All ccr genes were identified using blastp against all protein sequences and extracted from the cds fasta files. Only complete coding sequences were included.

Nucleotide sequences from a total of 18184 ccr genes were extracted and combined with 135 IWG reference sequences. These 19319 sequences were represented by 1363 unique sequences which were aligned using MUSCLE.

Definition of ccr types

A complete-linkage hierarchical clustering was performed based on pairwise dissimilarity between sequences in the MUSCLE alignment. Using a threshold of 50% similarity 11 clusters were defined including clusters for the three variants known to appear in the SCCmec element (ccrA, ccrB, ccrC) and eight unknown variants which were found at much lower frequencies. The eight unknown variants made up only 126 of 19184 identified genes or 39 out of 1363 unique sequnces.  
  
Based on this clustering all unique sequences from ccrA, ccrB and ccrC were extracted and aligned individually using MUSCLE for analysis of allotypes. For each of the three a complete-linkage hierarchical clustering was performed based on pairwise similarity between sequences, and allotypes were defined based on a 78% similarity threshold.