Comparison of Protein A Gene Sequencing with Pulsed-Field Gel Electrophoresis and Epidemiologic Data for Molecular Typing of Methicillin-Resistant Staphylococcus aureus

YI-WEI TANG,* MICHAEL G. WADDINGTON, DOUGLAS H. SMITH, JANICE M. MANAHAN, PEGGY C. KOHNER, LEANNE M. HIGHSMITH, HAUJING LI, FRANKLIN R. COCKERILL III, RODNEY L. THOMPSON, STACY O. MONTGOMERY, AND DAVID H. PERSING

Vanderbilt University School of Medicine, Nashville, Tennessee 37232; MIDI Labs, Inc., Newark, Delaware 19709; Hewlett-Packard, Palo Alto, California 94304; Mayo Clinic, Rochester, Minnesota 55905; Perkin-Elmer Biosystems, Foster City, California 94404; and Corixa Corporation and the Infectious Disease Research Institute, Seattle, Washington 98104

Received 1 September 1999/Returned for modification 22 November 1999/Accepted 7 January 2000

The epidemiologic relatedness of methicillin-resistant Staphylococcus aureus (MRSA) isolates is currently determined by analysis of chromosomal DNA restriction patterns by pulsed-field gel electrophoresis (PFGE). We have evaluated an alternative typing system (MicroSeq StaphTrack Kit; Perkin-Elmer Biosystems) based on the sequence analysis of the chromosomally encoded polymorphic repeat X region of the S. aureus protein A (spa) gene. A total of 69 clinical MRSA isolates were divided into 18 groups according to the number and nucleotide sequences of the spa repeats. Molecular typing results obtained both by spa sequencing and from the PFGE patterns were concordant except for one group, which contained 20 isolates recovered over a 2-year period from hospitalized patients at the Mayo Clinic. Although the spa typing patterns were indistinguishable for those isolates, PFGE analysis yielded seven related but distinguishable patterns. Further coagulase gene sequence analysis subtyped those 20 strains into four groups which followed distinct temporal and geographic distributions. During a 2-year epidemic period there were up to 7 fragment changes in PFGE patterns among epidemiologically related isolates, suggesting that PFGE may be unsuitable for long-term typing of strains involved in epidemics. Although more limited than PFGE in discriminatory power, spa sequencing analysis could be used as a screening method for typing of MRSA strains because of the shorter turnaround time, ease of use, and the inherent advantages of sequence analysis, storage, and sharing of information.

Since its first identification in the early 1960s (12), methicillin-resistant Staphylococcus aureus (MRSA) has become one of the most significant nosocomial pathogens throughout the world and is capable of causing a wide range of hospital infections (1, 17). It continues to spread through new communities wherever the methods and institutions of modern medical practice are adopted, while it regularly causes epidemics in places where it has been endemic for a decade or more (1, 35). Consequently, analysis of the dissemination of MRSA isolates has been a research focus for decades; more recently, the tools used to discriminate MRSA isolates have themselves disseminated (31, 34).

Analysis of genetic relatedness by molecular typing techniques is important for defining epidemics. Accurate and rapid typing of MRSA is crucial for the control of nosocomial outbreaks, and numerous methods have been described (4, 7, 10, 11, 18, 22, 24–26, 33). Pulsed-field gel electrophoresis (PFGE) has proven to be highly discriminatory for MRSA isolates, and it has been suggested that it is superior to other genotyping techniques (2, 23, 26, 29). However, this method is costly and technically complex, and there is a paucity of agreed-upon criteria for the interpretation of banding patterns (15, 31). Furthermore, interlaboratory standardization of PFGE is still problematic (4, 34) and it is not suited for the screening of large numbers of isolates simultaneously (20, 34). Thus, there is an ongoing need for rapid and definitive genotyping systems, especially ones that can be applied in clinical laboratories.

Protein A is a cell wall constituent of S. aureus. Nucleotide sequence analysis of region X of the staphylococcal protein A (spa) gene has demonstrated various numbers of degenerate 24-bp repeats (9). The repetitive region is highly polymorphic; the number and sequence of individual repeats may differ among strains, providing an alternative molecular method for the typing of MRSA strains (5, 6, 28). An MRSA typing system based on amplification and sequencing of the spa gene has been developed and distributed by Perkin-Elmer Biosystems (Foster City, Calif.). The MicroSeq StaphTrack Kit allows rapid typing of MRSA within 2 working days. The X region of the spa gene is amplified by PCR and is then sequenced by a fluorescent DNA sequencing technology. The relatedness of a group of MRSA strains is determined by variation of both the numbers and the sequences of individual repeats in region X of the spa gene (28). We report here on an evaluation of this system for its ability to type 69 clinical MRSA isolates to determine their relatedness in reference to the currently used PFGE technique. Epidemiologic data were collected for resolution of discrepant results.

(This study was presented in part at the 99th General Meeting of the American Society for Microbiology, Chicago, Ill., 30 May to 3 June 1999.)

MATERIALS AND METHODS

Bacterial strains. A total of 69 MRSA strains were included in the study. Sixty-three were isolated from patients seen at the Mayo Clinic, and six were isolated from horses and veterinarians at Oregon State University. Each isolate was from an individual subject, and no repeat isolates were included. All isolates...
were coagulase positive, as determined by a rapid slide agglutination procedure (Murex Diagnostics, Inc., Norcross, Ga.). Methicillin resistance was determined by the presence of a zone of 5-10 mm around a 1-μg oxacillin disk (BBL, Becton Dickinson Microbiology Systems, Sparks, Md.) after 24 h of incubation at 35°C on a Mueller-Hinton agar plate (21).

**Determination of epidemiologic relatedness.** The Mayo Clinic Infection Control Unit staff performed active surveillance and data collection for MRSA-colonized or -infected patients during this study period. A tracking log was used to identify epidemiologic relatedness between patients in order to select isolates for testing and thus document clusters. The key data used to determine relatedness were patient care unit or room, primary service, surgical procedure dates and the associated operating room, and other invasive procedures. When clusters were identified, control measures were then implemented for the patient population and the units or departments involved.

**Genomic DNA analysis by PFGE.** The genomic DNA was extracted from logarithmic-phase *S. aureus* cultures grown in brain heart infusion broth (BHI) as described previously (19, 36). The extracted DNA was prepared in low-melting-point agarose plugs and was digested with the *Smal* enzyme (New England Biolabs, Beverly, Mass.), as described previously (23). The DNA size standards used were a bacteriophage lambda ladder consisting of concatemers starting at 48.5 kbp and increasing to approximately 1,000 kbp (Bio-Rad Laboratories, Hercules, Calif.). Electrophoresis was performed with a Bio-Rad CHEF DR II system (Bio-Rad). Run conditions were 200 V with switching from 10 to 50 s for 15 h at 14°C (3). The gels were stained with ethidium bromide, rinsed, and photographed under UV light. Each strain was classified as indistinguishable, related, or dissimilar based on the pattern observed.

**Results**

**Table 1. spa characteristics of 69 MRSA strains included in the study**

<table>
<thead>
<tr>
<th>spa group</th>
<th>No. of isolates</th>
<th>No. of 24-bp repeats</th>
<th>Amplicon size (bp)</th>
<th>spa patterna</th>
<th>Length of outbreak (wk)</th>
<th>PFGE result</th>
<th>Epidemiologic relatedness</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>9</td>
<td>370</td>
<td>Z B C H R H C C H</td>
<td>NAa</td>
<td>Dissimilara</td>
<td>Unrelateda</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>9</td>
<td>370</td>
<td>A W H E H H J K</td>
<td>NA</td>
<td>Dissimilar</td>
<td>Unrelated</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>9</td>
<td>370</td>
<td>A W C H E E I E J K</td>
<td>1</td>
<td>Indistinguishable</td>
<td>Related</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>9</td>
<td>370</td>
<td>M F M D D H E M E</td>
<td>7</td>
<td>Indistinguishable</td>
<td>Related</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>8</td>
<td>346</td>
<td>Z K H R H P M E</td>
<td>3</td>
<td>Indistinguishable</td>
<td>Related</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>8</td>
<td>346</td>
<td>M F M D D H E M E</td>
<td>1</td>
<td>Indistinguishable</td>
<td>Related</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>10</td>
<td>394</td>
<td>A C B H E I E J K</td>
<td>NA</td>
<td>Dissimilar</td>
<td>Unrelated</td>
</tr>
<tr>
<td>8</td>
<td>5</td>
<td>9</td>
<td>394</td>
<td>A B W H E I E J K</td>
<td>61</td>
<td>Indistinguishable</td>
<td>Related</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>8</td>
<td>394</td>
<td>E O H E H R C H M</td>
<td>107</td>
<td>See Fig. 1</td>
<td>See Fig. 1</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>11</td>
<td>418</td>
<td>A A C T W H E I E J K</td>
<td>NA</td>
<td>Dissimilar</td>
<td>Unrelated</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>11</td>
<td>418</td>
<td>E O H E H R C H H M</td>
<td>NA</td>
<td>Dissimilar</td>
<td>Unrelated</td>
</tr>
<tr>
<td>12</td>
<td>15</td>
<td>11</td>
<td>418</td>
<td>B C M F M F K H I I</td>
<td>82</td>
<td>Indistinguishable</td>
<td>Related</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>3</td>
<td>226</td>
<td>C H M</td>
<td>NA</td>
<td>Dissimilar</td>
<td>Unrelated</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>5</td>
<td>274</td>
<td>D M C O E</td>
<td>NA</td>
<td>Dissimilar</td>
<td>Unrelated</td>
</tr>
<tr>
<td>15</td>
<td>7</td>
<td>6</td>
<td>298</td>
<td>E O H C H M</td>
<td>60</td>
<td>Indistinguishable</td>
<td>Related</td>
</tr>
<tr>
<td>16</td>
<td>1</td>
<td>7</td>
<td>322</td>
<td>E H G C H M</td>
<td>NA</td>
<td>Dissimilar</td>
<td>Unrelated</td>
</tr>
<tr>
<td>17</td>
<td>1</td>
<td>7</td>
<td>322</td>
<td>E G H C H M</td>
<td>NA</td>
<td>Dissimilar</td>
<td>Unrelated</td>
</tr>
<tr>
<td>18</td>
<td>1</td>
<td>7</td>
<td>322</td>
<td>B C D F K H I</td>
<td>NA</td>
<td>Dissimilar</td>
<td>Unrelated</td>
</tr>
</tbody>
</table>

*a For alphanumeric spa patterns, see reference 5. A newly assigned code, W, which was identified in the present study, represents DNA sequence 5’-GAA GAC AAC AAA AAG CCT GGC AAA-3’.

b NA, not applicable.

c Compared to those in other groups.

**TABLE 1. spa characteristics of 69 MRSA strains included in the study**

Genomic DNA was isolated from bacterial cultures by using the PrepMan extraction reagent from Perkin-Elmer Biosystems by the method provided in the kit for the extraction of DNA from gram-positive bacterial cells. PCR amplification of the *spa* gene was performed by adding 25 μl of diluted genomic DNA to 25 μl of the PCR master mixture. The samples were placed in a GeneAmp PCR System 9600 (Perkin-Elmer Biosystems). Thermal cycling parameters included an initial 10 min at 95°C, 30 cycles of 30 s at 95°C, 30 s at 60°C, and 45 s at 72°C, and a final extension at 72°C for 10 min. The PCR products were purified from excess PCR primers and nucleotides prior to sequencing with Microcon-100 microconcentrator columns (Millipore Corp., Bedford, Mass.). Forward and reverse cycle sequencing reaction mixtures were prepared by adding 3 μl of the PCR product to 13 μl of each sequencing master mixture and 4 μl of water. Cycle sequencing reactions were run on a GeneAmp PCR System 9600 (Perkin-Elmer Biosystems) by using the recommended thermal profile. The sequencing reaction mixtures were purified from excess dye-labeled terminators with Centri-Sep spin columns according to the manufacturer’s instructions. Samples were electrophoresed on an ABI Prism 377 automated fluorescent sequencer. The sequence data were analyzed by using the MicroSeq analysis software package to determine the numbers and the sequences of the 24-bp repeats in the X region of the *spa* gene.

**Results**

Sequence analysis of the X region of the *spa* gene resulted in 18 distinct groups. The number of 24-bp repeats within the gene varied between 3 and 11, corresponding to PCR products ranging in size from 226 to 418 bp (Table 1). According to convention, degenerate repeats are assigned alphabetical letters for distinct motifs. Eighteen such motifs were identified in this study; a new sequence (5’-GAA GAC AAC AAA AAG CCT GGC AAA-3’) was designated motif W, raising the total number of described motifs to 26 (5).

Typing results for the *spa* sequences were compared to the results obtained by the currently used PFGE typing method. Complete concordance, as well as epidemiologic relatedness, was found between the two methods for all except one group (designated group 9), which contained 20 isolates recovered from an unusual hospital effort at surveillance of MRSA isolates conducted over a 2-year period (Table 1). Although the majority of the latter isolates had identical gel patterns, PFGE analysis yielded seven related but distinguishable patterns, with up to seven fragment changes observed (Fig. 1). Although only 6 nucleotide sequence differences were found...
in the coa gene (<1%), 20 MRSA isolates from group 9 could be further divided into four groups (Fig. 1). The validities of the typing results from the PFGE, spa, and coa patterns were also compared with the epidemiologic data. Both M35 and M61 were recovered from two patients residing in the same long-term-care facility. Although both spa and coa sequence patterns were identical, PFGE analysis yielded two band changes (Fig. 1). In addition, M27, M35, and M49 were isolated from three patients treated from the same nursing station within a 2-week period; however, the PFGE patterns of M27 and M35 were different from that of M49 by three significant fragment changes (Fig. 1).

In an investigation of the group 9 isolates, there was no evidence of epidemiologic relatedness between isolates M7 and M59 or between isolates M18 and M50. However, the epidemiologic unrelatedness of these strains was confirmed by coa sequence analysis (Fig. 1). For the isolates other than the group 9 isolates spa sequence analysis placed epidemiologically distinct strains into the different groups in complete concordance with the groups into which they were placed by PFGE. Taken together, these data suggest that the combined sequence analysis of spa and coa genes provided greater concordance with epidemiologic information than PFGE.

**DISCUSSION**

Rapid and accurate differentiation of bacterial strains associated with nosocomial infections is central to epidemiologic surveillance and hospital infection control. We have shown that detection of genetic polymorphisms in the X region of the spa gene can be used as a typing method to determine the epidemiologic relatedness of MRSA isolates. Protein A is a component of the S. aureus cell wall and is covalently bound to the peptidoglycan. The spa gene is approximately 2,150 bp and contains three distinct regions: the Fc portion, the X region, and the C terminus. The polymorphic X region contains various numbers of 24-bp repeats with various sequences. The number of repeats varies between 3 and 15, and 26 different DNA sequences of repeats have been described (5). Previous studies have demonstrated that the X region of the spa gene is stable enough both in vitro and in vivo to discriminate between different clones (5). Our data indicate that the organization of the spa gene is stable among epidemiologically related MRSA strains recovered over a period of 2 years. The variability and stability of this gene indicate that sequence analysis of the spa gene could be used as an alternative system for the molecular typing of MRSA isolates.

PFGE has been used as a typing method in many clinical laboratories, including that at the Vanderbilt University Medical Center and the Mayo Clinic (2, 3, 23, 26, 29). PFGE was developed for the resolution of large (50- to 700-kb) fragments of DNA obtained by digesting genomic DNA with restriction enzymes that cleave DNA infrequently (27). Because it randomly samples the entire genome, genomic macrorestriction analysis by PFGE possesses a relatively high discriminatory power. However, PFGE is a complicated and time-consuming procedure. It involves growth of the organisms, usually overnight in broth, imbeddment of the washed cells in agarose, lysing of the cells in situ, digestion of the DNA with restriction enzyme in situ, loading of the agar block into gel, and then separation of the fragments on the gel by pulsed-field electrophoresis. Excluding culture, the whole procedure requires 2 to 5 days and approximately 6 to 8 h of technologist time. In addition, although the gel images can be scanned into computer systems and several commercial systems are available to facilitate comparison of isolates tested at different times or in different laboratories, interlaboratory standardization is still problematic, and problems of interpretation still exist.

PFGE analysis works accurately for the typing of isolates recovered from a localized outbreak of disease within a short period of time. However, individual loci or uncharacterized regions of the genome may be highly variable within the bacterial population; consequently, variation that is observed may be due to a variety of genetic mechanisms, the results of which can be misleading for long-term outbreaks (16). In theory, PFGE may be too discriminatory for long-term epidemiologic investigations because it fails to account for genetic changes within a single, globally distributed, clonal lineage of MRSA.
In our study, several epidemiologically related MRSA strains, as well as MRSA strains with identical spa and coa sequences, recovered over 2-year period had up to seven fragment changes in their PFGE patterns. These data suggest that during the course of a nosocomial epidemic of MRSA infection strain evolution occurs, and this may confound the use of PFGE for strain typing. Thus, when PFGE analysis is used for bacterial typing, the duration of the outbreak may need to be considered when patterns are interpreted.

The criteria set forth for determination of the suitability of typing schemes include typeability, reproducibility, discriminatory power, ease of use, and ease of interpretation. A parallel evaluation of up to 12 currently available typing systems with 60 staphylococcal isolates demonstrated that no single typing method, including PFGE, clearly prevailed over the others, and ultimately, a combination of two methods may be most efficacious (31). This has been shown for other organisms such as Escherichia coli O157 (8). Theoretically, one method that would be sensitive enough to identify all potential patients or sources may be used for the screening of isolates early in an epidemiologic evaluation, and another method for detailed strain differentiation may be used later. In our experience, typing systems based on spa sequencing can be completed within 2 days and possess a satisfactory discriminatory power. In addition, driven in part by the human and microbial genome projects, sequencing costs will probably continue their rapid trend downward, bringing this technology within reach of many microbiology laboratories. The potential advantages in terms of data analysis, storage, and sharing of information suggest that spa sequencing may soon become an alternative to PFGE for the typing of MRSA isolates.

Although spa sequencing appeared to be a quicker, unambiguous method for the typing of MRSA isolates, additional procedures may be necessary to provide maximum resolution. The discriminatory power of a single polymorphic gene is usually limited. Since the spa sequence is relatively stable both in vitro and in vivo, strains with different spa types are generally not epidemiologically related (5). However, the converse cannot be concluded; i.e., if strains have the same spa type, this does not imply that they are epidemiologically related. In our study, 20 MRSA isolates possessed an identical spa sequence; however, not all of them were epidemiologically related, as demonstrated by PFGE analysis, coa gene sequencing, and the epidemiologic data themselves. A recent study on molecular typing of meningococci indicated that multilocus sequence typing is a potentially viable approach to the identification of clones within populations of pathogenic microorganisms (16). The combination of two chromosomal genetic elements showed enough discriminatory power to provide a rich source of strain-specific genotypic heterogeneity (14). Thus, in our own clinical context, if spa sequencing fails to demonstrate a difference between strains, especially in the context of a long-term nosocomial outbreak, coa sequencing and PFGE could probably be used to resolve epidemiologically significant groups.

ACKNOWLEDGMENTS

We thank Scott Anderson, Margaret Richman, Eric Troop, John Bartell, Doug Bost, Deborah Dodge, Dan Chapman, Brenda Dylla, Garry Hall, Patty Schams, Sue Dana-Goorden, Kerri Smidt, Sharlene Allen, Nancy Haukom, Barbara Leyc, Barbara Meline, Karen Olson, and Alan Wright for excellent assistance.

REFERENCES


