J Forensic Sci, May 2005, Vol. 50, No. 3
Paper ID JFS2004357
Available online at: www.astm.org

TECHNICAL NOTE

Margaret C. Kline,[1] *M.S.; David L. Duewer,*[2] *Ph.D.; Janette W. Redman*[1] *; and John M. Butler,*[1] *Ph.D.*

Results from the NIST 2004 DNA Quantitation Study*

ABSTRACT: For optimal DNA short tandem repeat (STR) typing results, the DNA concentration ([DNA]) of the sample must be accurately determined prior to the polymerase chain reaction (PCR) amplification step in the typing process. In early 2004, the National Institute of Standards and Technology (NIST) conducted an interlaboratory study to help assess the accuracy of DNA quantitation in forensic DNA laboratories. This study was designed with four primary purposes: (1) to examine concentration effects and to probe performance at the lower DNA concentration levels that are frequently seen in forensic casework; (2) to examine consistency with various methodologies across multiple laboratories; (3) to examine single versus multiple source samples; and (4) to study DNA stability over time and through shipping in two types of storage tubes. Eight DNA samples of [DNA] from 0.05 ng/μL to 1.5 ng/μL were distributed. A total of 287 independent data sets were returned from 80 participants. Results were reported for 19 different DNA quantitation methodologies. Approximately 65% of the data were obtained using traditional slot blot hybridization methods; 21% were obtained using newly available quantitative real-time PCR (Q-PCR) techniques. Information from this interlaboratory study is guiding development of a future NIST Standard Reference Material for Human DNA Quantitation, SRM 2372.

KEYWORDS: forensic science, DNA typing, DNA quantitation, interlaboratory study, quantitative real-time PCR (Q-PCR)

The determination of the quantity of human DNA present in a sample prior to amplification with the polymerase chain reaction (PCR) is an important step in forensic DNA analysis using multiple short tandem repeat (STR) markers. There are both regulatory and measurement performance reasons for this importance. The U.S. Federal Bureau of Investigation's standards, which govern forensic DNA typing in the United States, require that a laboratory shall have and follow a procedure for evaluating the quantity of the human DNA in the sample when possible (1). More germanely, the best performance with multiplex PCR amplification of STR markers using commercially available kits occurs within a fairly narrow range of input DNA amount—typically on the order of 0.5 ng to 2.5 ng (2–5). Too much DNA can lead to off-scale fluorescent signal and a variety of PCR artifacts, including: imbalance in the multiplex PCR amplification, incomplete adenylation of PCR products, and enhanced strand-slippage or "stutter" of various forms (6,7). Too little DNA can result in stochastic amplification that causes imbalance within and between loci and even allele dropout (8,9).

Over the past decade, our group at NIST has been involved with a series of interlaboratory studies to help forensic DNA typing laboratories assess their performance with various technologies and issues relevant to forensic testing (10–17). The most recent of these studies indicated that the accuracy of DNA quantitation does impact the quality of STR typing, particularly when examining mixture samples (16,17). In order to better understand the measurement performance of different DNA quantitation techniques, we initiated the NIST 2004 DNA Quantitation Study (QS04). Results from this study enable evaluation of the relative precision and bias of current DNA quantitation methods at low ng/μL levels. QS04 also provided information on the stability of DNA samples under standard storage and shipping conditions that is guiding the development of a human DNA quantitation Standard Reference Material, SRM 2372.

Materials and Methods

DNA Samples

The QS04 sample materials consisted of eight extracted DNA samples, 100 μL each, in 10 mmol/L Tris, 0.1 mmol/L EDTA buffer (TE^{-4}). The samples were labeled "A" through "H". Samples A, B, C, and D were dilutions of a pooled lyophilized human DNA product D 7011 (Sigma, St. Louis, MO). A stock solution was made by reconstituting 5.06 mg of the lyophilized DNA in 50 mL of TE^{-4} and allowing it to equilibrate for 11 weeks in a Teflon container at 4°C. The final DNA concentration ([DNA]) of the stock was determined by UV absorbance at 260 nm on a Cary 100

[1] Biotechnology Division, National Institute of Standards and Technology, Gaithersburg, MD 20899.
[2] Analytical Chemistry Division, National Institute of Standards and Technology, Gaithersburg, MD 20899.

† Parts of this work were presented at the 2004 STR Megaplex and Research Technology Workshop (Virginia Beach, VA), March 30, 2004 and at the NIJ DNA Grantees Meeting (Washington, DC), June 30, 2004.

‡ National Institute of Justice (NIJ) funded this work in part through an interagency agreement with the NIST Office of Law Enforcement Standards.

Received 2 Sept. 2004; and in revised form 6 Nov. 2004; accepted 6 Nov. 2004; published 6 April 2005.

double-beam spectrophotometer (Varian Analytical Instruments, Walnut Creek, CA) to be 55.9 ng/μL (18) with measurement repeatability (6 sets of measurements over 11 weeks) of ≈1%. Quantitative volumetric dilutions of this stock with TE^{-4} provided solutions with nominal [DNA] of: A – 1.5 ng/μL, B – 0.5 ng/μL, C – 0.16 ng/μL, and D – 0.05 ng/μL. Aliquots of 100 μL were placed in appropriately labeled; sterile polypropylene limited volume micro tubes (SARSTEDT Inc., Newton, NC).

Samples E, F, G, and H were dilutions of a single-source anonymous male DNA obtained from whole blood collected in EDTA tubes (Becton Dickinson VACUTAINER Systems, Franklin Lakes, NJ) and extracted by a modified "salt out" procedure (19). The final [DNA] of the material was similarly determined using UV absorbance and determined to be 44.6 ng/μL with measurement repeatability (4 sets of measurements over 4 weeks) of ≈1%. Quantitative volumetric dilutions of this single source material with TE^{-4} provided nominal [DNA] of: E – 0.5 ng/μL, F – 0.16 ng/μL, G – 0.05 ng/μL, and H – 0.05 ng/μL. Samples E, F, and G were prepared as 100 μL aliquots in appropriately labeled, sterile polypropylene limited volume micro tubes (SARSTEDT Inc.). Sample H was prepared as 100 μL aliquots in UV-irradiated Teflon PFA autosampler vials with threaded caps (Savillex Corp., Minnetonka, MN). All materials were stored at −80°C prior to shipment on dry ice.

The lowest [DNA] samples, 0.05 ng/μL, were designed to be within the reported limits of detection for slot blot measurements (20) when using 5 μL of sample.

Participants

Solicitation for participation in QS04 began in September 2003 at the 14th International Symposium on Human Identification held in Phoenix, AZ and continued through December 2003 by email to previous participants in NIST interlaboratory studies. Eighty-four organizations elected to receive samples; 80 of these organizations participated by returning one or more sets of results. Participants from the U.S. included 59 state and local crime laboratories from 37 different states and Puerto Rico, four federal laboratories (including two non-forensic groups within the National Institutes of Health), seven private companies, and two universities. There were eight participants from Austria, Canada, Germany, South Africa, and the United Kingdom.

The majority of the QS04 materials were shipped mid-December 2003 through mid-January 2004. Regardless of when the samples were received, the final date for accepting results was April 5, 2004. Participants were asked to report sets of quantitative results for the eight samples (A-H) using as many methods as available in their laboratories and from as many analysts as time, resources, and sample permitted. As data were returned, they were transcribed into a database. A copy of the transcribed data was sent to the participants to check the accuracy of data entry.

Data Analysis

All statistical analysis and graphical presentations of the QS04 data were accomplished using Microsoft Excel 2000 (Redmond, WA).

Results and Discussion

Eighty of the 84 organizations that received samples returned one or more sets of [DNA] results. One organization shared their samples among analysts at three separate facilities, giving a total of 82 participating laboratories. This high participation rate and large number of participants is an indication of the level of forensic interest in the characterization of DNA quantitation measurement technologies.

Quantitation Methods and Frequency of Use

A total of 287 sets of results were reported. Of this total, 27 were within-analyst replicates (multiple results from the same analyst using the same methodology). More than half of the 260 among-analyst sets were within-laboratory duplicates (multiple results from the same laboratory using the same methodology but obtained by different analysts). There were 118 sets of unique among-laboratory results (different methods from the same laboratory and nominally identical methods from different laboratories).

Table 1 summarizes the nature and frequency of use of the various quantitation methods that were used to obtain these results. About 65% of the total number of results were obtained by direct imaging of sample DNA using various modifications of slot blot hybridization techniques based on the multi-copy D17Z1 probe (20), including in-house variants of the discontinued ACES technology (21,22). An additional 14% of the results were obtained by other more-or-less direct-imaging techniques, about equally distributed among: the AluQuant kit (23) from Promega Corporation (Madison, WI), various forms of a Picogreen assay that is not specific for human DNA (24), and yield gels (18). About 21% of the results were obtained using quantitative real-time PCR (Q-PCR) assays (including several not-yet publicly documented variants) that monitor the number of amplification cycles required to synthesize a given amount of DNA from a given template amount (25–31); more than half of these used the Quantifiler Human DNA kit released in November 2003 by Applied Biosystems (Foster City, CA) (25). One result set was obtained using an endpoint PCR method designed to indicate whether or not a sufficient amount of DNA is present to successfully amplify (32); results from this method are intrinsically qualitative for [DNA] below 1 ng/μL and so are not included in the following analyses.

Interpretation of Semi-Quantitative Data

Nearly 82% of the $(287\text{-}1) \times 8 = 2288$ QS04 quantitative data were reported as unambiguous concentrations (e.g., "1 ng/μL"). Slightly more than 11% were reported as "no result" indicating that no clear signal from DNA was observed for that sample. The residual 7% of the results were reported as semi-quantitative values that could not be directly used with routine data analysis tools. To include as many of the reported values as possible in the summary analyses, we have converted many of these semi-quantitative data to fully quantitative form (described below).

Six data were reported as the range between two contiguous factor-of-two calibration standards (e.g., (0.25 to 0.5) ng/μL). Given that a factor-of-two progression is inherently logarithmic, the geometric mean (the anti-log of the average of logarithmically-transformed values) of the range endpoints is the most appropriate mid-range value. Given the observed within-analyst variability (discussed below), conversion of these ranges does not distort the statistical distributions.

Two data were reported as lower limits (e.g., >2 ng/μL). Given that more than 75% of the fully quantitative values for the samples involved were less than the specified minimum value, assignment of any value equal to or greater than the limit has little or no influence on the results of the robust statistical methods used to

TABLE 1—*Quantitation Methods Used in QS04. Key for detection and instrument codes is listed below. There are 34 categories and 19 unique "codes" or quantitation methods examined.*

Class	Method	Principle	Signal	Instrument	Code	Labs	Set	Ref
Direct, non-blot	AluQuant	Luminescence	Probe	TLA	a	4	9	23
"	"	"	"	TD20/20	a	3	3	"
"	"	"	"	TR	a	1	1	"
"	Picogreen	Fluorescence	Picogreen	FA	p	2	3	24
"	"	"	"	LF320	p	1	1	"
"	"	"	"	PBC4000	p	1	1	"
"	"	"	"	PK_FL	p	1	2	"
"	"	"	"	SMAX	p	2	3	"
"	"	"	"	?	p	1	2	"
"	Yield Gel	"	EtBr	visual	y	3	14	18
Direct, blot	"ACES"	Luminescence	Solulink	visual	A	4	10	21
"	"	"	"	KI	A	1	4	"
"	QuantiBlot	"	ECL	visual	E	18	40	20
"	"	"	"	CCDBIO	E	2	11	"
"	"	"	"	KI	E	3	7	"
"	"	"	SSWD	visual	D	3	11	"
"	"	"	SSWF	visual	F	4	6	"
"	"	Colorimetry	TMB	visual	T	31	98	"
Endpoint PCR	BodeQuant	Fluorescence	Picogreen	CF4000	—	1	1	32
Real Time PCR	Quantifiler	Fluorescence	Probe	ABI7000	0	16	32	25
"	"	"	"	ABI7700	0	1	2	"
"	"	"	"	ABI7900	0	2	3	"
"	Alu Q-PCR	"	Sybr	RG3000	1	2	4	26
"	"	"	"	ABI7000	1	1	1	"
"	Alu_Sifis	"	"	i-Cycler	2	1	2	27
"	Alu_tqman	"	Probe	ABI7000	3	1	1	—
"	Aluprobe	"	"	ABI7700	4	1	1	—
"	BRCA 1	"	"	ABI7700	5	1	2	28
"	"	"	"	ABI7900	5	1	1	"
"	CFS-HUMRT	"	"	ABI7000	6	1	1	29
"	"	"	"	ABI7900	6	1	3	"
"	GB:L78833.1	"	"	i-Cycler	7	1	2	—
"	RB1	"	"	ABI7700	8	1	3	30
"	RTALU	"	"	ABI7000	9	1	2	31
						118	287	

Detection Codes

ECL	Enhanced Chemiluminescence (Amersham, Arlington Heights, IL)
EtBr	Ethidium bromide
Picogreen	Picogreen Reagent (Molecular Probes, Eugene, OR)
SSWD	Super Signal West Dura (Pierce Biotechnology Inc, Rockford, IL)
Sybr	SYBR Green I Reagent (Molecular Probes, Eugene, OR)
TMB	Chromgen:TMB 3,3′,5,5′- Tetramethylbenzidine (Applied Biosystems, Foster City, CA)
Solulink	ACES probes (Solulink Inc, San Diego, CA)
SSWF	Super Signal West Femto (Pierce Biotechnology Inc, Rockford, IL)

Instrument Codes

ABI7000	ABI Prism 7000 Sequence Detection System (Applied Biosystems, Foster City, CA)
ABI7700	ABI Prism 7700 Sequence Detection System (Applied Biosystems, Foster City, CA)
ABI7900	ABI Prism 7900 HT Sequence Detection System (Applied Biosystems, Foster City, CA)
CCDBIO	Hitachi CCD Bio Imager, (MiraiBio Inc., Alameda, CA)
CF4000	CytoFluor 4000 Multiwell Fluorescent Plate Reader (Applied Biosystems, Foster City, CA)
FA	Fluoroskan Ascent (Thermo Electron Corp., Waltham, MA)
i-Cycler	iCycler Opitical System (BIORAD, Hercules, CA)
KI	Kodak Imager 440, (Kodak, Rochester, NY)
LF320	Lambda Fluoro 320 Plus (MWG Biotech, Highpoint, NC)
PBC4000	Perspective Biosystems Cytofluor 4000 (Perspective Biosystems, Framington, MA)
PK_FL	Packard FluoroCount (Packard Instrument Co. Inc., Meriden, CT)
RG3000	Corbett Research Rotorgene 3000 (Phoenix Research, Hayward, CA)
SMAX	SpectraMax Gemini Spectrofluorometer (Molecular Devices Corp., Sunnyvale, CA)
TD20/20	Turner TD-20/20 Luminometer (Turner BioSystems, Inc, Sunnyvale, CA)
TLA	Thermo Luminoskan Ascent (Thermo Electron Corp., Waltham, MA)
TR	Turner Reporter Microplate Luminometer (Turner BioSystems, Inc, Sunnyvale, CA)

summarize these data. We have arbitrarily assigned these data to have quantitative values of twice the reported limit.

The remaining 160 semi-quantitative data were reported as upper limits (e.g., <0.031 ng/μL), with the majority of the limits at the lowest dilution of the calibration series in common use with the given method. More than 75% of the fully quantitative values for the given samples were greater than the specified maximum values. For these 123 data, assignment of any value less than the limit has little influence on robust statistical summaries. We have arbitrarily assigned these data to have quantitative values of

one-half the reported limit. The remaining 37 upper-limit data are within the central distribution of the fully quantitative values. No single value can be assigned to these data without potentially distorting the robust summaries. These data have been treated as "missing" values.

Method Sensitivities

Table 2 lists the percent success rate for obtaining quantitative values by method and sample [DNA]. The methods are in order of decreasing sensitivity. All Q-PCR assays, AluQuant, and in-house "ACES"-based methods provided quantitative results for all samples. Nearly all of the Picogreen and QuantiBlot assays provided quantitative results for all samples with target [DNA] of 0.16 ng/μL or greater. Colorimetric detection (TMB) appears less consistently able to provide quantitative results at target [DNA] of 0.05 ng/μL than chemilumenscence (ECL) or fluorescence (Picogreen). As expected, yield gels were too insensitive to reliably provide quantitative information even for the sample with the highest target [DNA] of 1.5 ng/μL.

TABLE 2—*The percent success rate reported for a sample.*

		% Quantitative Results*							
Target [DNA] ng/μL		1.5	0.5	0.5	0.16	0.16	0.05	0.05	0.05
Method	N_{anal}	A	B	E	C	F	D	G	H
Quantifiler	37	100	100	100	100	100	100	100	100
Other Q-PCR	23	100	100	100	100	100	100	100	100
"ACES"	14	100	100	100	100	100	100	100	100
AluQuant	13	100	100	100	100	100	100	100	100
PicoGreen	12	100	100	92	100	100	92	83	83
ECL	75	100	99	99	93	95	84	77	87
TMB	98	100	100	99	93	94	59	62	63
Yield gel	14	57	0	0	0	0	0	0	0
	286								

* Quantitative results are those that were reported as values, values reported as the range between contiguous calibration standards, values reported as less-than the lowest calibration standard if smaller than the target [DNA], or values reported as greater-than the highest calibration standard if larger than the target [DNA].

Combining Within-Analyst Replicates and Within-laboratory Duplicates

Since most of the within-analyst replicates reported appear to have been acquired to confirm a questioned result rather than resulting from designed experiments on analytical conditions, replicate results are unlikely to be statistically independent. To provide each unique combination of analyst and method with the same influence in the statistical analysis, each replicate set was replaced by the geometric means of the replicate values.

Likewise, since most analysts within a given laboratory can be assumed to have used similar reagents, equipment, and analytical protocols when carrying out a defined method, it is unlikely that the within-laboratory duplicates are truly independent. To provide each unique combination of laboratory and method with the same influence, each duplicate set was replaced by the geometric means of the duplicate values.

Distributions of the Among-laboratory Results

Figure 1 summarizes the among-laboratory results for the eight samples as a series of box-plots (33,34) for all quantitative results

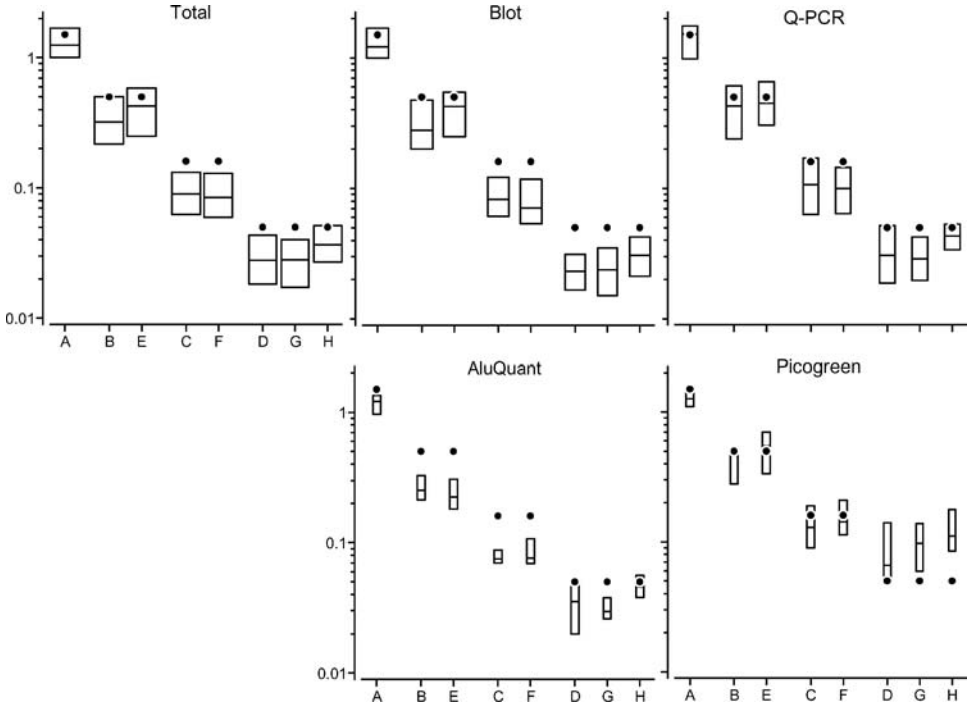

FIG. 1—*Box plot summary of the distribution of among-laboratory QS04 results. The "Total" subgraph summarizes all quantitative results for the eight samples regardless of measurement method; the other subgraphs provide summaries by method: all blot-based direct methods, all Q-PCR-based methods, AluQuant, and all of the Picogreen-based methods. The vertical length of each box within each of the subgraphs encompasses the central 50% of the results considered; the horizontal line within the box indicates the median value. The width of the box is a function of the relative number of results evaluated for the particular sample and method. The target values for the samples are denoted as solid circles. The samples are ordered by decreasing target [DNA]. Samples A to D were prepared from a multiple-source DNA; Samples E to H were prepared from a single-source male DNA. Samples A to G were distributed in polypropylene tubes; sample H was distributed in a Teflon tube.*

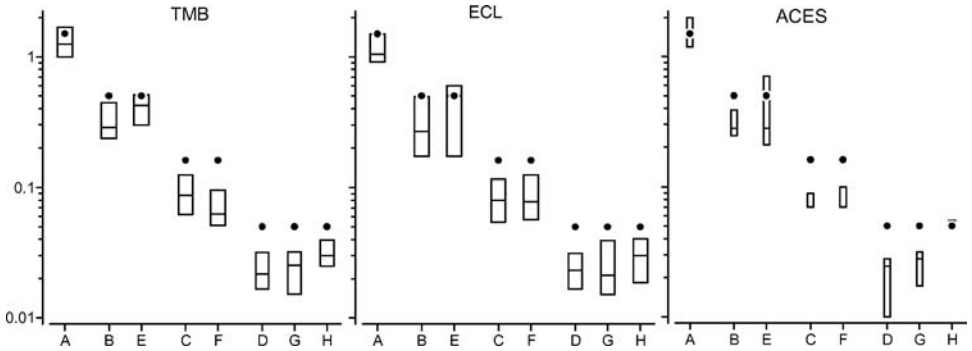

FIG. 2—*Box plot summary of the distribution of among-laboratory results for the major subgroups of the blot-based methods: TMB, ECL, and "ACES." The subgraph design is described in Fig. 1.*

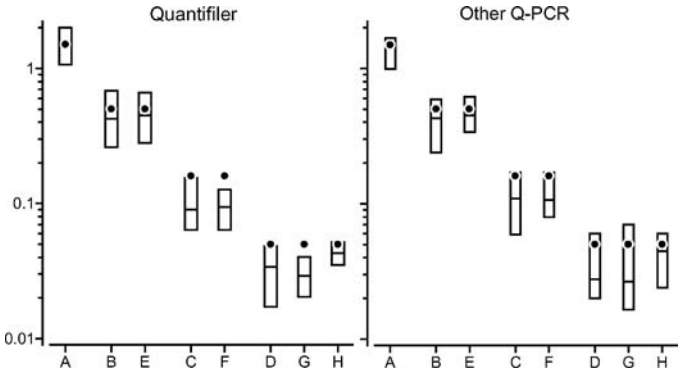

FIG. 3—*Box plot summary of the distribution of among-laboratory results for the Quantifiler method and for the aggregate of all other Q-PCR-based methods. The subgraph design is described in Fig. 1.*

combined and for the four major exclusive groups: all blot-based results, all Q-PCR results, and the two non-blot methods AluQuant and Picogreen. The consensus [DNA] for each set of results for each sample is indicated by the location of the median line within each box, the variability of the central 50% of results is indicated by the length of the box along the vertical axis, and the relative number of results is indicated by the width of the box along the horizontal axis. The pattern of results among the eight samples is similar for all methods, with the exception that the Picogreen-based methods over-estimate the [DNA] of the 0.05 ng/μL samples D, G, and H relative to the other assays. It is probable that this results from imperfect correction of background fluorescence with this assay at such low [DNA]; observation of a signal does not ensure that the signal has quantitative utility.

Figure 2 similarly summarizes the among-laboratory results for the three subgroups of blot-based direct methods: Quantiblot with TMB and ECL detection and the in-house "ACES"-based methods. Again, the pattern of results is quite similar, with the possible exception of the "ACES"-based methods with the Teflon-stored sample H. Figure 3 contrasts the among-laboratory results for Quantifiler with the collection of all other Q-PCR methods. The pattern of results is similar, although the variability of the collection is somewhat greater for the 0.05 ng/μL samples D, G, and H.

Table 3 lists the median [DNA] for the combined set of among-laboratory results. The median is the most commonly used robust estimate of consensus values for interlaboratory data, given the almost inevitable presence in such studies of a small number of "outlier" data (35). The percent recovery of the observed median relative to the target [DNA] range from 53% to 83% over the eight samples, for an average recovery of 66%. Potential causes of these systematically lower-than-expected consensus values and the effects of differences in DNA source, target [DNA], and the nature of the container are discussed below.

TABLE 3—*Consensus [DNA] and estimates of variability.*

[DNA], ng/μL				Multiplicative Standard Deviations			
Sample	Target	Median*	% Recovery†	s_{rep}‡	s_{dup}§	s_{among}‖	s'_{among}¶
A	1.50	1.25	83	1.3	1.5	1.5	—
B	0.50	0.32	64	1.5	1.6	1.9	1.4
E	0.50	0.42	84	1.4	1.6	1.8	1.6
C	0.16	0.090	56	1.6	1.6	1.7	1.4
F	0.16	0.084	53	1.6	1.5	1.8	1.5
D	0.050	0.028	56	1.5	1.5	1.9	1.6
G	0.050	0.028	56	1.5	1.6	1.8	1.6
H	0.050	0.037	73	1.8	1.5	1.6	1.7
		Average	66	1.5	1.5	1.8	1.5

* Median of the quantitative among-laboratory results for all methods, where the number of results varies from 103 to 117 depending on sample.

† 100 × (Median [DNA])/(Target [DNA]).

‡ Pooled multiplicative standard deviation of within-analyst replicates for all methods, where the number of replicate assays varies from 15 to 22 depending on sample.

§ Pooled multiplicative standard deviation of within-laboratory duplicates for all methods, where the number of duplicate assays varies from 52 to 64 depending on sample.

‖ Robust multiplicative standard deviation of the among-laboratory results for all methods, where the number of results varies from 103 to 117 depending on sample.

¶ Robust multiplicative standard deviation of the among-laboratory results for all methods after standardization to the value reported for sample A.

Measurement Variability

Given that the result distributions are quite similar for all adequately sensitive quantitative methods, the variability of current DNA quantitation methods can be estimated from the total set of interlaboratory data without regard for specific method. As noted in prior studies (16,17) and confirmed here by the approximate symmetry of the box-plot distributions when displayed on logarithmic axes, the reported [DNA] are lognormally distributed. The variability for such data is more readily understood when expressed in the units of measurement as multiplicative standard deviations, ×/÷SD, where the notation "×/÷" is employed as the multiplicative

analogue of "±". A [DNA] of 1.5 ng/μL with an associated multiplicative standard deviation of $\not{\times}2.0$ implies that the central 68% of results range from $1.5/2 = 0.75$ ng/μL to $1.5 \times 2 = 3.0$ ng/μL. The approximate 95% confidence intervals calculated using a coverage factor of 2.0 (36) are then $1.5/2^2 = 0.37$ ng/μL to $1.5 \times 2^2 = 6.0$ ng/μL.

Table 3 also details several variability components of the QS04 data. The variability of replicate assays (same analyst using the same method), s_{rep}, is estimated by pooling standard deviations calculated from the limited set of within-analyst replicates. The variability of duplicate assays (different analysts within the same laboratory using the same method), s_{dup}, is similarly estimated by pooling standard deviations calculated from the within-laboratory duplicates. The expected among-laboratory variability, s_{among}, is estimated as the appropriately scaled median absolute deviation from the median (MADe) of the among-laboratory results (35).

The within- and among-analyst variability estimates, s_{rep} and s_{dup}, are a remarkably similar $\not{\times}1.5$. The approximate 95% confidence factor for within-laboratory assays using a given method is thus $\not{\times}1.5^2 = \not{\times}2.3$. The among-laboratory variability, s_{among}, is on average $\not{\times}1.8$, with an associated 95% confidence factor of $\not{\times}1.8^2 = \not{\times}3.2$. This is identical to the among-laboratory variability observed in the 1999 Mixed Stain Study #2 for extracted DNA samples having nominal [DNA] ranging from 0.5 ng/μL to 5.0 ng/μL (16). However, it is slightly higher than the $\not{\times}1.6$ variability observed in the 2001 Mixed Stain Study #3 for extracted DNA samples of nominal [DNA] from 1.0 to 4.0 ng/μL (17). Given the $\not{\times}1.5$ result for the 1.5 ng/μL sample A, it is likely that the variability observed in this study for sub-ng/μL samples overestimates by 10% to 20% the variability expected for [DNA] of 1 ng/μL to 5 ng/μL.

Consensus Values and Variability as Functions of [DNA]

As noted above, the consensus [DNA] for the eight samples are on average about 66% of the target values. The better agreement for the higher-[DNA] samples A to C and for sample H distributed in Teflon tubes (see Table 3) suggests that some DNA becomes bound to the polypropylene tube. We are currently investigating the source of the difference between the consensus (1.25 ng/μL) and target (1.5 ng/μL) [DNA] for sample A. With the exception of the Picogreen methods, there is little evidence in Figs. 1 to 3 and in Table 3 for any increase in variability with decreasing [DNA] from the polypropylene-stored 0.50 ng/μL of samples B and E to the 0.05 ng/μL of samples D and G. The relative agreement between the within-laboratory, s_{rep} and s_{dup}, and among-laboratory, s_{among}, variability estimates for all samples suggests that there is little if any differential recovery of DNA bound to the walls of the sample container regardless of [DNA]: if it's bound, it apparently stays bound.

However, the among-laboratory variability of the Teflon-stored sample H is lower than for the polypropylene-stored samples with the same target [DNA]. Sample H is also the only sample for which the sample A-standardized among-laboratory variability, s'_{among}, is not smaller than the unstandardized s_{among}. It may be that while polypropylene-bound DNA is not differentially recoverable, the amount of DNA bound to different tubes may differ.

Measurement Performance Characteristics

While the accuracy of a single measurement can be characterized by the difference between the reported and the "true" value, the accuracy of a measurement process used to determine two or more values has two aspects: bias (or "trueness"), the difference between the measurement average and the putative true value, and precision, the standard deviation of the measurements (37). These two performance characteristics are rigorously defined only for a series of measurements on aliquots of a thoroughly characterized single material.

Given that the true "true value" for samples used in interlaboratory comparisons is seldom well known and that analysis of a number of similar but somewhat different samples can be more informative than the replicate analysis of one material, "comparability" is a more appropriate measurement performance concept for summarizing interlaboratory results (33). Like accuracy, comparability has two components: concordance, the average standardized difference between the reported and consensus values, and apparent precision, the standard deviation of the standardized differences (13,34). Concordance is the direct analogue of bias when evaluating a measurement process with similar but not identical samples. Likewise, apparent precision is analogous to precision but also incorporates sample-specific measurement differences or "matrix effects."

Figure 4 presents target plot (34) summaries of the measurement characteristics for all sets of among-laboratory results that include quantitative values for at least seven of the eight samples. For graphical clarity, the results are grouped by the three general categories of direct (non-blot), direct (blot), and Q-PCR. Each symbol represents the concordance of all quantitative measurements along the horizontal axis, their apparent precision along the vertical axis, and their total comparability as the distance from the target center. Three reference semi-circles are displayed: the inner-most "1-SD" semi-circle delimits a total comparability of one standard deviation from perfect agreement with the consensus medians for all samples, the middle "2-SD" delimits two standard deviations, and

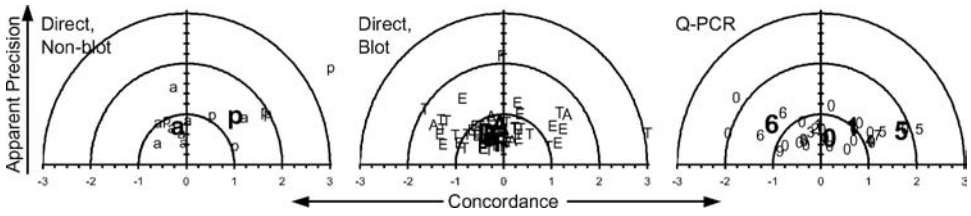

FIG. 4—*Target plot summary of measurement performance characteristics for the 118 QS04 among-laboratory results. For graphical clarity, the results are grouped by the three general categories of direct (non-blot), direct (blot), and Q-PCR. Each of the small symbols represents a single set ([DNA] of eight samples) of quantitative among-laboratory results: concordance is displayed along the horizontal axis, apparent precision along the vertical axis, and total comparability is the distance from the target center. Method codes are listed in Table 1. The large, bold-face symbols represent the median performance of the among-laboratory results for methods reported by two-or-more laboratories. Three reference semi-circles are displayed: the inner-most semi-circle delimits a total comparability of one standard deviation from perfect agreement with the consensus medians for all samples, the middle delimits two standard deviations, and the outer delimits three standard deviations.*

the outer "3-SD" delimits three standard deviations. Symbols that are to the left of center represent measurement processes that on average give results smaller than the consensus values; symbols to the right of center represent processes that on average give results larger than consensus. Symbols very close to the bottom of the target represent measurement processes that are quite consistently larger or smaller than consensus; symbols towards the top of each target represent processes that do not provide consistent results. Processes with measurement characteristics that fall within 2-SD can be considered as "exceptionally" to "acceptably" comparable; those between 2-SD and 3-SD as "marginally" comparable, and those outside 3-SD as "poorly" comparable.

As suggested by the variability reduction provided by standardizing the [DNA] reported for samples B to H to that reported for sample A, nearly all of the DNA quantitation processes employed in the QS04 study have remarkably good within-laboratory consistency but there are systematic among-laboratory differences. Only 2 of the 104 quantitatively complete data sets have apparent precisions above 2-SD, with a large majority below 1-SD. While only two of the data sets have concordances outside 3-SD, the majority are fairly uniformly distributed from −2-SD to +2-SD.

Figure 4 also displays the consensus performance characteristics for all quantitation methods reported by two or more participants, again estimating consensus as the median values. As expected from the distributions displayed in Fig. 1, the Picogreen results are on average somewhat greater than the consensus. While based upon too few data for confidence, one of the non-commercial Q-PCR methods appears to provide results that are consistently smaller than consensus while another provides results consistently larger than consensus. All of the commercial measurement systems have, on average, remarkably similar performance characteristics. While most methods thus appear capable of providing similar performance, the diversity of individual concordance values suggests that comparability is largely a function of particular analyst and/or laboratory practice.

Blot-Based Vs. Q-PCR Methods

As displayed in Fig. 4, the measurement performance characteristics of the Q-PCR and blot-based methods can be quite similar. As indicated in Table 2, the Q-PCR methods more consistently provide quantitative results than do the more commonly used blot methods for samples of target [DNA] 0.16 ng/μL and lower. Given that many forensic casework samples have very low amounts of DNA, it is likely that this superior sensitivity will encourage the adoption of Q-PCR methods.

Single-Source Vs. Multiple-Source Materials

Since many forensic casework samples are mixtures, it is valuable to explore whether there are discernable differences in the quantitation of single- and multiple-source materials. Samples E, F, G, and H are derived from a single DNA source while A, B, C, and D are derived from a multiple-source commercial material. The direct comparisons between samples B and E at 0.5 ng/μL, C and F at 0.16 ng/μL, and D and G at 0.05 ng/μL shown in Figs. 1 to 3 suggest that, at least for these two materials, none of the quantitative methods are particularly sensitive to differences in the number of DNA donors in a material.

Polypropylene Vs. Teflon Sample Containers

Conventional screw cap polypropylene tubes were used in this study with seven of the samples. To test an earlier hypothesis that a significant amount of DNA binds to polypropylene (17), Sample H was stored in Teflon tubes. As noted above and displayed in Figs. 1 to 3, the median results for the 0.05 ng/μL samples D and G (polypropylene) are lower than that for sample H (Teflon). The recovery of the target [DNA] for sample H, 73%, is considerably higher than the 56% observed for the other low-[DNA] samples (Table 3). This suggests that at this low [DNA], a significant proportion of the sample DNA does bind to the polypropylene walls.

Also as noted above, sample H is the only sample with an among-laboratory variability that is not decreased by standardization to the reported [DNA] of sample A. While as yet not understood, this suggests that polypropylene and Teflon storage differ qualitatively in some fashion beyond a simple difference in amount of DNA lost to the container surfaces.

Implications for the Development of the Proposed DNA Quantitation Standard, SRM 2372

The quality of the standards used to calibrate the various assays is a potential source of among-laboratory (and/or among-method for a given participant) variability, since all results are expressed relative to the putative [DNA] of the master standard used to calibrate a given assay. Re-expressing the [DNA] of all eight samples of a given reported sample set relative to the [DNA] of some one of the samples should largely remove this source of variability. Table 3 lists the MADe estimates for the variability of the among-laboratory results relative to the result reported for sample A, s'_{among}. The average variability for samples B to H is indeed reduced to $\times/1.5$, identical to that of the within-laboratory estimates.

The extra variability introduced by comparing results among methods and among different participants is approximately $\times/1.8/1.5 = \times/1.2$ or about 20%. An evaluation in our laboratory of four different commercially-available DNA standard materials, two of which were nominally identical but from different production batches and had very different storage histories, confirms this magnitude of difference among current DNA quantitation standards. On this evidence, we expect that calibration of working standards to a homogenous, stable, and properly characterized DNA quantitation reference material will help reduce among-laboratory variability. We are currently developing such a material, to be called NIST Standard Reference Material (SRM) 2372 DNA Quantitation Standard.

Conclusions

Regardless of DNA quantitation method employed, the one standard deviation within-laboratory variability for methods with adequate sensitivity for the samples evaluated is about $\times/1.5$. This applies both to within-analyst replicate and among-analyst duplicate measurements performed over a relatively short period of time. While there are a proportion of measurements that are not in good concordance with the majority, the expected one standard deviation among-laboratory variability for sub-ng/μL [DNA] is about $\times/1.8$. Much, if not all, of the excess 1.2-fold among-laboratory variability appears related to the quality of the master standards used to calibrate the methods.

To provide perspective, recall that PCR amplification nominally doubles the amount of DNA product every cycle. A factor-of-two uncertainty in the amount of template DNA is equivalent to ± a single PCR amplification cycle.

Acknowledgments

We express appreciation to all of the forensic scientists and their supervisors who took time out of their busy schedules to examine the samples provided as part of this interlaboratory study. This work would not be possible without their diligent efforts. We also gratefully acknowledge the technical assistance of Amy Decker and Jill Appleby of our laboratory in preparing some of the samples for shipping. This study would not have been possible without funding from the National Institute of Justice through an interagency agreement with the NIST Office of Law Enforcement Standards.

References

1. DNA Advisory Board, October 1998, Standards for Forensic DNA Testing Laboratories, Standard 9.3. See http://www.fbi.gov/hq/lab/codis/forensic.htm
2. Moretti TR, Baumstark AL, Defenbaugh DA, Keys KM, Smerick JB, Budowle B. Validation of short tandem repeats (STRs) for forensic usage: performance testing of fluorescent multiplex STR systems and analysis of authentic and simulated forensic samples. J Forensic Sci 2001;46: [PubMed] 647–60.
3. LaFountain MJ, Schwartz MB, Svete PA, Walkinshaw MA, Buel E. TWGDAM validation of the AmpFlSTR Profiler Plus and AmpFlSTR COfiler STR multiplex systems using capillary electrophoresis. J Foren- [PubMed] sic Sci 2001;46:1191–8.
4. Krenke BE, Tereba A, Anderson SJ, Buel E, Culhane S, Finis CJ, et al. Validation of a 16–locus fluorescent multiplex system. J Forensic Sci [PubMed] 2002;47:773–85.
5. Holt CL, Buoncristiani M, Wallin JM, Nguyen T, Lazaruk KD, Walsh PS. TWGDAM validation of AmpFlSTR PCR amplification kits for forensic [PubMed] DNA casework. J Forensic Sci 2002;47:66–96.
6. Butler JM. Forensic DNA typing. London: Academic Press, 2001.
7. Nicklas JA, Buel E. Quantitation of DNA in forensic samples. Anal [PubMed] Bioanal Chem. 2003;376:1160–7.
8. Gill P, Sparkes R, Fereday L, Werrett D. Report of the European Network of Forensic Science Institutes (ENSFI): formulation and testing of princi- [PubMed] ples to evaluate STR multiplexes. Forensic Sci Int 2000;108(1):1–29.
9. Nicklas JA, Buel E. Development of an Alu-based, QSY 7-labeled primer PCR method for quantitation of human DNA in forensic samples. [PubMed] J Forensic Sci 2003;48:282–91.
10. Duewer DL, Currie LA, Reeder DJ, Leigh SD, Liu HK, Mudd JL. Interlaboratory comparison of autoradiographic DNA profiling measurements. 2. Measurement uncertainty and its propagation. Anal Chem [PubMed] 1995;67:1220–31.
11. 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 [PubMed] pairs. Anal Chem 1996;68:1941–7.
12. Duewer DL, Currie LA, Reeder DJ, Leigh SD, Filliben JJ, Liu HK, et al. Interlaboratory comparison of autoradiographic DNA profiling [PubMed] measurements. 4. Protocol effects. Anal Chem 1997;69:1882–92.
13. Duewer DL, Lalonde SA, Aubin RA, Fourney RM, Reeder DJ. Interlaboratory comparison of autoradiographic DNA profiling measurements: [PubMed] precision and concordance. J Forensic Sci 1998;43:465–71.
14. Kline MC, Redman JW, Reeder DJ, Duewer DL. Intercomparison of DNA sizing ladders in electrophoretic separation matrices and their potential for accurate typing of the D1S8O locus. Appl Theor Electrophor [PubMed] 1996;6:33–41.
15. Kline MC, Duewer DL, Newall P, Redman JW, Reeder DJ, Richard M. Interlaboratory evaluation of short tandem repeat triplex CTT. J Forensic [PubMed] Sci 1997;42:897–906.
16. Duewer DL, Kline MC, Redman JW, Newall PJ, Reeder DJ. NIST mixed stain studies #1 and #2: interlaboratory comparison of DNA quantitation practice and short tandem repeat multiplex performance with multiple- [PubMed] source samples. J Forensic Sci 2001;46:1199–210.
17. Kline MC, Duewer DL, Redman JW, Butler JM. NIST Mixed Stain Study 3: DNA quantitation accuracy and its influence on short tandem repeat [PubMed] multiplex signal intensity. Anal Chem 2003;75:2463–9.
18. Maniatis T, Fritsch EF, Sanbrook J. Molecular cloning: A laboratory manual; Cold Spring Harbor, NY: Cold Spring Harbor Laboratory, 1982; Appendix A, pp. 468–9.
19. Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res 1988;16(3):1215. [PubMed]
20. Walsh PS, Varlaro J, Reynolds R. A rapid chemiluminescent method for quantitation of human DNA. Nucleic Acids Res 1992;20:5061–5. [PubMed]
21. Budowle B, Baechtel FS, Comey CT, Giusti AM, Klevan L. Simple protocols for typing forensic biological evidence: Chemiluminescent detection for human DNA quantification and RFLP analyses and manual typing of PCR amplified polymorphisms. Electrophoresis 1995;16:1559–67. [PubMed]
22. Solulink, Inc. HydraD17/AP Probe, 5′ Alkaline Phosphatase/D17Z1 Probe.
23. Mandrekar MN, Erickson AM, Kopp K, Krenke BE, Mandrekar PV, Nelson R, et al. Development of a human DNA quantitation system. Croat Med J 2001;42:336–9. [PubMed]
24. Singer VL, Jones LJ, Yue ST, Haugland RP. Characterization of PicoGreen reagent and development of a fluorescence-based solution assay for double-stranded DNA quantitation. Anal Biochem 1997;249:228–38. [PubMed]
25. Applied Biosystems Quantifiler kits: Quantifiler human DNA quantitation kit and Quantifiler Y human male DNA quantitation kit user's manual, Part Number 4344790 Rev A. Foster City, CA: Applied Biosystems, 2003.
26. Nicklas JA, Buel E. Development of an Alu-based, real-time PCR method for quantitation of human DNA in forensic samples. J Forensic Sci 2003;48:936–44. [PubMed]
27. Sifis ME, Both K, Burgoyne LA. A more sensitive method for the quantitation of genomic DNA by Alu amplification. J Forensic Sci 2002;47:589–92. [PubMed]
28. Haque KA, Pfeiffer RM, Beerman MB, Struewing JP, Chanock SJ, Bergen AW. Performance of high-throughput DNA quantitation methods. BMC Biotechnol 2003;3:20. [PubMed]
29. Richard ML, Frappier RH, Newman JC. Developmental validation of a real-time quantitative PCR assay for automated quantitation of human DNA. J Forensic Sci 2003;48:1041–6. [PubMed]
30. Andreasson H, Gyllensten U, Allen M. Real-time DNA quantitation of nuclear and mitochondrial DNA in forensic analysis. BioTechniques 2002;33:402–11. See also: Andreasson H, Allen M. Rapid quantitation and sex determination of forensic evidence materials. J Forensic Sci 2003;48:1280–7. [PubMed]
31. Nolde C, Reid P, Hartmann J. Development of an ALU-based real-time PCR assay for the quantitation of genomic (human) DNA. Poster presented at 13th International Symposium on Human Identification, Phoenix, AZ, 2002 October 7–10. Abstract published at: http://www.promega.com/geneticidproc/ussymp13proc/abstracts/49_hartmann.pdf
32. Fox JC, Cave CA, Schumm W. Development, characterization, and validation of a sensitive primate-specific quantitation assay for forensic analysis. BioTechniques 2003;34:314–22. [PubMed]
33. Tukey JW. Exploratory data analysis. Reading, MA: Addison-Wesley 1977.
34. 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(9):1870–1878. [PubMed]
35. Analytical Methods Committee, Royal Society of Chemistry. Robust statistics—How not to reject outliers. Part 1. Basics. Analyst 1989, 114, 1693–8. See also: www.rsc.org/lap/rsccom/amc/amc_index.htm
36. ISO. Guide to the expression of uncertainty in measurement. Geneva: ISO, 1995.
37. BIPM/IEC/IFCC/ISO/IUPAC/IUPAP/OIML. International vocabulary of basic and general terms in metrology. 2nd ed. Geneva: ISO, 1993.

Additional information and reprint requests:
Margaret Kline, M.S.
Research Biologist
National Institute of Standards and Technology
100 Bureau Dr., Mail Stop 8311
Gaithersburg, MD 20899-8311
Tel: 301-975-3134
Fax: 301-975-8505
E-mail: margaret.kline@nist.gov