STR Amplifications Using Dilutions of the NIST Human DNA Quantitation Standard SRM 2372A: Implications for Analysis and Validation

Mark Timken, Sonja Klein, Martin Buoncristiani
California Department of Justice
Jan Bashinski DNA Lab, Richmond, CA
mark.timken@doj.ca.gov
sonja.klein@doj.ca.gov
martin.buoncristiani@doj.ca.gov

August 6, 2014

NIST DNA Analyst Webinar Series: Validation Concepts and Resources - 1

Validation Questions

- How are the current and new STR/CE systems different in terms of … ?

sensitivity for allelic peak detection	susceptibility to "drop-in" and contamination	template (input) dynamic range
probability of allelic dropout at low template	heterozygous peak-height-ratio balance	stochastic threshold setting

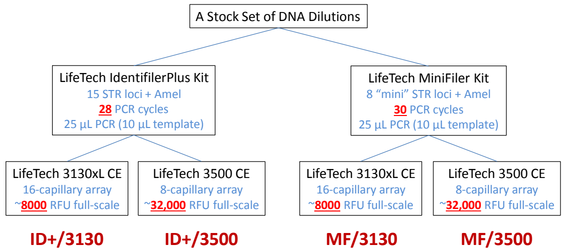
- These same questions can arise when we look at STR data obtained by different laboratories, even labs that use the same STR/CE combination, but that may use different amplification cycles or volumes, different post-amp purifications, different CE run parameters, or different analytical thresholds.

A Sensitivity Study Using Dilutions of NIST SRM 2372A DNA*

- Design
- Results
  - Heterozygous Peak-Height Ratios
  - Linear Signal Response (RFU vs. template)
  - Heterozygous Allelic Dropout Frequencies
- Implications for Analysis and Validation
  - Interpreting the Analytical Threshold
  - Predicting Allelic Dropout Probabilities
  - Setting a Stochastic Threshold
  - Comparing STR/CE Systems
  - Validation and "Standardization"

* M.D. Timken, S.B. Klein, M.R. Buoncristiani, Stochastic sampling effects in STR typing: Implications for analysis and interpretation, Forens. Sci. Int. Genet. 11 (2014) 195-204.

Design



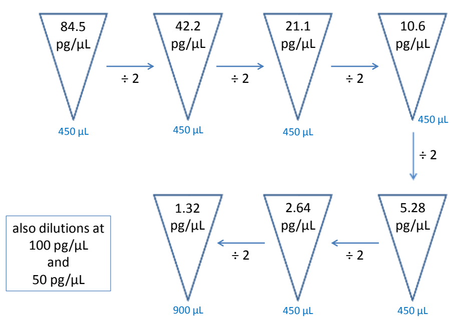
NOTE: Prior to this study, each of the four combinations had been previously validated for casework analysis at the CADOJ. This specific study was not designed for the purpose of validating either STR kit or CE platform.

DNA = NIST SRM 2372A

- 1 of 3 DNA components (A,B,C) in the NIST Human DNA Quantitation Standard (SRM 2372)
- 2372A DNA
  - ***known concentration*** = 57 ng/µL (absorbance, dPCR)
    - important for template-based simulation and modeling
  - single-source male donor (extracted from blood)
  - high quality (non-degraded, non-inhibited)
  - heterozygous at 11 of 15 ID+ STR loci (& Amel)
  - heterozygous at 8 of 8 MF STR loci (& Amel)
  - it's a ***standard*** – the *same sample* can be run by any lab

A Single Stock Dilution Series:

2-fold dilutions, "large" volumes, lo-bind tubes, same day amps



Replicate Amplifications of 2372A Dilutions

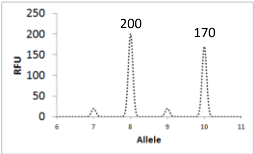
Template (pg)	Template (average # of diploid cell equivalents)*	Number of Replicate Amplifications (ID+ and MF)	Total Number of Heterozygous Loci		Total Number of Heterozygous Alleles	
			ID+	MF	ID+	MF
845 (ID+ only)	845/6.6 = 128	5	60		120	
422	64	9	108	81	216	162
211	32	12	144	108	288	216
106	16	16	192	144	384	288
52.8	8	16	192	144	384	288
26.4	4	16	192	144	384	288
13.2	2	16	192	144	384	288
1000 (ID+ only)	151.5	2	24		48	
500 (MF only)	75.8	2		18		36

* 6.6 pg = 1 human diploid cell equivalent (Butler) = 1 pair of sister heterozygous alleles

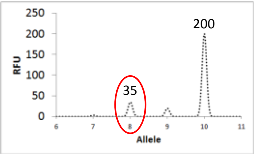
Results: Heterozygous Peak Height Ratios (PHR)

$$PHR = \frac{\text{peak height of less intense sister allele}}{\text{peak height of more intense sister allele}}$$

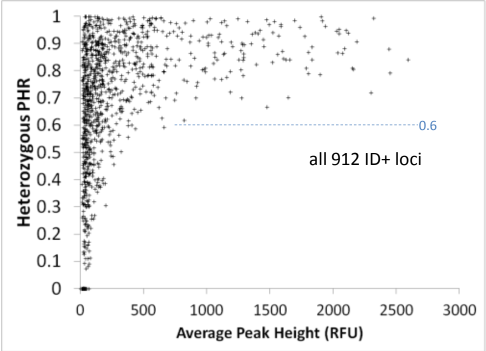
PHR = 170/200 = 0.85

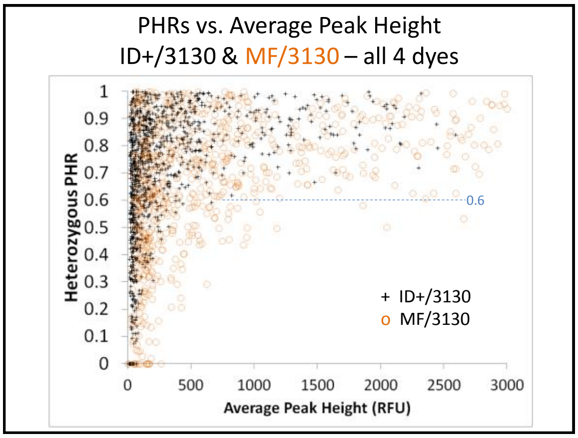


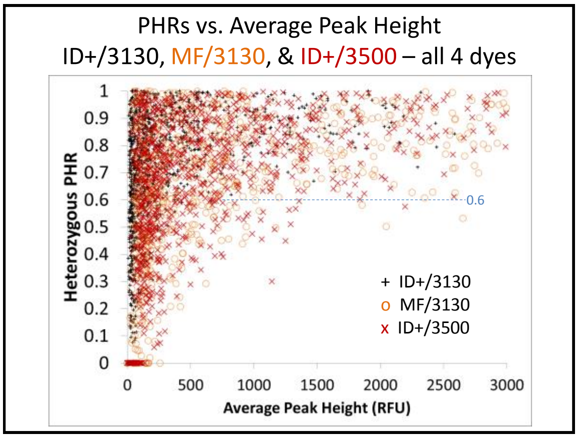
PHR = 35/200 = 0.175

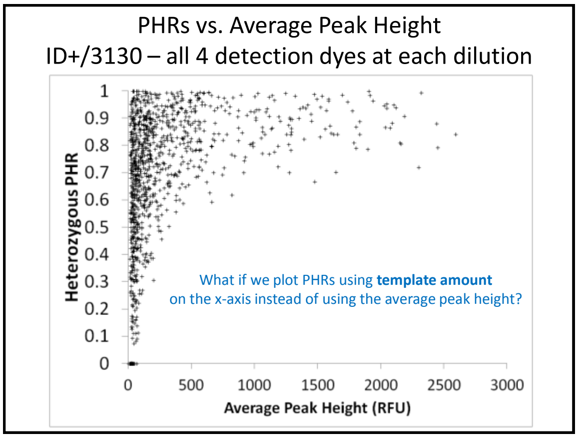


PHRs vs. Average Peak Height
ID+/3130 – all 4 detection dyes at each dilution

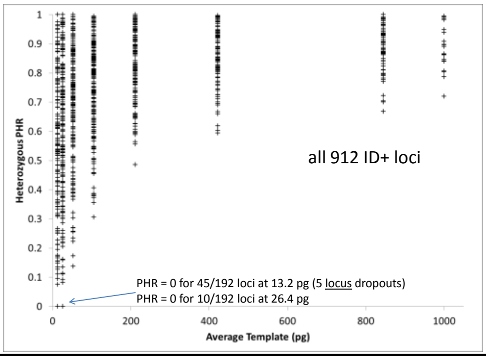



PHRs vs. Average Peak Height
ID+/3130 & MF/3130 – all 4 dyes
Heterozygous PHR
Average Peak Height (RFU)
0.6
+ ID+/3130
o MF/3130

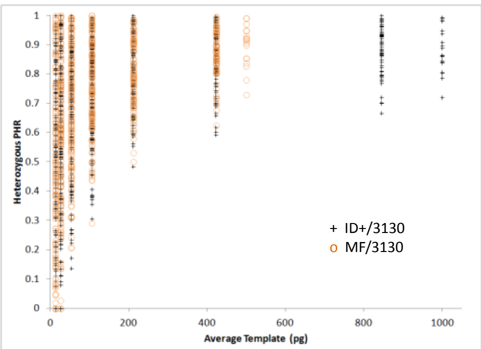

PHRs vs. Average Peak Height
ID+/3130, MF/3130, & ID+/3500 – all 4 dyes
Heterozygous PHR
Average Peak Height (RFU)
0.6
+ ID+/3130
o MF/3130
x ID+/3500


PHRs vs. Average Peak Height
ID+/3130 – all 4 detection dyes at each dilution
Heterozygous PHR
Average Peak Height (RFU)
What if we plot PHRs using template amount
on the x-axis instead of using the average peak height?

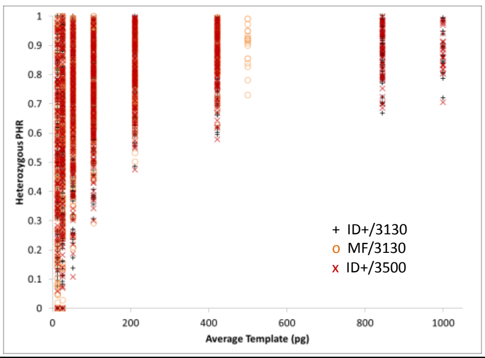
Distributions of PHRs vs. Template
ID+/3130 – all 4 detection dyes at each dilution



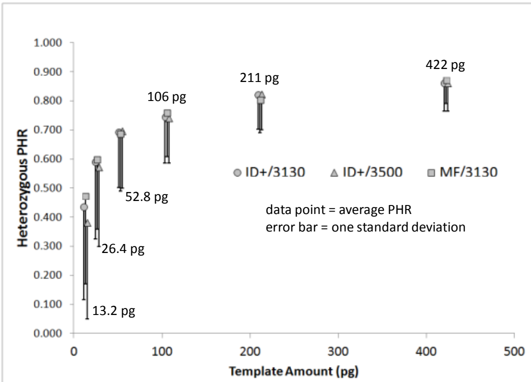
Distributions of PHRs vs. Template
ID+/3130 & MF/3130 – all 4 dyes



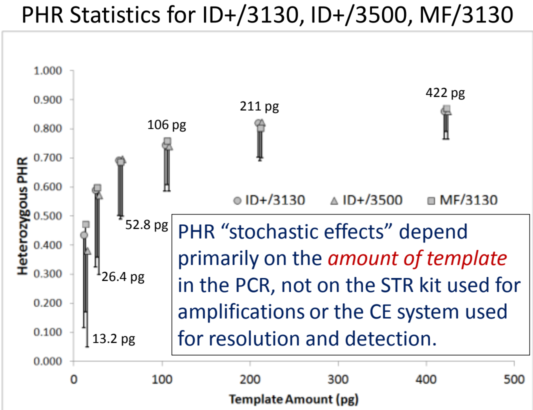
Distributions of PHRs vs. Template
ID+/3130, MF/3130, & ID+/3500 – all 4 dyes



PHR Statistics for ID+/3130, ID+/3500, MF/3130



PHR Statistics for ID+/3130, ID+/3500, MF/3130



An explanation …

For low-template amplifications of extracted DNA, ***pre-PCR*** stochastic sampling of the alleles into the amplification reaction is the primary source of post-PCR signal variance.

A. Jeffreys, V. Wilson, R. Neumann, J. Keyte, Amplification of human minisatellites by the polymerase chain reaction: towards DNA fingerprinting of single cells, Nucleic Acids Res. 16 (1988) 10953-10971.

P.S. Walsh, H.A. Erlich, R. Higuchi, Preferential PCR amplification of alleles: mechanisms and solutions, CSH Genome Res. 1 (1992) 241-250.

J. Stenman, A. Orpana, Accuracy in amplification, Nat. Biotechnol., 19 (2001) 1011-12.

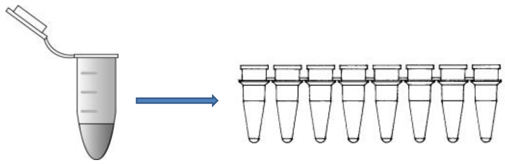
E.T. Lagally, I. Medintz, R.A. Mathies, Single-molecule DNA amplification and analysis in an integrated microfluidic device, Anal. Chem., 73 (2001) 565-570.

P.Gill, J. Curran, K. Elliot, A graphical simulation model of the entire DNA process associated with the analysis of short tandem repeat loci, Nucleic Acids Res., 33 (2005) 632-643.

An Example: Pre-PCR Stochastic Sampling with 52.8 pg Amps

- 8 diploid-cell equivalents per PCR amplification -

- ***on average***, there will be 8 template copies of each sister allele at each heterozygous locus per 10 μL sample used for PCR -



450 μL of
2372A template DNA
at 5.28 pg/ μL
(vortexed, homogeneous)

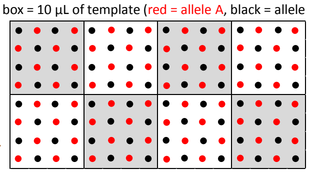
- transfer 10 μL into each PCR tube for 8 separate amplifications, each with 52.8 pg per amp

An Example: Pre-PCR Stochastic Sampling with 52.8 pg Amps

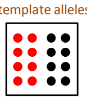
what we imagine: pre-PCR ***uniform*** sampling

1 box = 10 μL of template (red = allele A, black = allele B)

Every 10-μL transfer volume into the PCR will contain 8+8 sister alleles at each heterozygous locus.



Balanced template alleles.



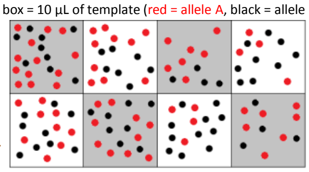
An Example: Pre-PCR Stochastic Sampling with 52.8 pg Amps

what we imagine: pre-PCR ***uniform*** sampling

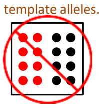
what we get: pre-PCR ***stochastic*** sampling

1 box = 10 μL of template (red = allele A, black = allele B)

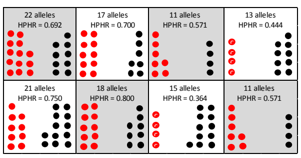
Every 10-μL transfer volume into the PCR will contain 8+8 sister alleles at each heterozygous locus.



Balanced template alleles.

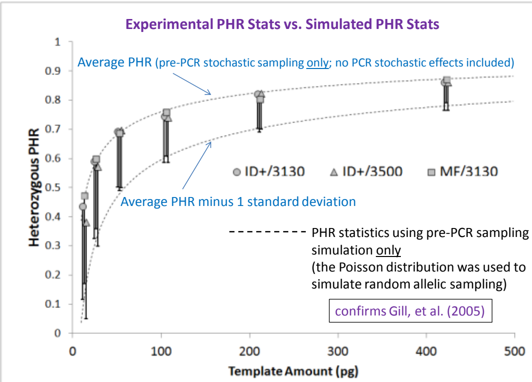


Pre-PCR stochastic sampling predicts that results will be different for each aliquot we take. (using the Poisson distribution (λ=8))



For extracted DNA, the variability due to **pre-PCR** stochastic sampling is the primary source of post-PCR signal variability and PHR variability.

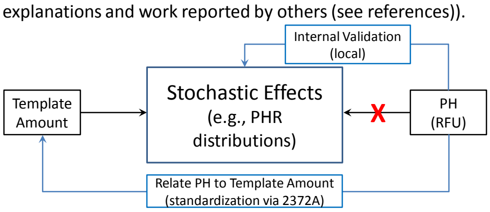
PHR Statistics for ID+/3130, ID+/3500, MF/3130



PHR Conclusions

(for extracted DNA dilutions)

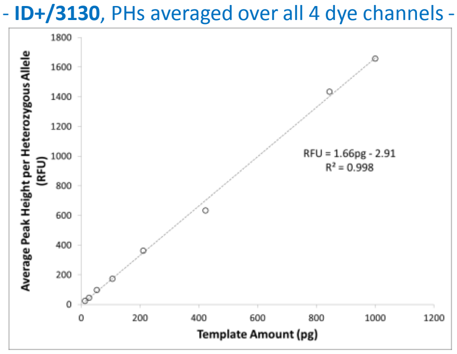
- PHR stochastic effects depend primarily on the *amount of template* … not on peak height (PH), not on the STR kit (or # of cycles), not on the CE platform.
- PHR distributions can be modeled by *pre-PCR* stochastic sampling of the sister alleles into the amplification reaction (confirming explanations and work reported by others (see references)).



Results: Linear Signal Response

Average PH for Heterozygous Alleles vs. Template

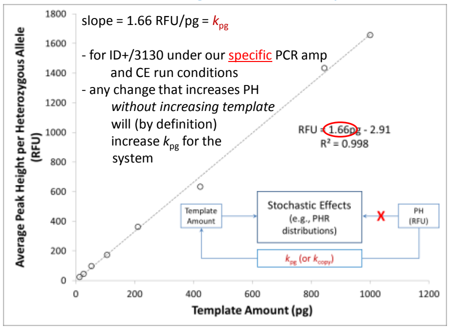
- **ID+/3130**, PHs averaged over all 4 dye channels -



Linear Signal Response (k_{pg})

Average PH for Heterozygous Alleles vs. Template

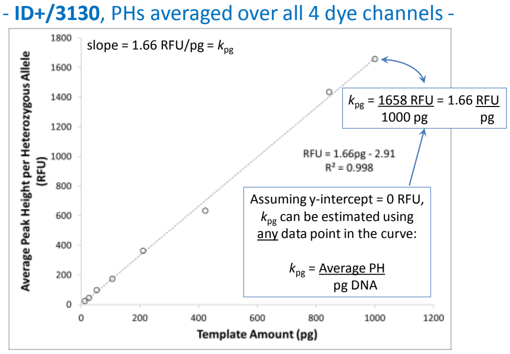
- **ID+/3130**, PHs averaged over all 4 dye channels -



Linear Signal Response (k_{pg})

Average PH for Heterozygous Alleles vs. Template

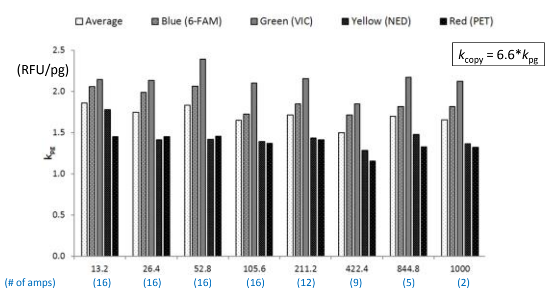
- **ID+/3130**, PHs averaged over all 4 dye channels -



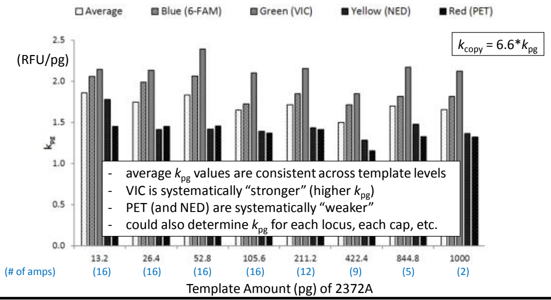
Linear Signal Response for **ID+/3130**

- k_{pg} calculated for each dilution level -

- averaged over all 4 dye channels and averaged for each dye channel separately -



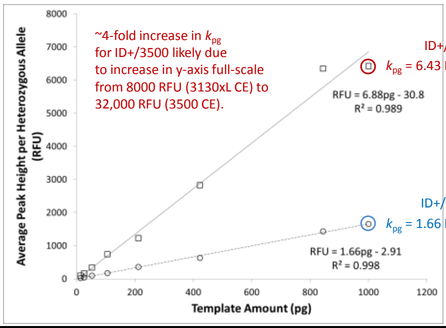
Linear Signal Response for **ID+/3130**
- k_{pg} calculated for each dilution level -
- averaged over all 4 dye channels and averaged for each dye channel separately -



Linear Signal Response (k_{pg})

Average PHs for Heterozygous Alleles vs. Template

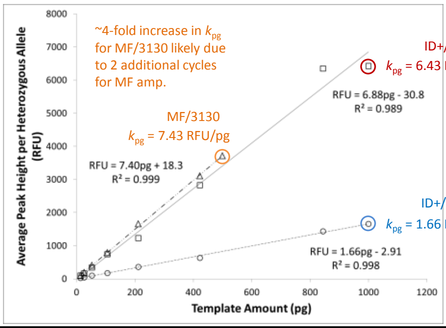
- **ID+/3130 & ID+/3500**, PHs averaged over all 4 dye channels -



Linear Signal Response (k_{pg})

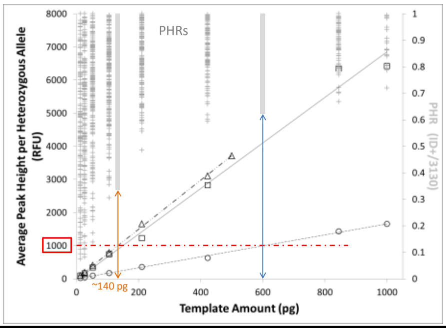
Average PHs for Heterozygous Alleles vs. Template

- **ID+/3130, ID+/3500**, & **MF/3130** PHs over all 4 dye channels -



Implications for Analysis

When compared at the same RFU level, why did the PHRs *appear* to be so much more "stochastic" for MF/3130 than for ID+/3130?



Implications for Analysis

- Comparing Allelic Detection Sensitivities for Different STR/CE Systems by Interpreting the Analytical Threshold (AT) in Terms of the Amount of Template*

$AT_{pg} = AT(RFU) \div k_{pg}$

* T. Tvedebrink, P.S. Eriksen, M. Asplund, H.S. Morgensen, N. Morling, Allelic drop-out probabilities estimated by logistic regression – further considerations and practical implementation. Forensic Sci. Int. Genet. 6 (2012) 263-267.

Implications for Analysis

- Comparing Allelic Detection Sensitivities for Different STR/CE Systems by Interpreting the Analytical Threshold (AT) in Terms of the Amount of Template

$AT_{pg} = AT(RFU) \div k_{pg}$

e.g., for **our** ID+/3130 system (validated at an AT = 50 RFU),

AT_{pg} = 50 RFU ÷ 1.66 RFU/pg = 30.1 pg

→ for repeated amps of 30.1 pg samples, expect to detect ~50% of heterozygous alleles at AT=50 RFU

(NOTE: 30.1 pg is an average of 4.6 diploid cell equivalents.)

→ AT_{pg} is a simple, useful measure of allelic detection sensitivity.

STR/CE System Comparisons†

V = CADOJ Validated; W = "what if?"	V	V	V	W	1997¥
STR Kit	ID+	ID+	MF	ID+	ProfilerPlus
Electrophoresis Platform	3130	3500	3130	3500	310
PCR Volume (µL)	25	25	25	25	***50***
PCR Cycles	28	28	30	28	28
Full-Scale RFU	8000	32000	8000	32000	8000
k_{pg} (RFU/pg) (from 2372A sensitivity study)	1.66	6.43	7.43	6.43	***0.83***
AT (RFU) (from baseline noise study)	50	150	50	***50***	***150¥***
AT_{pg} (pg) = AT(RFU)/k_{pg}	30.1	23.3	6.7	7.8	180.7
AT_{copy} (copies) = k_{pg}/6.6	4.6	3.5	1.0	1.2	27.4
~Input for 1/4 Full-Scale (pg)*	1205	1244	269	1244	2410
~Input for 1/2 Full-Scale (pg)**	2410	2488	538	2488	4819

* = (Full-Scale RFU/4)/k_{pg}
** = (Full-Scale RFU/2)/k_{pg}

ID+/3130

¥recommended by AB (1997)

†For comparisons, assuming all samples are prepared for CE by combining 1 µL of PCR product with 9 µL of formamide/size standard.

STR/CE System Comparisons

V = CADOJ Validated; W = "what if?"	V	V	V	W	1997¥
STR Kit	ID+	ID+	MF	ID+	ProfilerPlus
Electrophoresis Platform	3130	3500	3130	3500	310
PCR Volume (µL)	25	25	25	25	***50***
PCR Cycles	28	28	30	28	28
Full-Scale RFU	8000	32000	8000	32000	8000
k_{pg} (RFU/pg) (from 2372A sensitivity study)	1.66	6.43	7.43	6.43	***0.83***
AT (RFU) (from baseline noise study)	50	150	50	***50***	***150¥***
AT_{pg} (pg) = AT(RFU)/k_{pg}	30.1	23.3	6.7	7.8	180.7
AT_{copy} (copies) = k_{pg}/6.6	4.6	3.5	1.0	1.2	27.4
~Input for 1/4 Full-Scale (pg)*	1205	1244	269	1244	2410
~Input for 1/2 Full-Scale (pg)**	2410	2488	538	2488	4819

* = (Full-Scale RFU/4)/k_{pg}
** = (Full-Scale RFU/2)/k_{pg}

¥recommended by AB (1997)

Based on ***our*** AT_{pg} values, ***our*** ID+/3500 system is somewhat more sensitive than ***our*** ID+/3130 system, so it will *detect* elevated stochastic effects near the AT.
(NOTE: If our 3500 CE had been validated to have an AT = 194 RFU, then the allelic detection sensitivities would have been the same as the 3130 CE.)

STR/CE System Comparisons

V = CADOJ Validated; W = "what if?"	V	V	V	W	1997¥
STR Kit	ID+	ID+	MF	ID+	ProfilerPlus
Electrophoresis Platform	3130	3500	3130	3500	310
PCR Volume (µL)	25	25	25	25	***50***
PCR Cycles	28	28	30	28	28
Full-Scale RFU	8000	32000	8000	32000	8000
k_{pg} (RFU/pg) (from 2372A sensitivity study)	1.66	6.43	7.43	6.43	***0.83***
AT (RFU) (from baseline noise study)	50	150	50	***50***	***150¥***
AT_{pg} (pg) = AT(RFU)/k_{pg}	30.1	23.3	6.7	7.8	180.7
AT_{copy} (copies) = k_{pg}/6.6	4.6	3.5	1.0	1.2	27.4
~Input for 1/4 Full-Scale (pg)*	1205	1244	269	1244	2410
~Input for 1/2 Full-Scale (pg)**	2410	2488	538	2488	4819

* = (Full-Scale RFU/4)/k_{pg}
** = (Full-Scale RFU/2)/k_{pg}

¥recommended by AB (1997)

MF/3130 is significantly more sensitive:
→ ~single-copy sensitivity (~50% dropout)
→ can detect lower template, so will observe increased stochastic effects
→ more susceptible to contamination/drop-in
→ reduced input dynamic range

STR/CE System Comparisons

V = CADOJ Validated; W = "what if?"	V	V	V	W	1997[¥]
STR Kit	ID+	ID+	MF	ID+	ProfilerPlus
Electrophoresis Platform	3130	3500	3130	3500	310
PCR Volume (µL)	25	25	25	25	***50***
PCR Cycles	28	28	30	28	28
Full-Scale RFU	8000	32000	8000	32000	8000
k_{pg} (RFU/pg) (from 2372A sensitivity study)	1.66	6.43	7.43	6.43	***0.83***
AT (RFU) (from baseline noise study)	50	150	50	***50***	***150***[¥]
AT_{pg} (pg) = AT(RFU)/k_{pg}	30.1	23.3	6.7	7.8	180.7
AT_{copy} (copies) = k_{pg}/6.6	4.6	3.5	1.0	1.2	27.4
~Input for 1/4 Full-Scale (pg)*	1205	1244	269	1244	2410
~Input for 1/2 Full-Scale (pg)**	2410	2488	538	2488	4819

* = (Full-Scale RFU/4)/k_{pg}
** = (Full-Scale RFU/2)/k_{pg}

¥recommended by AB (1997)

What if we had set the AT for the 3500 CE at ***50*** RFU for ID+?

→ nearly single-copy sensitivity
→ detect lower template and see increased stochastic effects
→ more susceptible to contamination/drop-in

STR/CE System Comparisons

V = CADOJ Validated; W = "what if?"	V	V	V	W	1997[¥]
STR Kit	ID+	ID+	MF	ID+	ProfilerPlus
Electrophoresis Platform	3130	3500	3130	3500	310
PCR Volume (µL)	25	25	25	25	***50***
PCR Cycles	28	28	30	28	28
Full-Scale RFU	8000	32000	8000	32000	8000
k_{pg} (RFU/pg) (from 2372A sensitivity study)	1.66	6.43	7.43	6.43	***0.83***
AT (RFU) (from baseline noise study)	50	150	50	***50***	***150***[¥]
AT_{pg} (pg) = AT(RFU)/k_{pg}	30.1	23.3	6.7	7.8	180.7
AT_{copy} (copies) = k_{pg}/6.6	4.6	3.5	1.0	1.2	27.4
~Input for 1/4 Full-Scale (pg)*	1205	1244	269	1244	2410
~Input for 1/2 Full-Scale (pg)**	2410	2488	538	2488	4819

* = (Full-Scale RFU/4)/k_{pg}
** = (Full-Scale RFU/2)/k_{pg}

¥recommended by AB (1997)

Due to 50-µL PCR volume (which reduces k_{pg} by a factor of two[ǂ]) and higher AT(RFU), the "1997" version of ProfilerPlus on the 310 was running at an estimated AT_{pg} of ~180 pg !!

ǂM.L. Gaines, P.W. Wojkiewicz, J.A. Valentine, C.L. Brown, Reduced Volume PCR Amplification Reactions Using the AmpFlSTR Profiler Plus Kit, J. Forens. Sci. 47 (2002) 1224-1237.

Implications for Analysis

Predicting Allelic Dropout Probabilities

Pre-PCR stochastic sampling simulations, along with the STR/CE system's k_{pg} (or k_c) and its Analytical Threshold (AT), can be used to predict the probability of allelic dropout … see our FSIG paper for more, including how to estimate a stochastic threshold …

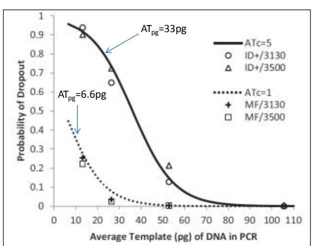


Fig. 6. Data points show empirical allelic dropout frequencies for 2372A dilutions with 13.2, 26.4, 52.8, and 105.6 pg of template, where dropouts were identified by analyzing Identifiler Plus data at an AT_c of 5 copies and MiniFiler data at an AT_c of 1 copy. Logistic regression curves show predicted dropout probabilities at AT_c = 5 and AT_c = 1 that are based solely on pre-PCR stochastic sampling simulations.

M.D. Timken et al. / Forensic Science International: Genetics 11 (2014) 195–204

Implications for Validation of a New STR Kit or CE Detection Platform

- Include 2-3 amps of 2372A* DNA as part of your internal validation study.
  - e.g., for ID+, you could include duplicated amps at 750 pg and 500 pg (*linear range*)
- Use the 2372A* results to determine k_{pg}, which will relate peak heights (RFU) to **template amount** (pg or starting allelic copies) for your STR/CE system.
- Use the k_{pg}, the AT, and the CE's full-scale RFU value to estimate many detection characteristics of the new system:
  - allelic detection sensitivity and susceptibility to contamination/drop-in (AT_{pg})
  - input dynamic range
  - expected PHR distribution vs. average peak height
  - probability of allelic dropout (vs. template or vs. detected peak height)
    - stochastic threshold setting
    - input required to detect a full profile
- Similarly "standardized" systems could be compared by using their k_{pg} and AT values as points of reference.

(*or some other *accurately* quantified template source)

Thank you for listening.

Thanks to the organizers for allowing me to present the CA DOJ results.

Thanks to Jeanette Wallin (CA DOJ) and Brian Harmon (CA DOJ) for assistance, as well as useful discussions and encouragement.

Forensic Science International: Genetics 11 (2014) 195–204
Contents lists available at ScienceDirect
Forensic Science International: Genetics
journal homepage: www.elsevier.com/locate/fsig

Stochastic sampling effects in STR typing: Implications for analysis and interpretation

Mark D. Timken[1], Sonja B. Klein[1,*], Martin R. Buoncristiani
State of California, Department of Justice, Jan Bashinski DNA Laboratory, 1001 W. Cutting Blvd., Richmond, CA 94804, USA