TECHNICAL NOTE

David L. Duewer,[1] Ph.D.; Kenneth T. Gary;[1] and Dennis J. Reeder,[1] Ph.D.

RFLP Band Size Standards: Cell Line K562 Values from 1991 to 1997 Proficiency Studies*

REFERENCE: Duewer DL, Gary KT, Reeder DJ. RFLP band size standards: cell line K562 values from 1991 to 1997 proficiency studies. J Forensic Sci 2000;45(5):1106–1118.

ABSTRACT: Cell line K562 is the *de facto* forensic control material for forensic restriction fragment length polymorphism (RFLP) DNA profiling in the U.S. Fifty-one proficiency tests conducted from 1991 through 1997 enable a detailed description of RFLP measurement performance during this period. Sufficient data are available to define reference distributions for all commonly utilized and many less commonly reported genetic loci, for both *Hae*III- and *Hinf* I-based RFLP systems. The average measured size of *Hae*III locus D1S7 and D5S110 bands has varied slightly over time; while relatively small, these temporal changes add to the overall interlaboratory measurement uncertainty. The characteristic standard deviation for *Hinf* I RFLP system measurements has a nearly identical dependence on expected band size as does the standard deviation for *Hae*III measurements. The ellipsoidal distance, K, is suggested for use as an RFLP data quality metric; the critical threshold value that on average excludes 1% of plausibly valid proficiency data for a given polymorphic locus is $K_{1\%} = 14.2$.

KEYWORDS: forensic science, DNA typing, data analysis, *Hae*III, *Hinf* I, D1S7, D1S339, D2S44, D2S92, D4S139, D4S163, D5S43, D5S110, D6S132, D7S21, D7S22, D7S467, D7Z2, D8S358, D10S28, D12S11, D14S13, D16S85, D16S309, D17S26, D17S79, K562 DNA, control, proficiency tests

Restriction fragment length polymorphism (RFLP) DNA profiling is currently being displaced in the U.S. forensic communities by polymerase chain reaction-based short tandem repeat (STR) methodologies. However, given the investment many jurisdictions have made in establishing RFLP databases and that complete STR profiling systems have only recently been defined (1) and are still being validated, RFLP and STR profiling may well be used in parallel for casework (when sample size is not limiting) until most archived offender samples are profiled by STR methods (2). Further, it is likely that some of the RFLP casework performed over the past decade has consumed the probative sample(s); new RFLP analyses will occasionally be required for any reopened cases. As RFLP becomes less routine and analysts lose familiarity with the art and techniques required, those analyses that are performed will need to be validated ever more rigorously.

The chronic myelogenous leukemia human cell line K562, established from a female donor in 1975 (3), is the *de facto* standard RFLP control material for most U.S. federal, state, and local forensic laboratories. Although having a modal karyotype of 55 normal chromosomes (while lacking chromosome #9) and 14 fragments, the K562 line CCL 243 deposited in the American Type Culture Collection (Manassas, VA) is apparently genomically stable (4). K562-derived materials are included as part of NIST's RFLP profiling standard, Standard Reference Material® (SRM®) 2390 (5). The K562 band sizes certified for SRM 2390 are used in a data validation stage preceding acceptance of RFLP data into the national level of the FBI's Combined DNA Indexing System (CODIS) (6).

RFLP band sizes for SRM 2390 were originally certified on the basis of interlaboratory comparison data collected in 1991 and will be recertified using interlaboratory data collected in 1997 (5,7). While small, there are differences in the two sets of certified values. Considerable evolutionary development of RFLP protocols, equipment, and reagents took place during the 1991 through 1997 time period. While these developments have been made in accord with the quality assurance guidelines established by the Technical Working Group for DNA Analysis Methods (TWGDAM) (8), protocol differences are known to have a small but reproducible influence on observed RFLP band sizes (9,10).

From 1991 through 1997, a total of 51 forensic DNA profiling proficiency tests have been conducted by three different commercial providers. In addition to quantitative sizing results for test materials, all three providers have requested, collected, and reported sizing results for control materials that participants include in their analyses. The K562 data collected in these studies provide a detailed history of RFLP sizing as performed in the U.S. forensic and paternity testing communities.

The following sections of this report describe the K562 RFLP sizing data available in these proficiency studies, statistically summarize these data for all reported *Hae*III and *Hinf* I bands, document changes with time at the most commonly reported genetic loci, and display the data in relation to various quality criteria. This documentation of K562 sizing performance may help assure continued reliable comparison of RFLP data over time and across analysts and laboratories.

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

* 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.

Data

Proficiency Tests

Members of the U.S. forensic DNA analysis communities regularly participate in one or more of three DNA profiling proficiency testing programs: the Forensic Identity Proficiency Testing Program (College of American Pathologists, 325 Waukegan Road, Northfield, IL 60093, USA), the International Quality Assessment Scheme, Forensic Analysis (Cellmark Diagnostics, PO Box 265, Abington, Oxon, OX14 1YX, UK), and the Forensic Testing Program, DNA Profiling (Collaborative Testing Services, Inc., PO Box 1049, 340 Herndon Parkway, Herndon, VA 20170, USA). Table 1 lists the 51 tests conducted by these providers from 1991 through 1997.

All RFLP band size data contained in the final reports of these proficiency studies, including manufacturer and reference laboratory data, were manually entered into a spreadsheet database at NIST. Where possible, statistical summary information for the database entries have been validated with those listed in the final reports. All data were visualized with the RFLP measurement quality demonstration tools elsewhere described (11,12). All observed inconsistencies were checked against the recorded data.

The proficiency test reports occasionally specify corrections to previous data (typically typographic or handwriting interpretation errors). All final reports were diligently examined for such information, and all noted corrections made. A number of "reporting blunders" having proficiency testing implication, but unrelated to the actual RFLP band sizing process, were rectified. These nonmeasurement errors include: reporting band sizes as kilo basepairs (kbp) rather than basepairs (bp), listing the smaller band before the larger, misstatement of the genetic locus, gross-error digit transpositions (e.g., "96" for "69"), and gross-error digit mistranscriptions (e.g., "36" for "56").

TABLE 1—*Proficiency tests.*

Year	Program*	Test
1991	CM	9102, 9103, 9104
	CTS	91-3, 91-15
1992	CM	9201, 9202, 9203, 9204
	CTS	92-3, 92-15
1993	CAP	1993 FID-A, 1993 FID-C
	CM	9301, 9302, 9303, 9304
	CTS	93C, 93Q
1994	CAP	1994 FID-A, 1994 FID-C
	CM	9401, 9402, 9403, 9404
	CTS	9403, 9415
1995	CAP	1995 FID-A, 1995 FID-C
	CM	9501, 9502, 9503, 9504
	CTS	9503, 9515
1996	CAP	1996 FID-A, 1996 FID-C
	CM	9601, 9602, 9603, 9604
	CTS	9603, 9615
1997	CAP	1997 FID-A, 1997 FID-B
	CM	9701, 9702, 9703, 9704
	CTS	9703, 9715

* CAP = College of American Pathologists: Forensic Identity Proficiency Testing Program.
CM = Cellmark Diagnostics: International Quality Assessment Scheme, Forensic Analysis.
CTS = Collaborative Testing Services, Inc.: Forensic Testing Program, DNA Profiling.

Restriction Endonucleases

Most of the proficiency test data were obtained using the *Hae*III endonuclease RFLP system, with less than 10% of the data obtained from the *Hinf*I system. A small set of test sample data were reported for the *Pvu*II system. However, since no cell line control data were reported for this system, our analyses encompass only *Hae*III- and *Hinf*I-based RFLP systems.

Genetic Loci

All polymorphic genetic loci that have been reported in any of the 51 proficiency tests for *Hae*III and *Hinf*I RFLP systems are included in our studies, along with the monomorphic locus D7Z2. Tables 2 and 3, respectively, list the loci and various summary statistics for both endonuclease systems. Figure 1 displays the number of participants who reported K562 data for each of the more commonly reported loci as a function of time.

Participants

Preservation of participant confidentiality is crucial to the success of any interlaboratory comparison exercise. All three forensic proficiency testing programs assign participants unique "code names." Only these participant codes are used in the reporting of qualitative and quantitative data. None of the programs reveal the identity of any participant, nor is their analytical background described. Summary data provided in the reports, patterns in the data themselves, and conversations with forensic professionals do suggest that: (1) essentially all of the participants who report K562 data are from the U.S., (2) the large majority of these participants are experienced forensic and/or paternity analysts, but a few are relatively inexperienced students participating as part of their training, and (3) the majority are affiliated with local, state, or Federal forensic laboratories with the minority from academic or private laboratories.

One of the proficiency programs has retained the same code for each laboratory for all tests conducted since 1994. This program distinguishes different analyst participants from the same laboratory with a laboratory-assigned letter sequence. The other two programs assign a unique code to each participating analyst in every test conducted. While close similarities among data supplied by some of the participants suggest their use of (at least) the exact same protocol, no attempt has been made to differentiate within-laboratory from among-laboratory measurements.

Invalid K562 Cell Line Control Data

Not all band sizes attributed to cell line control materials are valid measurements of K562 control material. While most non-K562 control materials are identified as such in the reports, some are not explicitly identified, and a few are misidentified. We therefore first entered all plausibly K562 cell line control data and, after data entry was complete, excluded sets of band sizes if bands for two-or-loci were not between 90% to 110% of known K562 values. While only occasionally encountered with *Hae*III data, a quite sizable minority of *Hinf*I control data are for non-K562 materials.

Occasionally participants reported a band size for only one of the two typically reported K562 bands, resulting in partial data. Only complete data have been used; all partial data have been excluded. This does not apply to the three loci that are typically reported as one-banded: locus D7Z2 is monomorphic (human-specific with a known sequence of size 2731 bp) (13), K562 is an apparent ho-

TABLE 2—*Summary statistics for* Hae*III K562 band sizing.*

Locus	N_{test}	$N_{invalid}$	N_{use}	$\overline{X}_1$ (bp)	S_1 (bp)	$\overline{X}_2$ (bp)	S_2 (bp)	R
D1S7	51	6	1718	4583	39	4234	36	0.83
D1S339	7	0	8	2870	32*	2781	21*	0.90
D2S44	51	10	2035	2912	19	1792	13	0.69
D2S92	8	0	11	9503	189*	4343	32*	0.66
D4S139	49	7	1828	6505	50	3447	22	0.54
D4S163	14	0	27	4467	41*	4032	29*	0.59
D5S110	39	7	1147	3720	24	2941	22	0.79
D6S132	8	1	9	3324	26*	1834	15*	0.69
D7S467	37	1	164	4689	43	3235	23	0.86
D7Z2†	21	0	96	2731	15	…	…	…
D8S358‡	5	2	3	≈6000	…	≈1300	…	…
D10S28	49	5	1683	1758	12	1185	11	0.63
D14S13§	46	1	170	1637	11	…	…	…
D16S85	10	0	16	1601	16*	<700‖	…	…
D17S26	41	3	116	4853	33	1366	9	0.44
D17S79	49	3	1087	1984	14	1522	13	0.82
	Total	46	10118					

All symbols are defined in text.
* Due to limited data, estimated as the maximum of S and $S(\overline{X})_{\beta_0 = 7.5}$.
† Human-specific monomorphic locus having a known sequence of length 2731 bp (13).
‡ Too few, too variable sizings for meaningful analysis.
§ K562 is an apparent homozygote at this locus.
‖ This band was quantitatively reported only once, with a value of 536 bp.

TABLE 3—*Summary statistics for* Hinf *I K562 band sizing.*

Locus	N_{test}	$N_{invalid}$	N_{use}	$\overline{X}_1$ (bp)	S_1 (bp)	$\overline{X}_2$ (bp)	S_1 (bp)	R
D1S7	49	36	189	4817	40	4458	34	0.85
D2S44	45	24	145	4025	26	2906	20	0.63
D4S139	10	5	15	5086	68*	2025	30*	0.85
D5S43	10	1	9	5550	53*	4782	42*	0.61
D5S110	10	5	15	4802	58*	4040	52*	0.93
D7S21	50	39	192	7896	95	7021	76	0.86
D7S22	47	28	92	7098	79	1991	18	0.57
D12S11	47	42	197	13778	372	5321	53	0.62
D14S13†	5	2	6	1861	17*	…	…	…
D16S309	19	2	21	2927	30*	2126	25*	0.94
	Total	184	881					

All symbols are defined in text.
* Due to limited data, estimated as the maximum of S and $S(\overline{X})_{\beta_0 = 8.3}$
† K562 is an apparent homozygote at this locus.

mozygote at D14S13, and one of the two K562 bands at locus D16S85 is too small to be quantitatively sized routinely.

A few participants (quite probably, judging from the patterns of less common loci employed, from just one laboratory) recorded all band size data to only two significant digits. While not invalid, such low-precision values are (i) not characteristic of forensic practice and (ii) distort estimates of measurement location and dispersion for smaller bands, particularly at the less commonly reported loci. Only data sets reported to three-or-more significant digits have been used.

In all 51 reported tests, four participants reported K562 and test sample data for a particular locus that were quite different from all the rest of the data in the proficiency report and that could not be ascribed to locus misidentification. We speculate that these data result from incorrect assignment of the sizing ladder bands or use of the wrong ladder. Only data from the improperly sized loci have been excluded from further analysis.

A very few data have been identified as "statistical outliers"; i.e., plausibly valid data that are implausible members of a consensus distribution of like measurements. The methods used to identify these outliers, and the data themselves, are presented in a later section. While a problem only for some of the less commonly employed loci, outliers by definition are not valid members of their nominal measurement populations. We have retained all outlier data for graphical displays and analysis, but have excluded them when estimating univariate and bivariate summary statistics. For polymorphic loci, both bands have been excluded, even when only one was an outlier.

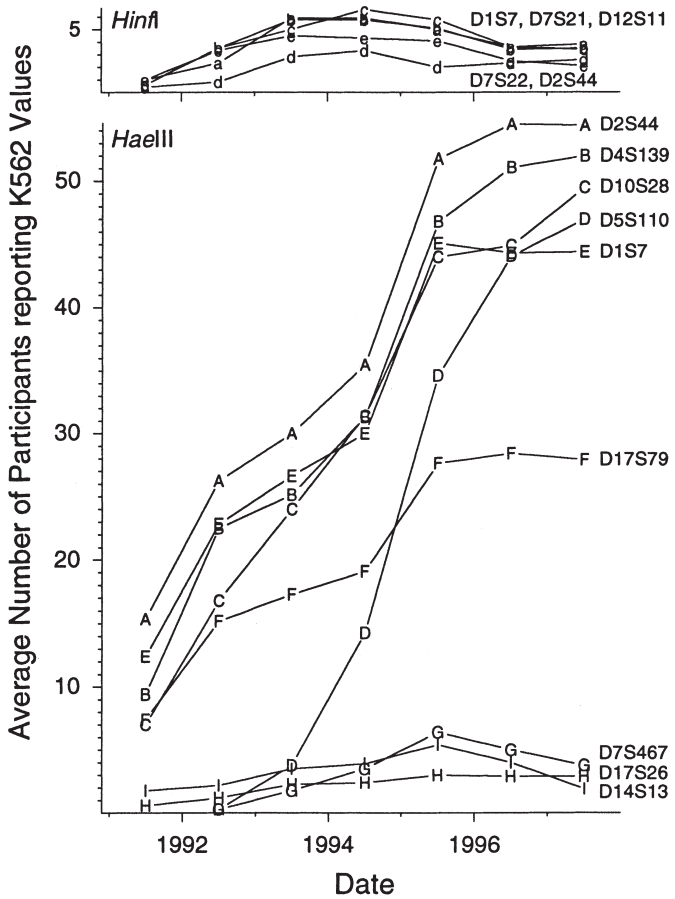

FIG. 1—*Number of participants reporting K562 RFLP band sizing data as a function of time. The lower graphical segment summarizes the most commonly reported* Hae*III-system loci; the upper segment summarizes the most commonly reported* Hinf *I-system loci. Values are the yearly averages for the proficiency tests listed in Table 1.*

Results and Discussion

Number of Informative Data

We have attempted to retain all plausibly valid K562 RFLP sizing data reported in any of the 51 proficiency tests. However, while validating the entered data, we noted several instances where two or more participants reported unusually similar data. We believe that some of these data may not be independent measurements but are "duplicates" of previously obtained values. Such duplicate data artificially increase the apparent correlation between different bands and decrease variance by inflating the number of nominally independent values ("degrees of freedom").

On the basis of an exact multi-locus match (all bands at two or more loci have exactly the same size) across all samples, it appears that two or more participants reported the same data in two of the 51 proficiency tests. In about 30% of the studies, two or more participants report exactly matching K562 values while the test sample data match considerably less well. These duplicate K562 values may result from two or more analysts using the same analytical gel. This duplication of K562 data has also been noted among participants in sequential studies from the same provider and among studies conducted in the same time period by different providers. We believe these duplicate K562 values reflect analysts reporting nominal rather than measured K562 band size. Whatever the true reason for these exact-match duplicate K562 measurements, only the first-encountered set is used in our analyses.

The numbers of reported values excluded from analysis for any reason (N_{invalid}) and the numbers used in the analyses (N_{use}) are listed in Table 2 for all *Hae*III bands and in Table 3 for *Hinf* I bands. However, we believe the true degrees of freedom for our analyses are modestly less than the N_{use} values reported in Tables 2 and 3. A small fraction of the *Hae*III data, but perhaps 20% of the *Hinf* I, are not exact multi-locus matches, but the band sizes among two or more participants in single tests are quite similar. We believe that many of these data represent multiple sizings of single sets of autoradio- or chemilumograms. While simple multivariate techniques could reliably identify such near-duplicates (14), isolation of within- from among-gel variation is beyond the scope of the current study.

Evolution of Use of Genetic Loci

Figure 1 depicts the average number of participants reporting K562 data per proficiency test for the period 1991 through 1997 for all commonly employed *Hae*III and *Hinf* I loci. The number of proficiency test participants using the *Hae*III RFLP system steadily increased from 1991 until 1996 and then stabilized. The number of participants using the *Hinf* I system increased slightly until 1994 and has since declined. This pattern is a composite result of a rapid increase in the percentage of *Hinf* I users who report K562 values and a steady decrease in the number of participants reporting *Hinf* I data. D5S110 is the only genetic locus that has become widely accepted since 1991, to some extent displacing use of locus D1S7.

Univariate Statistics

Tables 2 and 3 list the observed mean band size, $\overline{X}$, and standard deviation, S, for all *Hae*III and *Hinf* I RFLP bands. Figure 2 displays the functional dependence of S and its cousin the percent relative standard deviation, $\%\text{RSD} = 100 \times S/\overline{X}$, on $\overline{X}$ for the most commonly reported ($N_{\text{use}} \geq 92$) bands.

The functional relationship of S on $\overline{X}$ has been documented in both *Hae*III and *Hinf* I RFLP systems (15,16). We have described the short term among-laboratory standard deviation for laboratories adhering to TWGDAM quality control guidelines as

$$S(\overline{X}) = \beta_0 \left(1 + \frac{\overline{X}}{\beta_1}\right)^{\beta 2} \qquad (1)$$

with the constants estimated empirically as $\beta_0 = 7.5$, $\beta_1 = 19{,}500$, and $\beta_2 = 7.1$ from data provided by a designed interlaboratory study (17). With these parameters, Eq 1 describes both *Hae*III and *Hinf* I K562 data structure gratifyingly well.

The *Hinf* I S are on average 10% larger than predicted using the *Hae*III parameterization of Eq 1; increasing β_0 to 8.3 improves Eq 1's description of the current *Hinf* I data. This near equivalence of measurement performance between the two systems is compatible with the results from the European DNA profiling group's (EDNAP) interlaboratory comparison of *Hinf* I RFLP data obtained by laboratories using a standardized protocol (16).

Since the least commonly reported loci ($N_{\text{use}} < 30$) appear often reported by just one or a very few laboratories, the observed S may represent within-laboratory and not among-laboratory variability. The S listed in Tables 2 and 3 for these bands are the maximum of S and $S(\overline{X})$ calculated using the appropriate parameters.

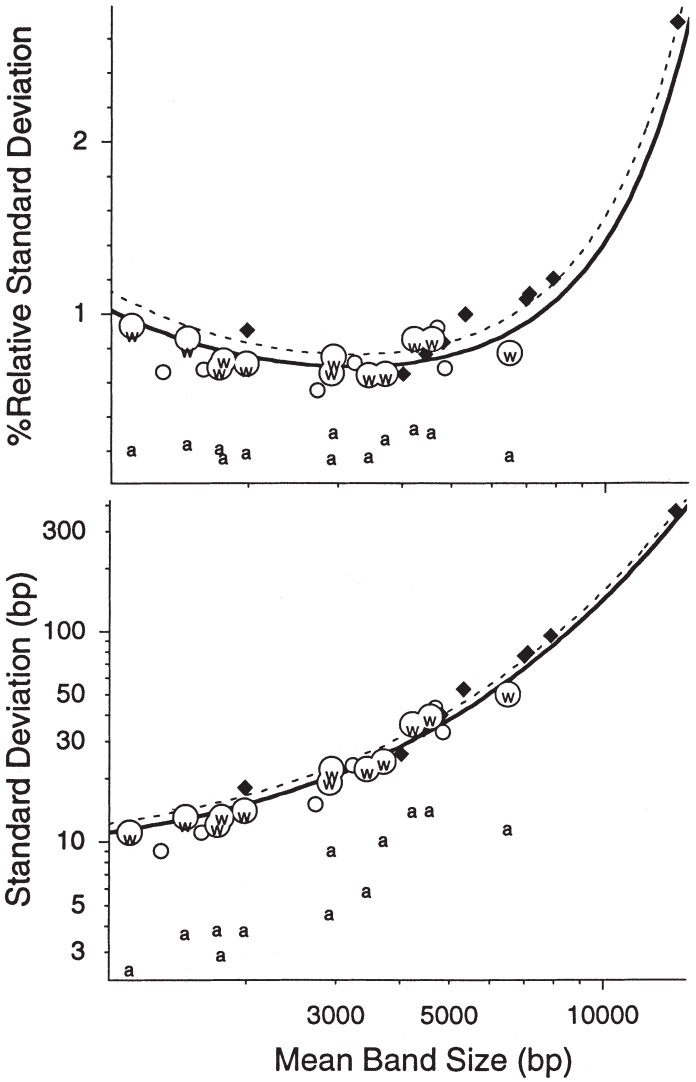

FIG. 2—*Total among-laboratory measurement uncertainty as a function of band size. The lower graphical segment presents relationships between* S *and band size; the upper segment presents the same relationships with* S *transformed to %RSD. The large open circles denote total uncertainties for the twelve bands of the commonly reported* Hae*III loci: D1S7, D2S44, D4S139, D5S110, D10S28, and D17S79). The small open circles denote total uncertainties for the six bands available for* Hae*III loci: D7S467, D7Z2, D17S26, and D14S13. The solid diamonds denote total uncertainties for the ten bands of the commonly reported* Hinf*I loci: D1S7, D2S44, D7S21, D7S22, and D12S11. The solid line denotes the expected relationship (Eq 1) between the total uncertainty and band size as parameterized for* Hae*III-system data; the dotted line denotes the same relationship adjusted to the better represent the current* Hinf*I results. The letters "a" and "w" denote among-test (Eq 2) and within-test (Eq 3) components of variance, respectively, for the most commonly reported* Hae*III bands.*

Among- and Within-Test Components of Variance

There are sufficient data to estimate the within-test and the among-test variance components (standard deviation ≡ $\sqrt{\text{variance}}$) for all the most commonly reported *Hae*III bands (N_{use} = 1147). The standard deviation of the mean band size of each individual proficiency test estimates the among-test component

$$S_{among} = \sqrt{\frac{\sum_{i=1}^{N_{test}} \left(\overline{X}_i - \sum_{i=1}^{N_{test}} \overline{X}_i / N_{test}\right)^2}{N_{test} - 1}} \qquad (2)$$

where N_{test} is the number of proficiency tests reporting relevant data and $\overline{X}_i$ is the mean band size in the *i*th such test. Pooling the individual test standard deviations estimates the within-test component

$$S_{within} = \sqrt{\frac{\sum_{i=1}^{N_{test}} (N_i - 1) S_i^2}{\sum_{i=1}^{N_{test}} N_i - N_{test}}} \qquad (3)$$

where N_i is the number of relevant band size data in the *i*th test and S_i is the standard deviation of these data. Figure 2 displays these variance components as a function of band size. The S_{within} are much larger than S_{among} for all twelve bands and are only slightly smaller than *S*. The S_{among} are relatively large for the three bands least well described by the *Hae*III parameterization of Eq 1.

Changes in Measured Band Size with Time

Figure 3 displays changes in mean band size as a function of time for all 12 commonly reported bands. The displayed data have been standardized to have the same zero mean and unit standard deviation

$$Z_i = \frac{(\overline{X}_i - \overline{X})}{S} \qquad (4)$$

Figure 3 also displays approximate one standard deviation uncertainty bars for the standardized differences

$$S_{z_i} = \frac{S_i/\sqrt{N_i}}{S} \approx \sqrt{\frac{1}{N_i}} \qquad (5)$$

While the mean values for loci D2S44, D4S139, D10S28, and D17S79 are nearly stationary over time, the mean size measurements of both D1S7 bands have steadily increased by about 1.5 *S* (or about 60 bp for the larger D1S7 band and 55 bp for the smaller). The D1S7 bands are by far the least widely separated bands of the commonly used loci. We speculate that the increase in the average measured sizes of the D1S7 bands may be related to evolutionary changes in the electrophoretic and optical resolution of RFLP bands. Alternatively, since one of the two commonly used sizing ladders has an anomalously migrating band that affects the sizing of both D1S7 bands (9), the unique increase in the measured D1S7 band sizes may reflect the decreasing frequency of use of this ladder.

The mean values for both of the locus D5S110 bands have also changed with time, increasing by nearly 2 *S* (or about 50 bp and 45 bp) from 1993 through 1995 but then slowly decreasing by about 1 *S* through 1997. Since these changes parallel the growth in use of this locus (Fig. 1), we believe that the observed D5S110 size measurement changes mostly reflect the changing number of participants and the type of laboratories employing this locus.

Bivariate Correlation

In addition to the univariate $\overline{X}$ and *S* statistics, Tables 2 and 3 list the observed between measurement correlation of each polymorphic loci

$$R = \frac{\sum_{i=1}^{N_{use}} \left(\frac{x_{i1} - \overline{X}_1}{S_1}\right)\left(\frac{x_{i2} - \overline{X}_2}{S_2}\right)}{N_{use} - 1} \qquad (6)$$

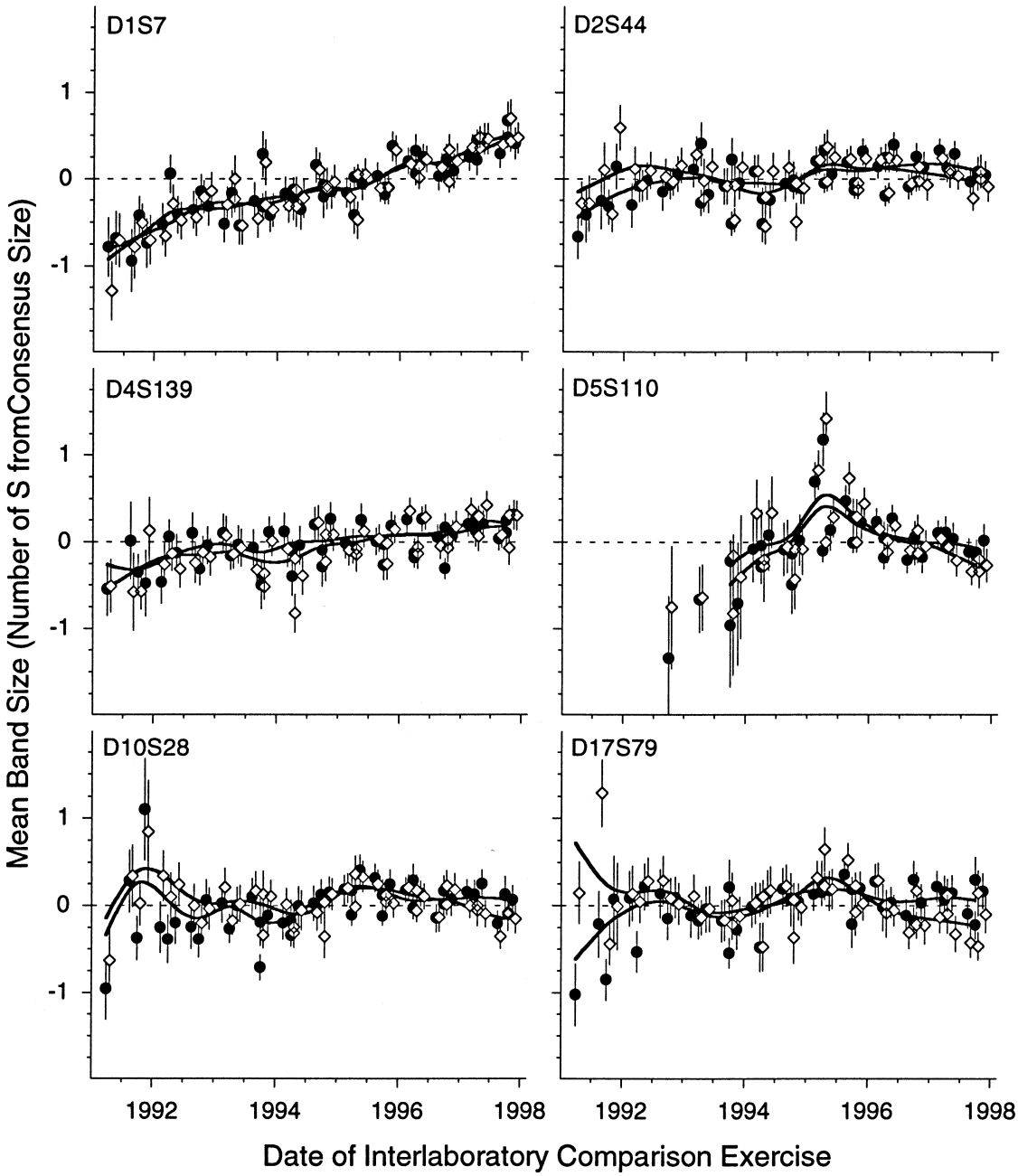

FIG. 3—*Change in apparent band size as a function of time. Each graphical segment presents the per-proficiency test mean band sizes for the high-band (solid circles) and low-band (open diamonds) of a given locus. The bars denote the one standard deviation uncertainty for each individual mean (Eq 5). All values have been standardized to have the same zero grand mean and unit standard deviation (Eq 4). The solid lines are smoothed yearly boxcar averages of the observed means. The dashed lines denote the expected zero grand mean.*

where subscript "1" indicates the larger band at each locus and subscript "2" the smaller and "x_i" is the ith measurement. Since correlation is extremely sensitive to the presence of some types of outlier data (18), such data were subjectively identified using single-locus charts (SLCs) of $\{x_1,x_2\}$ measurement pairs (11). The R values used in the SLCs were estimated by minimizing the area of the ellipsoid with center $\{\overline{X}_1,\overline{X}_2\}$ and scale $\{S_1,S_2\}$ that encloses 80% of the measurements (19). Figures 4 and 5 display the $\{x_1,x_2\}$ pairs for all but the least commonly employed polymorphic loci, along with the 80% minimum area ellipses and two data quality assessment criteria described below.

Multivariate Correlation

As displayed in Figs. 4 and 5, the $\{x_1,x_2\}$ measurement pairs for any given locus can be described as a single elliptical distribution without apparent substructure. Since all measurements for a given sample are typically traceable to the same lane(s) of a single (geometrically stabilized) gel, a similar "multivariate normal" structure could characterize all the band sizes of all loci for a given sample. There are 518 complete six-loci (D1S7, D2S44, D4S139, D5S110, D10S28, and D17S79) *Hae*III records in the proficiency data. Figure 6 displays the 36 $\{\text{band}_1, \text{band}_2\}$ pairs for these data. While

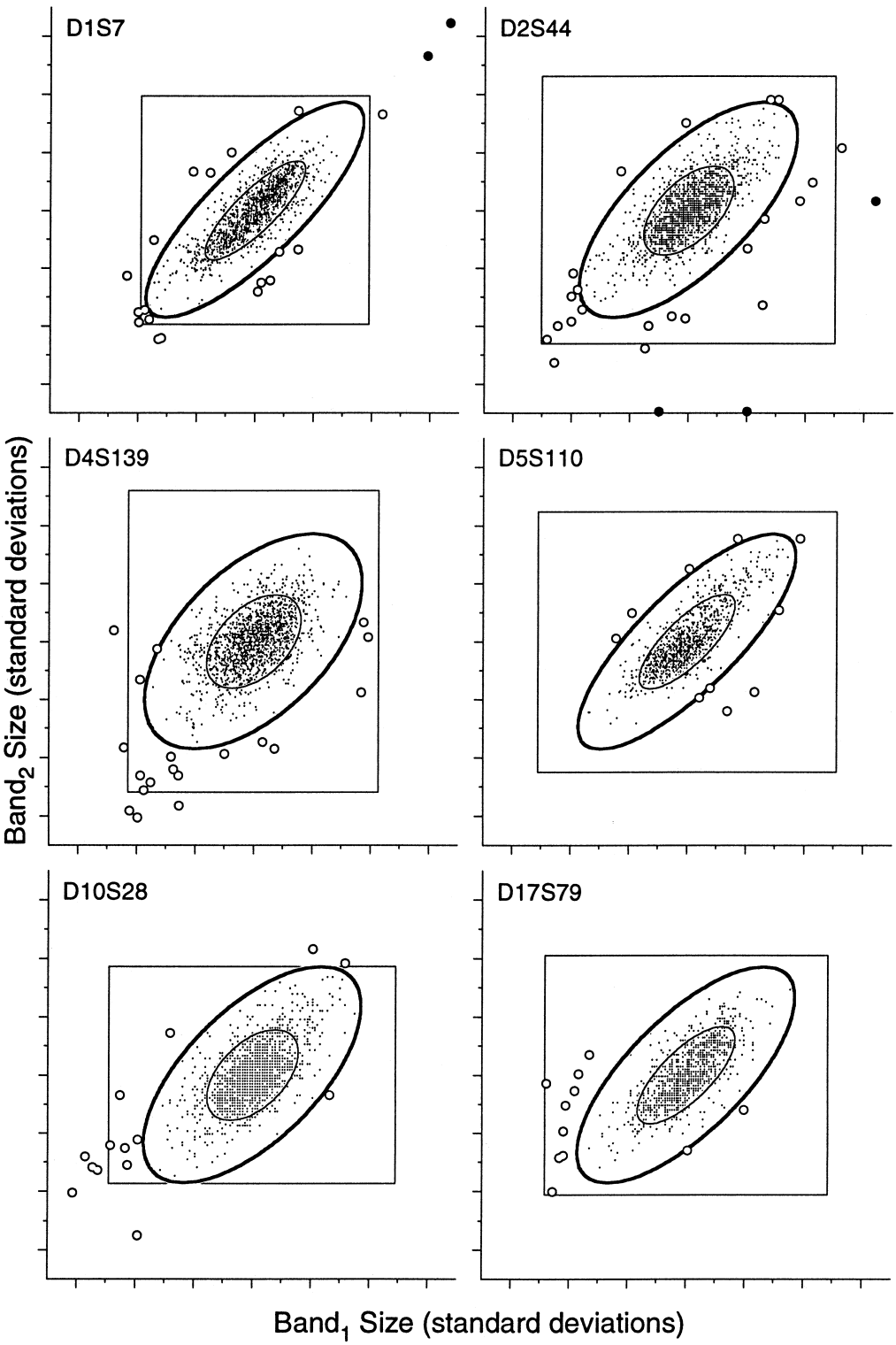

FIG. 4*a—Single locus K562 charts for commonly reported* Hae*III loci. Each graphical segment presents all K562 proficiency test measurement pairs for a given locus, with each chart axis spanning ±7 standard deviations about the mean value. The rectangles represent the %Δ 1% data quality criterion (Eq 7). The outer ellipses represent the K1% criterion (Eq 8); the inner ellipses are the minimum area ellipses that cover 80% of the data. "Dots" denote measurement pairs that are accepted by the K1% criterion. Open circles denote pairs that are included in the summary calculations but that are rejected by the K1% data quality criterion. Solid circles denote "outlier" measurements excluded from the summary calculations.*

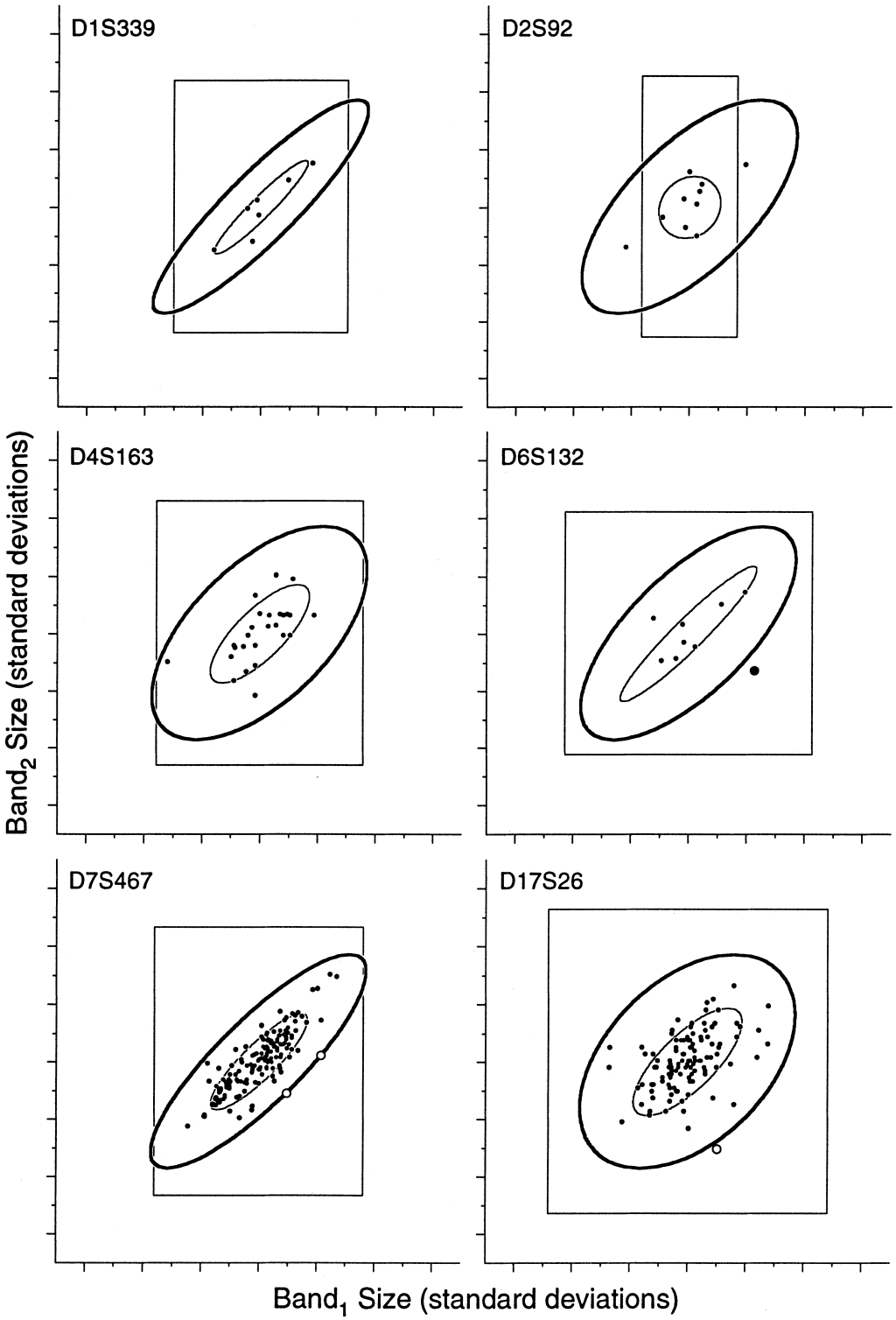

FIG. 4*b*—*Single locus K562 charts for less commonly reported two-banded* Hae*III loci. Legend as in Fig. 4a.*

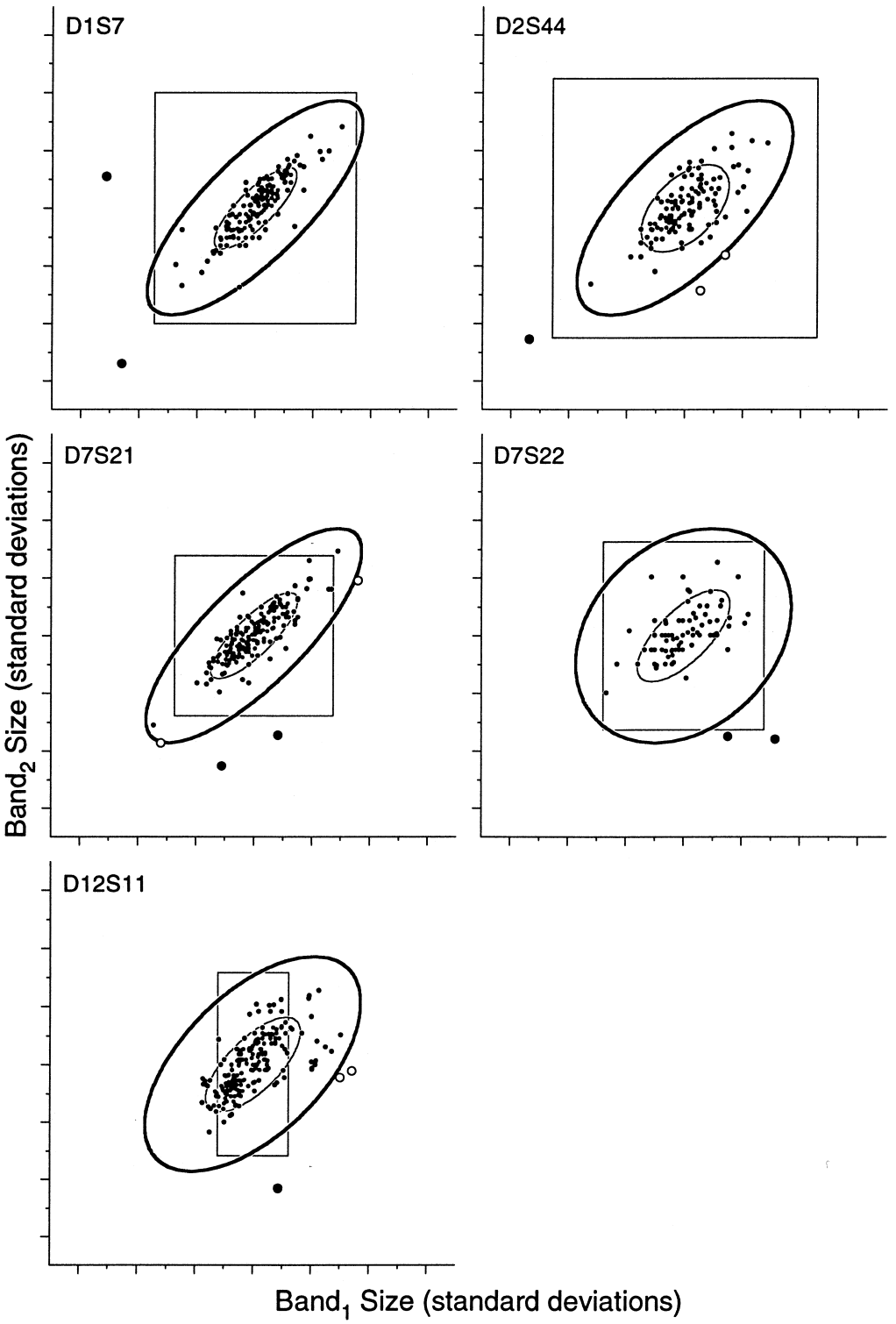

FIG. 5*a—Single locus K562 charts for commonly reported* Hinf*I loci. Legend as in Fig. 4a.*

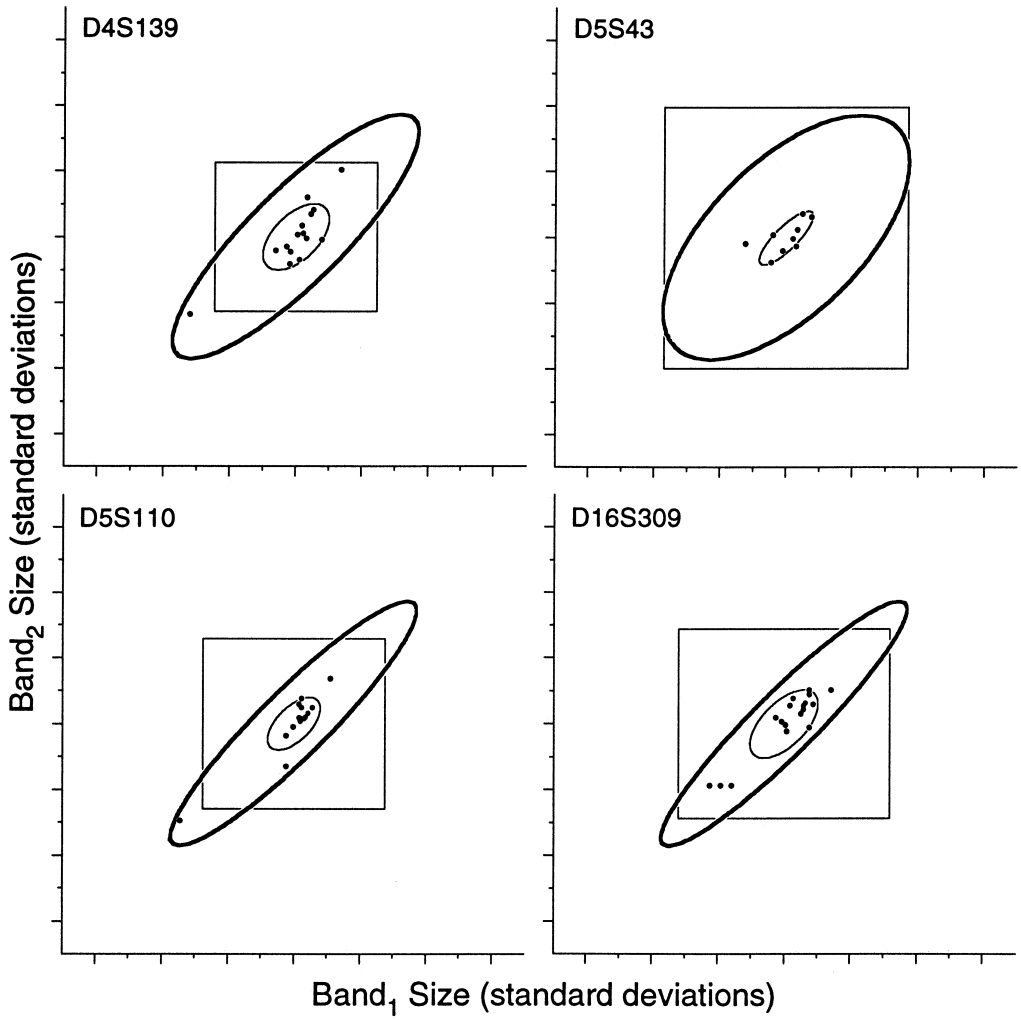

FIG. 5*b*—*Single locus K562 charts for less commonly reported* Hinf*I loci. Legend as in Fig. 4a.*

many of the among-locus pairings display the expected symmetric structure, several do not. Especially for pairs including D1S7 and D5S110, the distribution is asymmetric and appears to be divided into two or more subgroups.

Some or all of these "non bivariate normal" structures may be related to the changes in apparent band size over time for these two loci described above. At any rate, RFLP band sizes for two or more loci do not necessarily follow a multivariate normal distribution.

Data Quality Metrics

For effective assessment and control of any measurement system, ways of quantitatively summarizing data quality must be established. Percent relative difference is a common univariate metric; when applied to both bands of one locus, the quantity can be calculated as

$$\%\Delta = \text{maximum}\left(\frac{100\,|\,x_1 - \overline{X}_1\,|}{\overline{X}_1}, \frac{100\,|\,x_2 - \overline{X}_2\,|}{\overline{X}_2}\right) \qquad (7)$$

While the only explicit requirement for this metric is knowledge of the expected mean values, there is an implicit assumption that %RSD is constant for all bands. While %RSD is fairly—but not strictly—uniform for the 12 common *Hae*III K562 bands, %RSD rapidly increases as band size increases above about 8000 bp (Fig. 2). Moreover, this metric gives identical weight to both correlated "band shifts" (both the high and the low bands are larger or smaller than expected) and anticorrelated (one band is shifted high and the other shifted low).

The ellipsoidal shape formed when replicate single locus $\{x_1,x_2\}$ RFLP measurements are plotted (Figs. 4 and 5) has been frequently noted (20,21). The ellipsoidal distance, while computationally more complex

$$K = \frac{\left(\frac{x_1 - \overline{X}_1}{S_1}\right)^2 + \left(\frac{x_2 - \overline{X}_2}{S_2}\right)^2 - 2R\left(\frac{x_1 - \overline{X}_1}{S_1}\right)\left(\frac{x_2 - \overline{X}_2}{S_2}\right)}{(1 - R^2)} \qquad (8)$$

should be a more efficient RFLP data quality metric. The explicit incorporation of the between band correlation, R, makes this metric relatively forgiving of correlated band shifts and punishes anticorrelated shifts.

Data Validation

For both %Δ and K, the larger the magnitude of the metric the "less valid" the $\{x_1,x_2\}$ measurement pair. Traditional control charts can be used with either metric to assess, control, and docu-

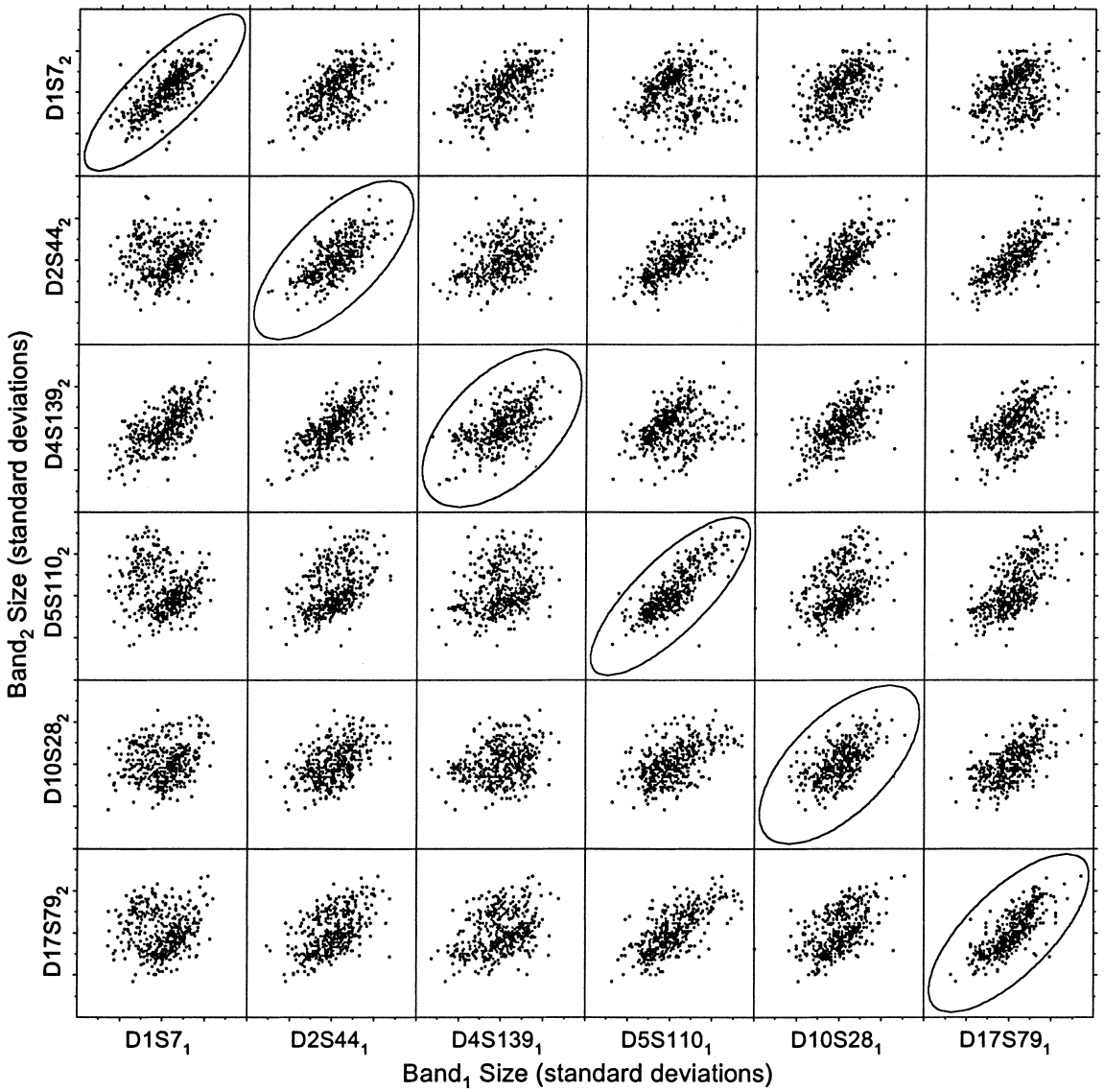

FIG. 6—*Multi-locus K562 scattergrams for the most commonly reported* Hae*III loci. Each segment displays one of the 36 possible combinations of {band1,band2} measurements among loci D1S7, D2S44, D4S139, D5S110, D10S28, and D17S79; the scale for each of the scattergrams is ±4 standard deviations about the mean value. Each segment displays all 518 measurement pairs available in the subset of data provided by participants reporting valid K562 values for both bands of all six loci. The diagonal segments display the same K1% criterion data quality ellipses presented in Fig. 4a.*

ment data quality (22). The metrics can also be used to validate data if decision threshold values, $\%\Delta_{crit}$ and K_{crit}, can be established. For a given quality metric, measurements are "insufficiently valid" when the magnitude of that metric becomes larger than the critical value.

The concept "insufficiently valid" is context-specific, subjective, and emotionally charged; no single critical value will be appropriate for all applications. The diversity of the RFLP proficiency test participants and the large number of measurements for the six most commonly employed *Hae*III loci suggest that critical values derived from these data may have general utility. While the "validity" characteristics of the proficiency data are unknown, about 0.5% of all the reported *Hae*III cell line control data were identified as qualitatively invalid. It is reasonable to assume that at least as many of the remaining data reflect quantitatively unacceptable measurement practice.

The upper section of Fig. 7 displays the minimum, average, and maximum $\%\Delta_{crit}$ that exclude a given percentage of $\{x_1,x_2\}$ pairs. The average values for three arbitrary decision thresholds are graphically emphasized: $\%\Delta_{5\%} = 2.26$, $\%\Delta_{1\%} = 3.27$, and $\%\Delta_{0.5\%} = 3.81$. These values are compatible with a ±3.0% intergel window recommended for bands of size between 500 bp and 10,000 bp and the ±1.6% to ±4.6% interlaboratory ranges observed for cell line GM9947 bands (23,24).

The lower section of Fig. 7 likewise displays the minimum, average, and maximum K_{crit} that exclude given percentages of the data. The observed average values for the same arbitrary thresholds are: $K_{5\%} = 6.80$, $K_{1\%} = 14.2$, and $K_{0.5\%} = 20.5$. To the extent that RFLP measurement pairs follow a bivariate normal distribution, K_{crit} has an expected inverse χ^2 distribution with two degrees of freedom (25). The lower segment of Fig. 7 also displays the expected relationship. Observation and the theory agree well for the

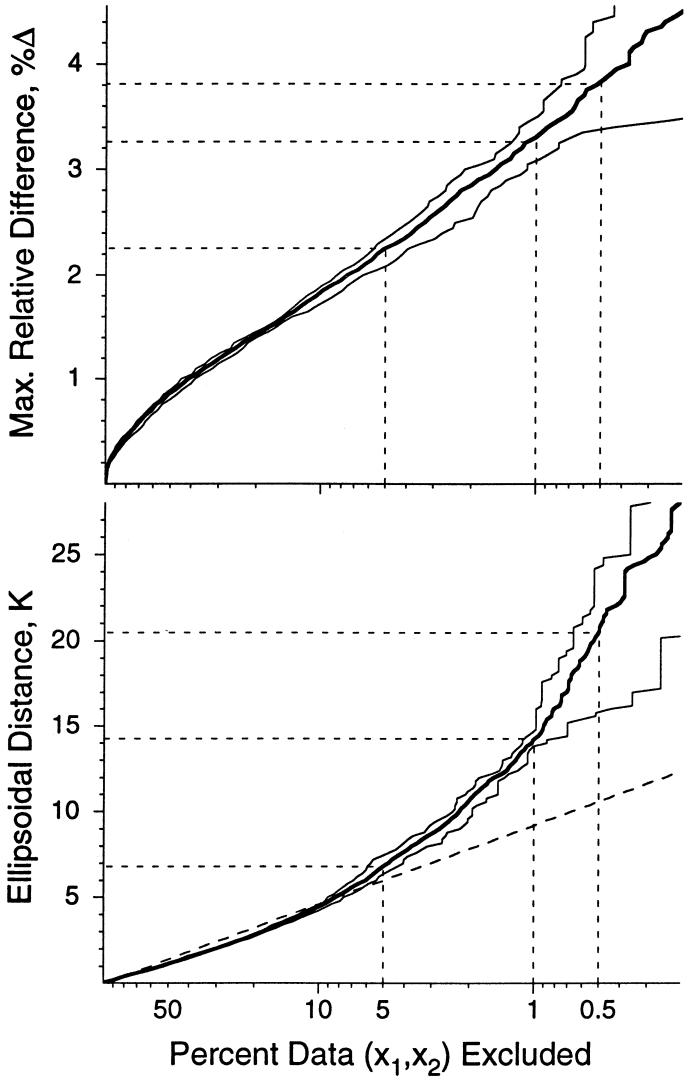

FIG. 7—*Data quality criteria as a function of percent data excluded. The upper graphical segment displays the %Δ (Eq 7) data quality metric for the six commonly reported* Hae*III loci; the lower segment displays the K (Eq 8) metric. The light solid lines bound the minimum and maximum values over the six loci for a given percentage of excluded data. The dark sold line denotes the average value over the six loci. The dotted lines connect three specific data exclusion thresholds (5%, 1%, and 0.5% of measurement pairs) with the corresponding critical values of each metric. The dashed line in the lower segment denotes the relationship expected for "truly" bivariate normal data.*

"most valid" 90 to 95% of the measurement pairs but the "least valid" are further from the ellipsoidal center ("thicker tails") than expected for a bivariate normal distribution.

Figures 4 and 5 display the $\%\Delta_{1\%}$ and $K_{1\%}$ thresholds for all loci, using the parameters listed in Tables 2 and 3. The differences between the two metrics are most apparent for loci with relatively large bands, such as *Hae*III locus D2S92 in Fig. 4*b* and *Hinf* I locus D12S11 in Fig. 5*a*. The intervals between the minimum and maximum critical value lines in Fig. 7 are much narrower for *K* than for %Δ in the most interesting 5 to 1% exclusion range. This superior consistency across the six loci suggests that *K* is the better estimate of data quality.

Conclusions and Suggestions

The K562 cell line control data provided in the numerous forensic DNA profiling proficiency tests from 1991 through 1997 enable a robust description of RFLP measurement performance. Approximately 95% of all data at polymorphic loci are well described with multivariate normal distributions. The univariate mean, univariate standard deviation, and bivariate correlations of these data should provide reliable reference distributions for establishing the quality of future RFLP measurements.

There has been some change in the measured size of several RFLP bands over time, most notably for *Hae*III locus D1S7. While relatively small, these temporal changes do add to the overall interlaboratory measurement uncertainty at some loci. The characteristic standard deviation for *Hinf* I RFLP system measurements has a nearly identical dependence on expected band size as do *Hae*III measurements.

Both percent relative difference, %Δ, and ellipsoidal distance, *K*, can be used as RFLP data quality metrics; *K* provides a more uniform representation of RFLP data structure over all relevant band sizes. Critical threshold values that on average exclude 1% of the plausibly valid proficiency data for a given polymorphic locus are: $\%\Delta_{1\%} = 3.27$ and $K_{1\%} = 14.2$.

Multivariate analysis techniques can identify unusual multiparticipant and multitest patterns in proficiency data, not necessarily just RFLP proficiency data. While these patterns may reflect routine and accepted practice, they suggest that not all laboratories require completely independent measurements of all analysts working the same proficiency test. We recommend that control *values* be reported only when they are true control *measurements*, made at the time and under the same conditions used for test samples.

Acknowledgments

We thank Cellmark Diagnostics, Collaborative Testing Services, Inc., and the College of American Pathologists for making complete sets of their forensic DNA proficiency testing program reports available to us.

We dedicate this paper to our late colleague and co-author, Kenneth T. Gary. While a high school senior, Ken transcribed, validated, and critically evaluated most of the data used in this study. We miss his help, his insights, and his energy. *Requiescat in pace*, 1979–1999.

References

1. Kerr D. Letter to crime laboratory directors receiving grant funds under the State Identification System grant program. FBI-Laboratory Division. 12 July 1998.
2. Coffman, D. Setting up a DNA database. Proceedings of IBC's International Conference on DNA Forensic Analysis, New Trends and Technologies; 30–31 July 1998; Annapolis, MD. Southborough, MA: IBC USA Conferences.
3. Lozzio CB, Lozzio BB. Human chronic myelogenous leukemia cell line with positive Philadelphia chromosome. Blood 1975;45:321–34.
4. Chen TR. Modal karyotype of human leukemia cell line, K562 (ATCC CCL 243). Cancer Genet Cytogenet 1985;17:55–60.
5. 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.
6. 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.
7. Duewer DL, Richie KL, Reeder DJ. RFLP band size standards. NIST Standard Reference Material® 2390. J Forensic Sci 2000; 45(5):1093–1105.
8. Technical Working Group for DNA Analysis Methods. Guidelines for a

quality assurance (QA) program for DNA analysis. Crime Lab Digest 1991;18:44–75.
9. 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.
10. Duewer DL, Lalonde SA, Aubin R, Fourney R, Reeder DJ. Interlaboratory comparison of autoradiographic DNA profiling measurements: precision and concordance. J Forensic Sci 1998;43:465–71.
11. Duewer DL, Liu H-K, Reeder DJ. Graphical tools for RFLP DNA profiling. Single-locus charts. J Forensic Sci 1999;44(5):969–77.
12. Gary KT, Duewer DL, Reeder DJ. Graphical tools for RFLP DNA profiling. Laboratory performance charts. J Forensic Sci 1999;44(5):978–82.
13. Waye JS, England SB, Willard HF. Genomic organization of alpha satellite DNA on human chromosome 7: evidence for two distinct alphoid domains on a single chromosome. Mol Cell Biol 1987;7:349–56.
14. Beebe KR, Pell RJ, Seasholtz MB. Chemometrics: A Practical Guide. New York: Wiley-Interscience 1998.
15. Eriksen B, Bertelsen A, Svensmark O. Statistical analysis of the measurement errors in the determination of fragment length in DNA-RFLP analysis. Forensic Sci Int 1992;52:181–91.
16. 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.
17. Stolorow AM, Duewer DL, Reeder DJ, Buel E, Herrin G. Interlaboratory comparison of autoradiographic DNA profiling measurements. 3. Repeatability and reproducibility of RFLP band sizing, particularly bands of molecular size >10k base pairs. Anal Chem 1996;68:1941–7.
18. Bebbington AC. A method of bivariate trimming for robust estimation of the correlation coefficient. Appl Stat 1978;27:221–6.
19. Titterington DM. Estimation of correlation coefficients by ellipsoidal trimming. Appl Stat 1978;27:227–34.
20. Eriksen B, Svensmark O. DNA-profiling of stains in criminal cases: analysis of measurement errors and band-shift. Discussion of match criteria. Forensic Sci Int 1993;61:21–34.
21. 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.
22. Juran JM, Gryna FM, editors. Juran's Quality Control Handbook, 4th ed. New York: McGraw-Hill, 1988.
23. Laber TL, Iverson JT, Liberty JA, Giese SA. The evaluation and implementation of match criteria for forensic analysis of DNA. J Forensic Sci 1995;40:1058–64.
24. 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.
25. Tracy ND, Young JC, Mason RL. A bivariate control chart for paired measurements. J Quality Tech 1995;27:370–6.

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