Promega

21st International Symposium on Human Identification
Mixture Interpretation Workshop:
Principles, Protocols, and Practice
October 11, 2010 – San Antonio, TX

NIJ National Institute of Justice

Amplification Variation and Stochastic Effects

Robin W. Cotton

BOSTON UNIVERSITY

NIST

Outline: Amplification Variation and Stochastic Effects

- **PRINCIPLES**
  - Sources of variation
  - Sampling
  - Probability of allele drop out
- **PROTOCOLS**
  - Approaches to characterizing variation and recognizing when stochastic effects are likely to occur
  - Use of this data in protocol development
- **PRACTICE**
  - How do we choose a threshold
  - Impact of that choice
  - SWGDAM Guidelines and ISFG Recommendations

PRINCIPLES

Ideally

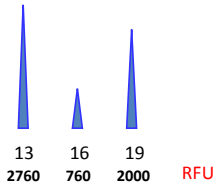
- Begin with extracted DNA
- If all were perfect there would be an equal number of copies of each allele at each locus and all copies would amplify equally

PRINCIPLES

Even in a mixture, amplification would be proportional to the number of alleles present.

- 2000 cells from person A who is 13, 19
- 760 cells from person B who is 13, 16

- 2760 copies of 13
- 2000 copies of 19
- 760 copies of 16



PRINCIPLES

Life is Beautiful!!

Interpretation is Easy

But there are storm clouds!

PRINCIPLES

"Easy" is hard to come by.

- *Why?*
- Multiple sources of variation
  - Different analysts (hands)
  - Different samples
  - Different pipettes
  - Different kit lots
  - Different instruments →

How do we increase consistency i.e. (reduce variation)?

1. One person does all the work
2. Quality assurance procedures
3. Protocols
4. Cute uniforms for DNA Analysts
5. Answers 2 and 3

Where do you think most of the amplification variation comes from?

1. Pipetting
2. Amplification chemistry
3. Injection and detection
4. Too much coffee before work

PRINCIPLES

Modified

Variation Summary

From the DNA Extract → Over Time

Amplification

Pipetting from extract
Sampling during amplification chemistry

Kit lots:
Component concentrations
Enzyme activity
Manufacturing processes
Company QC

Detection

Sample preparation for instrument; pipetting
Injection
Electrophoresis

Instrument changes over time:
Laser
CCD camera

Analysis

Protocol changes
New analysts ☺

See example in reference section under Peak Height Ratio: Debernardi et al. 2010

PRINCIPLES

How can we characterize variation?

- Look at total amount of variation at end of process
  - Follow the positive control over time
- Experimentally break process into components and characterize using appropriate statistics
  - E.g. Separate amplification variation from injection variation
- Analyze existing or new validation data, training sample data, SRM data, kit QC data
- Use casework data
  - E.g. Variation between knowns and matching single source evidence profiles

PRINCIPLES

Typical places we characterize variation associated with amplification

- % stutter
- Peak height
- Peak height ratio
- Calculated mixture ratio

PROTOCOLS

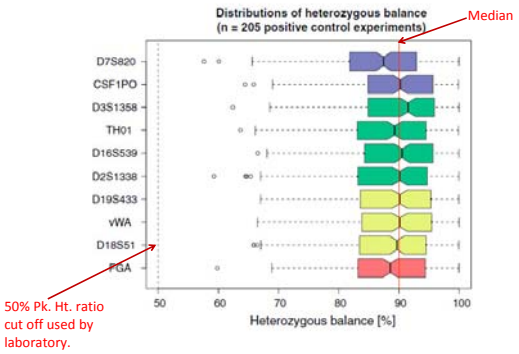
Example 1: Characterization of Variation in Positive Control

- From: "One year variability of peak heights, heterozygous balance and inter-locus balance for the DNA positive control of Identifiler STR Kit"
- This article looks at a compilation of data obtained at the end of the testing process
- Experimental design
  - 205 positive controls
    - Peak height
    - Heterozygous balance
    - Inter-locus balance

Debernardi, A., et al. (2010). One year variability of peak heights, heterozygous balance and inter-locus balance for the DNA positive control of AmpFlSTR Identifiler STR kit. *Forensic Science International: Genetics, (in press).* DOI: 10.1016/j.fsigen.2010.01.020

PROTOCOLS

Distribution of Peak height ratio



PROTOCOLS

Data from the positive control showed:

- High peak heights predicted ⟶ other peak heights would be high.
- ★ Peak heights and inter-locus balance varied more than peak height ratio.
- The two Genetic Analyzers (3130s) had different sensitivity in the blue channel.
- Heterozygous balance has no correlation with locus and "behaves as a random variable"… as "would be expected for a purely stochastic process…"

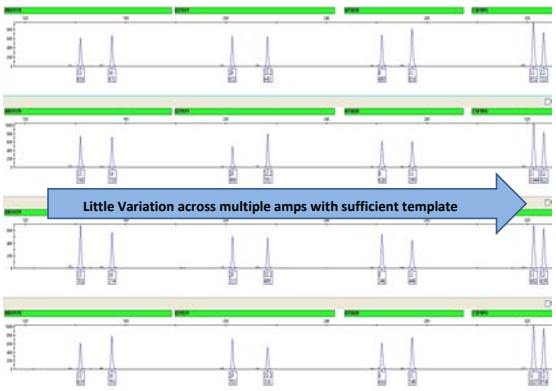
PROTOCOLS

Example 2: Design experiment to have minimum variation in amplification

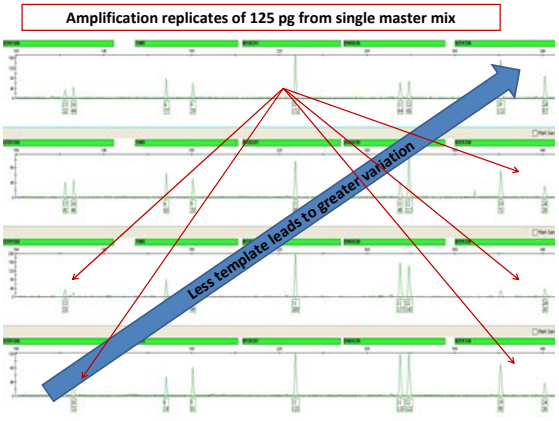
- Make single large master mix for each mass of DNA tested
  - DNA and master mix components added once and divided into multiple reaction tubes for amplification
- Therefore,
  - DNA concentrations are constant
  - master mix component concentrations are constant

BU data

Amplification replicates of 1 ng from single master mix



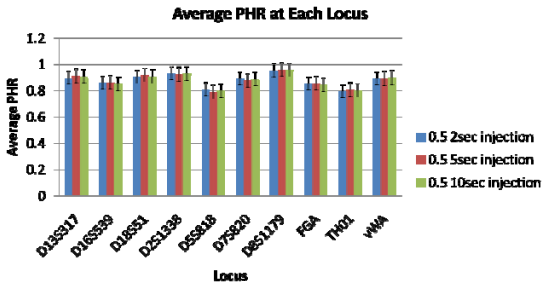
Amplification replicates of 125 pg from single master mix



PROTOCOLS

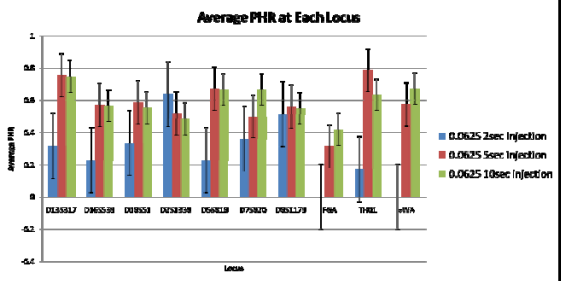
Modified

Four samples - 0.5 ng (500 pg)
Four replicate amplifications
2, 5 and 10 second injections



PROTOCOLS

Four samples- 0.0625 ng (62 pg)
Four replicate amplifications
Each sample injected at 2, 5 and 10 seconds



PROTOCOLS

In normal sample processing cannot avoid:

- Pipetting variation affect on:
  - Amount of template
  - Concentration of master mix components
- These concentration differences have a more profound affect when working with low template amounts

PROTOCOLS

Manage Variation: thru Rules

"The use of bounds applied to data that show continuous variation is common in forensic science and is often a pragmatic decision. However it should be borne in mind that applying such bounds has arbitrary elements to it and that there will be cases where the data lie outside these bounds."

From: Bright, Turkington and Buckleton, FSI Genetics, 2010

Bright, J.A., et al. (2010). Examination of the variability in mixed DNA profile parameters for the Identifiler multiplex. *Forensic Science International: Genetics, 4*, 111-114.

PROTOCOLS Modified

Managing Variation: Examples

- Analytical threshold
  - Single
  - Multiple set by amount of sample
  - Multiple set by dye emission color
- Stutter evaluation in %
  - Single value for entire profile
  - Single value for each locus
  - Value set by locus and size of allele within locus

PROTOCOLS Modified

Managing Variation: Examples

- Peak height ratio
  - One value above which two alleles can be considered to constitute a genotype
  - Two or more values used based on DNA mass
- Stochastic threshold
  - One threshold
  - Multiple thresholds reflecting different DNA amount
  - *Incorporate probability of drop out into likelihood ratio*

PRINCIPLES

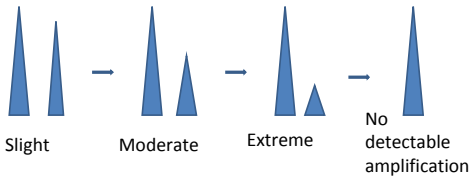
Stochastic Effects

- Allele drop out is an extension of the amplification disparity that is observed when heterozygous peaks heights are unequal

PRINCIPLES

Problem

- Occurs in single source samples and mixtures
- Analyst is unable to distinguish complete allele drop from a homozygous state



GUIDELINES

SWGDAM Definition of Stochastic

SWGDAM Guidelines glossary:

- **Stochastic effects**: the observation of intra-locus peak imbalance and/or allele drop-out resulting from random, disproportionate amplification of alleles in low-quantity template

PRINCIPLES

Broader Definition of Stochastic

- Stochastic is synonymous with "random." The word is of Greek origin and means "pertaining to chance". … Stochastic is often used as counterpart of the word "deterministic," which means that random phenomena are not involved. Therefore, stochastic models are based on random trials, while deterministic models always produce the same output for a given starting condition.
- http://mathworld.wolfram.com/Stochastic.html

PRINCIPLES

We cannot avoid stochastic effects and allele or locus drop out.

Why ?

WE DO NOT KNOW THE NUMBER OF CONTRIBUTORS TO A SAMPLE!

PRINCIPLES

Assume sample is a 1:3 mixture of two sources:

Amount of DNA	~ # of cells from major component	~ # of cells from minor component
1ng	107	36
0.5ng	53	18
0.25ng	27	9
0.125ng	12	4
0.0625ng	7	2

PRINCIPLES

Assume sample is a 1:9 mixture of two sources:

Amount of DNA	~ # of cells from major component	~ # of cells from minor component
1ng	129	14
0.5ng	64	7
0.25ng	32	4
0.125ng	16	2
0.0625ng	8	1

PRINCIPLES Modified

An overview of sampling:

(Modified outline from Gill et. al., NAR 2005)

Evidence has N cells

⬇

Extraction recovers n genome copies

⬇

DNA sample for PCR removes n_0 target copies

⬇

n_c amplicons of an allele are produced

⬇

n_{c1} amplicons are added to HiDi-LIZ sample prep

⬇

n_{c2} amplicons are injected & contribute to RFU

Gill, P., et al. (2005). A graphical simulation model of the entire DNA process associated with the analysis of short tandem repeat loci. *Nucleic Acids Research, 33,* 632-643.

PRINCIPLES

Accuracy in Amplification
From: Stenman and Orpana, Nature Biotechnology 2001

Table 1. True target copy number*

Target copy Number	95% Confidence interval Range low	Range high	Deviation
8	3	13	±62%
40	28	52	±30%
200	173	228	±14%
1,000	938	1,062	±6%

*Poisson-distribution-based estimates of the true target copy number in an aliquot pipetted from a highly diluted solution of known concentration.

Stenman, J., & Orpana, A. (2001). Accuracy in amplification. *Nature Biotechnology, 19,* 1011-1012.

PRINCIPLES

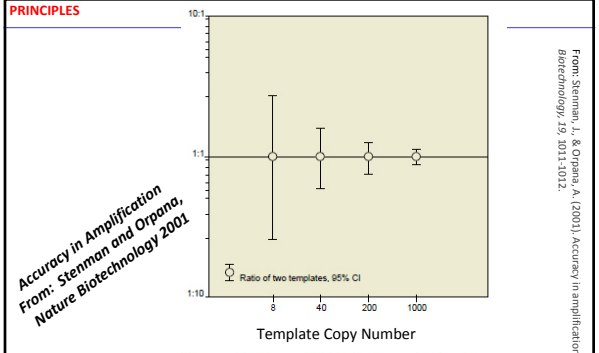
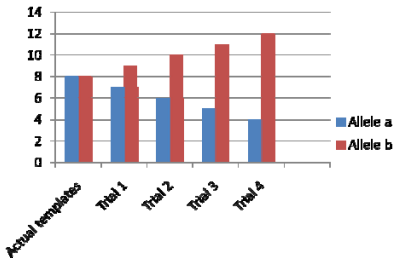


Figure 1. Poisson-distribution-based estimates of the ratio of two templates in an aliquot pipetted from a highly diluted solution containing the templates in a 1:1 ratio.

From: Stenman, J., & Orpana, A. (2001). Accuracy in amplification. *Nature Biotechnology, 19,* 1011-1012.

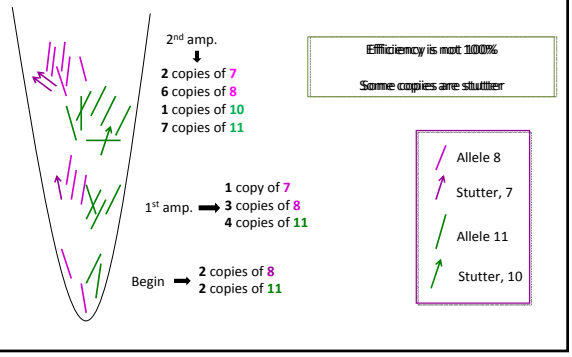
PRINCIPLES

Implications: Possible template amounts with replicate sampling



PROTOCOLS

Sampling during amplification occurs by imperfect success of making copies



PROTOCOLS Modified

Must experimentally define stochastic threshold(s) to address "drop out"

How do you develop data?

- Dilution series
- Replicate amplifications
- Casework data
  - Intimate samples
  - Replicates of evidence or known single-source samples

PRINCIPLES and PROTOCOLS

What is Drop Out?

- Scientifically
  - Failure to detect an allele within a sample or failure to amplify an allele during PCR. *From SWGDAM Guidelines, 2010*
  - Note that: Failure to detect ≠ failure to amplify
- Operationally
  - Setting a threshold(s) or creating a process, based on validation data and information in the literature, which allows assessment of the likelihood of drop out of an allele or a locus.

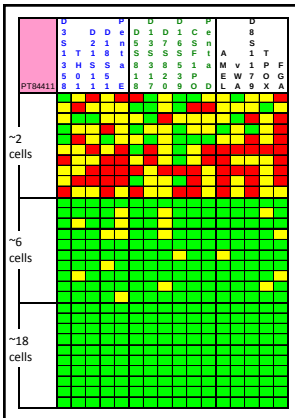
Developing data to make informed choice for stochastic threshold(s) determination

- Assuming some amplification of both peaks
  - Whether one or both peaks are observed will depend on:
    - the amount of amplified product injected
    - the analytical threshold used
- Look at different experimental approaches

PROTOCOLS

Data from NIST: Experimental Design

- Pristine DNA Samples
  - 2 single-source samples
  - **heterozygous for all loci tested** (permits peak height ratio studies)
- **Low DNA Temple Amounts**
  - Dilutions made after DNA quantitation against NIST SRM 2372
  - **100 pg, 30 pg, and 10 pg** (1 ng tested for comparison purposes)
- Replicates
  - **10 separate PCR reactions** for each sample
- STR Kits
  - **Identifiler and PowerPlex 16 HS** (half-reactions)
- **Increased Cycle Number**
  - Identifiler (**31 cycles**; 28 for 1 ng)
  - PowerPlex 16 HS (**31 cycles and 34 cycles**; 30 for 1 ng)



10 replicates of each [DNA]

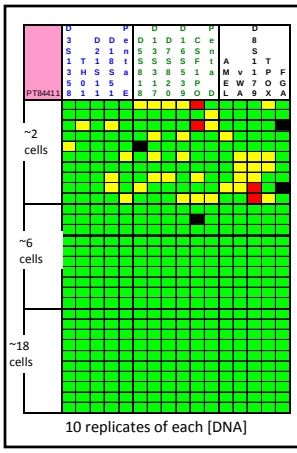
NIST data:
PowerPlex 16 HS
31 cycles

10 pg → 10 multiple problem profiles

30 pg
- 2 with 0 allele drop outs
- 2 with 1 allele drop out
- 2 with 2 allele drop outs
- 2 with 3 allele drop outs
- 1 with 4 allele drop outs

100 pg → 10 profiles with 0 allele drop outs

Green = both alleles observed
Yellow = allele dropout
Red = locus dropout



10 replicates of each [DNA]

NIST data:
PowerPlex 16 HS
34 Cycles

10 pg → 1 with 1 drop out
8 with multiple drop outs
3 with drop in

30 pg → One locus with drop in

100 pg → 10 profiles with 0 allele drop outs

Green = both alleles observed
Yellow = allele dropout
Red = locus dropout
Black = drop-in

Drop Out Probability as a Function of Present Allele Height From Gill et al., 2008 FSI Genetics

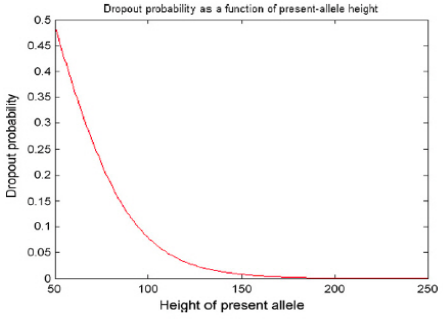


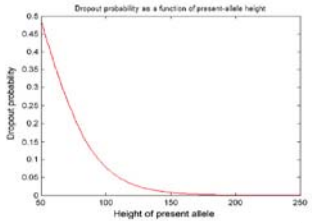
Fig. A.2. Probability drop-out as a function of present-allele height Pr(*D*|*h*).

Gill, P., et al. (2008). The *low-template* (stochastic) threshold-Its determination relative to risk analysis for national DNA databases. FSI Genetics, 104-111.

PROTOCOLS Modified

Data represents a probability distribution:
The smaller rfu of the observed peak the **greater probability** that drop out has occurred

Use data to:
Develop a single threshold
Develop multiple thresholds for different applications



With computer assistance or expert systems can use the probability distribution

PROTOCOLS

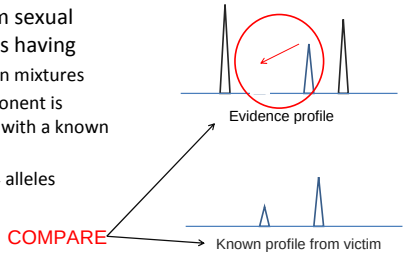
Can use some types of casework samples to look at drop out

- Provides examples of drop out observed in sample which you are likely to encounter
- Provides additional data to consider when setting thresholds

PROTOCOLS Modified

Can examine casework data from known heterozygous loci using:

- Profiles from sexual assault cases having
  - two person mixtures
  - one component is consistent with a known victim
  - loci with 4 alleles



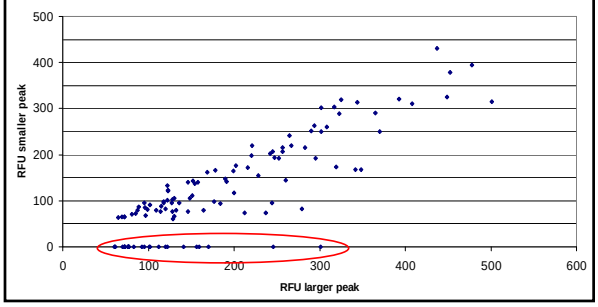
PROTOCOLS

Looking for Lost Alleles

- Casework samples exhibiting drop-out at a peak-calling threshold of 60 rfu were reanalyzed at 30 rfu to determine:
  - when no measurable amplification was truly occurring
  - when the dropped peak was simply below the 60 rfu peak-calling threshold.
- If an allele was not detected at 30 rfu = no amplification

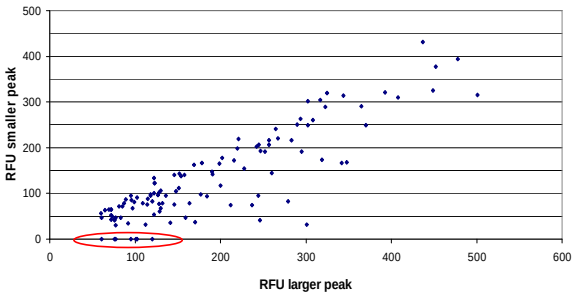
PROTOCOLS

Allelic dropout at peak-calling threshold of 60 rfu



PROTOCOLS

Allelic dropout at peak-calling threshold of 30 rfu



PROTOCOLS

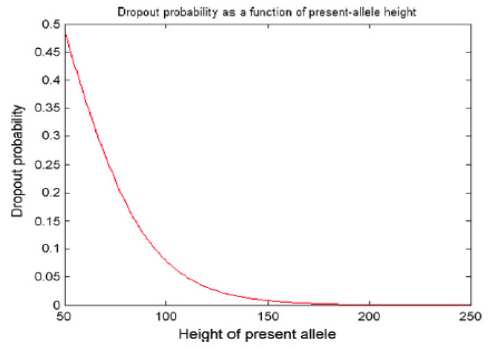
Comparison of data with rfu threshold set at 60 and 30 rfu:

- 72 % of peaks lost using a threshold of 60 rfu were observed between 60 and 30 rfu
- These peaks represents uneven amplification but not complete absence of amplification

Dropout percentages by larger minor peak height intensity (threshold: 60 rfu)

casework data				casework and validation data			
RFU Range:	Dropouts	Samples	% dropout (100d/n)	RFU Range:	Dropouts	Samples	% dropout (100d/n)
60-100	18	30	60.0%	60-100	29	61	47.5%
101-150	6	28	21.4%	101-150	13	74	17.6%
151-200	3	16	18.8%	151-200	4	64	6.3%
201-250	1	12	8.3%	201-250	1	46	2.2%
251-300	0	11	0.0%	251-300	0	42	0.0%
301-350	1	11	9.1%	301-350	1	30	3.3%
350-400	0	3	0.0%	350-400	0	19	0.0%
>400	0	6	0.0%	>400	0	47	0.0%
Total	29	117			48	383	

Considering this data…



Gill, P., et al. (2008). The *low-template* (stochastic) threshold-Its determination relative to risk analysis for national DNA databases. FSI Genetics, 104-111.

What value will you choose for today's stochastic threshold?

1. 100 RFU
2. 150 RFU
3. 200 RFU
4. 250 RFU
5. None of the above