TECHNICAL NOTE

David L. Duewer,[1] Ph.D.; Kristy L. Richie,[1] B.S.; and Dennis J. Reeder,[1] Ph.D.

RFLP Band Size Standards: NIST Standard Reference Material® 2390*

REFERENCE: Duewer DL, Richie KL, Reeder DJ. RFLP band size standards: NIST standard reference material® 2390. J Forensic Sci 2000;45(5):1093–1105.

ABSTRACT: The procedural standard for DNA profiling developed by the U.S. advisory board on DNA quality assurance methods mandates annual confirmation of forensic DNA measurement systems against an appropriate reference material supplied by or traceable to the National Institute of Standards and Technology (NIST). NIST Standard Reference Material® (SRM®) 2390 is a suitable and appropriate standard for *Hae*III restriction enzyme-based restriction fragment length polymorphism (RFLP) profiling systems. Originally issued in 1992, an among-laboratory SRM 2390 recertification study was initiated in 1997. Using data provided by the 20 state, local, or commercial forensic laboratory participants, quantitative band sizes values (expected mean values and associated bivariate tolerance intervals) are established for two different-source DNAs (female cell line K562 and healthy male "TAW") for genetic loci D1S7, D2S44, D4S139, D5S110, D10S28, and D17S79. Methods for validating an RFLP measurement system, validating a control material or other secondary standard, and for tracing a particular set of RFLP measurements to NIST SRM 2390 are described in detail.

KEYWORDS: forensic science, DNA typing, data analysis, *Hae*III, *Hinf*I, D1S7, D2S44, D4S139, D5S110, D7S467, D7Z2, D8S358, D10S28, D14S13, D17S26, D17S79, DYZI, standard reference material

The advisory board on DNA quality assurance methods (DAB) was authorized in 1994 to assist the Director of the Federal Bureau of Investigation (FBI) in establishing quality assurance and proficiency testing standards for forensic DNA testing laboratories (1). Standard 9.5 of the procedural standards developed by the DAB states "The laboratory shall check its DNA procedures annually or whenever substantial changes are made to the protocol(s) against an appropriate and available NIST standard reference material or standard traceable to a NIST standard" (2).

The National Institute of Standards and Technology (NIST) provides three Standard Reference Materials® (SRM®) for DNA testing: SRM 2390 for restriction fragment length polymorphism (RFLP) systems (3,4), SRM 2391 for polymerase chain reaction based systems (5), and SRM 2392 for mitochondrial sequencing (6). All three SRMs provide rigorously defined DNA materials that are certified to have various characteristics of forensic interest. Laboratories can demonstrate the validity of their DNA measurements by comparison of their results to the certified values provided with each SRM. They also can establish traceability through direct comparison of the SRM and appropriate proposed secondary standard materials through use of a validated measurement system.

The values certified in SRMs 2391 and 2392 reflect qualitative characteristics of the DNA materials themselves: allelic names for selected short tandem repeat loci and the nucleotide sequence of selected portions of the DNA, respectively. These characteristics are independent of any specific measurement system. While advances in DNA measurement technology may lead to certification of additional characteristics of these materials, we anticipate that these materials will retain their utility to the forensic community for the foreseeable future.

In contrast, the values certified in SRM 2390 reflect characteristics of both its DNA materials and the specific RFLP protocols developed by the FBI and members of the Technical Working Group for DNA Analysis Methods (TWGDAM) (7,8). The quantitative measurement characteristic ("band size") of a RFLP fragment is a function of the electrophoretic migration rate of each DNA fragment ("band") through a gel *relative* to the migration rates of neighboring internal standard "sizing ladder" components (9,10). The migration rate of a given band is predominantly proportional to the logarithm of the number of nucleotide base pairs (bp) included in the DNA fragment; however, nucleotide composition, sequence, and configuration also influence the migration rate (11). The relative rates of target and sizing ladder band migration can be differentially affected by virtually every post-extraction RFLP variable (12,13). We do not expect the currently certified characteristics of SRM 2390 to have forensic utility beyond current RFLP technology. However, given the number of RFLP profiles collected over the past decade, we do anticipate that RFLP measurements will continue for a number of years.

This report is intended to help U.S. forensic laboratories interpret and comply with the quality assurance requirements mandated by the DAB standards. The following sections describe the DNA components of SRM 2390 in detail and present methods for documenting RFLP measurement traceability to SRM 2390.

[1] Analytical Chemistry Division and Biotechnology Division, respectively, Chemical and Science Technology Laboratory, National Institute of Standards and Technology, Gaithersburg, MD.

* This work was supported in part by the Federal Bureau of Investigation and by the National Institute of Justice, U.S. Department of Justice.

Received 28 April 1999; and in revised form 10 Nov. 1999; accepted 15 Nov. 1999.

Experimental

SRM 2390 consists of 20 consumable materials required for evaluating *Hae*III-based RFLP measurement systems. It was designed to aid analysts qualitatively "troubleshoot" measurement difficulties as well as providing quantitative reference values for well characterized human DNA from two sources. Three different forms of DNA from both sources are provided to enable isolation of the extraction and *Hae*III-digestion components of a RFLP protocol; these forms are: cell pellet ("cell"), extracted but undigested genomic DNA ("genomic"), and *Hae*III restriction digest ("precut"). Component 12, 13, and 14 of SRM 2390 are cell, genomic, and precut samples derived from the immortal female myelogenous leukemia cell line K562. Components 15, 16, and 17 are cell, genomic, and precut samples derived from white blood cells donated by healthy male donor "TAW."

The K562 and TAW components were obtained from separate commercial sources.

1991 Certification Study

The measurements used in the original certification of the K562 and TAW *Hae*III RFLP band sizes were provided by a NIST-sponsored among-laboratory study initiated in 1991 and completed in 1992. Twenty-nine private, local, state, provincial, and U.S. and Canadian federal laboratories participated in this material certification exercise. Participants were asked to return band sizes for the three K562 and three TAW samples for D2S44 and as many other loci as they could work into their schedules. One participant reported results for two somewhat different protocols; one laboratory reported results from five analyses. Sufficient results were reported to certify *Hae*III RFLP band sizes at loci D1S7, D2S44, D4S139, D10S28, and D17S79. Details of this study are reported elsewhere (3).

Each certified value consisted of the expected band size in bp and an associated 95%/95% statistical tolerance interval. The expected values for the K562 and TAW DNA were calculated as the grand mean of the means of the cell, genomic, and precut samples. The tolerance intervals for the two DNAs were established using an analysis of variance design that appropriately allocated degrees of freedom to known within- and among-laboratory effects. With 95% confidence, at least 95% of measurements on SRM 2390 components made by a laboratory using RFLP protocols similar to those used by the laboratories in the study are expected to fall within the stated tolerance interval.

1997 Recertification Study

Twenty U.S. private, local, and state forensic laboratories participated in the among-laboratory SRM 2390 recertification exercise. Participation was solicited at DNA methodology conferences in June and September of 1997. SRM 2390 sets were provided to laboratories as they confirmed their intent to participate, starting in August 1997. The participants were asked to use their "normal" casework RFLP protocol, using as many of the SRM components as they could, and to report any observed anomalies in addition to the quantitative band sizes.

Data were entered into the master database at NIST as soon as they became available. Any missing, misrecorded, or anomalous information was obtained, corrected, or discussed with the participant. Participants who included their own K562 control material in the same gel as the SRM 2390 samples were requested to provide the band sizes for this material. To ensure that the K562-related band sizes were representative of routine measurement performance, participants were also requested to provide a statistical summary of their K562 casework data collected during "about the past two years." All laboratories receiving an SRM set provided their results and all other requested information. Table 1 lists the laboratories that participated in this exercise.

One participant provided two sets of data for the same gel, one set measured using the sizing ladder supplied with SRM 2390 and the other using the laboratory's routine ladder; only the values obtained with the SRM 2390 set are included in the data analysis. Several of the participants reported the average values of two or more independent analyses of the same images. Previous studies have documented that the variation attributable just to image analysis is a relatively small fraction of the observed among-laboratory variation for bands of size less than about 10 000 bp (9,14). All 20 data sets have therefore been given the same statistical weight in our analyses.

1997 NIST Quality Assurance Measurements

Ten SRM 2390 sets were examined at NIST to assay the performance of all SRM components. Autoradiograms for the six DNA samples were obtained at locus D5S110 to document the performance characteristics of all components of SRM 2390. Since NIST neither routinely performs forensic casework nor routinely performs RFLP analyses, the NIST sizing data are not included in the analysis of the among-laboratory data.

TABLE 1—*Participants in the 1997 SRM 2390 recertification study.*

Participating Laboratory	Location
Arizona Department of Public Safety Crime Laboratory DNA Unit	Phoenix, AZ
Broward County Sheriff's Office—Crime Laboratory	Ft. Lauderdale, FL
Connecticut State Police Forensic Science Laboratory DNA Unit	Meriden, CT
Illinois State Police, Forensic Science Center at Chicago	Chicago, IL
Illinois State Police, Springfield Forensic Science Laboratory	Springfield, IL
Indianapolis-Marion County Forensic Services Agency	Indianapolis, IN
Kentucky State Police Crime Laboratory	Frankfort, KY
LabCorp, Forensic Identity Testing	RTP, NC
Maryland State Police Crime Laboratory	Pikesville, MD
Metro-Dade Police Department Crime Laboratory Bureau	Miami, FL
Minnesota Department of Public Safety Bureau of Criminal Apprehension	St. Paul, MN
New York State Police Forensic Investigation Center	Albany, NY
North Carolina State Bureau of Investigation Crime Laboratory	Raleigh, NC
Orange County Sheriff-Coroner Department DNA Laboratory	Santa Ana, CA
Pennsylvania State Police DNA Laboratory	Greensburg, PA
South Carolina Law Enforcement Division DNA Laboratory	Columbia, SC
Vermont Forensic Laboratory, Department of Public Safety	Waterbury, VT
Washington State Patrol Crime Laboratory—Seattle	Seattle, WA
Washoe County Sheriff's Office Crime Laboratory	Reno, NV
Wisconsin State Crime Laboratory—Milwaukee	Milwaukee, WI

Results and Discussion

Table 2 lists the numbers of participants providing *Hae*III band sizes for all genetic loci reported in the 1991 and the 1997 material certification studies and summary statistics for all 1997 measurements. Figure 1 displays all K562 measurements reported in the 1997 recertification study for the six most commonly reported loci: D1S7, D2S44, D4S139, D5S110, D10S28, and D17S79. Figure 2 likewise displays all TAW data from the 1997 study for these loci. Each of the scattergram components of these figures is a modified single locus charts (SLC), displaying both measurements and data quality metrics for a given locus for a given source of DNA (15). The quality metrics displayed are the 95%/95% tolerance boxes for the 1991 certification study, 95%/95% bivariate tolerance ellipses for the 1997 recertification study, and 99% bivariate tolerance ellipses predicted from our analysis of the forensic proficiency test data. Detailed discussions of all statistical calculations are provided in following sections.

Figure 3 provides an alternative display of the same K562 and TAW data shown in Figs. 1 and 2. Rather than segregating the data by DNA source and locus, each of the scattergram components of Fig. 3 is a laboratory performance chart (LPC) displaying all measurements reported by a given participant relative to a correlation-adjusted 99% tolerance ellipse (16). In essence, each LPC in Fig. 3 is an overlay of all 12 SLCs in Figs. 1 and 2 showing just the data reported by the given participant. The LPCs in Fig. 3 are sorted in order of increasing apparent precision.

Outlier Data

Two sets of K562 data identify "outlier samples" in both Fig. 1 (the "boxed" values) and Fig. 3 (participants 97-01 and 97-05). All the values for these two samples are unusual with respect to all of the other data for that sample and to the other data provided by that participant. None of these data are used in any quantitative analysis. No other data are obviously inconsistent either by locus/source class (Figs. 1 and 2) or participant (Fig. 3).

Participant 97-01's SRM 2390 component 12 (K562 cell pellet) values are unusually large at all five loci reported; the value of at least one band of each pair is exterior to some or all measurement tolerance limits for three of the five loci reported by this participant. Visual inspection of the autoradiograms provided by the participant established that the width and optical density of all bands for this sample are very much greater than for the other sample and sizing ladder bands. Sizing of such "blowout" bands is known to be problematic (12). Inspection of all other autoradiograms and chemilumograms supplied by study participants confirmed that all other samples were in relative balance with each other and the sizing ladders.

Participant 97-05's K562 control data are unusually small at all six loci reported, although all values are within all tolerance limits. This participant sized the control K562 bands using the ladder supplied in the SRM 2390 set while using their own (nominally identical) ladder for all other samples. Visual inspection of the autoradiograms reveals that the SRM 2390 ladder (lanes 1 and 4, very dark bands; lane 1 bands "smile" upwards to the gel edge) were offset from the participant's (lanes 3, 5, 9, 13, 14; light bands) while the participant's control K562 bands (lane 2) aligned well with the SRM K562 bands (lanes 6, 7, 8). While the low control band sizes suggest some systematic protocol effect involving the SRM 2390 ladder and/or gel geometry, there is thus no evidence of true differences between the SRM 2390 and the participant's K562 values.

Differences Among Cell, Genomic, and Precut Samples

Other than the outlier samples described above, Fig. 3 provides no evidence of within-laboratory measurement differences between K562 and TAW samples. However, the values reported for the cell (components 12 and 15), genomic (13 and 16), and precut (14 and 17) samples are not completely random within the SLCs of Figs. 1 and 2; in particular, the precut component measurements tend to be somewhat larger than those for cell and genomic components. The precut measurements in the 1991 certification study data are similarly somewhat larger (12).

The relatively constant discordances (offsets from the center) visible in many of the Fig. 3 LPCs reflect systematic among-laboratory measurement differences that affect all samples. Figure 4 presents the 1991 and 1997 data for K562 and TAW SRM 2390 components as differences between the cell and genomic component measurements and between the precut and genomic measure-

TABLE 2—*Statistical summary of 1997 SRM 2390 K562 and TAW band size measurements.*

Locus	Number of Participants			Cell Line K562					TAW				
	1991	1997	$K_{bi}^{n97}{}_{95/95}$	$\bar{X}_1$ (bp)	$\bar{S}_1$ (bp)	$\bar{X}_2$ (bp)	$\bar{S}_2$ (bp)	$\bar{R}$	$\bar{X}_1$ (bp)	$\bar{S}_1$ (bp)	$\bar{X}_2$ (bp)	$\bar{S}_2$ (bp)	$\bar{R}$
D1S7	16	20	11.9	4585	29.9	4237	23.6	0.88	7773	63.2	6886	41.3	0.82
D2S44	29	20	11.9	2905	19.5	1788	14.3	0.81	3711	23.2	1288	8.7	0.65
D4S139	19	20	11.9	6474	49.2	3440	22.2	0.78	10854	117	8185	64.7	0.71
D5S110	0	17	13.2	3700	23.2	2926	17.4	0.85	3343	20.4	1444	11.5	0.60
D10S28	17	19	12.3	1754	11.4	1180	9.7	0.83	3935	24.5	1788	10.2	0.82
D17S79	12	15	14.5	1979	14.8	1514	11.2	0.88	1753	11.2	1515	11.7	0.95
D7S467	0	5	…	4677	…	3217	…	…	4496	…	4339	…	…
D7Z2*	1	1	…	2736	…	…	…	…	2747	…	…	…	…
D8S358	0	1	…	5878	…	1303	…	…	3417	…	2383	…	…
D14S13†	0	2	…	1642	…	…	…	…	1579	…	…	…	…
D17S26	1	1	…	4823	…	1358	…	…	5514	…	4852	…	…
DYZ1‡	1	1	…	…	…	…	…	…	3571	…	…	…	…

All symbols are defined in text.
* Human-specific monomorphic locus having a known sequence of length 2731 bp.
† Both K562 and TAW are apparent homozygotes at this locus.
‡ Y-chromosome locus.

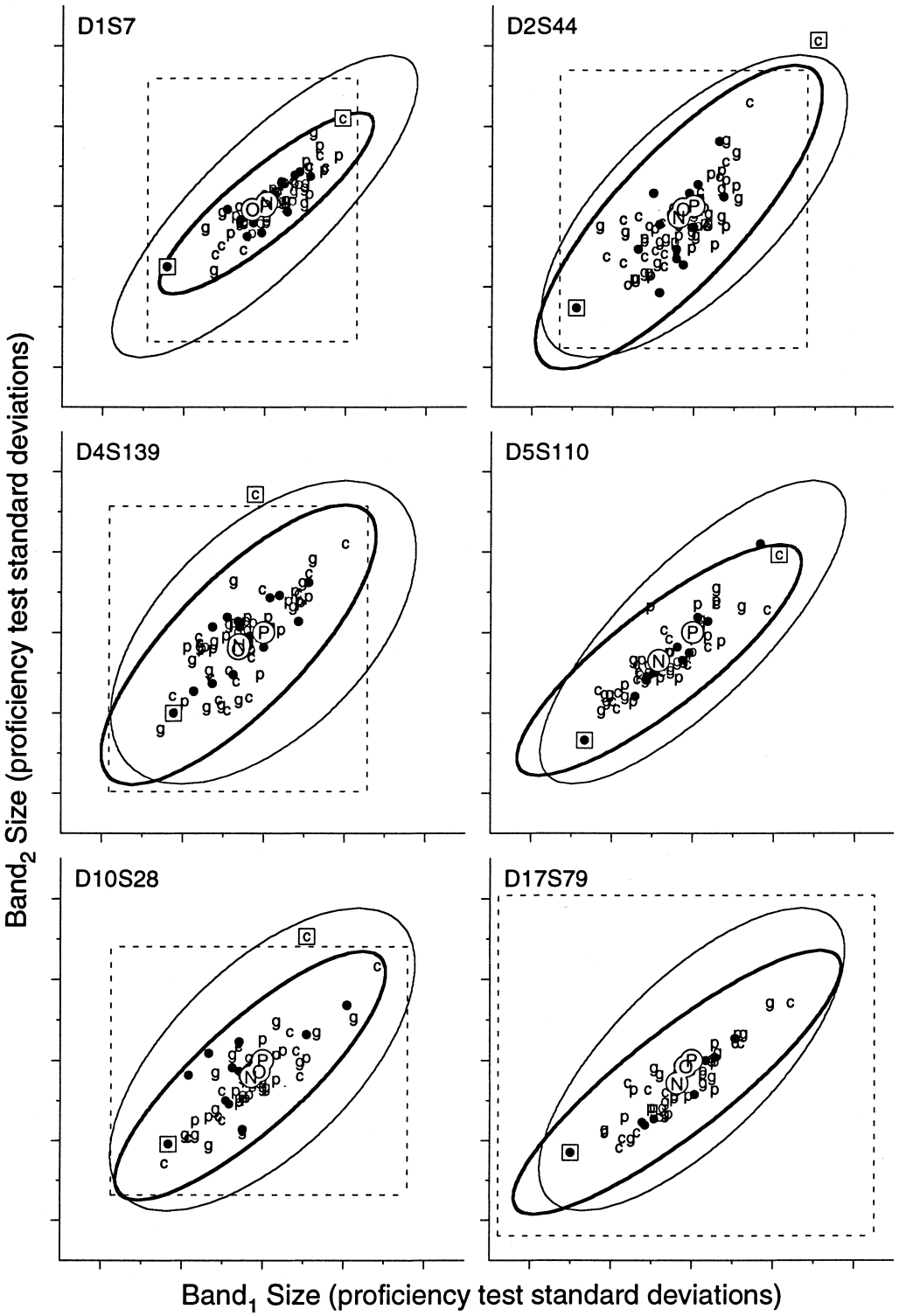

FIG. 1—*Single locus charts (SLC) for K562 components in the 1997 SRM 2390 recertification study. Each SLC displays all reported band sizes for the three K562 components of SRM 2390 for one genetic locus, plotting the smaller band size of each pair against the larger. Each SLC is centered on the average values for the 51 sets of 1991 through 1997 proficiency test data; both axes of each SLC span ±5 standard deviations observed for these proficiency test data. The measurements pairs reported for component 12 (cellular pellet) are denoted* "c," *those for component 13 (genomic) are denoted* "g," *and those for component 14 (*HaeIII *precut) are denoted* "p." *The open squares identify measurement pairs for two different "outlier samples." The three open circles denote the mean proficiency test values* ("P"), *the certified values for the 1991 study* ("O"), *and the certified values for the 1997 study* ("N"). *The light-line ellipse represents the 99% tolerance region for proficiency test data from 1991 through 1997. The dashed-line box represents the univariate 95%/95% tolerance region for the 1991 study. The dark-line ellipse represents the 95%/95% tolerance region for the 1997 study. No measurements for locus D5S110 were reported in the 1991 study.*

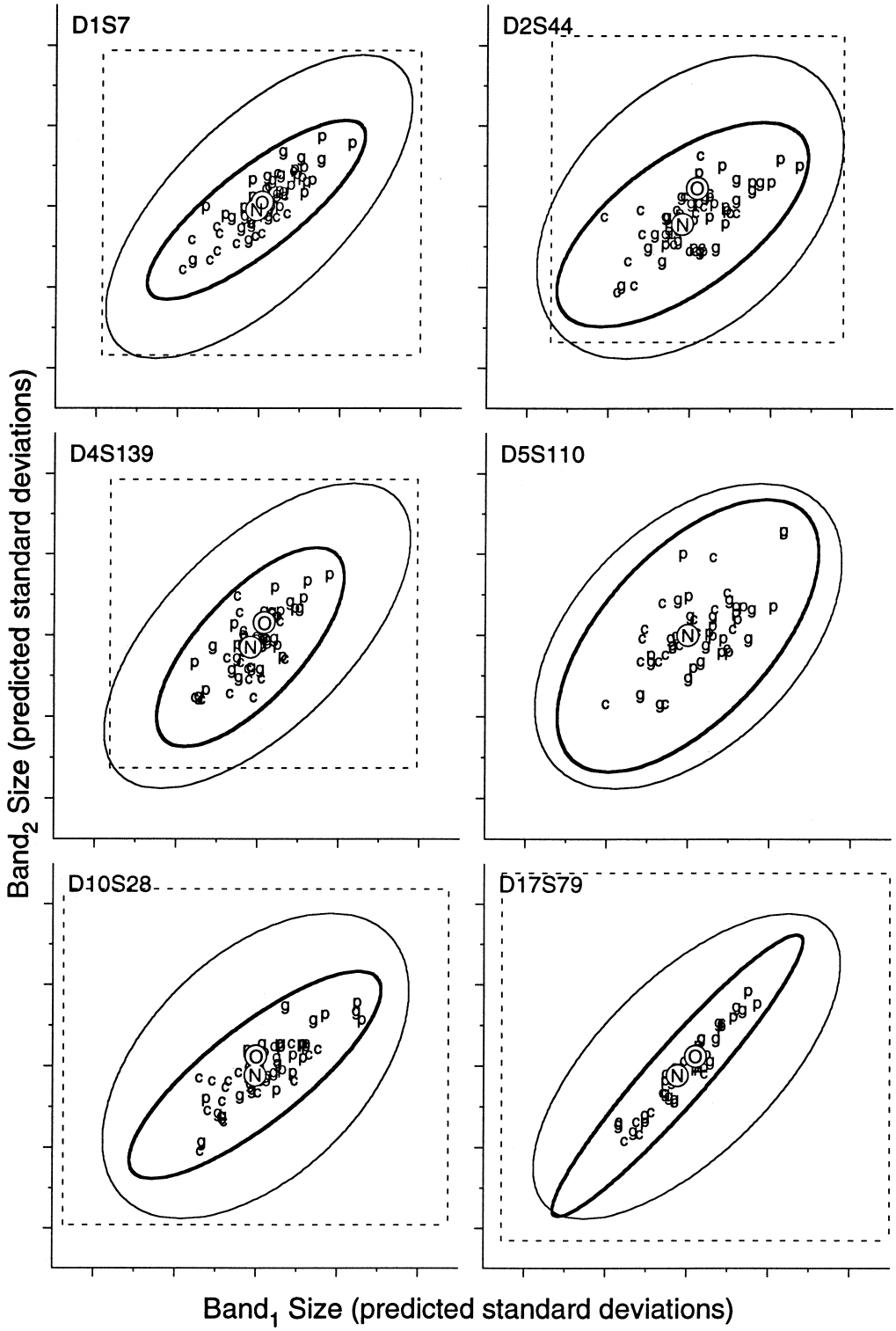

FIG. 2—*Single locus charts (SLC) for TAW components in the 1997 SRM 2390 recertification study. These SLCs are similar to those in Fig. 1, with each SLC centered on the average of the 1991 and 1997 material certification study measurements and both axes of each SLC spanning ±5 standard deviations predicted using Eq 9. The measurement pairs reported for component 15 (cellular pellet) are denoted* "c," *those for component 16 (genomic) are denoted* "g," *and those for component 17 (*HaeIII *precut) are denoted* "p." *The two open circles denote the mean proficiency test values* ("P"), *the certified values for the 1991 study* ("O"), *and the certified values for the 1997 study* ("N"). *The light-line ellipse represents the 99% tolerance region for proficiency test data from 1991 through 1997. The dashed-line box represents the univariate 95%/95% tolerance region for the 1991 study. The dark-line ellipse represents the 95%/95% tolerance region for the 1997 study.*

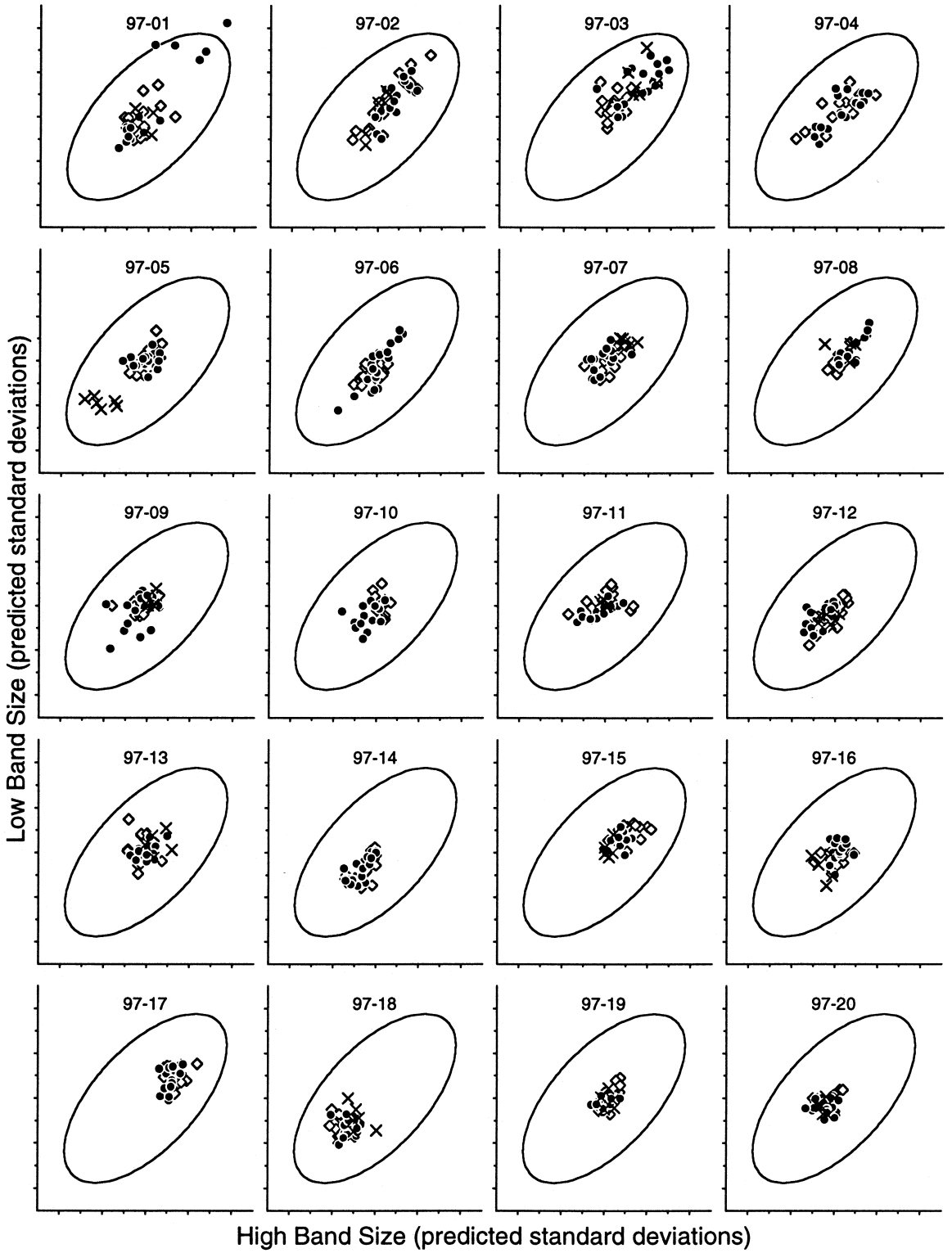

FIG. 3—*Laboratory performance charts (LPC) for participants in 1997 SRM 2390 recertification study. Each LPC displays all reported band sizes reported by one participant. The LPCs are presented in order of increasing apparent precision; this ordering is not related to the alphabetized listing of Table 1. The smaller band size of each measurement pair is plotted against the larger; each measurement is first standardized to the average value for the particular band and the standard deviation predicted by Eq 9. Each LPC is centered on the average measurement values for the 1997 recertification study; both axes of each LPC span ±5 predicted standard deviations. The K562 components of SRM 2390 are denoted as solid circles, the TAW components as open diamonds, and the participant's cell line K562 control as "x."*

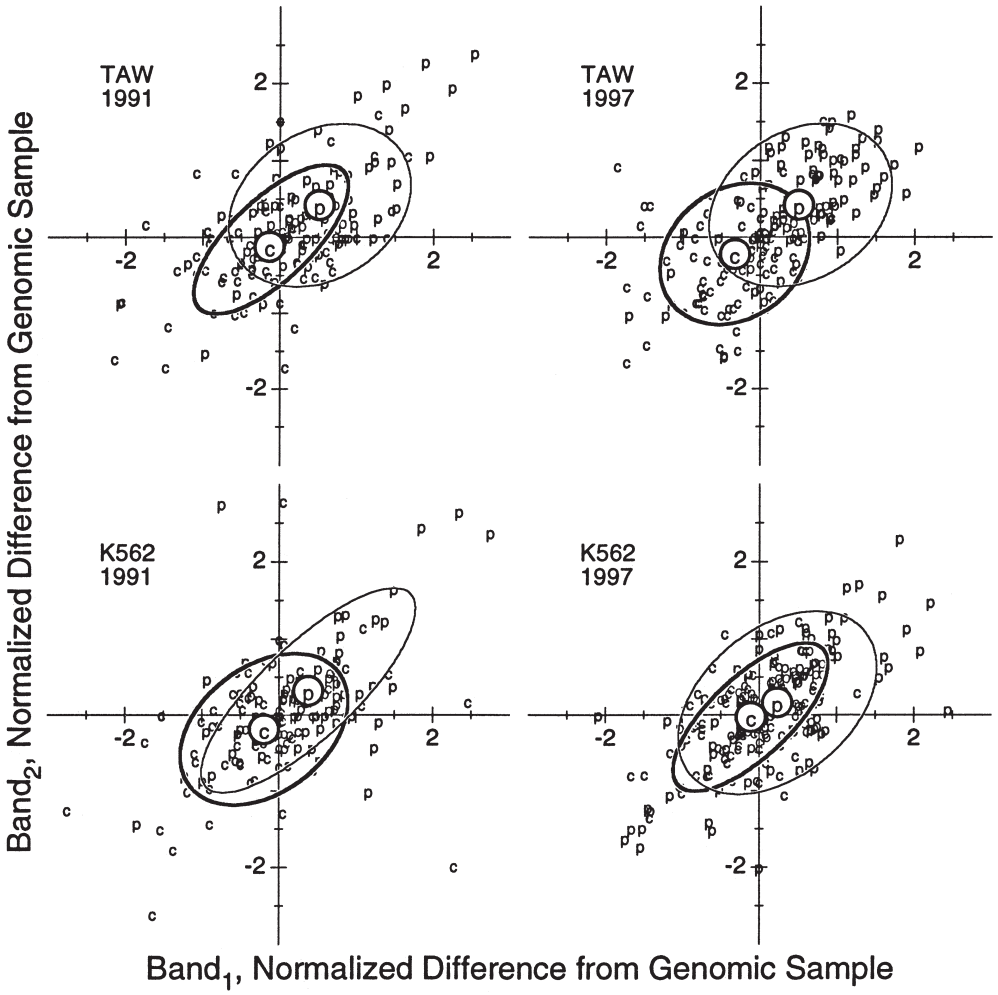

FIG. 4—*Differences among cell, genomic, and precut components of SRM 2390. These four scattergrams display all (cell–genomic) and (precut–genomic) measurement differences for K562 and TAW DNAs for each participant in the 1991 and the 1997 material certification studies. The standard deviation-normalized difference for the smaller band size of each measurement pair is plotted against that of the larger. The (cell–genomic) differences are denoted* "c"; *the dark-line ellipses are the minimum area ellipsoids for covering 80% of these differences. The (precut–genomic) differences are denoted* "p"; *the light-line ellipses are the minimum area ellipsoids for covering 80% of these differences. The open circles denote the mean values of the two differences for each DNA in the two studies.*

ments. These relative differences remove among-participant discordance variability from the data, enabling better evaluation of among component measurement differences. The measured band sizes in the cell components are slightly smaller than in genomic components, while the measured sizes in precut components are larger than in the genomic. The differences between the precut and genomic measurements are also more variable than those between the cell and genomic.

We believe that the observed measurement differences among the components are largely attributable to subtle composition differences in the loading buffers used with the different samples, since proprietary preservatives are included in commercial precut DNA solutions. Since there is the potential for band size measurements of precut DNA to be biased relative to samples that are extracted and/or digested in the same manner as evidential and reference samples, none of the precut component measurements are used to estimate the expected K562 and TAW band sizes. Although the precut component measurements average just slightly larger than the certified values (and within the measurement tolerance limits), the SRM 2390 certified values are strictly valid only for the cell and genomic components.

Expected Band Sizes

The band sizes listed in Table 2 are the mean of the mean cell and mean genomic component valid measurements. The mean for a given component "*j*" is estimated

$$X_{d\ell bj} = \sum_{i=1}^{n_{d\ell j}} x_{d\ell bji}/n_{d\ell j} \qquad (1)$$

where subscript d specifies the DNA source (K562 or TAW), ℓ the genetic locus, b the band (1 or 2), i the participant; $x_{d\ell bji}$ are the individual measurements reported for component j; and $n_{d\ell j}$ is the number of valid measurements. The expected band size for the DNA source is estimated

$$\overline{X} = \overline{X}_{d\ell b} = (X_{d\ell bc} + X_{d\ell bg})/2 \qquad (2)$$

where subscript c specifies the cell component and g the genomic

and $\overline{X}$ is the generic symbol when the specific DNA source, locus, and band do not require specification. Values of $\overline{X}$ are provided for all bands reported by any participant in the 1997-recertification study.

Expected Band Size Standard Deviations

The expected band size standard deviations listed in Table 2 represent the pooled variance for the individual cell, genomic, and precut component measurements. The standard deviation for a given component j is estimated

$$S_{d\ell b} = +\sqrt{\sum_{i=1}^{n_{d\ell j}} (x_{d\ell bji} - X_{d\ell bj})^2/(n_{d\ell j} - 1)} \quad (3)$$

The expected standard deviation for the DNA source is estimated

$$\overline{S} = \overline{S}_{d\ell b} = +\sqrt{\frac{(n_{d\ell c} - 1)S^2_{d\ell bc} + (n_{d\ell g} - 1)S^2_{d\ell bg} + (n_{d\ell p} - 1)S^2_{d\ell bp}}{n_{d\ell c} + n_{d\ell g} + n_{d\ell p} - 3}} \quad (4)$$

where subscript p specifies the precut component and $\overline{S}$ is the generic symbol for the expected among-laboratory standard deviation for replicate measurements of one particular band. Inclusion of the precut measurements improves the reliability of these estimates of within- and among-participant measurement variability. Values of $\overline{S}$ are provided only for bands reported by 15 or more participants in the 1997 recertification study.

Expected Bivariate Correlations

The expected correlations listed in Table 2 represent the observed correlation between the ordered pair of band measurements made at each locus, pooled over the cell, genomic, and precut components. The bivariate correlation for a given component j is estimated

$$R_{d\ell j} = \frac{\sum_{i=1}^{n_{d\ell j}} (x_{d\ell 1ji} - X_{d\ell 1j})(x_{d\ell 2ji} - X_{d\ell 2j})}{(n_{d\ell j} - 1)S_{d\ell 1j}S_{d\ell 2j}} \quad (5)$$

where subscript 1 designates the larger ("high") band and 2 designates the smaller ("low") band of the ordered pair. The expected correlation is pooled over the three different components using Fisher's normality transformation (17)

$$\overline{R} = \overline{R}_{d\ell} = \tanh\left(\frac{\tanh^{-1}(R_{d\ell c}) + \tanh^{-1}(R_{d\ell g}) + \tanh^{-1}(R_{d\ell p})}{3}\right) \quad (6)$$

where tanh is the hyperbolic tangent, $\tanh^{-1}$ is the "atanh" inverse hyperbolic tangent, and $\overline{R}$ is the generic symbol for the expected among-laboratory correlation between replicate $\{x_1, x_2\}$ measurements pairs of one particular locus. $\overline{R}$ values are provided only for loci reported by 15 or more participants in the 1997 recertification study.

95%/95% Tolerance Limits

The 1991 study established univariate 95%/95% tolerance intervals for individual bands. The "boxes" in Figs. 1 and 2 present the acceptance regions defined by simultaneous application of these univariate intervals to both bands at each locus. The ellipses in Figs. 1, 2, and 3 present more representative bivariate tolerance limits.

For normally distributed replicate measurements, the univariate 95%/95% tolerance interval contains 95% of all valid members of that distribution with 95% confidence. A given measurement, x, is within the interval when

$$\frac{|x - \overline{X}|}{\overline{S}} \le K^n_{\mathrm{uni}_{95/95}} \quad (7)$$

where $K^n_{\mathrm{uni}_{95/95}}$ is the two-tailed critical factor for $\overline{X}$ and $\overline{S}$ estimated from n measurements (18). Similarly, if the distribution of a set of replicate measurement pairs is bivariate normal, the bivariate 95%/95% tolerance interval contains 95% of all valid member pairs of that distribution with 95% confidence. A given measurement pair, $\{x_1,x_2\}$, is within the tolerance region when

$$\frac{\left(\frac{x_1 - \overline{X}_1}{\overline{S}_1}\right)^2 + \left(\frac{x_2 - \overline{X}_2}{\overline{S}_2}\right)^2 - 2\overline{R}\left(\frac{x_1 - \overline{X}_1}{\overline{S}_1}\right)\left(\frac{x_2 - \overline{X}_2}{\overline{S}_2}\right)}{1 - \overline{R}^2} \le K^n_{\mathrm{bi}_{95/95}} \quad (8)$$

where $K^n_{\mathrm{bi}_{95/95}}$ is the bivariate critical factor for all parameters estimated from n measurement pairs (19). Details on the construction of the tolerance ellipses and estimation of the critical values are provided elsewhere (15). The bivariate critical values for the 1997 SRM 2390 recertification study are listed in Table 2.

99% Tolerance Limits

In the second part of this series, we report our analysis of cell line K562 data from 51 forensic proficiency tests conducted from 1991 through 1997 (20). The numbers of measurement pairs (≥1000 pairs per locus) are sufficient to define the bivariate distribution of the six most commonly reported *Hae*III loci robustly. The distribution of the central 90% to 95% of measurement pairs at each locus is well approximated as bivariate normal; the remaining data are further from the distribution center than expected (the distributions have "heavier tails"). We find that 99% of all the qualitatively valid proficiency measurements for these loci fall inside ellipses having a bivariate critical factor of 14.2 (i.e., the left side of Eq 8 ≤ 14.2). We expect, with 100% confidence, that 99% of all valid cell line K562 measurement results will be within these empirical 99% tolerance ellipses regardless of the source or form of the K562 DNA.

The SRM 2390 results for K562 in Fig. 1 are displayed relative to these empirical distributions. The 95%/95% and 99% tolerance regions are remarkably similar, with the 95%/95% regions nearly completely contained within the corresponding 99% regions. All valid SRM 2390 K562 measurements are well inside the 99% ellipse. The mean band sizes for SRM 2390 K562 (center of the 95%/95% ellipse) are identical to slightly smaller than the mean for all proficiency test cell line control K562 (center of the 99% ellipse) and the correlation between the two bands at each locus is greater (manifest in the relative ratios of the minor to the major axes of the two ellipses).

While there is no proficiency test data available for TAW-source DNA, approximate 99% tolerance regions can be estimated using previously established empirical relationships. The expected among-laboratory measurement standard deviation for a given RFLP band can be estimated from the mean size of the band (9,14)

$$S(\overline{X}) = 7.5 \times \left(1 + \frac{\overline{X}}{19500}\right)^{7.1} \quad (9)$$

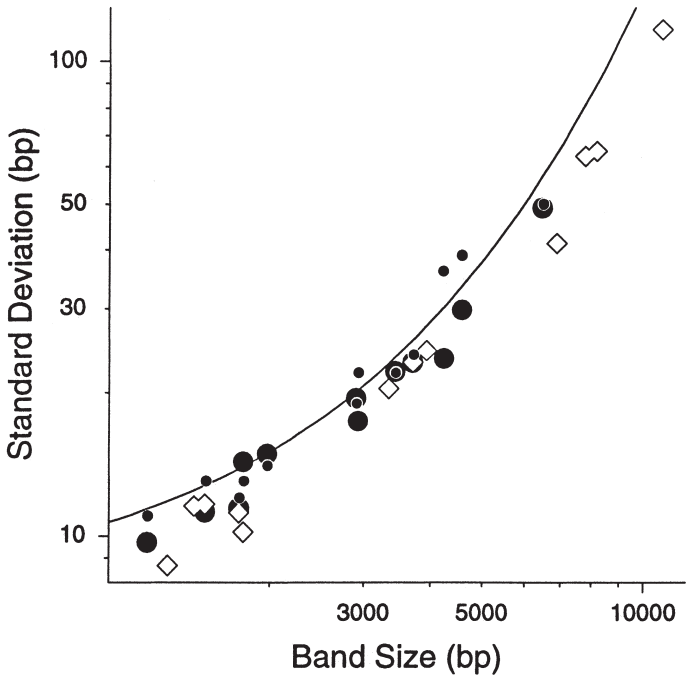

FIG. 5—*Among-laboratory standard deviation as a function of band size. The large solid circles denote the pooled among-laboratory standard deviations for the twelve bands of the commonly reported* Hae*III loci (D1S7, D2S44, D4S139, D5S110, D10S28, and D17S79) for the three SRM 2390 K562 components. The open diamonds denote the same quantities for the three TAW components. The small solid circles denote the total standard deviation observed for K562 control data reported in 51 proficiency tests conducted from 1991 through 1997. The solid line denotes the expected relationship (Eq 9) between among-laboratory standard deviation and band size.*

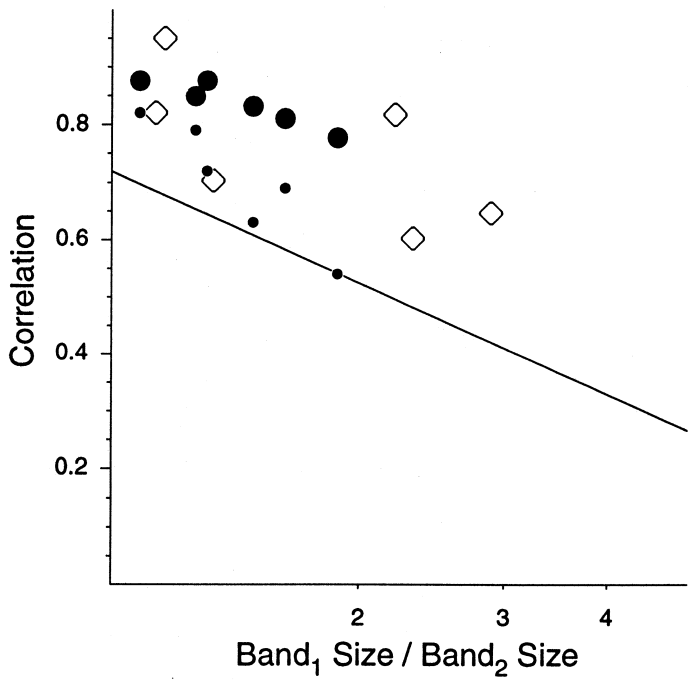

FIG. 6—*Between-band correlation as a function of band size differences. The large solid circles denote the pooled {band1, band2} correlation observed for the six commonly reported* Hae*III loci for the three SRM 2390 K562 components. The open diamonds denote the same quantity for the three TAW components. The small solid circles denote the correlations observed for the proficiency test K562 control data. The solid line denotes the expected relationship (Eq 10) between within-laboratory correlation and the ratio band1/band2.*

Figure 5 displays $\overline{S}$ versus $\overline{X}$ for the K562 and TAW 1997 recertification study and for the K562 proficiency test data. Both sets of $\overline{S}$ have the same qualitative dependence on $\overline{X}$, with the $\overline{S}$ for the recertification data somewhat smaller than observed with the proficiency data and/or predicted by Eq 9. The expected *within*-laboratory correlation between $\{x_1, x_2\}$ RFLP measurement pairs at one locus is an approximate function of the band size difference (15,21)

$$R(\overline{X}_1, \overline{X}_2) = 0.72 - \left[0.65 \times \log_{10}\left(\frac{\overline{X}_1}{\overline{X}_2}\right)\right] \qquad (10)$$

Since measurement biases among the laboratories generally inflate observed correlation, Eq 10 should provide a lower bound for *among*-laboratory correlations. Figure 6 displays $\overline{R}$ versus $\overline{X}$ for the 1997 recertification study and for the K562 proficiency test data. All the observed $\overline{R}$ display similar patterns of declining correlation with increasing difference in band size. All the proficiency test K562 correlations are smaller than observed in the 1997 recertification study. While the $\overline{R}$ for more-variable-than-expected loci D1S7 and D5S110 proficiency test measurements are underestimated, Eq 10 adequately predicts the other proficiency test correlations.

We define the approximate 99% confidence ellipses for TAW to be centered on the average of the 1991 and 1997 certified band sizes, to have the standard deviations predicted by Eq 9, to have the between-band correlation predicted by Eq 10, and to have the critical elliptical factor of 14.2 observed for the K562 proficiency test data. The SRM 2390 results for TAW in Fig. 2 are displayed relative to these approximate distributions. Table 3 lists all required 99% tolerance parameter values for both K562 and TAW DNA.

Traceability

All of the non-DNA components of SRM 2390 are provided to assist analysts troubleshoot intermediate stages of their *Hae*III RFLP procedure. Only after the reliability of the complete measurement system is well established should traceability validation measurements be performed. There are three laboratory prerequisites necessary for establishing that a particular set of RFLP measurements are traceable to NIST SRM 2390: (1) a validated RFLP measurement system, (2) validated control material(s), either K562 and/or in-house DNA, and (3) a compilation of all relevant measurements of all control material(s).

RFLP Measurement System Validation

Aliquots of the two SRM 2390 cell pellets components, 12 and 15, should be regarded as "normal" samples. Aliquots of the two genomic extract components, 13 and 16, should be appropriately diluted and regarded as "normal" DNA extracts. To the maximum extent possible, all four components should be treated and sized as if they were typical samples. All resulting band sizes should match the SRM 2390 certified values to within the stated 95%/95% bivariate tolerances (Table 2); the closer the measurements are to the certified values (the centers of the ellipses) the better. If any measurement pair is outside its tolerance ellipse, the RFLP measure-

TABLE 3—*Parameters for K562 and TAW 99% tolerance ellipses.*

		Cell Line K562*					TAW†				
Locus	K_{bi99}*	$\bar{X}_1$ (bp)	$\bar{S}_1$ (bp)	$\bar{X}_2$ (bp)	$\bar{S}_2$ (bp)	$\bar{R}$	$\bar{X}_1$ (bp)	$S(\bar{X}_1)$ (bp)	$\bar{X}_2$ (bp)	$S(\bar{X}_2)$ (bp)	$\bar{R}(\bar{X}_1,\bar{X}_2)$
D1S7	14.2	4583	39	4234	36	0.82	7768	82	6881	65	0.69
D2S44	14.2	2912	19	1792	13	0.69	3712	26	1291	12	0.42
D4S139	14.2	6505	50	3447	22	0.54	10862	174	8200	91	0.64
D5S110	14.2	3720	24	2941	22	0.79	3343	24	1444	13	0.48
D10S28	14.2	1758	12	1185	11	0.63	3929	28	1788	14	0.50
D17S79	14.2	1984	14	1522	13	0.72	1755	14	1517	13	0.68

All symbols are defined in text.
* Established using cell line K562 control proficiency test data, Part 2 of this series (20).
† Expected means are average of 1991 and 1997 certified values; standard deviations are calculated with Eq 9, correlations with Eq 10.

ment system is not adequately similar to the systems used to certify SRM 2390. The cause(s) of the measurement discordance should be identified and corrected before again attempting measurement system validation.

Plotting the observed measurement pairs in SLC scattergrams similar to those in Figs. 1 and 2 or as an LPC similar to those in Fig. 3 quantitatively demonstrates measurement validity. Detailed instructions for preparing SLC and LPC graphical tools are provided elsewhere (15,16).

Cell Line K562 Control Material Validation

Laboratories that routinely use a cell line K562-derived material for RFLP control can directly validate their control material against the SRM 2390 cell and genomic K562 components. One or more aliquots of the K562 control should be treated and sized along with the SRM 2390 K562 materials. The band sizes for the control should be very similar to the band sizes measured for the cell and genomic SRM K562 components; all of the band sizes should match the certified values within the 95%/95% bivariate tolerances.

A laboratory's routine K562 control measurements can be used to establish that the SRM 2390-validated control material measurements are truly "typical." Nearly all valid routine measurement pairs should be within the 99% bivariate tolerance ellipses (Table 3); the more tightly clustered the measurements are about the SRM 2390 K562 measurements the better.

Many of the participants in the 1997 interlaboratory recertification study included their laboratory's cell line K562 control material along with the SRM 2390 DNA components. Several of these laboratories also supplied their recent casework K562 control measurements. Figure 7 displays the quantitative validation of one laboratory's ("97-15" in Fig. 3) K562 control measurements. These SLC scattergrams simultaneously establish the close similarity of the K562 control material to the SRM 2390 cell and genomic K562 components and the similarity of the measurements of all three materials to the control measurements in routine casework. The inner ellipses are 95%/95% tolerance limits for the laboratory's casework control measurements, using calculated (Eq 10) rather than measured within-locus correlation.

Non-K562 Control and Secondary Standard Validation

DNA materials from sources other than cell line K562 also can be validated for use as a *Hae*III RFLP control or secondary standard through indirect comparison to SRM 2390 K562 and TAW components. The non-K562 material(s) must first be thoroughly characterized and shown to reliably provide bands of very reproducible size in a number (the more the better, but at least five) of different gels. One or more aliquots of the non-K562 material(s) should be treated and sized along with the SRM 2390 K562 and TAW cell and genomic components. The band sizes for all SRM 2390 K562 and TAW components must match their certified values. The band sizes of the non-K562 material(s) must be typical of routine measurements.

Figure 8 demonstrates the use of a modified LPC in combination with SLCs to quantitatively document indirect validation of a "non-K562" DNA (we use the same K562 control material measurements presented in Fig. 7 for this example, treating the measurements as if they are unrelated to cell line K562). The LPC shown at the upper left establishes the validity of the SRM 2390 K562 and TAW component measurements relative to their SRM certified band sizes and the expected among-laboratory measurement standard deviations. The SLC for each locus establishes that the validation measurements of the "non-K562" material are representative of routine measurements of this material relative to the mean of the routine measurements and the expected among-laboratory standard deviations.

Control Measurement Validation

After the RFLP measurement system and all control materials have been validated, a set of RFLP measurements derived from a given gel can be traced to NIST SRM 2390 by comparison of the gel's control measurements to the SRM-validated control measurements. The same quantitative tools used to validate the control materials can be used to demonstrate traceability. All but one set of measurements in Figs. 7 and 8 are tightly clustered and well centered in the tolerance ellipses and are therefore validly traceable to NIST SRM 2390. There is one set of measurements that is not representative of this laboratory's typical measurement performance; unless a specific problem affecting just the control material can be identified, none of the measurements from this gel should be considered validly traceable.

Acknowledgments

We thank all who have given of their time, effort, and expertise to help make SRM 2390 a useful RFLP standard for the forensic and paternity communities. We particularly thank the following in-

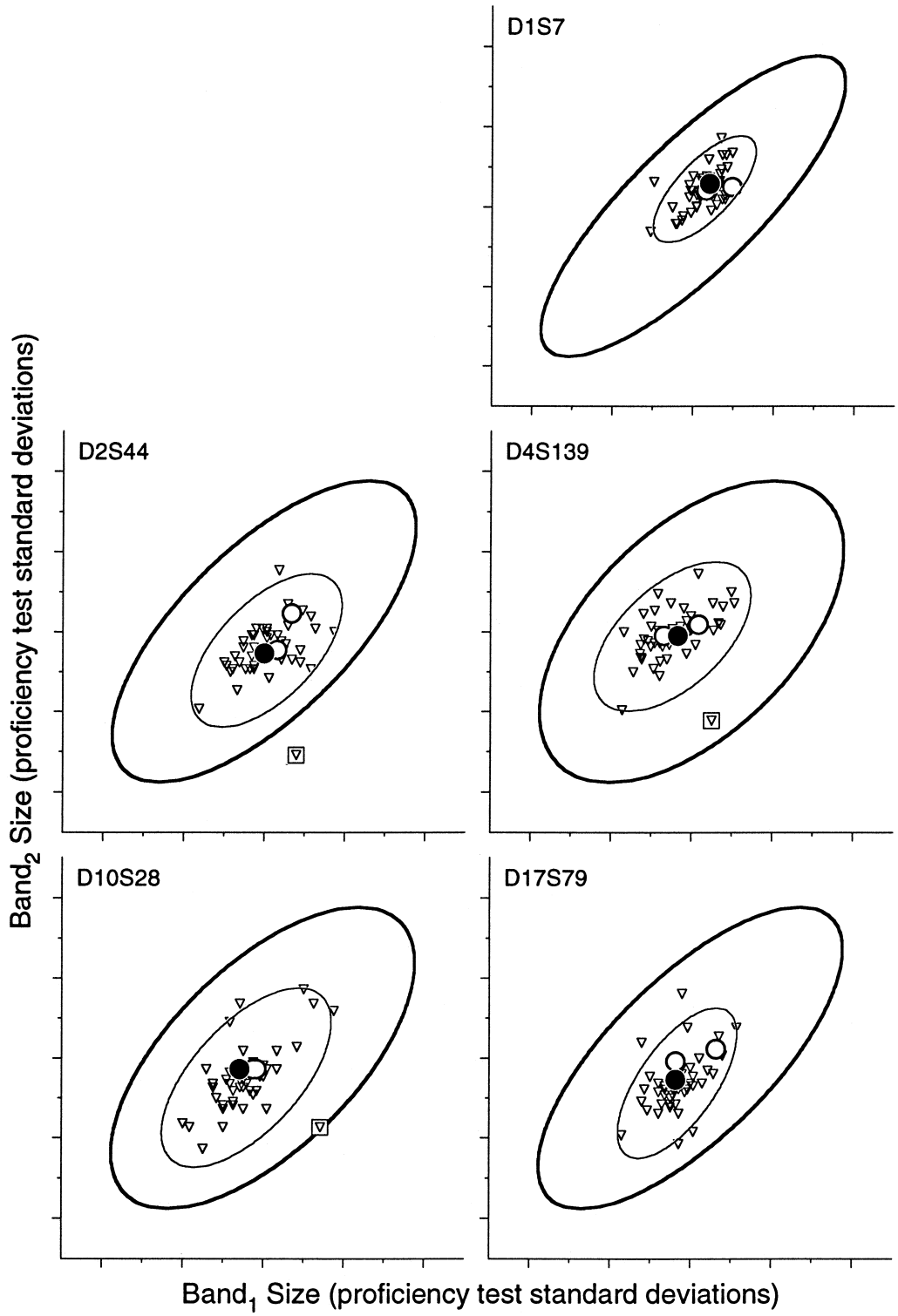

FIG. 7—*Direct demonstration of cell line K562 control measurement traceability. These SLCs simultaneously display one participant's ("97-05") {band1, band2} measurements for SRM 2390 K562 cell and genomic components (open circles), the participant's K562 cell line control included in the SRM gel (dark circle), and the participant's casework K562 control measurements for 1997 (triangles). Each SLC plots the smaller band size of each pair against the larger, is centered on the proficiency test average, and is scaled to span ±5 proficiency test standard deviations. The dark-line ellipse represents the 99% proficiency test tolerance region; the light-line ellipse represents the 95%/95% tolerance region for the participant's casework measurements. The open squares highlight one atypical set of casework control measurements.*

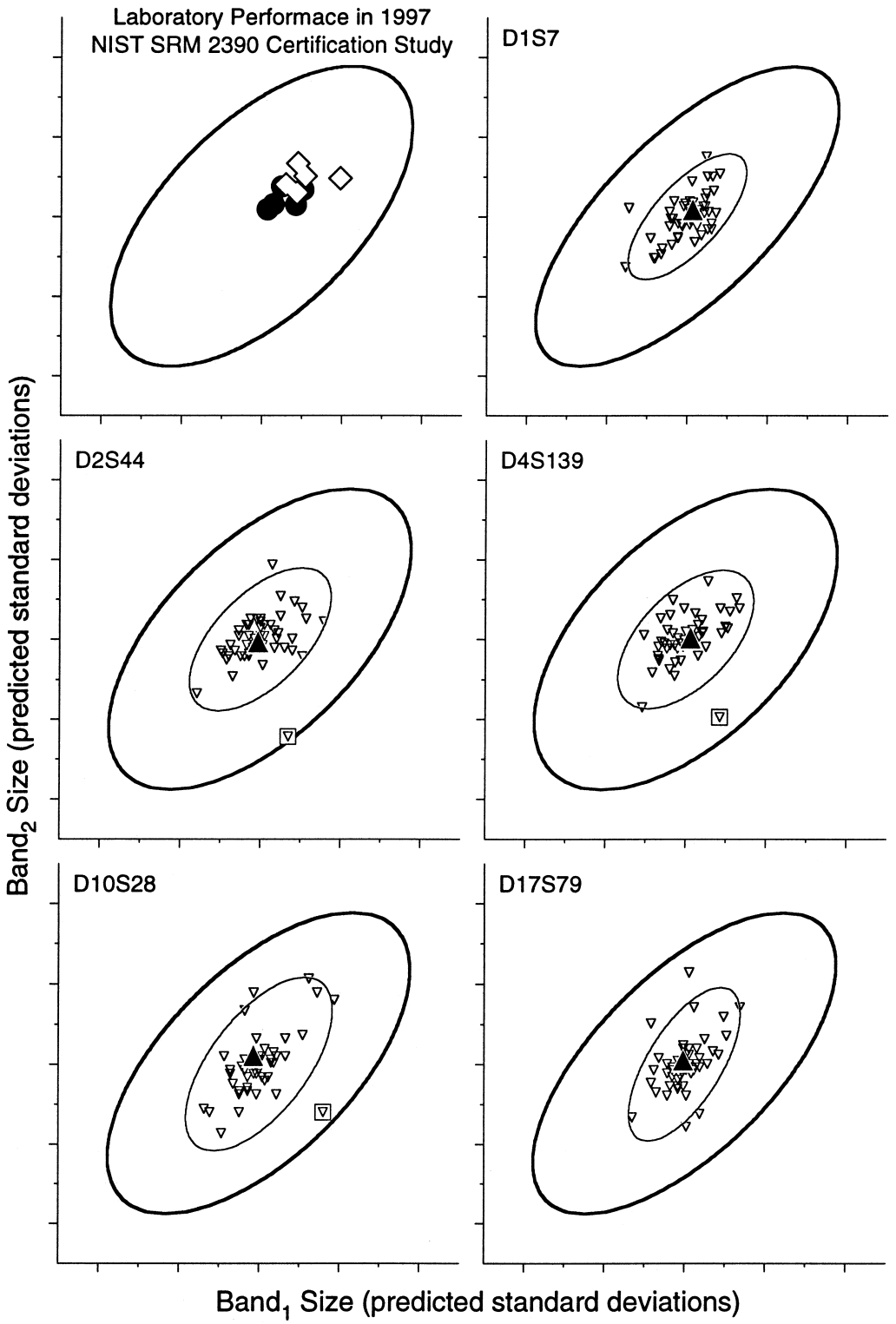

FIG. 8—*Indirect demonstration of "non-K562" control measurement traceability. This combination of LPC and SLCs also simultaneously displays one participant's SRM and control measurements, using the identical data displayed in Fig. 7 but analyzed as if the control material were not derived from cell line K562. The LPC is defined as in Fig. 3 with the exception that only the mean of the cell and genomic values is displayed. Each SLC displays "non-K562" control measurements made in the SRM gel (large solid triangle) along with the routine casework measurements for this material (open triangles). Each SLC is centered on the casework control average and is scaled to span ±5 expected standard deviations of those average band sizes (Eq 9). The dark-line ellipse represents the approximate 99% tolerance region expected for these average sizes; the light-line ellipse represents the 95%/95% tolerance region for the casework control measurements. The open squares highlight one atypical set of casework control measurements.*

dividuals for their patient assistance: Lisa D. Arnell, Larry Ayres, K. Balamurugan, Mark T. Boodée, Eric Buel, Sabrina Christian, Lucy A. Davis, Kristine S. Deters, Patricia A. Dobrowski, Barry W. Duceman, Morris Durbin, George Duncan, Carla J. Finis, Barbara Flowers, Mary Margaret Greer-Ritzheimer, Pat Hankla, Michael Hass, Sandra R. Hill, Jeffrey Johnson, Carll Ladd, Steve Lambert, Anita L. Matthews, Maria T. Mick, Meredith Monroe, Mark S. Nelson, Renee Romero, Roger S. Schneider, Taylor Scott, Mohammad A. Tahir, Edward L. Taylor, Christine S. Tomsey, Stacy C. Warnecke, Linda Winkle, and Jodi Zane.

References

1. Section 210303 "Quality Assurance and Proficiency Testing Standards." Conference Report on HR3355, Violent Crime Control and Law Enforcement Act of 1993 (As passed by the House on August 21, 1994 and subsequently approved by the Senate on August 25, 1994).
2. http://www.cstl.nist.gov/biotech/strbase/
3. Reeder DJ. Interlaboratory studies to certify NIST Standard Reference Material (SRM) 2390 for DNA quality assurance testing. In: Proceedings from the Third International Symposium on Human Identification 1992, Madison, WI: Promega Corporation 1992;271–87.
4. National Institute of Standards and Technology, Certificate of Analysis, Standard Reference Material 2390, DNA Profiling Standard, Standard Reference Materials Program, NIST, Gaithersburg, MD 20899, 1992.
5. National Institute of Standards and Technology, Certificate of Analysis, Standard Reference Material 2391, PCR-based DNA Profiling Standard, Standard Reference Materials Program, NIST, Gaithersburg, MD 20899, 1998.
6. National Institute of Standards and Technology, Certificate of Analysis, Standard Reference Material 2392, Mitochondrial DNA Profiling Standard, Standard Reference Materials Program, NIST, Gaithersburg, MD 20899, 1999.
7. Budowle B. A protocol for RFLP analysis of forensic biospecimens. Proceedings of the International Symposium on the Forensic Aspects of DNA Analysis, 19–23 June 1989, Quantico, VI, 57–62.
8. Budowle B, Baechtel FS. Modifications to improve the effectiveness of Restriction Fragment Length Polymorphism typing. Appl Theor Electrophoresis 1990;1:181–7.
9. Duewer DL, Currie LA, Reeder DJ, Leigh SD, Liu H-K, Mudd JL. Interlaboratory comparison of autoradiographic DNA profiling measurements. 2. Measurement uncertainty and its propagation. Anal Chem 1995;67:1220–31.
10. Waye JS, Fourney RM. Forensic DNA typing of highly polymorphic VNTR loci. In: Forensic Science Handbook, Vol III; R. Saferstein, editor, Englewood Cliffs, NJ: Regents/Prentice Hall 1993;358–415.
11. Holmes DL, Stellwagen NC. The electric field dependence of DNA mobilities in agarose gels: A reinvestigation. Electrophoresis 1990;11: 5–15.
12. Duewer DL, Currie LA, Reeder DJ, Leigh SD, Filliben JJ, Liu H-K, et al. Interlaboratory comparison of autoradiographic DNA profiling measurements. 4. Protocol effects. Anal Chem 1997;69:1882–92.
13. Duewer DL, Lalonde SA, Aubin RA, Fourney RM, Reeder DJ. Interlaboratory comparison of autoradiographic DNA profiling measurements: Precision and concordance. J Forensic Sci 1997;43:465–71.
14. Stolorow AM, Duewer DL, Reeder DJ, Buel E, Herrin G Jr. Interlaboratory comparison of autoradiographic DNA profiling measurements. 3. Repeatability and reproducibility of Restriction Fragment Length Polymorphism band sizing, particularly bands of molecular size >10k base pairs. Anal Chem 1996;68:1941–7.
15. Duewer DL, Liu H-K, Reeder DJ. Graphical tools for RFLP measurement quality assurance: single-locus charts. J Forensic Sci 1999;44:969–77.
16. Gary KT, Duewer DL, Reeder DJ. Graphical tools for RFLP measurement quality assurance: laboratory performance charts. J Forensic Sci 1999;44:978–82.
17. Hotelling H. New light of the correlation coefficient and its transforms. J R Statist Soc Ser B 1953;15:193–225.
18. Hahn GJ, Meeker WQ. Statistical intervals: a guide for practitioners. New York: Wiley, 1991.
19. Hall IJ, Sheldon DD. Improved bivariate normal tolerance regions with some applications. J Quality Tech 1979;11:13–9.
20. Duewer DL, Gary KT, Reeder DJ. RFLP band size standards. 2. Cell line K562 values from 1991–1997 proficiency studies. J Forensic Sci 2000;45: immediately following.
21. Duewer DL, Benzinger EA. Products of partial digestion with *Hae*III. Part 2. Quantification. J Forensic Sci 1997;42:864–72.

Additional information and reprint requests:
Dennis J. Reeder
NIST
100 Bureau Drive, Stop 8311
Gaithersburg, MD 20899-8311
Dennis.Reeder@NIST.gov