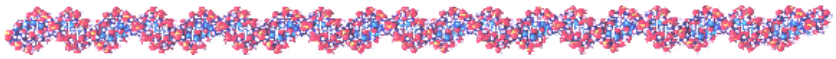
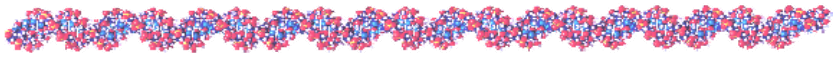


Validation



John M. Butler, PhD
National Institute of Standards and Technology

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

Webinar
August 26, 2008

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 mine 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.

What is **Validation** and Why Should It Be Done?

- Part of overall quality assurance program in a laboratory
- **We want the correct answer when collecting data…**
  - We want **analytical measurements made in one location to be consistent with those made elsewhere** (without this guarantee there is no way that a national DNA database can be successful).
- If we fail to get a result from a sample, we want to have confidence that the sample contains no DNA rather than there might have been something wrong with the detection method…

Want no false negatives…

Why is Method Validation Necessary?

- It is an important element of quality control.
- Validation helps provide assurance that a measurement will be reliable.
- In some fields, validation of methods is a regulatory requirement.
- …
- The validation of methods is **good science**.

Roper, P., *et al.* (2001) *Applications of Reference Materials in Analytical Chemistry*. Royal Society of Chemistry, Cambridge, UK, pp. 107-108.

Definition of Validation

- **Validation** is confirmation by examination and provision of objective evidence that the particular requirements for a specified intended use are fulfilled.
- **Method validation** is the process of establishing the performance characteristics and limitations of a method and the identification of the influences which may change these characteristics and to what extent. It is also the process of verifying that a method is fit for purpose, i.e., for use for solving a particular analytical problem.

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

Definitions

J.M. Butler (2005) *Forensic DNA Typing*, 2nd Edition, p. 389, 391

- **Quality assurance** (QA) – planned or systematic actions necessary to provide adequate confidence that a product or service will satisfy given requirements for quality
- **Quality control** (QC) – day-to-day operational techniques and activities used to fulfill requirements of quality
- **Validation** – the process of demonstrating that a laboratory procedure is **robust**, **reliable**, and **reproducible** in the hands of the personnel performing the test in that laboratory

Definitions

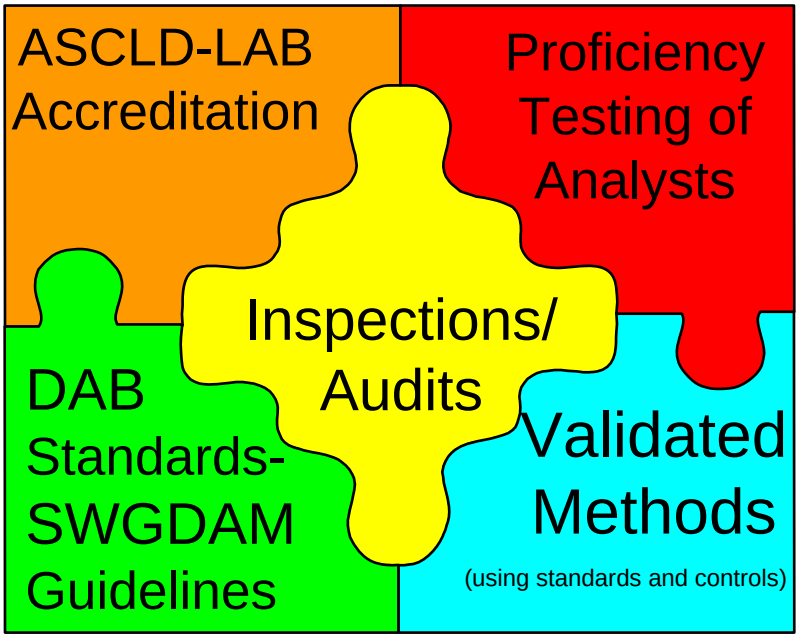
J.M. Butler (2005) *Forensic DNA Typing*, 2nd Edition, p. 391

- **Robust method** – successful results are obtained a high percentage of the time and few, if any, samples need to be repeated
- **Reliable method** – the obtained results are accurate and correctly reflect the sample being tested
- **Reproducible method** – the same or very similar results are obtained each time a sample is tested

General Levels of Validation

- **Developmental Validation** – commonly performed by commercial manufacturer of a novel method or technology (more extensive than internal validation)
- **Internal Validation** – performed by individual lab when new method is introduced
- **Performance Checks** – can be performed with every run (set of samples)

Ensuring Accurate Forensic DNA Results



Checks and Controls on DNA Results

Community	FBI DNA Advisory Board's Quality Assurance Standards *(also interlaboratory studies)* ISO17025
Laboratory	ASCLD/LAB Accreditation and Audits
Analyst	Proficiency Tests & Continuing Education
Method/Instrument	**Validation of Performance** ***(along with traceable standard sample)***
Protocol	Standard Operating Procedure is followed
Data Sets	Allelic ladders, positive and negative amplification controls, and reagent blanks are used
Individual Sample	Internal size standard present in every sample
Interpretation of Result	Second review by qualified analyst/supervisor
Court Presentation of Evidence	Defense attorneys and experts with power of discovery requests

When is Validation Needed?

- Before introduction of a new method into routine use
- Whenever the conditions change for which a method has been validated, e.g., instrument with different characteristics
- Whenever the method is changed, and the change is outside the original scope of the method

L. Huber (2001) Validation of Analytical Methods: Review and Strategy. Supplied by www.labcompliance.com

The VAM Principles

VAM = Valid Analytical Measurement

1. Analytical measurements should be made to satisfy an agreed requirement.
2. Analytical measurements should be made using methods and equipment that have been tested to ensure they are fit for their purpose.
3. **Staff making analytical measurements should be both qualified and competent to undertake the task**.
4. There should be a regular and independent assessment of the technical performance of a laboratory.
5. Analytical measurements made in one location should be consistent with those made elsewhere.
6. Organizations making analytical measurements should have well defined quality control and quality assurance procedures.

Roper P *et al.* (2001) *Applications of Reference Materials in Analytical Chemistry*. Royal Society of Chemistry: Cambridge UK, p. 2

How do you validate a method?

- Decide on analytical requirements
  - Sensitivity, resolution, precision, etc.
- **Plan a suite of experiments**
- **Carry out experiments**
- Use data to assess fitness for purpose
- Produce a statement of validation
  - Scope of the method

Roper, P., *et al.* (2001) *Applications of Reference Materials in Analytical Chemistry*. Royal Society of Chemistry, Cambridge, UK, pp. 108-109.

Assumptions When Performing Validation

- The equipment on which the work is being done is broadly suited to the application. It is clean, well-maintained and within calibration.
- The staff carrying out the validation are competent in the type of work involved.
- There are no unusual fluctuations in laboratory conditions and there is no work being carried out in the immediate vicinity that is likely to cause interferences.
- The samples being used in the validation study are known to be sufficiently stable.

Roper, P., *et al.* (2001) *Applications of Reference Materials in Analytical Chemistry*. Royal Society of Chemistry, Cambridge, UK, pp. 110-111.

Tools of Method Validation

- Standard samples
  - positive controls
  - NIST SRMs
- Blanks
- Reference materials prepared in-house and spikes
- Existing samples
- Statistics
- **Common sense**

Roper, P., *et al.* (2001) *Applications of Reference Materials in Analytical Chemistry*. Royal Society of Chemistry, Cambridge, UK, p. 110.

Recent Articles I Have Written on Validation

Profiles in DNA **(Promega Corporation), vol. 9(2), pp. 3-6**


PROFILES IN DNA

VALIDATION

http://www.promega.com/profiles/902/ProfilesInDNA_902_03.pdf

Debunking Some Urban Legends Surrounding Validation Within the Forensic DNA Community

By John Butler
National Institute of Standards and Technology, Gaithersburg, Maryland, USA

http://marketing.appliedbiosystems.com/images/forensic/volume8/PDFs_submitted/02A_CustomerCorner_Val_What_is_it.pdf


Applied Biosystems

Forensic News

January 2007 Customer Corner

Validation: What Is It, Why Does It Matter, and How Should It Be Done?
By John M. Butler, National Institute of Standards and Technology

Validation involves performing laboratory tests to verify that a particular instrument, software program, or measurement technique is working properly. These validation experiments typically examine precision, accuracy, and sensitivity, which all play a factor on the 3 R's of measurements: reliability, reproducibility, and robustness.[1]

Urban Legends of Validation…

Butler, J.M. (2006) *Profiles in DNA* vol. 9(2), pp. 3-6

#1: HUNDREDS OR THOUSANDS OF SAMPLES ARE REQUIRED TO FULLY VALIDATE AN INSTRUMENT OR METHOD

#2: VALIDATION IS UNIFORMLY PERFORMED THROUGHOUT THE COMMUNITY

#3: EACH COMPONENT OF A DNA TEST OR PROCESS MUST BE VALIDATED SEPARATELY

#4: VALIDATION SHOULD SEEK TO UNDERSTAND EVERYTHING THAT COULD POTENTIALLY GO WRONG WITH AN INSTRUMENT OR TECHNIQUE

#5: LEARNING THE TECHNIQUE AND TRAINING OTHER ANALYSTS ARE PART OF VALIDATION

#6: VALIDATION IS BORING AND SHOULD BE PERFORMED BY SUMMER INTERNS SINCE IT IS BENEATH THE DIGNITY OF A QUALIFIED ANALYST

#7: DOCUMENTING VALIDATION IS DIFFICULT AND SHOULD BE EXTENSIVE

#8: ONCE A VALIDATION STUDY IS COMPLETED YOU NEVER HAVE TO REVISIT IT

My Philosophy towards Validation

Ask first: Does the new method improve your capability?

- **C**oncordance – are the same typing results obtained with the new technique as with an older one?
- **C**onstant Monitoring – check multiple allelic ladders in a batch against one another to confirm precision and consistent lab temperature
- **C**ommon Sense – are replicate tests repeatable?

Common Perceptions of Validation

The goal is not to experience every possible scenario during validation…

"You cannot mimic casework because every case is different."

Effort

Lots of experiments are required

Many labs are examining far too many samples in validation and thus delaying application of casework and contributing to backlogs…

Significant time is required to perform studies

Time

Number of Samples Needed

Relationship between a sample and a population of data

How do you relate these two values?

Data collected in your lab as part of validation studies → All potential data that will be collected in the future in your lab

Student's *t*-Test associates a sample to a population

"Sample" of Typical Data

"Population" of All Data Obtained

Student's *t*-Tests

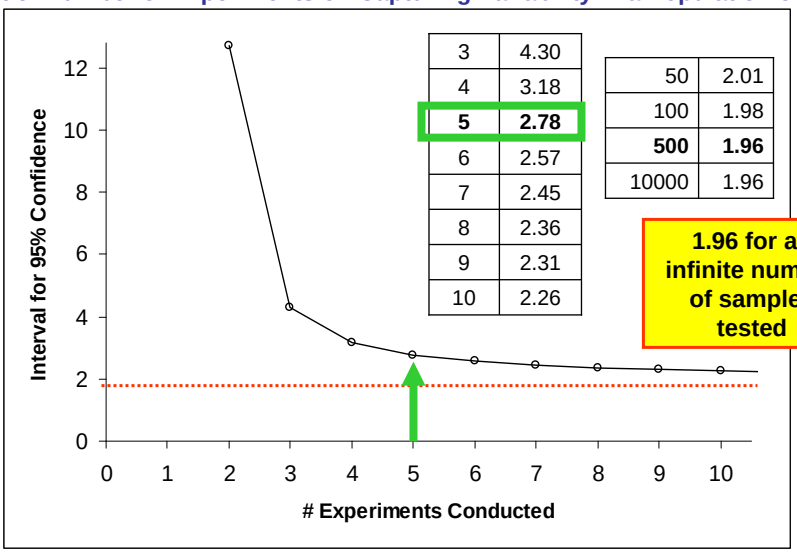
"Student" (real name: W. S. Gossett [1876-1937]) developed statistical methods to solve problems stemming from his employment in a brewery.

Student's *t*-test deals with the problems associated with inference based on "small" samples: the calculated mean (X_{avg}) and standard deviation (σ) may by chance deviate from the "real" mean and standard deviation (i.e., **what you'd measure if you had many more data items: a "large" sample**).

http://www.physics.csbsju.edu/stats/t-test.html

Student's t-Test Curve

Impact of Number of Experiments on Capturing Variability in a Population of Data



3	4.30
4	3.18
5	**2.78**
6	2.57
7	2.45
8	2.36
9	2.31
10	2.26

50	2.01
100	1.98
500	**1.96**
10000	1.96

1.96 for an infinite number of samples tested

Useful Resources on Validation

- Taylor JK. (1981) Quality assurance of chemical measurements. *Analytical Chemistry* 53(14): 1588A-1596A.
- Taylor JK. (1983) Validation of analytical methods. *Analytical Chemistry* 55(6): 600A-608A.
- Green JM. (1996) A practical guide to analytical method validation. *Analytical Chemistry* 68: 305A-309A.
- EURACHEM Guide (1998) *The Fitness for Purpose of Analytical Methods: A Laboratory Guide to Method Validation and Related Topics*; available at http://www.eurachem.ul.pt/guides/valid.pdf

See also STRBase Validation Section:
http://www.cstl.nist.gov/biotech/strbase/validation.htm

Overview of Developmental Validation Studies

2. Developmental Validation: The developmental validation process may include the studies detailed below. Some studies may not be necessary for a particular method.

2.1 Characterization of genetic markers
2.2 Species specificity
2.3 Sensitivity studies
2.4 Stability studies
2.5 Reproducibility
2.6 Case-type samples
2.7 Population studies
2.8 Mixture studies
2.9 Precision and accuracy
2.10 PCR-based procedures

SWGDAM Revised Validation Guidelines
http://www.fbi.gov/hq/lab/fsc/backissu/july2004/standards/2004_03_standards02.htm

PowerPlex Y Developmental Validation Experiments

Study Completed (17 studies done)	Description of Samples Tested (performed in 7 labs and Promega)	# Run
Single Source (Concordance)	5 samples x 8 labs	40
Mixture Ratio (male:female)	6 labs x 2 M/F mixture series x 11 ratios (1:0,1:1,1:10,1:100,1:300,1:1000,0.5:300, 0.25:300,0.125:300, 0.0625:300, 0.03:300 ng M:F)	132
Mixture Ratio (male:male)	6 labs x 2 M/M mixtures series x 11 ratios (1:0, 19:1, 9:1, 5:1, 2:1, 1:1, 1:2, 1:5, 1:9, 1:19, 0:1)	132
Sensitivity	7 labs x 2 series x 6 amounts (1/0.5/0.25/0.125/0.06/0.03)	84
Non-Human	24 animals	24
NIST SRM	6 components of SRM 2395	6
Precision (ABI 3100 and ABI 377)	10 ladder replicates + 10 sample replicated + [8 ladders + 8 samples for 377]	36
Non-Probative Cases	65 cases with 102 samples	102
Stutter	412 males used	412
Peak Height Ratio	N/A (except for DYS385 but no studies were noted)	
Cycling Parameters	5 cycles (28/27/26/25/24) x 8 punch sizes x 2 samples	80
Annealing Temperature	5 labs x 5 temperatures (54/58/60/62/64) x 1 sample	25
Reaction volume	5 volumes (50/25/15/12.5/6.25) x [5 amounts + 5 concentrations]	50
Thermal cycler test	4 models (480/2400/9600/9700) x 1 sample + [3 models x 3 sets x 12 samples]	76
Male-specificity	2 females x 1 titration series (0-500 ng female DNA) x 5 amounts each	10
TaqGold polymerase titration	5 amounts (1.38/2.06/2.75/3.44/4.13 U) x 4 quantities (1/0.5/0.25/0.13 ng DNA)	20
Primer pair titration	5 amounts (0.5x/0.75x/1x/1.5x/2x) x 4 quantities (1/0.5/0.25/0.13 ng DNA)	20
Magnesium titration	5 amounts (1/1.25/1.5/1.75/2 mM Mg) x 4 quantities (1/0.5/0.25/0.13 ng DNA)	20

Krenke *et al.* (2005) *Forensic Sci. Int.* 148:1-14 TOTAL SAMPLES EXAMINED **1269**

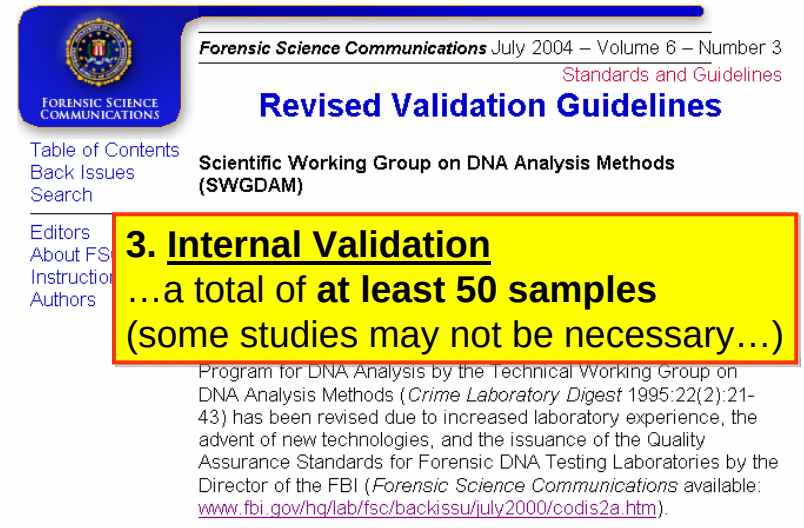
General Steps for Internal Validation

- **Review literature and learn the technique**
- Obtain equipment/reagents, if necessary
- Determine necessary validation studies (there can be overlap and you only need to run a total of 50 samples)
- Collect/obtain samples, if necessary
- **Perform validation studies maintaining all documentation**
- Summarize the studies and submit for approval to Technical Leader
- Write-up the analytical procedure(s). Include quality assurance (controls, standards, critical reagents and equipment) and data interpretation, as applicable
- Determine required training and design training module(s)
- Design qualifying or competency test

From Robyn Ragsdale (FDLE), Validation Workshop (Aug 24-26, 2005 at NFSTC)
http://www.cstl.nist.gov/biotech/strbase/validation/validationworkshop.htm

Revised SWGDAM Validation Guidelines
(July 2004)

http://www.fbi.gov/hq/lab/fsc/backissu/july2004/standards/2004_03_standards02.htm



The document provides validation guidelines and definitions approved by SWGDAM July 10, 2003.

Overview of Internal Validation Studies

3. Internal Validation: The internal validation process should include the studies detailed below encompassing **a total of at least 50 samples**. Some studies may not be necessary due to the method itself.

   3.1 Known and nonprobative evidence samples
   3.2 Reproducibility and precision
   3.3 Match criteria
   3.4 Sensitivity and stochastic studies
   3.5 Mixture studies
   3.6 Contamination
   3.7 Qualifying test

SWGDAM Revised Validation Guidelines
http://www.fbi.gov/hq/lab/fsc/backissu/july2004/standards/2004_03_standards02.htm

Design of Experiments Conducted for Validation Studies

- Before performing a set of experiments for validation, ask yourself:
  - **What is the purpose of the study?**
  - **Do we already know the answer?**
  - **Can we write down how we know the answer?**
- Think before you blindly perform a study which may have no relevance (e.g., extensive precision studies)
- **Too often we do not differentiate learning, validation, and training**

Points for Consideration

- Remove as many variables as possible in testing an aspect of a procedure
  - e.g., create bulk materials and then aliquot to multiple tubes rather than pipeting separate tubes individually during reproducibility studies
- Who can do (or should do) validation…
  - Outside contractor?
  - Summer intern?
  - Trainee?
  - Qualified DNA analyst

From a validation standpoint, having an outside group perform the validation studies on your instruments is legitimate, **but valuable experience and knowledge are lost…**

Practical Examples

- Profiler Plus/COfiler kit switch to Identifiler
- ABI 3100 upgrade to ABI 3130xl
- GeneScan/Genotyper to GeneMapperID
- New allelic ladder provided by company
- Bringing Quantifiler “on-line” (from Quantiblot)
- DNA IQ
- Corbett robot
- FSS-i3 expert system software
- Reduced volume reactions

Discuss each example - participants to provide what they would do…

Suggestions for an Internal Validation of an STR Kit

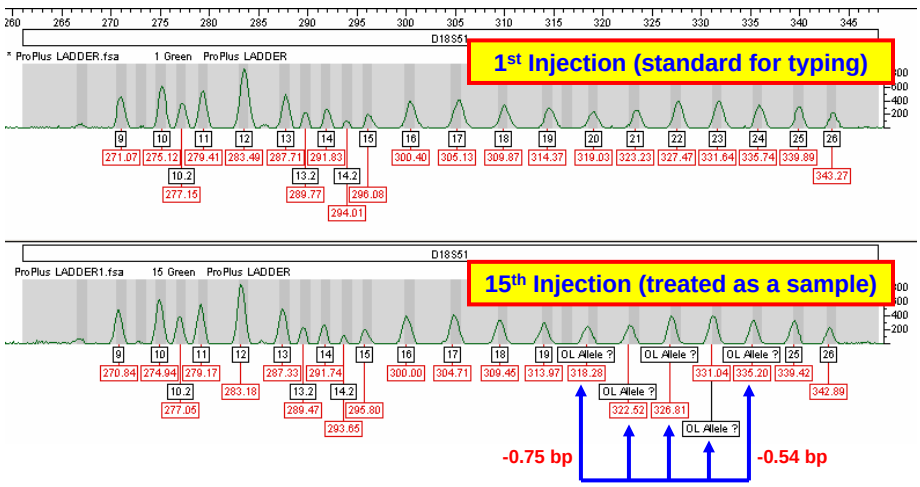
- Standard samples (3.1) **Between 1 and ~20 samples**
  - Verify correct type with positive control or NIST SRM samples
  - Concordance study with 5-10 (non-probative casework) samples previously typed with other kit(s)
- Precision samples (3.2) **5-10 samples**
  - Run at least 5-10 samples (allelic ladder or positive control)
- Sensitivity samples (3.4) **14 samples**
  - Run at least 2 sets of samples covering the dynamic range
  - 5 ng down to 50 pg—e.g., 5, 2, 1, 0.5, 0.2, 0.1, 0.05 ng
- Mixture samples (3.5) **10 samples**
  - Run at least 2 sets of samples
  - Examine 5 different ratios—e.g., 10:1, 3:1, 1:1, 1:3, 1:10

>50 samples

Additional Suggestions for Meeting the SWGDAM Revised Validation Guidelines

- Match Criteria (3.3)
  - As part of running a batch of samples (e.g., 10 or 96), run one allelic ladder at the beginning and one at the end
  - If all alleles are typed correctly in the second allelic ladder, then the match criteria (i.e., precision window of +/-0.5 bp) has likely been met across the entire size range and duration of the run
- Contamination Check (3.6)
  - Run negative controls (samples containing water instead of DNA) with each batch of PCR products
- Qualifying Test (3.7)
  - Run proficiency test samples

Use of Second Allelic Ladder to Monitor Potential Match Criteria Problems



These alleles have drifted outside of their genotyping bins due to temperature shifting over the course of the sample batch

Example: ABI 3130

- Evaluation of a new ABI 3130 when a laboratory already has experience with ABI 310
- STR kits used in lab will remain the same

Recommendations:

- Precision studies to evaluate instrument reproducibility
- Sensitivity studies
- **Do not need new stutter, mixture ratio, peak height ratio, etc. (these relate to dynamics of the the kit used)**

Instrument/Software Upgrades or Modifications

- What should be done to "validate" new upgrade?
  - ABI 7000 to ABI 7500
  - ABI 3100 to ABI 3130xl
  - GeneScan/Genotyper to GeneMapper*ID*
- Try to understand what is different with the new instrument or software program compared to the one you are currently using (e.g., ask other labs who may have made the switch)
- If possible, try to retain your current configuration for comparison purposes for the validation period

Run the same plate of samples on the original instrument/software and the new one

ABI 3130xl vs ABI 3100

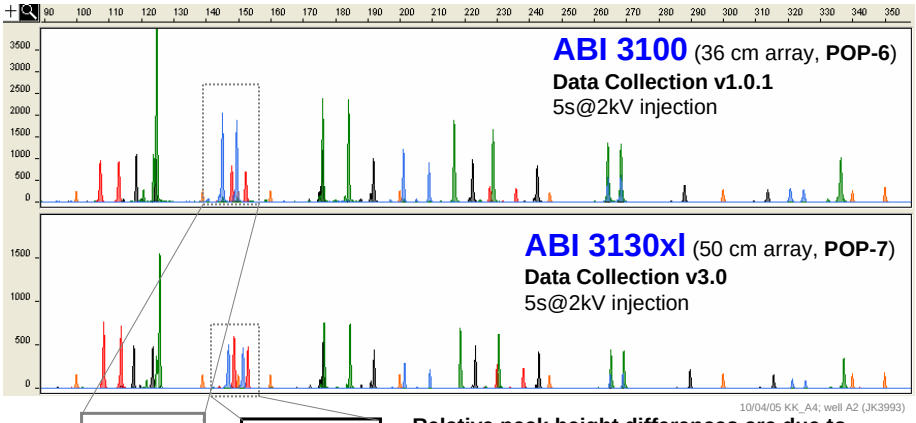
What NIST did to "validate" a 3130xl upgrade

- Ran plates of samples on both instruments with same injection and separation parameters and compared results
  - Data Collection version 1.0.1 (3100) vs 3.0 (3130xl)
  - POP-6 (3100) vs POP-7 (3130xl)
  - 36 cm array (3100) vs 50 or 80 cm array (3130xl)
- Ran several plates of Identifiler samples and compared allele calls (noticed a sensitivity difference with equal injections and relative peak height differences between dye colors) – **all obtained allele calls were concordant**
- Ran a plate of Profiler Plus samples and compared sizing precision – **precision was not significantly different**
- Also examined SNaPshot products and mtDNA sequencing data

Environmental conditions may change over time so original validation is no longer valid…

Comparison of ABI 3100 Data Collection Versions

Same DNA sample run with Identifiler STR kit (identical genotypes obtained) GeneScan display



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.

Validation Section of the DNA Advisory Board Standards
issued July 1998 (and April 1999); published in *Forensic Sci. Comm.* July 2000

STANDARD 8.1 The laboratory shall use validated methods and procedures for forensic casework analyses *(DNA analyses)*.

8.1.1 Developmental validation that is conducted **shall be appropriately documented**.

8.1.3 Internal validation **shall be performed and documented by the laboratory**.

FORENSIC SCIENCE COMMUNICATIONS JULY 2000 VOLUME 2 NUMBER 3

Example of Validation Documentation

Alabama Department of Forensic Sciences
Birmingham DNA

ABI Prism® 7000 Validation

TABLE OF CONTENTS

TAB	TITLE
8.1.1	Developmental Validation
8.1.3.1(a)	Known and Non-Probative Samples
8.1.3.1(b)	Reproducibility • Quantitation • Peak Height Experiment – ABI 310
8.1.3.1(b)	Precision

The validation studies referenced above have been reviewed and provide the necessary documentation required by the FBI Director's "Quality Assurance Standards for Forensic DNA Testing Laboratories" for a quantitation method to be used in the forensic casework section of the Alabama Department of Forensic Sciences Birmingham DNA laboratory.

Angelo Della Manna, MSFS, D-ABC
Forensic Biology Discipline Chief
Statewide Technical Leader

Date

Available on STRBase Validation Website:
http://www.cstl.nist.gov/biotech/strbase/validation/ADFS-BH_7000val.pdf

Documentation of Alabama Validation for ABI 7000 and Quantifiler Assay

What Section of QAS Validation Requirements

Experiments Performed

Summary of Results

Conclusions

Known and Non-Probative Samples

8.1.3.1(a) Has the procedure been tested using known and non-probative evidence samples?

Experiment:
Eleven (11) evidence samples comprised of various origins that are encountered in routine casework were analyzed with the Quantifiler Human Kit on the ABI 7000. These quantitation results were then compared to the previously obtained Quantiblot results. All samples were then amplified using the Identifiler Kit and the Quantifiler Human results in an effort to determine the preferred amount of sample template to add to the PCR process.

Additionally, this laboratory participated in a NIST Quantitation study aimed at evaluating shipping conditions of standard DNA samples. Each of the NIST samples was analyzed with the Quantifiler Human Kit as well as the Quantiblot Kit, with results compared and tabulated as well.

Results:

Non-Probative Samples

Sample Name	Sample Type	Quantifiler Human Result	Quantiblot Result	Percent Difference 100-[(QF/QB)*100]
96MB84292-1D	vaginal swab	0.0914	0.12	23.83
00BH01157-1A	victim standard	6.03	2	-201.50
96MB84292-1A	victim standard	1.36	1	-36.00
98BH00578-3	blood from currency	0.0993	0.03	-231.00
96BH29999-1C	vaginal swab	0.303	0.24	-26.25
96BH29999-1A1	victim standard	0.684	1	31.6
96BH32136-2A	semen on comforter	0.103	0.12	14.17
96BH32136-3A	suspect standard	2.14	1	-114.00
96BH32136-1A	victim standard	4.14	2	-107.00
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
H	0.0365	0.056	0.0396	0.04	0.03	0.03	0.03	-46.78

Note: All quantitation results are in units of ng/µl
Note: The red numbers indicate the lower quant value

The experimental results demonstrate that the Quantiblot method of quantitating DNA typically underestimated the amount of DNA present in a sample. An accurate quantitation result is critical to obtaining an adequate DNA profile downstream with the Identifiler Kit. If DNA quantities greater than the optimal range are added to the PCR mix, the analyst will likely have a more imbalanced PCR product as well as possible saturation of the detection system causing pull-up and a greater likelihood of Stochastic effects. When utilizing the Quantifiler results to determine DNA template addition, the resulting peak heights on the ABI 310 from the Identifiler amplicon were acceptable and produced no excessive pull-up or stochastic related issues.

http://www.cstl.nist.gov/biotech/strbase/validation/ADFS-BH_7000val.pdf

Acknowledgments

National Institute of Justice
The Research, Development, and Evaluation Agency of the U.S. Department of Justice

- **NIJ Funding** for NIST Project Team through NIST Office of Law Enforcement Standards
- **Robyn Ragsdale (FDLE)**
- **Chris Tomsey (PSP), Margaret Kline (NIST)**
- Dave Duewer (NIST)
- Kari Tontarski (Montgomery County Crime Lab)
- Robin Cotton (Orchid Cellmark)
- Tim McMahon (AFDIL)

- **Many members of forensic DNA typing community for their input on our 2004 validation questionnaire**