JOURNAL OF CLINICAL MICROBIOLOGY, May 1996, p. 1176–1179 Vol. 34, No. 5
0095-1137/96/$04.00+0

Novel BOX Repeat PCR Assay for High-Resolution Typing of *Streptococcus pneumoniae* Strains

ALEX VAN BELKUM,[1]* MARCEL SLUIJTER,[2] RONALD DE GROOT,[2] HENRI VERBRUGH,[1] AND PETER W. M. HERMANS[2]

Department of Clinical Microbiology[1] and Department of Paediatrics,[2] University Hospital Rotterdam, Erasmus University, 3015 GD Rotterdam, The Netherlands

Received 25 September 1995/Returned for modification 10 January 1996/Accepted 1 February 1996

Typing data obtained by specifically targeting a single, high-stringency PCR at the pneumococcal BOX repeat element for 28 strains of *Streptococcus pneumoniae* completely corroborated the resolution attained by five genotypic procedures as described by Hermans et al. (P. W. M. Hermans, M. Sluijter, T. Hoogenboezem, H. Heersma, A. van Belkum, and R. de Groot, J. Clin. Microbiol. 33:1606–1612, 1995). All pairs of strains, except one, derived from both the cerebrospinal fluid and blood of the same individual were shown to be identical. Moreover, other, epidemiologically unrelated isolates were demonstrated to be unique. Considering the combined data from the five typing techniques applied previously as the "gold standard," the single BOX PCR test demonstrated excellent resolving powers while maintaining epidemiological linkage.

DNA typing by amplification of regions of tandemly repeated DNA has been applied for different purposes (6, 16). This approach enables digitalization of the human genome by amplification of defined patches of satellite DNA (2). Typing by amplification of DNA bordered by invertedly repeated elements has also been described for a variety of strictly eukaryotic organisms (7). For members of the prokaryotic kingdom, examples of a similar type of analysis have been presented. By targeting the PCR at conserved elements such as ribosomal genes or genes encoding tRNAs, restriction fragment length variation in the various spacer regions can be visualized (5). Also, PCR-mediated amplification of regions bordered by enterobacterial repetitive intergenic consensus (ERIC) sequences or repetitive extragenic palindrome (REP) motifs proved to be a valuable tool for determination of genetic variation among an extensive range of bacterial species (15). However, relatively relaxed primer annealing conditions have been used in these studies, and it remains to be determined whether the enterobacterial repetitive intergenic consensus and repetitive extragenic palindrome PCRs are basically different from arbitrary primed PCR or analysis by random amplification of polymorphic DNA (17, 18). In contrast to the situation for eukaryotes, there are no clear-cut examples of the successful demonstration of DNA polymorphism among bacterial strains by use of high-stringency PCR that targets repetitive DNA motifs.

Typing of microbial pathogens in a swift and efficient manner is of importance for medical microbiologists and hospital epidemiologists. By adequate laboratory analysis of bacterial strains, valuable insight into the local, national, or international spread of pathogens can be obtained. Several molecular genetic tools have been adapted successfully for application in the field of bacterial infection surveillance (4, 9, 13, 14). Although most of the techniques generally perform satisfactorily, improvement in resolution, reproducibility, and/or speed is required in several instances (8, 12).

It was recently demonstrated for strains of *Streptococcus pneumoniae* that the combination of the results of five molecular fingerprinting techniques leads to nearly complete resolution of strains that have been obtained from epidemiologically unrelated sources (1). However, in a routine microbiology laboratory, the application of multiple typing techniques is not feasible because of limitations in time, expertise, and/or budget. For that reason, a novel, high-stringency PCR assay, targeted at regions within the pneumococcal genome and bordered by invertedly repeated BOX elements (3), was developed.

MATERIALS AND METHODS

Bacterial strains. Twenty-eight strains of *S. pneumoniae* were obtained from the Dutch National Reference Center for Bacterial Meningitis (L. Spanjaard, Academic Medical Center, Amsterdam, The Netherlands). These strains have been analyzed by DNA fingerprinting in a previous study (1). For specificity studies DNA was extracted from various other bacterial species, including *Acinetobacter anitratus*, *Acinetobacter baumannii*, *Acinetobacter lwoffi*, *Citrobacter freundii*, *Enterobacter aerogenes*, *Enterobacter cloacae*, *Enterococcus faecalis*, *Escherichia coli*, *Haemophilus influenzae*, *Haemophilus parainfluenzae*, *Klebsiella oxytoca*, *Klebsiella pneumoniae*, *Listeria monocytogenes*, *Moraxella catarrhalis*, *Proteus mirabilis*, *Proteus vulgaris*, *Pseudomonas aeruginosa*, *Pseudomonas cepacia*, *Pseudomonas fluorescens*, *Pseudomonas putrefaciens*, *Salmonella enteritidis*, *Salmonella typhimurium*, *Serratia marcescens*, *Staphylococcus aureus* ($n = 10$), *Staphylococcus epidermidis*, *Staphylococcus haemolyticus*, *Staphylococcus hominis*, *Staphylococcus warneri*, *S. pneumoniae*, *Streptococcus sanguis* ($n = 3$), *Streptococcus intermedius*, and *Xanthomonas maltophilia.*

The genotypically different strains of *S. aureus* were used for assessing whether genetic polymorphisms in nonpneumococcal strains could be detected by the BOX PCRs. These strains were randomly chosen, epidemiologically nonrelated clinical isolates from the Danish Statens Seruminstitute (courtesy of Niels Rieswerts Eriksen). The strains had previously been typed by three different random-amplification-of-polymorphic-DNA assays and restriction fragment length polymorphism analysis targeted at the coagulase and protein A genes (unpublished data). All selected strains represented distinct genotypes.

DNA extraction and PCR typing. Bacteria were cultured in Todd-Hewitt broth containing 0.5% yeast extract, and genomic DNA was isolated with the use of cetyltrimethylammonium bromide (14). PCR was performed with a buffer and thermostable *Taq* polymerase (SuperTaq; Sphaero Q, Leiden, The Netherlands) as described and under conditions employed previously (10, 11). Primers were designed on the basis of the primary structure of the BOX repeat motif, and template DNA was included in amounts of 50 ng/100 μl. The DNA sequences of the primers and their position within the BOX element are as follows: BOX A, 5′-ATACTCTTCGAAAATCTCTTCAAAC-3′, positions 2 to 26; BOX B, 5′-CACGAGAAACTAAAAGTAACTCATAT-3′, positions 77 to 59; BR-1, 5′-CTGACTTCGTCAGTTCTATC-3′, positions 80 to 99; and BR-2A, 5′-TTTGTGACAAAACTCCAACA-3′, positions 152 to 128 (3). The PCR program was designed as follows. The incubation mixtures were predenatured for 4 min at 94°C, after which 40 cycles of repeated denaturation (1 min at 94°C), primer annealing

* Corresponding author. Mailing address: University Hospital Rotterdam, Department of Bacteriology, Dr. Molewaterplein 40, 3015 GD Rotterdam, The Netherlands. Phone: 31-10-4635813. Fax: 31-10-4633875. Electronic mail address: vanbelkum@bacl.azr.nl.

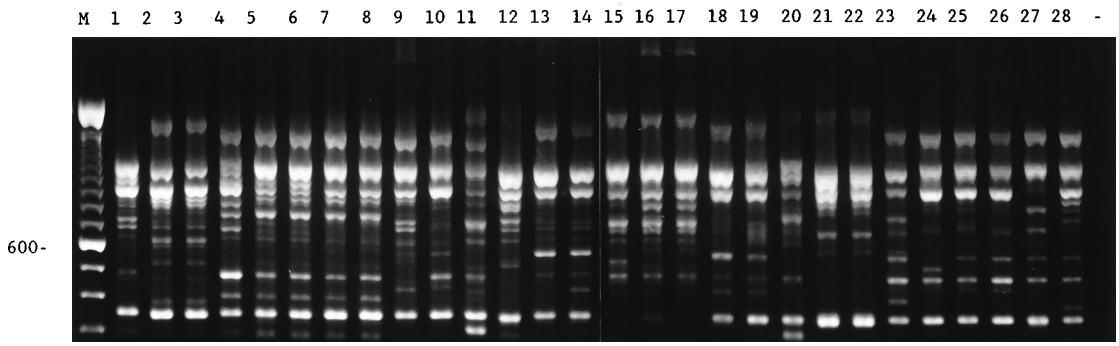

FIG. 1. BOX PCR for 28 strains of *S. pneumoniae*. The lane numbers correspond to those in Table 1. Lane M, molecular length marker (100-bp ladder; Promega, Leiden, The Netherlands). The position of the major 600-bp fragment is indicated on the left. Lane N, control reaction with no template DNA included during PCR. Lanes 1 to 28, strains 175B, 181B, 181C, 200C, 310B, 310C, 369B, 369C, 015C, 019B, 174B, 189C, 241C, 353B, 078B, 797B, 797C, 242B, 242C, 187C, 063B, 063C, 352C, 079B, 022B, 022C, 021B, and 064C, respectively.

(2 min at 50 or 60°C) and chain extension (2 min at 74°C) took place. Finally, a postcycling incubation of 5 min at 74°C was performed. In order to determine the influence of buffer components and the brand of thermostable DNA polymerase, BOX PCRs were also performed with Goldstar *Taq* polymerase (Eurogentec, Seraing, Belgium) and its appropriate buffer [75 mM Tris-HCl (pH 9.0), 20 mM $(NH_4)_2SO_4$, 1.5 mM $MgCl_2$, 0.01% Tween 20]. The amplified products were separated by length on 1.5% agarose gels run in 0.5× Tris-borate-EDTA at a constant current of 100 mA for 3 h. The data were evaluated by visual inspection of the banding patterns, and the individual fingerprints were indexed with a capital letter as a genocode. In addition, phylogenetic trees were constructed with the help of Gelcompar 3.10 (Applied Maths, Kortrijk, Belgium). Pictures were first digitized by use of a scanner (HP Scanjet; Hewlett-Packard, Amsterdam, The Netherlands). Comparison of the different lanes was performed by the unweighted pair group method with arithmetic averages clustering method, with the Jaccard coefficient being applied to peaks. A tolerance of 2% was allowed during comparison of band positions.

RESULTS

Amplification of pneumococcal DNA by using the four different BOX motif primers in individual assays demonstrated that only the BOX A primer generated satisfactory amounts of amplicons in highly variable numbers and sizes. The BR-1 and BR-2 primers did not give rise to DNA amplification, whereas the PCR with the BOX C primer led to the generation of a single 2,000-bp amplicon with any pneumococcal DNA template (results not shown). Figure 1 displays the results obtained for the 28 pneumococcal strains when primer BOX A was employed at an annealing temperature of 60°C (see Table 1 for the interpretation of Fig. 1). When the annealing temperature was lowered to 50°C, it appeared that the synthesis of longer DNA fragments was favored slightly. No increase in resolving power was observed, although the patterns generated at 60°C seemed to be somewhat more clear (comparative data not shown). Under the conditions described above, DNAs from a number of additional bacterial species were amplified as well. DNAs from most of the control microorganisms (see Materials and Methods) could not be amplified. With three strains of *S. sanguis*, a small number of amplimers were synthesized. The patterns encountered among these viridans streptococci were variable (data not shown). The single strain of *S. intermedius* that was included in the control panel did not give rise to the synthesis of amplimers, indicating that the BOX elements may be restricted to certain species of streptococci only. The largest number of DNA fragments synthesized and the highest degree of banding pattern variation were observed for strains of *S. pneumoniae*. Moreover, altering the buffer conditions or the brand of *Taq* DNA polymerase can also lead to alteration of the specificity of the BOX PCR (more amplicons synthesized, even with DNA from gram-negative organisms as the template), which may in addition decrease its resolving power with respect to *S. pneumoniae*. However, the epidemiological clustering remained correct, albeit at a lower resolution. Figure 2 clearly demonstrates that when the *Tth* polymerase with its appropriate (high-magnesium) buffer was used, none of the *S. aureus* DNA samples gave rise to DNA fingerprints. However, when Goldstar *Taq* polymerase (Eurogentec) combined with its appropriate, low-stringency buffer was used, fingerprints were generated (Fig. 2B). Interestingly, the banding patterns obtained were all completely different, confirming the lack of epidemiological relatedness among the *S. aureus* strains.

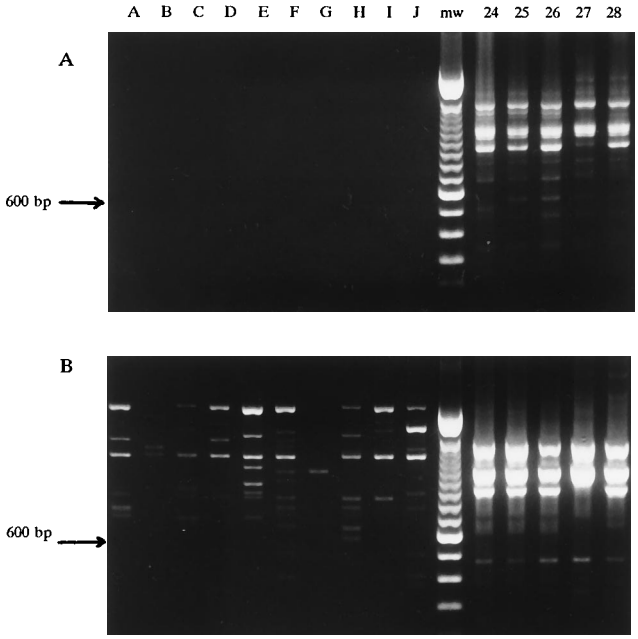

FIG. 2. BOX PCR for 10 strains of *S. aureus* (lanes A to J) and 5 strains of *S. pneumoniae* (lanes 24 to 28) (see Fig. 1 and Table 1). (A) Data obtained with *Tth* polymerase; (B) results obtained by amplification with Goldstar *Taq* polymerase. *S. aureus* DNA is not amplified under the more stringent conditions employed during the experiment resulting in the data shown in panel A. When the experimental conditions are changed, the PCR loses its specificity for pneumococci and the staphylococcal DNA is amplified as well. All 10 banding patterns obtained are different, confirming the epidemiological independence of the strains. Lane mw, molecular length marker (100-bp ladder; Pharmacia, Woerden, The Netherlands). The position of the prominent 600-bp fragment is indicated on the left.

TABLE 1. DNA fingerprint results for 28 *S. pneumoniae* strains analyzed by serology, five different DNA typing strategies, and the BOX PCR[a]

Strain[b]	Lane[c]	Serotype	Strain type[d]					Overall DNA type[e]	BOX PCR[f]
			16S	BOX[g]	RAPD	PFGE	RFEL		
022B	25	7F	A	A	A	A	A	**1**	**I**
022C	26	7F	A	A	A	A	A	**1**	**I**
019B	10	8	A	B	A	A	A	**1A**	**II**
079B	24	7F	A	C	A	A	A	**1B**	**II**
078B	15	4	B	D	B	B	B	2	III
181B	2	6B	B	E	C	C	C	**3**	**IV**
181C	3	6B	B	E	C	C	C	**3**	**IV**
063B	21	23F	C	F	D	D	D	**4**	**V**
063C	22	23F	C	F	D	D	D	**4**	**V**
174B	11	19F	D	G	E	E	E	5	VI
187C	20	18C	D	H	F	F	F	6	VII
200C	4	6B	D	I	G	G	G	7	VIII
015C	9	8	D	J	H	H	H	8	IX
241C	13	9V	D	K	I	I	I	9	X
353B	14	9V	D	L	I	J	I	10	XI
797B	16	4	D	M	J	K	J	**11**	**XII**
797C	17	4	D	M	J	K	J	**11**	**XII**
242B	18	18C	D	L	K	J	I	12	XIII
242C	19	18C	D	N	L	J	K	13	XIV
021B	27	14	D	O	M	—[h]	L	**14**	XV
064C	28	14	D	P	M	L	L	**14A**	XVI
189C	12	9F	D	Q	B	M	M	15	XVII
310B	5	6B	E	R	N	N	N	**16**	**XVIII**
310C	6	6B	E	R	N	N	N	**16**	**XVIII**
369B	7	6B	E	S	O	O	O	**17**	**XIX**
369C	8	6B	E	S	O	O	O	**17**	**XIX**
352C	23	1	F	T	N	P	P	18	XX
175B	1	6B	G	U	A	Q	Q	19	XXI
No. of types		10	7	21	15	17	17	19	21

[a] Part of this table was presented previously (1).
[b] The numbers in the strain designations represent the patient, and the letters indicate the origin of the strain (B, blood; C, cerebrospinal fluid).
[c] Lanes are numbered as in Fig. 1.
[d] 16S, ribotyping; BOX, BOX fingerprinting (Southern hybridization with oligonucleotide probe); RAPD, random amplification of polymorphic DNA; PFGE, pulsed-field gel electrophoresis; RFEL, restriction fragment end labeling.
[e] Overall DNA types were determined by combining the results of all of the individual DNA-mediated procedures except the BOX PCR. The DNA types shared by different strains are shown in boldface. Types 1A and 1B were found to differ from type 1 on the basis of the results of a single technique only.
[f] BOX types are defined as described in Materials and Methods. The DNA types shared by different strains are shown in boldface.
[g] A cutoff value of 90% similarity was used to classify identical and distinct strain types.
[h] —, DNA was degraded.

The interpretation of the results displayed in Fig. 1 is given in Table 1. It is concluded that the BOX PCR displays excellent resolving capacities. In general, epidemiological relationships are identified correctly. When the combined data from the five genotypic approaches (1) are considered the "gold standard" (expressed as overall DNA type in Table 1), the BOX PCR displays aberrant results in two instances. The pair of 242B and 242C was discriminated (Fig. 1, lanes 18 and 19), whereas that of 019B and 079B was not (Fig. 1, lanes 10 and 24). This number of potential misinterpretations, which is possibly a reflection of a relatively high level of background mutation in or among the BOX loci, equals the number of potentially incorrect results of the combined strategies (1).

The phylogenetic tree presented in Fig. 3 illustrates that automated data analysis can be performed and that data generated in this way are congruent with those obtained by visual inspection. When the homology cutoff value is set at 90%, which is feasible given the limited number of DNA fragments generated, all pairs are clustered correctly. The only independent isolates which are clustered are strains 241C and 353B (Fig. 1, lanes 13 and 14). It has to be emphasized that these strains were also considered identical on the basis of ribotyping, PCR fingerprinting, and pulsed-field gel electrophoresis (Table 1). The strains were separated on the basis of visual inspection because of the presence of an additional 450-bp DNA fragment in the PCR fingerprint of strain 353B (Fig. 1, lane 14).

DISCUSSION

A unique class of highly conserved, repetitive DNA elements within the genome of *S. pneumoniae* was described previously (3). Although no clear-cut function has been proposed to date, these elements may represent regulatory DNA domains possibly controlling competence-specific and virulence-related genes. Despite this lack of precise functional knowledge, we here have used the apparent genetic variability in the vicinity of the BOX motifs for DNA-mediated typing for *S. pneumoniae*. The specificity of the newly developed assay is high; only in the case of *S. sanguis* can DNA amplification be observed for another organism as well. For the other species (*S. intermedius*) included in the present reference panel, no amplimers were observed upon BOX PCR. This implies that the BOX element may be present only in a subset of the streptococcal species. When the

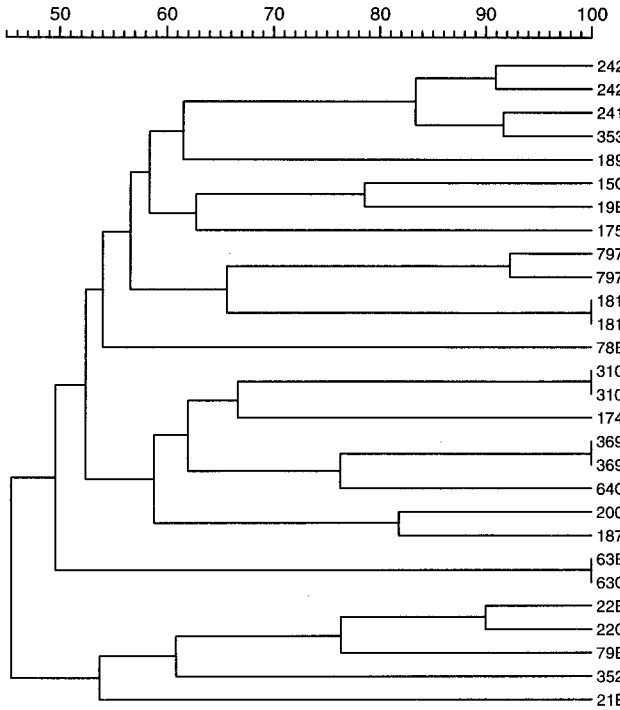

FIG. 3. Dendrogram derived from computerized analysis of the BOX PCR-generated DNA fingerprints. Strain codes, corresponding to those in Fig. 1, are given on the right. The percent homology is given at the top. The dendrogram was constructed as described in Materials and Methods.

stringency of the reaction is lowered, aspecific DNA amplification also occurs when DNAs of other microorganisms (e.g., *S. aureus*) are used as templates. This phenomenon may indicate that within the staphylococcal genome, BOX-like sequences are present, although these sequence motifs must be different from their pneumococcal counterparts.

A lack of interinstitute reproducibility with respect to random amplification of polymorphic DNA was recently demonstrated in a multicenter study on *S. aureus* (12), indicating the necessity for improving the standardization of PCR-mediated typing studies. Unfortunately, the high-stringency assay described in the present paper does not provide a solution to this problem. Although it was expected that with the high-stringency PCR conditions banding patterns could be reproduced in detail in different laboratories, it was demonstrated that this was not the case. Variation in the brand of thermostable DNA polymerase and the nature of the buffer caused the banding patterns to change. However, the resolution of the PCRs remained unaffected, rendering the BOX PCR well suited for intralaboratory typing of *S. pneumoniae*. Moreover, upon standardization of both the brand of polymerase and the buffer, the present assay may be of additional value for interinstitute comparisons of *S. pneumoniae* typing data.

It has been demonstrated here that a single high-stringency PCR assay targeting a dispersed repetitive motif in the genome of *S. pneumoniae* has equal or even greater resolving power than any of the other individual DNA typing strategies. Only the Southern hybridization test employing BOX oligonucleotide motifs as DNA probes can withstand direct comparison with BOX PCR, although the hybridization methodology is laborious and technically complex. It is suggested that the newly developed BOX PCR may be the optimal tool for high-speed, high-resolution typing of pneumococcal strains, especially in day-to-day practice within the clinical microbiology laboratory. Our current efforts are focused on the automated digitalization of the DNA banding patterns and on the implementation of the BOX PCR in larger-scale epidemiological studies of *S. pneumoniae.*

REFERENCES

1. **Hermans, P. W. M., M. Sluijter, T. Hoogenboezem, H. Heersma, A. van Belkum, and R. de Groot.** 1995. Comparative study of five different DNA fingerprint techniques for molecular typing of *Streptococcus pneumoniae* strains. J. Clin. Microbiol. **33:**1606–1612.
2. **Jeffreys, A. J., A. McLeod, K. Tamaki, D. L. Neil, and D. G. Monckton.** 1991. Minisatellite repeat coding as a digital approach to DNA typing. Nature (London) **354:**204–209.
3. **Martin, B., O. Humbert, M. Camara, E. Guuenzi, J. Walker, T. Mitchell, P. Andrew, M. Prudhomme, G. Alloing, R. Hakenbeck, D. A. Morrison, G. J. Boulnois, and J. P. Claverys.** 1992. A highly conserved repeated DNA element located in the chromosome of *Streptococcus pneumoniae*. Nucleic Acids Res. **20:**3479–3483.
4. **Maslow, J., and R. D. Arbeit.** 1995. Molecular epidemiology for the clinician: selected techniques. Infect. Dis. Clin. Pract. **4:**128–133.
5. **McClelland, M., C. Petersen, and J. Welsh.** 1992. Length polymorphisms in tRNA intergenic spacers detected by using the polymerase chain reaction can distinguish streptococcal strains and species. J. Clin. Microbiol. **30:** 1499–1504.
6. **Perring, T. M., A. D. Cooper, R. J. Rodriguez, C. A. Farrar, and T. S. Bellows, Jr.** 1993. Identification of a whitefly species by genomic and behavioral studies. Science **259:**74–77.
7. **Riley, D. E., M. Samadpour, and J. N. Krieger.** 1991. Detection of variable DNA repeats in diverse eukaryotic microorganisms by a single set of PCR primers. J. Clin. Microbiol. **29:**2746–2751.
8. **Tenover, F. C., R. Arbeit, G. Archer, J. Biddle, S. Byrne, R. Goering, G. Hancock, G. A. Hebert, B. Hill, R. Hollis, W. R. Jarvis, B. Kreiswirth, W. Eisner, J. Maslow, L. K. McDougal, J. M. Miller, M. Mulligan, and M. A. Pfaller.** 1994. Comparison of traditional and molecular methods of typing of *Staphylococcus aureus*. J. Clin. Microbiol. **32:**407–415.
9. **van Belkum, A.** 1994. DNA fingerprinting of medically important microorganisms by use of PCR. Clin. Microbiol. Rev. **7:**174–184.
10. **van Belkum, A., R. Bax, P. Peerbooms, W. Goessens, N. van Leeuwen, and W. Quint.** 1993. Comparison of phage typing and DNA fingerprinting by PCR for discrimination of methicillin-resistant *Staphylococcus aureus* isolates. J. Clin. Microbiol. **31:**798–803.
11. **van Belkum, A., R. Bax, and G. Prevost.** 1994. Comparison of four genotyping assays for epidemiological study of methicillin resistant *Staphylococcus aureus*. Eur. J. Clin. Microbiol. Infect. Dis. **13:**420–424.
12. **van Belkum, A., J. Kluytmans, W. van Leeuwen, R. Bax, W. Quint, A. Peters, A. Fluit, C. Vandenbroucke-Grauls, A. Van den Brule, H. Koeleman, W. Melchers, J. Meis, A. Elaichouni, M. Vaneechoutte, F. Moonens, N. Maes, M. Struelens, F. Tenover, and H. Verbrugh.** 1995. Multicenter evaluation of arbitrarily primed PCR for typing of *Staphylococcus aureus* strains. J. Clin. Microbiol. **33:**1537–1547.
13. **van Embden, J. D. A., M. D. Cave, J. T. Crawford, J. W. Dale, K. D. Eisenach, B. Gicquel, P. W. M. Hermans, C. Martin, R. Mcadam, T. M. Shinnick, and P. M. Small.** 1992. Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. J. Clin. Microbiol. **31:**406–409.
14. **van Soolingen, D., P. E. W. de Haas, P. W. M. Hermans, P. M. A. Groenen, and J. D. A. van Embden.** 1993. Comparison of various repetitive DNA elements as genetic markers for strain differentiation and epidemiology of *Mycobacterium tuberculosis*. J. Clin. Microbiol. **31:**1987–1995.
15. **Versalovic, J., T. Koeuth, and J. R. Lupski.** 1991. Distribution of repetitive DNA sequences in eubacteria and application to fingerprinting bacterial genomes. Nucleic Acids Res. **19:**6823–6831.
16. **Warburton, P. E., and H. F. Willard.** 1992. PCR amplification of tandemly repeated DNA: analysis of intra- and interchromosomal sequence variation and homologous unequal crossing-over in human alpha satellite DNA. Nucleic Acids Res. **20:**6033–6042.
17. **Welsh, J., and M. McClelland.** 1990. Fingerprinting genomes using PCR with arbitrary primers. Nucleic Acids Res. **18:**7213–7218.
18. **Williams, J. G., A. R. Kubelik, K. J. Livak, J. A. Rafalski, and S. V. Tingey.** 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res. **18:**6531–6535.