# New Autosomal STR Loci to Address Challenges in Human Identity Testing

Dr. Peter M. Vallone, **Becky Hill**,
Margaret C. Kline and Dr. John M. Butler
National Institute of Standards and Technology

Centre of Forensic Sciences
Toronto Canada
June 11, 2008

## Disclaimers

**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 authors and do not necessarily represent the official position or policies of the US Department of Justice.

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.

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

## Questions to Address

- **Which loci were selected?**
  - Utility of additional loci
  - Characterization of additional STR loci
- **How was the STR multiplex developed?**
  - Initial work with smaller "miniplexes"
  - The "Autoplex" single amplification reaction
- **How can this information be useful?**

## Aren't the Current STR Loci Good Enough?

- For general forensic matching of evidence to suspect, the 13 CODIS STR loci are sufficient
- For other human identity/relationship testing questions, more autosomal loci can be beneficial or even necessary

## More Loci are Useful in Situations Involving Relatives

- **Missing Persons** and Disaster Victim Identification (kinship analysis)
- Immigration Testing (often limited references)
  - Recommendations for 25 STR loci
- Deficient Parentage Testing
  - often needed if only one parent and child are tested

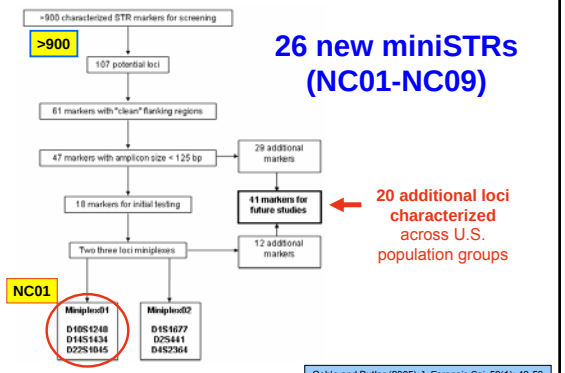
Relationship testing labs are being pushed to answer more difficult genetic questions…and **we want to make sure the right tools are in place**

## Selection of New Autosomal Loci

- Aim to have candidate sets for optimal miniSTRs
- Using ~900 STR loci with some literature data as a starting point…
  - Loci with high heterozygosities (>0.7)
  - Loci with small allele ranges (<24 bp) – low mutation?
  - Tetra (some tri-)nucleotide repeats without variants
  - Clean flanking regions (PCR products <140 bp)
- **26 loci** met criteria and fully characterized…

Coble and Butler (2005) Characterization of new miniSTR loci to aid analysis of degraded DNA *J. Forensic Sci.* 50(1): 43-53

## Initial Testing Results with Potential miniSTR Loci




**26 new miniSTRs (NC01-NC09)**

**20 additional loci characterized** across U.S. population groups

Coble and Butler (2005) *J. Forensic Sci.* 50(1): 43-53

## Characterizing New Loci

- ✓ Select genetic loci
- ✓ Design primers – optimize multiplex assay
- ✓ Type population samples to examine variation
- ✓ Sequence alleles to establish nomenclature
- ✓ Develop bins and panels for genotyping
- ✓ Construct allelic ladders
- ✓ Evaluate RMP or ability to separate common types
- ✓ Perform mutation rate studies
- ✓ Perform concordance studies (when applicable)
- ✓ Calibrate genotypes with NIST SRM components
- ✓ Work with companies/collaborators
- ✓ Publish details on loci and assays

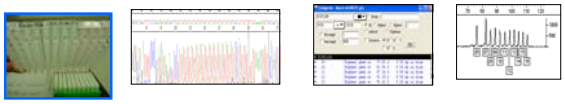
## Characterization of New Loci

**"Computer Work"**

Candidate STR marker selection (e.g. literature searches) → Pull down sequence data from the web (e.g. NCBI) → Identify Chromosome Location (e.g. Human BLAT Search ) → Screen for PCR Primers (e.g. Primer3) → Test primers for Multiplex-ability (e.g. AutoDimer - NIST )

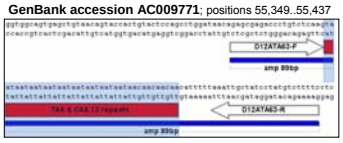
**"Laboratory Work"**

Test Markers on Population samples → Sequence homozygotes to determine allele sizes → Build Macros for Genotyping → Construct Allelic Ladders



## Characterization of miniSTR **D12ATA63**

**GenBank accession AC009771**; positions 55,349..55,437




**[FAM]** – GAGCGAGACCCTGTCTCAAG
GAAAAGACATAGGATAGCAATTT

**Chr 12** 106.825 Mb (12q23.3)

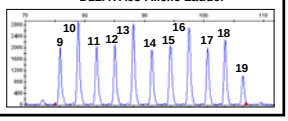
**Trinucleotide**
[TAA][CAA] repeat
**76 -106 bp**
Alleles 9 -19

| Allele | Caucasian (N = 260) | African Am (N = 259) | Hispanic (N = 140) |
|---|---|---|---|
| 9 | -- | -- | 0.0036 |
| 10 | 0.0019 | 0.0154 | 0.0036 |
| 11 | 0.1385 | 0.1525 | 0.1500 |
| 12 | 0.2154 | 0.1004 | 0.1786 |
| 13 | 0.0173 | 0.1564 | 0.0286 |
| 14 | 0.1615 | **0.3340** | 0.2214 |
| 15 | 0.0577 | 0.0772 | 0.0714 |
| 16 | **0.2981** | 0.1004 | **0.2643** |
| 17 | 0.0981 | 0.0521 | 0.0679 |
| 18 | 0.0096 | 0.0058 | 0.0071 |

**Heterozygosity Values**

| | |
|---|---|
| U.S. Caucasian | **0.842** |
| African American | **0.788** |
| U.S. Hispanic | **0.879** |

**D12ATA63 Allelic Ladder**




## European Labs Have Adopted the NIST-Developed NC (non-CODIS) miniSTRs

*FSI* (2006) **156(2):** 242-244

Short communication

The evolution of DNA databases—Recommendations for new European STR loci

Peter Gill[a,*], Lyn Fereday[b], Niels Morling[c], Peter M. Schneider[d]

[a] Forensic Science Service, Birmingham, UK
[b] Forensic Science Service, London, UK
[c] Department of Forensic Genetics, Institute of Forensic Medicine, University of Copenhagen, Denmark
[d] Institute of Legal Medicine, University of Cologne, Germany

Received 25 May 2005; accepted 26 May 2005

**…recommended that existing multiplexes are re-engineered to enable small amplicon detection, and that three new mini-STR loci with alleles <130 bp (D10S1248, D14S1434 and D22S1045) are adopted as universal. This will increase the number of European standard Interpol loci from 7 to 10.**

## In Jan 2008 Issue of *J. Forensic Sci.*

**J. Forensic Sci., Jan 2008, 53(1):73-80**

*J Forensic Sci*, January 2008, Vol. 53, No. 1
doi: 10.1111/j.1556-4029.2008.00595.x
Available online at: www.blackwell-synergy.com

*Carolyn R. Hill, M.S.; Margaret C. Kline, M.S.; Michael D. Coble,[†] Ph.D.; and John M. Butler, Ph.D.*

Characterization of 26 MiniSTR Loci for Improved Analysis of Degraded DNA Samples

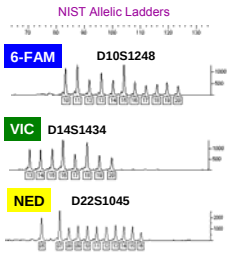
- Characterization of **26** new autosomal loci
- Primer sequences, GeneMapper bins and panels, genotypes on common samples, and allele frequency information **already available on STRBase**

**http://www.cstl.nist.gov/div831/strbase/miniSTR.htm**
**http://www.cstl.nist.gov/div831/strbase/newSTRs.htm**

## Multiple Miniplexes

- **26 characterized loci** divided into 10 miniplexes
- One locus per dye color
- Allelic ladders created
- **Amplicons <140 bp**
- miniSTRs
- Work with 100 pg DNA
- **For degraded samples**
  **(bones in missing persons cases)**

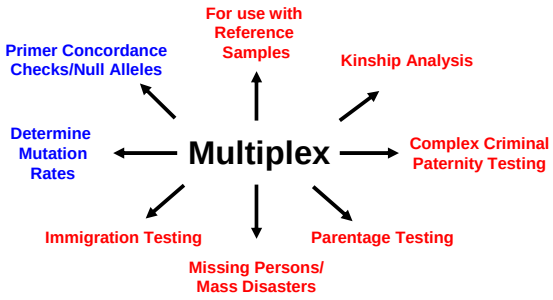
**NC = Non-CODIS or non-core**




**NC01 Loci**

See Dixon et al. (2006) *Forensic Sci. Int.* 164: 33-44.

## Why Build a Large Multiplex?

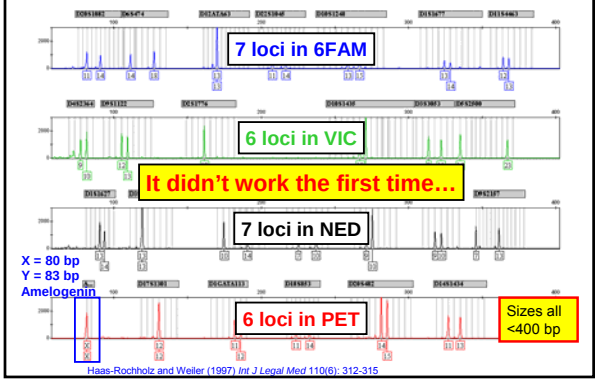



## The Multiplex Design Process

## The Design of the Multiplex

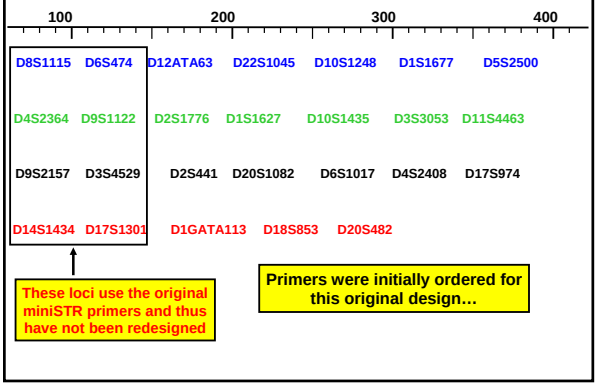
- **Goal**: A single amplification multiplex combining the 26 autosomal loci + Amelogenin in one reaction (27plex)
- How was this achieved?
  - PCR amplicons labeled with 6FAM, VIC, NED, and PET dyes
  - Primer redesign for non "miniSTR loci"
  - Empirical trial and error of primer compatibility, as well as balancing the primer concentrations

## Current Autoplex – 26 Loci

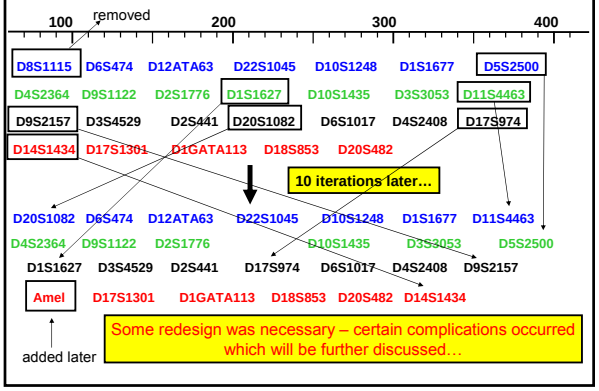



Haas-Rochholz and Weiler (1997) *Int J Legal Med* 110(6): 312-315

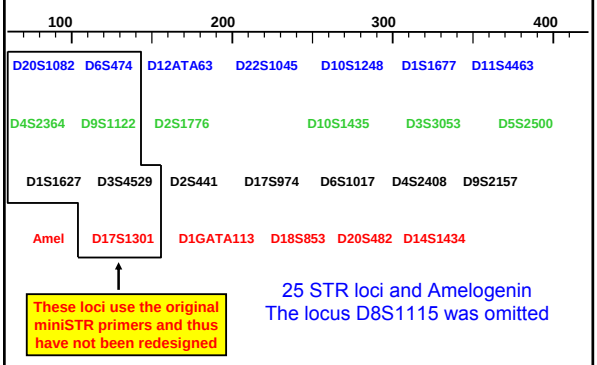
## 1st Autoplex Design



## Some Loci Had to be Rearranged




## Current Autoplex Design

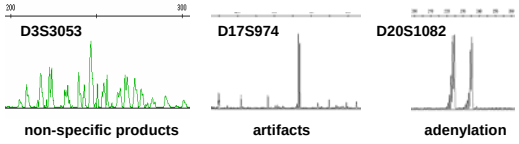



## Where do we begin?

- PCR primers were designed using Primer3
- Primers were screened with the AutoDimer software
- Designed primers were mapped to confirm amplicon size and ensure primers flank the repeat
- Primers ordered
  - Forward fluorescent dye-labeled from AB
  - Reverse non-dye labeled from Operon

## What Next?

- 1st step: primers were amplified in singleplex
  - If the primers fail in singleplex, they will most likely fail in multiplex – Candidates for redesign
  - Initially, 7 loci out of 26 failed – 19 worked in singleplex
- How do we define failure?
  - Incomplete adenylation, presence of artifacts, low signal, non-specific products, or no PCR product

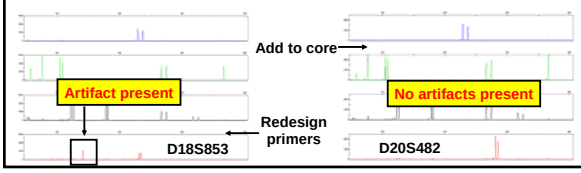



## Building the Multiplex

- Successful primers were tested in multiplex – 19 loci were combined together
- Then, the minimum amount of loci that work well together were determined
  - Can isolate by dye color – amplify loci in one dye channel at a time to hone in on failures
  - 9 loci: add remaining loci one at a time


## Trial and Error

- Serially add a locus to the "core-9" multiplex and determine compatibility
  - Check for artifacts and adenylation
- Increase the size of the multiplex
  - Continue adding to a 'core' multiplex and redesign incompatible primers



## Finalizing the Multiplex

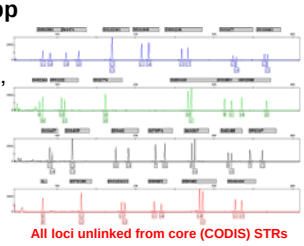
- The size of the multiplex increased
  - 18, 19, and 23plex
- Locus to locus primer balancing was performed with each multiplex
- Concordance and Mutation Rate studies were performed with a 23plex
- Currently, the multiplex has all but one locus (D8S1115) 26plex

## Lessons Learned from Primer Redesign

- Some loci had to be redesigned with PCR products in a different size range
  - If artifacts are present or if it is a noisy baseline
- The fluorescent dye label can be switched to the reverse primer to mask an artifact
  - Can check forward and reverse primers separately
- Adenylation issues: a 'PIGTAIL' (GTTTCTT) can be added to the 5' end of reverse primers
  - D1S1677, D3S3053, D11S4463, and D12ATA63
- Dye artifacts can be filtered out with post-PCR cleanup (Edge Columns)
  - Especially in the PET dye channel

## Autoplex

- **25 STRs and amelogenin** in single multiplex
  Multiple loci in four dye channels
- **Amplicons 70 to 400 bp**
  (No longer 'miniSTRs')
- Typically use 1 ng DNA, 30 cycles



**26plex**

All loci unlinked from core (CODIS) STRs

## PCR Parameters

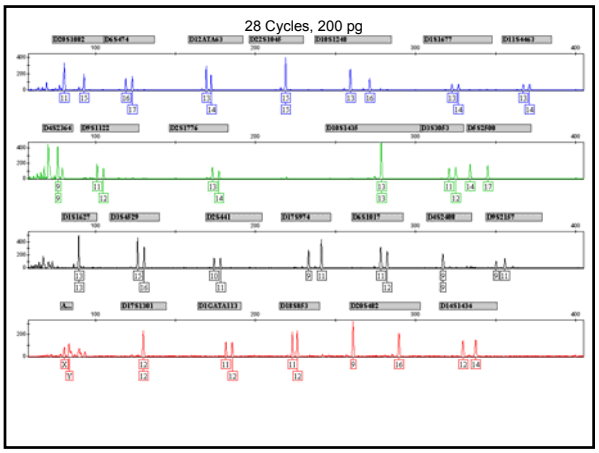
- Master Mix
  - 2 mM $MgCl_2$
  - 1x PCR Buffer
  - 1 Unit TaqGold
  - 0.2 µM Primer mix
  - 250 mM dNTPs
  - 0.16 mg/mL BSA
- 20µL reaction volume = 19µL MM + 1µL DNA sample (~1ng)

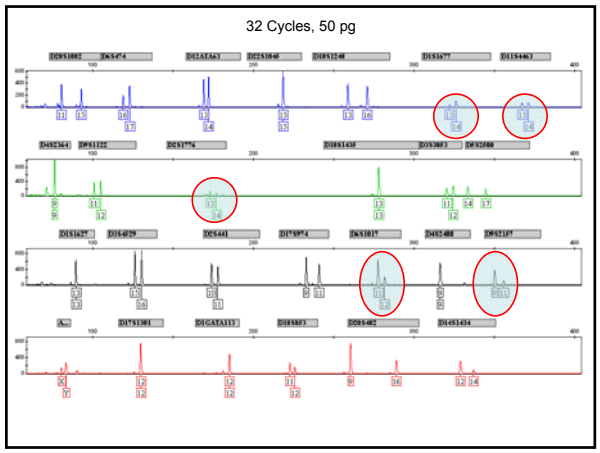
## Thermal Cycling Conditions

- Conditions for ABI 9700 in 9600 emulation mode
- 95˚C Hot Start for 11 min
- 30 cycles
  - 94˚C for 45 sec Denaturation
  - 59˚C for 2 min Annealing
  - 72˚C for 1 min Elongation
- 60˚C soak for 60 min
- 25˚C hold (∞)

## Sensitivity Study

- A highly characterized sample was used for this study at a wide range of concentrations:
  - **2 ng, 1 ng, 750 pg, 500 pg, 400 pg, 300 pg, 250 pg, 200 pg, 100 pg, 50 pg, and 25 pg**
- 3 different PCR cycles were tested:
  - **28, 30, and 32 cycles**
- Ideal DNA concentration and cycling:
  - **1 ng DNA for 30 cycles**
- Multiplex sensitivity (lowest concentration where all peaks were detected above 50 RFUs):
  - **28 cycles, 200 pg**
  - **30 cycles, 100 pg**
  - **32 cycles, 50 pg**




## Further Work with the Autoplex

Studies were Performed with the 23plex

## Evaluation of Autoplex (23plex)

- **660 U.S. population samples**
  - U.S. Caucasian, African American, Hispanic
  - **Concordance testing** compared to miniSTR results
- **790 father/son samples**
  - U.S. Caucasian, African American, Hispanic, Asian
  - **Mutation rate determination**
- 12 samples for **extended family testing**

**>1450 samples examined so far**
(multiple primer batches prepared)

## Concordance Study to Check for Null Alleles

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

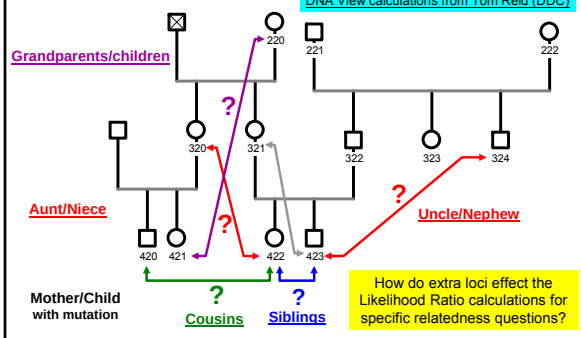
- **639 samples** compared
- 14,058 total types (639 x 22 loci)
- 28 types discordant (0.20%)
- **99.80%** concordance
- Discordance has not yet been confirmed by sequencing

## Mutation Rates Measured for New STRs

- **395 father/son pairs** (790 samples total)
- 22 STR loci examined
- 8690 allelic transfers
- Only **6 mutations** were observed in total
- **0.069%** mutation rate
- 2-3 times less than typical 0.2% for common STRs

## Extended Family Sample Testing

**15 vs 37 STRs**
DNA View calculations from Tom Reid (DDC)




How do extra loci effect the Likelihood Ratio calculations for specific relatedness questions?

## Comparison of Likelihood Ratios

| Relationship Examined | **15 STRs** (Identifiler, ID15) | ID15 + **Autoplex 22** STRs = **37 loci** (A37) |
|---|---|---|
| Mother/Child* (*with single mutation) | 0.214 | **5,200,000** Extra loci help… |
| Siblings | 477 | **113,000** Extra loci help… |
| Uncle/Nephew | 824 | **247,000** Extra loci help… |
| Cousins | 0.45 | **2.25** |
| Grandparents/ Grandchildren | 0.53 | **1.42** |

**Conclusions:** Longer distance multi-generational questions cannot usually be solved with additional autosomal STRs…

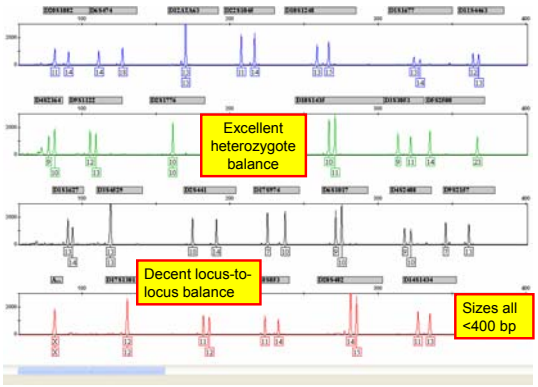
## SRM 2391b: DNA Profiling Standard Certificate of Analysis Update

- Genotyping and sequencing have been performed with SRM 2391b components (#1-12) for all 26 additional loci
- Certified and reference values have been assigned to all resulting alleles
- The Certificate of Analysis is in the process of being updated for all 26 loci (coming up in near future…)
- Using these values, bins and panels have been written in GeneMapper*ID*
- **Purpose**: No commercial allelic ladders are available, but all genotypes are certified for the components of SRM 2391b

http://www.cstl.nist.gov/biotech/strbase/miniSTR/miniSTR_NC_loci_types.htm

## Current Autoplex (26 loci)

Comp 9
9947A




## Summary/Conclusions

- **26 unlinked loci** have been characterized and we have developed multiple miniplexes and an Autoplex (26plex)
- The Autoplex is a robust single amplification 5-dye multiplex reaction that can benefit the forensic community for reference purposes and relationship testing
- **NIST SRM 2391b** will include certified and reference values on these 26 additional autosomal STR loci

## In the Future…

- All information will be available on STRBase
- A manuscript is currently being prepared for submission to a forensic journal
- **Use loci in new applications such as Rapid PCR**

## Further Evaluation of the Loci for Rapid PCR

- Amplified (25/26) each locus in singleplex under rapid cycling conditions
- Evaluate each locus for signal, adenylation and artifacts
- Rank and test candidate loci in a rapid multiplex

## Rapid PCR

- What do we mean by rapid PCR?
  - Rapid hot start polymerases (save ~10min)
  - Shortening cycling hold times (5 sec vs 1 min)
  - Utilizing existing thermal cycling technology (AB 9700)
    - Eliminating 1 ºC/sec ramp rate (9600 emulation)
    - Utilize the 9700 4 ºC/sec ramp rate
  - Using commercial polymerases that are 'faster'

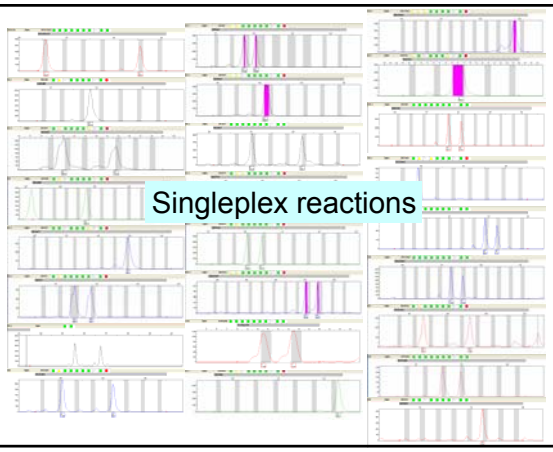
Obtain results in less than 45 minutes
Trying simple things first…

## Thermal Cycling

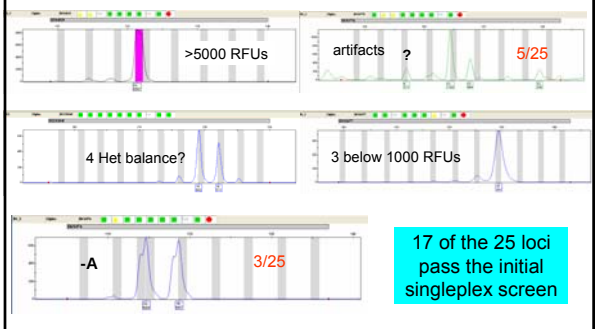
| Parameter | Unit | Trad | Rapid | Difference (min) | % |
|---|---|---|---|---|---|
| **Hot Start** | Min | 10 | 1 | 9.0 | 6.3 |
| **Hold** | Sec | 60 | 5/10 | 72.3 | 50.6 |
| **Soak** | Min | 60 | 1 | 59.0 | 41.2 |
| **Ramp rate** | (deg/sec) | 1 | 4 | 22.4 | 15.7 |
| **Cycles** | | 28 | 28 | | |
| **Time** | | 2:58:41 | 0:35:38 | 2:23:03 | |

| Parameter | Purpose |
|---|---|
| Hot Start | Primer Dimer, non-specific amplification |
| Hold | Denature, annealing, elongation, Inter and intra locus balance |
| Soak | Full adenylation of PCR products |

Evaluate robustness and reproducibility




## Singleplex Evaluation




17 of the 25 loci pass the initial singleplex screen

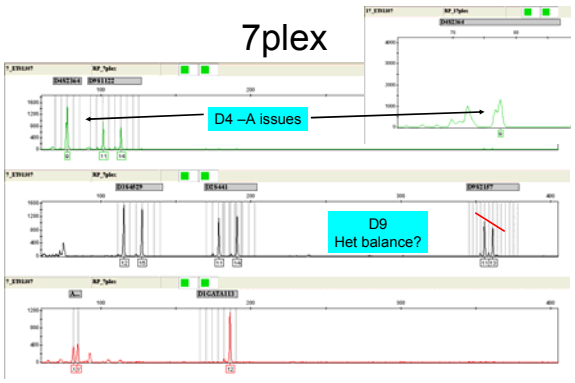
## Testing 4 Multiplexes

- After singleplex evaluation 4 multiplexes were tested (empirical balancing)
  - 17plex
  - 14plex
  - 12plex
  - 7plex

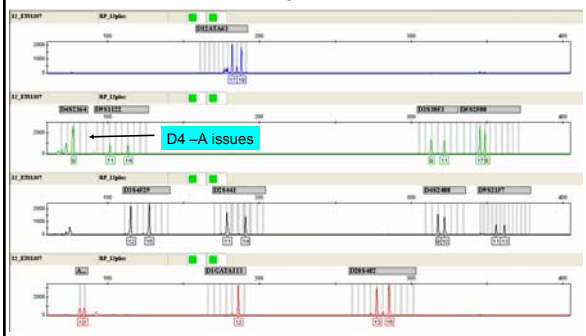
  } Subset of the 17plex

- Run under rapid cycling conditions
- 1 ng DNA, 28 cycles

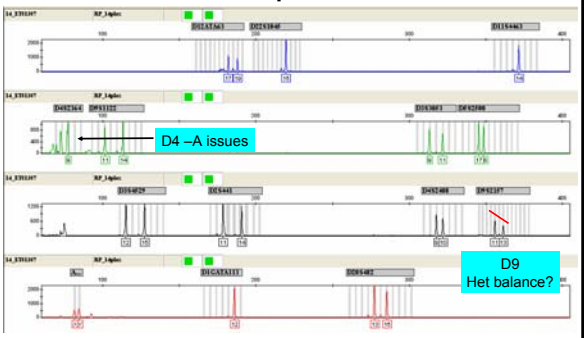
## 7plex




## 12plex




## 14plex




## 17plex


Boxed loci have pigtailed primers

## Rapid Multiplex Concordance

- Results for the rapid multiplexes were compared with previously run assays (Standard cycling – TaqGold)
- N = 16 samples
- D4S2364 adenylation issues/artifacts
- D9S2157 severe het imbalance – allele drop out in 2 samples (13,13 vs 13,14) and (7,7 vs 7,11)
- Evidence that heterozygote imbalance does not directly track with amplicon size

Future: continuing developing rapid PCR protocols

## Thank you for your attention…

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

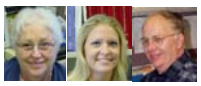
http://www.cstl.nist.gov/biotech/strbase

Peter.Vallone@nist.gov

301-975-4872

Becky Hill

John Butler | Margaret Kline



Jan Redman | Amy Decker | Dave Duewer

Funding from the **National Institute of Justice (NIJ)** through NIST Office of Law Enforcement Standards

**Collaborators**

**Mike Coble** (now AFDIL) – early miniSTR work

**Tom Reid** (DDC) - father/son samples