Characterization of Clonal Relatedness among the Natural Population of *Staphylococcus aureus* Strains by Using *spa* Sequence Typing and the BURP (Based upon Repeat Patterns) Algorithm

Alexander Mellmann,1* Thomas Weniger,2 Christoph Berssenbrügge,1 Ursula Keckevoet,1 Alexander W. Friedrich,1 Dag Harmsen,2 and Hajo Grundmann3

Institute for Hygiene1 and Department of Periodontology,7 University Hospital Münster, D-48149 Münster, Germany, and Centre for Infectious Diseases Epidemiology, National Institute for Public Health and the Environment, Bilthoven 3720BA, The Netherlands3

Received 14 January 2008/Returned for modification 27 March 2008/Accepted 27 May 2008

We evaluated the BURP (based upon repeat patterns) algorithm, which relies on sequencing of the *Staphylococcus aureus* protein A gene (*spa*), for its ability to infer clonal relatedness within a population of 110 wild-type strains. BURP clustering of the resulting 66 *spa* types was highly concordant with multilocus sequence typing (96.5% concordance) and pulsed-field gel electrophoresis (94.9%).

*Staphylococcus aureus*, the leading cause of nosocomial infections worldwide, causes a wide variety of infections (14, 18). In recent decades, the worldwide spread of methicillin-resistant *S. aureus* (MRSA) has become a major challenge in health care (5). More recently, community-acquired MRSA has intensified public health concerns (26). To understand these changes in epidemiology, different typing methods have been applied, including phage typing, pulsed-field gel electrophoresis (PFGE), and sequence-based typing methods. Sequence-based typing offers the advantage that results are easy to compare and communicate between different laboratories.

Multilocus sequence typing (MLST) has become the “gold standard” for population analysis (15), but it has low discriminatory power and is expensive, so this method is mainly restricted to reference laboratories. As a result of the predominantly clonal evolution of *S. aureus*, sequencing of the repeat region of the protein A gene (*spa*) generates informative typing results and has quickly been established as a robust and highly discriminatory method (1, 3, 10, 16, 21, 24). The *spa* region consists of a variable number of 21- to 27-bp repeats with differing nucleotide compositions that result in different *spa* types. It has been observed that this region provides information not only about short-term epidemiology but also about long-term phylogeny and contains a reliable signal that could be utilized for the determination of clonal relatedness among individual strains (12).

Here we investigated a well-characterized collection of methicillin-sensitive *S. aureus* (MSSA) carriage strains by *spa* typing. To evaluate the ability of *spa* typing to determine the clonal relatedness of a natural population of *S. aureus* strains, we used the recently described grouping algorithm that is “based upon repeat patterns” (BURP) to cluster related *spa* types (17) and compared the results with MLST clonal complexes (CCs) and PFGE groups.

*Published ahead of print on 4 June 2008.*
BURP, the resulting spa types were clustered into 9 spa CCs and 19 singletons. Six spa types with fewer than five repeats were excluded (spa types t026, t232, t233, t287, t362, and t398). Figure 1 shows the spa/BURP and corresponding MLST/eBURST and PFGE grouping results for each isolate. The clonal relatedness of all BURP-grouped spa types is illustrated in a population snapshot (Fig. 2). Four of the nine spa CCs had designated group founders (spa CC401, spa CC382/399, spa CC084/346, and spa CC005). The group founder within BURP clusters of at least three different spa types is defined as the spa type with the highest founder score (assigned to the spa type to which the relevant spa types and strains are most closely related). In two spa CCs (spa CC382/399 and spa CC084/346), spa types t382 and t399 and spa types t084 and t346 had identical founder scores, respectively.

All STs belonging to CC217 and CC15 were grouped into spa CC005 and spa CC084/346. PFGE subdivided both spa CCs into two different groups (CC c and CC d) that shared only one of the seven MLST loci (aroE). PFGE subdivided spa CC401 into three different groups. The 21 strains of spa CC382/399, exhibiting 12 different spa types, all clustered in MLST CC30 (ST30, ST37, ST39, and ST1005) by eBURST.

FIG. 1. UPGMA dendrogram illustrating the similarity (based on Dice coefficients) of SmaI macrorestriction profiles (PFGE) of 110 S. aureus strains carried by healthy individuals in the community. The UPGMA dendrogram was constructed using MEGA (version 4) software (23). The scale bar at the bottom shows percent similarity. The PFGE groups are arbitrarily numbered. CCs with a designated founder are named after the founder ST or spa type(s). CCs without an assigned founder are arbitrarily named with lowercase letters. spa types shorter than five repeats were excluded from BURP clustering. Abbreviations: r, repeat; t, type.

Six other strains of CC30 were clustered by BURP in spa CC c, in one singleton (No260), and in one excluded spa type (No335). PFGE corroborated the group diversity (14 PFGE types). However, PFGE grouping resulted in five groups. Overall, grouping by BURP was highly concordant with that by eBURST (96.5%) and PFGE (94.9%). PFGE groups were 93.8% concordant with eBURST CCs. On the level of types instead of groups, concordance between spa types and ST or PFGE types was 96.8% or 97.1%, respectively, while PFGE types were 95.9% concordant with ST.

Two recent studies with strain collections that did not represent diverse natural populations of S. aureus (Fig. 2), but contained predominantly clinical isolates or mainly MRSA strains, showed a strong sampling bias (9, 22). In both studies, high concordances between BURP-grouped spa types and MLST and PFGE clusters were found, and only a few discrepancies were detected. In this study, similar discrepancies became apparent by using MLST-based grouping as a reference, e.g., in MLST CC30. MLST data are not greatly influenced by the effects of recombination, due to the use of BURST, which deduces CCs from allelic profiles. In contrast, spa and PFGE grouping algorithms lack any transformation of the original data and are therefore more sensitive to recombination events.

FIG. 2. Population snapshot based on BURP analysis of 110 S. aureus strains. BURP grouping using default parameters resulted in 9 spa CCs, 20 singletons, and 6 excluded spa types (spa types t026, t232, t233, t287, t362, and t398). Each dot represents a unique spa type. The diameter of a dot is proportional to the quantity of the corresponding spa type. Blue dots represent group founders, defined as the spa type(s) with the highest founder score within a CC. Note that the spacing between linked spa types and between unlinked spa types and spa CCs provides no information concerning the genetic distance between them.
Therefore, large chromosomal replacements, which affect macrorestriction patterns and spa typing substantially, are likely within CC30. Such events have already been shown in different clonal lineages, including CC30, by a previous study (20). These cases could be clarified by using an additional target from another genomic region (e.g., clumping factor B gene) (13). In some other instances, the use of BURP clustering is limited because short spa types are excluded from further analysis. However, analysis of the SpaServer content, comprising more than 42,000 isolates with more than 3,100 different spa types (10), shows that fewer than 7% of all spa types were affected by exclusion from BURP clustering.

In summary, BURP determines clonal relatedness, yielding results highly congruent with MLST and PFGE groupings, within an unbiased sampled population of MSSA strains based on spa data. spa typing and BURP should be considered for phylogenetic studies in addition to strain typing of MSSA.

A. Mellmann was funded by a grant from the Deutsche Forschungsgemeinschaft (ME 3205/1-1). D. Harmsen is one of the developers of the Ridom StaphType software mentioned in the article. The software is distributed and sold by the company Ridom GmbH, of which he is a partial owner. All other authors have declared no conflicts of interest.

We thank Phillip I. Tarr (Washington University School of Medicine, St. Louis, MO) for critical review of the manuscript.

REFERENCES


