# Forensic Applications of Insertion-Deletion (InDel) Markers

Peter M. Vallone Ph.D.
Group Leader, Applied Genetics Group
National Institue of Stanards and Technology (NIST)
Gaithersburg, MD

## Collaborators

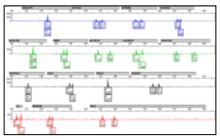
- Dr. Manuel Fondevila
  - Guest research at NIST
  - 2011 - Summer 2012
- Dr. Rui Pereira
  - HID-38plex

## Outline

- Forensic Markers (STRs, SNPs, InDels)
- Typing SNPs and InDels
- 30 and 38plex InDel Assays
  - Characterizing assay performance
  - Allele frequencies for U.S. population samples

## Forensic Markers

- Length Variation
  - short tandem repeats (STRs)



CTAGTCGT**(GATA)(GATA)(GATA)**GCGATCGT

- Core STR Loci in national database
- PCR product sizes range from 100-500bp
- Commercial multiplex PCR kits (Promega, Life Tech, Qiagen)
- Excellent for 1-to-1 matching

## Forensic Markers

- Sequence Variation
  - **Single nucleotide polymorphisms** (SNPs)

GCTAGTCGATGCTC**(G/A)**GCGTATGCTGTAGC

- **Length Variation**
  - **Insertions-deletions** (InDels)

GCTAGTCGATGCTC-GCGTATGCTGTAGC

GCTAGTCGATGCTC**($N_x$)**GCGTATGCTGTAGC

*Typically biallelic*

## InDels

Why are we interested in using InDels?

- What are the benefits?
- What are the challenges?

## InDels

### Forensic Issues/Questions

- How many InDels = 13 to 15 STR loci?
- Multiplexing (25-50plex < 1 ng DNA)
- Databases (core loci legacy concerns)
- Platforms for InDel typing? Kits?
- Unique interpretation issues – mixtures
- Validation
- Sensitivity
- Cost

## InDels

### Advantages/Benefits

- Small PCR amplicon sizes perform better with **degraded samples**
- Lower mutation rate compared with STRs
  - ($10^{-8}$ vs. $10^{-3}$)
- Abundant in the human genome ($2 \times 10^6$)
- Can provide alternative information to STRs
  - (identity, ancestry, lineage)
- Fragment analysis typing provides a familiar workflow to STRs

## InDels

### Limitations/Challenges

- Only one commercial kit (Qiagen DIPplex)
- InDels are not currently represented in national DNA databases
  - No widely established core loci
- Mixture resolution issues/interpretation
- Larger multiplex PCR assays

## Forensic InDel Literature *2012-2013*


- A 21 marker insertion deletion polymorphism panel to study biogeographic ancestry. Zaumsegel D, Rothschild MA, Schneider PM. Forensic Sci Int Genet. 2013 Feb;7(2):305-12.
- Population genetic data of 30 autosomal indels in Central Spain and the Basque Country populations. Martín P, García O, Heinrichs B, Yurrebaso I, Aguirre A, Alonso A. Forensic Sci Int Genet. 2013 Feb;7(2):e27-30.
- Automated Genotyping of a Highly Informative Panel of 40 Short Insertion-Deletion Polymorphisms Resolved in Polyacrylamide Gels for Forensic Identification and Kinship Analysis. Pena HB, Pena SD. Transfus Med Hemother. 2012 Jun;39(3):211-216.
- **Forensic performance of two insertion-deletion marker assays. Fondevila M, Phillips C, Santos C, Pereira R, Gusmão L, Carracedo A, Butler JM, Lareu MV, Vallone PM. Int J Legal Med. 2012 Sep;126(5):725-37.**
- Indel markers: genetic diversity of 38 polymorphisms in Brazilian populations and application in a paternity investigation with post mortem material. Manta F, Caiafa A, Pereira R, Silva D, Amorim A, Carvalho EF, Gusmão L. Forensic Sci Int Genet. 2012 Sep;6(5):658-61.
- Straightforward inference of ancestry and admixture proportions through ancestry-informative insertion deletion multiplexing. Pereira R, Phillips C, Pinto N, Santos C, dos Santos SE, Amorim A, Carracedo Á, Gusmão L. PLoS One. 2012;7(1):e29684.
- A validation study of the Qiagen Investigator DIPplex® kit; an INDEL-based assay for human identification. LaRue BL, Ge J, King JL, Budowle B. Int J Legal Med. 2012 Jul;126(4):533-40.
- Capillary electrophoresis of 38 noncoding biallelic mini-Indels for degraded samples and as complementary tool in paternity testing. Pereira R, Gusmão L. Methods Mol Biol. 2012;830:141-57.
- Application of the new insertion-deletion polymorphism kit for forensic identification and parentage testing on the Czech population. Zidkova A, Horinek A, Kebrdlova V, Korabecna M. Int J Legal Med. 2013 Jan;127(1):7-10.
- Typing short amplicon binary polymorphisms: supplementary SNP and Indel genetic information in the analysis of highly degraded skeletal remains. Romanini C, Catelli ML, Borosky A, Pereira R, Romero M, Salado Puerto M, Phillips C, Fondevila M, Freire A, Santos C, Carracedo A, Lareu MV, Gusmao L, Vullo CM. Forensic Sci Int Genet. 2012 Jul;6(4):469-76.
- Discrimination power of Investigator DIPplex loci in Finnish and Somali populations. Neuvonen AM, Palo JU, Hedman M, Sajantila A. Forensic Sci Int Genet. 2012 Jul;6(4):e99-102.
- Typing of 30 insertion/deletions in Danes using the first commercial indel kit--Mentype® DIPplex. Friis SL, Børsting C, Rockenbauer E, Poulsen L, Fredslund SF, Tomas C, Morling N. Forensic Sci Int Genet. 2012 Mar;6(2):e72-4.
- A method for the analysis of 32 X chromosome insertion deletion polymorphisms in a single PCR. Pereira R, Pereira V, Gomes I, Tomas C, Morling N, Amorim A, Prata MJ, Carracedo A, Gusmão L. Int J Legal Med. 2012 Jan;126(1):97-105.
- Selection of 29 highly informative InDel markers for human identification and paternity analysis in Chinese Han population by the SNPlex genotyping system. Li C, Zhang S, Li L, Chen J, Liu Y, Zhao S. Mol Biol Rep. 2012 Mar;39(3):3143-52.
- Insertion-deletion polymorphisms--utilization on forensic analysis. da Costa Francez PA, Rodrigues EM, de Velasco AM, dos Santos SE. Int J Legal Med. 2012 Jul;126(4):491-6.

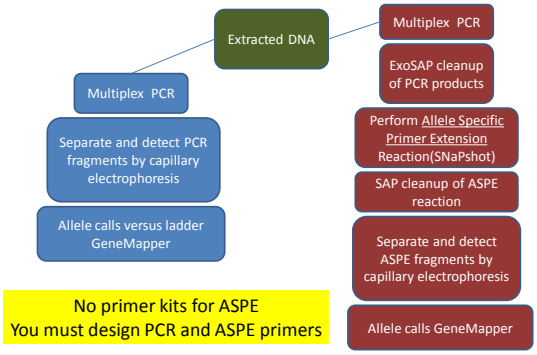

## SNP Typing Platforms

- RT-PCR (TaqMan, Light Cycler, Molecular Beacon)
- Allele Specific Primer Extension (SNaPshot)
- Mass Spectrometry (Electrospray, IBIS)
- Sequencing
- Flow Cytometry (Luminex)
- Pyrosequencing
- Ligation (SNPplex/GenPlex, Illumina)
- Invader assay
- ARMS assay (Forensic Science Service - UK)
- RFLP

Sensitivity, multiplexing, accurate typing

*Budowle 2004 FSI 139-142; Sobrino et al., 2005 FSI (epub); Dixon et al., 2005 FSI (epub)*

## STR and SNP Typing Workflows




No primer kits for ASPE
You must design PCR and ASPE primers

## Allele-Specific Primer Extension

SNaPshot assay

SNP Primer is extended by one base unit

ABI PRISM® SNaPshot™ Multiplex System

Fluorescently labeled ddNTPs + polymerase

"tail" used to vary electrophoretic mobility

Oligonucleotide primer 18-28 bases

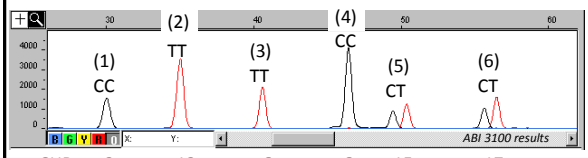
5′ 3′

PCR Amplified DNA Template

| ddNTP | Dye label | Color |
|---|---|---|
| A | dR6G | Green |
| C | dTAMRA | Black |
| G | dR110 | Blue |
| T | dROX | Red |

25 Cycles
96°C 10s
50°C 5s
60°C 30s

## 6plex SNP Assay



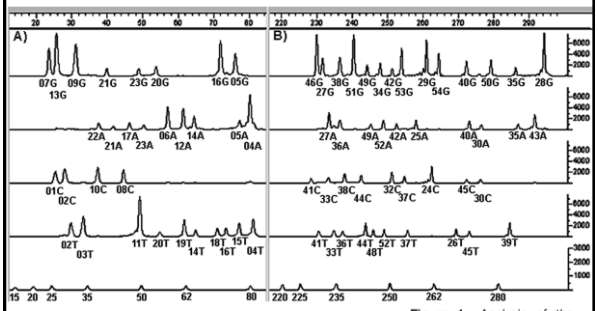

CHR: 2 13 6 6 15 17

Extension primers for autosomal SNP6plex

(1) TTTTTAGCTCCTAATTTCTTGATGGG
(2) TTTTTTTTTTCATCTGATGCCATGAGAAAGC
(3) TTTTTTTTTTTTTTGTTCTGCTTTAATACAAAACCAG
(4) TTTTTTTTTTTTTTTTTTTTTTTTATAAAGGGCAGAATGAGGATTA
(5) TTTTTTTTTTTTTTTTTTTTTTTTTTTTAGAAAGTATCTTGCAAAAGGTCCA
(6) TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTCATAATCACAGCTTTTTTCTCCCAA

## SNP*for*ID 52 plex

Sanchez, J. J., Phillips, C., Borsting, C., Balogh, K., Bogus, M., Fondevila, M., Harrison, C. D., Musgrave-Brown, E., Salas, A., Syndercombe-Court, Schneider, P. M., Carracedo, A., and Morling, N. [2006] A multiplex assay with 52 single nucleotide polymorphisms for human identification, *Electrophoresis* 27, 1713-1724.

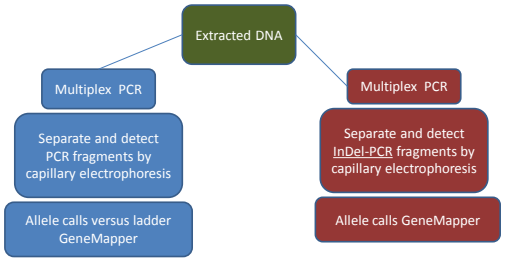



Mean match probability
European $5.06 \times 10^{-21}$

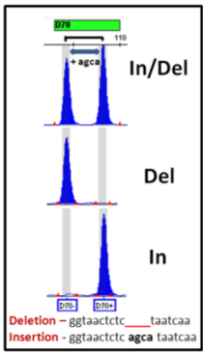
59 to 115 bp
500 pg
2 ASPE reaction

**Figure 1.** Analysis of the 52 SNP-plex assay. (A) Results of the first injection (23 SNPs). (B) Results of the second injection (29 SNPs).

## STR and InDel Typing Workflows




## Typing InDels




- As length polymorphisms, InDels can be typed with a simple direct PCR-to-CE genotyping strategy, using a single multiplexed PCR with dyed-labeled primers immediately followed by capillary electrophoresis
- Benefits of short amplicon markers (like SNPs) with a simpler workflow

## InDel Work at NIST

- Type NIST U.S. population samples (n > 700)
  - Commercial DIPplex kit
  - HID-38plex assay (from Portugal)
- Generate allele frequencies for U.S. population groups, evaluate random match probabilities
- Evaluate performance with degraded samples
- Characterization of 'off ladder' alleles
  - DIPplex kit

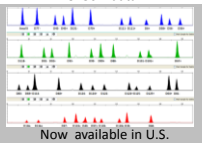
## InDel assays used in this study

3130xl data

**Qiagen Investigator DIPplex kit**

http://www.qiagen.com/products/investigatordipplexkit.aspx

- 31plex PCR
- 30 InDel markers plus Amelogenin (on 18 chromosomes)
- PCR amplicons ranging from 75 to 150 bp

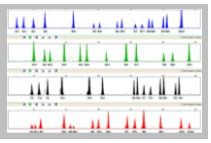


Now available in U.S.

**HID-38plex**

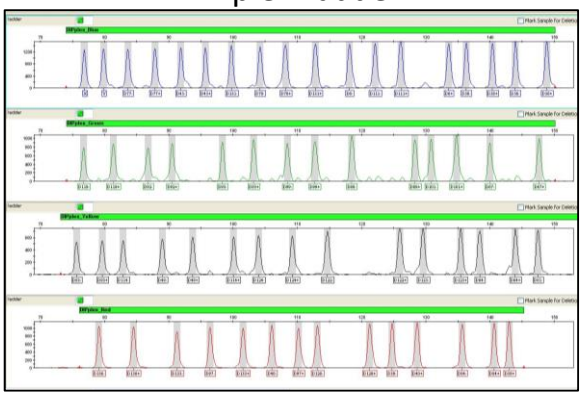
*R. Pereira Electrophoresis (2009)*

- 38plex PCR
- 38 InDel markers (on 22 chromosomes)
- PCR amplicons ranging from 50 to 155 bp



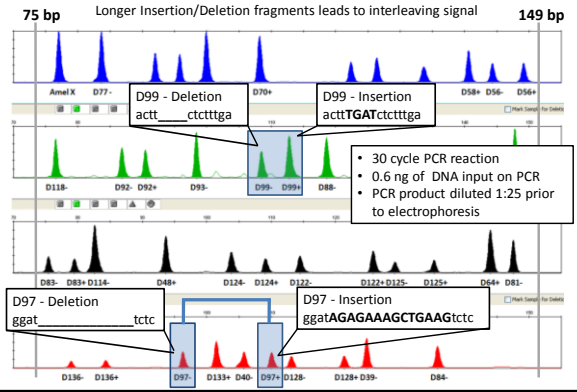
**68 InDel markers in total**

## DIPplex Ladder

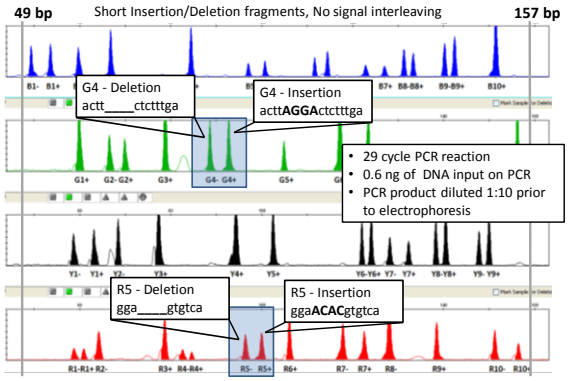


## DIPplex profile for 9947a

75 bp — Longer Insertion/Deletion fragments leads to interleaving signal — 149 bp



## HID-38plex profile for 9947a




## Allele Frequency Analysis

- We performed population allele frequency analysis with both InDel multiplexes typing the NIST collection of 712 U.S. population samples.
- Samples from the four representative human groups of the U.S. population have been used. All of them independent and of self-declared ancestry.
  - **262 African Americans**
  - **260 U.S. Caucasians**
  - **140 U.S. Hispanics**
  - **50 U.S. Asians**

Allele frequencies for each U.S. population group were generated

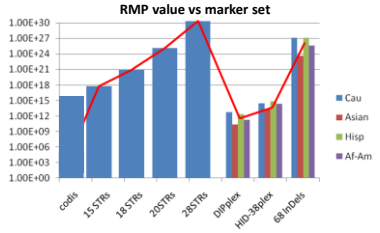
Working under the assumption of full independence of the markers, the following RMP values were calculated.

| | U.S. Cauc | U.S. Asian | U.S. Hisp | Af-Am |
|---|---|---|---|---|
| Mean DIPplex RMP | 1.86E-13 | 4.67E-11 | 4.88E-13 | 5.88E-12 |
| Mean HID-38plex RMP | 3.67E-15 | 5.11E-14 | 1.47E-15 | 4.74E-15 |
| Mean Combined RMP* | 6.79E-28 | 2.43E-24 | 7.20E-28 | 2.54E-26 |

Forensic performance of two insertion-deletion marker assays. Fondevila M, Phillips C, Santos C, Pereira R, Gusmão L, Carracedo A, Butler JM, Lareu MV, Vallone PM. Int J Legal Med. 2012 Sep;126(5):725-37.

## RMP Comparisons

- At least 50 biallelic polymorphisms are needed to reach the same power as 15 STRs
- Divergence in the power among population was detected for both InDel assays
- In spite of population divergence, 68 InDels supply with RMP as high as 20 STRs

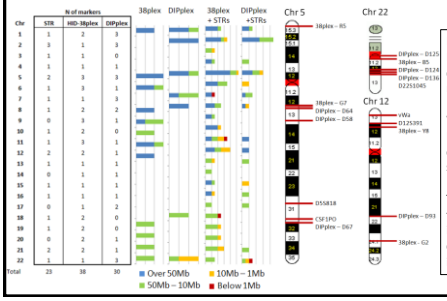



| | U.S. Cauc | U.S. Asian | U.S. Hisp | Af-Am |
|---|---|---|---|---|
| Mean DIPplex RMP | 1.86E-13 | 4.67E-11 | 4.88E-13 | 5.88E-12 |
| Mean HID-38plex RMP | 3.67E-15 | 5.11E-14 | 1.47E-15 | 4.74E-15 |
| Mean Combined RMP* | 6.79E-28 | 2.43E-24 | 7.20E-28 | 2.54E-26 |

## Genomic position of the markers: Linkage Disequilibrium possibility

- The 68 InDels are not evenly distributed on the genome
- Markers with close positions may not be independent, and could not be used together
- There is also risk of linkage with currently applied STR loci
- The accumulation of markers in small chromosomes greatly increases this risk

**Fondevila, M.**, Santos, C., Phillips, C., Carracedo, A., Butler, J.M., Lareu, M.V., Vallone, P.M. (2011) An assessment of linkage between forensic markers: core STRs, mini-STRs and InDels. *Proceedings of the 22nd International Symposium on Human Identification*




Neither HID-38plex nor DIPplex individually display great risk of marker linkage.

When the loci from both InDel assays are combined there may be significant risk of linkage.

The inclusion of STR loci to the analysis is of importance when combining STR and InDel information.

## Genomic position of the markers: Linkage Disequilibrium possibility

**Markers separated by less than 10 Mb**

6 loci from each InDel assay that are less than 10 Mb from a core STR locus.

**DIPplex**

| CHR | STR | InDel | Physical Distance |
|---|---|---|---|
| 5 | CSF1PO | rs1305056 | 6,158,834 |
| 6 | SE33 | rs2307652 | 8,521,842 |
| 8 | D8S1179 | rs3081400 | 5,959,018 |
| 15 | PentaE | rs2307433 | 7,509,680 |
| 22 | D22S1045 | rs6481 | 1,747,100 |
| 22 | D22S1045 | rs16363 | 39,169 |

**HID-38plex**

| CHR | STR | InDel | Physical Distance |
|---|---|---|---|
| 7 | D7S820 | rs2307978 | 311,150 |
| 11 | TH01 | rs10688868 | 1,890,820 |
| 12 | D12S391 | rs1610919 | 2,352,263 |
| 12 | vWA | rs1610919 | 8,838,263 |
| 16 | D16S539 | rs2067208 | 1,804,212 |
| 21 | D21S11 | rs35605984 | 4,919,264 |

When contemplating the possibility of combining the information contained in these InDel markers systems with each other or with core STR loci, we should keep in mind that the proximity between some of these markers could lead to a linkage disequilibrium state

## Artificially Degraded DNA Assay

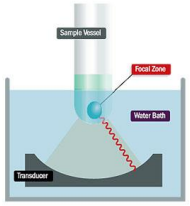
- Objective: to compare the short-amplicon InDel typing reactions to established and short amplicon STRs kit performance
- Mimic DNA degradation in a **controlled** manner
- DNA fragmentation processes were simulated
  - Temperature/time degradation process through natural oxidative damage was not predictable
  - Sonication even in long time treatments (more than 5 hours of straight sonication) didn't break the DNA to a point on which any of the STRs significantly affected

COVARIS **focused acoustic DNA shearing** evaluated as an option

# Artificially Degraded DNA Assay

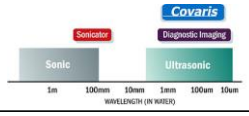
## COVARIS system

- Originally intended to be used to prepared genomic DNA libraries for next generation sequencing
- A process called Adaptive Focused Acoustics (AFA) that works by creating shock waves from a conical shaped transducer focused to converge on a small localized area
- Induces compression and expansion in a high rate cycle. This would cause collapsing forces within the liquid medium that would disrupt the DNA.




Enables precise control of the compress-expansion process in isothermal conditions.

Application of wavelengths much shorter than sonication, allows to concentrate the mechanical energy on a small volume sample.



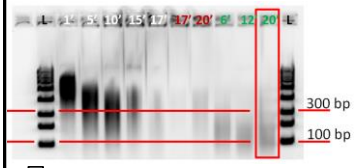

*Images from COVARIS webpage*

# Artificially Degraded DNA Assay

Treatment is defined by the following parameters:

- Cycles per Burst - the number of acoustic oscillations contained in each burst.
- Duty Factor - the percentage of active burst time in the acoustic treatment.
- Time
- Sample volume and container
- Viscosity of the medium

Several protocols were tested before reaching the desired DNA fragmentation (100-250 bp fragments)



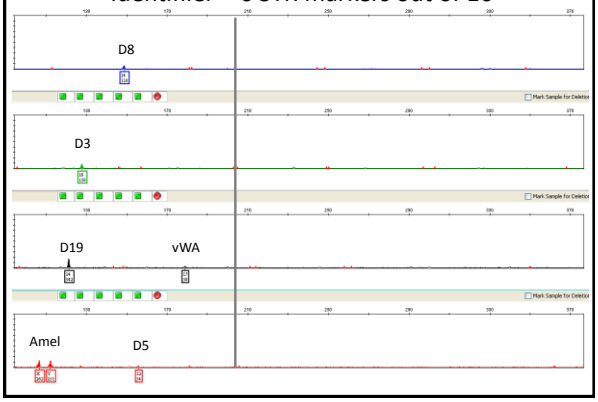

100 cycles per burst / 1 mL container
1000 cycles per burst / 1 mL container
1000 cycles per burst / 100 µL container

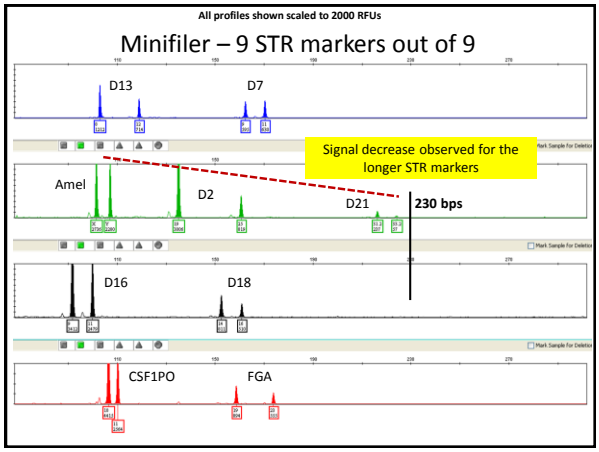
Temperature: 5 ºC
Mode: Frequency sweeping
Duty Cycle: 10%
Intensity: 10%
Cycle/Burst: 1000
Time: 20 minutes
DNA: 50 ng
Dilution volume: 100 µL
Tube: glass, 100 µL tube

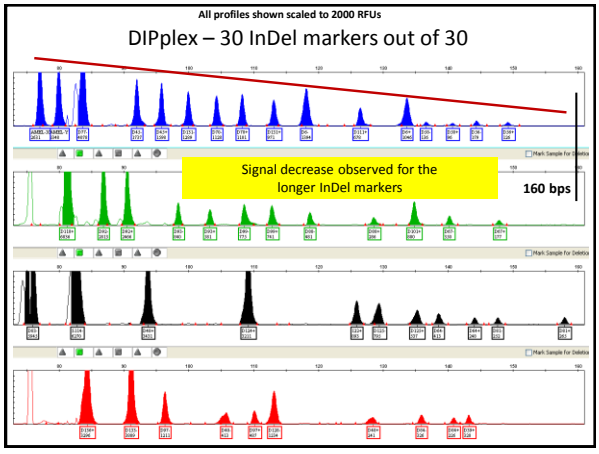
Only sample corresponding to this conditions (20') was used for the final analysis

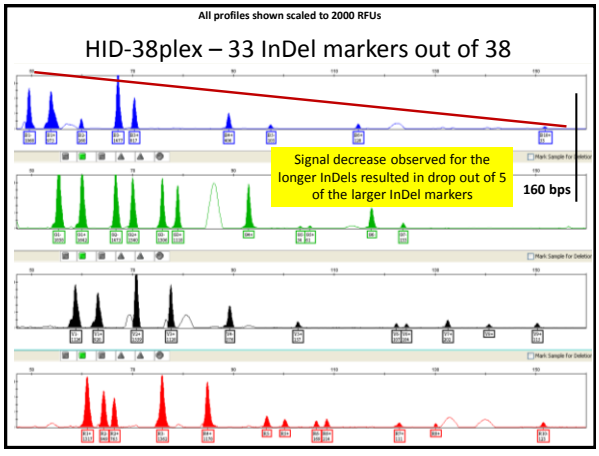
**All profiles shown scaled to 2000 RFUs**

## Identifiler – 6 STR markers out of 16




All profiles shown scaled to 2000 RFUs
Minifiler – 9 STR markers out of 9
D13
D7
Signal decrease observed for the longer STR markers
Amel
D2
D21
230 bps
D16
D18
CSF1PO
FGA


All profiles shown scaled to 2000 RFUs
DIPplex – 30 InDel markers out of 30
Signal decrease observed for the longer InDel markers
160 bps


All profiles shown scaled to 2000 RFUs
HID-38plex – 33 InDel markers out of 38
Signal decrease observed for the longer InDels resulted in drop out of 5 of the larger InDel markers
160 bps

## Artificially Degraded DNA Assay

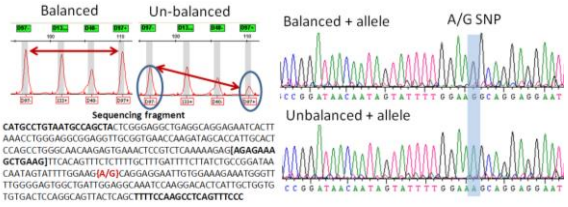
With the number of observed alleles on each kit, we obtained the following RMP values

| Assay | Exp. Alleles | Obs. Alleles | Loci total | Amp. Loci | RMP |
|---|---|---|---|---|---|
| Identifiler | 10 | 5 | 15 | 5 | n/a |
| Minifiler | 16 | 16 | 9 | 9 | $2.89\ e^{-12}$ |
| DIPplex | 49 | 49 | 30 | 30 | $4.77\ e^{-14}$ |
| HID-38plex | 43 | 43 | 38 | 33 | $1.03\ e^{-14}$ |

- Application of short amplicon markers such as DIPplex and Minifiler to challenging DNA samples would be of great interest for casework
- In case of limited amount of sample, InDel marker amplification should be considered versus other short amplicon assays, such as minifiler -- unless core STRs are needed
- For future sample preparation, increased shearing times could be tried in order to achieve a further level of fragmentation

## Further Characterization of DIPplex Loci

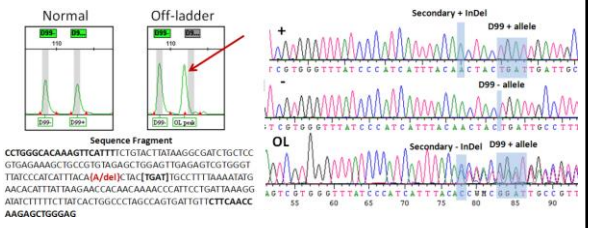
## D97- rs17238892
### heterozygote peak imbalance




- A neighboring SNP (**A/G**), located 61 bp downstream from the main InDel site. This is a SNP referenced in the dbSNP database as **rs17245568**. The A allele of this SNP corresponds to the samples carrying the observed imbalance
- We do not have the Qiagen PCR primer sequences It is reasonable to assume that the A/G SNP 61 bases downstream from the insertion is the cause of the peak imbalance

## D99 – rs2308163

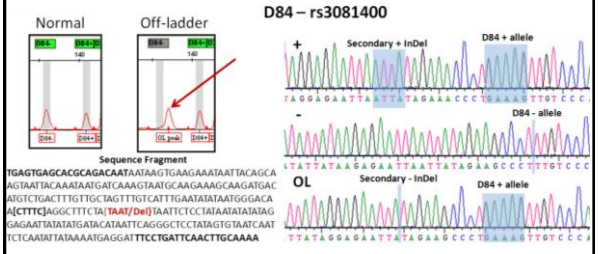
### Off-ladder allele




The off-ladder band proven a regular **Insertion** allele with a neighboring single-base deletion of an **A** residue, located 4 bp upstream from the InDel site. This is a referenced InDel on dbSNP database as **rs11346981**.

## D84 – rs3081400

### Off-ladder allele




The off-ladder band proven to be a regular **Insertion** allele with a neighboring 4 bp (**ATTA**) deletion located 10 bases downstream the main InDel site. This is a referenced InDel on dbSNP database as **rs11573892**.

## Sequencing of previously unreported variation

**Observed frequency of the unreported variation**

| Frequency | European | Population African | Hispanic | Asian |
|---|---|---|---|---|
| D97 inbalance | **0,044** | **0,22** | **0,062** | **0,06** |
| D83 inbalance | 0 | **0,08** | **0,015** | 0 |
| D99 OL allele | 0 | **0,0766** | **0,0156** | 0 |
| D84 OL allele | 0 | **0,0443** | 0 | 0 |

- We would suggest a reformulation of the reverse primer for the marker D97, as nearly as much as a quarter of the analyzed African-American samples displayed imbalance
- This situation may lead, especially in degraded DNA samples, to the drop-out of the Insertion allele of this marker
- The 'off ladder' variants observed in the Qiagen DIPplex InDel set have proven to be stable and due to a single characterized polymorphic variant
- The characterization of such rarer mobility variants, far from being a hindrance, can further contribute to the informative power of InDel typing

## Summary

- InDel genotyping can be applied to forensic DNA typing as complement to STR typing
  - Simpler workflow compared to traditional SNP typing
  - Performs well with highly degraded samples
- Population frequency data for U.S. population samples have been calculated and published
- A successful protocol for artificial DNA fragmentation mimicking challenging DNA samples has been developed.
  - A comparison between long and short amplicon amplification assays has been carried out
  - InDel assays have proven to be more informative for these samples
- Unreported variation on Qiagen's DIPplex Investigator kit have been characterized
  - The characterization of such mobility variants would contribute to raise the informative power of the test

## Acknowledgements


- **Qiagen** for providing early access to DIPplex kits
- **Margaret Kline and Becky Hill** (NIST) for assistance with allele sequencing
- **Jennifer McDaniel** NIST) for assistance with the COVARIS system
- **Carla Santos** (USC) for providing HID-38plex primer mix
- **FBI** – Lab and Biometrics Center of Excellence for funding


## Thanks for your attention!

Questions?


peter.vallone@nist.gov

1-301-975-4872