TECHNICAL NOTE

David L. Duewer,[1] Ph.D.; Margaret C. Kline,[2] M.S.; Janette W. Redman[2]; Pamela J. Newall,[3] M.A.; and Dennis J. Reeder,[2,4] Ph.D.

NIST Mixed Stain Studies #1 and #2: Interlaboratory Comparison of DNA Quantification Practice and Short Tandem Repeat Multiplex Performance with Multiple-Source Samples*,†

REFERENCE: Duewer DL, Kline MC, Redman JW, Newall PJ, Reeder DJ. NIST mixed stain studies #1 and #2: interlaboratory comparison of DNA quantification practice and short tandem repeat multiplex performance with multiple-source samples. J Forensic Sci 2001;46(5):1199–1210.

ABSTRACT: The Mixed Stain Study 1 (MSS1, Apr.–Nov. 1997) and Mixed Stain Study 2 (MSS2, Jan.–May 1999) evaluated multiplexed short-tandem repeat (STR) DNA typing systems with samples containing DNA from more than one source. These interlaboratory challenge studies evaluated forensic STR measurement, interpretation, and reporting practice using well-characterized samples of very different analytical difficulty. None of the relatively few errors reported in either exercise resulted in a false identification of a reference source; several errors in evaluating the unknown source in three-source samples would hinder matching the profile in any archival database. None of the measurement anomalies reported is associated with any particular STR multiplex; all DNA amplification anomalies are associated with inefficient DNA extraction, inaccurate DNA quantitation, and/or analytical threshold policies.

KEYWORDS: forensic science, blood stains, DNA fingerprinting, DNA typing, evaluation studies, interlaboratory comparison, polymerase chain reaction, reproducibility of results, semen stains, short tandem repeat (STR) alleles

Here, we report results from two interlaboratory comparison exercises conducted by the National Institute of Standards and Technology (NIST). These studies explored the performance of multiplexed short-tandem repeat (STR) DNA typing systems with samples containing DNA from more than one source. Both the Mixed Stain Study #1 (MSS1) and #2 (MSS2) were motivated by concerns that multiple-source samples could present measurement and interpretation challenges to systems requiring the simultaneous amplification of DNA at many different STR loci: i.e., STR multiplexes (1,2).

Validation studies addressing the above concerns have and continue to be performed by individual forensic laboratories as well as by the STR systems' manufacturers. Further, forensic DNA analysts regularly participate in proficiency tests that, while primarily intended to evaluate and document the sample evaluation skills of analysts, also evaluate analytical methods. However, both MSS1 and MSS2 were "challenge studies" designed to illuminate potential measurement weaknesses. By recruiting users of diverse analytical systems and analysts of varying experience, and challenging them with unusual and difficult samples, errors and analytical failures were elicited. Most measurement abnormalities were recognized as such by the performing analysts; these failures waste analytical resources. There were a very few instances of true alleles excluded from the "match profile" of an unknown donor in a mixed-source sample. While unlikely to affect (or go unrecognized) in any direct casework comparison, such errors could lead to less efficient lead prioritization by the Federal Bureau of Investigation's CODIS, the Forensic Science Service's National DNA Database, or other DNA database systems (3,4).

None of the clerical, measurement, or interpretation anomalies observed in MSS1 or MSS2 can be attributed to particular STR systems or instrumentation. We attribute the few clerical errors to analyst inattention. All allele measurement failures (true alleles not called, stutter called as an allele) are attributable to inefficient extraction of DNA from the samples, inaccurate estimation of the amount of DNA used in the PCR mixture, and/or analytical threshold policies. To the extent that simultaneous amplification of multiple alleles demands better control of initial conditions as the complexity of the system increases, highly multiplexed STR systems may well require improved DNA quantification technology.

[1] Analytical Chemistry Division, Chemical Science and Technology Laboratory, National Institute of Standards and Technology, Gaithersburg, MD 20899.
[2] Biotechnology Division, Chemical Science and Technology Laboratory, National Institute of Standards and Technology, Gaithersburg, MD 20899.
[3] Head of Research, Development and Defence Projects, The Centre of Forensic Sciences, 25 Grosvenor Street, Toronto, Ontario M7A 2G8.
[4] Current address: Applied Biosystems, 850 Lincoln Centre Drive, Foster City, CA 94404.
* This work was supported in part by the National Institute of Justice, U.S. Department of Justice.
*,† Certain commercial equipment, instruments, or materials are identified in this report to specify adequately experimental conditions or reported results. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology or the Centre of Forensic Sciences, nor does it imply that the equipment, instruments, or materials identified are necessarily the best available for the purpose.
Received 1 June 2000; and in revised form 9 Nov. 2000; accepted 14 Nov. 2000.

Summary results of both the MSS1 and MSS2 studies were provided to all participants as each study was completed. Preliminary interpretations of the studies' results were presented as soon thereafter as possible (5,6). This report describes the critical aspects of the study designs and results, with a focus on the performance and importance of DNA extraction and DNA quantification technologies.

Materials and Methods

To better anticipate the reference material and measurement quality assurance needs of the human identity community, NIST assesses real-world performance of DNA typing technologies via interlaboratory comparison exercises. In collaboration with members of the Scientific Working Group for DNA Analysis Methods (SWGDAM, previously termed T(echnical)WGDAM), commercial "CTT" (CFS1PO, TH01, TPOX) triplex and "CTTv" (CTT + VWA) quadraplex STR systems were evaluated by 34 participants during the period of Jan.–May, 1996 (7). Several participants in this CTT study reported unequal amplification of alleles at one or more loci.

Since our laboratories also occasionally observed unequal allele amplification with these and other STR systems, MSS1 and then MSS2 were designed to document the extent and forensic implications of this potential measurement problem. The MSS1 samples were distributed to 28 laboratories beginning in Apr. 1997; the final data set was received in Nov. 1997. The MSS2 samples were distributed to 52 laboratories beginning in Jan. 1999; the final data set was received in May 1999. Table 1 lists the participants in the MSS1 and MSS2 studies.

Mixed Stain Study #1 (MSS1)

This study was designed around a standard forensic casework problem: identification of all sources of DNA in multiple-source samples given a complete reference set of potential sources. Eleven sets of samples were prepared on S&S 903 paper (Schleicher & Schuell, Inc., Keene, NH) from human Buffy coat cells (QC Products, Inc., Pompano Beach, FL). Six of the samples were single-source reference samples, four were two-source mixtures, and one was a three-source mixture. Only the reference sources were

TABLE 1—*MSS1 and MSS2 participants.*

Laboratory	CTT	MSS1	MSS2
Alabama Department of Forensic Sciences, Huntsville Laboratory	X	X	X
Alabama Department of Forensic Sciences, Mobile Laboratory			X
Arizona Department of Public Safety, Phoenix		X	X
Arkansas State Crime laboratory, Little Rock		X	X
Armed Forces DNA Identity Laboratory, Rockville, MD	X		X
Broward County Crime Laboratory, Fort Lauderdale, FL			X
California Department of Justice DNA Laboratory, Berkeley	X	X	X
Central Utah Criminalistics Laboratory, Salt Lake City			X
Centre of Forensic Sciences, Toronto, Ontario	X	X	X
City of Phoenix Police Department, AZ			X
Commonwealth of Virginia Department of Criminal Justice Services, Richmond	X	X	X
Detroit Police Department, Forensic Services Division, MI		X	X
FBI Laboratory, DNA Analysis Unit I, DC			X
Genelex Corporation, Seattle, WA		X	X
Illinois State Police R&D Laboratory, Springfield	X	X	X
Indiana State Police Laboratory, Indianapolis			X
Kentucky State Police Forensic Laboratory, Frankfort	X		X
Laboratory Corporation of America, RTP, NC	X		X
Las Vegas Metropolitan Police Department, NV			X
Maine State Police Crime Laboratory, Augusta		X	
Maryland State Police Crime Laboratory, Pikesville			X
Miami-Dade Police Department, Miami, FL			X
Minnesota Forensic Science Laboratory, St. Paul	X	X	X
New York City, Office of the Chief Medical Examiner, NY	X	X	X
New York State Police Forensic Investigation Center, Albany			X
North Carolina State Bureau of Investigation, Raleigh	X	X	X
North Louisiana Criminalistics Laboratory, Shreveport	X		X
Oklahoma State Bureau of Investigation, Oklahoma City			X
Orange County Sheriff-Coroner Department, Santa Ana, CA	X	X	X
Oregon State Police Forensic laboratory, Portland		X	X
Orlando Regional Crime Laboratory, Orlando, FL	X	X	
Palm Beach County Sheriff's Office Crime Laboratory, West Palm Beach, FL	X	X	X
PE Biosystems, Foster City, CA	X		X
Pennsylvania State Police DNA Laboratory, Greensburg			X
Promega Corporation, Madison, WI	X	X	X
Royal Canadian Mounted Police, Central Forensic Laboratory, Ottawa, Ontario	X	X	X
State of Connecticut Department of Public Safety, Meridian	X		X
State of Delaware, Office of the Chief Medical Examiner, Wilmington			X
State of Michigan, Department of State Police DNA Laboratory, East Lansing	X	X	X
State of Ohio, Bureau of Criminal Identification and Investigation, London			X
Suffolk County Crime Laboratory, Hauppauge, NY	X	X	X
Tallahassee Regional Crime Laboratory, Tallahassee, FL		X	
Texas Department of Public Safety, Austin			X
Texas Department of Public Safety, Garland			X
The Bode Technology Group, Inc, Springfield, VA.			X
United States Army Criminal Investigation Laboratory, Forest Park, GA			X
Vermont Forensic Laboratory, Waterbury			X
Washoe County Sheriff's Office, Reno, NV	X		X

TABLE 2—*MSS1 and MSS2 samples.*

Study	Samples	Matrix	Composition		
MSS1	349	Buffy coats, S&S 903 paper	30–50 ng ♀$_{349}$		
	350	″	30–50 ng ♀$_{350}$		
	351	″	30–50 ng ♀$_{351}$		
	352	″	30–50 ng ♂$_{352}$		
	353	″	30–50 ng ♀$_{353}$		
	354	″	30–50 ng ♀$_{354}$		
	A	″	30–50 ng ♀$_{349}$	30–50 ng ♀$_{350}$	
	B	″	30–50 ng ♀$_{350}$	30–50 ng ♀$_{353}$	
	C	″	30–50 ng ♂$_{352}$	30–50 ng ♀$_{354}$	
	D	″	30–50 ng ♀$_{349}$	30–50 ng ♀$_{351}$	
	E	″	30–50 ng ♀$_{349}$	30–50 ng ♀$_{350}$	30–50 ng ♂$_{352}$
MSS2, Set 1	F	Whole blood, cotton cloth	0.9(2) μg ♀$_{N01}$		
	G	Semen, cotton cloth	1.2(1) μg ♂$_{1087}$		
	H	Blood & semen, cotton cloth	0.8(2) μg ♀$_{N01}$	1.3(1) μg ♂$_{1087}$	1.2(1) μg ♂$_{1039}$
MSS2, Set 2	J	Whole blood, cotton cloth	1.7(3) μg ♀$_{N02}$		
	K	Semen, cotton cloth	0.8(1) μg ♂$_{1131}$		
	L	Blood & semen, cotton cloth	1.7(3) μg ♀$_{N02}$	0.5(1) μg ♂$_{1140}$	
MSS2, Set 3	M	Extracted DNA, TE buffer	2.5(3) ng/μL {♀$_{N03}$, ♂$_{N04}$}		
	N	″	1.0(1) ng/μL {♀$_{N03}$, ♂$_{N04}$}		
	O	″	5.0(5) ng/μL {♀$_{N03}$, ♂$_{N04}$}		
	P	″	1.0(1) ng/μL {♀$_{N03}$, ♂$_{N04}$}		
	Q	″	0.5(1) ng/μL {♀$_{N03}$, ♂$_{N04}$}		

present in the multiple-source samples; all of the reference sources were used in one or more of the multiple-source samples. The particular source combinations were selected to minimize allelic overlap; i.e., to provide as uncomplicated a set of samples as possible given the source materials available. Table 2 lists the composition of all MSS1 samples.

Each stain was targeted to contain a minimum of 30 ng DNA per source. The concentration of DNA extractable from the Buffy coat cells for each of the six DNA sources was estimated as the average of replicate slot blot evaluations (8) of organic (9) and Chelex® (Bio-Rad Industries, Hercules, CA) extracts (10). All stains were prepared from volumetric aliquots of continuously stirred suspensions of the individual source materials in Phosphate Buffered Saline (PBS) (11). All stains for each sample set were thoroughly air-dried in a laminar-flow hood, labeled, and stored at −20°C before preparing the next sample. The multiple-source samples were prepared by sequential addition of the individual sources, with thorough drying under laminar flow between additions.

Two stains of each of the 11 samples were supplied to each participant. Several participants who did not achieve satisfactory amplification for one or more of the samples in the initial distribution were provided with a complete second set of stains. Participants were asked to report complete allelic profiles for all samples, using as many STR systems and instruments as possible. Participants were further asked to identify, using their standard casework protocol, all of the reference sources represented in all of the multiple-source samples.

Mixed Stain Study #2 (MSS2)

This study was designed around an emerging forensic opportunity: defining a "searchable profile" for an unknown DNA source in a multiple-source stain given an incomplete reference set of potential sources. It was also designed to provide quantitative information on pre-amplification (DNA extraction and quantity determination) stages of the STR measurement protocols. Two qualitatively different kinds of quantitatively well-defined samples were used to achieve these goals.

Sets 1 and 2, Stains on Cotton Cloth—Two sample sets of three stains each were prepared to present different unknown-source scenarios. Both sets consisted of a female reference prepared from whole blood, a male reference prepared from commercially obtained semen (Cryogenics Laboratories, Inc., North Roseville, MN), and a "questioned-sample" prepared from a mixture of whole blood and semen. The three-source sample of the first scenario contained DNA from both the female and the male reference sources plus DNA from a second "unknown" male. The two-source sample of the second scenario contained DNA from the reference female and a male source different from that of the male reference. All stains were prepared under laminar flow on 4 cm by 4 cm squares of white cotton cloth (Jo Ann Fabrics, Hagerstown, MD) that had been bleached, twice washed, and UV sterilized. After preparation, stains were dried for 2 h at ambient temperature, labeled, sealed under vacuum in aluminized Mylar® bags (MIL-B-131F Class I, Columbus Packages Co., Columbus, GA), and stored at −20°C. Table 2 lists the composition of these samples.

Each stain was targeted to contain about 1 μg (1000 ng) of DNA per source. The DNA concentration of each whole blood was estimated from a white cell count obtained shortly before blood donation. The DNA concentration of each semen was estimated from the original sperm count; the number of intact sperm at the time of sample preparation was not determined. Unlike the MSS1 samples, the multiple-source MSS2 samples were prepared from single aliquots of a continuously stirred PBS mixture of blood and semen. There were no detected qualitative or quantitative differences between stains produced at the beginning of production and those produced near the end for any of the six samples. Complete sample preparation details are provided elsewhere (6).

For these six samples, participants were asked to: 1) specify all possible profiles for all sources in each sample, 2) provide a "CODIS search profile" for the unknown source in the two multiple-source samples, and 3) estimate the total amount of recoverable DNA in each sample (ng/stain). We did not specify the stringency or the format of the search profile. Participants were also asked to provide all "relevant details" of their extraction and DNA quantitation protocols, again without further specification.

Set 3, Buffered DNA Solutions in Sealed Vials—A four-level concentration series was prepared to evaluate the accuracy of each participant's quantitative DNA measurements. All samples were prepared from a two-source stock solution of extracted DNA in tris-EDTA (TE) buffer (12). Aliquots of this stock solution were individually diluted with TE buffer to produce sample solutions at the desired DNA concentrations. Approximately 30 μL of the continuously-stirred sample solutions were aliquoted into 500 μL limited-volume vials (SARSTEDT, Inc., Newton, NC), sealed under argon, labeled, and stored at −20°C. Table 2 lists the composition of these samples.

The DNA used in the stock solution was an approximately equal blend of material extracted from blood donated by a female and a male source. The mixed DNA was repurified (13). The final purity and total DNA concentration of the mixture were verified by spectrophotometry and yield gels. The transmittance scale of the UV/Vis spectrophotometer was confirmed using NIST SRM® 2031a, Metal-on-Fused-Silica Filters for Spectrophotometry; the wavelength scale was verified using NIST SRM® 2034, Holmium Oxide Solution Wavelength Standard from 240 nm to 650 nm.

Note that the 1.0 ng/μL DNA concentration level was labeled as two separate samples, N and P. All vials for these two samples were produced as a single lot; the dual labeling established the samples as true replicates for assessing within-laboratory measurement performance characteristics.

For these five samples, participants were asked to estimate the concentration of DNA in each vial (ng/μL). Participants were informed that the concentration of DNA in all vials was within the range 0.2 ng/μL to 20 ng/μL.

Results

Multiplexes and Instrumentation

Table 3 lists the STR multiplexes used in the MSS1 and MSS2 and the number of participants using each. Many participants used more than one multiplex, particularly in MSS2. More than half of the MSS2 participants reported alleles at all 13 CODIS core loci (14) plus amelogenin; only one MSS2 participant reported alleles for fewer than seven of these core loci. Table 4 lists the instrumentation employed. Several participants in MSS1 used more than one instrument. None of the multiplexes or any of the instruments were systematically associated with any measurement artifact.

TABLE 3—*STR Multiplexes used by participants.*

STR Multiplex	#Loci	MSS1	MSS2
AmpFℓSTR Blue™	3	4	2
AmpFℓSTR COfiler™	5	…	23
AmpFℓSTR CTT	3	3	1
AmpFℓSTR Green™ I	3	3	2
AmpFℓSTR Green™ II	3	4	…
AmpFℓSTR Profiler™	10	6	…
AmpFℓSTR Profiler Plus™	10	2	30
AmpFℓSTR Yellow™	3	1	…
BHO Quad	4	2	…
Promega CTTv	4	1	…
Promega FFv	3	1	…
Promega PowerPlex™ 1.1	8	9	11
Promega PowerPlex™ 1.2	8	1	1
Promega PowerPlex™ 2.1	9		
	Total	37	70

STR Profiles

Two participants in MSS1 and four in MSS2 incorrectly specified one or more alleles for one or more samples in their initial data submission. One of the MSS1 and three of the MSS2 errors were made copying correct data from a worksheet into the final report, with the one and only error involving multiple alleles arising from a misaligned spreadsheet column. Two participants misassigned one allele each while analyzing their gel images. In all cases, the analysts involved found and corrected the error after reexamination of their primary data. Additionally, five MSS2 participants called to our attention one or more errors we made when transcribing their data into our database—and two errors made in the entry of MSS1 data were identified while preparing this manuscript. Given that MSS1 and MSS2 both required atypical data evaluations and report formats, this does not represent an error rate for casework or other "routine" analyses. Table 5 summarizes profiling performance for all nondifferentially extracted samples (i.e., all MSS1 samples and the single sources samples of MSS2) after correction of all known

TABLE 4—*Instrumentation used by participants.*

Instruments	MSS1	MSS2
ABI 310	5	21
ABI 373	1	1
ABI 377	7	11
Hitachi FMBio	10	11
MD FLuorImager	1	1
Silver stain	1	
Total	25	45

TABLE 5—*Profiling performance for nondifferentially extracted samples.*

Sample	Complete*	Partial†	No Result‡	No Signal§	Extra‖	Total
349	34					34
350	33		1	1		35
351	32		1	1		34
352	31			4		35
353	35					35
354	35					35
A	23	20	1	1		45
B	36	8			1	45
C	27	17				44
D	42	3				45
E	8	36			1	45
MSS1 Total	336	84	3	7	2	432
F	41		1			42
G**	41		1			42
J	40	2¶				42
K**	42					42
MSS2 Total	164	2	2	0	0	168

* Exact specification plus profiles with extra possible alleles clearly defined as weak (stutter) bands.
† At least one allele not specified for at least one locus.
‡ No result reported for one locus.
§ No result reported for any locus.
‖ One excess allele reported without comment for one locus of profile.
¶ Minor band of three-banded pattern not reported.
** Approximately 15% of the MSS2 participants differentially extracted these samples; only male fraction profiles were reported.

TABLE 6—*Profiling performance for differentially extracted samples.*

Sample	Exact*	Extra†	Partial‡	No Result§	No Signal‖	Total
H_{Sperm}	30	5	5	1	1	42
L_{Sperm}	37	3		2		42
$H_{Nonsperm}$	3	35		2	2	42
$L_{Nonsperm}$	4	31	3	1	3	42

* Exact specification.
† All true alleles specified plus one or more alleles from incompletely differentiated source.
‡ At least one true allele not specified for at least one locus.
§ No result reported for one or more locus.
‖ No result reported for any locus.

TABLE 7—*Identification of known sources in MSS1 multiple source samples.*

Sample	Complete*	Partial†	Missing Reference‡	Inconclusive§	Total
A	19	1(♀$_{349}$)		2	22
B	20			2	22
C	18		1(♂$_{352}$)	2	22
D	20			2	22
E	14	5 1(♀$_{349}$) 4(♂$_{352}$)	1(♂$_{352}$)	2	22

* All sources correctly identified.
† All specified sources correct; listed source not specified due to weak or missing alleles in multiple-source sample.
‡ One or more source(s) not specified due to absence of signal from reference sample(s).
§ No source identification attempted due to unbalanced peak heights in multiple-source samples or no signal obtained from three of the six reference samples.

participant and NIST clerical errors. Table 6 likewise summarizes performance for the sperm and nonsperm components of the two samples requiring differential extraction (i.e., the multiple source samples of MSS2).

A number of MSS1 participants reported multiple profiles. While some are replicate assays by different analysts of the same laboratory, a number of participants did not obtain sufficient signal for one or more STR loci with their initial extractions. All but two of these participants obtained sufficient signal after reamplification, reextraction of their original samples, or complete reanalysis of a second set of MSS1 samples.

Only true alleles were reported for all single source samples. A number of participants in both studies reported extra alleles but appropriately labeled them as "stutter" or weak minor-component alleles. Two participants reported a "stutter" allele for one or another of the multiple-source samples without explicitly noting that the peak was of relatively low intensity or otherwise unlikely to be a true allele. Two participants in MSS2 did not report a true minor allele of a known three-banded pattern. About half of the MSS1 participants did not explicitly report all alleles at all loci for at least one of the two-source samples and most did not report all alleles for the three-source sample.

The majority of participants reported only male-source alleles for the sperm-fraction of the differential extracts of the MSS2 multiple-source samples. Several participants reported minor alleles attributable to the female source for one or both samples. Several participants failed to observe some true alleles in the sample having two male-sources. Nearly all participants reported male-source alleles in the nonsperm fractions of both multiple-source samples. (We attribute the strong male-source signal in the nonsperm fraction to sperm lysis prior to sample preparation.) A number of participants noted that male-source contamination of the nonsperm fraction is not typical of casework samples, with further note that casework sperm-fractions are seldom as free of female-source contamination as the MSS2 samples.

The reported signals ranged continuously from essentially all above baseline events to just major component electropherogram peaks or gel-image bands. The majority of participants who did report some "probably stutter" signals used a variety of methods to indicate relative intensities: listing order, nested parentheses, and/or a wide selection of footnotes.

Identification of Known Sources

Table 7 summarizes source identification results for the five multiple-source MSS1 samples. All identified sources were true contributors. Due to incomplete profiling of the multiple-source samples, one participant chose not to specify one of the sources in one two-source sample and five participants chose not to specify one source in the three-source sample. Due to a single incomplete reference profile, one participant chose not to specify the unrecognized source in two of the samples although all other samples were successfully excluded as possible sources. Two participants chose not to attempt source identification, one due to multiple incomplete reference profiles and one due to inexperience with casework samples.

Most of the MSS1 profiling difficulties were encountered with the only male source of the six used to prepare the samples. Since many participants reported difficulty in obtaining good signal for this source's reference stain, it is probable that less than the target amount of this DNA was actually delivered to the stains involved.

A number of participants who identified all sources in the three-source sample noted that the three nonexcludable source profiles did not fully account for all features of the multiple source profile. Many of the differences between the expected and the observed multiple-source profiles are attributable to incomplete profiles for the multiple-source samples. However, several participants noted among-loci inconsistencies in the relative intensities of the allelic signals.

Specification of Unknown Sources

MSS2 participants were asked to provide "CODIS profile(s) to search for the suspect(s)" for the two multiple-source samples. The responses to this request were quite diverse, with many participants stating that they had little or no experience with CODIS. We have grouped the responses into the categories summarized in Table 8. These categories do not necessarily reflect CODIS nomenclature or practice.

About 25% of the participants chose not to profile the unknown male source in either sample, explaining that: 1) their laboratory did not perform this type of analysis; 2) it was against their laboratory's policy to profile a source in the absence of a reference; or 3) they were inexperienced in this type of analysis. A number of participants who did specify a profile for one or both unknown sources also noted that it was against their laboratory's policy to perform this type of analysis on casework samples.

All participants who profiled the unknown male in the two-source (known female, unknown male) sample specified all true

TABLE 8—*Specification of unknown source in multiple source samples.*

Sample	Match Stringency			Partial§	Ambiguous‖	Not Attempted¶	Total
	High*	Medium†	Low‡				
$H_{Unknown}$	4	11	8	4	5	10	42
$L_{Unknown}$	29	1			1	11	42

* Exact specification.
† All true alleles specified plus one-to-several alleles from other sources.
‡ All true alleles specified plus many-to-all alleles from other sources.
§ One or more true alleles not specified.
‖ No profile specified because of inconsistent data or too complex given laboratory experience.
¶ No profile specified by laboratory policy.

alleles; all but one participant specified only true alleles. The one participant who did not specify the exact profile for the unknown source explicitly identified all alleles unique to the unknown-source and implicitly identified all possible alleles for the loci with less than two unique alleles. One participant did not achieve a sufficiently good signal for the nonsperm fraction of the sample to confirm that the reference female was a true source and chose not to proceed with the analysis.

Specification of the unknown male in the three-source (known female, known male, unknown male) sample proved more problematic. Only four participants provided the exact profile. Eleven participants specified exact alleles at most loci and narrowly defined the possible allele combinations at the other loci. Eight participants chose to specify most-to-all signals observed in the sperm-fraction. Five participants found the problem too complex and chose not to proceed with the analysis. Several participants who specified large numbers of possible multiple allele pairs for one or more loci did not exhaustively specify all of the possibilities implied by the pairs that were specified.

Four participants did not specify all true alleles for the unknown profile. Two heterozygous loci were specified as homozygous and one homozygous locus was explicitly specified as heterozygous. None of these miss-specifications is attributable to unusual peak shape or intensities. Both alleles at one heterozygous locus were mis-specified due to stochastic "dropout" of one of the two alleles contributed by the unknown male. This event occurred in a sperm-fraction characterized by generally low-intensity signals at a locus with four male-source alleles. While present and not a possible "stutter" peak, the height of the "missing" allele peak was less than 25% of that expected and was very similar to the heights of true "stutter" at other loci.

A number of participants who specified a less than exact profile for the three-source sample, and at least one of the participants who chose not to attempt specification, noted among-loci inconsistencies in the relative intensities of the allelic signals. These inconsistent allelic signal intensity ratios for the different loci, also noted by a few MSS1 participants, were not related to any specific STR multiplex or manufacturer.

DNA Concentration Estimates

Figure 1 presents both the consensus and individual participant results reported for the MSS2 Set 3 extracted-DNA samples. The DNA concentrations in this four-level series ranged from 5.0 ng DNA per μL solution to 0.5 ng/μL. The results for all levels are well described as lognormal distributions. Given both the 10-fold span of concentration among the samples and the lognormal distribution of results at each level, all calculations have been performed on logarithmically transformed ($Y = \log_{10}(X)$) concentrations. All summary statistics have been back-transformed ($X = 10^Y$) to report the results in units of concentration. Measures of location, like the median, are qualitatively unchanged by this manipulation. However, measures of dispersion, like the robust standard deviation (SD), change from symmetrically additive terms to symmetrically multiplicative factors. That is, about 68% of normally distributed concentrations are expected to be in the interval (median $-$ SD $\leq$ median $\leq$ median $+$ SD); 68% of lognormally distributed concentrations are expected to be within the interval (median / SD $\leq$ median $\leq$ median $\times$ SD) (15).

There is excellent agreement between the nominal concentration (what we believe went into the tubes) and the median of the estimated concentrations (a robust consensus estimate of what came out (16)) for the highest three levels. At the lowest level of the series, the median is about 50% of the nominal; this difference may be due to DNA binding to the sample tube.

The concentration series reported by individual participants are typically collinear with the consensus values (i.e., straight lines with different intercepts but generally unit slope). The majority of participants reported concentrations clustered close to the consensus values, with a robust estimate for the SD being a factor of 1.8 for all four levels (16). Three participants reported quite similar values that are about five-fold higher than consensus, four reported values 4- to 40-fold lower than consensus for two or more levels, and three reported both very high and very low values. We use the term "concordance" to characterize the average difference of measurements from the consensus values; those measurement series that are consistently higher or lower than the consensus values are thus very positively or negatively discordant (17,18). We use the term "apparent precision" to characterize the SD among the differences; those measurement series that are inconsistently higher and lower than the consensus values are thus very "apparently imprecise" (18).

Figure 2 is a Youden plot (19) detailing all results reported for the duplicate samples. As with the Fig. 1 concentration series, participants who reported one very high or very low value tend to be consistently high or low. There is, however, little correlation between duplicate results for the concordant majority of participants. The robust SD for these samples is the same factor of 1.8. Since these two samples are true independently analyzed duplicates, this factor of 1.8 SD represents the intrinsic precision of the DNA concentration measurements. That is, among-participant apparent precision is as good as possible given the current measurement technologies.

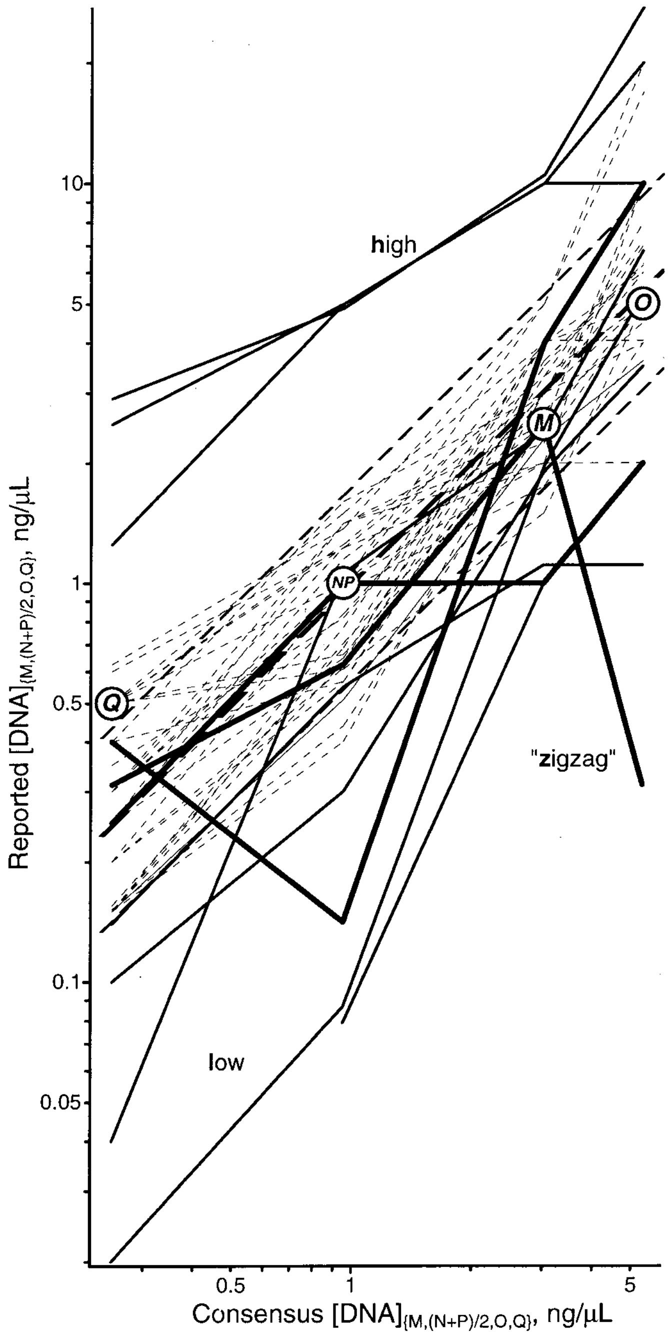

FIG. 1—*Measurements for MSS2 Set 3 DNA Concentration Series. Measurements reported by all participants are displayed as functions of the consensus median at each of four DNA concentration levels. The open circles denote the nominal concentrations for the five samples. The dashed lines at 45° represent the median response and the one SD factor interval about the median (median/1.8 ≤ median ≤ median × 1.8). The light dashed lines denote participants who reported values mostly within this one SD factor interval. The heavy solid lines (labeled "zigzag") denote participants who reported values irregularly much higher and much lower than the median. The solid lines above and below the one SD factor interval denote participants who reported values consistently much higher (high) or much lower (low) than the median.*

Figure 3 is a "target" plot (18) displaying the concordance and apparent precision characteristics of all MSS2 Set 3 sample measurements for each participant. The innermost "ring" of the target represents a combined discordance and apparent imprecision of one SD (here, a factor of 1.8) from complete agreement with the consensus values. The middle ring likewise represents the two SD

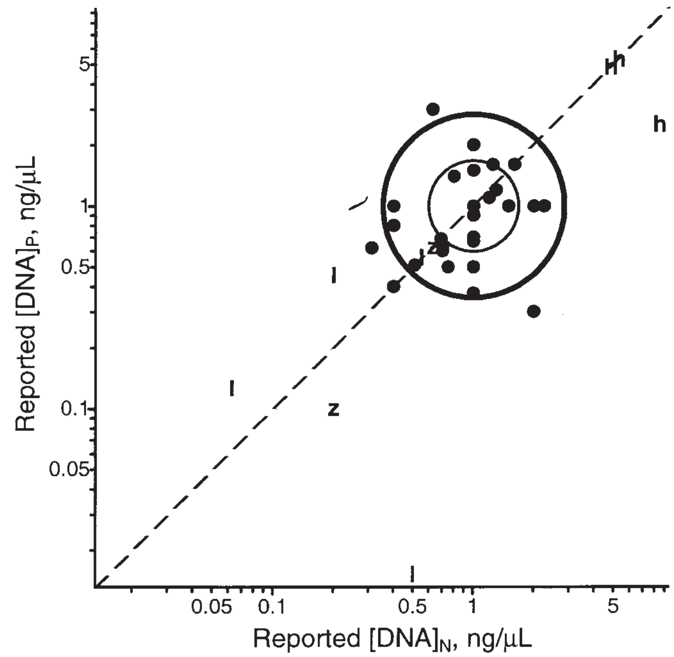

FIG. 2—*Measurements for MSS2 Set 3 Concentration Duplicates. Measurement pairs for the duplicate 1.0 ng/μL samples, N and P, are displayed for all participants: solid circles denote participants whose measurements were in good concordance with the median over the entire MSS2 Set 3 concentration series, "h" denote participants whose measurements were consistently much higher than the median, "l" denotes participants consistently much lower, and "z" denotes participants irregularly higher and lower. The dashed line at 45° represents equality. The inner solid circle denotes the one SD factor interval about the joint median, the outer circle denotes the two SD factor interval.*

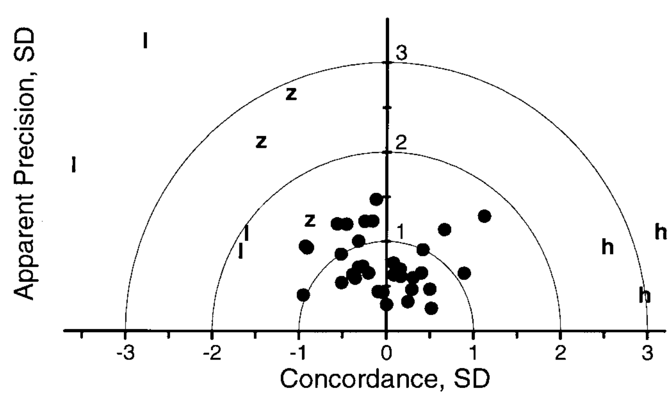

FIG. 3—*Measurement Characteristics of MSS2 Set 3 DNA Concentration Measurements Concordance (horizontal axis) and apparent precision (vertical axis) measurement characteristics are displayed for all participants, using the same symbol legend as in Fig. 2. The inner ring encloses a combined concordance and apparent precision of a one SD factor about the consensus medians, the middle ring encloses a two SD factor, and the outer ring encloses a three SD factor. Approximately 95% of all participants with measurement characteristics qualitatively similar to the consensus should plot within the middle ring.*

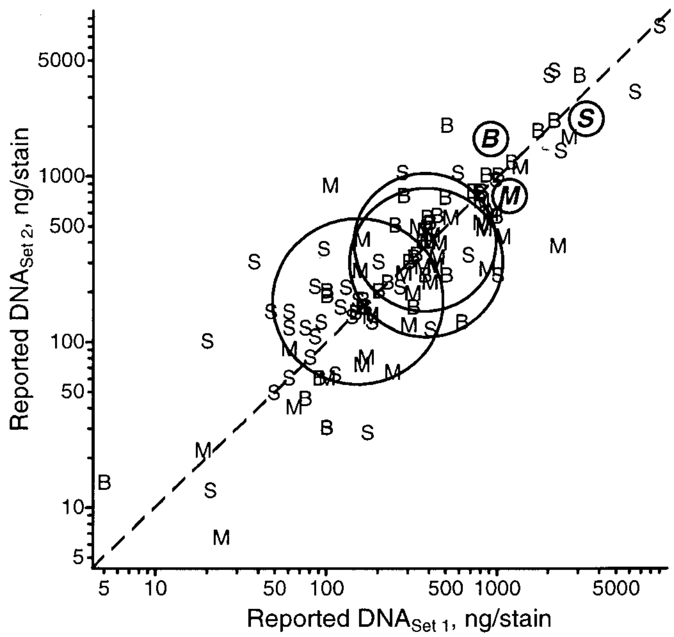

FIG. 4—*DNA quantity estimates for MSS2 Set 1 and Set 2 Samples. DNA quantity estimate pairs for the blood (F Vs J), semen (G Vs K), and multiple source (H Vs L) stains are displayed for all participants. Blood stain results are denoted "B," semen stains are denoted "S," and multiple source stains are denoted "M." The open circles denote nominal quantities for the three pairs. The dashed line at 45° represents equality. The circles enclose one SD factors for the three different types of stain.*

($1.8^2 = 3.2$) and the outermost ring the three SD ($1.8^3 = 5.8$) disagreement factors. Roughly, 95% of all participants with measurement characteristics "similar to the consensus" should plot within the two SD ring.

Total DNA Quantity Estimates

The total recoverable DNA quantities reported for the six MSS2 stains are presented in Fig. 4, a Youden variant plotting results for the blood, semen, and multiple-source stains of Set 1 against their analogues in Set 2. While different extraction protocols were typically used for blood and multiple-source stains (most, but not all, participants extracted the semen stains in the same manner as they did the blood stains), all participants used the same protocols for the Set 2 samples as they did for the Set 1 samples. Since all samples were prepared, randomized, and packaged independently, the strong correlation between the Set 1 and Set 2 estimated DNA quantities is a function of the participants and not of sample preparation.

The quantity estimates for the blood stains averaged about 30% of the nominal amount that we believe was actually present; the semen stains averaged about 20% of nominal; and the mixed-source stains about 10%. Three participants re-extracted the multiple-source stain matrix (cotton cloth) after differential extraction and estimated the residual quantity of DNA. All three reported about as much DNA in the residual as in the combined sperm and nonsperm fractions.

The variation in the estimated quantity of DNA in the stains, ranging from an average factor of 2.6 for the blood reference samples to a factor of 3.2 for the multiple-source stains, is much larger than the factor of 1.8 characteristic of DNA concentration measurements. Some of this variation may be attributed to stain subsampling. A few

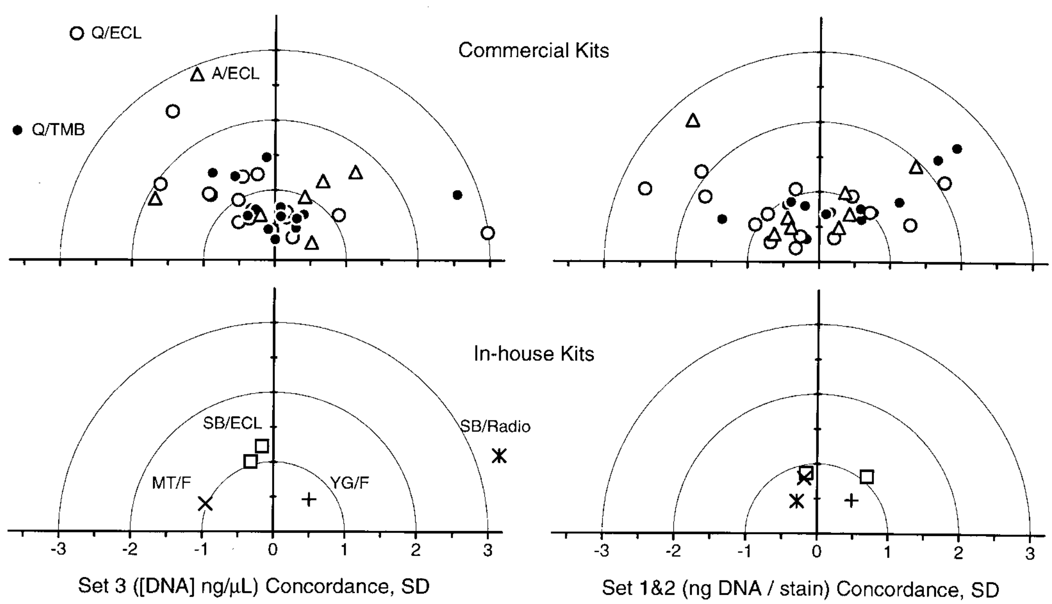

FIG. 5—*Measurement characteristics different DNA quantitation techniques. Concordance and apparent precision characteristics are displayed for MSS2 Set 3 DNA concentration measurements (the two left-hand plots) and for Set 1 and 2 DNA quantity estimates (the two right-hand plots). The upper two plots present the measurement characteristics of all participants who used one of three types of commercial DNA quantitation kit: solid circles denote QuantiBlot® (PE Applied Biosystems, Inc., Foster City, CA) with colorimetric detection, open circles denote QuantiBlot® with chemiluminescence detection, and open triangles denote ACES™ (Life Technologies, Gaithersburg, MD) with chemiluminescence detection. The two lower plots present the characteristics of all participants who used noncommercial techniques: open squares denote in-house slot blot assay with chemiluminescence detection, "*" denotes in-house slot blot with radiographic detection, "x" denotes microtiter plate with fluorescence detection, and "+" denotes yield gel with fluorescence detection.*

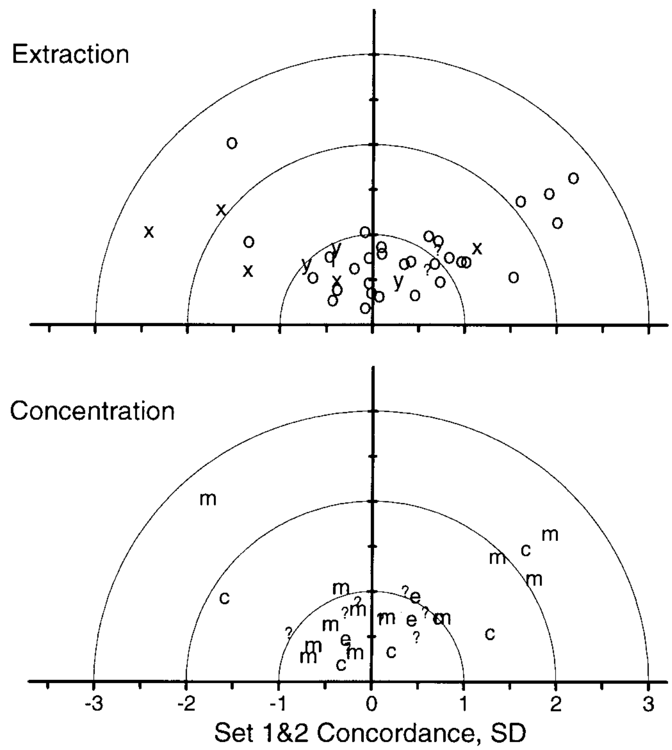

FIG. 6—*Measurement characteristics of different DNA extraction and concentration techniques. The upper plot presents concordance and apparent precision characteristics for MSS2 Sets 1 and 2 DNA quantity estimates for participants as a function of the DNA extraction technique used: "o" denotes participants who used an organic extraction protocol for all samples, "q" denotes use of QIAamp® spin columns (Qiagen Inc., Valencia, CA), "x" denotes use of a Chelex® protocol, and "y" denotes participants who used an organic protocol for some samples and a Chelex protocol for others. The lower plot presents the measurement characteristics as a function of DNA concentration technique used: "e" denotes ethanol precipitation, "c" denotes use of Centricon® filters (Millipore Corp., Bedford, MA), and "m" denotes Microcon® filters (Millipore Corp., Bedford, MA). In both plots, "?" denotes participants who did not specify their extraction and/or concentration techniques.*

participants directly estimated DNA quantities from the extracts of entire stains. However, most participants extracted a fraction of the stain and adjusted their estimate by the relative area of the subsample to the entire stain. Unless care is taken to proportionally sample all regions of the stain, this adjustment is valid only if the DNA is uniformly distributed over the entire stain. Given that leukocytes, intact sperm, and free DNA doubtless have different affinities for the cloth matrix, they may differentially concentrate at the stain center or the stain edge. Participants who subsampled from the center of each stain may well have sampled less representatively than those who divided the stain into halves or quarters.

Comparison of DNA Measurement Methods

The majority of MSS2 participants used one of two commercial human DNA quantitation kits with either colorimetric or chemiluminescence detection. The top section of Fig. 5 displays the concordance and apparent precision characteristics for these participants for both the Set 3 DNA concentration and the Set 1 and 2 DNA quantity measurements. There is no clear measurement-performance difference between the kits or between the detection methods.

The lower section of Fig. 5 likewise displays the concordance and apparent precision characteristics for the participants that do not use a commercial DNA quantification kit. While the Set 3 results for one participant using an in-house method are extremely discordant, they are very similar to the results reported by two participants using commercial kits. Since the Sets 1 and 2 results for this participant are quite concordant, we believe that the extreme Set 3 discordance is analyst—rather than method—related. There is no other clear measurement-performance difference among the in-house methods or between the in-house and commercial methods.

Comparison of DNA Extraction and Concentration Methods

Figure 6 displays the concordance and apparent precision characteristics for the Sets 1 and 2 total DNA quantity estimates as a function of extraction and concentration methods. Most of the participants who systematically reported very much less than the consensus DNA quantities used Chelex extraction with all samples. However, not all participants using a Chelex protocol were negatively discordant. There are no other clear performance differences related to extraction or concentration methods.

Influence of Measurement Discordance on STR Profiles

Thirty-seven participants reported both Set 3 DNA concentrations and Sets 1 and 2 DNA quantities. Figure 7 displays the characteristic concordances for these participants, again as a Youden variant plotting the Sets 1 and 2 total DNA quantity concordance as a function of the Set 3 DNA concentration concordance. There is no systematic relationship between the two measurement characteristics. However, the one participant with highly positive con-

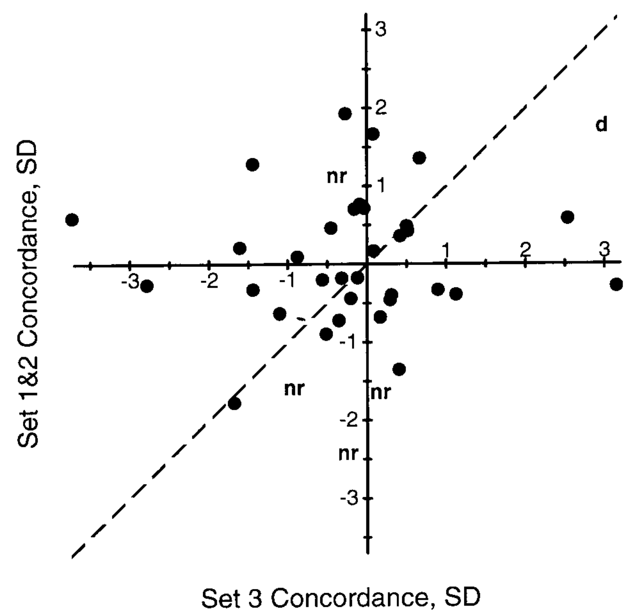

FIG. 7—*Comparison of DNA concentration and quantity concordances. The MSS2 Set 1 and 2 DNA quantity concordance is displayed as a function for the MSS2 Set 3 DNA concentration concordance for all participants: solid circles denote participants who reported profiles for all loci for all samples, "d" denotes the participant who experienced allele dropout at one locus of a multiple-source sample, "nr" denotes participants who did not report alleles at one or more loci for at least one sample, and "p" denotes participants who reported partial profiles for the unknown male source of the three-source stain.*

centration and quantity discordance is the only laboratory experiencing allele dropout. This participant's consistent five-fold overestimation of available DNA may have resulted in too-little DNA being used in the amplification reaction mixture.

Most of the participants who reported amplification failure at one or more loci reported much less than the consensus quantity of DNA in the Set 1 and Set 2 samples. If these highly negative quantity discordances truly represent very inefficient DNA extractions, the failures may reflect too little DNA being available in the amplification reaction. Alternatively, if the low quantitation results are erroneous and the DNA extractions were "typically" efficient, the failures may reflect too much DNA being present in the initial reaction mixture for optimal amplification.

Discussion and Recommendations

The MSS1 and MSS2 interlaboratory comparisons challenged multiplex STR measurement systems with intentionally difficult samples. About 10% of the participants in both studies experienced measurement failures, misinterpreted some part of their analytical signals, or made clerical errors. Most of these problems occurred with the most challenging samples of each exercise—those containing DNA from three sources. Furthermore, few participants specified unique profiles for all three components of these complex samples.

While none of the measurement anomalies in either exercise resulted in a false identification of a reference source, measurement difficulties waste laboratory resources. While few of the anomalies reported could result in a false exclusion if they occurred in casework, database matching is most efficient when the target profile is uniquely defined and complete. Thus, while the amplification and detection stages of STR multiplex DNA profiling appear robust, we believe that the efficiency of the overall measurement systems can be further improved. The following recommendations address the potential measurement difficulties suggested by the MSS1 and MSS2 results.

Clerical Errors

The number of clerical errors reported in the MSS1 and MSS2 challenge studies is larger than we expect for "routine" information transfers. However, by definition, any clerical error represents a failure (or circumvention) of the laboratory's internal technical review process. Clerical errors in specialized reports prepared for interlaboratory comparisons should be regarded as an early warning sign of a laboratory's need to review its review process (or the need to apply it to all external reports).

Transcription errors can be minimized by direct electronic transfer of single source profiles from instrument through local database to final report. This can be best achieved by insisting that such linkages among information systems be a design goal for all forensic software.

Independent data analysis by two or more analysts followed by rigorous comparison and immediate resolution of differences although time consuming can minimize interpretation as well as transcription errors. Use of standardized report templates for casework scenarios would minimize the clerical perils intrinsic to the production of "once only" reports. Rigorous internal technical review of all data, interpretation, and final reports as is done routinely for casework will minimize the number of errors.

Turnabout—Clerical errors by interlaboratory study providers must also be minimized, using the same tools as above. To better approach the desired zero error rate, the NIST analyst-authors recently switched from review of each other's results to completely independent analysis of the raw data.

Whenever information must be manually transcribed, transformed, and/or interpreted before analysis, it is essential that the analysts responsible for the original data have the opportunity to review their modified results.

Terminology and Communication Issues

To enable assessment of current forensic STR reporting practice, the MSS1 and MSS2 exercises did not specify that the requested information be provided in any particular format. While all reports were quite interpretable, given the solid experimental background of most of the authors, the terminology and style of reports were unexpectedly diverse. This diversity includes descriptions of analytical techniques, materials and equipment, differential extraction fractions, stutter and minor-component peaks, homozygous loci, and nonuniquely determined profiles.

Information exchange among all members of the forensic community (especially those without a strong technical background) would be facilitated if the forensic community adopted consensus terms and symbols. Consensus development of report templates appropriate for the different casework scenarios could also facilitate communication and potentially lower the clerical error rate for less-common situations.

Turnabout—Development of consensus standards and report templates can only be accomplished through collaborative effort of the (already overburdened) forensic analysts directly involved. Should the community start such a project, interlaboratory studies could be used to help evaluate and refine the proposed communication tools.

Extraction, Quantitation, and Amplification Issues

Several participants did not obtain complete amplification of all components of all samples. These analytical failures are almost all associated with difficulties in preamplification stages of the STR measurement process: sample extraction, DNA concentration, and DNA quantitation. With one exception, the MSS2 study results indicate that these difficulties are not attributable to specific techniques or methods but rather reflect possible variability within the quantitation systems and/or analyst experience and ability. Chelex extraction systems appear to be less efficient for mixed-stains than the organic systems.

The participants' estimates of DNA quantity were quite discordant for many of the MSS2 stains. However, nearly all participants reported the same relative discordance for the three {Set 1, Set 2} pairs of stains: blood, semen, and mixed. The between-set uniformity of within-participant results suggests large and consistent among-participant differences in DNA recovery (extraction and/or concentration) efficiency. No DNA extraction or concentration method was consistently associated with either higher or lower results. The few participants who re-extracted the mixed stains after differential extraction reported significant amounts of DNA in the second extract (20). We speculate that some of the recovery efficiency differences are related to sample digestion conditions, particularly agitation ("poke it with a toothpick") and duration ("long enough"). Since nearly all participants who reported total or multiple-locus amplification failure for any sample also reported very low DNA recovery, analysts experiencing amplification failures with pristine samples (such as used in MSS1 and MSS2) should re-evaluate their sample preparation techniques.

Some of the recovery differences can also be attributed to how the stains were subsampled. At least for interlaboratory comparisons, stains should be divided into roughly equal "pie slices." This will help ensure that all pieces of the sample contain about the same quantity and distribution of tissues; it will also greatly simplify estimating the quantity of DNA/stain.

The consensus DNA concentrations for the MSS2 Set 3 samples are nearly identical to the nominal values for the 1 nL, 2.5 nL, and 5 ng/μL samples. No DNA quantitation method consistently produced results very different from the consensus result. The SDs of the intra- and all interlaboratory DNA quantitation results are of nearly identical magnitude, a factor of 1.8 about the consensus results. Thus, all of the current quantitation methods appear to be capable of providing very comparable and accurate results, with expected bias of zero and precision of about a factor of two. The close agreement between the intra- and interlaboratory SD factors suggests that this 2-fold imprecision is intrinsic to the process and cannot be easily improved.

While the majority were in remarkable agreement, more than 10% of participants consistently reported DNA concentration results much higher or much lower than consensus. Few of these participants reported similarly discordant Sets 1 and 2 DNA quantities. Many of the anomalous Set 3 results may not indicate routinely biased DNA quantitation but rather pipetting or dilution errors specific to these unfamiliar samples. However, the only serious amplification anomaly reported in either MSS1 or MSS2 was experienced by a participant who consistently reported 4 to 10 fold more DNA than consensus. We believe that this resulted in too little DNA being added to the reaction mixture for reliable amplification of all alleles from all sources.

Turnabout—There are no control or reference materials currently available for documenting forensic DNA quantitation performance or evaluating DNA extraction and recovery efficiencies. Since these pre-amplification measurements impact the overall performance of STR measurement systems, this lack should be rectified.

The consensus result for the nominal 0.5 ng/μL MSS2 Set 3 quantitation sample is 2-fold low. This suggests a sample preparation or storage problem for low-DNA-concentration samples that must be solved before producing any further DNA quantitation materials.

Several participants noted that the MSS2 multiple-source samples provided "cleaner" sperm fractions and "less clean" nonsperm fractions than typical of casework. We speculate that the lack of female-source contamination of the sperm fraction is due to our use of fresh whole blood rather than epithelial cells. We speculate that the male-source contamination of the nonsperm fraction is due to our use of over-age fertility-clinic semen. Since the nonsperm fraction profile primarily provides confirmation of the female-source, male-source contamination of this profile may be unrealistic but of little forensic consequence.

The MSS2 established that DNA extraction efficiency and quantitation accuracy do at least qualitatively (signal/no signal) affect "real life" STR measurement systems, even with pristine samples. There should be quantitative relationships among all stages of the measurement process from the DNA amount to the allelic signal intensities; it should be possible to design a study to evaluate these linkages.

Policy Issues

As noted above, there was considerable variation in *how* stutter peaks were described. More importantly, there was considerable variation in *whether* stutter peaks were described. While stutter is undesirable, it is frequently observed and may not be avoidable. We believe that peaks unambiguously attributable to stutter do not require notation in the final report. If the interpretation is ambiguous and stutter cannot be excluded as the source of a peak (that is, the peak could be either a true allele or a stutter peak) that peak should be reported, if reported at all, only as a *potential* allele. A notation that distinguishes between true and potential alleles would facilitate communication and help ensure consideration of all relevant possibilities. In any case, we believe some consensus policy on the evaluation and reporting of stutter would benefit the entire forensic community.

Again as noted above, we ascribe most observed differential allele amplification to inefficient DNA extraction and inaccurate quantitation (competitive kinetics with too much DNA in the reaction mixture, stochastic dropout with too little). However, we occasionally encounter profiles having unbalanced allelic signals at some loci, up to and including complete absence of one ("null") allele. While of little importance when evaluating multiple-source profiles against reference profiles, it could well complicate specification of an "unknown" profile. When multiple interpretations cannot be excluded, "moderate stringency" profiles—unique where possible and explicitly describing the ambiguities where required—should be specified.

The *raison d'être* of any forensic DNA database system is identification of leads to perpetrator identity from evidence containing DNA that is not accounted for otherwise. Given the utility of such information, a surprisingly large fraction of MSS2 participants noted that specifying a profile in the absence of a reference profile was against their laboratory's policy. To ensure that such evidence can be recognized and efficiently exploited when needed, the evaluation of unknown-source profiles and use of DNA database match systems should become a routine component of forensic training and competency evaluation.

Turnabout—Complete and unique specification of an "unknown" profile will not always be possible. A comparative evaluation of how current DNA database matching systems process incomplete and ambiguous profiles could enable more effective use of these forensic tools.

Conclusion

The MSS1 and MSS2 interlaboratory comparisons challenged multiplex STR measurement systems with difficult samples representing unusual forensic scenarios. These studies were explicitly designed to elicit measurement problems related to unbalanced amplification of DNA from multiple-source samples. None of the relatively few analysis problems encountered can be attributed to abnormal STR multiplex performance; all DNA amplification anomalies reported are associated with inefficient DNA extraction, inaccurate DNA quantitation, and/or analytical threshold policies. Given an appropriate total amount of DNA in the reaction mixture, current STR multiplex systems reliably amplify multiple-source DNA.

Acknowledgments

Interlaboratory studies are made possible by the cooperation of many analysts and laboratory supervisors. We thank them for sharing with us their time and resources, and for their willingness to tackle our somewhat contrived samples. Their voluntary and open participation in challenge exercises speaks as much to pursuit of analytical truth as to confidence in analytical systems.

References

1. Bercovich D, Regev Z, Ratz T, Luder A, Plotsky Y, Gruenbaum Y. Quantitative ratio of primer pairs and annealing temperature affecting PCR products in duplex amplification. Biotechniques 1999; 27(4):762–70.
2. Henegariu O, Heerema NA, Dlouhy SR, Vance GH, Vogt PH. Multiplex PCR: critical parameters and step-by-step protocol. Biotechniques 1997;23(3):504–11.
3. Baechtel FS, Monson KL, Forsen GE, Budowle B, Kearney JJ. Tracking the violent criminal offender through DNA typing profiles—a national database system concept. EXS 1991;58:356–60.
4. Werrett DJ. The National DNA Database. Forensic Sci Int 1997; 88:33–42.
5. Reeder DJ. Short tandem repeats: experiences with interlaboratory testing of mixed DNA stains. Presented at the DNA Forensics: Science, Evidence, and Future Prospects, Cambridge Healthtech Institute; 1997 Nov. 17–18; McLean (VA).
6. Kline MC, Redman JW, Duewer DL, Reeder DJ. Results from the 1999 NIST mixed-stain study #2: DNA quantification, differential extraction, and identification of the unknown contributors. Presented at the 10th International Symposium on Human Identification; 1999 Sept. 29–Oct. 2; Orlando (FL).
7. Kline MC, Duewer DL, Newall P, Redman JW, Reeder DJ. Interlaboratory evaluation of STR triplex CTT, including manual and automated methods: understanding the differences. J Forensic Sci 1997;42:897–906.
8. Waye JS, Presley LA, Budowle B, Shutler GG, Fourney RM. A simple and sensitive method for quantifying human genomic DNA in forensic specimen extracts. Biotechniques 1989;7(8):852–5.
9. Walsh PS, Metzger DA, Higuchi R. Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. Biotechniques 1991;10(4):506–13.
10. Budowle B, Baechtel FS. Modifications to improve the effectiveness of restriction fragment length polymorphism typing. Appl Theor Electrophor 1990;1(4):181–7.
11. Kirby LT. DNA fingerprinting: an introduction. New York: Stockton Press, 1990; 72.
12. Maniatis T, Fritsch EF, Sanbrook J. Molecular cloning: a laboratory manual. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory, 1982.
13. Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acid Research 1988;16:1215.
14. Budowle B. Studies for selecting core STR loci for CODIS. Presented at the DNA Forensics: Science, Evidence, and Future Prospects, Cambridge Healthtech Institute; 1997 Nov. 17–18; McLean (VA).
15. Sharpless KE, Duewer DL. Population distributions and intralaboratory reproducibility for fat-soluble vitamin-related compounds in human serum. Anal Chem 1995;67:4416–22.
16. Analytical Methods Committee, Royal Society of Chemistry. Robust statistics—how not to reject outliers. Part 1, Basic Concepts. Analyst 1989;114:1693–7.
17. 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.
18. Duewer DL, Kline MC, Sharpless KE, Brown Thomas J, Gary KT, Sowell AL. Micronutrients measurement quality assurance program: helping participants use interlaboratory comparison exercise results to improve their long-term measurement performance. Anal Chem 1999;71:1870–8.
19. Youden WJ. Graphical diagnosis of interlaboratory test results. Ind Quality Control 1959;15:1–5.
20. Baechtel FS, Presley KW, Smerick JB. D1S80 typing of DNA from simulated forensic specimens. J Forensic Sci 1995;40:536–45.

Additional information and reprint requests:
Margaret C. Kline
NIST
100 Bureau Drive, Stop 8311
Gaithersburg, MD 20899-8311
Margaret.Kline@nist.gov