David L. Duewer,[1] Ph.D.; Sylvain A. Lalonde,[2] B.Sc.; Rémy A. Aubin,[3] Ph.D.; Ron M. Fourney,[4] Ph.D.; and Dennis J. Reeder,[5] Ph.D.

Interlaboratory Comparison of Autoradiographic DNA Profiling Measurements: Precision and Concordance*

REFERENCE: Duewer DL, Lalonde SA, Aubin RA, Fourney RM, Reeder DJ. Interlaboratory comparison of autoradiographic DNA profiling measurements: precision and concordance. J Forensic Sci 1998;43(3):465–471.

ABSTRACT: Knowledge of the expected uncertainty in restriction fragment length polymorphism (RFLP) measurements is required for confident exchange of such data among different laboratories. The total measurement uncertainty among all Technical Working Group for DNA Analysis Methods laboratories has previously been characterized and found to be acceptably small. Casework cell line control measurements provided by six Royal Canadian Mounted Police (RCMP) and 30 U.S. commercial, local, state, and Federal forensic laboratories enable quantitative determination of the within-laboratory precision and among-laboratory concordance components of measurement uncertainty typical of both sets of laboratories. Measurement precision is the same in the two countries for DNA fragments of size 1000 base pairs (bp) to 10,000 bp. However, the measurement concordance among the RCMP laboratories is clearly superior to that within the U.S. forensic community. This result is attributable to the use of a single analytical protocol in all RCMP laboratories. Concordance among U.S. laboratories cannot be improved through simple mathematical adjustments. Community-wide efforts focused on improved concordance may be the most efficient mechanism for further reduction of among-laboratory RFLP measurement uncertainty, should the resources required to fully evaluate potential cross-jurisdictional matches become burdensome as the number of RFLP profiles on record increases.

KEYWORDS: forensic science, DNA typing GM9947, GM9948, K562 cell line controls, electrophoresis, graphical data analysis, restriction fragment length polymorphism

[1] Analytical Chemistry Division, Chemical Science and Technology Laboratory, National Institute of Standards and Technology, Gaithersburg, MD 20899.

[2] Biology Section, Forensic Laboratory Edmonton, Royal Canadian Mounted Police, Edmonton, Alberta T5V 1B7.

[3] Life Sciences Division—Biotechnology, Health Canada, Sir F.G. Banting Research Center, Ottawa, Ontario K1A OL2.

[4] DNA Methods and Data Base Unit, Central Forensic Laboratory, Royal Canadian Mounted Police, Ottawa, Ontario K1G 3M8.

[5] Biotechnology Division, Chemical Science and Technology Laboratory, National Institute of Standards and Technology, Gaithersburg, MD 20899.

* Certain commercial equipment, instruments, or materials are identified in this report to specify adequately the results of the interlaboratory study and the experimental procedures used. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, the Royal Canadian Mounted Police, or Health Canada; nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

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

Received 27 May 1997; and in revised form 18 July 1997; accepted 2 Oct. 1997.

The electrophoretic mobility of DNA through an agarose gel is a function of the mass, composition, and three-dimensional configuration each DNA fragment (1). DNA profiling measurements obtained with restriction fragment length polymorphism (RFLP) techniques do *not* estimate any absolute quantity but rather produce a characteristic ''apparent size'' that is a complex function of molecular properties, gel composition, and electrophoretic parameters. Reproducible RFLP profiling measurements across different analysts, different times, and different laboratories thus require similar use of similar measurement systems and even greater than typical attention to measurement quality control and assurance.

The Technical Working Group on DNA Analysis Methods (TWGDAM) has sponsored a series of studies designed to document the evolution of DNA profiling protocols, to identify opportunities for improvement, and to communicate results and suggestions to the human identification communities (2). Using these studies and control sample results from routine casework, we have documented that RFLP measurement uncertainties are small enough to permit RFLP profiling to be reliably exchanged among laboratories using similar analytical protocols (assuming that the laboratories appropriately monitor their results through control and reference samples, internal quality assurance programs, and external proficiency demonstrations) (3–6).

While the RFLP protocols used by U.S. and Canadian TWGDAM member laboratories differ in detail, their overall approaches are quite similar: *Hae*III restriction endonuclease enzymatic digestion of extracted DNA, agarose gel electrophoresis, Southern blot immobilization of separated DNA fragments, autoradiographic or lumographic imaging of DNA fragments (bands) containing selected genetic loci, determination of the relative electrophoretic migration distance of sample and calibrant bands using digital image analysis, and conversion of the relative migration distances to an approximate molecular size (band size) expressed as the number of DNA basepairs (bp) (7–9).

U.S. commercial, local, state, and Federal laboratories use protocols derived from a system developed by the Federal Bureau of Investigation (FBI). These measurement systems are characterized by the use of Tris-acetate EDTA (TAE) buffers and the female cell line K562 as the control sample. A variety of materials, different types of equipment, and diverse analyst training requirements exist in the various U.S. forensic laboratories.

The Canadian Federal Laboratories use a protocol developed by and originally implemented in a central Royal Canadian Mounted Police (RCMP) research and training laboratory. This system is characterized by use of Tris-borate EDTA (TBE) buffers, cell lines

GM9948 (male) and GM9947a (female) as control samples, and inclusion of a bloodstain from one of six donors as a blind internal standard (BIS). In contrast to the U.S. laboratories, the six RCMP laboratories that make RFLP measurements adhere to a single protocol: all materials and equipment are nominally identical, all equipment is installed, maintained, and used by individuals initially trained by a core group of instructors.

No differences of forensic significance have been observed between U.S. and Canadian laboratories for any RFLP measurement, either in TWGDAM-sponsored studies or in formal proficiency demonstrations. Given the readily apparent differences in electrophoretic migration patterns attributable to the TAE and TBE buffer systems (10), this consistent agreement testifies to the robust nature of the basic RFLP profiling system and the care and attention to detail of the many analysts involved. However, a quantitative comparison of the measurement characteristics of the U.S. and Canadian systems is necessary prior to international exchange of DNA casework data.

We present here an analysis of within-laboratory measurement precision and among-laboratory measurement concordance, using cell line control data from casework. We contrast these measurement accuracy characteristics of the RCMP laboratories with those observed for a group of U.S. laboratories.

Methods and Materials

Data from RCMP Forensic Laboratories

The RCMP computerized their collection of RFLP data in late 1993. Cell line control and BIS sample data from casework and proficiency gels analyzed from that time through mid-1996 have been used in this study. The only editing of the data has been removal of exact duplicate records.

RCMP laboratories—There are six RCMP forensic laboratories using RFLP in casework: the Central Laboratory at Ottawa and the Regional laboratories at Halifax, Winnipeg, Regina, Edmonton, and Vancouver. The amount of data available from each laboratory varies widely, with data from more than 900 casework gels from Edmonton to just 40 from newly on-line Regina. To date, 44 RCMP forensic specialists have contributed to RFLP casework.

Electrophoretic conditions and gel format—The RCMP protocol specifies 14-cm-long by 20-cm-wide, 1% low electroendoosmality (EEO) agarose gels with 1×TBE (89 mmol/L Tris, 89 mmol/L borate, 2 mmol/L EDTA, pH 8.0) buffer with ethidium bromide (EB) added after electrophoresis for photographic purposes. Relative to similar gels using the 1×TAE (40 mmol/L Tris, 20 mmol/L acetate, 1 mmol/L EDTA, pH 8.0) buffer of FBI-derived protocols, the RCMP system was designed to provide greater separation among bands of size 600 to about 6000 bp. However, the RCMP system provides less separation for band sizes above 10,000 bp and does not resolve sizing ladder components larger than 15,000 bp.

Each RCMP casework gel uses a maximum of 22 lanes, with the BRL 23,000 bp sizing ladder (Life Technologies, Gaithersburg, MD) in the first and last lanes and every third to fifth lane in between. Two cell line controls and a BIS sample are usually located in lanes 2, 3, and 4. Known and questioned samples are loaded in the remaining lanes; sizing ladders are never separated by more than five samples.

Genetic loci—Data are available for monomorphic α-satellite DNA loci D7Z2 (human) and DYZ1 (Y chromosome) for nearly 100% of casework gels evaluated by the RCMP laboratories. Polymorphic loci are evaluated in approximately the following percentage of gels: D5S110 (95% of gels), D2S44 (85%), D10S28 (80%), D1S7 (75%), D4S139 (40%), and D17S79 (20%). Some casework data are available for locus D16S85 (about 5% of gels); however, only three of the six RCMP laboratories had evaluated five or more casework D16S85 autoradiographs as of late 1996.

Cell line controls—The immortalized female line GM9947a and the male line GM9948 are used as restriction, loading, electrophoresis, and sizing controls in every casework gel (11). These lines were adopted for use as controls after extensive evaluation in 1989 and 1990. Both GM9947a and GM9948 have complete, gender-normal sets of chromosomes. Five or more values are available from all six RCMP laboratories for thirteen GM9947a bands (two each for loci D1S7, D2S44, D4S139, D5S110, D10S28, and D17S79; one for D7Z2) and 14 for GM9948 (the same plus one for DYZ1).

Blind internal standards—Every RCMP casework gel also includes a BIS bloodstain, providing control for the sample extraction and DNA quantification components of the protocol. The BIS sample used in a given casework gel is randomly assigned from a set of six, each from a different donor. The identity of the BIS sample is unknown to the examiner at the time of extraction.

A total of eight donors have provided BIS samples since 1993, six male and two female. To reflect better the nature of typical casework, use of the female-donor BIS-B and BIS-D samples was discontinued in late 1994. They were replaced with the male-donor BIS-G and BIS-H. In consequence, there are five or more values available from all six RCMP laboratories for only 48 (of the 109 possible) BIS bands, mostly from BIS-A, -C, -E, and -F.

Data from U.S. Forensic Laboratories

A number of U.S. commercial, local, state, and Federal forensic laboratories have provided K562 results for our use in various studies (3–6,12). Thirty sets of autoradiographic casework data were available by late 1996. Some of the data sets contain data from as early as 1989, but the majority consist of data collected from about 1992 through 1995. A very few ''outlier'' data (less than 0.1%), defined as one or more bands at a given locus of size more than ±5% different from the data set's median, have been deleted from the original data sets. All such outliers were traced to manual data entry errors. While of concern with regard to database quality assurance, such errors do not reflect components of measurement uncertainty germane to this study.

Although complete protocol descriptions were provided with many of these sets, we have no detailed information on more than half of them. However, previous studies have documented that the U.S. forensic community used both medium and low EEO agarose, a range of EB levels in the running buffer, at least four different commercial sizing ladders, many different gel formats, a variety of electrophoretic equipment and conditions, and several different image analysis systems (6). Therefore, each of the 30 data sets represents an essentially unique implementation of the basic FBI (1×TAE buffer) protocol.

Cell line K562 is used as for sizing control by all U.S. TWGDAM laboratories and is the only control or standard sample used by many. It is the only control currently recognized by the FBI's CODIS system (13), and is a component of NIST's Standard Reference Material® 2390 DNA Profiling Standard (14). The K562 control is typically loaded in lane 2, with sizing ladders in the first and last lanes as well as interspersed between every two to four known and unknown samples.

Five or more values from six or more data sets are available for

12 cell line K562 bands: two each from genetic loci D2S44 (7746 data in 30 sets), D1S7 (5664 in 30), D4S139 (6932 in 28), D10S28 (6151 in 26), D17S79 (4019 in 22), and D5S110 (929 in 12). While some data have been provided for D14S13 (1251 data in 5 sets), D17S26 (624 in 4), D7S467 (208 in 3), D7Z2 (91 in 2), and D16S85 (14 in 1), more data sets are needed to make quantitative use of this additional information.

Results and Discussion

The total interlaboratory uncertainty in a given measurement has two components, within-laboratory precision and among-laboratory concordance (15). *Precision* describes the extent of agreement among all measurements from a given laboratory and may be thought of as summarizing random fluctuations afflicting any measurement process (16). *Concordance* describes the extent of agreement among the measurement average characteristic of each laboratory and may be thought of as summarizing fixed differences among the different laboratories.

The histograms in Fig. 1 display cell line control sizing data distributions for the RCMP controls GM9947a and GM9948 and for the U.S. control K562. The bp sizes for the BRL 23,000 bp ladder components are shown to provide context: the total distribution of every band is contained in less than the average spacing between ladder components. Distributions for several bands are shown at 10-fold higher resolution, along with Gaussian curves having the same means and SDs. All histograms are sufficiently ''normal'' for application of traditional statistical analysis tools (17).

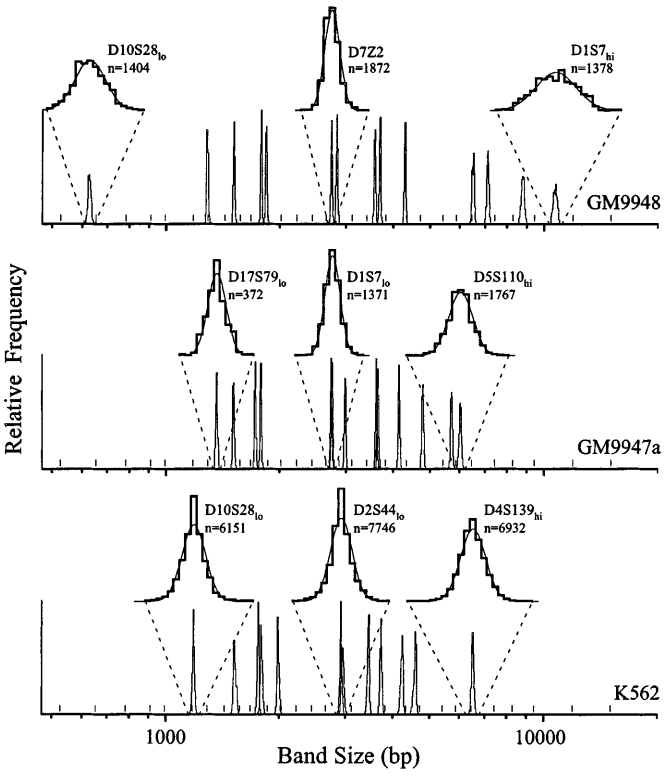

FIG. 1—*Size distributions for cell line control measurements. Each of the graphical segments presents histogram representations of all locus D1S7, D2S44, D4S139, D5S110, D10S28, D17S79, D7Z2, and DYZ1 band measurements used in this study. The top segment represents the male cell line GM9948, the middle segment represents the female line GM9947a, and the bottom segment represents the female line K562. Each histogram has been scaled to have the same area. Tic marks above each X-axis represent the nominal size of a band in the BRL 23,000 bp sizing ladder. Histograms for the smallest, largest, and a mid-size band for each cell line are shown at 10-fold higher bp resolution. Each of these expanded histograms is labeled with the band code and the number of measurements available, along with the Gaussian distribution for the observed band size mean and SD.*

Within-Laboratory Precision

The measurement precision characteristic of a particular laboratory can be estimated as the standard deviation (SD) of a set of measurements made repeatedly on identical samples:

$$\mathrm{SD}_i = \sqrt{\frac{\sum_{j=1}^{N_j} (X_{ij} - \overline{X}_i)^2}{N_i - 1}} \qquad (1)$$

where index i designates a given laboratory, index j designates a given measurement, X_{ij} is a particular measurement value, N_i is the number of values, and $\overline{X}_i$ is the mean of the values for the ith laboratory:

$$\overline{X}_i = \frac{\sum_{j=1}^{N_i} X_{ij}}{N_i} \qquad (2)$$

The smaller the value of SD_i, the better the laboratory's measurement precision.

In general, different laboratories will have different measurement precision. However, the measurement precision ''typical'' of the sampled laboratories can be estimated by appropriately pooling all the individual SD_i:

$$\mathrm{SD}_{\mathrm{prec}} = \sqrt{\frac{\sum_{i=1}^{M} (N_i - 1)\mathrm{SD}_i^2}{\sum_{i=1}^{M} N_i - M}} \qquad (3)$$

where M is the total number of laboratories (18). The smaller $\mathrm{SD}_{\mathrm{prec}}$, the better the precision expected for laboratories ''typical'' of the ones studied.

Figure 2 displays $\mathrm{SD}_{\mathrm{prec}}$ for the RCMP laboratories, $\mathrm{SD}_{\mathrm{prec}}^{\mathrm{RCMP}}$, and for the U.S. laboratories, $\mathrm{SD}_{\mathrm{prec}}^{\mathrm{US}}$. Very nearly the same measurement precision for a given band is expected for RCMP and U.S. laboratories. The $\mathrm{SD}_{\mathrm{prec}}^{\mathrm{RCMP}}$ are somewhat smaller than $\mathrm{SD}_{\mathrm{prec}}^{\mathrm{US}}$ for bands of size less than 4000 bp, perhaps reflecting the improved band separation in this size range provided by use of the TBE buffer.

The available data are insufficient for quantitative comparison of sizing precision above 6500 bp, the size of the largest K562 band. The $\mathrm{SD}_{\mathrm{prec}}^{\mathrm{RCMP}}$ and $\mathrm{SD}_{\mathrm{prec}}^{\mathrm{US}}$ are very similar in this size region. The two largest components of the routinely used RFLP sizing ladders have nominal size of about 23,000 bp and 15,000 bp. The 15,000 band is the largest ladder component routinely resolved in the RCMP protocol, with the next largest component of size 11,919 bp. Since the relationship between nominal ladder component size and electrophoretic migration distance is less well defined in the extreme end segments, we expect that $\mathrm{SD}_{\mathrm{prec}}^{\mathrm{RCMP}}$ for bands of size 12,000 bp to 15,000 bp and $\mathrm{SD}_{\mathrm{prec}}^{\mathrm{US}}$ for bands of size 15,000 bp to

23,000 bp may be somewhat greater than expected from extrapolation of the available data. We also expect that SD_{prec}^{RCMP} may exceed SD_{prec}^{US} for bands of size 12,000 bp to 15,000 bp.

Among-Laboratory Concordance

The concordance of the entire set of measurements over all the laboratories can be estimated as the SD over all the individual mean values:

$$SD_{con} = \sqrt{\frac{\sum_{i=1}^{M} (\overline{X}_i - \overline{\overline{X}})^2}{M - 1}} \quad (4)$$

where $\overline{\overline{X}}$ is the mean-of-means:

$$\overline{\overline{X}} = \frac{\sum_{i=1}^{M} \overline{X}_i}{M} \quad (5)$$

The smaller SD_{con}, the better the expected measurement concordance among typical laboratories.

Figure 3 displays SD_{con} for the RCMP laboratories, SD_{con}^{RCMP}, and for the U.S. laboratories, SD_{con}^{US}. There is much higher concordance (smaller SD_{con}) among the RCMP laboratories than among the U.S. laboratories, surely a direct result of the RCMP's diligent use of a common protocol among all their laboratories.

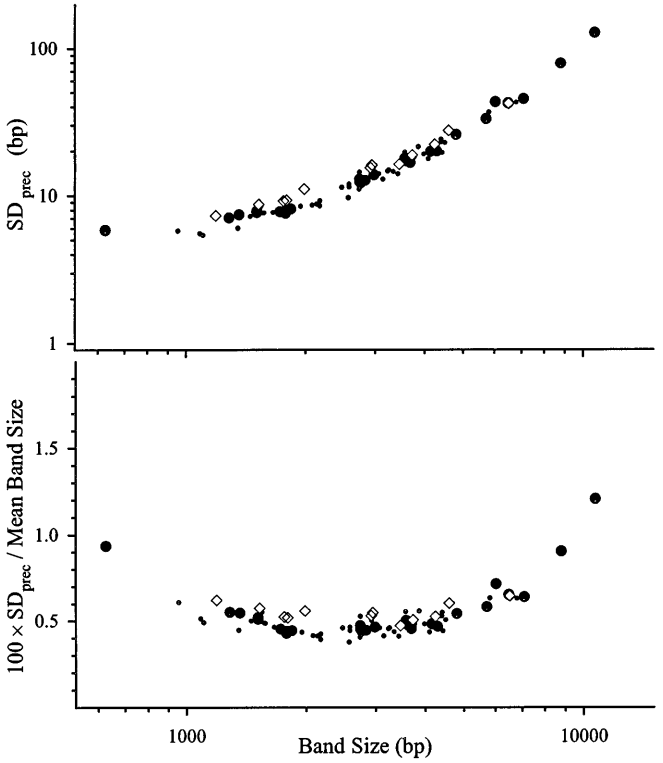

FIG. 2—*Within-laboratory measurement precision. The upper graphical segment presents the observed SD_{prec} for all bands where at least six laboratories provided at least five measurements. There are 27 SD_{prec}^{RCMP} values for cell lines GM9947a and GM9948 (large solid circles), 48 SD_{prec}^{RCMP} values for BIS bloodstains (small solid circles), and 12 SD_{prec}^{US} for cell line K562 (open diamonds). The lower graphical segment presents the same data as the upper, expressing the SD values as a percentage of the mean size of the band.*

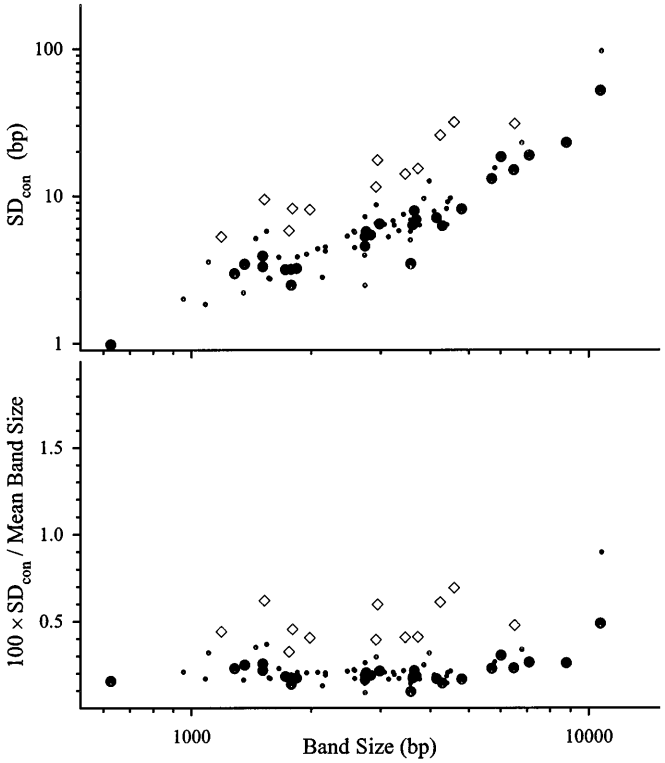

FIG. 3—*Among-laboratory measurement concordance. Legend as in Fig. 2, displaying SD_{con} rather than SD_{prec}.*

Total Measurement Uncertainty

The total interlaboratory measurement uncertainty expected for any given measurement is estimated as a composite SD, defined by appropriately combining SD_{prec} and SD_{con}

$$SD_{tot} = \sqrt{SD_{prec}^2 + SD_{con}^2} \quad (6)$$

The smaller SD_{tot}, the lower the measurement uncertainty expected for a given value reported by a typical laboratory.

Figure 4 displays SD_{tot}^{RCMP} for the RCMP laboratories and SD_{tot}^{US} for the U.S. laboratories. While the total measurement uncertainty observed for both the RCMP and the U.S. laboratories is less than 1% of band size over most of the measurement domain, the excellent interlaboratory concordance of the RCMP data clearly facilitates comparing measurements from different RCMP laboratories.

The relationship between SD_{tot} and bp can be approximated as

$$\hat{SD}_{tot} = A\left(1 + \frac{bp}{B}\right)^C \quad (7)$$

where A, B, and C are empirical coefficients estimated using nonlinear regression (4,5). (The functional form of Eq 7 is the uncertainty-propagation rule for sigmoidal calibration, here applied to the observed relationship between DNA fragment band size and electrophoretic migration distance (19).) In a designated interlaboratory study using bloodstains from donors known to have at least one D4S139 or D17S79 allele larger than 10,000 bp, the expected SD_{tot} for a group of 20 TWGDAM laboratories (18 U.S. laboratories, 2 Canadian) was estimated to be

$$\hat{\mathrm{SD}}_{\mathrm{tot}}^{\mathrm{TWG}} = 7.5\left(1 + \frac{\mathrm{bp}}{19500}\right)^{7.1} \quad (8)$$

for bands of size 1000 to 20,000 bp (5). The regression analysis used in that study can be applied to the RCMP data displayed in Fig. 4, resulting in the following estimate:

$$\hat{\mathrm{SD}}_{\mathrm{tot}}^{\mathrm{RCMP}} = 4.1\left(1 + \frac{\mathrm{bp}}{10200}\right)^{5.0} \quad (9)$$

for RCMP laboratory measurement of bands of size 600 to 11,000 bp. The relationships described by Eqs 8 and 9 are displayed in Fig. 4.

Concordance Stability Among Bands

$\mathrm{SD}_{\mathrm{tot}}$ estimates the magnitude of interlaboratory measurement differences, but it does not address whether the pattern of differences among the laboratories is the same for all DNA fragments. If the differences are predictable, measurement concordance could be improved through mathematical rather than experimental standardization.

Recalling that $\overline{X}_i$ is the expected size of a given fragment measured in laboratory i, the pattern of potentially laboratory-specific differences should be reflected in the pattern of differences among the $\overline{X}_i$. Comparison of differences among the various known fragments is simplified by ''standardizing'' each fragment's set of M $\overline{X}_i$ values to have the same location and scale:

$$Z_i = \frac{\overline{X}_i - \overline{\overline{X}}}{\mathrm{SD}_{\mathrm{tot}}} \quad (10)$$

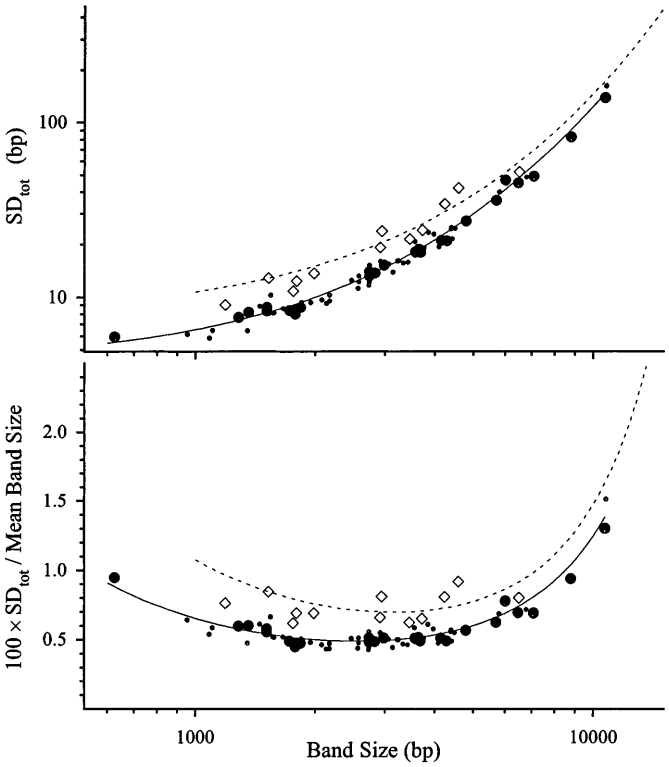

FIG. 4—*Total interlaboratory measurement uncertainty. Legend as in Fig. 2, displaying SD_{tot} rather than SD_{prec}. The Eq 8 estimate of SD_{tot} for all TWGDAM laboratories, derived in Ref 5, is denoted (dotted line). The Eq 9 estimate for SD_{tot}^{RCMP} is denoted (solid line).*

These Z_i represent the number of expected SD between a given $\overline{X}_i$ and $\overline{\overline{X}}$ (or other location metric, such as a reference laboratory's mean values or the NIST certified K562 values (14)). If the among-laboratory measurement difference patterns are constant, all Z_i for all laboratories should have about the same value.

Table 1 details the application of Eq 10 to the K562 data provided by three U.S. laboratories, where $\overline{\overline{X}}$ is estimated as the NIST-certified K562 values and $\mathrm{SD}_{\mathrm{tot}}$ is estimated using Eq 8, $\hat{\mathrm{SD}}_{\mathrm{tot}}^{\mathrm{TWG}} = 7.51(1 + \overline{\mathrm{X}}/19500)^{7.1}$. Table 1 also displays the mean and SD of the standardized differences (the Z_i) for the three laboratories, with the mean estimating the ''expected offset'' for each laboratory from the NIST values and the SD providing a guide to the consistency of the offset over all the different K562 bands. These three laboratories represent the extreme range of average differences from the NIST values of the 21 laboratories examined. The laboratory having the greatest negative average difference (-0.7 $\mathrm{SD}_{\mathrm{tot}}$) is denoted ''L,'' the laboratory closest to the NIST values (0.1 $\mathrm{SD}_{\mathrm{tot}}$) is denoted ''M,'' and the laboratory with the greatest positive average difference (0.9 $\mathrm{SD}_{\mathrm{tot}}$) is denoted ''R.''

The lower panel of Fig. 5 displays the distributions for all 21 U.S. laboratories that provided data for at least 10 of the 12 available K562 bands. The upper panel of Fig. 5 displays the distribution of Z_i for each of the six RCMP laboratories, with each distribution representing the sum of 27 control Z_i and 48 BIS Z_i. The Z_i for the RCMP laboratories are calculated using the grand mean of all six RCMP laboratories as $\overline{\overline{X}}$, but with the same Eq 8 estimate for $\mathrm{SD}_{\mathrm{tot}}$. Using the same estimate of variability for both U.S. and RCMP laboratories facilitates the comparison between the two groups. For graphical clarity, the distributions are shown in smooth ''probability density function'' form rather than as histograms (20).

All six RCMP laboratory distributions, each composed of 75 individual Z_i, are of remarkably similar shape: unimodal, symmetric, and narrow. With such high stability in these among-laboratory measurement differences, mathematical adjustment (albeit unnecessary) would further improve measurement comparability. In contrast, the distributions for U.S. laboratories, each the sum of just 10 or 12 individual Z_i, are quite dissimilar: unimodal to multimodal, symmetric to strongly skewed, narrow to very wide. (See Table 1; the SD for the Z_L is more than twice that for Z_M and Z_R.) With such instability in these among-laboratory differences in K562, mathematical adjustment (however desirable) would not be beneficial.

The causative agent(s) of the small remaining differences among the RCMP laboratories has yet to be identified; whatever its origin, the fixed pattern of differences suggests that all DNA fragments are similarly affected regardless of size or composition. The relatively large and band-specific differences among U.S. laboratories suggest that many different variables are involved. Previous studies have identified fragment-specific differences among different sizing ladders, agarose grades, and gel formats as well as fragment-independent differences apparently related to minute differences in loading buffer composition (5,6).

Extraction and DNA Quantity Contributions to Measurement Uncertainty

The dependencies of $\mathrm{SD}_{\mathrm{prec}}^{\mathrm{RCMP}}$, $\mathrm{SD}_{\mathrm{con}}^{\mathrm{RCMP}}$, and $\mathrm{SD}_{\mathrm{prec}}^{\mathrm{RCMP}}$ on band size (Figs. 2–4) are virtually identical for the cell line and BIS bands. Thus, at least within the RCMP laboratories and for unweathered samples, sample extraction and DNA quantity determination stages of the RFLP measurement process do not appreciably contribute to band size measurement uncertainty.

TABLE 1—*Standardized among-laboratory measurement differences for three U.S. laboratories.*

K562	$\overline{\overline{X}}^a$	SD_{tot}*	Laboratory$_L$			Laboratory$_M$			Laboratory$_R$		
			$\overline{X}_L$*	$X_L - \overline{\overline{X}}^a$	Z_L†	$\overline{X}_M$*	$X_M - \overline{\overline{X}}^a$	Z_M†	$\overline{X}_R$*	$X_R - \overline{\overline{X}}^a$	Z_R†
D10S28$_{lo}$	1182	11	1180	−2	−0.18	1186	4	0.35	1193	11	0.97
D17S79$_{lo}$	1520	13	1524	4	0.31	1521	1	0.08			
D10S28$_{hi}$	1757	14	1744	−13	−0.94	1757	0	0.00	1767	10	0.72
D2S44$_{lo}$	1791	14	1785	−7	−0.46	1788	−3	−0.21	1795	4	0.29
D17S79$_{hi}$	1982	15	1977	−5	−0.34	1982	0	0.00			
D2S44$_{hi}$	2907	20	2887	−21	−1.02	2912	5	0.22	2931	24	1.19
D5S110$_{lo}$	2942	20	2956	14	0.69	2936	−6	−0.29	2965	23	1.11
D4S139$_{lo}$	3438	24	3411	−27	−1.14	3447	9	0.38	3461	23	0.97
D5S110$_{hi}$	3714	26	3714	0	0.00	3704	−10	−0.39	3738	24	0.91
D1S7$_{lo}$	4231	30	4162	−69	−2.28	4225	−6	−0.20	4240	9	0.30
D1S7$_{hi}$	4571	33	4490	−81	−2.42	4582	11	0.33	4611	40	1.20
D4S139$_{hi}$	6474	57	6440	−35	−0.60	6518	44	0.77	6569	95	1.65
				Mean	−0.70		Mean	0.09		Mean	0.93
				SD	0.94		SD	0.34		SD	0.42

*Units: bp.
†Units: number of SD_{tot}.

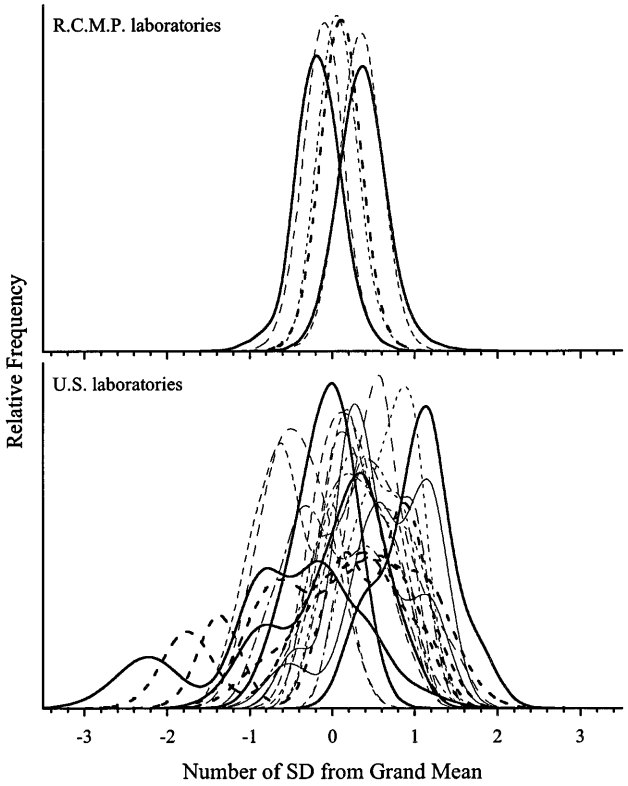

FIG. 5—*Stability of among-laboratory measurement differences. The upper graphical segment presents the observed average measurement differences among the 6 RCMP laboratories, based on the 27 cell line control and 48 BIS bands where all laboratories provided at least five measurements. The lower graphical segment presents the observed average measurement differences among the 21 U.S. laboratories that provided at least 5 measurements for at least 10 bands.*

Conclusions

The RCMP's inclusion of two genomically normal cell line controls (GM9948 and GM9947a) and a blind internal standard in every casework gel enables complete demonstration of RFLP measurement performance at all loci. It also generates data of sufficient number and diversity for efficient evaluation of measurement characteristics. Given its greater range of band sizes across genetic loci of forensic interest, inclusion of the male cell line GM9948 as a second control would enhance the ability to evaluate (and demonstrate) the analytical quality of U.S. RFLP casework gels.

The RCMP among-laboratory measurement concordance is superior to that achieved within the U.S. forensic community. This excellent concordance results from use of a common protocol (the same methods, materials, equipment, equipment maintenance, analyst training, master database, and management) in all RCMP laboratories. While complete experimental standardization is probably not feasible among the diverse U.S. forensic community, consensus standards for the most critical equipment and materials may be possible. Concordance among U.S. forensic laboratories cannot be improved by simple mathematical standardization of the data.

Within-laboratory RFLP measurement precision is nearly the same for RCMP and U.S. forensic laboratories, at least for bands of size 1000 bp to 11,000 bp. This level of precision may well represent a practical limit for multi-analyst laboratories using autoradiographic RFLP technologies. Should it become desirable to further improve the total measurement uncertainty among U.S.

laboratories using such RFLP methods, community-wide efforts towards improving concordance are likely to be more productive than individual efforts at improving precision.

Acknowledgments

We thank all the forensic analysts and laboratory staff, both within the RCMP and U.S. laboratories, who make and keep track of DNA profiling measurements. It is their professional care and diligence, unsung but not unappreciated, and passion for doing the job *right* that make real the promise of forensic human identification technologies.

References

1. Holmes DL, Stellwagen NC. The electric field dependence of DNA mobilities in agarose gels: A reinvestigation. Electrophoresis 1990; 11:5–15.
2. Kearney JJ. Summary of the First Meeting of the Technical Working Group for DNA Analysis Methods. Crime Laboratory Digest 1988;15:115–7.
3. Mudd JL, Baechtel FS, Duewer DL, Currie LA, Reeder DJ, Leigh SD, et al. Interlaboratory comparison of autoradiographic DNA profiling measurements. 1. Data and summary statistics. Anal Chem 1994;66:3303–17.
4. 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.
5. Stolorow AM, Duewer DL, Reeder DJ, Buel E, Herrin G. Interlaboratory comparison of autoradiographic DNA profiling measurements. Part 3. Repeatability and reproducibility of RFLP band sizing, particularly bands of molecular size >10k base pairs. Anal Chem 1996;68:1941–7.
6. Duewer DL, Currie LA, Reeder DJ, Leigh SD, Liu H-K, Mudd JL. Interlaboratory comparison of autoradiographic DNA profiling measurements. 4. Protocol effects. Anal Chem 1997;69: 1882–1892.
7. Budowle B, Baechtel FS. Modifications to improve the effectiveness of Restriction Fragment Length Polymorphism typing. Appl Theor Electrophor 1990;1:181–7.
8. Waye JS, Fourney RM. Forensic DNA typing of highly polymorphic VNTR loci. In: Forensic Science Handbook, Vol III; Saferstein R, Ed; Englewood Cliffs (NJ):Regents/Prentice-Hall 1993; 358–415.
9. Technical Working Group for DNA Analysis Methods. Guidelines for a Quality Assurance (QA) Program for DNA Analysis. Crime Laboratory Digest 1991;18:44–75.
10. Allen RC, Budowle B. Gel Electrophoresis of Proteins and Nucleic Acids: Selected Techniques. Berlin:Walter de Gruyter, 1994.
11. Fregeau CJ, Aubin RA, Elliot JC, Gill SS, Fourney RM. Characterization of human lymphoid cell lines GM9947 and GM9948 as intra- and interlaboratory reference standards for DNA typing. Genomics 1995;28:184–97.
12. Duewer DL, Benzinger EA. Products of partial digestion with Hae III. Part 2. Quantification. J Forensic Sci 1997;42(5):864–872.
13. Federal Bureau of Investigation. Draft, National DNA Index System (NDIS) Standards for CODIS Acceptance of DNA RFLP Data at NDIS. Washington:FBI Laboratory Room 3658, 20 May 1996.
14. National Institute of Standards and Technology. Certificate of Analysis, Standard Reference Material 2390, DNA Profiling Standard. Gaithersburg (MD):NIST Standard Reference Materials Program 1992.
15. Eisenhart C. Realistic evaluation of the precision and accuracy of instrument calibration systems. J Res NBS 1963;67C:21–47.
16. International Organization for Standardization. Precision of Test Methods—Determination of Repeatability and Reproducibility for a Standard Test by Inter-laboratory Tests, ISO 5725. Geneva: ISO 1986.
17. Youden WJ. Experimentation and Measurement, NIST Special Publication 672. Gaithersburg (MD):NIST Standard Reference Materials Program 1991.
18. Mandel J. Evaluation and Control of Measurements. New York: Marcel Dekker, 1991;48–51.
19. O'Connell MA, Belanger BA, Haaland PD. Calibration and assay development using the four-parameter logistic model. Chemom Intell Lab Syst 1993;20:97–114.
20. Silverman BW. Density Estimation for Statistics and Data Analysis. London:Chapman and Hall, 1986.

Additional information and reprint requests:
Dennis J. Reeder
Chemistry/222, Room A353
National Institute of Standards and Technology
Gaithersburg, MD 20899