Performance Assessment of DNA Fragment Sizing by High-Sensitivity Flow Cytometry and Pulsed-Field Gel Electrophoresis

Matthew M. Ferris,¹ Xiaomei Yan,¹ Robbert C. Habbersett,¹ Yulin Shou,¹ Cheryl L. Lemanski,¹
James H. Jett,¹ Thomas M. Yoshida,² and Babetta L. Marrone^{1*}

Bioscience¹ and Chemistry² Divisions, Los Alamos National Laboratory, Los Alamos, New Mexico 87545

Received 6 October 2003/Returned for modification 1 December 2003/Accepted 27 January 2004

The sizing of restriction fragments is the chief analytical technique utilized in the production of DNA fingerprints. Few techniques have been able to compete with pulsed-field gel electrophoresis (PFGE), which is capable of discriminating among bacteria at species and strain levels by resolving restriction fragments. However, an ultrasensitive flow cytometer (FCM) developed in our lab has also demonstrated the ability to discriminate bacteria at species and strain levels. The abilities of FCM warrant a quantitative parallel comparison with PFGE to assess and evaluate the accuracy and precision of DNA fragment sizing by both techniques. Replicate samples of *Staphylococcus aureus* Mu50 were analyzed along with two clinical *S. aureus* isolates. The absolute fragment sizing accuracy was determined for PFGE ($5\% \pm 2\%$) and FCM ($4\% \pm 4\%$), with sequence-predicted Mu50 SmaI fragment sizes used as a reference. Precision was determined by simple arithmetic methods (relative standard deviation for PFGE [RSD_{PFGE}] = $3\% \pm 2\%$ and RSD_{FCM} = $1.2\% \pm 0.8\%$) as well as by the use of dendrograms derived from Dice coefficient-unweighted pair group method with arithmetic averages (UPGMA) and Pearson-UPGMA analyses. All quantitative measures of PFGE and FCM precision were equivalent, within error. The precision of both methods was not limited by any single sample preparation or analysis step that was tracked in this study. Additionally, we determined that the curve-based clustering of fingerprint data provided a more informative and useful assessment than did traditional band-based methods.

DNA fragment sizing is arguably the most widely used analytical method in molecular biology, biochemistry, and microbiology. Specific applications of DNA fragment sizing include microbe identification and discrimination, genotyping, and sequencing. Traditionally, DNA fragments are characterized via size-dependent separation methods such as gel electrophoresis.

Conventional gel electrophoresis, using a solid polymer (e.g., agarose or polyacrylamide) and a static electric field, is ubiquitous in today's molecular biology and biochemistry labs. Such methods are routinely used to separate and size DNA fragments of ≤ 20 kb (4).

Pulsed-field gel electrophoresis (PFGE), developed in 1984 by Schwartz et al. (44), extended the fragment sizing range to over 1 Mb (4, 6, 9, 19, 44). For PFGE, large DNA fragments (i.e., >10 kb) are separated in an agarose gel by a pulsed electric field. A popular application of PFGE has been strain-level bacterial fingerprinting through the sizing of DNA fragments resulting from the digestion of whole genomic DNAs with rare-cutting restriction endonucleases (5, 11, 28, 35, 47, 49). Macrorestriction-based bacterial fingerprinting has found applications primarily in the public health and food safety industries. For example, the Centers for Disease Control and Prevention (CDC) have formed PulseNet, the National Molecular Subtyping Network for Foodborne Disease Surveillance (46; <http://www.cdc.gov/pulsenet>). PulseNet utilizes mac-

rorestriction fingerprinting of bacteria and a database of PFGE fingerprints to provide a means for the early recognition, timely investigation, and reduction of food-borne disease outbreaks across the nation.

While PFGE is generally applicable for routine bacterial identification, it has practical limitations. For instance, PFGE requires relatively large quantities of DNA (≥ 200 ng of DNA or $\sim 10^7$ bacterial cells [33] or at least 1 ng of DNA per band [42]). Typically, bacterial samples must be cultured to obtain this amount of DNA (culturing can require up to 48 h and some bacteria cannot be cultured). PFGE is also constrained by extremely slow analysis (e.g., times of >20 h/gel). Rapid bacterial analysis by restriction fragment sizing is impossible because the total preparation and analysis times for PFGE are typically >40 h.

Alternatives to gel electrophoresis techniques have emerged in an effort to provide accurate DNA fragment sizing while addressing the limitations of these gel-based methods. Many of the alternatives rely on separation-based sizing like that employed by gel electrophoresis. Examples of alternative separation-based DNA fragment sizing methods include microfluidic entropic traps (24, 25) as well as capillary and microchannel electrophoresis (13, 14, 18, 21, 22, 34). These methods considerably reduce the amount of DNA required for analysis and provide shortened analysis times relative to traditional gel electrophoresis methods. However, these techniques have not demonstrated the fragment sizing range of PFGE that is required for bacterial fingerprinting. Advances in mass spectrometry (MS) have provided yet another analytical alternative that is capable of sizing DNA fragments (2, 48). While MS techniques can provide sequence-level information and have

* Corresponding author. Mailing address: Mail Stop M888, Bioscience Division, Los Alamos National Laboratory, Los Alamos, NM 87545. Phone: (505) 667-3279. Fax: (505) 665-3024. E-mail: blm@lanl.gov.

superior accuracy over gel electrophoresis methods (e.g., <0.01% error) (48), they are restricted to relatively small fragment sizes and have not demonstrated the capability of bacterial discrimination that PFGE provides.

Another DNA fragment sizing technique, which does not rely on size-dependent fragment separation, has been developed. This method utilizes fluorescence-based, single-molecule detection technology to size individual DNA fragments in flowing streams (1, 7, 8, 15, 20, 23, 43). Through the use of intercalating dyes that selectively bind to DNA, the fragment length can be determined from its fluorescence intensity (50). Our lab has developed a high-sensitivity flow cytometer (FCM) that utilizes this technology to size DNA fragments as they pass through a laser beam (1, 20, 23, 29–31, 33, 39, 50). Fragments ranging from 0.125 to 500 kb have been sized by this technique, and strain-level bacterial differentiation has been demonstrated (23, 29, 31, 33). Bacterial fingerprints by FCM are obtained in approximately 30 min from $\sim 10^3$ cells. FCM techniques are approximately 50 times faster and require $\sim 200,000$ times less DNA than PFGE.

Our FCM technology has progressed to the point that it warrants a thorough and direct quantitative comparison with PFGE. Here we report a comparative performance assessment of FCM and PFGE accuracy and precision for an analysis of bacterial fingerprinting based on macrorestriction fragments. A sequenced organism, *Staphylococcus aureus* Mu50, was chosen to provide absolute fragment sizes as needed to properly assess accuracy. The precision and accuracy associated with FCM and PFGE were quantitatively evaluated based on the analysis of 45 replicates of Mu50 genomic DNA digested with the restriction endonuclease *Sma*I. We correlated precision with sample preparation and experimental parameters in order to evaluate which, if any, of these variables limits the fragment-sizing precision. Additional considerations, including sensitivity, analysis time, throughput, cost, and availability are also discussed in this comparative assessment of bacterial fingerprinting techniques.

MATERIALS AND METHODS

PFGE and FCM sample preparation. (i) Preparation of agarose plugs. A vancomycin-resistant strain of *S. aureus* (Mu50) was obtained from the American Type Culture Collection (ATCC 700699; 2,878,084-kb genome [27]). Mu50 cells were grown in tryptic soy broth overnight at 37°C with aeration before being collected by centrifugation for 10 min at 4°C and 3,700 rpm in a tabletop refrigerated centrifuge (model GS-6R; Beckman Instruments, Inc., Fullerton, Calif.). Cells were washed in 5 ml of PET-IV buffer (10 mM Tris-HCl [Sigma, St. Louis, Mo.] and 1 M NaCl [Sigma] at pH 7.5) and then resuspended in PET-IV buffer to achieve a final concentration of $\sim 1.6 \times 10^9$ cells/ml. Cell concentrations were determined by measuring the optical densities of cultures at 610 nm. Clinical *S. aureus* isolates (nares strains obtained from patients starting maintenance hemodialysis at the Algemeen Ziekenhuis St.-Jan Hospital in Bruges, Belgium [3]) were prepared in the same manner.

The diluted cells were mixed with an equal volume of molten (55°C) 1.5% low-melting-point agarose (InCert; FMC BioProducts, Rockland, Maine) prepared in EC buffer (6 mM Tris-HCl, 1 M NaCl, 100 mM EDTA [Sigma], 0.5% Brij 58 [Sigma], 0.2% deoxycholate [Sigma], 0.5% *N*-lauroylsarcosine [Sigma], pH 7.6). Aliquots (100 μ l) of this mixture were then formed into individual agarose plugs, with each containing $\sim 1.6 \times 10^8$ Mu50 cells, using commercially available plug molds (1.5 $\times$ 5 by 10 mm; Bio-Rad Laboratories, Inc., Hercules, Calif.).

(ii) Cell lysis. The *S. aureus* Mu50 cells embedded in agarose plugs were lysed with egg white lysozyme (1 mg/ml [Sigma]) and lysostaphin (0.1 mg/ml [Sigma]) at 37°C for 105 min with gentle shaking in EC buffer containing RNase A (20 μ g/ml [Sigma]). Lysis was followed by a heat shock at 62°C for 15 min. The

plugs were then treated with proteinase K (1 mg/ml [Fisher Scientific, Fair Lawn, N.J.]) at 55°C for 60 min in ES buffer (0.5 M EDTA, 1% sodium laurylsarcosine, pH 8.5). Any proteinase K remaining in the plug was inactivated with 1 mM phenylmethylsulfonyl fluoride (Sigma) at 37°C for 30 min. Plugs were washed two times with 18 M Ω H₂O (25°C, 20 min) and twice (25°C, 20 min) with 1 $\times$ TE buffer (10 mM Tris-HCl, 1 mM EDTA at pH 8.0). The final washing buffer was decanted and the plugs were resuspended in TE buffer and stored at 4°C for up to 4 weeks with no observed degradation in results.

(iii) Restriction enzyme digestion. The *S. aureus* Mu50 genomic DNA remaining in the plugs after cell lysis was digested with the *Sma*I restriction enzyme (New England Biolabs, Beverly, Mass.). This restriction endonuclease was chosen because it results in a manageable number of restriction fragments (25 cut sites within the *S. aureus* Mu50 genome). Restriction was carried out with 40 U of *Sma*I for 2 h under the manufacturer's recommended conditions.

Samples (i.e., plugs) prepared according to the procedures described above were analyzed by both PFGE and FCM performed according to the methods described below. The common preparation steps described here have been optimized and, excluding culturing time, require only 6 h (H. Cai, C. J. Lemanski, J. R. Hakovirta, X. Yan, J. H. Jett, R. H. Keller, and B. L. Marrone, unpublished data). This represents a time savings of approximately 18 h compared to the methods outlined by the CDC for their PulseNet analysis.

PFGE. After the lysis and restriction steps, agarose plugs containing *S. aureus* Mu50 cells were cut into six sections (1.5 $\times$ 5 by ~ 1.5 mm, with each containing $\sim 2.7 \times 10^7$ cells). Plug sections were loaded and aligned inside the wells of a 1% SeaKem Gold agarose slab (FMC BioProducts) prepared in 0.5 $\times$ Tris-borate-EDTA buffer. A total of 15 samples were loaded on each gel along with four equally distributed aliquots of a sizing standard (low-range PFGE marker; New England Biolabs). Gel wells were sealed with a small amount of molten agarose and run on the CHEF-Dr II system (Bio-Rad Laboratories) with the following running parameters: 200 V, 2.2-s initial switching time, 54.2-s final switching time, and 22 h of total running time at 14°C.

After the separation step, agarose gels were stained with ethidium bromide (Fisher Scientific) and digitally photographed with a Gel-Pro Imager (Media Cybernetics, Carlsbad, Calif.).

Flow cytometry. (i) Sample preparation. To facilitate the flow cytometry fragment sizing of restriction fragments embedded in agarose plugs, we needed to liquefy the plug. This was accomplished by use of GELase enzyme (Epicentre, Madison, Wis.) as described below.

After restriction enzyme digestion, $\sim 1/6$ of an agarose plug ($\sim 2.7 \times 10^7$ cells) was equilibrated twice for 30 min in 500 μ l of the freshly diluted, ice-cold 1 $\times$ GELase reaction buffer (40 mM bis-Tris, 50 mM NaCl at pH 6.0) provided with the enzyme. The plug portion was placed in a 0.5-ml microcentrifuge tube with 5 μ l of 500 mM NaCl and heated at 70°C for 5 min to melt the agarose before being transferred to 45°C for 5 min. Freshly diluted GELase enzyme (0.5 U in 1 $\times$ GELase buffer) was added to the sample, which was then incubated at 45°C for 60 min. Samples were treated with GELase within 8 h of staining and FCM analysis. At this stage, the restriction fragments in solution were very susceptible to shearing by mechanical force and were handled only with wide-bore pipettes and without vortexing.

The staining of GELase-treated samples for FCM analysis was accomplished with SYTOX Orange nucleic acid stain (Molecular Probes, Inc., Eugene, Oreg.). One hundred forty microliters of TTP buffer (10 mM Tris-HCl, 0.05% Tween 20 [Sigma], 0.5% polyvinylpyrrolidone [molecular weight of 10^6] [Polysciences, Inc., Warrington, Pa.], pH 8.0) and 20 μ l of diluted SYTOX Orange stain (1:1,000 dilution of a 5 mM stock) were carefully added to each GELase-treated bacterial DNA sample (~ 40 μ l) and allowed to equilibrate for 2 min prior to analysis. Stained samples were diluted 10- to 30-fold in TTP buffer ($\sim 10^9$ DNA fragments/ml) prior to analysis to minimize the simultaneous detection of two or more DNA fragments.

(ii) Flow cytometry. An ultrasensitive flow cytometer, designed and built at Los Alamos National Laboratory, was used to measure DNA fragment sizes. This instrument has been described previously (23, 50, 51).

The data obtained directly from the FCM instrument are burst area (total detected photons) distribution histograms. A calibration procedure using sizing standards is required to accurately relate the burst area to the fragment length in kilobases. While the fluorescence intensity obtained from a stained DNA fragment depends primarily on the fragment size, the DNA concentration can also affect the measured fluorescence intensity (51). Since the DNA concentration in many samples was only approximately known, we used a set of three equally represented internal standards to calibrate the fluorescence intensity and provide an accurate internal calibration for each sample. These standards were (i) a linearized plasmid (17.4 kb) (provided by Mark MacInnes, Bioscience Division, Los Alamos National Laboratory), (ii) lambda DNA (48,502 kb) (Promega

Corp.), and (iii) bacteriophage T4GT7 DNA (165.6 kb [i.e., T4 DNA with a 3.255-kb deletion between the sites at 165.251 and 168.506 kb]) (Wako/Nippon Gene, Toyama, Japan).

For each sample, a data set was first collected from the bacterial sample and then a second data set was collected after the addition of prestained internal standards, with time for the mixture to equilibrate. Histograms were fit with a fifth-order polynomial to estimate the background produced by random length fragments, which was subtracted prior to further data processing. A visual overlay of the two data sets readily identified histogram peaks corresponding to the internal standards, and each internal standard's position was numerically determined by using the centroids of Gaussian fits for the standard peaks in the histogram data. The resulting calibration slope and intercept (from a linear regression fit to the burst area peak values versus the known standards' sizes) were then applied to the data set containing the three standards and the analyte. The histogram from the corresponding sample, without calibration standards, was then aligned with the calibrated histogram by using a Pearson cross-correlation function (17). The calibrated scale was transferred to the original data set without internal standards. Once data were collected from the sample, with and without internal standards, this calibration process required <5 min per sample.

Comparative data processing. A commercially available relational database software package, BioNumerics (version 3.00; Applied Maths, Austin, Tex.), was used to process, analyze, and compare PFGE gel images and FCM data sets.

(i) **PFGE data.** Inverted, 16-bit, grayscale images of PFGE gels, each containing 15 sample and 4 marker lanes, were imported into BioNumerics for processing and analysis. Image tone curves were adjusted by using an offset linear curve algorithm to optimize contrast prior to a Fourier analysis of pixel intensity values. This spectral analysis provides an estimate of the background that is used to determine the disk size applied in subsequent "rolling disk" background subtractions and also provides parameters for Wiener (40) noise filtering. Additionally, median and spot removal filters were applied to each image before further processing or analysis.

Adjusted PFGE gel images were normalized by using four equally spaced reference lanes containing low-range PFGE molecular weight markers. Reference bands from 9.42 to 533.4 kb were identified and marked in each of the reference lanes by the software's snap-to-peak algorithm. A cubic-spline fit for the assignment of these band positions in each reference lane was used to normalize and calibrate each gel.

Normalized and calibrated PFGE gels were analyzed by both band and densitometric curve-based methods. Bands corresponding to known restriction fragments between 18 and 550 kb were identified within each gel lane by the BioNumerics snap-to-peak function. Average band positions, absolute sizing errors, and relative standard deviations (RSD) were calculated for each band by using exported band position information and simple arithmetic methods.

Band-based dendrograms were produced by using Dice coefficients (10) and an unweighted pair group method using arithmetic averages (UPGMA) (47). Default clustering settings of 0.50% optimization (i.e., the relative distance an entire lane is allowed to shift in matching attempts) and 1.00% band position tolerance (i.e., the relative distance a single band within a lane is allowed to shift during matching attempts) were used.

Curve-based dendrograms were created by using Pearson product-moment correlation coefficients (38) and a UPGMA clustering method. This curve-based method is insensitive to relative differences in intensity and background but is sensitive to local differences in background. An optimization setting of 0.50% was also used for the creation of curve-based dendrograms.

(ii) **FCM data.** FCM-generated fingerprint histograms were imported into BioNumerics by use of a custom script provided by Applied Maths. This script allows delimited text files, exported from the FCM data processing software, to be converted into a "histogel" format (i.e., analogous to a gel image). Histogels created from calibrated FCM data sets were processed and analyzed by the methods described above for the PFGE data, and both band and curve-based dendrograms were created from the FCM histogels using the same parameters that were used for the PFGE data analysis.

RESULTS AND DISCUSSION

Study overview. This study was designed to test the accuracy and precision of DNA fragment sizing by both PFGE and FCM while correlating the results to critical steps in the sample preparation and analysis processes. Specifically, 45 replicates of *S. aureus* Mu50, each digested with SmaI, were analyzed by both techniques. Further details regarding these samples are

given below. This comparative performance assessment of PFGE and FCM was done in the context of each method's ability to provide reliable macrorestriction fingerprints.

Eighteen agarose plugs, intended to be identical, were made from a single bacterial culture. The plugs were evenly divided into three tubes (designated L1 to L3), which were processed through the cell lysis, protein digestion, and cleanup steps in parallel. The six plugs in each lysis batch were then individually digested with the SmaI restriction enzyme (each digestion batch was designated D1 to D6). Following restriction digestion, each plug was cut into six equally sized pieces. Plug portions were loaded and run in adjacent lanes of one of three PFGE gels (designated G1 to G3). Either two or three portions of each plug were analyzed by PFGE (replicates were designated R1 to R3), giving a total of 45 SmaI-digested Mu50 replicates that were analyzed by PFGE. Thus, the PFGE results can be correlated to cell lysis batch (L), restriction digestion batch (D), gel number (G), and replicate number (R).

One portion of each plug was digested with GELase, stained, and diluted to a final fragment concentration of $\sim 10^9$ fragments/ml for analysis by FCM. Five replicates of each GELase-treated sample (designated R1 to R5) were analyzed. A total of 45 SmaI-digested Mu50 samples were analyzed by FCM and tracked according to cell lysis batch (L), restriction digestion batch (D), and replicate number (R).

GELase procedure. The main purpose of the agarose plug preparation was to reduce the shearing of DNA during lysis and restriction. However, FCM analysis of DNA fragments requires an aqueous solution. It is thus necessary for the restriction fragments within the agarose plugs to be extracted to facilitate FCM analysis. In the past, members of our laboratory used electroelution to accomplish this task (29, 31). However, electroelution is time-consuming (i.e., it requires >12 h) (29). In efforts to reduce the sample preparation time, we adopted an alternative procedure, described in detail above, to solubilize the agarose plugs containing the restriction fragments. This procedure uses the GELase enzyme and is approximately six times faster than electroelution.

To determine if the GELase procedure altered the restriction fragments, we selected a single plug containing SmaI-digested Mu50 restriction fragments and used PFGE to analyze two portions of the plug before and after GELase treatment. The PFGE results for both samples (data not shown) indicated that the observed bands, as well as the overall background, were consistent between the two samples (i.e., 13 of 13 bands matched when a 1% band tolerance setting was used). These results demonstrated that the GELase procedure does not alter the restriction fragments or the resulting DNA fingerprint. Thus, GELase treatment is a reliable and effective method for obtaining a liquid sample from agarose plugs for FCM analysis.

PFGE results. The 45 SmaI-digested *S. aureus* Mu50 replicates were analyzed over a 2-day period in three PFGE gels under identical separation conditions. A representative sample (L3D4G2R2) is shown in Fig. 1 as both a calibrated gel lane image and a densitometric curve which was obtained from the gel lane. A virtual digest histogram for SmaI-digested Mu50 is also shown in Fig. 1 for reference. The fragments comprising the virtual digest histogram were determined from the organism's published sequence (accession no. BA 000017) (32) and

FIG. 1. Comparison of PFGE and FCM RFLP fingerprints of *SmaI*-digested *S. aureus* Mu50. Representative data sets for PFGE and FCM are shown, with each data set represented as both a densitometric curve (or histogram) and a gel (or histogel) lane. Additionally, a histogram depicting the virtual digest fragment sizes (≥ 3.104 kb) is shown for reference. The PFGE and FCM peaks and bands at approximately 23, 58, and 110 kb are attributed to unresolved RFLP fragments (i.e., doublets), while the remaining peaks have been assigned to single RFLP fragments. Histograms have been vertically offset and normalized to allow for simultaneous viewing. The aspect ratios of the gel lane images were modified to match the histogram distributions and visually aligned with the curves for demonstration purposes.

the known recognition site of the *SmaI* restriction enzyme. The virtual digest, produced by using Kodon software (Applied Maths, Austin, Tex.), predicted that 25 restriction fragments, ranging from 2.966 to 537.122 kb, would be produced when

Mu50 DNA was digested with *SmaI*. All of the virtual digest fragments of ≥ 3.104 kb ($n = 20$) are included in the virtual digest histogram displayed in Fig. 1.

The majority of *SmaI*-digested Mu50 replicates analyzed by PFGE produced a total of 13 bands. Band positions were determined for each of the 45 samples, and the average band sizes, in kilobases, are reported in Table 1 along with the RSD of each band. The smallest detected PFGE fragment (19.873 kb) could not be confidently identified for six of the samples run in G1 but was seen for all samples run in G2 and G3.

The 13 bands observed in PFGE fingerprints of *SmaI*-digested Mu50 were attributed to 16 virtual digest fragments, of between 19.873 and 537.283 kb. The band assignments are indicated in Table 1. Three of the PFGE bands (23.43, 54.99, and 115.97 kb) were attributed to doublets corresponding to unresolved restriction fragments. The relative intensities of these three bands supported these assignments, and the averages of the unresolved fragments are reported in Table 1. We should note that a restriction fragment predicted by the virtual digest (77.388 kb) was only sporadically observed visually in the PFGE gels and was never identified by the image analysis software with the settings used in this study. Therefore, we chose to exclude this fragment from the remaining analyses. Except for the omission of this 77-kb fragment, each of the observed PFGE bands was attributed to a restriction fragment predicted by the genomic sequence.

FCM results. The 45 *SmaI*-digested *S. aureus* Mu50 replicates examined by FCM, along with each sample's corresponding calibration set, were analyzed over a period of 9 days. As discussed above, each FCM sample was calibrated by using a

TABLE 1. Accuracy and precision of FCM and PFGE fragment sizing

Known fragment size (kb) ^a	PFGE fragment sizing results				FCM fragment sizing results			
	<i>N</i>	Average size (kb)	% RSD ^b	% Error ^c	<i>N</i>	Average size (kb)	% RSD ^b	% Error ^c
537.283	45	562.29	0.94	4.7				
495.734	45	532.19	1.2	7.4				
320.685	45	341.93	1.2	6.6				
314.119	45	327.54	1.4	4.3	45	313.59	1.4	0.17
269.842	45	283.78	1.3	5.2	45	263.42	1.5	2.4
216.623	45	226.54	1.4	4.6	45	189.20	0.82	13
127.148	45	132.94	1.5	4.6	45	125.89	0.53	0.99
110.165 ^d	45	115.97	1.9	5.3	45	111.52	0.53	1.2
57.601 ^e	45	54.99	3.9	4.5	45	58.68	0.53	1.9
41.365	45	37.40	5.3	9.6	45	42.15	0.77	1.9
32.506	45	31.49	6.3	3.1	45	35.26	0.70	0.76
22.969 ^f	45	23.43	3.3	2.0	45	24.68	0.85	7.5
19.873	39	18.10	3.3	8.9	45	21.96	1.0	10
16.294					45	16.09	1.0	1.3
9.445					45	9.26	2.1	1.9
3.104					43	2.97	3.6	4.2

^a Known fragment sizes (≥ 3.104 kb), excluding a 77.388-kb fragment that was not consistently observed by either PFGE or FCM, were determined from the published genomic sequence of *S. aureus* Mu50 and a virtual restriction enzyme digest with *SmaI*.

^b % RSD = $\left(\frac{\sigma}{\mu}\right) \times 100$, where μ and σ are the mean and standard deviation of the measured fragment sizes (in kilobases), respectively. The means $\pm$ standard deviations were as follows: for PFGE, 3 ± 2 ; and for FCM, 1.2 ± 0.8 .

^c % Error = $\left(\frac{|\mu - x_c|}{x_c}\right) \times 100$, where μ and x_c are the experimentally measured mean and calculated fragment sizes (in kilobases), respectively. The means $\pm$ standard deviations were as follows: for PFGE, 5 ± 2 ; and for FCM, 4 ± 4 .

^d The reported fragment size is the average of two unresolved restriction fragments (111.887 and 108.443 kb).

^e The reported fragment size is the average of two unresolved restriction fragments (58.740 and 56.45 kb).

^f The reported fragment size is the average of two unresolved restriction fragments (23.881 and 22.057 kb).

FIG. 2. FCM-generated fragment size histograms corresponding to the *Sma*I-digested Mu50 sample L3D1R2. The histogram used for calibration (a) contains three internal standards (17.4, 48.5, and 165.6 kb) in addition to the bacterial restriction fragments. Peaks corresponding to the internal standards were fit with a sum of three Gaussian peaks, and the peak centroid values (triangles) were used to create the calibration curve shown in the overlay. The linear regression fit to this curve, shown as a dashed line, allowed the raw FCM histogram to be calibrated in kilobase units and was obtained with a correlation coefficient (R) of 0.999. The calibrated fragment size scale was transferred to the histogram of the sample without internal standards added (b) after alignment of the two histograms was optimized by cross-correlation.

portion of the sample spiked with three internal standards. Figure 2 shows an example of an FCM-generated fragment size histogram for *Sma*I-digested Mu50 (panel b) and its corresponding histogram with internal standards (panel a). The peaks attributed to the internal standards are identified with triangles and the size of each standard is given. Experimentally determined average burst area values (in detected photons) and known standard sizes (in kilobases) were used to create the inset calibration curve describing the linear relationship between the two quantities. The alignment of the two histograms shown in Fig. 2 was optimized by use of a Pearson cross-correlation algorithm, and the calibration was then transferred to the histogram shown in Fig. 2b. This process was repeated for each of the 45 *Sma*I-digested Mu50 samples.

A representative calibrated FCM sample (L3D1R4) is shown in Fig. 1 as both a fragment size histogram and a histogram lane image created from the histogram. A total of 13 peaks were quantified for all but two of the FCM data sets, for which the intensity did not permit confident quantification of the smallest (3.104 kb) restriction fragment detected by FCM. A summary of the peak statistics determined by FCM is given in Table 1.

As was done with the PFGE data, the 13 fragments quantified by FCM were attributed to 16 virtual digest fragments, and the same fragments that were unresolved by PFGE were also unresolved by FCM. However, the range of virtual digest fragments (3.104 to 314.119 kb) assigned to the observed FCM fragments was shifted toward smaller fragment sizes than those for PFGE. The 77.388-kb fragment that was predicted by the virtual digest, but sporadically observed by PFGE, was also

only weakly and inconsistently apparent visually in the FCM data sets. The band assignments, shown in Table 1, for each technique are supported by agreement with each other and the virtual digest.

Accuracy. Because we were interested in assessing accuracy, we chose a sequenced pathogen for this study (Mu50), which provided absolute fragment sizes to analytically evaluate the fragment sizes measured by PFGE and FCM. This unique approach allowed us to independently evaluate the accuracy of each method. The accuracy (i.e., the percent error) of fragment sizing results obtained by PFGE and FCM was evaluated by comparing restriction fragment sizes calculated from the DNA sequence and its *Sma*I virtual digest. The average fragment sizes determined by each technique are shown in Table 1 along with corresponding assignments to virtual digest fragments and corresponding sizing errors. The average sizing error associated with PFGE measurements was $5\% \pm 2\%$, while that associated with FCM measurements was $4\% \pm 4\%$. Given these values, the sizing errors for PFGE and FCM were equivalent within the measured uncertainty.

Previous reports place the uncertainty of PFGE fragment sizing near 10% (29, 31, 33). However, a recent study by Duck et al., who did not use sequence-determined fragment sizes for reference, showed an average fragment sizing error of 27% (11). The accuracy of PFGE fragment sizing determined in this study ($5\% \pm 2\%$) represents an improvement of approximately two- to fivefold over most previous estimates. Operator skill and computer-assisted data analysis have previously been attributed with having the largest influence on PFGE accuracy (37). Prior reports of fragment sizing errors for single-mole-

cule FCM varied from 2 to 10% depending on fragment size (29, 31, 33). The FCM error measured here ($4\% \pm 4\%$) is in agreement with these previous studies, and any enhancement in FCM accuracy can be attributed to an improved calibration system and several improvements in the data analysis software.

The largest observed error, for either method, was the underestimation of the 216.623-kb fragment by FCM (13% error). This discrepancy is noteworthy given the relatively accurate sizing of this fragment by PFGE (4.6% error) and the linear response of our FCM method in previous studies (20, 29, 31, 33, 39). In an attempt to understand this inconsistency, we examined the sequence of this 216-kb restriction fragment more closely. The only apparent anomalous feature of this fragment is that it spans the genome's origin of replication. While this fact alone cannot explain the observed FCM results, it may provide a clue to this seemingly unique discrepancy.

Precision. The simplest measure of precision for PFGE and FCM is the consistency of the restriction fragment sizes assigned among the 45 replicate samples analyzed by each technique. This precision was gauged by using the RSD values of the measured fragment sizes (a lower RSD value indicates superior precision). The RSD values for each fragment are shown in Table 1 along with the average RSD value associated with each sizing method. The mean RSD values associated with FCM and PFGE were $1.2\% \pm 0.8\%$ and $3\% \pm 2\%$, respectively. These values are again in agreement with previously reported literature values (11, 20, 23, 29, 31, 33). While the mean numerical precision of FCM was slightly better than that of PFGE, the two values were equivalent, within 1 standard deviation. This result indicates that the overall numerical levels of precision of FCM and PFGE were roughly equivalent based on the examination of replicate samples analyzed under identical conditions over multiple days and in multiple gels.

Because macrorestriction fingerprints are commonly used to create phylogenetic comparisons in the form of dendrograms, we also chose to evaluate the precision of FCM and PFGE by using a dendrogram-derived method. Generally, dendrograms created from macrorestriction fingerprints contain data from multiple strains and/or species of bacteria and provide an estimation of the genetic similarities of the samples (for an example, see reference 28). Quantitative statistical measures indicating the significance of the groups within a dendrogram, as well as the overall dendrogram structure, can be obtained. For this study, we created dendrograms relating the 45 replicate samples analyzed by each fingerprinting method and used these dendrograms to evaluate the precision of fragment sizing by each technique. This dendrogram-derived method of evaluating precision is somewhat unorthodox, but it provides a quantitative measure of precision in a context that is relevant to the way in which macrorestriction fingerprint data are typically used and interpreted. These dendrograms also allowed us to easily determine if the precision of either method was limited by any of the sample preparation or analysis steps that were tracked in this study.

As stated above, the grouping of macrorestriction fingerprints based on the similarities of their features creates a dendrogram. More specifically, each fingerprint is compared to all others in a given data set to create a matrix of correlation coefficients determined from one of many possible correlation methods (e.g., Dice, Ochiai, Jaccard, or Pearson product-mo-

ment correlation). Another set of algorithms (e.g., UPGMA, Ward, or neighbor joining) is then used to create a dendrogram from the matrix of correlation coefficients. The number of fingerprint features and the algorithms utilized in the clustering drastically affect the resulting dendrogram (16).

For macrorestriction fingerprinting by PFGE, the most common clustering method used to create dendrograms is the Dice-UPGMA combination. The CDC uses this method in their PulseNet system (41, 46). The Dice correlation (10) considers the presence or absence of PFGE bands that have been identified in the fingerprints to create the correlation matrix used by the UPGMA algorithm (45). The Dice correlation and other band-based correlations use binary (e.g., the band is present or absent) comparisons, and the resulting correlation coefficients depend on the number of bands identified as well as the number of common bands between two fingerprints. Despite the popularity of this clustering method, there are drawbacks that should be considered. First, band-based methods use only a small portion of the available data. Because of the use of only the average band sizes, the majority of the data (i.e., band shapes and intensities) are neglected. Additionally, band values used in these comparisons are determined by user-defined parameters (28). Because the user is able to determine what is or is not a band, there may be an introduction of an irreproducible bias that affects the resulting dendrogram.

An alternate method of creating dendrograms from macrorestriction fingerprints is a Pearson-UPGMA method. The Pearson product-moment correlation (38) uses the entire data set by correlating points along the densitometric curves obtained from each sample. This and other curve-based methods utilize more features of a data set than do band-based algorithms. With more possible states at each comparison point, the average correlation coefficients as well as the resulting dendrogram's similarity values will be lower (16). Regardless of these lower similarity values, evidence suggests that the Pearson-UPGMA clustering method is superior to band-based methods for analyzing DNA fingerprints (26). Additionally, curve-based clustering methods do not require user-defined input that could be biased.

A dendrogram produced by any of the clustering methods discussed above provides a graphical representation of sample similarity. Each node (i.e., the point at which samples intersect in the dendrogram) describes the average similarity of the samples contributing to that node. Error flags on each node represent $\pm 1\sigma$ value from the average similarity. The σ values used to make error flags are determined by reconstructing similarity values from the dendrogram branch defined by the node of interest and comparing these values with the original correlation coefficient matrix. For this study, the node that joined all replicate samples was used to judge the reproducibility of PFGE and FCM. All dendrograms presented below were intentionally produced with BioNumeric's default clustering settings (e.g., 0.5% optimization and 1.0% band position tolerance). While these settings are arbitrary, holding them constant allowed for direct comparisons of PFGE and FCM reproducibility.

A band-based dendrogram produced from Dice-UPGMA clustering of PFGE results for the 45 replicate SmaI-digested Mu50 samples and two unrelated clinical *S. aureus* samples is shown in Fig. 3. In this dendrogram, many of the replicate

FIG. 3. Band-based PFGE dendrogram for 45 replicates of SmaI-digested Mu50 and two clinical *S. aureus* isolates. The dendrogram shows the corresponding gel lane image for each sample beside its key, which indicates the experimental parameters. The error flags at each node indicate $\pm 1 \sigma$ from the average similarity, and cophenetic coefficients are provided, in parentheses, for important nodes.

samples are considered identical (i.e., 100% similarity) to at least one other sample. However, the 45 samples are divided into seven groups, with each group containing 1 to 23 samples which are not considered identical. The node that joins all 45 samples, which has a similarity of $93\% \pm 6\%$, quantifies the reproducibility of macrorestriction fingerprinting by PFGE. While this quantity is not an absolute measure comparable with the RSD values in Table 1, it is directly comparable to the corresponding value obtained from the band-based clustering of the FCM data discussed below. Figure 3 also includes the SmaI macrorestriction fingerprints of two clinically isolated *S. aureus* samples for comparison purposes. The dendrogram groups these two clinical isolates together, while the node that joins these samples with the 45 SmaI-digested Mu50 replicates has an average similarity of only $61\% \pm 9\%$. This illustrates that strain-level discrimination is possible while simultaneously evaluating the precision of discrimination within a single strain.

The dendrogram produced from band-based clustering (Dice-UPGMA) of FCM results, using the same optimization and band tolerance settings as were used for the PFGE data, is shown in Fig. 4. The dendrogram shows that all 45 SmaI-digested Mu50 replicates are considered identical. Again, the two clinical samples were considered identical to each other while being clearly discriminated from the 45 SmaI-digested Mu50 replicates. The average similarity of the node joining the replicates with the clinical isolates was $76.6\% \pm 0.6\%$. These results show that our FCM technique is absolutely (i.e., 100%)

reproducible when using band-based clustering methods. However, it is clear from the histogel lanes shown in Fig. 4 that the 45 replicate SmaI-digested Mu50 data sets are not identical. These results illustrate that the band-based clustering of DNA fingerprint data can, in some instances, be misleading and provide a higher sense of certainty than is actually present in the data. Curve-based clustering methods, discussed below, address these problems.

Figures 5 and 6 show the curve-based dendrograms corresponding to the same samples shown in Fig. 3 and 4, respectively. Curve-based dendrograms were created by using the Pearson product-moment correlation and the UPGMA clustering algorithm. Key differences are apparent when the band- and curve-based dendrograms are compared.

First, no two samples are considered identical when using the Pearson-UPGMA clustering method. The additional data that are used in curve-based clustering methods provide additional details that are not utilized in band-based methods. The Pearson-UPGMA clustering method creates a dendrogram that shows explicitly and quantitatively just how similar replicates are without calling any two samples identical. The nodes connecting all of the Mu50 replicates in the PFGE and FCM dendrograms have average similarity values of $92\% \pm 2\%$ and $91\% \pm 4\%$, respectively. Within the measured uncertainty, these values are equivalent. While the curve-based clustering results for these Mu50 replicates agree with the RSD values and indicate that the levels of precision of the PFGE and FCM

FIG. 4. Band-based FCM dendrogram for 45 replicates of SmaI-digested Mu50 and two clinical *S. aureus* isolates.

methods are equivalent, band-based clustering methods suggest that FCM is more precise.

Second, the clinical *S. aureus* isolates are again clearly discriminated from the Mu50 replicates in the curve-based dendrograms. This indicates that strain-level bacterial discrimination is accomplished by both the Dice-UPGMA and Pearson-UPGMA clustering methods. For this particular data set, band-based methods do not appear to provide an advantage in strain discrimination. When replicate samples are included, the curve-based clustering method provides the advantage of quantitatively evaluating precision, which can aid in the interpretation of the resulting dendrograms.

Third, the overall node similarities tend to be lower for the curve-based dendrograms. The nodes connecting the Mu50 replicate samples to the clinical *S. aureus* isolates in the curve-based PFGE and FCM dendrograms have average similarity values of $26\% \pm 6\%$ and $13\% \pm 3\%$, respectively. These values are substantially lower, as expected, than the corresponding values from the band-based dendrograms. The other significant node is that which connects all of the Mu50 replicates. For the FCM dendrograms, the average similarity of this node dropped from $100\% \pm 0\%$ for the Dice-UPGMA method to $91\% \pm 4\%$ for the Pearson-UPGMA method. The corresponding replicate joining node similarities for the PFGE dendrograms were $93\% \pm 6\%$ and $92\% \pm 2\%$.

Correlating precision with experimental steps. To determine if the curve-based dendrogram-derived precision of PFGE ($92\% \pm 2\%$) and/or FCM ($91\% \pm 4\%$) is limited by or

attributable to key sample preparation or handling steps, we tracked each replicate SmaI-digested Mu50 sample according to labels identifying its preparation and analysis history. For PFGE, the tracked steps included lysis batch, digestion batch, gel number, and gel replicate number. FCM samples were tracked according to lysis batch, digestion batch, and replicate number. Day-to-day variations were not tracked for the FCM samples because each sample was independently calibrated. In contrast, each gel calibration was applied to multiple lanes. Statistical methods were used to correlate these experimental parameters with the dendrograms discussed above. For example, the 45 SmaI-digested Mu50 replicates analyzed by both PFGE and FCM were lysed in three batches (i.e., L1, L2, and L3). Is there any correlation between the lysis batch number and the dendrogram groupings? If so, do batch differences limit precision?

A visual examination of curve-based dendrogram structures does not appear to indicate that any of the groupings within the 45 replicate samples is at all correlated with any one of the experimental parameters tracked in this study. Statistical methods were utilized to provide quantitative measures of the correlation between each experimental variable and the dendrogram groupings. The first method used to examine group statistics was a jackknife analysis (12, 36). For this, each sample was assigned to a group based on a single experimental variable (L, D, G, or R). The average percent similarity was then calculated for each group. Each sample was then removed from the data set, treated as an unknown, and matched to a

FIG. 5. Curve-based PFGE dendrogram for 45 replicates of SmaI-digested Mu50 and two clinical *S. aureus* isolates.

group based on average percent similarities. The average percentages of entries correctly matched to their assigned groups are reported in Table 2. A value closer to 100% indicates that the experimental variable has a larger effect on dendrogram groupings and the overall precision of each method. While a jackknife value of >90% indicates a strong correlation, only one of the values calculated for our replicate samples was above 50%. This value, 84%, indicated that gel-to-gel differences were the most significant limitation on PFGE precision within the variables examined in this study. However, variations between gels did not limit the PFGE precision (i.e., 84% is not considered significant, and samples within a single gel

showed more variability than those from two different gels in some cases).

The second statistical tool used to determine the correlation between experimental parameters and the experimentally determined precision of each method was the group violation method. This method evaluates the overlap of similarity values within and between assigned groups. For this method, the percentages of nonoverlaps, or nonviolations, were measured and are reported in Table 2. Again, the group violation values showed that all of the groups overlapped significantly and that there was no significant correlation between any one experimental parameter and the dendrogram groupings. The largest value, 75%, again indicated that the highest correlation linked the PFGE gel number to dendrogram structure and overall PFGE precision. Since none of the statistical values obtained by either the jackknife or group violation methods were significant, the overall precision of PFGE and FCM does not appear to be dependent on any of the experimental steps tracked in this study. It would then follow that the measured precision of PFGE and FCM is limited by data processing or some untracked experimental parameter.

Additional measures of dendrogram confidence. Dendrogram interpretations are often not guided by statistical measures. In addition to including replicate samples and using curve-based clustering methods, we applied a statistical tool to help with the interpretation of dendrograms summarizing the macrorestriction fingerprint data.

TABLE 2. Group statistics correlating Pearson-UPGMA-derived precision to sample preparation and analysis steps					
Analysis method	Grouping statistic	Grouping criterion (%)			
		Lysis	Digestion	Gel	Replicate
PFGE	Jackknife ^a	47	46	84	19
	Group violation ^b	35	40	75	29
FCM	Jackknife	49	22		11
	Group violation	36	23		17

^a Jackknife values are the average percentages of Mu50 samples matched to their correct group. Groups were assigned, according to only one grouping criterion, before each jackknife analysis.

^b The average percentage of nonviolators for each grouping criterion is reported.

FIG. 6. Curve-based FCM dendrogram for 45 replicates of *SmaI*-digested Mu50 and two clinical *S. aureus* isolates.

Cophenetic coefficients (45) provide a measure of cluster consistency. These coefficients are obtained by correlating the dendrogram-derived similarities and correlation matrix values. Cophenetic coefficients, which can be calculated for each dendrogram node, are shown in Fig. 3 through 6 for important nodes. First, coefficients for the dendrogram root (i.e., the node connecting all samples) provide a measure of the confidence for the overall structure. The root cophenetic values were 90 and 99% for the band-based dendrograms derived from the PFGE and FCM data, respectively. The corresponding coefficients for the curve-based dendrograms were both 99%. These values indicate that the overall dendrogram structure is valid and that the strain-level discrimination between the clinical *S. aureus* samples and the 45 Mu50 samples is significant.

Second, cophenetic coefficients for the nodes joining all 45 of the Mu50 replicates were examined. For the band- and curve-based PFGE dendrograms, these values were 50 and 69%, respectively. The corresponding values for the FCM nodes were 100 and 56%. Only one of these coefficients, the 100% coefficient from the band-based FCM data, indicated any significant degree of confidence in the dendrogram structure for the 45 replicates. This means that there is no meaningful relationship in the subgroups for the 45 samples. Since the replicate sample data sets were all very similar without being identical, this result was expected. These results also confirm the conclusion that experimental parameters were not

correlated with precision, as determined from the curve-based dendrograms.

Additional considerations. In assessing the relative performance of PFGE and FCM, factors such as sensitivity, analysis time, throughput, cost, and availability must be considered. The sensitivity (i.e., the amount of DNA required for analysis) was previously evaluated by Larson et al. (33), who determined that FCM is $\sim 200,000$ times more sensitive than PFGE. That is, only picogram quantities of DNA are required for FCM analysis, while $\geq \text{ng}$ of DNA is required for PFGE analysis.

Throughput is another factor to consider when evaluating analytical methods. Because the sample preparation time is roughly equivalent for both PFGE and FCM, the following estimates and discussion of throughput do not include sample preparation time. PFGE requires ~ 24 h of running time to resolve macrorestriction fragments. However, 15 samples are simultaneously run with calibration standards in that time. Therefore, the throughput of PFGE is estimated to be ~ 1.6 h/sample (i.e., 15 samples per 24 h). It should be noted, however, that 24 h is required to test even a single sample by PFGE. In contrast, FCM can collect and calibrate a macrorestriction fingerprint from a single sample in ~ 30 min. Therefore, FCM currently has a throughput advantage of approximately threefold. This is important for time-critical applications such as bioforensics.

Cost and availability are two more considerations when choosing an analytical method. PFGE setups, like the one used

in this study, are commercially available for ~\$6,000 (4). However, a digital imaging system (~\$11,500) and data processing software (BioNumerics was used for both PFGE and FCM data processing) are also necessary to realize the accuracy and precision measured in this study. While the ultrasensitive flow cytometer used in this study is one of a kind, its capabilities are available to the research community through the National Flow Cytometry Resource at Los Alamos National Laboratory. Therefore, PFGE currently has the advantage of being more accessible and affordable.

Conclusions. In this study, the accuracy and precision of bacterial macrorestriction fingerprinting by PFGE and FCM were quantitatively evaluated. Precision estimates, determined from Pearson-UPGMA dendrograms of replicate samples, were shown to agree with simple arithmetic calculations, and all quantitative measures of both accuracy and overall precision were, within error, equivalent for PFGE and FCM. The precision of FCM was found to have no significant correlation to any of the experimental parameters tracked in this study, and PFGE precision was slightly, but not significantly, dependent on gel-to-gel variations. The effective sizing range of PFGE is somewhat larger than that of FCM, while the latter is more adept at sizing smaller restriction fragments. Thus, both PFGE and FCM are reliable methods for producing bacterial macrorestriction fingerprints for the discrimination of bacteria at the species and strain levels.

Pearson-UPGMA clustering methods were shown to be more reliable and informative, for the limited data set used in this study, than band-based methods for predicting significant dendrogram groupings. Cophenetic coefficients were also shown to provide valuable interpretations of Pearson-UPGMA dendrograms that included sample replicates. We conclude that multiple replicates should be included when conducting macrorestriction fingerprinting experiments and that Pearson-UPGMA clustering of the data with cophenetic coefficients is helpful for interpreting the results accurately.

BioNumerics software, which has been adopted by the CDC in PulseNet to analyze PFGE data, was used here to analyze the FCM data. This allowed us to directly compare the data obtained from the PFGE and FCM approaches for sizing DNA fragments. Because the accuracy and precision of PFGE and FCM were comparable, macrorestriction fingerprints from FCM should be as reliable as those from PFGE for matching in a relational database such as BioNumerics.

Although both the FCM and PFGE approaches are comparable in terms of precision and accuracy, there are several advantages to the FCM approach. The first is the small amount of sample needed. In a diagnostic setting, this may allow fingerprints to be obtained more quickly from a clinical sample or a small culture. Another advantage of FCM is the speed of analysis per sample: 30 min versus 24 h. However, the fast sample analysis is counterbalanced by the relatively low throughput of the FCM. With PFGE, multiple lanes can be run in parallel. In contrast, only one sample can be run at a time on the FCM instrument, as it is currently configured with a single capillary. Currently in our lab, the agarose plug preparation is used for both PFGE and FCM, even though the FCM approach requires a relatively small amount of liquid sample. Our future plans call for scaling down the FCM sample preparation and making a transition away from the agarose plug

format and into a microfluidics format. We expect that this will reduce the sample preparation time per sample at least by half and will facilitate protocol standardization.

Currently, FCM provides an attractive alternative to PFGE for DNA fragment sizing. We believe that with continued improvements and enhancements, the FCM approach to fragment sizing will find more widespread use.

ACKNOWLEDGMENTS

We thank Johan R. Boelaert of Algemeen Ziekenhuis St-Jan in Belgium and Naomi Balaban of the University of California—Davis for the clinical isolates of *S. aureus* used in this study. We also thank Mark MacInnes of Los Alamos National Laboratory for providing the 17.4-kb plasmid.

This work was performed under the auspices of the U.S. Department of Energy (contract W-7405-ENG-36). Funding was provided by the DOE Office of Nonproliferation and National Security, the NIH National Center for Research Resources (grant RR-01315) for the National Flow Cytometry Resource, and the Federal Bureau of Investigation (FBI).

The statements and conclusions herein are those of the authors and do not necessarily represent the views of the FBI.

REFERENCES

1. Ambrose, W. P., H. Cai, P. M. Goodwin, J. H. Jett, R. C. Habbersett, E. J. Larson, W. K. Grace, J. H. Werner, and R. A. Keller. 2003. Flow cytometric sizing of DNA fragments, p. 239–270. In J. R. Lackowicz (ed.), Topics in fluorescence: DNA technology, vol. 7. Kluwer Academic/Plenum Publishers, New York, N.Y.
2. Bai, J., Y. H. Liu, D. M. Lubman, and D. Siemieniak. 1994. Matrix-assisted laser-desorption ionization mass-spectrometry of restriction enzyme-digested plasmid DNA using an active nafen substrate. *Rapid Commun. Mass Spectrom.* **8**:687–691.
3. Balaban, N., Y. Gov, A. Bitler, and J. R. Boelaert. 2003. Prevention of *Staphylococcus aureus* biofilm on dialysis catheters and adherence to human cells. *Kidney Int.* **63**:340–345.
4. Birren, B., and E. Lai. 1993. Pulsed field gel electrophoresis: a practical guide. Academic Press, Inc., San Diego, Calif.
5. Busch, U., and H. Nitschko. 1999. Methods for the differentiation of microorganisms. *J. Chromatogr. B* **722**:263–278.
6. Cantor, C. R., C. L. Smith, and M. K. Mathew. 1988. Pulsed-field gel electrophoresis of very large DNA molecules. *Annu. Rev. Biophys. Chem.* **17**:287–304.
7. Castro, A., F. R. Fairfield, and E. B. Shera. 1993. Fluorescence detection and size measurement of single DNA molecules. *Anal. Chem.* **65**:849–852.
8. Chou, H. P., C. Spence, A. Scherer, and S. Quake. 1999. A microfabricated device for sizing and sorting DNA molecules. *Proc. Natl. Acad. Sci. USA* **96**:11–13.
9. Chu, G., D. Vollrath, and R. W. Davis. 1986. Separation of large DNA molecules by contour-clamped homogeneous electric fields. *Science* **234**:1582–1585.
10. Dice, L. R. 1945. Measures of the amount of ecological association between species. *Ecology* **26**:297–302.
11. Duck, W. M., C. D. Steward, S. N. Banerjee, J. E. J. McGowan, and F. C. Tenover. 2003. Optimization of computer software settings improves accuracy of pulsed-field gel electrophoresis macrorestriction fragment pattern analysis. *J. Clin. Microbiol.* **41**:3035–3042.
12. Efron, B. 1979. 1977 Rietz Lecture. Bootstrap methods—another look at the jackknife. *Ann. Stat.* **7**:1–26.
13. Figey, D., and N. J. Dovichi. 1995. Multiple separations of DNA sequencing fragments with a non-cross-linked polyacrylamide-filled capillary: capillary electrophoresis at 300 V/cm. *J. Chromatogr. A* **717**:113–116.
14. Fister, J. C., S. C. Jacobson, L. M. Davis, and J. M. Ramsey. 1998. Counting single chromophore molecules for ultrasensitive analysis and separations on microchip devices. *Anal. Chem.* **70**:431–437.
15. Foquet, M., J. Korfach, W. Zipfel, W. W. Webb, and H. G. Craighead. 2002. DNA fragment sizing by single molecule detection in submicrometer-sized closed fluidic channels. *Anal. Chem.* **74**:1415–1422.
16. Forey, P. L., C. J. Humphries, I. L. Kitching, R. W. Scotland, D. J. Siebert, and D. M. Williams. 1992. Cladistics: a practical course in systematics. Oxford Press, New York, N.Y.
17. Fuller, W. A. 1995. Introduction to statistical time series, 2nd ed. John Wiley & Sons, New York, N.Y.
18. Gao, Q. F., and E. S. Yeung. 1998. A matrix for DNA separation: genotyping and sequencing using poly(vinylpyrrolidone) solution in uncoated capillaries. *Anal. Chem.* **70**:1382–1388.

19. Gardiner, K. 1991. Pulsed field gel electrophoresis. *Anal. Chem.* **63**:658–665.
20. Goodwin, P. M., M. E. Johnson, J. C. Martin, W. P. Ambrose, B. L. Marrone, J. H. Jett, and R. A. Keller. 1993. Rapid sizing of individual fluorescently stained fragments by flow-cytometry. *Nucleic Acids Res.* **21**:803–806.
21. Haab, B. B., and R. A. Mathies. 1997. Optimization of single-molecule fluorescence burst detection of ds-DNA: application to capillary electrophoresis separations of 100–1000 basepair fragments. *Appl. Spectrom.* **51**: 1579–1584.
22. Haab, B. B., and R. A. Mathies. 1999. Single-molecule detection of DNA separations in microfabricated capillary electrophoresis chips employing focused molecular streams. *Anal. Chem.* **71**:5137–5145.
23. Habbersett, R. C., J. H. Jett, and R. A. Keller. 2000. Molecular flow detection system, p. 115–137. In G. Durack and J. P. Robinson (ed.), *Emerging tools for single-cell analysis: advances in optical measurement technologies*. John Wiley & Sons, New York, N.Y.
24. Han, J., and H. G. Craighead. 2000. Separation of long DNA molecules in a microfabricated entropic trap array. *Science* **288**:1026–1029.
25. Han, J. Y., and H. G. Craighead. 2002. Characterization and optimization of an entropic trap for DNA separation. *Anal. Chem.* **74**:394–401.
26. Hane, B. G., K. Jager, and H. G. Drexler. 1993. The Pearson product-moment correlation coefficient is better suited for identification of DNA fingerprint profiles than band matching algorithms. *Electrophoresis* **14**:967–972.
27. Hiramatsu, K., N. Aritaka, H. Hanaki, S. Kawasaki, Y. Hosoda, S. Hori, Y. Fukuchi, and I. Kobayashi. 1997. Dissemination in Japanese hospitals of strains of *Staphylococcus aureus* heterogeneously resistant to vancomycin. *Lancet* **350**:1670–1673.
28. Huang, X. Z., M. C. Chu, D. M. Engelthaler, and L. E. Lindler. 2002. Genotyping of a homogeneous group of *Yersinia pestis* strains isolated in the United States. *J. Clin. Microbiol.* **40**:1164–1173.
29. Huang, Z., J. H. Jett, and R. A. Keller. 1999. Bacterial genome fingerprinting by flow cytometry. *Cytometry* **35**:169–175.
30. Huang, Z. P., J. T. Petty, B. Oquinn, J. L. Longmire, N. C. Brown, J. H. Jett, and R. A. Keller. 1996. Large DNA fragment sizing by flow cytometry: application to the characterization of P1 artificial chromosome (PAC) clones. *Nucleic Acids Res.* **24**:4202–4209.
31. Kim, Y., J. H. Jett, E. J. Larson, J. R. Penttila, B. L. Marrone, and R. A. Keller. 1999. Bacterial fingerprinting by flow cytometry: bacterial species discrimination. *Cytometry* **36**:324–332.
32. Kuroda, M., T. Ohta, I. Uchiyama, T. Baba, H. Yuzawa, I. Kobayashi, L. Cui, A. Oguchi, K. Aoki, Y. Nagai, J. Lian, T. Ito, M. Kanamori, H. Matsumaru, A. Maruyama, H. Murakami, A. Hosoyama, Y. Mizutani-Ui, N. K. Takahashi, T. Sawano, R. Inoue, C. Kaito, K. Sekimizu, H. Hirakawa, S. Kuhara, S. Goto, J. Yabuzaki, M. Kanehisa, A. Yamashita, K. Oshima, K. Furuya, C. Yoshino, T. Shiba, M. Hattori, N. Ogasawara, H. Hayashi, and K. Hiramatsu. 2001. Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*. *Lancet* **357**:1225–1240.
33. Larson, E. J., J. R. Hakovirta, H. Cai, J. H. Jett, S. Burde, R. A. Keller, and B. L. Marrone. 2000. Rapid DNA fingerprinting of pathogens by flow cytometry. *Cytometry* **41**:203–208.
34. Lindberg, P., P. G. Righetti, C. Gelfi, and J. Roeraade. 1997. Electrophoresis of DNA sequencing fragments at elevated temperature in capillaries filled with poly(*N*-acryloylaminopropanol) gels. *Electrophoresis* **18**:2909–2914.
35. Maule, J. 1998. Pulsed-field gel electrophoresis. *Mol. Biotechnol.* **9**:107–126.
36. Miller, R. G. 1974. Jackknife—review. *Biometrika* **61**:1–15.
37. Murchan, S., M. E. Kaufmann, A. Deplano, R. De Ryck, M. Struelens, C. E. Zinn, V. Fussing, S. Salmenlinna, J. Vuopio-Varkila, N. El Solh, C. Cuny, W. Witte, P. T. Tassios, N. Legakis, W. Van Leeuwen, A. Van Belkum, A. Vindel, I. Laconcha, J. Garaizar, S. Haeggman, B. Olsson-Liljequist, U. Ransjo, G. Coombes, and B. Cookson. 2003. Harmonization of pulsed-field gel electrophoresis protocols for epidemiological typing of strains of methicillin-resistant *Staphylococcus aureus*: a single approach developed by consensus in 10 European laboratories and its application for tracing the spread of related strains. *J. Clin. Microbiol.* **41**:1574–1585.
38. Pearson, K. 1926. On the coefficient of racial likeness. *Biometrika* **18**:105–117.
39. Petty, J. T., M. E. Johnson, P. M. Goodwin, J. C. Martin, J. H. Jett, and R. A. Keller. 1995. Characterization of DNA size determination of small fragments by flow-cytometry. *Anal. Chem.* **67**:1755–1761.
40. Press, W. H. 1992. Numerical recipes in C: the art of scientific computing, 2nd ed. Cambridge University Press, Cambridge, United Kingdom.
41. Ribot, E. M., C. Fitzgerald, K. Kubota, B. Swaminathan, and T. J. Barrett. 2001. Rapid pulsed-field gel electrophoresis protocol for subtyping of *Campylobacter jejuni*. *J. Clin. Microbiol.* **39**:1889–1894.
42. Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
43. Schins, J. M., A. Agronskaya, B. G. De Grooth, and J. Greve. 1998. New technique for high resolution DNA sizing in Epi-illumination. *Cytometry* **32**:132–136.
44. Schwartz, D. C., and C. R. Cantor. 1984. Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. *Cell* **37**:67–75.
45. Sokal, R. R., and P. H. A. Sneath. 1963. Principle of numerical taxonomy. Freeman, San Francisco, Calif.
46. Swaminathan, B., T. J. Barrett, S. B. Hunter, R. V. Tauxe, and C. P. T. Force. 2001. Pulsenet: the molecular subtyping network for foodborne bacterial disease surveillance, United States. *Emerg. Infect. Dis.* **7**:382–389.
47. 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**:2233–2239.
48. Van Baar, B. L. M. 2000. Characterization of bacteria by matrix-assisted laser desorption/ionization and electrospray mass spectrometry. *FEMS Microbiol. Rev.* **24**:193–219.
49. Van Belkum, A., W. Van Leeuwen, M. E. Kaufmann, B. Cookson, F. Forey, J. Etienne, R. Goering, F. Tenover, C. Steward, F. O'Brien, W. Grubb, P. Tassios, N. Legakis, A. Morvan, N. El Solh, R. De Ryck, M. Struelens, S. Salmenlinna, J. Vuopio-Varkila, M. Kooistra, A. Talens, W. Witte, and H. Verbrugh. 1998. Assessment of resolution and intercenter reproducibility of results of genotyping *Staphylococcus aureus* by pulsed-field gel electrophoresis of *Sma* I macrorestriction fragments: a multicenter study. *J. Clin. Microbiol.* **36**:1653–1659.
50. Yan, X. M., W. K. Grace, T. M. Yoshida, R. C. Habbersett, N. Velappan, J. H. Jett, R. Keller, and B. L. Marrone. 1999. Characteristics of different nucleic acid staining dyes for DNA fragment sizing by flow cytometry. *Anal. Chem.* **71**:5470–5480.
51. Yan, X. M., R. C. Habbersett, J. M. Cordek, J. P. Nolan, T. M. Yoshida, J. H. Jett, and B. L. Marrone. 2000. Development of a mechanism-based, DNA staining protocol using Sytox orange nucleic acid stain and DNA fragment sizing flow cytometry. *Anal. Biochem.* **286**:138–148.