

**Analysis of typing methods for epidemiological surveillance of both methicillin-**

**resistant and susceptible *Staphylococcus aureus***

Nuno A. Faria<sup>1,§</sup>, João A. Carrico<sup>2,§,\*</sup>, Duarte C. Oliveira<sup>1</sup>, Mário Ramirez<sup>3</sup> and Hermínia de Lencastre<sup>1,4</sup>

6

<sup>1</sup> Laboratório de Genética Molecular, Instituto de Tecnologia Química e Biológica,

Oeiras, Portugal. <sup>2</sup> Knowledge Discovery in Bioinformatics Group (KDBIO), Instituto de Engenharia de Sistemas e Computadores: Investigação e Desenvolvimento

(INESC-ID), Lisboa, Portugal. <sup>3</sup>Instituto de Medicina Molecular e Instituto de

Microbiologia, Faculdade de Medicina de Lisboa, Lisboa, Portugal <sup>4</sup>Laboratory of

Microbiology, The Rockefeller University, New York, USA

Running title: Quantitative congruence of typing methods in *S. aureus*

Key-words: *Staphylococcus aureus*, *spa* typing, PFGE, MLST, Typing methods congruence, Adjusted Rand and Wallace coefficients

18

§ N. A. F. and J. A. C. contributed equally to this work.

20

\* Corresponding author. Mailing address: Instituto de Engenharia de Sistemas e

Computadores: Investigação e Desenvolvimento (INESC-ID), Rua Alves Redol 9, Apartado 13069, 1000-029 Lisboa, Portugal. Phone: (+351) 213 100 272. Fax: (+351)

213 145 843. E-mail: jcarrico@kdbio.inesc-id.pt

**ABSTRACT**

2

Sequence based methods for typing *Staphylococcus aureus*, such as Multilocus  
4 Sequence Typing (MLST) and *spa* typing, have increased inter-laboratory  
reproducibility, portability, and speed in obtaining results, but Pulsed-Field Gel  
6 Electrophoresis (PFGE), remains the method of choice in many laboratories due to the  
extensive experience with this methodology and the large body of data accumulated  
8 using this technique. Comparisons between typing methods have been  
overwhelmingly based on a qualitative assessment of the overall agreement of results  
10 and the relative discriminatory indexes. In this study, we quantitatively assess of the  
congruence of the major typing methods for *S. aureus*, using a diverse collection of  
12 198 *S. aureus* strains previously characterized by PFGE, *spa* typing, MLST, and, in  
the case of methicillin-resistant *S. aureus* (MRSA), *SCCmec* typing, in order to  
14 establish the quantitative congruence between the typing methods.

The results of most typing methods agree in that MRSA and methicillin-susceptible *S.*  
16 *aureus* (MSSA) differ in terms of diversity of genetic backgrounds, with MSSA being  
more diverse. Our results show that *spa* typing has a very good predictive power over  
18 the clonal relationships defined by eBURST while PFGE is less accurate in that  
purpose, providing nevertheless better typeability and discriminatory power. The  
20 combination of PFGE and *spa* typing provided even better results. Based on these  
observations we suggest the use of the conjugation of *spa* typing and PFGE typing for  
22 epidemiological surveillance studies, since this combination provides the ability to  
infer long-term relationships, while maintaining the discriminatory power and  
24 typeability needed in short-term studies.

## INTRODUCTION

2

*Staphylococcus aureus* is a leading human pathogen and remains a major cause of infections worldwide (16, 22, 38), causing a high rate of hospital-acquired infections in several countries (2, 10). Recently, the epidemiology of *S. aureus*, in particular for methicillin-resistant *Staphylococcus aureus* (MRSA), has changed with the emergence of community-acquired MRSA (CA-MRSA), as reported by several studies (10, 13, 16, 22, 38). The epidemiology of infectious diseases relies on typing methods as tools for the characterization and discrimination of isolates based on either their genotypic or phenotypic characteristics, which may be used to establish clonal relationships between strains and trace the geographic dissemination of bacterial clones. Nowadays, the classification of isolates is mostly based on molecular methods, which usually provide better discriminatory power than phenotypic methods. Pulsed-field gel electrophoresis (PFGE), after *Sma*I digestion of total bacterial DNA (33), is still regarded by many authors as the gold standard used for benchmarking new typing methods, although it was originally proposed for outbreak investigation (37). Recently, due to the availability and affordability of DNA sequence technology, several sequenced-based typing methods have been developed and are now widely used, such as Multilocus Sequence Typing (MLST) (23) and *spa* typing (34), which are the most frequently used for *S. aureus*. DNA sequence based typing methods generate unambiguous and portable data, amenable to the creation of central databases, which enable the comparison of local data with data from previous studies in different geographical locations.

Apart from factors such as discriminatory indexes, reproducibility and standardization, typing techniques differ dramatically in associated costs (32, 39), which may restrict the choice of typing methods due to budget limitations. For

instance, MLST, which relies on the sequence of the internal fragments of seven  
2 housekeeping genes, is much more expensive than *spa* typing, which is based on the  
sequence of an internal fragment of a single gene. Although PFGE is a labor-intensive  
4 and may be a more economical alternative, it has several drawbacks: it requires  
unique technical skills, has an high set-up cost and the inter-laboratory comparison of  
6 results is not straightforward.

According to a proposal by Enright and colleagues (12) which was accepted by a  
8 subcommittee of the International Union of Microbiology Societies in Tokyo, 2002,  
MRSA clones are named according to their MLST and *SCCmec* types (e.g. clone  
10 ST5-MRSA-II). However, the amount of sequencing required for MLST typing and  
the increasing number of primers need to define *SCCmec* types (25, 29) as new types  
12 and variants are found, hampers the use of this combination of methods for clonal  
characterization of large collections, mainly due to cost related reasons. Other  
14 combinations of methods that provide a similarly fine resolution of the accepted  
clonal group definition can be explored.

16 In line with this rationale, SeqNet ([www.seqnet.org](http://www.seqnet.org)), the European Network of  
Laboratories for Sequence Based Typing of Microbial Pathogens, has proposed *spa*  
18 typing as the sequence-based method of choice to determine the genetic relatedness of  
*S. aureus* isolates. An online database is now available featuring automated curation  
20 of submitted sequences and assignment to *spa* types (3).

Molecular epidemiology studies of clinical microorganisms often rely on the  
22 application of typing methods that produce different type assignments. From the  
comparison and analysis of these assignments, a classification of the isolate in terms  
24 of clonal type or lineage is generated. What is more, since different laboratories may  
use different combinations of methods and over time implement new typing schemes,

the definition of clones is neither universal nor static. Since different typing schemes  
2 analyze different phenotypic and genotypic properties of bacteria, if a congruent result  
is obtained between different methods, it suggests that a phylogenetic signal is being  
4 recovered by both methods, allowing a high confidence in the assignment of clonal  
types. Therefore, the quantification of congruence between different methods, with an  
6 assessment of the confidence for predicting an unknown character from another  
typing method, can be a useful tool in epidemiology (6), enabling an informed choice  
8 between typing methods for a given study taking into account the degree of  
discrimination needed and the available budget.

10 At the Molecular Genetics Laboratory, ITQB, Oeiras, Portugal, different *S. aureus*  
strains from different worldwide locations have been collected since the late 80's.  
12 These strains have been analyzed by different typing methods over time. Pulsed-field  
gel electrophoresis has been the standard typing technique during this period due to its  
14 high discriminatory power and relative low cost per isolate. However, with the  
introduction of *spa* typing and MLST typing schemes, the characterization of *S.*  
16 *aureus* isolates involves now a combination of different techniques [including the  
SCC*mec* type for the characterization for MRSA strains (20, 27, 28)]. We have  
18 previously extended the work of Robinson et al. (31), by proposing the use of  
measures of clustering concordance – Adjusted Rand and Wallace coefficients – to  
20 compare type assignments, allowing a quantitative approach for exploring the  
concordance between typing methods (6). In this study, we have implemented the use  
22 of that methodological framework to a set of 198 *S. aureus* strains, which were  
previously characterized by PFGE, *spa* typing, MLST, and SCC*mec* typing (for  
24 MRSA) in order to quantify the congruence between methods and the discriminatory  
power of each method and combination of methods.

## MATERIAL and METHODS

### 2 Strain collection

A collection of 198 *S. aureus* strains (116 MRSA and 82 MSSA) was included in this study – see Tables 1S and 2S of the online supplementary material. These strains were chosen from a large (> 5,000 isolates) international collection of *S. aureus* isolated in several parts of the world, mainly in hospitals in Southern and Eastern Europe, Latin America and USA since the late 1980's, and included representatives of early isolates from the United Kingdom and Denmark isolated between 1957 and 1972. Overall, among the selected 198 *S. aureus* strains, 19 countries are represented as follows: Argentina (n=9; 4.5%), Brazil (n=1; 0.5%), Cabo Verde (n=12; 6%), Chile (n=1; 0.5%), Colombia (n=2; 1%), Czech Republic (n=3; 1.5%), Denmark (n=46; 23%), Egypt (n=10; 5%), Greece (n=5; 2.5%), Hungary (n=18; 9%), Italy (n=2; 1%), Japan (n=7; 3.5%), Mexico (n=4; 2%), Poland (n=4; 2%), Portugal (n=63; 32%), Spain (n=1; 0.5%), Taiwan (n=5; 2.5%), UK (n=8; 4.0%), USA (n=6; 3%). The criteria used in the strain selection process excluded duplicated outbreak strains, in order to minimize sampling bias, and tried to maximize the diversity represented in the analyzed collection relative to that present in the >5,000 isolates screened. All strains included in this study were characterized by PFGE (7), MLST (9, 11) and *spa* typing (21, 34). MRSA strains were also characterized by *SCCmec* typing (27, 28).

20

### PFGE data analysis

22 PFGE patterns were analyzed in Bionumerics version 4.61 from Applied Maths (Sint-Martens-Latem, Belgium). Gel photos were acquired using Polaroid black & white  
24 instant pack film 667 and the negative was digitalized as an 8-bit grayscale TIFF images to use in the above software. Each image was then analyzed using the

resources of the Bionumerics™ software. A spectral analysis was performed for each  
2 image in order to obtain the background subtraction (Background scale) and the cutoff  
threshold for least squares filtering (wiener cutoff scale). After this process, PFGE  
4 runs were normalized, intra- and inter-gels, using *S. aureus* strain NCTC8325 loaded  
in each gel as reference. Band assignments were manually curated after automatic  
6 band detection for all gel images; bands ranging from 10 kb to ~674 kb were  
considered for analysis in this study. For band pattern comparisons within and  
8 between different gels, the following settings were used: optimization of 0.5% and  
position tolerance of 1.25%. PFGE types and subtypes were defined respectively by  
10 groups formed at an 80% and 95% Dice similarity cut-off on a dendrogram  
constructed by Unweighted Pair Group Method with Arithmetic mean (UPGMA). The  
12 groups defined at these thresholds were previously shown to approximate those  
defined using Tenover's criteria for visual PFGE type definition (5, 24, 37).

## 14 DNA sequence data analysis

### 16 *spa* typing

Ridom StaphType software, version 1.4 (Ridom GmbH, Würzburg, Germany) was  
18 used for *spa* type analyzes. The new *spa* type assignments were provided  
automatically through the Ridom SpaServer (<http://spa.ridom.de/index.shtml>). BURP  
20 algorithm (Ridom StaphType software) was used to calculate *spa* clonal complexes  
(*spa*CC) with the default parameters for the group/cluster definition: “exclude *spa*  
22 types that are shorter than 5 repeats” and “*spa* types are clustered if cost is less or  
equal to 6” (30).

24

### **Multi Locus Sequence Typing (MLST)**

2 MLST sequence types (ST) were assigned through the MLST database  
(<http://www.mlst.net>). e-BURST algorithm was used to assign MLST clonal  
4 complexes (CC) (<http://www.e-burst.net>).

### **SCC*mec* typing**

6 SCC*mec* types were determined by a multiplex PCR strategy, which generated a  
8 specific amplification pattern for each SCC*mec* structural type (28). SCC*mec* type  
assignments were confirmed by *ccrAB* typing, as previously described by Okuma et  
10 al. (27).

### **Diversity indices**

12 Hunter and Gaston (18) proposed the use of Simpson's index of diversity (35) to  
14 measure the discriminatory ability of typing systems. This index indicates the  
probability of two strains sampled randomly from a population belonging to two  
16 different types. Grundmann et al. (17) introduced a method for determining  
confidence intervals (CIs) of Simpson's index, thereby improving the objective  
18 assessment of the discriminatory power of typing techniques.

### **Comparison of typing methods**

20 A framework for assessing the quantitative correspondence between typing methods  
22 was proposed by Carriço et al. (6). It is based on two coefficients developed to  
compare two ways to partition a given dataset: Adjusted Rand (AR) (17) and Wallace  
24 (W) (40).

Adjusted Rand's coefficient corrects Rand's coefficient, commonly used for  
2 quantifying the congruence for typing methods (31) for the presence of chance  
agreement; i.e., that the two sets of results match by chance alone. The use of Rand  
4 coefficient, which is formally equivalent to the concordance measure proposed by  
Robison et al. (31), leads to overestimation of the agreement between two typing  
6 methods and should be avoided. For a more detailed discussion on the use of these  
indices in the context of microbial typing see Carriço et al (6). Wallace coefficient can  
8 provide an even finer comparison between two methods, since the value indicates the  
probability that two strains classified as the same type by one method, are also  
10 classified in the same type by the other method, or vice-versa. A high value of  
Wallace's coefficient (it can assume any value from 0 to 1) indicates that partitions  
12 defined by a given method could have been predicted from the results of another  
method, suggesting that the use of both methodologies could be redundant. The  
14 combined use of the two coefficients can provide further information: two methods  
can have a low global agreement (assessed by AR) and yet, one of those methods can  
16 predict very well the results of another, which can be assessed by Wallace coefficient.  
To facilitate the use of these coefficients, a web page with the Bionumerics™ scripts  
18 used in this study was made available at [www.comparingpartitions.info](http://www.comparingpartitions.info).

## RESULTS

2 Our strain collection can be divided into two distinct groups – MRSA and MSSA –  
that differ on the “broad-spectrum” resistance to  $\beta$ -lactams as a consequence of the  
4 acquisition of the *SCCmec* element. The following analysis will always be applied to  
the two groups and to the overall collection, since the MRSA and MSSA populations  
6 are expected to differ. For each typing method two levels of discrimination were  
considered: one corresponding to the direct result of the method itself [MLST  
8 sequence type (ST), *spa* type, PFGE subtype and, for MRSA, *SCCmec* type] and  
another resulting of the application of an algorithm that generates groups of related  
10 isolates from this primary data (eBURST for MLST data, BURP for *spa* and the 80%  
cutoff for defining PFGE type). In order to explore if there was an improvement in  
12 discriminatory power when compared with classifications obtained with individual  
typing method results, classifications based on the combination of typing methods  
14 were also considered. One such conjugation of typing methods evaluated was the  
currently accepted ST/*SCCmec* combination for the definition of MRSA lineages. We  
16 also considered the conjugation of PFGE type and subtype with *spa* type, the two  
methods recently shown to be suitable for long range epidemiological studies (15),  
18 and in the MRSA subset, PFGE type and subtype together with *SCCmec* type. The  
results of the number of groups found for each method or conjugation of method are  
20 presented in Table 1.

### 22 Pulsed-field gel electrophoresis

The type and subtype definition in PFGE is commonly obtained by determining a cut-  
24 off value on a dendrogram of distances constructed with Dice coefficient and  
Unweighted Pair Group Mean Average (UPGMA) method. Several cut-off values and

parameters for dendrogram construction have been proposed (5, 24, 26, 36). Ideally, the determination of the cut-off level should be supported by other epidemiologically relevant data referring to the strains. To evaluate the groups defined by PFGE two cut-off levels were considered: 80% to define PFGE types and 95% to define PFGE subtypes, which were shown to be adequate for the analysis of large and diverse collections of strains (24). Considering the entire collection of 198 strains, 56 PFGE types (80% cut-off) and 153 PFGE subtypes (95% cut-off) were detected, corresponding to which can be split in 32 types and 92 subtypes if only MRSA are considered and 30 types and 63 subtypes for MSSA.

### ***spa* typing**

*spa* types were assigned using Ridom StaphType software and BURP algorithm was run in all the datasets using the settings previously described in Materials and Methods. For the 198 strains, 98 distinct *spa* types were found. For the MRSA and MSSA subsets, 51 and 55 types were determined, respectively. Some *spa* types were shared by both subsets: t002, t008, t012, t018, t021, t024, t127, and t148.

When applying the BURP algorithm for the inference of clusters of related isolates, 192 strains (96.7%) were distributed in 29 groups, where 12 of those groups were singletons (groups represented by a single type), and six strains (3 MRSA and 3 MSSA strains) were excluded due to the rules described in the Material and Methods section.

### **MLST**

Sequence types (ST) were assigned through the MLST database (<http://www.mlst.net>) and MLST clonal complexes (CC) were calculated using the e-BURST algorithm.

Overall, the 198 strains were distributed among 61 ST's, belonging to 21 CC's. Thirty-two ST's, belonging to 11 CC's, were found among the 116 MRSA strains. MSSA strains were more diverse since, among the 82 MSSA strains, 35 ST's belonging to 17 CC's were detected.

## 6 Comparing Discriminatory power for different methods

Simpson's index (SID) of diversity provides a measure of the discriminatory power of the different typing methods or conjugation of typing methods within a confidence interval. If the confidence intervals of any two methods overlap, one cannot exclude the hypothesis that they have similar discriminatory power at 95% confidence level. The SID values obtained for our datasets are presented in Table 1. When analyzing either the entire collection or only the MRSA subset, PFGE has the highest SID, whenever PFGE subtype assignments are compared with *spa* types or ST's or PFGE types or subtypes are compared to eBURST or BURP. However when comparing the PFGE type assignment with the *spa* types, as is usually done in literature, similar levels of discriminatory power are found (15). This observation is valid for the entire dataset, or any of the MRSA or MSSA subsets. As expected for MRSA strains, SCC*mec* typing was the least discriminatory technique due to the restricted number of variants generated by this method. Nevertheless, if SCC*mec* types are conjugated with ST's, the level of discrimination is similar to that of *spa* typing or conjugation of PFGE type and SCC*mec* type. Among all conjugations of two typing methods, the highest SID's were obtained for the PFGE subtype with either *spa* type or SCC*mec*. It is interesting to note, that for MSSA the three typing methods (PFGE, MLST and *spa*) show similar SID's in contrast to MRSA, where ST's do not perform so well. Also, when comparing the MLST discriminatory power at the ST or eBURST level,

between MRSA and MSSA, the MSSA presented always have higher values. This is also true when comparing the discriminatory power of groups made by the BURP algorithm, further confirming the more diverse genetic structure of the MSSA subset.

#### Typing methods results global concordance: Adjusted Rand

By analyzing the values of Adjusted Rand (AR) coefficient between all typing methods, an overall concordance value can be reached (see Tables 2 – 4). For comparison purposes with other published values of concordance between typing methods (31), tables with Rand coefficient values are provided as supplemental material (see Supplemental Material - Tables 3S to 5S), although these values clearly overestimate the concordance between typing methods as discussed in the Materials and Methods section and previous publications (6). For the entire collection (Table 2) the highest AR value found between two single typing methodologies was between eBURST and BURP groups – 0.551. This value suggests that clustering ST's in clonal complexes using eBURST or *spa* types in groups using BURP extracts roughly similar phylogenetic signals. All other methods and combination of methods presented AR values lower than 0.5 except, as expected, one comparing a method with a conjugation of itself with another method (e.g. PFGE type+*spa* type vs *spa* type, 0.581 of concordance).

A similar situation was found when only the MRSA subset was considered (Table 3). For the MSSA subset (Table 4), AR values were higher, although always below 0.66. PFGE type has the highest agreement values when compared to other methods: 0.53 to *spa* type, 0.61 to BURP groups, 0.66 to ST's, and 0.59 to e-BURST clonal complexes. Similarly, to the entire collection and the MRSA subset, for the MSSA

subset an agreement of 0.56 was found between BURP and eBURST and 0.58  
2 between ST's and BURP groups.

#### 4 **Directional agreement between typing methods groupings: Wallace coefficient**

In order to determine how the results of one method map onto the results of the other  
6 methods, we calculated the Wallace coefficients. The results are presented in tables  
6S to 8S as Supplementary Material, and are summarized in Figures 1, 2 and 3.  
8 Overall, Wallace coefficients show that MRSA and MSSA datasets differ in the  
confidence of the predictions of the results of other typing methods knowing those of  
10 a single method.

Concerning *spa* type performance, it was found that, in all datasets, if any two strains  
12 share the same *spa* type, they have a high probability (over 94%) of belonging to the  
same eBURST group. *spa* typing is also able to predict the PFGE type in MSSA  
14 strains with a 92% probability while for MRSA strains the agreement is only about  
40% (Figures 2 and 3). In all datasets, *spa* typing is also not able to clearly predict the  
16 ST (W=0.61 for the all dataset, W=0.60 for MRSA and W=0.52 for MSSA).

PFGE is able to predict the BURP group much better for MRSA than for MSSA:  
18 W=0.83 for PFGE type and W=0.97 for PFGE subtype in MRSA versus W=0.69 and  
W=0.83, respectively in MSSA strains. Concerning the ability of PFGE (type or  
20 subtype) to predict eBURST complexes, the values are similar for MRSA and MSSA  
as well as for the overall dataset: around W=0.84 for PFGE type and W=1.0 for PFGE  
22 subtype. Similar results, at a lower level of agreement, were found for the PFGE  
subtype's capability to predict ST's: W=0.82 for MRSA dataset and W=0.79 for  
24 MSSA and entire dataset. Conversely, if two strains have the same ST, when  
considering the whole dataset, they have only 33% probability of sharing the same

PFGE type. This value is lower in the MRSA dataset ( $W=0.21$ ), but interestingly higher for the MSSA dataset ( $W=0.79$ ). In summary, knowing the ST information one can only predict with some certainty the PFGE type if an isolate is a MSSA. In the case of MRSA strains, PFGE subtypes also predict quite well the *SCCmec* type ( $W=0.88$ ).

When conjugating the results of PFGE and *spa* typing, it was found that if two strains are classified together in the same PFGE-*spa* type, there is 100% probability of sharing the same eBURST for MSSA or 99.5% for the entire collection (it is not 100% because there are two strains classified together in PFGE type 6 / *spa* type t030 but one belongs to CC 8 while the other is a singleton). For MRSA the agreement between PFGE-*spa* types and eBURST complexes is slightly lower ( $W=0.94$ ), which is explained by the fact that five strains grouped together in PFGE type 18 / *spa* type t001 were divided in CC 228 (2 strains) and CC 5 (3 strains) and two strains grouped together in PFGE type 4- *spa* type t030 that were classified as CC8 and as a singleton. The combination of PFGE subtype with *spa* type does not add over the PFGE subtype alone for the prediction of eBURST groups, although it slightly improves the prediction of ST's (Figure 1). However, in MRSA strains, the PFGE subtype / *spa* type combination was found to perform better in the prediction of *SCCmec* type ( $W=0.91$ ) than PFGE subtype alone ( $W=0.88$ ).

We have also analyzed the performance of PFGE type-*SCCmec* type and PFGE subtype-*SCCmec* type combinations for MRSA strains and, in both cases, there was no significant differences when comparing with the results of the PFGE type-*spa* type conjugation, although the later performs better for the prediction of eBURST clonal complexes ( $W=0.85$  vs  $W=0.94$ ). It is also interesting to note that if two MRSA strains share the same PFGE type-*SCCmec* type, they have a 94% probability of

- sharing the same BURP group, which is a higher probability than that of sharing the
- 2 same eBURST group ( $W=0.85$ ).

ACCEPTED

## DISCUSSION

2

Currently a wide variety of genotype-based typing methods are available for  
4 classifying *S. aureus* isolates for epidemiological studies. Molecular methods based  
on the analysis of band patterns, such as PFGE, are now being replaced by more  
6 portable sequence-based methods such as *spa* typing and MLST. However, the  
advantages of these newer methods in terms of discriminatory power and their  
8 relationship between the groups defined by the once dominant typing methods and  
these new ones, now frequently more used, was not fully explored.

10 In this study, we have analyzed a collection of 198 epidemiologically unrelated *S.*  
*aureus*, composed of 82 MSSA and 116 MRSA, in order to quantify the congruence  
12 between the most frequently used genotyping methods for the characterization of *S.*  
*aureus* strains: PFGE subtypes (defined at an 95% cut-off on the Dice/UPGMA  
14 dendrogram), MLST, *spa* typing and, for MRSA strains, SCCmec typing. We have  
also evaluated the congruence between techniques in assigning strains to larger groups  
16 using eBURST for MLST, BURP for *spa* type and PFGE type defined as the groups  
formed at an 80% cut-off on the Dice /UPGMA dendrogram. For PFGE , the type  
18 definition at the 80% cut-off approximates Tenover's criteria for possibly related  
isolates (up to 6 bands difference) (5) while the 95% cut-off for subtype definition is a  
20 more discriminatory cut-off usually allowing for up to 1 or 2 band difference  
(depending on the total number of bands for each isolate), that would correspond to  
22 indistinguishable or closely related strains in Tenover's classification.

Not only the overall congruence of results was evaluated by Adjusted Rand, but we  
24 also evaluated the capability of one method to predict the results of any other in terms  
of Wallace coefficients. Our goal was to determine, among the methods used or any

combination of them, which is the best and most cost-effective method to infer genetic relatedness.

The discriminatory power is an important parameter for the evaluation of any typing method performance. Our results show that, in contrast with some previous studies (8), PFGE at the subtype level is the most discriminatory technique, with Simpson's ID values over 99.69% (Table 1). Comparing the MRSA and MSSA sub-datasets, it was found that all methods have higher discriminatory power for the latter (always above over 90%), supporting the notion that MSSA have a more diverse genetic structure than MRSA and in agreement with the hypothesis that MRSA derived recently from a limited number of MSSA lineages, by the acquisition of the *SCCmec* element (4, 19).

When assessing the overall concordance between typing methods using the Adjusted Rand coefficient, a distinction between MRSA and MSSA is also apparent. Although, overall the levels of concordance were not high (below 80%), they are higher for the MSSA subset than for MRSA. Assuming that MRSA are largely confined to the clinical settings where intense antibiotic selective pressure may favor the exchange of genetic material. Since different methods probe different areas of the genome, the higher levels of concordance for MSSA, may indicate that the MSSA sub-population strains is constituted by more stable clones whereas among MRSA recombination is a more frequent event. This is further supported by the analyzes of Wallace coefficients, where, for instance, it was found that if two MSSA strains share the same *spa* type they have a 92% probability of sharing the same PFGE type, whereas for MRSA this probability drops to 40%. This doesn't exclude the fact that recombination is an infrequent event in *S. aureus* (14).

*spa* typing and BURP analysis have been proposed as the sequence based method of choice to determine the genetic relatedness of *S. aureus* (1, 36). Our results show that *spa* types alone are able to infer eBURST clonal complexes (W=0.96 for MRSA and W=0.94 for MSSA). The overall agreement between eBURST and BURP clonal complexes was also low (0.53 for MRSA strains and 0.56 for MSSA), although it was the highest found for any two methods in the MRSA subset. This suggests that the BURP algorithm retrieves a similar phylogenetic signal as eBURST but, since the latter interrogates housekeeping genes and BURP is based on the alignment of sequence repeats found in the polymorphic region of the *spa* gene, BURP is likely to reflect a faster evolutionary clock. This is supported by the fact that if two strains belong to the same BURP group they have 74% probability of belonging to the same eBURST clonal complex if the strain are MSSA and 76% if the strain are MRSA, but the reverse (strains belonging to the same eBURST CC, also having the same BURP group) is only 50% for MSSA and 56% for MRSA, indicating that some eBURST groups have strains that belong to more than one BURP group. However, one should bear in mind that the typeability of *spa* typing is not 100%, as shown in Table 1, which could negatively influence an epidemiological study.

Our results also demonstrate that PFGE, while being a labor-intensive and time-consuming technique, shows high levels of agreement with other methods. For instance, if any two strains share the same PFGE type, they have over 80% probability of belonging to the same eBURST CC (W=0.85 for MSSA, W=0.83 for MRSA and W=0.86 for the whole dataset); whereas for PFGE subtypes there is an absolute agreement (W=1.00; fact expected given the high discriminatory power of PFGE at subtype level: SID = 99.69%), which is higher than the values found for agreement between PFGE subtypes and *spa* typing (W=0.97 for MRSA; W=0.94 for

MSSA;  $W=0.82$  for the entire collection) or ST ( $W=0.82$  for MRSA;  $W=0.79$  for MSSA;  $W=0.77$  for the entire collection) .

We have also found that PFGE type-*spa* type combination, for the MRSA subset, improves the predictive power of each single technique for determining SCC*mec* type ( $W=0.71$ ). For the MSSA subset, if two strains share the same PFGE type-*spa* type, they have 100% probability of belonging to the same eBURST CC; for MRSA strains this value is slightly lower ( $W=0.94$ ), and there is no gain of predictive power over *spa* type alone ( $W=0.96$ ). In terms of costs, and for MRSA, the PFGE-SCC*mec* combination is possibly the best option: while being less expensive than *spa* typing alone (data not shown), it was found that if two strains share the same PFGE type-SCC*mec* type they have 94% probability of sharing the same BURP group and 85% probability for sharing the same eBURST clonal complexes (in the last case it is only marginally better than PFGE type alone). However, since SCC*mec* characterization can only be applied to MRSA strains, the PFGE-*spa* type combination remains a more broadly applicable technique to the study of *S. aureus*. Also SCC*mec* characterization does not provides 100% typeability, which can hamper the correct characterization of some MRSA strains even when combined with PFGE types, similarly to *spa* typing.

Based in the data here presented, we can provide a rationale for the informed choice of the most appropriate typing scheme for the characterization of *S. aureus* isolates. Since the MRSA and MSSA subsets were shown to differ in terms of diversity and stability, they could be characterized by different typing schemes, although an unified methodology could be desirable. In our dataset PFGE subtypes were shown to predict the CC at 100% and are also the best method for the ST and SCC*mec* prediction, they have, nevertheless, the disadvantage of requiring a normalized and curated database of patterns, the complexity of which the complexity increases dramatically with the

number of isolates. Therefore, we suggest that the most suitable method to infer clonal relationships between isolates, taking MLST and eBURST as reference method, is *spa* typing. Moreover, for the MSSA dataset *spa* types also predict the PFGE type with 92% confidence rendering PFGE analysis redundant. Therefore, for MSSA strains, our data support the use solely of *spa* typing, whereas for MRSA a combination of PFGE with *spa* typing would offer additional discriminatory power. Overall the most cost effective combination of techniques for a detailed characterization of *S. aureus* isolates, irrespective of resistance to methicillin, would be a combination of PFGE and *spa* typing, which allows a very accurate ( $W = 0.995$ ) assignment of strains to eBURST clonal complexes without performing MLST. This combination also provides the necessary discriminatory power and typeability for local epidemiological studies as well as the possibility of defining more distant relationships between isolates suitable for long-term epidemiological studies.

#### Acknowledgements

Partial support for this study was provided by contract from Fundação para a Ciência e a Tecnologia, Portugal: POCTI/BIA-MIC/58416/2004 to Hermínia de Lencastre. Partial support for this study was provided by project POCTI/BIA-MIC/58416/2004 from Fundação para a Ciência e Tecnologia (FCT), Lisbon, Portugal to H. de Lencastre. N.A.F. and D.C.O. were supported by grants SFRH/BD/19208/2004 and SFRH/BPD/9374/2002, respectively from FCT.

## REFERENCES

- 2 1. **Aires de Sousa, M., T. Conceicao, C. Simas, and H. de Lencastre.** 2005.  
Comparison of genetic backgrounds of methicillin-resistant and susceptible  
4 *Staphylococcus aureus* isolates from Portuguese hospitals and the community.  
J Clin Microbiol **43**:5150-7.
- 6 2. **Aires de Sousa, M., and H. de Lencastre.** 2004. Bridges from hospitals to  
the laboratory: genetic portraits of methicillin-resistant *Staphylococcus aureus*  
8 clones. FEMS Immunol Med Microbiol **40**:101-11.
- 10 3. **Aires-de-Sousa, M., K. Boye, H. de Lencastre, A. Deplano, M. C. Enright,**  
**J. Etienne, A. Friedrich, D. Harmsen, A. Holmes, X. W. Huijsdens, A. M.**  
**Kearns, A. Mellmann, H. Meugnier, J. K. Rasheed, E. Spalburg, B.**  
12 **Strommenger, M. J. Struelens, F. C. Tenover, J. Thomas, U. Vogel, H.**  
**Westh, J. Xu, and W. Witte.** 2006. High interlaboratory reproducibility of  
14 DNA sequence-based typing of bacteria in a multicenter study. J Clin  
Microbiol **44**:619-21.
- 16 4. **Beck, W. D., B. Berger-Bachi, and F. H. Kayser.** 1986. Additional DNA in  
methicillin-resistant *Staphylococcus aureus* and molecular cloning of mec-  
18 specific DNA. J. Bacteriol. **165**:373-378.
- 20 5. **Carrico, J. A., F. R. Pinto, C. Simas, S. Nunes, N. G. Sousa, N. Frazao, H.**  
**de Lencastre, and J. S. Almeida.** 2005. Assessment of band-based similarity  
coefficients for automatic type and subtype classification of microbial isolates  
22 analyzed by pulsed-field gel electrophoresis. J Clin Microbiol **43**:5483-90.
- 24 6. **Carrico, J. A., C. Silva-Costa, J. Melo-Cristino, F. R. Pinto, H. de**  
**Lencastre, J. S. Almeida, and M. Ramirez.** 2006. Illustration of a common

- framework for relating multiple typing methods by application to macrolide-resistant *Streptococcus pyogenes*. J Clin Microbiol **44**:2524-32.
- 2
7. **Chung, M., H. de Lencastre, P. Matthews, A. Tomasz, I. Adamsson, M. Aires de Sousa, T. Camou, C. Cocuzza, A. Corso, I. Couto, A. Dominguez, M. Gniadkowski, R. Goering, A. Gomes, K. Kikuchi, A. Marchese, R. Mato, O. Melter, D. Oliveira, R. Palacio, R. Sa-Leao, I. Santos Sanches, J. H. Song, P. T. Tassios, and P. Villari.** 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**:189-98.
- 4
- 6
- 8
- 10
8. **Cookson, B. D., D. A. Robinson, A. B. Monk, S. Murchan, A. Deplano, R. de Ryck, M. J. Struelens, C. Scheel, V. Fussing, S. Salmenlinna, J. Vuopio-Varkila, C. Cuny, W. Witte, P. T. Tassios, N. J. Legakis, W. van Leeuwen, A. van Belkum, A. Vindel, J. Garaizar, S. Haeggman, B. Olsson-Liljequist, U. Ransjo, M. Muller-Premru, W. Hryniewicz, A. Rossney, B. O'Connell, B. D. Short, J. Thomas, S. O'Hanlon, and M. C. Enright.** 2007. Evaluation of molecular typing methods in characterizing a European collection of epidemic methicillin-resistant *Staphylococcus aureus* Strains: the HARMONY collection. J Clin Microbiol **45**:1830-7.
- 12
- 14
- 16
- 18
- 20
- 22
- 24
9. **Crisostomo, M. I., H. Westh, A. Tomasz, M. Chung, D. C. Oliveira, and H. de Lencastre.** 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**:9865-70.

10. **EARSS** 2006, posting date. European Antimicrobial Resistance Surveillance System 2005, 23 August 2006. <http://www.earss.rivm.nl>. [Online.]
11. **Enright, M. C., N. P. Day, C. E. Davies, S. J. Peacock, and B. G. Spratt.** 2000. Multilocus sequence typing for characterization of methicillin-resistant and methicillin-susceptible clones of *Staphylococcus aureus*. *J Clin Microbiol* **38**:1008-15.
12. **Enright, M. C., D. A. Robinson, G. Randle, E. J. Feil, H. Grundmann, and B. G. Spratt.** 2002. The evolutionary history of methicillin-resistant *Staphylococcus aureus* (MRSA). *Proc Natl Acad Sci U S A* **99**:7687-92.
13. **Faria, N. A., D. C. Oliveira, H. Westh, D. L. Monnet, A. R. Larsen, R. Skov, and H. de Lencastre.** 2005. Epidemiology of emerging methicillin-resistant *Staphylococcus aureus* (MRSA) in Denmark: a nationwide study in a country with low prevalence of MRSA infection. *J Clin Microbiol* **43**:1836-42.
14. **Feil, E. J., J. E. Cooper, H. Grundmann, D. A. Robinson, M. C. Enright, T. Berendt, S. J. Peacock, J. M. Smith, M. Murphy, B. G. Spratt, C. E. Moore, and N. P. Day.** 2003. How clonal is *Staphylococcus aureus*? *J Bacteriol* **185**:3307-16.
15. **Hallin, M., A. Deplano, O. Denis, R. De Mendonca, R. De Ryck, and M. J. Struelens.** 2007. Validation of pulsed-field gel electrophoresis and *spa* typing for long-term, nationwide epidemiological surveillance studies of *Staphylococcus aureus* infections. *J. Clin. Microbiol.* **45**:127-133.
16. **Ho, P. L., C. Cheung, G. C. Mak, C. W. Tse, T. K. Ng, C. H. Cheung, T. L. Que, R. Lam, R. W. Lai, R. W. Yung, and K. Y. Yuen.** 2007. Molecular epidemiology and household transmission of community-associated

- methicillin-resistant *Staphylococcus aureus* in Hong Kong. *Diagn Microbiol Infect Dis* **57**:145-51.
17. **Hubert, L., and P. Arabie.** 1985. Comparing partitions. *J Classification* **2**:193-218.
18. **Hunter, P. R., and M. A. Gaston.** 1988. Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. *J Clin Microbiol* **26**:2465-6.
19. **Ito, T., Y. Katayama, and K. Hiramatsu.** 1999. Cloning and nucleotide sequence determination of the entire *mec* DNA of pre-methicillin-resistant *Staphylococcus aureus* N315. *Antimicrob. Agents Chemother.* **43**:1449-1458.
20. **Kondo, Y., T. Ito, X. X. Ma, S. Watanabe, B. N. Kreiswirth, J. Etienne, and K. Hiramatsu.** 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**:264-74.
21. **Koreen, L., S. V. Ramaswamy, E. A. Graviss, S. Naidich, J. M. Musser, and B. N. Kreiswirth.** 2004. *spa* typing method for discriminating among *Staphylococcus aureus* isolates: implications for use of a single marker to detect genetic micro- and macrovariation. *J Clin Microbiol* **42**:792-9.
22. **Kuehnert, M. J., D. Kruszon-Moran, H. A. Hill, G. McQuillan, S. K. McAllister, G. Fosheim, L. K. McDougal, J. Chaitram, B. Jensen, S. K. Fridkin, G. Killgore, and F. C. Tenover.** 2006. Prevalence of *Staphylococcus aureus* nasal colonization in the United States, 2001-2002. *J Infect Dis* **193**:172-9.

23. **Maiden, M. C., J. A. Bygraves, E. Feil, G. Morelli, J. E. Russell, R. Urwin,**  
2 **Q. Zhang, J. Zhou, K. Zurth, D. A. Caugant, I. M. Feavers, M. Achtman,**  
4 **and B. G. Spratt.** 1998. Multilocus sequence typing: a portable approach to  
the identification of clones within populations of pathogenic microorganisms.  
Proc Natl Acad Sci U S A **95**:3140-5.
- 6 24. **McDougal, L. K., C. D. Steward, G. E. Killgore, J. M. Chaitram, S. K.**  
8 **McAllister, and F. C. Tenover.** 2003. Pulsed-field gel electrophoresis typing  
of oxacillin-resistant *Staphylococcus aureus* isolates from the United States:  
establishing a national database. J Clin Microbiol **41**:5113-20.
- 10 25. **Milheirico, C., D. C. Oliveira, and H. de Lencastre.** 2007. Multiplex PCR  
12 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.
- 14 26. **Murchan, S., M. E. Kaufmann, A. Deplano, R. de Ryck, M. Struelens, C.**  
16 **E. Zinn, V. Fussing, S. Salmenlinna, J. Vuopio-Varkila, N. El Solh, C.**  
18 **Liljequist, U. Ransjo, G. Coombes, and B. Cookson.** 2003. Harmonization  
of pulsed-field gel electrophoresis protocols for epidemiological typing of  
20 strains of methicillin-resistant *Staphylococcus aureus*: a single approach  
developed by consensus in 10 European laboratories and its application for  
22 tracing the spread of related strains. J Clin Microbiol **41**:1574-85.
- 24 27. **Okuma, K., K. Iwakawa, J. D. Turnidge, W. B. Grubb, J. M. Bell, F. G.**  
**O'Brien, G. W. Coombs, J. W. Pearman, F. C. Tenover, M. Kapi, C.**  
**Tiensasitorn, T. Ito, and K. Hiramatsu.** 2002. Dissemination of new

- methicillin-resistant *Staphylococcus aureus* clones in the community. J Clin Microbiol **40**:4289-94.
28. **Oliveira, D. C., and H. de Lencastre.** 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**:2155-61.
29. **Oliveira, D. C., C. Milheirico, S. Vinga, and H. de Lencastre.** 2006. Assessment of allelic variation in the *ccrAB* locus in methicillin-resistant *Staphylococcus aureus* clones. J Antimicrob Chemother **58**:23-30.
30. **Ridom-GmbH.** 2006. Ridom StaphType - User Guide, Wurzburg, Germany.
31. **Robinson, D. A., S. K. Hollingshead, J. M. Musser, A. J. Parkinson, D. E. Briles, and M. J. Crain.** 1998. The IS1167 Insertion Sequence Is a Phylogenetically Informative Marker Among Isolates of Serotype 6B *Streptococcus pneumoniae*. J Mol Evol **47**:222-29.
32. **Sa-Leao, R., A. Tomasz, and H. de Lencastre.** 2001. Multilocus sequence typing of *Streptococcus pneumoniae* clones with unusual drug resistance patterns: genetic backgrounds and relatedness to other epidemic clones. J Infect Dis **184**:1206-10.
33. **Schwartz, D. C., and C. R. Cantor.** 1984. Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. Cell **37**:67-75.
34. **Shopsin, B., M. Gomez, S. O. Montgomery, D. H. Smith, M. Waddington, D. E. Dodge, D. A. Bost, M. Riehman, S. Naidich, and B. N. Kreiswirth.** 1999. Evaluation of protein A gene polymorphic region DNA sequencing for typing of *Staphylococcus aureus* strains. J Clin Microbiol **37**:3556-63.
35. **Simpson, E. H.** 1949. Measurement of species diversity. Nature **163**:688.

36. **Strommenger, B., C. Kettlitz, T. Weniger, D. Harmsen, A. W. Friedrich, and W. Witte.** 2006. Assignment of *Staphylococcus* isolates to groups by *spa* typing, *Sma*I macrorestriction analysis, and multilocus sequence typing. *J Clin Microbiol* **44**:2533-40.
37. **Tenover, F. C., R. D. Arbeit, R. V. Goering, P. A. Mickelsen, B. E. Murray, D. H. Persing, and B. Swaminathan.** 1995. Interpreting chromosomal DNA restriction patterns produced by pulsed-field gel electrophoresis: criteria for bacterial strain typing. *J Clin Microbiol* **33**:2233-9.
38. **Tiemersma, E. W., S. L. Bronzwaer, O. Lyytikainen, J. E. Degener, P. Schrijnemakers, N. Bruinsma, J. Monen, W. Witte, and H. Grundman.** 2004. Methicillin-resistant *Staphylococcus aureus* in Europe, 1999-2002. *Emerg Infect Dis* **10**:1627-34.
39. **van Belkum, A., M. Struelens, A. de Visser, H. Verbrugh, and M. Tibayrenc.** 2001. Role of genomic typing in taxonomy, evolutionary genetics, and microbial epidemiology. *Clin Microbiol Rev* **14**:547-60.
40. **Wallace, D. L.** 1983. A method for comparing two hierarchical clusterings: comment. *J Am Stat Assoc* **78**:569-576.

## TABLES

2

**Table 1** - Resolution of typing methods for the three data sets analyzed.

	Typing technique	Number of types	Typeability (%)	Simpson's I. D.	C. I. (95%)
Entire collection (n=198)	PFGE type	56	100	95.92	(94.89-96.95)
	PFGE subtype	153	100	99.69	(99.55-99.82)
	<i>spa</i> type	98	99	97.31	(96.30-98.33)
	BURP ( <i>spa</i> )	29	97	87.35	(84.88-89.81)
	ST (MLST)	61	100	94.82	(93.42-96.21)
	e-BURST (MLST)	21	100	82.24	(78.19-86.29)
	PFGE type + <i>spa</i> type	129	100	98.88	(98.29-99.47)
	PFGE subtype + <i>spa</i> type	175	100	99.85	(99.74-99.95)
MRSA strains (n=116)	PFGE type	32	100	94.27	(92.70-95.84)
	PFGE subtype	92	100	99.51	(99.20-99.81)
	<i>spa</i> type	51	98.3	95.85	(94.34-97.35)
	BURP ( <i>spa</i> )	14	97.4	78.62	(74.98-82.26)
	ST (MLST)	34	100	91.36	(88.63-94.10)
	e-BURST (MLST)	12	100	70.84	(63.99-77.69)
	SCC <i>mec</i>	11	97.4	83.13	(79.89-86.38)
	PFGE type + <i>spa</i> type	72	100	98.32	(97.47-99.17)
	PFGE subtype + <i>spa</i> type	101	100	99.67	(99.38-99.96)
	PFGE type +SCC <i>mec</i>	56	100	97.32	(96.21-98.43)
	PFGE subtype + SCC <i>mec</i>	96	100	99.57	(99.26-99.87)
	SCC <i>mec</i> + ST (MLST)	50	100	96.42	(95.18-97.65)
MSSA strains (n=82)	PFGE type	30	100	94.85	(92.85-96.86)
	PFGE subtype	63	100	99.28	(98.83-99.72)
	<i>spa</i> type	55	100	97.83	(96.27-99.39)
	BURP ( <i>spa</i> )	26	96.3	93.83	(91.73-95.92)
	ST (MLST)	35	100	96.18	(94.78-97.58)
	e-BURST (MLST)	17	100	90.88	(88.19-93.56)
	PFGE type + <i>spa</i> type	59	100	98.01	(96.43-99.59)
	PFGE subtype + <i>spa</i> type	74	100	99.76	(99.55-99.97)

4 I. D. – Index of Diversity  
 C. I. – Confidence Interval

6

8

**Table 2**– Adjusted Rand values for the entire collection (n=198).

Typing technique	PFGE type	PFGE subtype	<i>spa</i> type	BURP ( <i>spa</i> )	ST (MLST)	e-BURST (MLST)	PFGE type + <i>spa</i> type	PFGE subtype + <i>spa</i> type	<i>spa</i> type + ST (MLST)
PFGE type									
PFGE subtype	0.1373								
<i>spa</i> type	0.3079	0.0975							
BURP ( <i>spa</i> )	0.3215	0.0376	0.3201						
ST (MLST)	0.3372	0.0823	0.3915	0.4915					
e-BURST (MLST)	0.2719	0.0286	0.2189	0.5507	0.4041				
PFGE type + <i>spa</i> type	0.4198	0.2112	0.5810	0.1448	0.2266	0.0990			
PFGE subtype + <i>spa</i> type	0.0699	0.6586	0.1057	0.0210	0.0471	0.0142	0.2399		

**Table 3** – Adjusted Rand values for the MRSA strains (n=116).

Typing technique	PFGE type	PFGE subtype	<i>spa</i> type	BURP ( <i>spa</i> )	ST (MLST)	e-BURST (MLST)	SCC <sub>mec</sub>	PFGE type + <i>spa</i> type	PFGE subtype + <i>spa</i> type	PFGE type +SCC <sub>mec</sub>	PFGE subtype + SCC <sub>mec</sub>	SCC <sub>mec</sub> + ST (MLST)
PFGE type												
PFGE subtype	0.1513											
<i>spa</i> type	0.3065	0.1343										
BURP ( <i>spa</i> )	0.2862	0.0345	0.2749									
ST (MLST)	0.1996	0.0801	0.3529	0.4897								
e-BURST (MLST)	0.1935	0.0239	0.1798	0.5310	0.3735							
SCC <sub>mec</sub>	0.1663	0.0409	0.1579	0.1599	0.1883	0.0571						
PFGE type + <i>spa</i> type	0.4389	0.2981	0.5654	0.1182	0.1804	0.0726	0.1002					
PFGE subtype + <i>spa</i> type	0.1033	0.7992	0.1419	0.0240	0.0576	0.0160	0.0286	0.3246				
PFGE type +SCC <sub>mec</sub>	0.6244	0.2675	0.3245	0.1698	0.2087	0.0998	0.2393	0.5333	0.1943			
PFGE subtype + SCC <sub>mec</sub>	0.1341	0.9352	0.1238	0.0316	0.0750	0.0210	0.0421	0.2787	0.7835	0.2734		
SCC <sub>mec</sub> + ST (MLST)	0.2605	0.1767	0.3873	0.2264	0.5644	0.1656	0.3096	0.3206	0.1250	0.4175	0.1802	

ACCEPTED

**Table 4** – Adjusted Rand values for the MSSA strains (n=82).

Typing technique	PFGE type	PFGE subtype	<i>spa</i> type	BURP ( <i>spa</i> )	ST (MLST)	e-BURST (MLST)	PFGE type + <i>spa</i> type	PFGE subtype + <i>spa</i> type
PFGE type								
PFGE subtype	0.2365							
<i>spa</i> type	0.5288	0.1575						
BURP ( <i>spa</i> )	0.6055	0.1639	0.5039					
ST (MLST)	0.6560	0.2424	0.4160	0.5764				
e-BURST (MLST)	0.5890	0.1352	0.3395	0.5623	0.5674			
PFGE type + <i>spa</i> type	0.5439	0.1690	0.9556	0.4712	0.4094	0.3361		
PFGE subtype + <i>spa</i> type	0.0852	0.4982	0.1965	0.0708	0.0996	0.0470	0.2128	

## Figure Legends

**FIGURE 1.** - Representation of correspondences between typing methods and combination of typing methods, used in the 198 strain collection, calculated by Wallace coefficients. The arrows represent Wallace coefficients of  $>0.60$ , excluding the obvious relations.

**FIGURE 2.** - Representation of correspondences between typing methods and combination of typing methods, used in the 116 MRSA subset, calculated by Wallace coefficients. The arrows represent Wallace coefficients of  $>0.60$ , excluding the obvious relations.

**FIGURE 3.** - Representation of correspondences between typing methods and combination of typing methods, used in the 82 MSSA subset, calculated by Wallace coefficients. The arrows represent Wallace coefficients of  $>0.60$ , excluding the obvious relations.





