David L. Duewer,[1] *Ph.D. and Elizabeth A. Benzinger,*[2] *Ph.D.*

Products of Partial Digestion with *Hae* III. Part 2. Quantification*

REFERENCE: Duewer DL, Benzinger EA. Products of partial digestion with *Hae* III. Part 2. Quantification. J Forensic Sci 1997;42(5):864–872.

ABSTRACT: The base pair size of the excess DNA in the smallest three partial digestion bands for the variable number of tandem repeat loci D1S7, D2S44, D4S139, D5S110, D10S28, and D17S26 has been quantitatively evaluated using data obtained from intentional partial digestion of liquid blood DNA. Restriction fragment length polymorphism (RFLP) measurement characteristics specific to the performing laboratory were evaluated from that laboratory's historical K562 cell line control data. The expected size of the excess DNA is estimated as the weighted mean of the differences between the measured size of the partially digested bands and the fully digested band, with the weights predicted using knowledge of RFLP measurement characteristics. Confidence limits are developed for evaluating whether the size differences among a set of RFLP band multiplets observed in pristine samples are consistent with those expected from partial digestion. The base pair size of excess DNA for partials observed in evidentiary samples appears to be somewhat less than that from pristine samples.

KEYWORDS: forensic science, band shift, band sizing, DNA typing, gel electrophoresis, restriction fragment length polymorphism, variable number of tandem repeat

Occasionally some of the *Hae* III sites near DNA variable number of tandem repeat (VNTR) loci are not cut during *Hae* III digestion. A number of electrophoresis bands attributable to partial digestion ("partials") that contain the VNTR region plus additional DNA from the 5′ and/or 3′ adjoining sequences may result. Although such partials do not normally interfere with the forensic interpretation of restriction fragment length polymorphism (RFLP) analysis of single-donor samples, they may confound the analysis of multiple-donor samples. As discussed in Part 1 (1), we have observed strong regularities in the electrophoretic size of the partials for a number of commonly used VNTR loci.

It is possible to estimate both the apparent number of extra base pairs (bp) in a given partial and the expected uncertainty in the estimate, given: (1) the measured size in bp of the limit digest VNTR DNA fragment ("true band" or "T"), (2) the measured size in bp of the partial digestion band ("P"), and (3) the quantitative RFLP measurement characteristics of the laboratory performing the RFLP analysis. The extra bp size of a given partial P_i relative to T is simply the difference in the measured band sizes:

$$D_i = P_i - T \quad (1)$$

To the extent that the DNA sequences between the VNTR and the neighboring 5′ and/or 3′ *Hae* III sites are conserved among individuals, all "true" D_i for a given partial should be constant.

Because T and P_i are measurements and thus are known with limited certainty, Eq 1 actually yields *estimates* of the bp difference, $\hat{D}_i$.[3] An estimate of one standard deviation (SD) uncertainty for these $\hat{D}_i$ is given by (2):

$$SD_{\hat{D}_i} = \sqrt{SD_{P_i}^2 + SD_T^2 - 2(SD_{P_i})(SD_T)R_{TP_i}} \quad (2)$$

where SD_T is the repeatability SD (3) for measuring a band of size T, SD_{P_i} is the repeatability SD for measuring a band of size P_i, and R_{TP_i} is the expected correlation between the measurements of T and P_i. Thus the uncertainty associated with $\hat{D}_i$ is a function of the uncertainties in the measurement of the (widely varying) T and P_i band sizes and not of the (potentially constant) true D_i. Because many forensically useful *Hae* III VNTR loci yield T bands ranging in size from <1000 to >20,000 bp, the influence of measurement uncertainty on the interpretation of band size differences must be carefully examined. In particular, appropriate treatment of measurement uncertainty is required for evaluating the constancy of any particular set of $\hat{D}_i$ and the geometry of partial digestion *Hae* III sites.

We present here a quantitative interpretation of the observations on partial digestion bands in pristine samples reported in Part 1 of this series. We determine the expected additional size of the smallest three partial bands in such samples for VNTR loci: D1S7, D2S44, D4S139, D5S110, D10S28, and D17S26. We express the expected measurement SD for these additional sizes as functions of the size of the fully digested VNTR band. We demonstrate that the smallest partial digestion bands can be attributed to incomplete digestion at both the 3′ and 5′ ends of the VNTR for the majority of the loci studied. We compare results derived from pristine samples with those observed in available evidentiary samples.

Methods and Materials

Partial Digestion Bands

The partial digestion data used in this study were generated from a set of 122 pristine samples at the Illinois State Police (ISP) Forensic Sciences Command Research and Development Laboratory. Data were collected for alleles at genetic loci D1S7, D2S44, D4S139, D5S110, D10S28, and D17S26. A complete

[1]Analytical Chemistry Division, Chemical Science and Technology Laboratory, National Institute of Standards and Technology, Gaithersburg, MD.
[2]Illinois State Police, 2060 Hill Meadows, Springfield, IL. Present address: Ohio Bureau of Criminal Identification and Investigation, Box 365, London, OH 43140.
*This study was supported in part by a grant from the Midwestern Association of Forensic Scientists.
Received 9 Aug. 1996; and in revised form 27 Dec. 1996; accepted 24 Jan. 1997.

[3]In standard statistical notation, all estimated values are "hatted" (ˆ). As all the terms used in the following discussion are estimates of one form or another, we simplify our notation and "hat" only those parameters for which the "true value" and estimates of the value are discussed.

description of the experimental methods and materials used is provided in Part 1 of this series (1).

When possible, the completely digested T band and the smallest three partial digestion P_i bands were quantitatively sized. Many samples expressed fewer than three partial bands for each of the two T bands at some loci; some samples at some loci expressed such a multitude of partial bands that fewer than three could be quantitatively sized or unambiguously assigned to a given T. All P_i and T band associations were assigned by one of the authors (EAB). A summary description of the available data is presented in Table 1 of Part 1. This complete data set is available from the corresponding author

Partial digestion band data characteristic of casework were obtained from 27 evidentiary samples analyzed by the ISP (mostly for D2S44, some at D10S28, and few at the other loci) and from 37 evidentiary samples analyzed at the Federal Bureau of Investigation (all for D2S44).

K562 Bands—The RFLP measurement characteristics were assessed from K562 cell line control bands collected from casework, offender, and population studies performed by the ISP from 1992 to 1995.

Data Analysis—The analysis of these data was performed at the National Institute of Standards and Technology (NIST) as part of an ongoing effort to characterize DNA measurement methods.

Results and Discussion

Quantitative Estimation of Individual Band Size Differences

Using data provided by member laboratories of the Technical Working Group for DNA analysis methods (TWGDAM), we have previously shown that the SD for RFLP DNA fragment bands of size 1000 to 20,000 bp can be estimated as (4–6):

$$SD = \beta_1\left(1 + \frac{MBS}{\beta_2}\right)^{\beta_3} \quad (3)$$

TABLE 1—*Long-term repeatability of the Illinois State Police K562 RFLP measurements.*

		Number of Data						
Locus	Allele	Case*	Pop†	Off‡	Total	$\overline{MBS}$	$\overline{SD}$	β_i'§
D1S7	High	44	73		117	4602	17	3.9
D1S7	Low	44	73		117	4250	19	4.8
D2S44	High	78	72	860	1010	2914	13	4.8
D2S44	Low	78	72	860	1010	1789	8	4.2
D4S139	High	78	73	873	1024	6515	41	5.3
D4S139	Low	78	73	873	1024	3454	14	4.3
D5S110	High	40			40	3721	13	3.9
D5S110	Low	40			40	2933	9	3.5
D10S28	High	76	75	839	990	1759	10	5.5
D10S28	Low	76	75	839	990	1180	8	5.6
D17S26	High	30	73		103	4839	22	4.6
D17S26	Low	30	73		103	1362	6	3.7
								4.5

*Casework.
†Population.
‡Offender.
§$\beta_i' = \dfrac{\overline{SD}}{(1 + \overline{MBS}/19500)^{7.1}}$

where β_1, β_2, and β_3 are empirically determined parameters. The mean band size (MBS) and SD are determined for a given DNA fragment band from a given set of data as:

$$MBS = \sum_i^N bp_i \Big/ N$$

$$SD = \sqrt{\sum_i^N (bp_i - MBS)^2 \Big/ (N - 1)}$$

where bp_i are independent measurements of the DNA fragment and N is the number of such measurements.

Using data for 28 bands from a designed set of mixed donor samples analyzed in 20 laboratories, the interlaboratory measurement reproducibility (3) SD is estimated as (6):

$$SD_{Inter} = 7.5\left(1 + \frac{MBS}{19500}\right)^{7.1} \quad (4)$$

Equation 3 also describes expected long-term (months to years) intralaboratory repeatability SD for all single-laboratory data sets evaluated thus far; however, the values of the parameters differ for different laboratories. The repeatability SD characteristic of the ISP measurements, SD_{ISP}, must be evaluated from replicated measurements of bands from gels similar to those from which the partial band measurements were obtained. This can be accomplished with K562 cell line control data.

ISP Repeatability SD

Figure 1 presents MBS and SD_{ISP} for several years of accumulated K562 cell line control data with summaries for casework, offender, and population gels plotted separately. The solid line in Fig. 1 represents SD_{Inter} as predicted by Eq 4; the SD_{ISP} for all gel formats are clearly smaller (less variable, better). However, the available K562 data do not include sufficiently large band sizes for reliable direct estimation of all three parameters of Eq 3. Observing that the ISP data are approximately parallel to the Eq 4 curve in the log-linear coordinates of Fig. 1, we estimate an ISP-specific value for the β_1 term while retaining the previously determined values for the β_2 and β_3 terms:

$$\beta_i' = \frac{\sum_{j=1}^N \dfrac{\overline{SD}_{ISPj}}{(1 + \overline{MBS}_j/19500)^{7.1}}}{N} \quad (5)$$

where $\overline{MBS}_j$ and $\overline{SD}_{ISPj}$ are the mean and SD of all available data for a given K562 allele regardless of gel format, and N is the number of groups of such data. For the 12 K562 bands summarized in Table 1, β_i' has a value of 4.5. Substituting this estimate into Eq 4 and simplifying the denominator, the ISP's expected long-term repeatability SD for a given bp band size is:

$$SD_{ISP} = 4.5\left(\frac{19500 + MBS}{19500}\right)^{7.1} \quad (6)$$

The dashed line in Fig. 1 represents Eq 6. The assumption of constant proportionality between the predicted SD_{Inter} and the

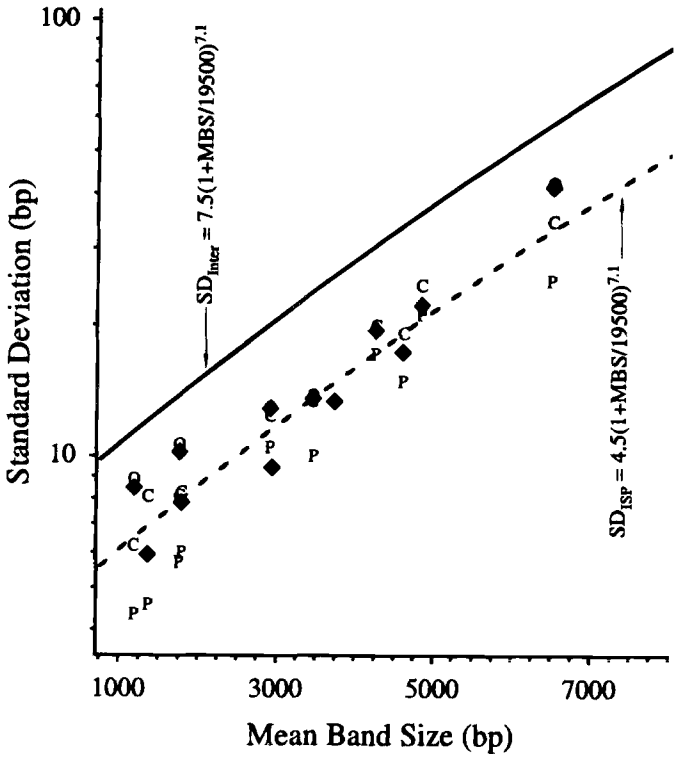

FIG. 1—*SD as a function of MBS. The long-term repeatability SD for K562 alleles at loci D1S7, D2S44, D4S139, D5S110, D10S28, and D17S26 is shown for casework (denoted "offender ("O"), and population ("P") format gels and for all gel formats combined ("◆"). The solid line represents interlaboratory reproducibility SD predicted from Eq 4. The dashed line represents the ISP repeatability SD predicted from Eq 6.*

observed SD_{ISP} provides the sole justification for extrapolating above the largest observed K562 MBS of 6400 bp.

ISP Measurement Correlation

If two measurements are not completely independent, then the uncertainty in the difference between the two measurements must be adjusted to account for their interaction: if the correlation is positive (i.e., if one measurement is larger than expected, then both are likely to be larger than expected), the uncertainty in the difference will be somewhat less than if the measurements were independent; if the correlation is negative (i.e., if one is larger than expected, then the other is likely to be smaller than expected), the uncertainty in the difference will be somewhat greater than if the measurements were independent. Previous studies have observed that same-lane, same-gel RFLP band size measurements are positively correlated (7–9). The uncertainty in determination of same-lane, same-gel partial digestion band $\hat{D}_i$ should thus be less than for two bands measured in completely different gels.

We have observed that the strength of measurement correlation for given K562 allele pairs varies considerably among laboratories, with larger (closer to 1.0) correlation in data from laboratories with larger (more variable, worse) repeatability characteristics (10). The median interlaboratory correlation between high and low K562 bands at loci D1S7, D2S44, D4S139, D5S110, D10S28, and D17S79 was estimated as a locus-independent 0.62. Others have found that the strength of the correlation declines with increasing difference in fragment size (7–9). Although research in this area is incomplete, we believe that RFLP measurement correlation for a given laboratory is a function of the laboratory-specific measurement repeatability and the gel-specific spatial geometry of the sample and calibration bands. Until a more satisfactory model is available, we approximate the correlation expected between a pair of bands measured in the same lane of a given gel empirically as a linear function of the bp size difference (7,8). After assembling all the K562 cell line control data derived from the same analytical gel into a single record (and eliminating all data from gels for which data from only one or two loci are available), the same data used to estimate SD_{ISP} can be used to estimate the expected correlation between ISP measurements of two DNA fragments located in the same lane of a given gel.

Figure 2 presents the absolute value of the observed K562 control band size differences

$$D_{mn} = |MBS_m - MBS_n|$$

plotted against the observed correlation for all pairs of K562 allele measurements derived from the same analytical gel

$$R_{mn} = \frac{\sum_{i=1}^{N} (bp_{mi} - MBS_m)(bp_{ni} - MBS_n)}{\sqrt{\left(\sum_{i=1}^{N} (bp_{mi} - MBS_m)^2\right)\left(\sum_{i=1}^{N} (bp_{ni} - MBS_n)^2\right)}}$$

where MBS_m and MBS_n are the mean values for the m^{th} and n^{th} K562 alleles, bp_{mi} and bp_{ni} are measurements for the two alleles in the i^{th} analytical gel considered, and N is the number of gels. Correlations for casework, offender, and population gels are plotted separately. Table 2 lists the correlations of all K562 allele pairs

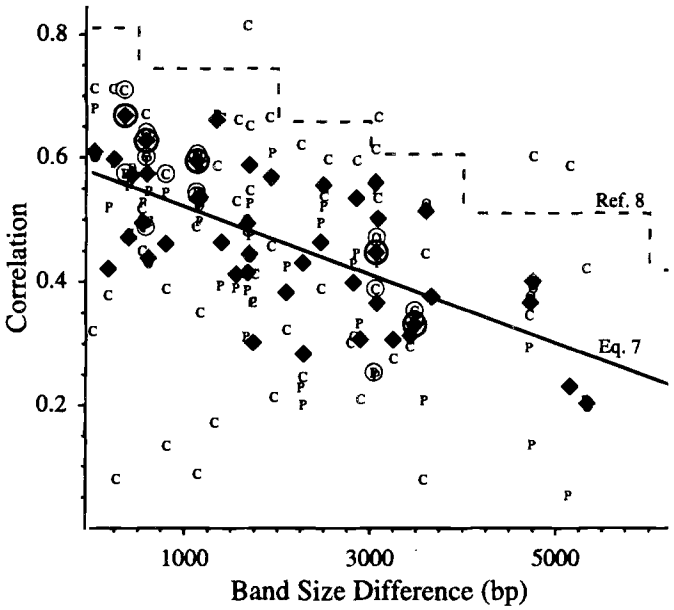

FIG. 2—*Correlation as a function of band size difference. The correlation between measurements for all pairs of K562 allele among loci D1S7, D2S44, D4S139, D5S110, D10S28, and D17S26 is shown for casework (denoted "C"), offender ("O"), and population ("P") format gels. The correlation for all pairs with data from at least 80 different analytical gels of any format is denoted "◆". Correlations for same-allele pairs are circled. The solid line represents the relationship between correlation and bp size difference of Eq 7. The dashed stair-step curve approximates the relationship reported in Ref 8.*

TABLE 2—*Measurement correlation of Illinois State Police K562 RFLP measurements.*

m		n		Number of Data					
Locus	Allele	Locus	Allele	Case*	Pop†	Off‡	Total	D_{mn}	R_{mn}
D1S7	High	D1S7	Low	38	72	0	110	352	0.67
D1S7	Low	D2S44	High	32	72	0	104	1336	0.66
D10S28	High	D10S28	Low	62	72	660	794	579	0.63
D10S28	High	D10S28	Low	62	72	660	794	579	0.63
D2S44	Low	D10S28	High	51	72	660	783	30	0.61
D1S7	High	D17S26	High	12	72	0	84	237	0.60
D1S7	High	D2S44	High	32	72	0	104	1688	0.59
D2S44	High	D2S44	Low	61	72	660	793	1125	0.59
D2S44	High	D17S26	High	15	72	0	87	1925	0.57
D1S7	Low	D17S26	High	12	72	0	84	589	0.57
D2S44	Low	D17S26	Low	15	72	0	87	426	0.57
D2S44	Low	D17S26	High	15	72	0	87	3050	0.56
D1S7	Low	D10S28	High	31	72	0	103	2490	0.56
D2S44	High	D10S28	High	51	72	660	783	1155	0.54
D1S7	High	D10S28	High	31	72	0	103	2843	0.53
D1S7	High	D4S139	Low	34	72	0	106	1148	0.53
D2S44	High	D4S139	High	55	72	660	787	3599	0.51
D10S28	High	D17S26	High	20	72	0	92	3080	0.50
D2S44	High	D4S139	Low	55	72	660	787	540	0.50
D2S44	Low	D4S139	Low	55	72	660	787	1665	0.49
D10S28	High	D17S26	Low	20	72	0	92	397	0.47
D1S7	Low	D2S44	Low	32	72	0	104	2461	0.46
D4S139	Low	D17S26	High	16	72	0	88	1385	0.46
D1S7	Low	D4S139	Low	34	72	0	106	796	0.46
D4S139	High	D4S139	Low	64	72	660	796	3059	0.45
D4S139	Low	D10S28	High	55	72	660	787	1695	0.45
D2S44	Low	D10S28	Low	51	72	660	783	608	0.44
D1S7	Low	D4S139	High	34	72	0	106	2263	0.43
D4S139	High	D17S26	High	16	72	0	88	1674	0.42
D10S28	Low	D17S26	Low	20	72	0	92	182	0.42
D2S44	High	D17S26	Low	15	72	0	87	1551	0.41
D4S139	High	D10S28	High	55	72	660	787	4753	0.40
D1S7	High	D2S44	Low	32	72	0	104	2813	0.40
D10S28	Low	D17S26	High	20	72	0	92	3659	0.38
D4S139	Low	D17S26	Low	16	72	0	88	2091	0.38
D2S44	Low	D4S139	High	55	72	660	787	4724	0.37
D1S7	Low	D10S28	Low	31	72	0	103	3069	0.37
D17S26	High	D17S26	Low	22	72	0	94	3476	0.33
D1S7	High	D10S28	Low	31	72	0	103	3421	0.31
D1S7	High	D17S26	Low	12	72	0	84	3239	0.31
D1S7	Low	D17S26	Low	12	72	0	84	2887	0.31
D2S44	High	D10S28	Low	51	72	660	783	1733	0.30
D4S139	Low	D10S28	Low	55	72	660	787	2274	0.28
D4S139	High	D17S26	Low	16	72	0	88	5150	0.23
D4S139	High	D10S28	Low	55	72	660	787	5332	0.20

*Casework.
†Population.
‡Offender.

for which at least 80 data pairs from any gel format are available. The solid line in Fig. 2 represents the linear least squares regression to this data:

$$R_{mn} \cong 0.58 - 0.000{,}055 \times D_{mn} \tag{7}$$

While the correlations among the ISP data are less strong than those reported in Ref 8, the trend is identical.

ISP Repeatability SD for $\hat{D}_i$

Substituting Eqs 6 and 7 into Eq 2 and noting that P_i must always be greater than T, the ISP's expected SD for a given $\hat{D}_i$ is:

$$SD_{\hat{D}_i} \cong 4.5 \sqrt{\left(\frac{19500 + P_i}{19500}\right)^{14.2} + \left(\frac{19500 + T}{19500}\right)^{14.2} - (1.16 - 0.000{,}11(P_i - T)) \left(\left(\frac{19500 + P_i}{19500}\right)\left(\frac{19500 + T}{19500}\right)\right)^{7.1}} \tag{8}$$

Expected Values for D_i

If the DNA sequence between the VNTR complete digestion and a partial digestion site is conserved, then all $\hat{D}_i$ should equal a constant number of bp, $\hat{D}_i$, within measurement uncertainty. If we can assume that the $\hat{D}_i$ follow a Gaussian distribution about D_i, we can estimate D_i with specified confidence from the observed

data. Further, we can quantitatively evaluate the probability that difference $\hat{D}_i$ between the measured size of a given P_i and T is, within expected measurement uncertainty, equal to D_i.

Calculation of Expected Values

The mean and SD of the $\hat{D}_i$ are not necessarily the best location and dispersion estimates for the expected bp size difference between a given P_i and T, given that: (1) there may be outlier values well away from the majority of the data, and (2) $\hat{D}_i$ for large T are known with less certainty than are those for smaller T (Eq 3). The median and interquartile range (IQR) location and dispersion estimates help alleviate the first concern because they are robust to outlier data. (The median and the mean are identical for symmetrical distributions. The SD equals 0.741 × IQR for Gaussian distributions (11)). A more complete, if more complex, approach is use of the predicted $SD_{\hat{D}_i}$ of Eq 8 as weighting factors (12) to give the weighted mean:

$$D_i = \sum_{i=1}^{N} \frac{\hat{D}_i}{SD^2_{\hat{D}i}} \Big/ \sum_{i=1}^{N} \frac{1}{SD^2_{\hat{D}i}} \tag{9}$$

and the SD of the weighted mean:

$$SD_{D_i} = \sqrt{1 \Big/ \sum_{i=1}^{N} \frac{1}{SD^2_{\hat{D}i}}} \tag{10}$$

From the relationship between the raw SD of a sample population and the SD of the unweighted sample mean, a "weighted SD" for the sample population can be approximated as:

$$SD_i = \sqrt{N} \times SD_{D_i} \tag{11}$$

Table 3 compares the ordinary, robust quartile, and weighted location and dispersion estimates for the smallest three partials at the six loci studied. The different location estimates (mean, median, and weighted mean) are in good agreement for all partials, with a maximum difference of 3%. The three SD_i estimates (SD, 0.741 × IQR, and weighted SD) agree to within 20% for most partials, with a maximum difference of 50%. Because the weighted estimates most completely utilize the available information, we use the weighted D_i and weighted SD_i estimates throughout the rest of this discussion.

TABLE 3—*Expected value of partial band size difference, $D_i = P_i - T$, in pristine samples.*

			Ordinary*		Quartile†		Weighted‡	
Locus	Partial	n	D_i	SD_i	D_i	SD_i	D_i	SD_i
D1S7	D_1	115	204	14	203	12	205	13
	D_2	114	394	15	397	13	398	14
	D_3	107	610	20	613	16	612	15
D2S44	D_1	135	576	11	576	10	577	10
	D_2	165	1770	15	1771	16	1771	14
	D_3	132	2329	21	2331	19	2333	18
D4S139	D_1	35	270	24	265	22	269	22
	D_2	17	1068	57	1065	34	1065	40
	D_3	27	1559	36	1557	41	1561	42
D5S110	D_1	48	189	12	190	11	191	11
	D_2	43	307	12	305	13	305	11
	D_3	20	498	16	499	11	497	14
D10S28	D_1	83	267	12	267	10	267	10
	D_2	70	1396	27	1397	22	1399	16
	D_3	41	1665	29	1665	32	1675	19
D17S26	D_1	42	248	26	246	25	239	18
	D_2	75	980	21	979	10	980	15
	D_3	28	1275	28	1280	25	1289	18

*Mean and SD.
†Median and 0.741 × interquartile range.
‡Weighted mean and weighted SD.

Distribution Normality

Figure 3 displays the observed $\hat{D}_i$ distributions for all partials in histogram form, with each distribution standardized to have zero mean and unit variance

$$\hat{D}_i' = \frac{\hat{D}_i - D_i}{SD_i} \tag{12}$$

The standard Gaussian of mean zero and unit variance is shown for each set of $\hat{D}_i'$ histograms. All distributions at loci D1S7, D2S44, and D10S28 are well described as Gaussian distributions.

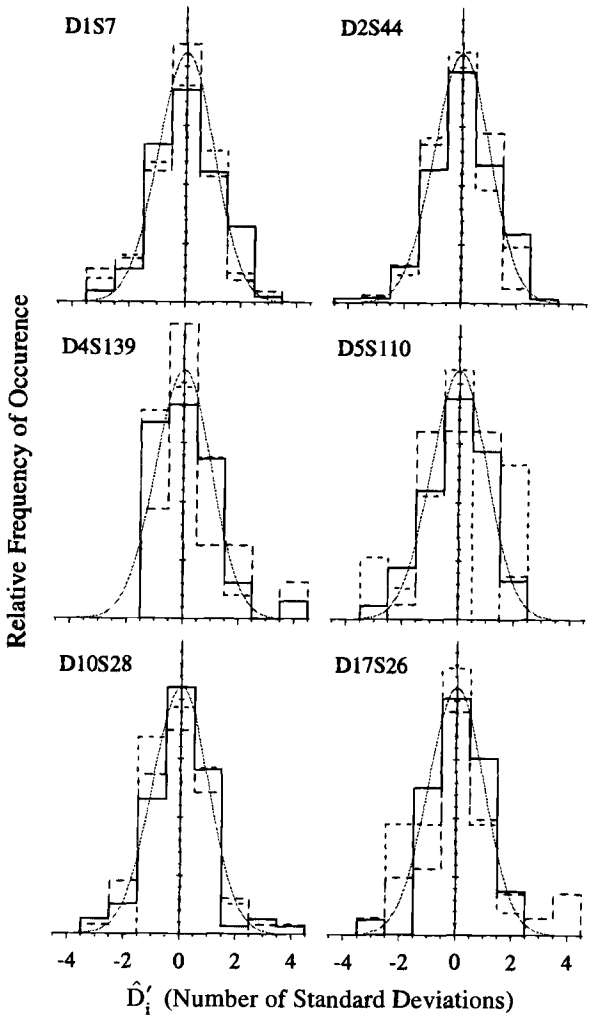

FIG. 3—*ISP $\hat{D}_i$ distributions. Unit-area histograms for $\hat{D}_1'$, $\hat{D}_2'$, and $\hat{D}_3'$ (the smallest three $\hat{D}_i$ standardized to zero mean and unit variance) at loci D1S7, D2S44, D4S139, D5S110, D10S28, and D17S26 are denoted with solid, long-dashed, and short-dashed lines, respectively. The dotted line represents the unit-area standard Gaussian. The locus designation is provided in the upper left-hand corner of each subgraph.*

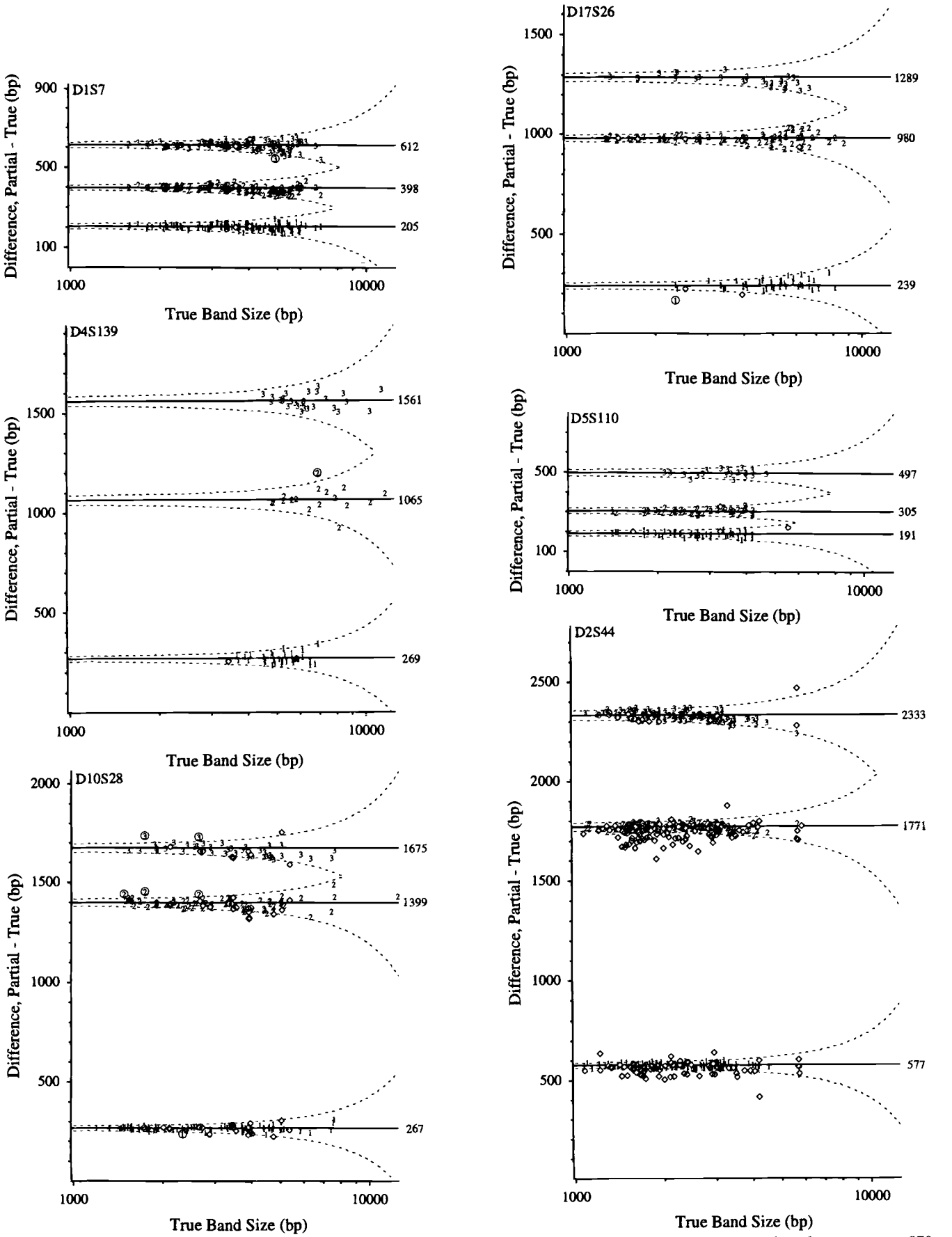

FIG. 4

For caption please see page 870.

Some of the distributions for loci D5S110 and D17S26 may be composites of two or more different groups of data, although one group is dominant in each distribution. All distributions for locus D4S139 appear to be positively skewed.

Confidence Intervals for Pristine Samples

Assuming approximately symmetric distributions, $\hat{D}_i$ that are not within measurement uncertainty of the expected D_i bp size can be identified with specified confidence. An approximate confidence interval about D_i can be derived from the hypothesis test for the difference between two means:

$$|\hat{D}_i - D_i| \leq z_{\alpha/2}\sqrt{SD^2_{\hat{D}_i} + \frac{SD_i^2}{N}} \qquad (13)$$

where $z_{\alpha/2}$ is the number of standard deviations about the mean of a Gaussian distribution such that the area of the distribution within the interval (mean $\pm$ $z_{\alpha/2}$ SD) is $(1 - \alpha)$ of the total area. The value of $z_{\alpha/2}$ for given confidence coefficients, $1 - \alpha$, is widely tabulated and typically available as a spreadsheet function: $z_{\alpha/2}$ is approximately 2 for 95% ($\alpha = 0.05$) confidence intervals and 2.6 for 99% ($\alpha = 0.01$) confidence intervals.

These confidence intervals can be used to screen the data used in their calculation for gross violations of the underlying assumptions. When considered as a whole, the expected 1% (13 of the 1297 individual $\hat{D}_i$) have less than a 1% probability of being from a symmetric population of mean D_i. However, these data are not uniformly distributed among the different partials; there are no 99% outliers in the 111 locus D5S110 $\hat{D}_i$ although there are six in the 194 locus D10S28 $\hat{D}_i$. Figure 4 provides 95% confidence intervals for the six genetic loci considered in this work. The data that are outside 99% confidence intervals are labeled; while most appear to be the tail-members of their distribution, the few that are graphically distinct could be miss-assignments or represent mutant sequences.

The confidence intervals can be used to evaluate quickly partial digestion as a potential cause of anomalous data. While the intervals shown for each locus in Fig. 4 are strictly applicable only to pristine sample data from RFLP measurements with repeatability and correlation characteristics similar to those of the ISP, they should be approximately correct for most laboratories in good analytical control. Laboratories with significantly different measurement characteristics should calculate intervals appropriate to their RFLP measurement characteristics. The tables in Part 1 were generated using the interlaboratory measurement reproducibility estimate of Eq 4 and the approximately constant $R_{mn} = 0.62$ value observed in a preliminary analysis of intralaboratory correlation (10).

Evidentiary Samples

All partial band data for evidentiary samples are shown in Fig. 4, with summary results given in Table 4. Far too many data do not fall within the RFLP measurement uncertainty 99% confidence bounds established using pristine samples for measurement uncertainty to account for the differences between sample types. For the loci with sufficient data to enable meaningful statistical summation (D2S44 and D10S28), the excess size of the evidentiary partials is somewhat less than that observed in pristine sample partials by 1 to 35 bp. The data are too limited to identify any functional relationships between the excess size of the partial and the differences in excess size between sample types.

TABLE 4—*Expected value of partial band size difference, $D_i = P_i - T$, in evidentiary samples.*

Locus	Partial	n*	>1%†	Weighted‡ D_i	Weighted‡ SD_i	Weighted‡ Δ§	Weighted‡ %Δ‖
D1S7	D_1	2	0				
	D_2	2	0				
	D_3	2	0				
D2S44	D_1	82	27	559	10	−17	−3.0
	D_2	103	37	1737	14	−34	−1.9
	D_3	16	1	2319	17	−14	−0.6
D4S139	D_1	1	0				
D5S110	D_1	3	0				
	D_2	1	0				
D10S28	D_1	24	1	266	13	−1	−0.3
	D_2	18	2	1381	18	−19	−1.3
	D_3	9	0	1657	20	−18	−1.1
D17S26	D_1	2	1				
	D_2	5	0				

*Total number of number of partial band data available for evidentiary samples.

†Number of evidentiary partials outside 99% confidence intervals for pristine samples.

‡Weighted mean and SD for evidentiary data.

§Difference, evidentiary—pristine weighted D_i.

‖Percent difference, $100(\Delta)$/pristine D_i.

The weighted SD_i's characteristic of the evidentiary samples are very similar to those characteristic of pristine samples. Although the evidentiary sample partials appear to have less excess bp size, that excess size is not much more variable among different samples than it is for pristine samples.

We speculate that this reduction in excess size is related to the "anodal band shift" frequently observed in evidentiary samples (13,14). Partial digestion studies of samples intentionally exposed to a variety of environmental insults could help provide insight into the sources of these changes in electrophoretic mobility. Unfortunately, little information on the nature and history of the evidentiary samples studied here is available to us.

Relative Location of Partial *Hae* III Digestion Sites

The geometric relationship between the fully digested and the three smallest partial bands can be visualized as one of the patterns listed in Table 5 (and presented graphically in Fig. 1 of Part 1). For a given locus, pattern I (two partial digestion sites, one on each side of the VNTR) must produce a D_3 that is equal to the

FIG. 4—*See page 869 for Fig. 4. 95% confidence limits for ISP D_i as functions of true band size. Approximate 95% confidence limits for the expected size differences are shown for loci D1S7, D2S44, D4S139, D5S110, D10S28, and D17S26 as dotted lines. The weighted mean estimates of D_i are denoted as solid lines; the quantitative value is listed at the right end of each line. The $\hat{D}_1$, $\hat{D}_2$, and $\hat{D}_3$ for products of intentional partial digestion of pristine samples are denoted "1", "2", and "3", respectively; $\hat{D}_i$ that are outside the 99% confidence limits are circled. The $\hat{D}_1$, $\hat{D}_2$, and $\hat{D}_3$ from evidentiary samples are denoted "◇".*

TABLE 5—*Possible geometries of partial digestion sites at a given locus.*

	Pattern*	T	P_1	D_1	P_2	D_2	P_3	D_3	$D_3 - (D_1 + D_2)$
I	...a-A-B--b...	AB	aB	aA	Ab	Bb	ab	aA + Bb	0
IIa	...a-A-B--b-c...	AB	aB	aA	Ab	Bb	Ac	Bc	bc-aA ⇔ 0
IIb	...b-a-A-B---c...	AB	aB	aA	bB	bA	Ac	Bc	Bc-aA-bA ⇔ 0
IIc	...c--a-A-B--b...	AB	aB	aA	Ab	Bb	cB	cA	ca-Bb ⇔ 0
III	...c-b-a-A-B...	AB	aB	aA	bB	bA	cB	cA	cb > 0

*"A" and "B" designate complete digest sites; "a", "b", and "c" designate partial digestion sites; the spacing between symbols indicates the minimum relative number of bp between sites to insure $P_1 < P_2 < P_3$. See Fig. 1 of Part 1 of this series (1).

TABLE 6—*Plausible geometries for observed partial digestion sites.*

	All Data*					Complete Data†				
Locus	N‡		Δ§	SD	P(Δ = 0)‖	N¶	Δ§	SD	P(Δ = 0)‖	Pattern**
D1S7	107	115	9	25	0.71	106	10	25	0.68	I, II
D2S44	132	165	−15	25	0.56	114	−14	25	0.57	I, II
D4S139	17	35	227	62	0.00	6	232	53	0.00	II, III
D5S110	20	48	2	21	0.94	16	−1	22	0.98	I, II
D10S28	41	83	9	26	0.72	40	9	28	0.74	I, II
D17S26	28	75	70	30	0.02	16	85	32	0.01	II, III

*Calculated from weighted mean and SD summary statistics of Table 3.
†Calculated from weighted mean and SD summary statistics for the subset of samples expressing all three partial digestion bands.
‡Smallest and largest number of data available for D_1, D_2, or D_3.
§$D_3 - (D_1 + D_2)$.
‖Probability of observing a Δ this large given a "true" Δ of zero.
¶Number of samples with complete D_1, D_2, and D_3 data.
**Plausible geometries; see Table 5.

sum $D_1 + D_2$. Pattern II (three partial digestion sites, two on one side of the VNTR region and one on the other) is the only geometry that can produce a D_3 that is smaller than the sum of $D_1 + D_2$, but pattern II can also produce a D_3 that is equal to or greater than $D_1 + D_2$. Pattern III (three partial digestion sites, all on the same side of the VNTR) must produce a D_3 that is larger than the sum $D_1 + D_2$.

The expected difference in numbers of bp between D_3 and $(D_1 + D_2)$ can be calculated from the summary values listed in Table 3

$$\Delta = D_3 - (D_1 + D_2) \quad (14)$$

where all three D_i for each locus are calculated from all available data. The individual D_i are positively correlated; however, the magnitude of the correlations is modest (0.1 to 0.4) and follows no discernible pattern. We therefore approximate the expected SD for Δ by

$$SD_\Delta = \sqrt{SD_1^2 + SD_2^2 + SD_3^2} \quad (15)$$

Table 6 lists the Δ and SD_Δ values, along with the probability of observing each Δ assuming that the "true" value for Δ is zero. Table 6 also summarizes identical calculations based on the subset of samples that gave all three partial bands for the given locus.

The two sets of summary values are very similar for all loci, with D1S7, D2S44, D5S110, and D10S28 following patterns I or II and D4S139 and D17S26 following patterns II or III. Given that three partial bands were observed at the D1S7 locus for most of the intentionally partially digested DNA samples and that no more than three partial bands were ever observed there, it is likely that the geometry of partial digestion at D1S7 is pattern I. Given the composite appearance of the D_i histograms for D17S26, the data for this locus may well arise from a number of closely spaced partial digestion sites.

Additional information regarding the geometry of the partial digestion products can be obtained from analysis of the three banded patterns occasionally observed with limit digests. At least some of these patterns arise from *Hae* III sites internal to the VNTR block. We have used such information to demonstrate that the D5S110 geometry is pattern I (1).

Acknowledgments

We wish to thank Taylor C. Scott of the Illinois State Police for supplying the historical K562 cell line control data and Stefan Leigh of the Statistical Engineering Division, NIST, for his assistance with statistical concepts and notation. This study was supported in part by a grant to EAB from the Midwestern Association of Forensic Scientists.

References

1. Benzinger EA, Emerek E, Grigsby N, Duewer D, Lovekamp ML, Deadman H, et al. Products of partial digestion with *Hae* III. Part 1. Characterization, casework experience, and confirmation of the theory of three-, four- and five-banded RFLP pattern origins using partial digestion. J Forensic Sci 1997;42(5):850–863.
2. Bevington PR. Data reduction and error analysis for the physical sciences. New York: McGraw-Hill Book Co., 1969;58–65.
3. International Organization for Standardization. ISO 3534-1: statistics—vocabulary and symbols, ISO, Geneva, Switzerland. 1993; definitions 3.14–3.25.
4. Mudd JL, Baechtel FS, Duewer DL, Currie LA, Reeder DJ, Liu H-K, et al. Interlaboratory comparison of autoradiographic DNA profiling measurements. 1. Data and summary statistics. Anal Chem 1994;66:3303–17.
5. 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.
6. 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 >10 k base pairs. Anal Chem 1996;68:1941–47.
7. 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.
8. Evett IW, Scrange JK, Pinchin R. An efficient procedure for interpreting DNA single locus profiling data in crime cases. J Forensic Sci Soc 1992;32:307–24.
9. Devlin B, Risch N, Roeder K. Statistical evaluation of DNA fingerprinting—a critique of the NRC's report. Science 1993;259:1096.
10. Duewer DL, Reeder DJ, Liu H-K, Baechtel FS. Proposed CODIS K562 bivariate tolerance limits. Proceedings of the Fifth International Symposium on Human Identification; 1994 Oct 9–11; Scottsdale, AZ. Promega Corporation, Madison, WI 53711, 1995;167.
11. Stuart A, Ord JK, Eds. Kendall's Advanced Theory of Statistics, 5th Edition, Volume 1. Oxford University Press, NY 1987.
12. Meyer SL. Data Analysis for Scientists and Engineers. New York: John Wiley & Sons, 1975;146.
13. McNally L, Baird M, McElfresh K, Eisenberg A, Balazs I. Increased migration rate observed in DNA from evidentiary material precludes the sue of sample mixing to resolve forensic cases of identity. Appl Theor Electrophor 1990;1:267–72.
14. 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.

Additional information and reprint requests:

The original sizing data for all partial bands described in this report are available upon request from the authors. Requests for this data, any additional information, and/or reprints should be addressed to:

Elizabeth A. Benzinger, Ph.D.
Ohio Bureau of Criminal Identification and Investigation
P.O. Box 365
London, Ohio 43140
Email:ebenzinger@ag.ohio.gov