Systematic Derivation of Marker Sets for Staphylococcal Cassette Chromosome mec Typing††

Alex J. Stephens, Flavia Huygens, and Philip M. Giffard*

Institute of Health and Biomedical Innovation, Queensland University of Technology, Kelvin Grove QLD 4059, Australia

Received 24 October 2006/Returned for modification 11 December 2006/Accepted 18 April 2007

The aim of this study was to identify optimized sets of genotyping targets for the staphylococcal cassette chromosome mec (SCCmec). We analyzed the gene contents of 46 SCCmec variants in order to identify minimal subsets of targets that provide useful resolution. This was achieved by firstly identifying and characterizing each available SCCmec element based on the presence or absence of 34 binary targets. This information was used as input for the software “Minimum SNPs,” which identifies the minimum number of targets required to differentiate a set of genotypes up to a predefined Simpson’s index of diversity (D) value. It was determined that 22 of the 34 targets were required to genotype the 46 SCCmec variants to a D of 1. The first 6, 9, 12, and 15 targets were found to define 21, 29, 35, and 39 SCCmec variants, respectively. The genotypes defined by these marker subsets were largely consistent with the relationships between SCCmec variants and the accepted nomenclature. Consistency was made virtually complete by forcing the computer program to include ccr1 and ccr5 in the target set. An alternative target set biased towards discriminating abundant SCCmec variants was derived by analyzing an input file in which common SCCmec variants were repeated, thus ensuring that markers that discriminate abundant variants had a large effect on D. Finally, it was determined that mecA single nucleotide polymorphisms (SNPs) can increase the overall genotyping resolution, as different mecA alleles were found in otherwise identical SCCmec variants.

Comparative bacterial genomics is revealing numerous hypervariable regions in bacterial chromosomes. Interrogation of such regions can efficiently provide an epidemiological fingerprint, or insight into pathogenic capability, antimicrobial resistance phenotype, or vaccine susceptibility (7, 9, 11). However, the often complex nature of the variations can make it difficult to devise standardized protocols for subtyping such regions.

An excellent example of a clinically relevant hypervariable region in a major bacterial pathogen is the staphylococcal cassette chromosome mec (SCCmec) (16, 19). This element defines methicillin-resistant Staphylococcus aureus (MRSA) through carriage of the beta-lactam resistance gene mecA. A combination of multilocus sequence type (MLST) and SCCmec type is frequently used as an identifier for MRSA clones (8, 31, 36). As a consequence, in recent years there has been considerable interest in developing SCCmec subtyping methods.

A number of broadly similar but distinct SCCmec typing methods have been described. These methods include PCR-based detection of type-specific targets (ccr variants, mec gene complexes, and junkyard regions [10, 29, 30, 46]), restriction digests (43, 45), and full sequence analysis (15–17). In general, SCCmec is divided into five major types based on the combination of mec class and recombinase-encoding (ccr) genes present (15, 17, 25). The mec classes are themselves defined by the arrangement of genes adjacent to mecA (38). Subtypes are defined by binary variation (presence or absence) of sequence blocks in the “junkyard” regions (4). The various typing methods and associated terminology have evolved in a somewhat ad hoc fashion as more SCCmec variants have been discovered. Recently, a proposal for a rationalized SCCmec typing nomenclature was published (4). This proposal has merit, as it incorporates the core structural features and the variably junkyard regions. However, there is currently no standardized SCCmec genotyping method that systematically and efficiently utilizes all known variations.

Our research group has previously developed the computer program “Minimum SNPs” (35). This was designed to derive sets of single nucleotide polymorphisms (SNPs) from MLST databases on the basis of maximization of resolving power. Resolving power may be measured either on the basis of the power to discriminate a user-defined sequence type from all other sequence types or on the basis of maximization of the Simpson’s index of diversity (D) (13). More recently, this approach was adapted to the identification of resolution-optimized sets of binary markers. Price and coworkers derived a set of such markers from Campylobacter jejuni comparative genome hybridization data on the basis of D maximization, and these markers were shown to have considerable utility as genotyping targets (33). SCCmec diversity can also be considered to be a database of binary gene variation and is therefore amenable to a similar analysis. Accordingly, we have carried this out, with the central hypothesis of the study being that the sets of markers derived by our systematic approach would efficiently provide resolving power.

* Corresponding author. Mailing address: Institute of Health and Biomedical Innovation, Queensland University of Technology, 60 Musk Ave., Kelvin Grove QLD 4059, Australia. Phone: 61 7 3138 6194. Fax: 61 7 3138 6030. E-mail: p.giffard@qut.edu.au.
† Supplemental material for this article may be found at http://aac.asm.org/.
‡ Published ahead of print on 21 May 2007.
FIG. 1. Illustration of the strategy used to identify sets of resolution-optimized binary markers. Each genotype of interest is characterized based on the presence (T) or absence (A) of the total set of binary markers (five are used in this example). Thus the binary gene configuration is converted into a pseudo-DNA sequence composed of A’s and T’s. The alignment of pseudo-DNA sequences is then analyzed using Minimum SNPs for combinations of binary markers that maximize the Simpson’s index of diversity (D) (33). In the example shown above, binary genes 2 and 4 provide a D of 1, i.e., they completely resolve the three genotypes. In the present study, 46 SCCmec types were defined by the presence or absence of 34 binary markers. Even with this relatively small data set, manual identification of resolution-optimized marker sets is extremely difficult.

MATERIALS AND METHODS

Identification of SCCmec variants. A literature and NCBI database search was undertaken to identify SCCmec variants for this study. Variants were named according to the proposed SCCmec nomenclature system (4). This procedure was complicated by the fact that described SCCmec variants differ in the detail to which they have been assessed; some are completely sequenced, while others are classified only through PCR amplification. In a small number of instances, SCCmec variants reported in the literature lacked sufficient structural information to be included in this study.

Identification of resolution-optimized sets of binary markers. Sets of binary markers were identified using the computer program “Minimum SNPs” by a strategy illustrated in Fig. 1 (33, 35). This program extracts resolution-optimized sets of SNPs from DNA sequence alignments. It does this by identifying the single SNP with the highest resolving power, labeling this as SNP 1, and then identifying the SNP that in combination with SNP 1 gives the highest resolving power and labeling that as SNP 2, etc. A valuable feature of the software is that the user can force the program to include or exclude any SNP in/from the SNP set.

Minimum SNPs can measure resolving power in more than one way, but the most generally applicable method is by calculation of D with respect to the sequence alignment. This algorithm was used throughout this study. Resolution-optimized sets of binary markers were identified by first converting the binary marker data for the SCCmec variants into a string of “A’s” and “T’s,” with “A” denoting binary marker absence and “T” denoting binary marker presence. In this way the binary data for each SCCmec variant becomes a pseudo-DNA sequence that can be aligned with other pseudo-DNA sequences representing other SCCmec variants. This alignment was then mined by Minimum SNPs in order to identify sets of binary markers that give a high D value with respect to that alignment. The alignment input file reflecting estimated SCCmec abundance included 49 extra copies of variants: 1B.1.1 (I), 2A.1.1 (II), 3A1.1.1 (III), 3A1.2 (IIIA), 3A1.3 (IIIB), 2B.1 (IVa), 2B.2.1 (IVb), 2B.3.1 (IVc), and 5C.1 (V).

mecA nucleotide sequence determination and SNP analysis. The mecA genes from 19 diverse Australian MRSA isolates from nine MLST types were amplified using primers mecA-F1 and mecA-R3 (Sigma-Prolog, Lismore, Australia) (39, 41). The amplicons were purified using Exo-SapIt (Amersham Biosciences, Castle Hill, Australia) for 15 min at 37°C and 15 min at 80°C and then sequenced for 1,400 bp from the 3’ end using primers mecA-F1 and mecA-R2. The sequence traces were viewed and analyzed in SeqMan II version 4.06 (DNAStar, Madison, Wis.).

The literature and NCBI databases were searched for S. aureus mecA gene sequences whose corresponding SCCmec type was known. Twenty-five mecA sequences were identified, cropped to 1,440 bp, and added to a sequence alignment with the 19 partial mecA sequences identified in this study. Clustal X (v. 1.64b) was used to align the partial mecA gene sequences to identify SNPs. In order to investigate the possibility of combinatorial genotyping methods based on SNPs and binary genes, the SNP data were added to the pseudo-DNA sequences derived from the binary marker genotypes, so as to construct an input file for Minimum SNPs that would allow the identification of resolution-optimized SNP-binary marker combinations.

For full details of the mecA sequences used in this analysis and the isolates from which they were derived, see the supplemental material.

Nucleotide sequence accession numbers. The new mecA sequences have been deposited in GenBank under accession numbers EF692630, EF692631, and EF692632.

RESULTS

Collation of currently available data concerning SCCmec binary diversity. The overall aim of this study was to apply a

<table>
<thead>
<tr>
<th>Region</th>
<th>Target</th>
<th>ccr</th>
<th>mec class</th>
<th>J1</th>
<th>J2/J3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Core</td>
<td>ccr1</td>
<td>A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ccr2</td>
<td>A1&quot;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ccr3</td>
<td>A3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ccr4</td>
<td>A4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ccr5</td>
<td>B</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ccrC-VT</td>
<td>B1b</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>C2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>E</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>F</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>G</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Junkyard</td>
<td>a</td>
<td>pT181, dcs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>p1258, pls</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>c</td>
<td>pUB110, kdp</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>d</td>
<td>Tn4001, 3A2.1 (unique)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>g</td>
<td>Tn554 MLS, ¶Tn554 (cad)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>IS256-dcs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>IS256-Tn4001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>IS256-mecI</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a mec class A1 is characterized by a 166-bp deletion in mecRI (15, 24, 32).
b Two separate mec classes have been termed B1. For this analysis, the mec class described by Lim et al. remains B1 and the mec configuration described by Shukla et al. has been renamed mec class F (23, 39).
<table>
<thead>
<tr>
<th>SCCmec class</th>
<th>SCCmec (n = 46)</th>
<th>Uniform nomenclature (a)</th>
<th>Defining characteristics</th>
<th>PCR genotype (a)</th>
<th>Strain/isolate and/or sequence (accession no.)</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I (1B)</td>
<td>1B.1.1</td>
<td>ccr1, mec class B</td>
<td>pIs</td>
<td>I</td>
<td>NCTC10442 (AB053763), COL (CP000046)</td>
<td>15</td>
</tr>
<tr>
<td>I variant</td>
<td>(1B.2.1)</td>
<td>ccr1, mec class B</td>
<td>No pIs</td>
<td>I</td>
<td>PHeI</td>
<td>38</td>
</tr>
<tr>
<td>IA</td>
<td>1B.1.2</td>
<td>ccr1, mec class B</td>
<td>pIs</td>
<td>I</td>
<td>PER34</td>
<td>31</td>
</tr>
<tr>
<td>IA variant</td>
<td>(1B.1.3)</td>
<td>ccr1, mec class B</td>
<td>pIs</td>
<td>I</td>
<td>PER31</td>
<td>31</td>
</tr>
<tr>
<td>IA variant</td>
<td>(1B.1.4)</td>
<td>ccr1, mec class B</td>
<td>pIs</td>
<td>I</td>
<td>PER88</td>
<td>31</td>
</tr>
<tr>
<td>II (2A)</td>
<td>IIa</td>
<td>ccr2, mec class A</td>
<td>kalp operon</td>
<td>II</td>
<td>N315 (D86934), MRSA252 (BX571856)</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>II variant</td>
<td>ccr2, mec class A</td>
<td>kalp operon</td>
<td>II</td>
<td>Not described</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>II variant</td>
<td>ccr2, mec class A</td>
<td>J1 of IVb (2B.2.1)</td>
<td>II</td>
<td>Not described</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>IIa</td>
<td>ccr2, mec class A</td>
<td>J1 of IVb (2B.2.1)</td>
<td>II</td>
<td>Not described</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td>IIB</td>
<td>ccr2, mec class A</td>
<td>J1 of IVb (2B.2.1)</td>
<td>II</td>
<td>AJ810123</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td>IID</td>
<td>ccr2, mec class A</td>
<td>J1 of IVb (2B.2.1)</td>
<td>II</td>
<td>Not described</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td>IIIE</td>
<td>ccr2, mec class A</td>
<td>J1 of IVb (2B.2.1)</td>
<td>II</td>
<td>AR13.1/330.2 (AJ810120)</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td>IIb</td>
<td>ccr2, mec class A</td>
<td>Unique sequence for 2A2</td>
<td>II</td>
<td>JCS3063 (AB127982)</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>IIb variant</td>
<td>ccr2, mec class A</td>
<td>dcs and no pUB110</td>
<td>NV</td>
<td>Not described</td>
<td>40</td>
</tr>
<tr>
<td>III (3A)</td>
<td>III</td>
<td>ccr3, mec class A</td>
<td>Tn554 MLS, pT181, and pl258</td>
<td>III</td>
<td>HUSA304</td>
<td>23, 31</td>
</tr>
<tr>
<td></td>
<td>III</td>
<td>ccr3, mec class A</td>
<td>Tn554 MLS, pT181, and pl258</td>
<td>III</td>
<td>85/2082 (AB037671), ANS46</td>
<td>15, 31</td>
</tr>
<tr>
<td></td>
<td>III</td>
<td>ccr3, mec class A</td>
<td>No pT181 or ptp</td>
<td>III</td>
<td>HU25 (AF422696)</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>III</td>
<td>ccr3, mec class A</td>
<td>No pT181, pl258 or Tn554</td>
<td>III</td>
<td>HDG2</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>III</td>
<td>ccr3, mec class A</td>
<td>No Tn554</td>
<td>III</td>
<td>R35</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>III</td>
<td>ccr3, mec class A</td>
<td>No Tn554</td>
<td>III</td>
<td>15814-9852</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td>III</td>
<td>ccr3, mec class A</td>
<td>Tn554 cad with tsp4</td>
<td>III</td>
<td>85/907</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>III</td>
<td>ccr3, mec class A</td>
<td>Tn554 MLS</td>
<td>III</td>
<td>85/961</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>III</td>
<td>ccr3, mec class A</td>
<td>dcs</td>
<td>III</td>
<td>DOM068</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>III</td>
<td>ccr3, mec class A</td>
<td>dcs</td>
<td>III</td>
<td>HSA10</td>
<td>1</td>
</tr>
<tr>
<td>IV (2B)</td>
<td>IVa</td>
<td>ccr2, mec class B</td>
<td>Specific for 2B.1 (IVa)</td>
<td>Iva</td>
<td>MW2 (BA000033), CA05 (AB063172)</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>IVb</td>
<td>ccr2, mec class B</td>
<td>Specific for 2B.2 (IVb)</td>
<td>IVb</td>
<td>8/6-3P (AB063173)</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>IVc</td>
<td>ccr2, mec class B</td>
<td>Specific for 2B.3.1 (IVc)</td>
<td>IVc</td>
<td>MR108 (AB096217)</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>IVc variant</td>
<td>ccr2, mec class B</td>
<td>No Tn6001</td>
<td>IVc</td>
<td>2314 (AY271717)</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>IVd</td>
<td>ccr2, mec class B</td>
<td>Identical to 2B.4 (IVd)</td>
<td>IVd</td>
<td>JCS346 (AB097677)</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>IVE</td>
<td>ccr2, mec class B</td>
<td>Unique left extremity</td>
<td>IVe</td>
<td>AR43 (33330.2 (AJ810121)</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td>IVF</td>
<td>ccr2, mec class B</td>
<td>Unique left extremity</td>
<td>IVe</td>
<td>AR43</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td>IVG</td>
<td>ccr2, mec class B</td>
<td>Specific for 2B.5 (IVg)</td>
<td>IVb</td>
<td>M03-68 (DQ1006887)</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>IV variant</td>
<td>ccr2, mec class B</td>
<td>Different from IVa, -b, -c, or -d</td>
<td>IVb</td>
<td>SD179-1</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>IV variant</td>
<td>ccr4, mec class B</td>
<td>Is256</td>
<td>IVa</td>
<td>HDE286 (AF419135)</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>IVA</td>
<td>ccr2, mec class B</td>
<td>Is256</td>
<td>IVA</td>
<td>BARGH117</td>
<td>31, 34</td>
</tr>
<tr>
<td></td>
<td>IVA</td>
<td>ccr2, mec class B</td>
<td>pUB110</td>
<td>IVA</td>
<td>PER2</td>
<td>3, 30</td>
</tr>
<tr>
<td>V (SC)</td>
<td>V</td>
<td>ccr5, mec class C2</td>
<td>hsd</td>
<td>V</td>
<td>WIS (AB121219)</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>V variant</td>
<td>ccr5, mec class B</td>
<td>No details described</td>
<td>NV</td>
<td>04-17489</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>V variant</td>
<td>ccr5, mec class B</td>
<td>No details described</td>
<td>NV</td>
<td>WBG3804</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>V variant</td>
<td>ccr5, mec class E</td>
<td>No details described</td>
<td>NV</td>
<td>WBG10198</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>VT</td>
<td>ccr2, mec class C2</td>
<td>hsd</td>
<td>V</td>
<td>TSGH17</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>V variant</td>
<td>ccr5 and ccr2, mec class C2</td>
<td>hsd</td>
<td>NV</td>
<td>ZH 47 (AM292304) Not published This study: 41</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td>V variant</td>
<td>ccr5 and ccr1, mec class C2</td>
<td>No details described</td>
<td>NV</td>
<td>B827549</td>
<td>28</td>
</tr>
<tr>
<td>Other</td>
<td>NV</td>
<td>ccr2, mec class C2</td>
<td>No details described</td>
<td>NV</td>
<td>B827549</td>
<td>28</td>
</tr>
</tbody>
</table>

(a) Table adapted from Chongtrakool et al. (4) with permission.
(b) Uniform nomenclature values in parentheses denote tentative naming of unnamed variants as defined by Chongtrakool et al. (4).
(c) Hypothetical SCCmec genotype using standard PCR assays for ccr, mec class, and the 2B J1 regions a, b, c, and d. NV and NT indicate a new variant and a nontypeable variant, respectively.
(d) Shukla et al. described this isolate as a SCCmec 2A variant (40).
(e) The mec class of isolate 85/2082 features a 166-bp deletion of the mecR1 membrane spanning domain, which is characteristic for mec class A1 (15, 24, 32).

Downloaded from http://aac.asm.org on September 7, 2017 by guest
systematic approach for the identification of genetic targets to efficiently genotype SCCmec. In order to do this effectively, it was necessary to extract all available SCCmec data from the literature and the publicly available databases. Diversity can be in the form of binary variability (gene presence/absence) or SNPs. As complete sequences are known for only a subset of the genes that make up all the known SCCmec variants, binary diversity data were collated and analyzed first.

Analysis of the literature and online databases resulted in the definition of 34 SCCmec-associated binary targets (Table 1). Unique sequences associated with the targets (see the supplemental material) were used to facilitate the design of genotyping methods. The 34 binary targets in turn defined 46 SCCmec variants (Table 2). Each SCCmec element differing at one or more of the 34 binary targets (Table 1) was considered a separate variant. This extensive compilation of SCCmec variants reveals the extensive rearrangement and mutation that the five major types have undergone during their evolutionary histories, with diversity in the junkyard regions defining the majority of the variants.

**Identification of a resolution-optimized set of binary targets, without constraints.** The database of binary gene variation was converted into a pseudo-sequence alignment (see Materials and Methods) and analyzed using Minimum SNPs for sets of markers that provide a high D. For the first experiment, a set of binary markers was derived from the data in Table 2 simply on the basis of maximization of D, with no attempt to make the results consistent with previously described typing methods. It was determined that all 46 SCCmec variants can be completely resolved by interrogating 22 of the 34 binary targets. Table 3 shows the resolving powers of the six-, nine-, 12-, and 15-target sets, which differentiate 21, 29, 35, and 39 of the 46 variants, respectively.

These targets were then compared with those used in currently extant SCCmec typing methods, with respect to their identity, resolving power, and whether the variants they define correspond with current SCCmec classification schemes. Current PCR-based SCCmec typing generally makes use of 10 targets (ccrl, -2, -3, and -5, mec classes A, B, and C, and three junkyard regions from type 2B) to assign a SCCmec element to one of the five major types and also provide some subtyping information. When tested against the data set containing 46 variants, these 10 targets discriminated 18 genotypes. In comparison, the first 10 targets derived on the basis of D maximization discriminated 32 genotypes. This supports our conjecture that binary target identification on the basis of computerized D maximization will provide a superior result to an ad hoc approach. In order to further test this, the resolving powers of randomly selected sets of markers were determined. This always gave a much lower resolving power than marker sets selected on the basis of D maximization (data not shown), thus supporting the utility of our method for target selection.

**A resolution-optimized set of binary targets nucleated by ccrl and ccrr5.** It would be desirable for a SCCmec genotyping method to define genotypes consistent with accepted terminology and current models concerning the degrees of relatedness of the different SCCmec variants. In other words, a genotyping method that fails to discriminate some pairs of distantly related SCCmec variants may be regarded as problematic. The genotypes defined by the D-maximized marker set largely met this requirement. However, the six- and nine-target sets were unable to discriminate the unrelated 4B and truncated SCCmec from 1B (I) variants and also the 5/2G.1 and 2C.1 from 2B (IV) variants (Table 3). Accordingly, a second set of resolution-optimized targets was derived using the “Include” function in Minimum SNPs. This forces the program to include a chosen marker(s) in the set. In this instance, ccrl and ccrr5 were forced into the set, because these discriminate the variants that are not discriminated with the original marker set. In this analysis, Minimum SNPs builds marker sets using the set of ccrl plus ccrr5 as a starting point, so in effect the derived marker sets are nucleated by ccrl and ccrr5.

The derived marker sets are very similar to the unconstrained set and provided almost the same resolution. However, the occurrence of incorrectly grouped variants was largely rectified. The only instances of disparate variants being grouped together were 2C.1 with 2B variants and 2B.N.2 with 2A variants when using the six-target set (Table 4).

**A resolution-optimized target set derived from an input file reflecting SCCmec variant abundances.** An assumption inherent in our approach to identifying resolution-optimized sets of genotyping targets is that the database is a useful surrogate of the population structure. If this is the case, then the resolving power of the targets with respect to the database is a useful measure of their resolving power on actual collections of isolates. This assumption, although unlikely to be completely wrong, may be simplistic. This is because all of the SCCmec variants are not similarly abundant. This could result in a difference between the D value calculated from the database and the D value obtained from actual collections of isolates. A corollary of this is that markers that discriminate between the abundant genotypes should be preferentially included in the marker set if the D for actual collections of isolates is to be maximized.

We addressed this issue by creating an input file for “Minimum SNPs” that contained multiple copies of abundant SCCmec variants. A comprehensive and accurate determination of relative abundances in nature was not practical, so this exercise was carried out somewhat crudely; the published literature was used to make a judgment as to which SCCmec variants are abundant, and these variants were repeated 50 times in the Minimum SNPs input file. It was hypothesized that the derived marker sets would provide a very high performance at discriminating the abundant variants because a marker that discriminated between abundant variants would have a big impact on the D value. It was also predicted that the derived markers would still be effective for discriminating the less-abundant variants from the abundant variants and from each other.

The resolution-optimized marker sets from this exercise are shown in Table 5. The target sets identified are significantly different from the unconstrained and ccrl- and ccrr5-nucleated target sets, and interestingly, do not discriminate as many genotypes. However, as expected, this marker set provided a performance superior to the unconstrained and ccrl- and ccrr5-nucleated marker sets in resolving the nine SCCmec variants identified as being abundant. There were, however, a small number of instances where the six-target set failed to discriminate rare SCCmec variants from other unrelated SCCmec variants. This was partially rectified by manually promoting the
<table>
<thead>
<tr>
<th>SCCmec class</th>
<th>6 targets (ccr2, Tn554 MLS, dcs, pUB110, mec class A, and J1b)</th>
<th>9 targets (previous 6 plus mec class C2, IS256-dcs, and pT181)</th>
<th>12 targets (previous 9 plus mec class B, ccr1, and mec class A.3)</th>
<th>15 targets (previous 12 plus J1a, J1c, and pls)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1B (I)</td>
<td>1B.1.1</td>
<td>1B.1.1</td>
<td>1B.1.1</td>
<td>1B.1.1</td>
</tr>
<tr>
<td></td>
<td>(1B.2.1)</td>
<td>(1B.2.1)</td>
<td>(1B.2.1)</td>
<td>(1B.2.1)</td>
</tr>
<tr>
<td></td>
<td>4B*</td>
<td>4B*</td>
<td>4B*</td>
<td>4B*</td>
</tr>
<tr>
<td></td>
<td>(1B.1.3)</td>
<td>(1B.1.3)</td>
<td>(1B.1.3)</td>
<td>(1B.1.3)</td>
</tr>
<tr>
<td></td>
<td>1B.1.2</td>
<td>1B.1.2</td>
<td>1B.1.2</td>
<td>1B.1.2</td>
</tr>
<tr>
<td></td>
<td>Truncated*</td>
<td>Truncated*</td>
<td>Truncated*</td>
<td>Truncated*</td>
</tr>
<tr>
<td></td>
<td>(1B.1.4)</td>
<td>(1B.1.4)</td>
<td>(1B.1.4)</td>
<td>(1B.1.4)</td>
</tr>
<tr>
<td>2A (II)</td>
<td>2A.3.1</td>
<td>2A.3.1</td>
<td>2A.3.1</td>
<td>2A.3.1</td>
</tr>
<tr>
<td></td>
<td>2A.3.3</td>
<td>2A.3.3</td>
<td>2A.3.3</td>
<td>2A.3.3</td>
</tr>
<tr>
<td></td>
<td>2A.3.4</td>
<td>2A.3.4</td>
<td>2A.3.4</td>
<td>2A.3.4</td>
</tr>
<tr>
<td></td>
<td>2A.3.5</td>
<td>2A.3.5</td>
<td>2A.3.5</td>
<td>2A.3.5</td>
</tr>
<tr>
<td></td>
<td>2A.1.1</td>
<td>2A.1.1</td>
<td>2A.1.1</td>
<td>2A.1.1</td>
</tr>
<tr>
<td></td>
<td>2A.3.2</td>
<td>2A.3.2</td>
<td>2A.3.2</td>
<td>2A.3.2</td>
</tr>
<tr>
<td></td>
<td>2A.2</td>
<td>2A.2</td>
<td>2A.2</td>
<td>2A.2</td>
</tr>
<tr>
<td></td>
<td>2A.1.2</td>
<td>2A.1.2</td>
<td>2A.1.2</td>
<td>2A.1.2</td>
</tr>
<tr>
<td></td>
<td>(2A.1.3)</td>
<td>(2A.1.3)</td>
<td>(2A.1.3)</td>
<td>(2A.1.3)</td>
</tr>
<tr>
<td></td>
<td>(2F.1.1)</td>
<td>(2F.1.1)</td>
<td>(2F.1.1)</td>
<td>(2F.1.1)</td>
</tr>
<tr>
<td>3A (III)</td>
<td>3A.1.1</td>
<td>3A.1.1</td>
<td>3A.1.1</td>
<td>3A.1.1</td>
</tr>
<tr>
<td></td>
<td>(3A.2.1)</td>
<td>(3A.2.1)</td>
<td>(3A.2.1)</td>
<td>(3A.2.1)</td>
</tr>
<tr>
<td></td>
<td>(3A.3.1)</td>
<td>(3A.3.1)</td>
<td>(3A.3.1)</td>
<td>(3A.3.1)</td>
</tr>
<tr>
<td></td>
<td>3A.1.2</td>
<td>3A.1.2</td>
<td>3A.1.2</td>
<td>3A.1.2</td>
</tr>
<tr>
<td></td>
<td>3A.1.3</td>
<td>3A.1.3</td>
<td>3A.1.3</td>
<td>3A.1.3</td>
</tr>
<tr>
<td></td>
<td>(3A.1.5)</td>
<td>(3A.1.5)</td>
<td>(3A.1.5)</td>
<td>(3A.1.5)</td>
</tr>
<tr>
<td></td>
<td>(3A.1.6)</td>
<td>(3A.1.6)</td>
<td>(3A.1.6)</td>
<td>(3A.1.6)</td>
</tr>
<tr>
<td></td>
<td>(3A.4.1)</td>
<td>(3A.4.1)</td>
<td>(3A.4.1)</td>
<td>(3A.4.1)</td>
</tr>
<tr>
<td></td>
<td>(3A.1.1.1)</td>
<td>(3A.1.1.1)</td>
<td>(3A.1.1.1)</td>
<td>(3A.1.1.1)</td>
</tr>
<tr>
<td></td>
<td>(3A.1.4)</td>
<td>(3A.1.4)</td>
<td>(3A.1.4)</td>
<td>(3A.1.4)</td>
</tr>
<tr>
<td>2B (IV)</td>
<td>2B.4</td>
<td>2B.4</td>
<td>2B.4</td>
<td>2B.4</td>
</tr>
<tr>
<td></td>
<td>(2B.6)</td>
<td>(2B.6)</td>
<td>(2B.6)</td>
<td>(2B.6)</td>
</tr>
<tr>
<td></td>
<td>2B.1</td>
<td>2B.1</td>
<td>2B.1</td>
<td>2B.1</td>
</tr>
<tr>
<td></td>
<td>2B.3.2</td>
<td>2B.3.2</td>
<td>2B.3.2</td>
<td>2B.3.2</td>
</tr>
<tr>
<td></td>
<td>(2B.1.2)</td>
<td>(2B.1.2)</td>
<td>(2B.1.2)</td>
<td>(2B.1.2)</td>
</tr>
<tr>
<td></td>
<td>2B.3.1</td>
<td>2B.3.1</td>
<td>2B.3.1</td>
<td>2B.3.1</td>
</tr>
<tr>
<td></td>
<td>2B.3.3</td>
<td>2B.3.3</td>
<td>2B.3.3</td>
<td>2B.3.3</td>
</tr>
<tr>
<td></td>
<td>(5/2G.1)* (2C.1)*</td>
<td>(5/2G.1)* (2C.1)*</td>
<td>(5/2G.1)* (2C.1)*</td>
<td>(5/2G.1)* (2C.1)*</td>
</tr>
<tr>
<td></td>
<td>2B.2.1</td>
<td>2B.2.1</td>
<td>2B.2.1</td>
<td>2B.2.1</td>
</tr>
<tr>
<td></td>
<td>2B.5</td>
<td>2B.5</td>
<td>2B.5</td>
<td>2B.5</td>
</tr>
<tr>
<td></td>
<td>2B.N.2</td>
<td>2B.N.2</td>
<td>2B.N.2</td>
<td>2B.N.2</td>
</tr>
<tr>
<td></td>
<td>2B.2.2</td>
<td>2B.2.2</td>
<td>2B.2.2</td>
<td>2B.2.2</td>
</tr>
<tr>
<td>5C (V)</td>
<td>5C.1</td>
<td>5C.1</td>
<td>5C.1</td>
<td>5C.1</td>
</tr>
<tr>
<td></td>
<td>(5C2.1)</td>
<td>(5C2.1)</td>
<td>(5C2.1)</td>
<td>(5C2.1)</td>
</tr>
<tr>
<td></td>
<td>(5B1.1) (5E.1)</td>
<td>(5B1.1) (5E.1)</td>
<td>(5B1.1) (5E.1)</td>
<td>(5B1.1) (5E.1)</td>
</tr>
<tr>
<td></td>
<td>(5B.1) (5C1.1)</td>
<td>(5B.1) (5C1.1)</td>
<td>(5B.1) (5C1.1)</td>
<td>(5B.1) (5C1.1)</td>
</tr>
<tr>
<td>No. of variants resolved (D)</td>
<td>21/46 (0.9507)</td>
<td>29/46 (0.9768)</td>
<td>35/46 (0.9855)</td>
<td>39/46 (0.9923)</td>
</tr>
</tbody>
</table>

Horizontal lines indicate discrimination by the respective sets of binary targets. The discriminatory power increases as additional targets are added, so the number of horizontal lines increases from left to right. SCCmec variants in parentheses were named in this study. Asterisks indicate variants grouped with unrelated variants.
<table>
<thead>
<tr>
<th>SCCmec class</th>
<th>6 targets (ccr1, ccr5, Ts554 MLS, dcs, pUB110, and ccr2)</th>
<th>9 targets (previous 6 plus J1b, mec class C2, and IS256-dcs)</th>
<th>12 targets (previous 9 plus mec class A, pT181, and mec class A.3)</th>
<th>15 targets (previous 12 plus mec class B, J1a, and J1c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1B (I)</td>
<td>1B.1.1 (1B.2.1)</td>
<td>1B.1.1 (1B.2.1)</td>
<td>1B.1.1 (1B.2.1)</td>
<td>1B.1.1 (1B.2.1)</td>
</tr>
<tr>
<td></td>
<td>(1B.1.3)</td>
<td>(1B.1.3)</td>
<td>(1B.1.3)</td>
<td>(1B.1.3)</td>
</tr>
<tr>
<td></td>
<td>1B.1.2 (1B.1.4)</td>
<td>1B.1.2 (1B.1.4)</td>
<td>1B.1.2 (1B.1.4)</td>
<td>1B.1.2 (1B.1.4)</td>
</tr>
<tr>
<td>2A (II)</td>
<td>2A.1.1</td>
<td>2A.1.1</td>
<td>2A.1.1</td>
<td>2A.1.1</td>
</tr>
<tr>
<td></td>
<td>2A.3.1</td>
<td>2A.3.1</td>
<td>2A.3.1</td>
<td>2A.3.1</td>
</tr>
<tr>
<td></td>
<td>2A.3.3</td>
<td>2A.3.3</td>
<td>2A.3.3</td>
<td>2A.3.3</td>
</tr>
<tr>
<td></td>
<td>2A.3.4</td>
<td>2A.3.4</td>
<td>2A.3.4</td>
<td>2A.3.4</td>
</tr>
<tr>
<td></td>
<td>2A.3.5</td>
<td>2A.3.5</td>
<td>2A.3.5</td>
<td>2A.3.5</td>
</tr>
<tr>
<td></td>
<td>2A.1.2</td>
<td>2A.1.2</td>
<td>2A.1.2</td>
<td>2A.1.2</td>
</tr>
<tr>
<td></td>
<td>(2F.1.1)</td>
<td>(2F.1.1)</td>
<td>(2F.1.1)</td>
<td>(2F.1.1)</td>
</tr>
<tr>
<td></td>
<td>2A.3.2</td>
<td>2A.3.2</td>
<td>2A.3.2</td>
<td>2A.3.2</td>
</tr>
<tr>
<td></td>
<td>2B.N.2*</td>
<td>2B.N.2</td>
<td>2B.N.2</td>
<td>2B.N.2</td>
</tr>
<tr>
<td></td>
<td>2A.2</td>
<td>2A.2</td>
<td>2A.2</td>
<td>2A.2</td>
</tr>
<tr>
<td></td>
<td>(2A.1.3)</td>
<td>(2A.1.3)</td>
<td>(2A.1.3)</td>
<td>(2A.1.3)</td>
</tr>
<tr>
<td>3A (III)</td>
<td>3A.1.1 (3A.2.1) (3A.3.1)</td>
<td>3A.1.1 (3A.2.1) (3A.3.1)</td>
<td>3A.1.1 (3A.2.1) (3A.3.1)</td>
<td>3A.1.1 (3A.2.1) (3A.3.1)</td>
</tr>
<tr>
<td></td>
<td>(3A.1.2) (3A.1.1.1)</td>
<td>(3A.1.2) (3A.1.1.1)</td>
<td>(3A.1.2) (3A.1.1.1)</td>
<td>(3A.1.2) (3A.1.1.1)</td>
</tr>
<tr>
<td></td>
<td>3A.1.3</td>
<td>3A.1.3</td>
<td>3A.1.3</td>
<td>3A.1.3</td>
</tr>
<tr>
<td></td>
<td>3A.1.5</td>
<td>3A.1.5</td>
<td>3A.1.5</td>
<td>3A.1.5</td>
</tr>
<tr>
<td></td>
<td>(3A.1.4)</td>
<td>(3A.1.4)</td>
<td>(3A.1.4)</td>
<td>(3A.1.4)</td>
</tr>
<tr>
<td></td>
<td>(3A.1.6) (3A.4.1)</td>
<td>(3A.1.6) (3A.4.1)</td>
<td>(3A.1.6) (3A.4.1)</td>
<td>(3A.1.6) (3A.4.1)</td>
</tr>
<tr>
<td>2B (IV)</td>
<td>2B.1</td>
<td>2B.1</td>
<td>2B.1</td>
<td>2B.1</td>
</tr>
<tr>
<td></td>
<td>2B.3.2</td>
<td>2B.3.2</td>
<td>2B.3.2</td>
<td>2B.3.2</td>
</tr>
<tr>
<td></td>
<td>2B.2.1</td>
<td>2B.2.1</td>
<td>2B.2.1</td>
<td>2B.2.1</td>
</tr>
<tr>
<td></td>
<td>2B.5</td>
<td>2B.5</td>
<td>2B.5</td>
<td>2B.5</td>
</tr>
<tr>
<td></td>
<td>(2B.1.2)</td>
<td>(2B.1.2)</td>
<td>(2B.1.2)</td>
<td>(2B.1.2)</td>
</tr>
<tr>
<td></td>
<td>2B.3</td>
<td>2B.3</td>
<td>2B.3</td>
<td>2B.3</td>
</tr>
<tr>
<td></td>
<td>2B.3.3</td>
<td>2B.3.3</td>
<td>2B.3.3</td>
<td>2B.3.3</td>
</tr>
<tr>
<td></td>
<td>2B.2.2</td>
<td>2B.2.2</td>
<td>2B.2.2</td>
<td>2B.2.2</td>
</tr>
<tr>
<td></td>
<td>(2C.1)*</td>
<td>(2C.1)</td>
<td>(2C.1)</td>
<td>(2C.1)</td>
</tr>
<tr>
<td>5C (V)</td>
<td>5C.1 (5C2.1)</td>
<td>5C.1 (5C2.1)</td>
<td>5C.1 (5C2.1)</td>
<td>5C.1 (5C2.1)</td>
</tr>
<tr>
<td></td>
<td>(5B.1)</td>
<td>(5B.1)</td>
<td>(5B.1)</td>
<td>(5B.1)</td>
</tr>
<tr>
<td></td>
<td>(5B1.1) (5E.1)</td>
<td>(5B1.1) (5E.1)</td>
<td>(5B1.1) (5E.1)</td>
<td>(5B1.1) (5E.1)</td>
</tr>
<tr>
<td></td>
<td>(5/2G.1) (5/1C.1.1)</td>
<td>(5/2G.1) (5/1C.1.1)</td>
<td>(5/2G.1) (5/1C.1.1)</td>
<td>(5/2G.1) (5/1C.1.1)</td>
</tr>
<tr>
<td>Other</td>
<td>4B</td>
<td>4B</td>
<td>4B</td>
<td>4B</td>
</tr>
</tbody>
</table>

No. of variants resolved ($D$) | 18/46 (0.9391) | 28/46 (0.9729) | 34/46 (0.9836) | 37/46 (0.9903) |

* Horizontal lines indicate discrimination by the respective sets of binary targets. SCCmec variants in parentheses were named in this study. Asterisks indicate variants grouped with unrelated variants.
TABLE 5. Resolution-optimized binary target sets derived from an input file reflecting SCCmeC variant abundance.

<table>
<thead>
<tr>
<th>SCCmeC</th>
<th>7 targets (ccr1, dcs, Tn554 MLS, ccr2, mec class A, J1b, and pT181)</th>
<th>9 targets (previous 7 plus pUB110 and mec class B)</th>
<th>11 targets (previous 9 plus mec class C2 and J1a)</th>
<th>13 targets (previous 11 plus pls and IS256-dcs)</th>
<th>Abundant variant</th>
</tr>
</thead>
<tbody>
<tr>
<td>1B (I)</td>
<td>1B.1.1</td>
<td>1B.1.1</td>
<td>1B.1.1</td>
<td>1B.1.1</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>(1B.1.3)</td>
<td>(1B.1.3)</td>
<td>(1B.1.3)</td>
<td>(1B.1.3)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1B.2.1)</td>
<td>(1B.2.1)</td>
<td>(1B.2.1)</td>
<td>(1B.2.1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1B.1.2</td>
<td>1B.1.2</td>
<td>1B.1.2</td>
<td>1B.1.2</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>(1B.1.4)</td>
<td>(1B.1.4)</td>
<td>(1B.1.4)</td>
<td>(1B.1.4)</td>
<td></td>
</tr>
<tr>
<td>2A (II)</td>
<td>2A.1.1</td>
<td>2A.1.1</td>
<td>2A.1.1</td>
<td>2A.1.1</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>2A.1.2</td>
<td>2A.1.2</td>
<td>2A.1.2</td>
<td>2A.1.2</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>2.2</td>
<td>2.2</td>
<td>2.2</td>
<td>2.2</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>(2A.1.3)</td>
<td>(2A.1.3)</td>
<td>(2A.1.3)</td>
<td>(2A.1.3)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2A.3.2</td>
<td>2A.3.2</td>
<td>2A.3.2</td>
<td>2A.3.2</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>(2F.1.1)</td>
<td>(2F.1.1)</td>
<td>(2F.1.1)</td>
<td>(2F.1.1)</td>
<td></td>
</tr>
<tr>
<td>3A (III)</td>
<td>3A.1.1</td>
<td>3A.1.1</td>
<td>3A.1.1</td>
<td>3A.1.1</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>(3A.2.1)</td>
<td>(3A.2.1)</td>
<td>(3A.2.1)</td>
<td>(3A.2.1)</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>(3A.3.1)</td>
<td>(3A.3.1)</td>
<td>(3A.3.1)</td>
<td>(3A.3.1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3A.1.1)</td>
<td>(3A.1.1)</td>
<td>(3A.1.1)</td>
<td>(3A.1.1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3A.1.2</td>
<td>3A.1.2</td>
<td>3A.1.2</td>
<td>3A.1.2</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>(3A.1.4)</td>
<td>(3A.1.4)</td>
<td>(3A.1.4)</td>
<td>(3A.1.4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3A.1.3</td>
<td>3A.1.3</td>
<td>3A.1.3</td>
<td>3A.1.3</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>(3A.1.5)</td>
<td>(3A.1.5)</td>
<td>(3A.1.5)</td>
<td>(3A.1.5)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3A.1.6)</td>
<td>(3A.1.6)</td>
<td>(3A.1.6)</td>
<td>(3A.1.6)</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>(3A.4.1)</td>
<td>(3A.4.1)</td>
<td>(3A.4.1)</td>
<td>(3A.4.1)</td>
<td></td>
</tr>
<tr>
<td>2B (IV)</td>
<td>2B.1</td>
<td>2B.1</td>
<td>2B.1</td>
<td>2B.1</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>(2B.1.2)</td>
<td>(2B.1.2)</td>
<td>(2B.1.2)</td>
<td>(2B.1.2)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2B.3.2</td>
<td>2B.3.2</td>
<td>2B.3.2</td>
<td>2B.3.2</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>2B.4</td>
<td>2B.4</td>
<td>2B.4</td>
<td>2B.4</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>(2B.6)</td>
<td>(2B.6)</td>
<td>(2B.6)</td>
<td>(2B.6)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2B.N.2</td>
<td>2B.N.2</td>
<td>2B.N.2</td>
<td>2B.N.2</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>2B.2.1</td>
<td>2B.2.1</td>
<td>2B.2.1</td>
<td>2B.2.1</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>2B.5</td>
<td>2B.5</td>
<td>2B.5</td>
<td>2B.5</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>2B.3.1</td>
<td>2B.3.1</td>
<td>2B.3.1</td>
<td>2B.3.1</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>2B.3.3</td>
<td>2B.3.3</td>
<td>2B.3.3</td>
<td>2B.3.3</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>(5/2G.1)*</td>
<td>(5/2G.1)*</td>
<td>(5/2G.1)</td>
<td>(5/2G.1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2C.1)*</td>
<td>(2C.1)*</td>
<td>(2C.1)</td>
<td>(2C.1)</td>
<td></td>
</tr>
<tr>
<td>5C (V)</td>
<td>5C.1</td>
<td>5C.1</td>
<td>5C.1</td>
<td>5C.1</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>(5C2.1)</td>
<td>(5C2.1)</td>
<td>(5C2.1)</td>
<td>(5C2.1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(5B.1.1)</td>
<td>(5B.1.1)</td>
<td>(5B.1.1)</td>
<td>(5B.1.1)</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>(5E.1)</td>
<td>(5E.1)</td>
<td>(5E.1)</td>
<td>(5E.1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(5B.1)</td>
<td>(5B.1)</td>
<td>(5B.1)</td>
<td>(5B.1)</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>(5/1C.1.1)</td>
<td>(5/1C.1.1)</td>
<td>(5/1C.1.1)</td>
<td>(5/1C.1.1)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>4B*</td>
<td>4B</td>
<td>4B</td>
<td>4B</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>Truncated*</td>
<td>Truncated</td>
<td>Truncated</td>
<td>Truncated</td>
<td></td>
</tr>
</tbody>
</table>

No. of variants resolved: 19/46 (0.8976) 28/46 (0.9012) 31/46 (0.9034) 36/46 (0.9047)*

* Horizontal lines indicate discrimination by the respective sets of binary targets. SCCmeC variants in parentheses were named in this study. Asterisks indicate variants grouped with unrelated SCCmeC variants.

b Due to the input data set containing repeats of identical sequences, a D value close to 1 cannot be obtained.
The data used as input into Minimum SNPs was in effect an variants with corresponding mecA (39). 612 (39), 675 (17), and 737 (39). A new SCC data set was created containing only variants, a new SCC mecA analysis as additional data points. Because no precise mecA sequence data could be assigned to many of the SCC mecA elements. In total, 25 sequences were retrieved from the NCBI database, and in addition, partial mecA sequence analysis was undertaken on 19 selected isolates from our collection of previously characterized SCCmec elements (41). From this set of 44 partial mecA sequences, eight SNPs in total were identified, three of which were novel (Table 6).

To determine whether mecA SNPs add resolution to the binary genes, they were included in a Minimum SNPs analysis as additional data points. Because no precise mecA sequence data could be assigned to many of the SCCmec variants, a new SCCmec data set was created containing only variants with corresponding mecA sequences. In this case, the data used as input into Minimum SNPs was in effect an alignment that contained a region of pseudo-DNA sequence derived from binary gene variation and a region of actual DNA sequence derived from mecA sequences. This data set consisted of 18 of the original 46 SCCmec variants (Table 6). As different mecA SNP profiles were found within identical SCCmec variants, the 34 binary markers in combination with the eight SNPs defined 24 variants. These expanded mixed SNP/binary target profiles were analyzed using “Minimum SNPs,” which revealed that a D of 1.0 was achieved with 9 binary targets and five mecA SNPs. Overall, of these five SNPs, only SNP 737 could be considered a possible replacement for the binary targets, as it was selected at the third position in the Minimum SNPs output. The remaining SNPs were selected at positions 9 and 12 to 14, which demonstrates minimal D-value contribution. The entire marker set is ccr2, Tn554 MLS, mecA737, mec class A, dcs, pT181, ccrC, Jyb, mecA438, ccrC-VT, Tn4001, mecA75, mecA415, and mecA448.

**DISCUSSION**

It is now generally accepted that a powerful strategy for bacterial genotyping is to interrogate the genome backbone plus one or more hypervariable regions. This generalized approach has been termed “phylogenetic hierarchical assays using nucleic acids” (20). The practice of identifying MRSA clones by a combination of MLST and SCCmec type is an example of this. The widespread adoption of this MRSA genotyping method has proven an impetus to the accumulation of considerable information regarding SCCmec diversity and the development of several specific SCCmec subtyping schemes.

**SCCmec** has been classified into five major “types,” composed of mec classes and ccr gene identity, and several SCCmec genotyping methods classify the element to this level only (10, 45). The results of our comprehensive search of the literature and the databases emphasized the high diversity of SCCmec

### TABLE 6. SCCmec variants and associated mecA SNP profiles

<table>
<thead>
<tr>
<th>SCCmec class</th>
<th>SCCmec variant</th>
<th>Nucleic acid at mecA SNP position</th>
</tr>
</thead>
<tbody>
<tr>
<td>1B (I)</td>
<td>1B.1.1</td>
<td>A/C G T G T T G</td>
</tr>
<tr>
<td>2A (II)</td>
<td>2A.1.1</td>
<td>A/C G T G T T G</td>
</tr>
<tr>
<td></td>
<td>2A.3.5</td>
<td>A/C G T G T T G</td>
</tr>
<tr>
<td>2B (IV)</td>
<td>2B.1</td>
<td>A/C G T G T T G T G</td>
</tr>
<tr>
<td></td>
<td>2B.2.1</td>
<td>A/C G T G T T G T G</td>
</tr>
<tr>
<td></td>
<td>2B.3.1</td>
<td>A/C G T G T T G T G</td>
</tr>
<tr>
<td></td>
<td>2B.3.2</td>
<td>A/C G T G T T G T A</td>
</tr>
<tr>
<td></td>
<td>2B.3.3</td>
<td>A/C G T G T G</td>
</tr>
<tr>
<td>3A (III)</td>
<td>3A.1.1</td>
<td>A/C G T/A G/A G T A</td>
</tr>
<tr>
<td></td>
<td>3A.1.2</td>
<td>A A G T G T G</td>
</tr>
<tr>
<td></td>
<td>3A.1.3</td>
<td>A/C G T G T G</td>
</tr>
<tr>
<td></td>
<td>3A.1.4</td>
<td>A/C G/A G T G</td>
</tr>
<tr>
<td></td>
<td>3A.2.1</td>
<td>A/C G T G T G</td>
</tr>
<tr>
<td>5C (V)</td>
<td>5/C.1.1</td>
<td>A/C G T G T G</td>
</tr>
<tr>
<td></td>
<td>5/C.1.2</td>
<td>A/C G T G T G</td>
</tr>
<tr>
<td></td>
<td>5/C.1.3</td>
<td>A/C G T G T A</td>
</tr>
<tr>
<td></td>
<td>5/C.2.1</td>
<td>A/C G T G T A</td>
</tr>
<tr>
<td></td>
<td>5/C.2.2</td>
<td>A/C G T G T A</td>
</tr>
</tbody>
</table>

* Asterisks indicate nonsynonymous codon changes.

---

mecA SNPs can increase resolving power. It has previously been reported that the MRSA mecA gene contains several SNPs (37, 39, 44). These are potential genotyping markers that could be used either as replacements for the binary targets selected or to define more SCCmec subtypes. The literature and sequence databases were searched for mecA sequences from characterized SCCmec elements. In total, 25 sequences were retrieved from the NCBI database, and in addition, partial mecA sequence analysis was undertaken on 19 selected isolates from our collection of previously characterized SCCmec elements (41). From this set of 44 partial mecA sequences, eight SNPs in total were identified, three of which were novel (Table 6).

To determine whether mecA SNPs add resolution to the binary genes, they were included in a Minimum SNPs analysis as additional data points. Because no precise mecA sequence data could be assigned to many of the SCCmec variants, a new SCCmec data set was created containing only variants with corresponding mecA sequences. In this case, the data used as input into Minimum SNPs was in effect an alignment that contained a region of pseudo-DNA sequence derived from binary gene variation and a region of actual DNA sequence derived from mecA sequences. This data set consisted of 18 of the original 46 SCCmec variants (Table 6). As different mecA SNP profiles were found within identical SCCmec variants, the 34 binary markers in combination with the eight SNPs defined 24 variants. These expanded mixed SNP/binary target profiles were analyzed using “Minimum SNPs,” which revealed that a D of 1.0 was achieved with 9 binary targets and five mecA SNPs. Overall, of these five SNPs, only SNP 737 could be considered a possible replacement for the binary targets, as it was selected at the third position in the Minimum SNPs output. The remaining SNPs were selected at positions 9 and 12 to 14, which demonstrates minimal D-value contribution. The entire marker set is ccr2, Tn554 MLS, mecA737, mec class A, dcs, pT181, ccrC, Jyb, mecA438, ccrC-VT, Tn4001, mecA75, mecA415, and mecA448.

**DISCUSSION**

It is now generally accepted that a powerful strategy for bacterial genotyping is to interrogate the genome backbone plus one or more hypervariable regions. This generalized approach has been termed “phylogenetic hierarchical assays using nucleic acids” (20). The practice of identifying MRSA clones by a combination of MLST and SCCmec type is an example of this. The widespread adoption of this MRSA genotyping method has proven an impetus to the accumulation of considerable information regarding SCCmec diversity and the development of several specific SCCmec subtyping schemes.

**SCCmec** has been classified into five major “types,” composed of mec classes and ccr gene identity, and several SCCmec genotyping methods classify the element to this level only (10, 45). The results of our comprehensive search of the literature and the databases emphasized the high diversity of SCCmec.
and revealed that much of the diversity is invisible to previously published SCCmec typing methods. In addition to the initial characterization of types 1B, 2A, and 3A (I, II, and III) (15), the community-acquired type 2B (IV) (25), and the most recent type 5C (V) (17), more variable and unusual variants have been identified. Of particular interest are variants carrying multiple copies of ccr and instances of new combinations of mec and ccr classes (2, 28).

Our approach to identifying sets of binary markers for genotyping SCCmec differs greatly from previously published methods (21). Genotyping approaches reported to date yield approximately one mec class or ccr gene per marker interrogated. This is needlessly inefficient as the extensive recombination in SCCmec means that resolving power can potentially increase logarithmically rather than arithmetically as more targets are interrogated. Accordingly, rather than identifying binary markers diagnostic for particular classes/types/subtypes, we used a computerized approach that identifies sets of markers that maximize D. What this algorithm does is attempt to identify a marker that splits the known variants into two equal halves and then attempts to find a single marker that splits each of the groups defined by marker 1 into two equal halves, and so forth. In effect, it is a search for markers that are maximally unlinked. This approach proved to be valuable. It provided sets of markers with greater resolving power than equivalently sized sets of markers identified by the traditional approach. Also, the analyses clearly defined a range of options regarding the numbers of markers interrogated and the resolution obtained. These results bore out the prediction that resolution could increase exponentially with the number of targets interrogated; when the numbers of targets was graphed against log(1 − D), the points formed a straight line (data not shown).

The markers identified have obvious potential to inform the design of specific SCCmec genotyping methods around the multiplexing capacity of the technology to be used and/or the resolution required. One intriguing result was the strong consistency between the SCCmec classes and the groups defined by small numbers of resolution-optimized marker sets, even when there were no constraints on marker selection. The fact that maximally unlinked markers define these groups supports the notion that the SCCmec classes indeed represent distinct phylogenetic clusters of this element. The strategy to increase the consistency between genotypes and the relationships between the SCCmec variants by nucleating the marker set with ccr1 and ccr5 proved successful, as was the analysis to identify markers especially efficient at discriminating the abundant SCCmec variants. Overall, our approach to marker selection proved effective and flexible. The marker sets identified provide a wide choice of well-understood options for the design of SCCmec genotyping procedures. This general approach could be applied to any genome region displaying a high degree of binary variability, e.g., the loci encoding the enzymes that synthesize complex antigenically active cell wall-associated polysaccharides.

Rapid interpretation of the results of a genotyping procedure based on the marker sets described here is potentially problematic. It can be done by manually correlating the results with the information in Table 2, but this is quite laborious. However, it can also be done more rapidly using Minimum SNPs, which is able to run in reverse and return all sequences in an alignment that correspond to a user-defined SNP profile.

The utility of including mecA SNPs in the marker set was explored, and it was found that such SNPs define additional SCCmec variants. However, the SNPs are not effective substitutes for binary markers. Of interest were the observations that identical SNP allelic combinations were found in different SCCmec variants while, conversely, different mecA alleles were found in identical SCCmec variants. For example, the first four SCCmec classes (1B, 2A, 3A, and 2B) each carry the most abundant mecA sequence, while seven of the nine mecA sequences are found in type 3A. This suggests that some SCCmec variants are very much older than others or that there has been recombination between different SCCmec variants. Other observations of interest include variant 2B.1 (IVa) exclusively carrying the most numerous mecA sequence, which is different from the mecA sequence in the closely related variant 2B.3.1 (IVc), and the complete linkage found between ccr5 and a particular mecA sequence (Table 6). It was also observed that two SNP profiles were primarily associated with mec class B, while seven SNP profiles were associated with mec classes A and C. It was concluded that mecA SNPs constitute a possible means of increasing resolving power, with inclusion of the SNP data increasing the number of genotypes. However, the mecA SNPs have only a limited potential use as replacements for the binary markers, especially considering that it is more techni-

ACKNOWLEDGMENTS

We thank E. P. Price for scientific advice.

This work was supported by the Cooperative Research Centres Program of the Australian Federal Government.

REFERENCES


