), their genome size, number of contigs, and N50 value, where genomes with a size of <2.5 Mb, an N50 value of <10,000, and >1,000 contigs were discarded. All genomes were subjected to MLST (http://github.com/tseemann/mlst) and subsequently submitted to SCCmecFinder by using default thresholds.

Accession number(s). All of clinical sequences included in this study are available at the European Nucleotide Archive under project accession number PRJEB22380.
SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at https://doi.org/10.1128/mSphere.00612-17.

TABLE S1, PDF file, 0.1 MB.
TABLE S2, PDF file, 0.1 MB.
TABLE S3, PDF file, 0.1 MB.
TABLE S4, PDF file, 0.1 MB.

ACKNOWLEDGMENTS

This work was supported by the Danish Ministry of Health and Prevention as part of the surveillance of MRSA, the Danish Council for Strategic Research (grant 09-067103), and the European Union Horizon 2020 research and innovation program under grant agreement 643476 to the COMPARE project (http://www.compare-europe.eu).

REFERENCES