Addressing STR Data Interpretation Concerns

John M. Butler, PhD

October 11, 2006
MAFS Workshop
(Indianapolis, IN)

NIST and NIJ Disclaimer

Funding: Interagency Agreement 2003-IJ-R-029 between the National Institute of Justice and NIST Office of Law Enforcement Standards

Points of view are those of the author and do not necessarily represent the official position or policies of the US Department of Justice or the National Institute of Standards and Technology.

Certain commercial equipment, instruments and materials are identified in order to specify experimental procedures as completely as possible. In no case does such identification imply a recommendation or endorsement by the National Institute of Standards and Technology nor does it imply that any of the materials, instruments or equipment identified are necessarily the best available for the purpose.

Understanding the Audience Here

- Where is everyone from?
  - Which states?
  - State lab?
  - Local lab?
  - Private lab?
- Experience level?
  - Less than 1 year?
  - 1-3 years?
  - >3 years?
- STR kits in use?
  - Profiler Plus/COfiler
  - Identifiler
  - PowerPlex 16
- Instrumentation is use?
  - ABI 310
  - ABI 3100/3130xl
- Software in use?
  - GeneScan/Genotyper
  - GeneMapperID

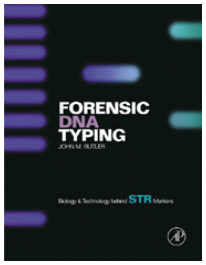
This one's from NIST…

My Background

- PhD **(Analytical Chemistry)** from University of Virginia (Aug 1995)
- Research conducted at FBI Academy under Bruce McCord doing CE for STR typing
- NIST Postdoc – developed STRBase website
- GeneTrace Systems – private sector experience validating assays
- NIST Human Identity Project Leader since 1999
- Invited guest to SWGDAM since 2000
- Member of SWGDAM Validation Subcommittee
- Served on WTC KADAP and helped evaluate and validate new miniSTR, mtDNA, and SNP assays
- Author of *Forensic DNA Typing: Biology, Technology, and Genetics of STR Markers* (2nd Edition)
- Married with 6 children – I have "validated" that they are mine using STR typing…

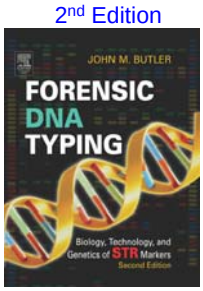
The Field Is Growing and Changing Rapidly…

1st Edition



Jan 2001
335 pp.

2nd Edition



Feb 2005
688 pp.

3rd Edition

If my wife lets me write it…

2009
~1000 pages

A Few Thoughts from What Has Been Discussed Today

- **LOD vs LOQ** – instruments will differ yet lab-wide SOPs are used for practical purposes; statistical vs. empirical
- **Measuring noise** – is it practical to do so? Relationship to stochastic effects? Theory vs. practical application?
- **Setting a threshold** – will it remain constant over time? (value of Multiplex_QA if you want to measure)
- **Chemical artifacts** – why do blobs and spikes occur? Troubleshooting and improving data quality…
- **Low-levels of DNA** – issues with LCN, data quality changes at low levels
- **Potential for human error** – how do you measure it? Will it be constant over time?
- **Case context** – DNA results do not come in a vacuum; some injections will fail and samples be re-injected

AAFS Workshop #6 (Feb 2006, Seattle)

Advanced Topics in STR DNA Analysis

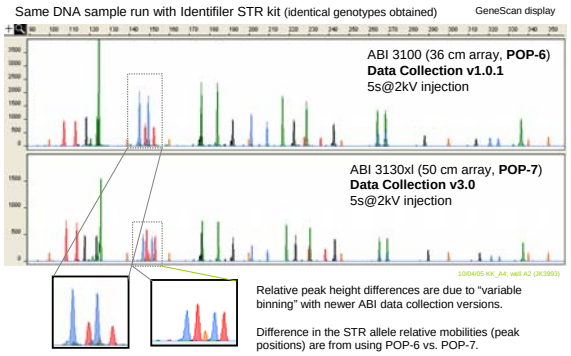
Instructors: John Butler and Bruce McCord

For DNA analysts using the ABI 310 or ABI 3100 who would like to better understand the underlying issues and science involved with STR DNA typing

- **STR Biology, Markers, and Methods**
- **Capillary Electrophoresis Instrumentation: Theory and Application**
- **Validation Aspects to Consider in Bringing a New STR Kit "On-line"**
- **CE Troubleshooting**
- **STR Mixture Interpretation**
- **DNA Quantitation with Real-Time qPCR**
- **Low-copy Number Issues**
- **Y-STRs and mtDNA**

Available at http://www.cstl.nist.gov/biotech/strbase/training.htm

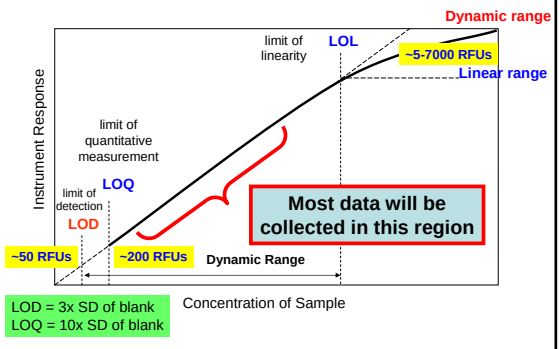
Comparison of ABI 3100 Data Collection Versions



Relative peak height differences are due to "variable binning" with newer ABI data collection versions.

Difference in the STR allele relative mobilities (peak positions) are from using POP-6 vs. POP-7.

Useful Range of an Analytical Method



Adapted from Figure 1-7 in Skoog, D.A., *et al.* (1998) *Principles of Instrumental Analysis* (5th Edition). Thomson Learning, Inc.

Why are empirical thresholds used?

- GeneScan/Genotyper and GeneMapperID do not permit analysis of the noise

BatchExtract

BatchExtract

- **Developed by Stephen Sherry's group** at the National Center for Biotechnology Information (NCBI) – same branch of NIH doing PubMed and GenBank
- Available at from **ftp://ftp.ncbi.nih.gov/pub/forensics/BATCHEXTRACT**
- Came out of World Trade Center data review
- Front end for OSIRIS (an expert system under development)
- Outputs are as ASCII data

http://www.ncbi.nlm.nih.gov/IEB/Research/GVWG/OSIRIS/index.htm

Identifying Victims of Mass Disasters

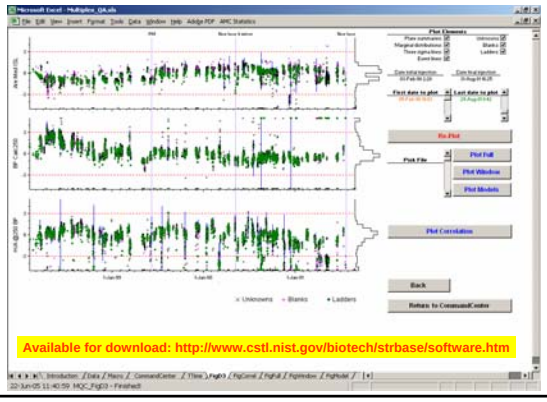
Butler, J.M. (2005) *Forensic DNA Typing, 2nd Edition*, Chapter 24

***Science* (2005) 310: 1122-1123**

Highly Degraded DNA Was Obtained from the Human Remains Recovered

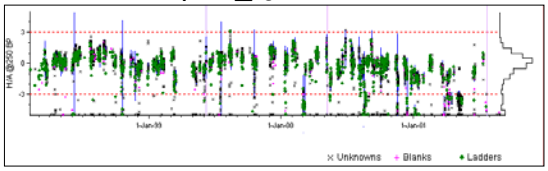
Largest Forensic Case in History
~20,000 bone fragments were processed
>6,000 family reference samples and personal effects samples were analyzed

NIST Multiplex_QA Program for Monitoring Performance Over Time



Available for download: http://www.cstl.nist.gov/biotech/strbase/software.htm

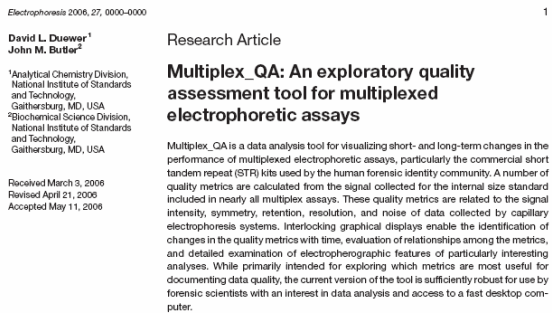
Multiplex_QA Overview



- **Research tool** that provides quality metrics to review instrument performance over time (e.g., examines resolution on internal size standard peaks)
- Runs with Microsoft Excel macros. Requires STR data to be converted with NCBI's BatchExtract program into numerical form.

Available for download from STRBase:
http://www.cstl.nist.gov/biotech/strbase/software.htm

Multiplex_QA Article in press



Electrophoresis 2006, 27, 0000–0000 1

David L. Duewer[1]
John M. Butler[2]

[1]Analytical Chemistry Division, National Institute of Standards and Technology, Gaithersburg, MD, USA
[2]Biochemical Science Division, National Institute of Standards and Technology, Gaithersburg, MD, USA

Received March 3, 2006
Revised April 21, 2006
Accepted May 11, 2006

Research Article

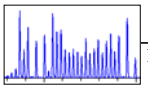
Multiplex_QA: An exploratory quality assessment tool for multiplexed electrophoretic assays

Multiplex_QA is a data analysis tool for visualizing short- and long-term changes in the performance of multiplexed electrophoretic assays, particularly the commercial short tandem repeat (STR) kits used by the human forensic identity community. A number of quality metrics are calculated from the signal collected for the internal size standard included in nearly all multiplex assays. These quality metrics are related to the signal intensity, symmetry, retention, resolution, and noise of data collected by capillary electrophoresis systems. Interlocking graphical displays enable the identification of changes in the quality metrics with time, evaluation of relationships among the metrics, and detailed examination of electropherographic features of particularly interesting analyses. While primarily intended for exploring which metrics are most useful for documenting data quality, the current version of the tool is sufficiently robust for use by forensic scientists with an interest in data analysis and access to a fast desktop computer.

NIST QA/QC Software

Tool being developed by Dave Duewer for STR Process Control

Tracks allelic ladders and positive controls and internal size standards



NCBI Program

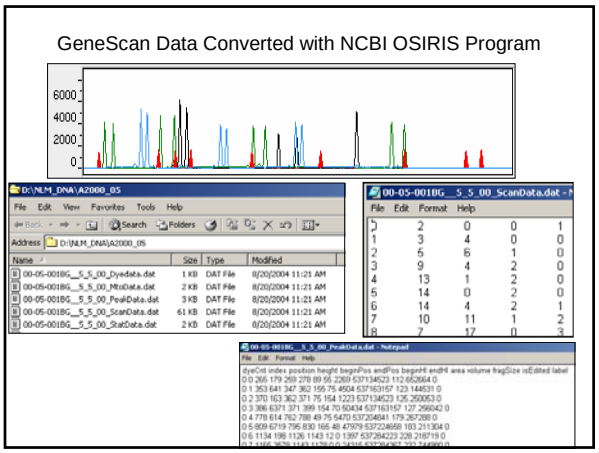
Peak Height, Area, Size

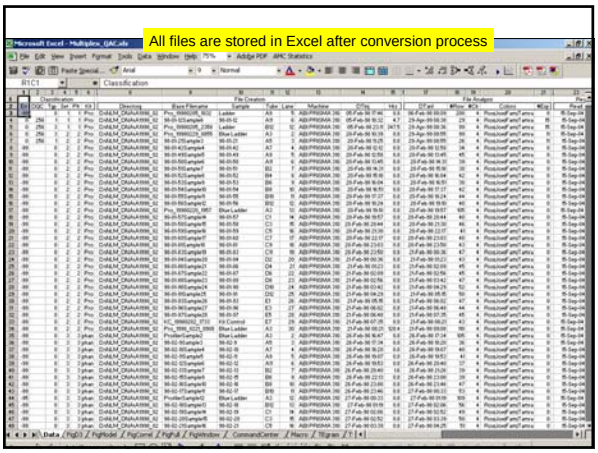
X | Y
103 | 436
104 | 569

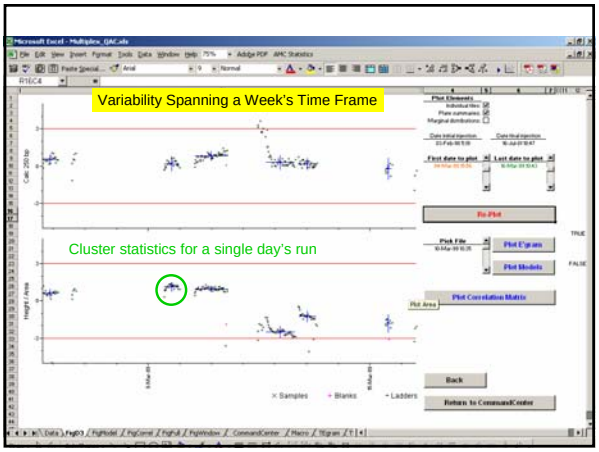
NIST ControlCharter → Date vs Sensitivity, Resolution, Precision

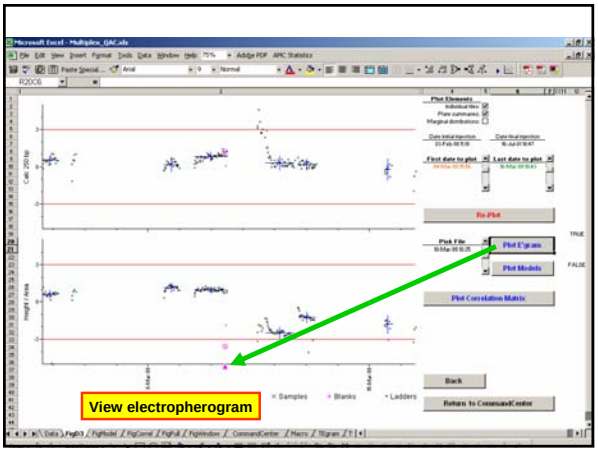
NIST ControlCharter → Date vs Signal/Noise

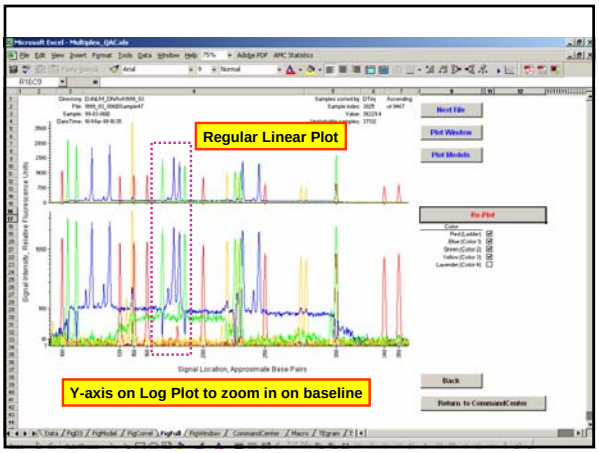
This software does not perform genotyping.
It merely permits a view of analytical parameters over time.











Pete Vallone Dave Duewer

Software Tools from NIST

- AutoDimer – multiplex PCR primer screening tool
  http://www.cstl.nist.gov/biotech/strbase/AutoDimerHomepage/AutoDimerProgramHomepage.htm
- mixSTR – mixture component resolution tool
- **Multiplex_QA** – quality assessment tool for monitoring instrument performance over time
- Tools to aid Expert System data review
  - DNA_FSSi3_Convert.xls (converts data format)
  - STR_MatchSamples.xls (compares samples)

  http://www.cstl.nist.gov/biotech/strbase/software.htm

DNA Testing Requires a Reference Sample

A DNA profile by itself is fairly useless because it has no context…

DNA analysis for identity only works by comparison – **you need a reference sample**

Crime Scene Evidence compared to **Suspect(s)** (Forensic Case)
Child compared to **Alleged Father** (Paternity Case)
Victim's Remains compared to **Biological Relative** (Mass Disaster ID)
Soldier's Remains compared to **Direct Reference Sample** (Armed Forces ID)

Issues with Low Amounts of DNA

MAAFS DNA Workshop

Introduction to Low Copy Number (LCN) DNA Testing Issues

John M. Butler, PhD

john.butler@nist.gov
NIST

Richmond, VA
May 3, 2006

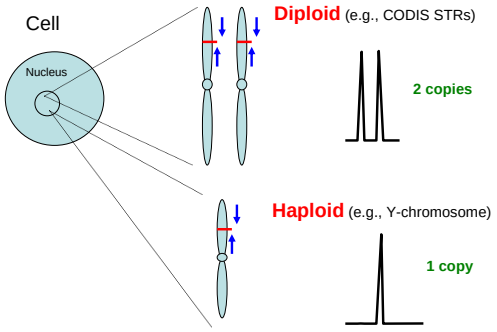
Some Definitions of Low-Copy Number (LCN)

- Work with **<100 pg genomic DNA** (~15-17 diploid copies of nuclear DNA markers such as STRs)
- Below stochastic threshold level where PCR amplification is not as reliable (determined by each laboratory; typically 150-250 pg)
- Enhancing sensitivity of detection (34 cycles instead of 28 cycles)
- Too few copies of DNA template to ensure reliable PCR amplification
- Other terms for LCN:
  - Low-level DNA
  - Trace DNA
  - Touch DNA

LCN is dependent on the amount of DNA present NOT the number of PCR cycles performed; LCN conditions may exist with 28 or 34 cycles

DNA quantity in samples

Diploid vs. Haploid



Calculation of the Quantity of DNA in a Cell

1. **Molecular Weight of a DNA Base Pair = 618 g/mol**

   A = 313 g/mol; T = 304 g/mol; A-T base pairs = 617 g/mol
   G = 329 g/mol; C = 289 g/mol; G-C base pairs = 618 g/mol

2. **Molecular Weight of DNA = 1.98 x10^{12} g/mol**

   There are **3.2 billion base pairs** in a haploid cell ~3.2 x 10^9 bp
   (~3.2 x 10^9 bp) x (618 g/mol/bp) = 1.98 x 10^{12} g/mol

3. **Quantity of DNA in a Haploid Cell = 3 picograms**

   1 mole = 6.02 x 10^{23} molecules
   (1.98 x 10^{12} g/mol) x (1 mole/6.02 x 10^{23} molecules)
   = 3.3 x 10^{-12} g = 3.3 picograms (pg)
   A diploid human cell contains ~6.6 pg genomic DNA

4. **One ng of human DNA comes from ~152 diploid cells**

   1 ng genomic DNA (1000 pg)/6.6pg/cell = **~303 copies of each locus**
   (2 per 152 diploid genomes)

Adapted from D.N.A. Box 3.3, J.M. Butler (2005) Forensic DNA Typing, 2nd Edition (Elsevier Academic Press), p. 56

At the 2003 AAFS LCN Workshop (Chicago,IL), **Robin Cotton** from Orchid Cellmark presented a talk entitled "Are we already doing low copy number (LCN) DNA analysis?"

Where does low copy number start?

<100 pg template DNA

(Butler, 2001, Fregeau & Fourney 1993, Kimpton *et al* 1994)

Amount of DNA	~ # of cells
1 ng	152
0.5 ng	76
0.25 ng	38
0.125 ng	19
0.0625 ng	10

Values for # of cells adjusted to reflect updated DNA quantitation numbers

Robin Cotton, AAFS 2003 LCN Workshop
"Are we already doing low copy number (LCN) DNA analysis?"

Assume sample is from a **single source**:

Amount of DNA	Total Cells in sample	~ # of copies of each allele if het.
1 ng	152	152
0.5 ng	76	76
0.25 ng	38	38
0.125 ng	19	19
0.0625 ng	10	10

Robin Cotton, AAFS 2003 LCN Workshop
"Are we already doing low copy number (LCN) DNA analysis?"

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

Amount of DNA	Total Cells in sample	~ # of cells from each component
1 ng	152	76
0.5 ng	76	38
0.25 ng	38	19
0.125 ng	19	10
0.0625 ng	10	5

Robin Cotton, AAFS 2003 LCN Workshop
"Are we already doing low copy number (LCN) DNA analysis?"

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

Amount of DNA	~ # of cells from **major component**	~ # of cells from **minor component**
1 ng	114	38
0.5 ng	57	19
0.25 ng	28	10
0.125 ng	14	5
0.0625 ng	7	2

Robin Cotton, AAFS 2003 LCN Workshop
"Are we already doing low copy number (LCN) DNA analysis?"

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

Amount of DNA	~ # of cells from **major component**	~ # of cells from **minor component**
1ng	137	15
0.5ng	68	8
0.25ng	34	4
0.125ng	17	2
0.0625ng	9	1

Robin Cotton, AAFS 2003 LCN Workshop
"Are we already doing low copy number (LCN) DNA analysis?"

Low copy number situations exist in many samples

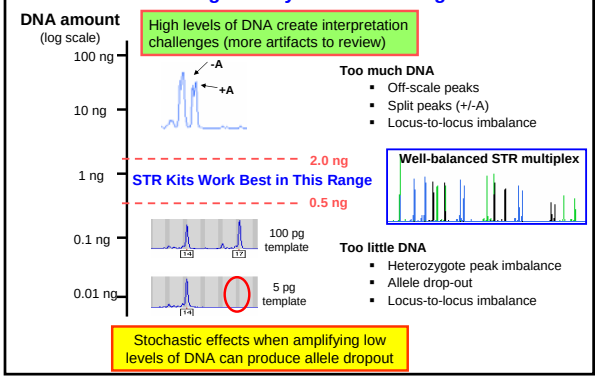
- In a 1:1 mixture, each DNA source is at LCN when the total amount of DNA in the amplification reaction is ~ 0.125 ng.
- In a 1:9 mixture, the minor component could be at LCN **even when the total amount of DNA in the amplification is 1 ng**.

Two different amplifications would be useful with a 1:9 mixture situation:
Normal level of total DNA (e.g., 1 ng) so that major component is on-scale
High level of total DNA (e.g., 5 ng) so that minor (e.g., ~500 pg) is out of LCN realm – yes, the major component will be off-scale…

Robin Cotton, AAFS 2003 LCN Workshop
"Are we already doing low copy number (LCN) DNA analysis?"

Impact of DNA Amount into Multiplex PCR Reaction

We generally aim for 0.5-2 ng



Stochastic PCR amplification

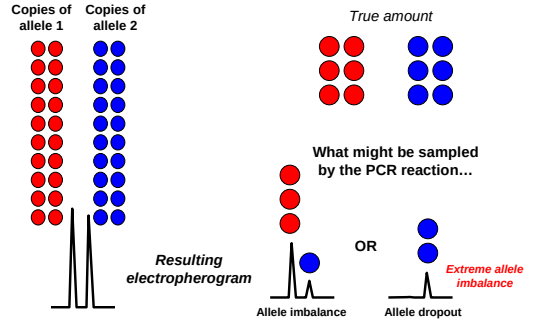
Stochastic = random selection

Stochastic Fluctuation Effects

- Unequal sampling of the two alleles present in a heterozygous individual can occur when low levels of input DNA are used (**results in allele drop-out**)
- PCR reactions **with <100 pg** (~17 diploid copies)
- Walsh *et al.* (1992) – propose avoiding stochastic effect by adjusting the number of PCR cycles in an assay so that the sensitivity limit is around 20 or more copies of target DNA (i.e., a full profile is obtained with ~125 pg)

Walsh PS, Erlich HA, Higuchi R. Preferential PCR amplification of alleles: Mechanisms and solutions. *PCR Meth Appl* 1992; 1:241-250.

Stochastic Statistical Sampling

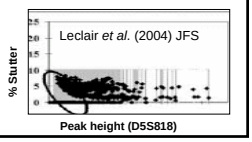


>20 copies per allele

6 copies copies per allele **(LCN)**

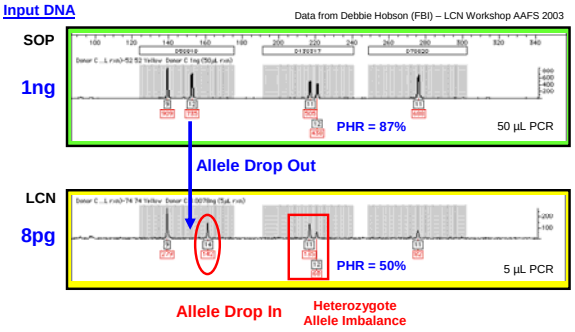
Stochastic Effect

- Sometimes called "preferential amplification" – not really a correct term since either allele may be amplified if the other drops-out…not related to allele size
- Stutter product amounts may go up…
  - If in an early cycle of PCR, the stutter product is amplified more (due to sampling effect)
- Contaminating DNA can also be amplified giving rise to allele "drop-in" or a mixture



Comparison of STR Kit Amplification SOP with LCN

Using the Same DNA Donor

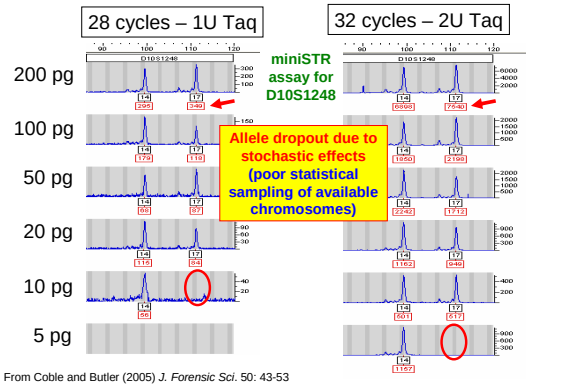


Balance of Assay Sensitivity and Potential for Stochastic Effects

- **One of the ways that assays can be made more sensitive is by increasing the number of PCR amplification cycles**
- Optimal cycle number will depend on desired assay sensitivity
- The number of PCR cycles was set to 28 for ABI STR kits to limit their sensitivity for generating full profiles to ~125 pg or 20 cells
- Sensitivity is a combination of fluorescent dye characteristics (relative to the instrument and laser excitation used) and PCR amplification conditions such as primer concentration and amount of polymerase used

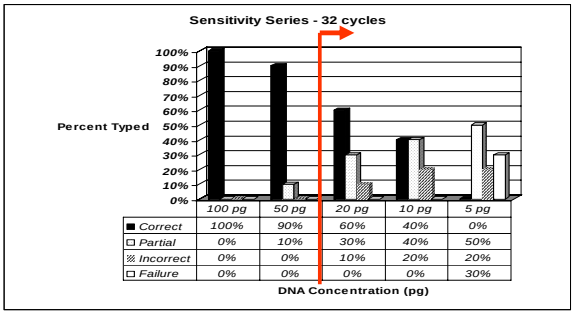
Note that Promega STR kits use higher numbers of cycles to generate roughly equivalent sensitivity to ABI kits because they have less efficient dye labels and lower primer and polymerase concentrations

Higher Sensitivity with More Polymerase and Cycle Numbers



From Coble and Butler (2005) *J. Forensic Sci.* 50: 43-53

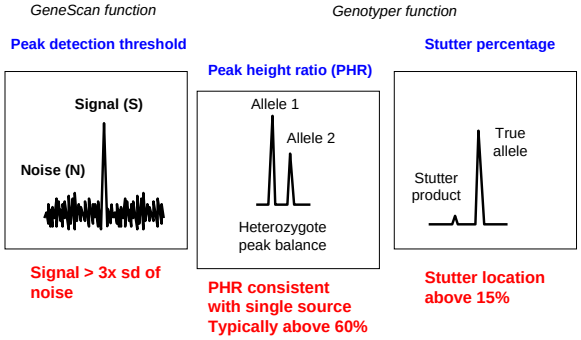
Problems with Obtaining Correct Allele Calls at Low DNA Levels



	100 pg	50 pg	20 pg	10 pg	5 pg
Correct	100%	90%	60%	40%	0%
Partial	0%	10%	30%	40%	50%
Incorrect	0%	0%	10%	20%	20%
Failure	0%	0%	0%	0%	30%

Coble, M.D. and Butler, J.M. (2005) *J. Forensic Sci.* 50: 43-53

What is a true peak (allele)?



Threshold Settings for the ABI 310/3100

Detection Limit: 3x the standard deviation of the noise.
Estimated using **2x peak to peak noise**. (approximately 35 - 50 RFUs)

Limit of Quantitation: 10x the standard deviation of the noise
Estimated using **7x peak to peak noise** (150-200 RFUs)
Below this point estimates of peak area or height are unreliable.

Dynamic Range: The range of sample quantities that can be analyzed from the lowest to the highest **(linear range is also important)**

Stochastic Threshold: Level of quantifiable DNA below which peaks can show severe imbalance (peak height ratios below 60%) Approximately 150 -200 RFUs. Enhanced stutter also occurs at these signal levels.

The Scientific Reasoning behind the Concept of an Analytical Threshold (limit of detection)

- This is fundamentally an issue of reliability
- **For a peak intensity three times the standard deviation of the noise there is a limited chance that such a signal is the result of a random fluctuation**
- This is because 99.7 percent of all noise signals fall below this value (from the definition of a Gaussian curve)
- **Below this point the very real possibility exists that what you think is a peak is simply a statistical fluctuation in the baseline noise**.

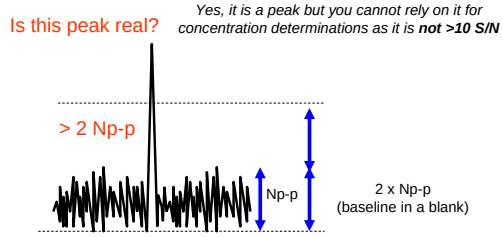
Sensitivity

- Limit of detection (**LOD**) – "the lowest content that can be measured with reasonable statistical certainty."
- Limit of quantitative measurement (**LOQ**) – "the lowest concentration of an analyte that can be determined with acceptable precision (repeatability) and accuracy under the stated conditions of the test."
- How low can you go?

EURACHEM Guide (1998) *The Fitness for Purpose of Analytical Methods: A Laboratory Guide to Method Validation and Related Topics*, p. 43; available at http://www.eurachem.ul.pt/guides/valid.pdf

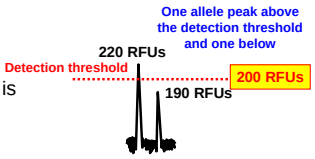
Limit of Detection (LOD)

- Typically 3 times the signal-to-noise (based on standard deviation of the noise) or 2x Np-p

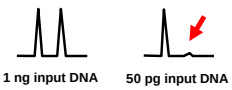


Types of Results at Low Signal Intensity (Stochastic amplification potential)

Straddle Data
- Only one allele in a pair is above the laboratory stochastic threshold



Allelic Drop-out
- one or more sets of alleles do not amplify



TWGDAM validation of AmpFlSTR Blue

Wallin *et al.* (1998) *J. Forensic Sci.* 43(4): 854-870

- Minimum cycle # (27-30 cycles examined)
- **Amplification adjusted to 28 cycles so that quantities of DNA below 35pg gave very low peaks or no peaks** (below the analytical threshold!)
- 35 pg is approx 5 cells
- (but is 35pg the analytical threshold?) Determining this value might be a useful goal of a validation study

TWGDAM validation of AmpFlSTR Blue

Wallin *et al.* (1998) *J. Forensic Sci.* 43(4): 854-870

Determination of Minimum Sample

- Goal: avoid situations where peak imbalance results in only one detectable allele from a heterozygous pair.
- Perform serial dilution (1ng- 8pg) of 2 control samples which were heterozygous at all 3 loci
  - **Samples above 125pg had peak height RFUs above 150**
  - **Below 125pg peak heights were not significantly above background**
  - **At 31 pg peaks were very low or undetectable**
- **"Peaks below 150 RFU should be interpreted with caution"** **Why? Noise and stochastic fluctuation!**

Sensitivity of Detection

Moretti et al, JFS, 2001, 46(3), 661-676

- Different 310 instruments have different sensitivities; determination of stochastic threshold should be performed following in-house studies
  - Variations in quantitation systems
  - Variations in amplification systems
  - Variations in instrument sensitivity
- Peaks with heights below the threshold should be interpreted with caution
  - Caution should be used before modification of
    - Amplification cycles
    - Electrophoretic conditions

How to determine the stochastic threshold

- Examine intensity and peak height ratio of 5 samples at three different low concentrations (e.g., 60, 75, and 125 pg)
- Observe variation in peak height ratio and peak intensity
- The stochastic threshold is the point at which this variation begins a rapid increase (change in slope of line relating std dev vs concentration)
- This can also be defined as the concentration at which a set percentage of peak height ratio values fall below 60%

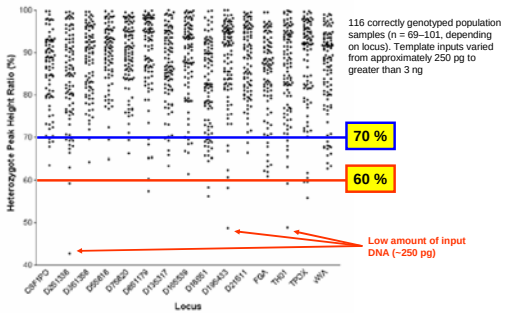
TWGDAM validation of AmpFlSTR BluePCR

Wallin et al.JFS, 1998 43(4) 854-870

- In approximately 80 heterozygous loci in population samples:
  - Average peak height ratio was 92% for each locus – D3, vWA, FGA
  - Standard deviation was 7%
- Thus 99.7% of all samples should show a peak height ratio (PHR) above 71%
- Those that have a PHR of <70% may result from mixtures, low [DNA], inhibition, degradation or poor primer binding

Heterozygote Peak Height Ratios

Identifiler STR Kit Developmental Validation



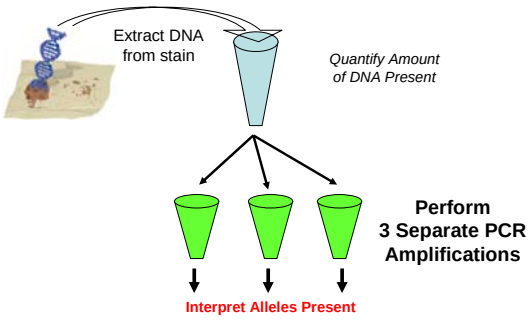
Collins PJ, Hennessy LK, Leibelt CS, Roby RK, Reeder DJ, Foxall PA. Developmental validation of a single-tube amplification of the 13 CODIS STR loci, D2S1338, D19S433, and amelogenin: the AmpFlSTR Identifiler PCR amplification kit. *J. Forensic Sci.* 2004; 49(6): 1265-1277.

Peak height ratios

Moretti et al., JFS 2001, 46(3) 647-660

- PP + Cofiler gave PHR >88% n= 230+ samples with a lower range PHR (-3sd) of 59%
- Suggest using 59% as a guide
- 2% of single source samples were below this value
- Many validation studies focus on 1ng input DNA. What happens with lower amounts?

Typical LCN Procedure



Develop a Consensus Profile
(based on replicate consistent results)

New Interpretation Rules Required for LCN

ELSEVIER
Forensic Science International
112 (2000) 17–40
Forensic Science International
www.elsevier.com/locate/forsciint

An investigation of the rigor of interpretation rules for STRs derived from less than 100 pg of DNA

Peter Gill[a,*], Jonathan Whitaker[a], Christine Flaxman[a], Nick Brown[a], John Buckleton[b]

Forensic Science Service, Priory House, Gooch Street North, Birmingham B56QQ, UK
[b]ESR, Private Bag 92021, Auckland, New Zealand

Received 9 December 1999; received in revised form 12 February 2000; accepted 13 February 2000

Suggestions to Optimal Results with LCN

- At least two* PCR amplifications from the same DNA extract (if enough DNA is present to do more than 4-5 amplifications, then most likely a single aliquot would be run under standard STR typing conditions)
- An allele cannot be scored (considered real) unless it is present at least twice in replicate samples
- Extremely sterile environment is required for PCR setup to avoid contamination from laboratory personnel or other sources

*five is better; results are investigative

Other methods for higher sensitivity and signal enhancements

Improving Sensitivity

- Improved recovery of biological material and DNA extraction
- Longer injection on CE
- Salt removal from CE sample – enhances electrokinetic injection
- Reduced volume PCR – concentrates amplicon
- Increase number of cycles in PCR and/or TaqGold concentration
- Use miniSTRs – shorter amplicons amplify better
- Use mtDNA – higher copy number per cell

Modifications in DNA Analysis Process to Improve LCN Success Rates

- **Collection** – better swabs for DNA recovery
- **DNA Extraction** – into smaller volumes
- **DNA Quantitation** – qPCR helps with low DNA amounts
- **PCR Amplification** – increased number of cycles
- **CE Detection** – longer electrokinetic injection; more sensitive fluorescent dyes
- **Interpretation** – composite profile from replicate analyses with at least duplicate results for each reported locus
- **Match** – is it even relevant to the case?

The Wisdom of Obi Wan Kenobi

http://www.starwars.com/kids/explore/lore/img/news20000902_1.jpg

Just before entering the Mos Eisley spaceport in Episode IV, Ben (Obi Wan) Kenobi warned Luke Skywalker, "You will never find a more wretched hive of scum and villainy…

WE MUST BE CAUTIOUS!"

Thank you for your attention…

Questions?

http://www.cstl.nist.gov/biotech/strbase
john.butler@nist.gov
301-975-4049

Our team publications and presentations are available at:
http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm