

DNA Mixture Interpretation Principles: Observations from a NIST Scientific Foundation Review
AAFS 2019 Workshop #10 (February 18, 2019; Baltimore, MD)

The Potential of New Technologies

Peter M. Vallone, PhD
Leader, Applied Genetics Group, NIST



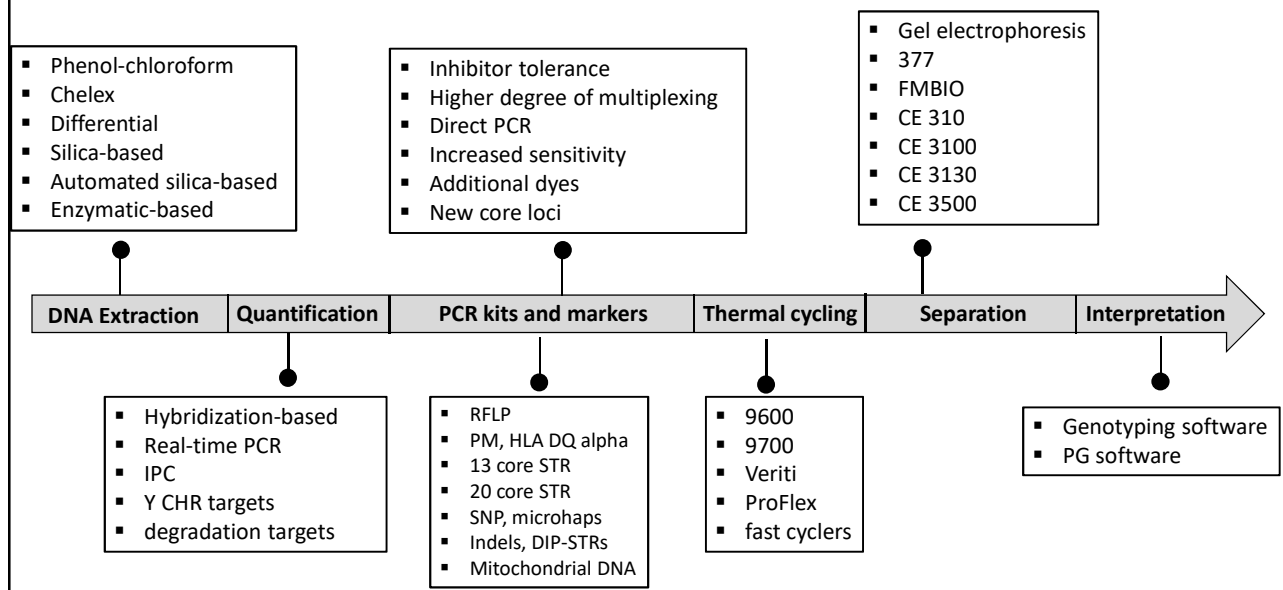
Disclaimer

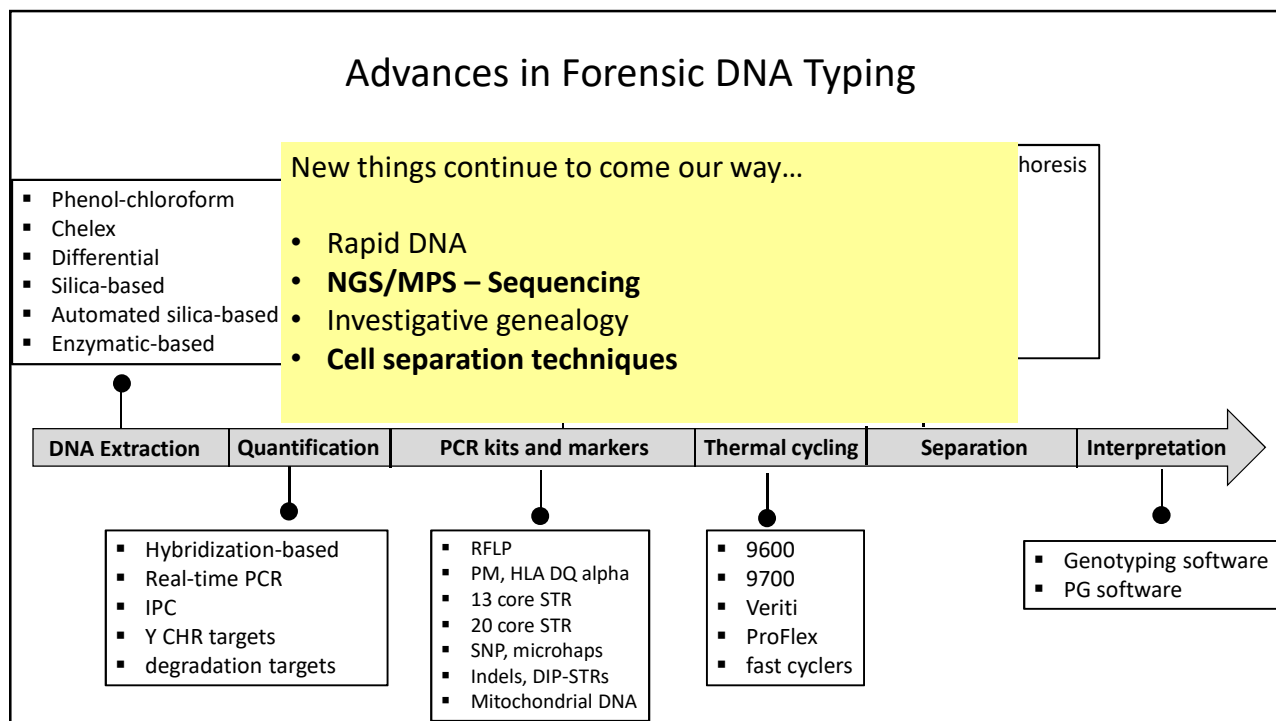
- *Points of view in this document are those of the authors and do not necessarily represent the official position or policies of the National Institute of Standards and Technology*
- *Certain commercial equipment, instruments, and materials are identified in order to specify experimental procedures as completely as possible. In no case does such identification imply a recommendation or endorsement by NIST, nor does it imply that any of the materials, instruments, or equipment identified are necessarily the best available for the purpose.*

Outline

- Adoption and implementation of new technologies
 - Advances in forensic DNA typing
 - Idealized process
- The mixture problem
 - General illustration
- Sequencing
 - STR and microhaplotype examples
- Cell separation techniques

Advances in Forensic DNA Typing

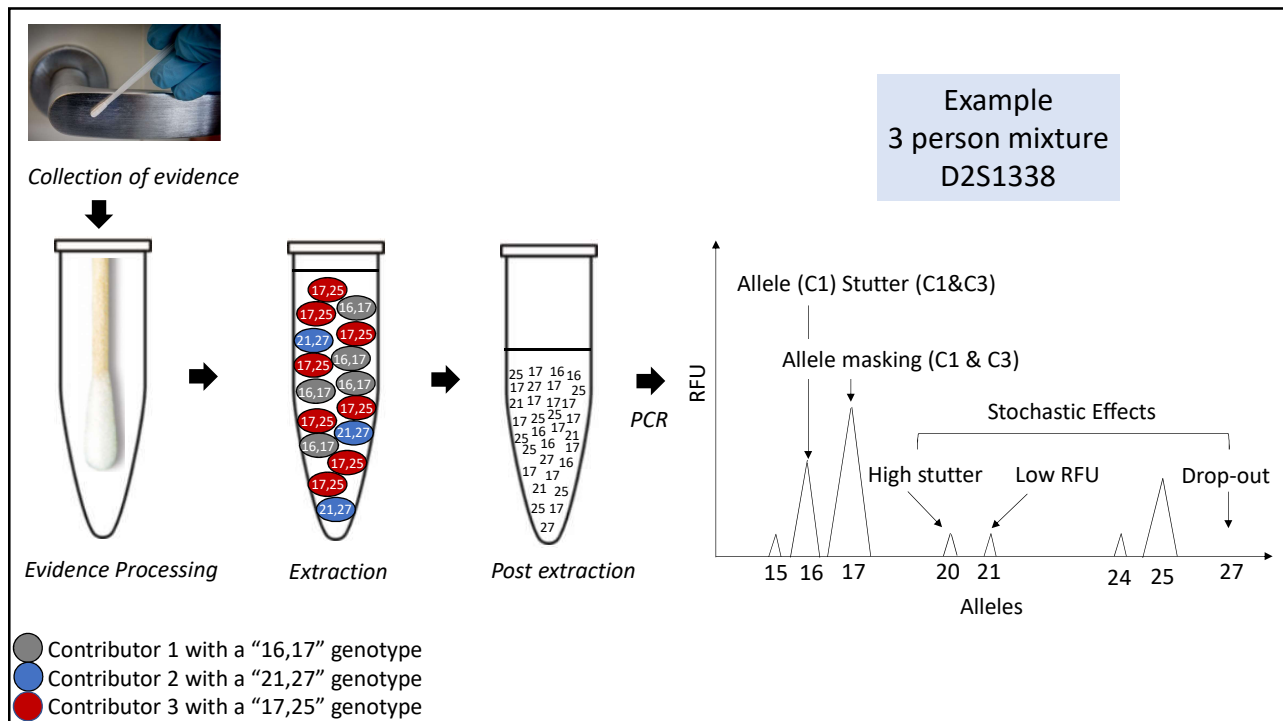
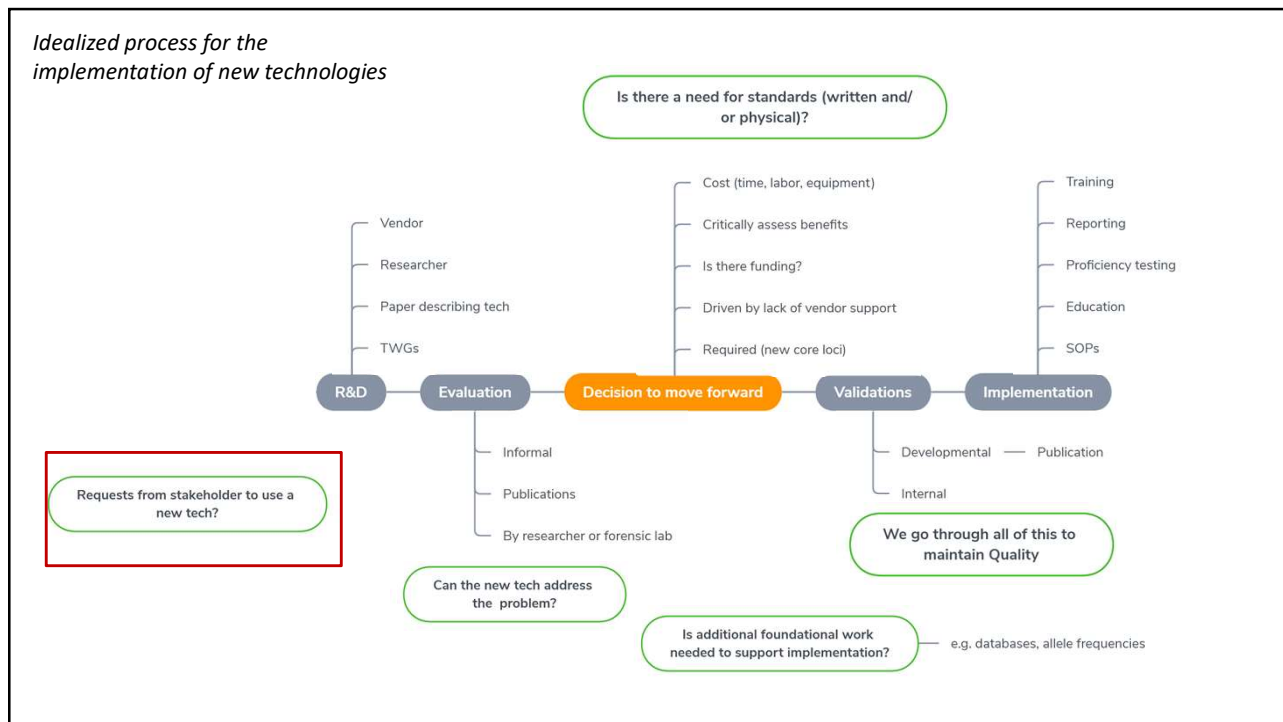




Analysis of a DNA Mixture

- DNA from two or more contributors are deposited
 - DNA may be in the cell or cell free
- Post DNA **extraction**, the alleles from all contributors are mixed together
 - DNA may be lost/reduced in the purification process
- **PCR** amplifies the alleles present post DNA extraction
 - Stochastic effects, degradation, inhibitors
- Currently PCR amplicons are **separated and detected** by CE methods
- **Interpretation** of the data (community is moving toward probabilistic genotyping)

Challenges as we address more difficult cases: touch DNA, lower template amounts, more contributors...



Sequencing

Next-generation sequencing (NGS)
Massively parallel sequencing (MPS)

Current NGS/MPS platforms and assays allow for the typing of forensically-relevant STR and SNP marker systems



*Verogen FGx platform
ForenSeq panel (STRs, SNPs, mito)*



*Thermo Fisher S5 platform
Precision ID panels (STRs, SNPs, mito)*



*PowerSeq™ 46GY System
PowerSeq™ CRM Nested System*



*Qiagen GeneReader platform
Large SNP panels and mito*

Not comprehensive

Benefits of sequencing (general)

- Can provide further resolution of STR alleles

- Increased polymorphic content
- Length (CE) -> Sequence (NGS)

19 allele -> [GGAA]11 [GGCA]8

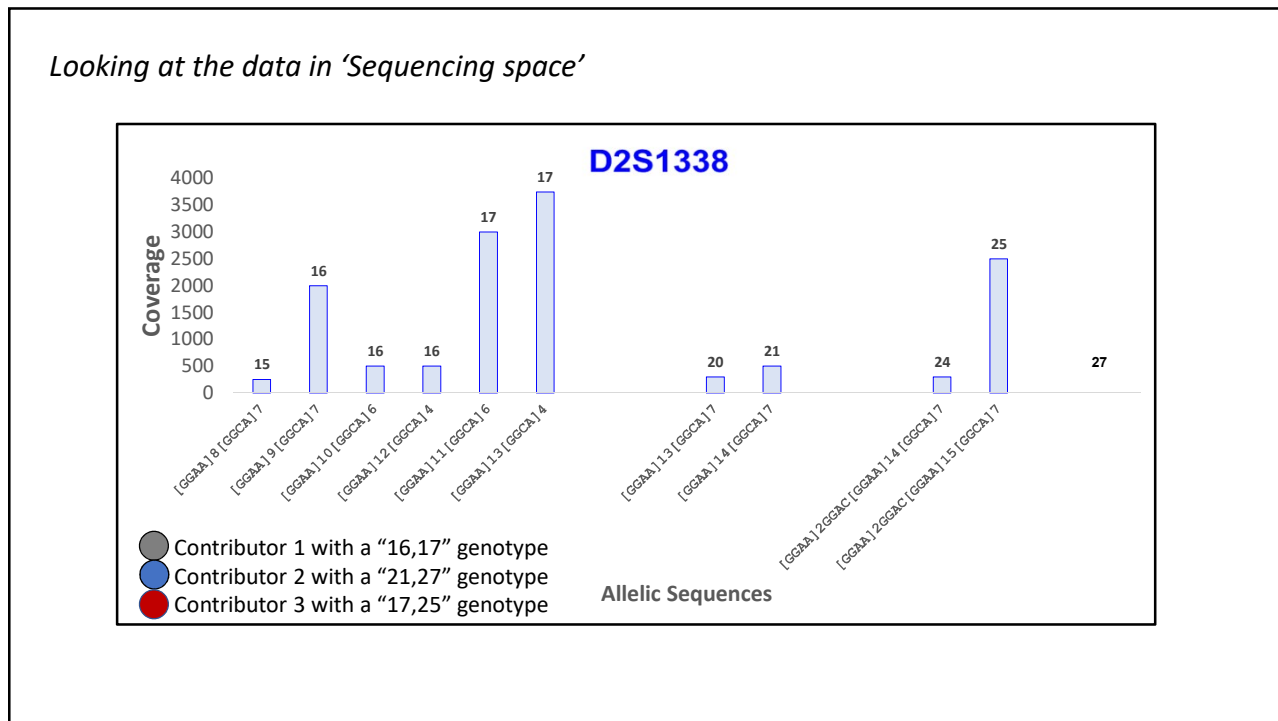
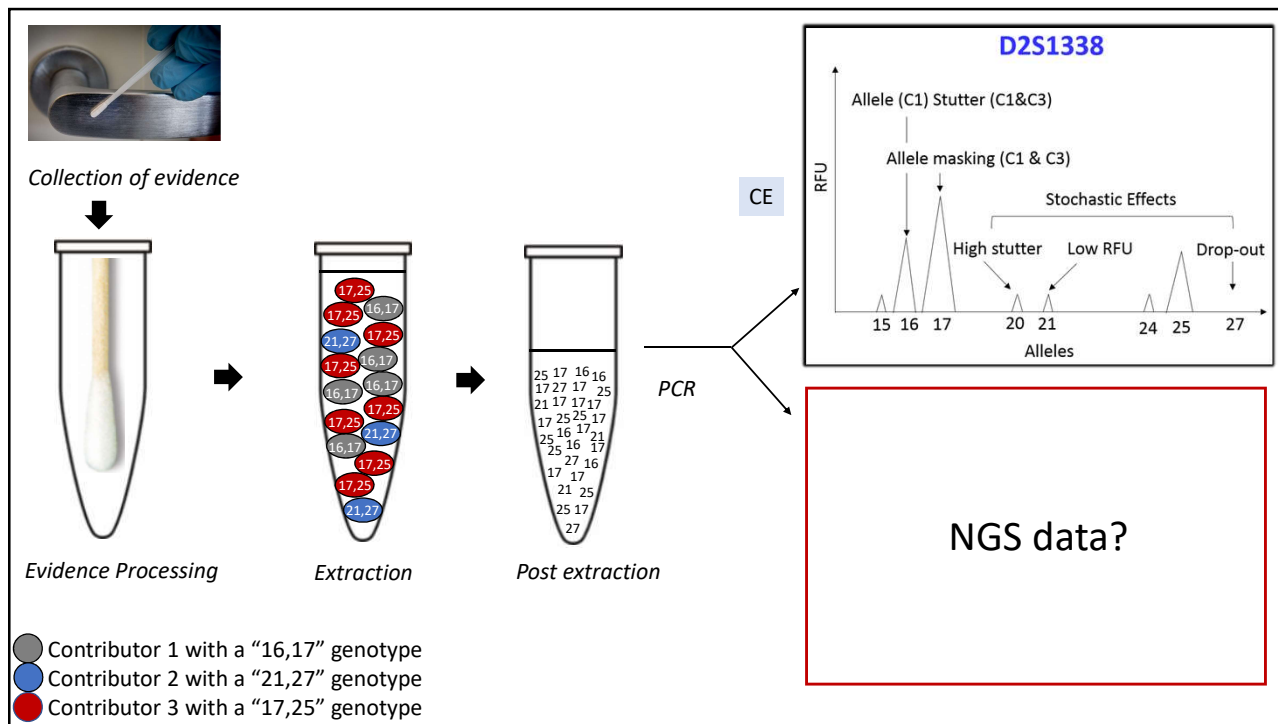
More markers
Multiplexing of samples per run
More information per run

- Technology can be applied to type additional markers systems

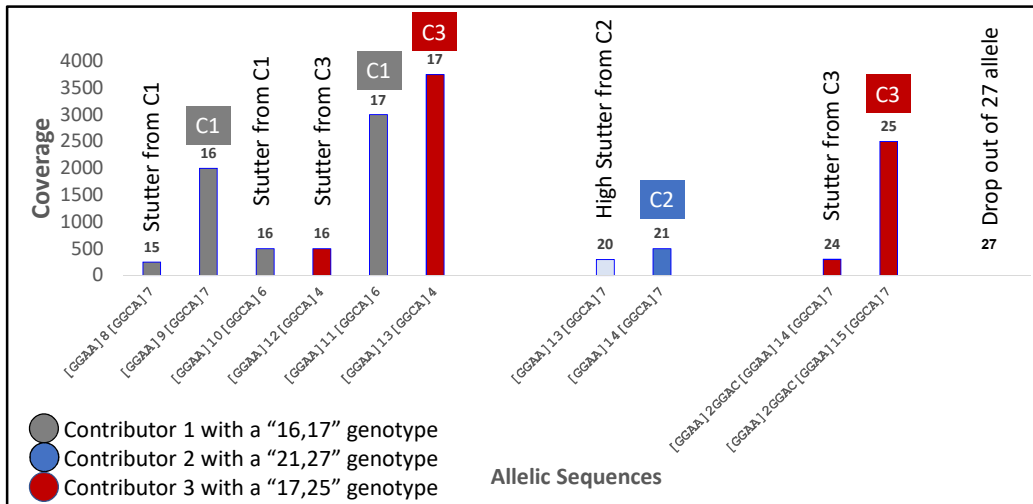
- Additional non-CODIS STR markers
- Insertion=deletion markers
- Mitochondrial (control region, whole mitochondrial genome)
- SNPs
 - Ancestry, Phenotype, ID, **Microhaplotypes**

Sequencing *STRs for Mixtures*

In Comparison to CE methods	Comment
<ul style="list-style-type: none"> • Additional alleles 	<ul style="list-style-type: none"> • “Unmasking” of alleles identical by length • Improve number of contributor estimates • Sequence-based allele frequencies are applied • Length-based alleles are back compatible to current databases
<ul style="list-style-type: none"> • Currently using targeted PCR 	<ul style="list-style-type: none"> • Comparable sensitivity to CE • May have an increased input range (> 2 ng) • Subject to stochastic effects
<ul style="list-style-type: none"> • Stutter products are sequenced 	<ul style="list-style-type: none"> • <i>Potential to correlate</i> stutter product(s) to parent allele • Allow for a more accurate modeling of stutter products
<ul style="list-style-type: none"> • Signal thresholds 	<ul style="list-style-type: none"> • Discern noise (from instrument, PCR, seq, library error) from an allele; determine an AT
<ul style="list-style-type: none"> • Artifacts 	<ul style="list-style-type: none"> • No dye artifacts; other concerns?
<ul style="list-style-type: none"> • Shorter PCR amplicons 	<ul style="list-style-type: none"> • Improved performance with degraded samples
<ul style="list-style-type: none"> • Larger multiplexes 	<ul style="list-style-type: none"> • More loci can be analyzed (autosomal, Y, X, mito)
<ul style="list-style-type: none"> • Interpretation 	<ul style="list-style-type: none"> • A NGS-based probabilistic genotyping model for STRs?

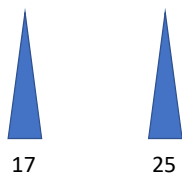


Looking at the data in 'Sequencing space'



Locus D2S1338 - Length-based allele frequencies

CE length-based information



Length	Freq	
15	0.0010	0.1%
16	0.0401	4.0%
17	0.1419	14.2%
18	0.0705	7.0%
19	0.1486	14.9%
20	0.1327	13.3%
21	0.0666	6.7%
22	0.0752	7.5%
23	0.1182	11.8%
24	0.0970	9.7%
25	0.0835	8.3%
26	0.0227	2.3%
27	0.0019	0.2%

13 alleles observed

Heterozygous genotype (17, 25)

$$2pq = 2(0.1419)(0.0835) = 0.0237$$

...1 in 42

Combined set of allele frequencies taken from Gettings et al. Sequence-based U.S. population data for 27 autosomal STR loci. Forensic Sci Int Genet. 2018 37:106-115

Locus D2S1338 – Sequence-based allele frequencies

Sequence-based information

Heterozygous genotype

17 [GGAA]13 [GGCA]4
 25 [GGAA]2 GGAC [GGAA]15 [GGCA]7

$$2pq = 2(0.0029)(0.0734) = 0.0004$$

...1 in 2,349

15[GGAA]10 [GGCA]5	0.0010	21[GGAA]14 [GGCA]7	0.0256
16[GGAA]10 [GGCA]6	0.0217	21[GGAA]2 GGAC [GGAA]11 [GGCA]7	0.0232
16[GGAA]12 [GGCA]4	0.0145	21[GGAA]13 [GGCA]8	0.0082
16[GGAA]9 [GGCA]7	0.0019	21[GGAA]2 GGAC [GGAA]12 [GGCA]6	0.0068
16[GGAA]11 [GGCA]5	0.0019	21[GGAA]15 [GGCA]6	0.0014
17[GGAA]11 [GGCA]6	0.1366	21[GGAA]12 [GGCA]9	0.0005
17[GGAA]13 [GGCA]4	0.0029	21[GGAA]16 [GGCA]5	0.0005
17[GGAA]10 [GGCA]7	0.0014	21[GGAA]17 [GGCA]4	0.0005
17[GGAA]12 [GGCA]5	0.0010	22[GGAA]2 GGAC [GGAA]12 [GGCA]7	0.0410
18[GGAA]12 [GGCA]6		22[GGAA]2 GGAC [GGAA]13 [GGCA]6	0.0145
18[GGAA]11 [GGCA]7	= 14.2%	22[GGAA]15 [GGCA]7	0.0101
18[GGAA]14 [GGCA]4	0.0024	22[GGAA]14 [GGCA]8	0.0043
18[GGAA]13 [GGCA]5	0.0019	22[GGAA]13 [GGCA]9	0.0039
18[GGAA]8 GAAA [GGAA]2 [GGCA]7	0.0010	22[GGAA]16 [GGCA]6	0.0010
18[GGAA]15 [GGCA]3	0.0005	22[GGAA]2 GGAC [GGAA]14 [GGCA]5	0.0005
19[GGAA]12 [GGCA]7	0.1076	23[GGAA]2 GGAC [GGAA]13 [GGCA]7	0.0960
19[GGAA]13 [GGCA]6	0.0333	23[GGAA]2 GGAC [GGAA]14 [GGCA]6	0.0130
19[GGAA]11 [GGCA]8	0.0024	23[GGAA]16 [GGCA]7	0.0029
19[GGAA]14 [GGCA]5	0.0024	23[GGAA]14 [GGCA]9	0.0024
19[GGAA]2 GGAC [GGAA]10 [GGCA]6	0.0014	23[GGAA]2 GGAC [GGAA]12 [GGCA]8	0.0019
19[GGAA]9 GAAA [GGAA]2 [GGCA]7	0.0005	23[GGAA]15 [GGCA]8	0.0019
19[GGAA]11 GAAA [GGCA]7	0.0005	24[GGAA]2 GGAC [GGAA]14 [GGCA]7	0.0835
19[GGAA]16 [GGCA]3	0.0005	24[GGAA]2 GGAC [GGAA]15 [GGCA]6	0.0106
20[GGAA]13 [GGCA]7	0.0893	24[GGAA]2 GGAC [GGAA]13 [GGCA]8	0.0024
20[GGAA]2 GGAC [GGAA]10 [GGCA]7	0.0121	24[GGAA]15 [GGCA]9	0.0005
20[GGAA]14 [GGCA]6	0.0092	25[GGAA]2 GGAC [GGAA]15 [GGCA]7	0.0734
20[GGAA]10 GAAA [GGAA]2 [GGCA]7	0.0087	25[GGAA]2 GGAC [GGAA]14 [GGCA]8	0.0072
20[GGAA]12 GAAA [GGCA]7	0.0068	25[GGAA]2 GGAC [GGAA]16 [GGCA]6	0.0029
20[GGAA]12 [GGCA]8	0.0043	26[GGAA]2 GGAC [GGAA]16 [GGCA]6	
20[GGAA]16 [GGCA]4	0.0010	26[GGAA]2 GGAC [GGAA]15 [GGCA]7	= 8.3%
20[GGAA]2 GGAC [GGAA]11 [GGCA]6	0.0005	26[GGAA]2 GGAC [GGAA]17 [GGCA]6	0.0014
20[GGAA]2 GGAC [GGAA]9 AGAA [GGCA]7	0.0005	26[GGAA]2 GGAC [GGAA]18 [GGCA]5	0.0005
20[GGAA]15 [GGCA]5	0.0005	27[GGAA]2 GGAC [GGAA]17 [GGCA]7	0.0014
		27[GGAA]2 GGAC [GGAA]16 [GGCA]8	0.0005

Combined set of allele frequencies taken from Gettings et al. Sequence-based U.S. population data for 27 autosomal STR loci. *Forensic Sci Int Genet.* 2018 37:106-115

67 alleles observed

Microhaplotypes

Investigative Genetics
 Criteria for selecting microhaplotypes: mixture detection and deconvolution
 Kenneth K. Kidd^a and William C. Speed

Forensic Science International: Genetics 12 (2014) 215–224
 Contents lists available at ScienceDirect
 Forensic Science International: Genetics
 journal homepage: www.elsevier.com/locate/fig

Current sequencing technology makes microhaplotypes a powerful new type of genetic marker for forensics
 Kenneth K. Kidd^{a,*}, Andrew J. Pakstis^a, William C. Speed^a, Robert Lagacé^b, Joseph Chang^b, Sharon Wootton^b, Eva Haigh^a, Judith R. Kidd^a

^a Department of Genetics, Yale University School of Medicine, New Haven, CT 06520-8005, USA
^b Human Identification Group, Thermo Fisher Scientific, 1300 Oyster Point Blvd., South San Francisco, CA 94080

Forensic Science International: Genetics 28 (2019) 54–69
 Contents lists available at ScienceDirect
 Forensic Science International: Genetics
 journal homepage: www.elsevier.com/locate/fig

Microhaplotypes in forensic genetics
 Fabio Oldoni^a, Kenneth K. Kidd^a, Daniele Podini^{b,*}

^a Department of Forensic Science, New College of Florida, 2000 Rock Road NW, Palm Bay, FL 32909, United States
^b First Laboratory of Medicine, Department of Forensic Medicine, University of Turin, 10126, Turin, Italy

ABSTRACT
 Microhaplotype loci (microhaplotypes, MHL) are a novel type of molecular marker of less than 100 nucleotides, defined by two or more closely linked SNPs associated to multiple alleles combinations. The value of these markers is enhanced by automatic parallel sequencing (MPS), which allows the sequencing of high spatial haplotypes at each of the many analyzed loci. This review describes the features of these novel DNA markers and discusses their value in forensic genetics, focusing on identification, kinship analysis, mixture deconvolution, and forensic applications. Forensic applications also include missing person identification, vulnerability testing, and medical diagnostic applications. The technique is not restricted to humans.

1. Introduction: historical background of haplotype discovery
 1.1. Discovery of haplotype blocks in the human genome
 The term ‘haplotype’ was first introduced by Ruggieri Cappellini in the late 1910s to describe alleles within the human Y-chromosomal (HSA) region that are inherited together as a block [1]. Twenty years later the human genome project (HGP) launched an unprecedented international collaboration [2] fundamental to the study of human genetics and biomedical research. The early work focused on mapping of human and mouse genes and sequencing the genomes of significantly smaller and easily studied organisms [3–6]. This was of paramount importance for the understanding of the hereditary architecture of disease [7–9] and provided an essential scaffold for the assembly and annotation of the human genome. The publication that proclaimed the completion of the sequencing of the first draft of the human genome in

1.2. Large scale international haplotype map project
 With the goal of establishing a resource for the study of global variation in the human genome, the HapMap (2002) Human Genome Diversity Panel – Center of Excellence for Population Genomics (CEPOP) project was initiated in 2002. In a case the HapMap (2002) [10] is a collection of 1064 cultured lymphoblastoid cell lines from 52 populations of different parts of the world distributed at the Fondazione Human Genome in Paris.
 Later studies, aimed at understanding human evolution and population structure [11,12] focused on linkage disequilibrium (LD). Core haplotypes (hap maps) [13] studies identified distinct LD blocks, defined as haplotype blocks or specific genomic regions with a restricted number of haplotypes occurring due to the limited number of recombination events within these regions [13,15]. With these considerations in mind, a large-scale HapMap International [14,16,

International Journal of Legal Medicine
<https://doi.org/10.1007/s00414-019-02010-7>

ORIGINAL ARTICLE

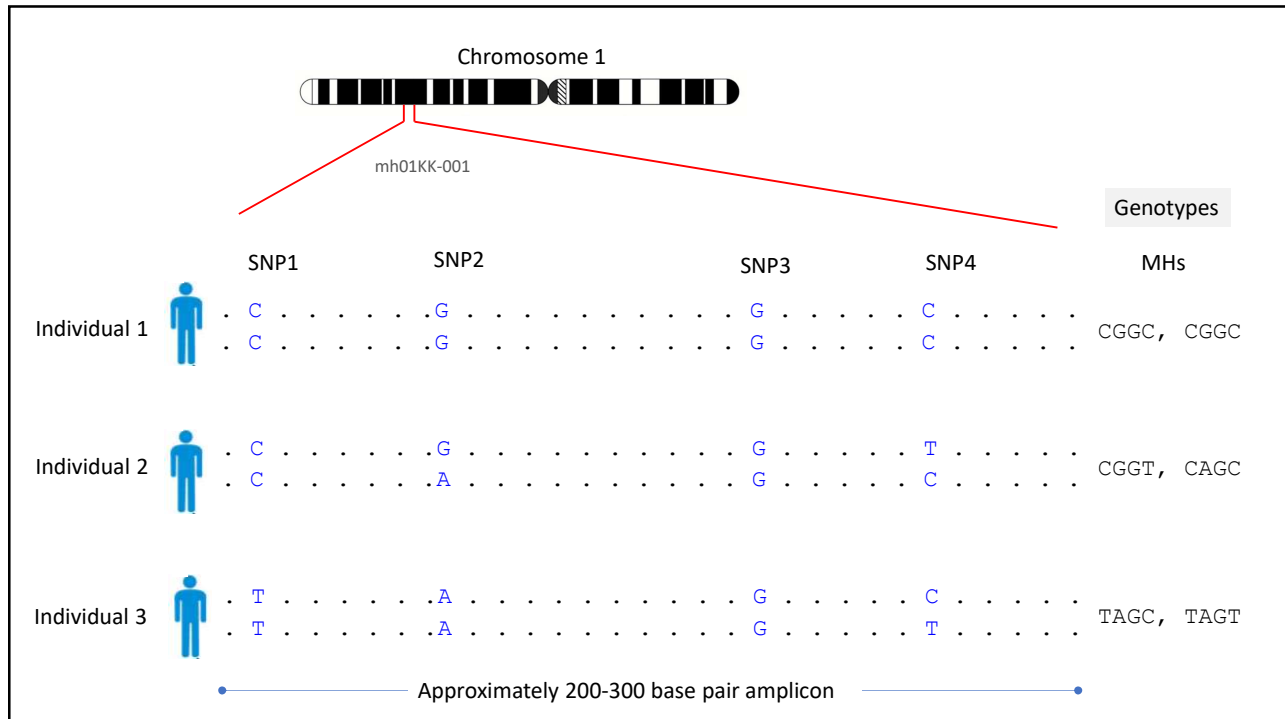
Mixture deconvolution by massively parallel sequencing of microhaplotypes
 Lindsay Bennett¹ · Fabio Oldoni² · Kelly Long² · Selena Cisana² · Katrina Madella² · Sharon Wootton³ · Joseph Chang³ · Ryo Hasegawa³ · Robert Lagacé³ · Kenneth K. Kidd⁴ · Daniele Podini²

Microhaplotypes

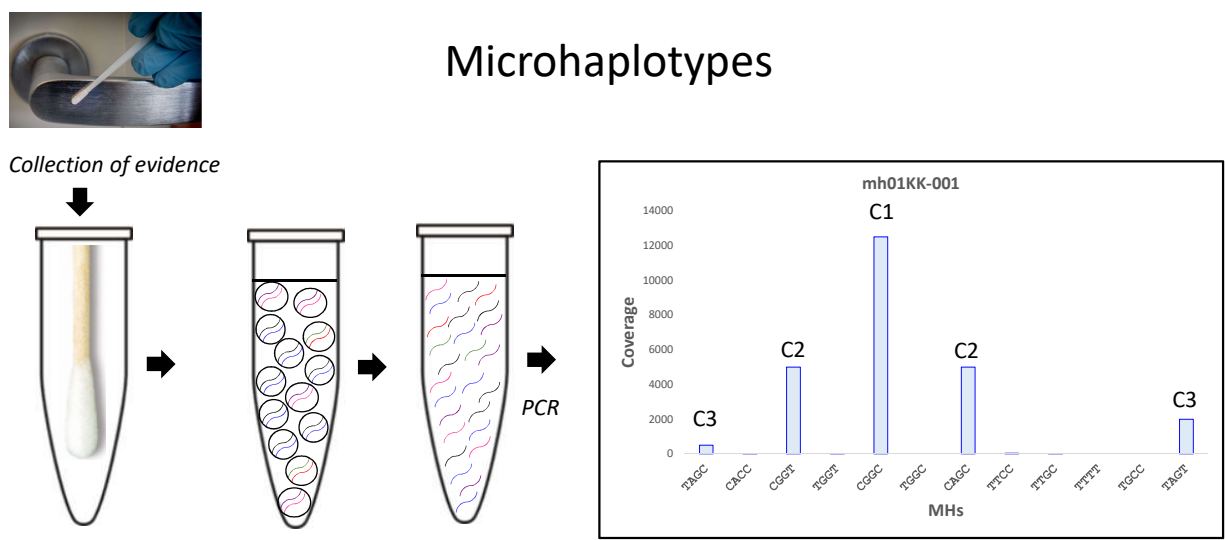
- Novel type of marker of < 300 base pairs
- Defined by two or more closely linked SNPs associated in multiple allelic combinations
 - Similar size amplicons can be used in a multiplex (not length-based alleles)
 - Each allele from a locus will be the same size
- Can also be used for ancestry prediction
- (Typically) fewer alleles than a STR locus

Absence of stutter as the alleles are SNP-defined versus a repeating motif

- Still enriched by targeted PCR
- Allele frequencies are published (and still being generated)
- No core law enforcement database
- Will require a framework for interpretation (probabilistic genotyping?)



Microhaplotypes

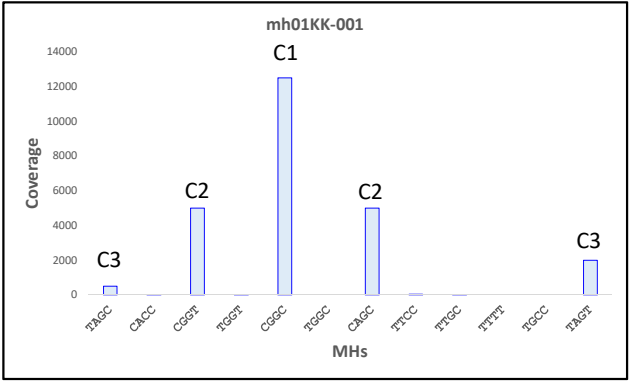


Collection of evidence

Evidence Processing *Extraction* *Post extraction* *PCR*

Legend:

- Contributor 1 with a "CGGC, CGGC" microhaplotype
- Contributor 2 with a "CGGT, CAGC" microhaplotype
- Contributor 3 with a "TAGC, TAGT" microhaplotype



mh01KK-001

Coverage

MHs

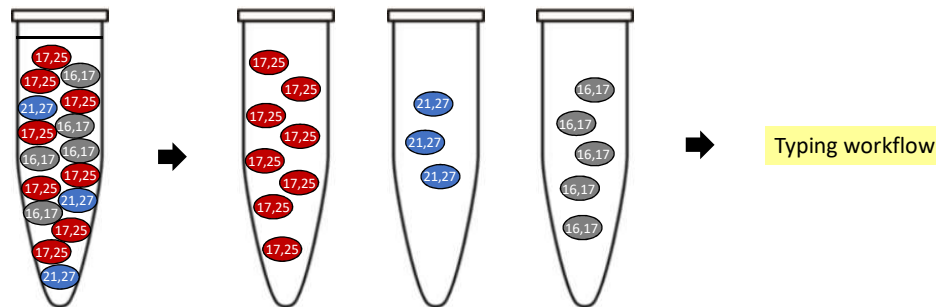
Will still observe some 'noise'
 Hb imbalance at lower levels of DNA input (C3)
No stutter artifacts

Sequencing *Microhaplotypes for Mixtures*

In Comparison to CE/STR methods	Comment
<ul style="list-style-type: none"> Additional Microhaplotype-based alleles 	<ul style="list-style-type: none"> "Unmasking" of alleles identical by length • Improve number of contributor estimates • Microhaplotype-based allele frequencies are applied Length-based alleles are back compatible to current databases
<ul style="list-style-type: none"> • Currently using targeted PCR 	<ul style="list-style-type: none"> • Comparable sensitivity to CE • May have an increased input range (> 2 ng) • Subject to stochastic effects
<ul style="list-style-type: none"> Stutter products are sequenced • No stutter artifacts 	<ul style="list-style-type: none"> Potential to correlate stutter product(s) to parent allele Allow for a more accurate modeling of stutter products.
<ul style="list-style-type: none"> • Signal thresholds 	<ul style="list-style-type: none"> • Discern noise (from instrument, PCR, seq, library error) from an allele; determine an AT
<ul style="list-style-type: none"> • Artifacts 	<ul style="list-style-type: none"> • No dye artifacts; other concerns?
<ul style="list-style-type: none"> • Shorter PCR amplicons (compared to CE) 	<ul style="list-style-type: none"> • Improved performance with degraded samples
<ul style="list-style-type: none"> • Larger multiplexes 	<ul style="list-style-type: none"> • A need for microhaplotype loci
<ul style="list-style-type: none"> • Interpretation 	<ul style="list-style-type: none"> • A NGS-based probabilistic genotyping model for MHs?

Physical Separation of Cells

General concept – physical separation/sorting of cells before DNA typing workflow



Physical Separation of Cells – How?!

- Some proposed methods
 - Partition into microreactors
 - Micro-manipulation (needle, laser)
 - Sort based on cell morphology or tagging

Challenges

- Dried cells are more challenging than fresh solutions
- Single cell sampling methods are lower throughput, may require PCR optimization
- Consider the specificity/sensitivity of reagents that bind cells (antibodies)
- Is there DNA in the cell (and what about cell free DNA)?

Short communication Forensic Science International: Genetics 17 (2015) 8–16

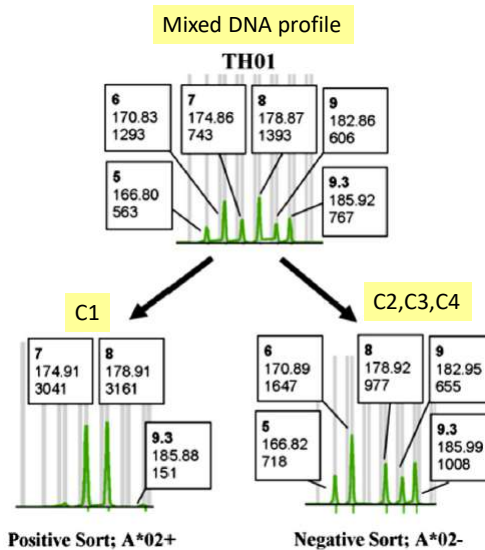
Separation of uncompromised whole blood mixtures for single source STR profiling using fluorescently-labeled human leukocyte antigen (HLA) probes and fluorescence activated cell sorting (FACS)

Lee Dean^a, Ye Jin Kwon^a, M. Katherine Philpott^a, Cristina E. Stanciu^a, Sarah J. Seashols-Williams^a, Tracey Dawson Cruz^a, Jamie Sturgill^b, Christopher J. Ehrhardt^{a,b}

^a Department of Forensic Science, Virginia Commonwealth University, 1015 Floyd Ave, Richmond, VA 23284, USA
^b School of Nursing, Virginia Commonwealth University, Medical College of Virginia, Richmond, VA 23284, USA

- Four person mixture (C1, C2, C3, C4)
- Binding cells with HLA allele (A*02) antibody
 - C1 is HLA allele (A*02) positive
- FACS separated
- Fractions typed

Contributor genotypes	
C1	7, 8
C2	6, 9
C3	5, 6
C4	8, 9.3



Single source DNA profile recovery from single cells isolated from skin and fabric from touch DNA mixtures in mock physical assaults

Katherine Farash^a, Erin K. Hanson^b, Jack Ballantyne^{a,b,c,*}

^a Graduate Program in Forensic Science, Biochemistry Track, Department of Chemistry, University of Central Florida, PO Box 162366, Orlando, FL 32816-2366, USA

^b National Center for Forensic Science, PO Box 162367, Orlando, FL 32816-2367, USA

^c Department of Chemistry, University of Central Florida, PO Box 162366, Orlando, FL 32816-2366, USA

Science & Justice 58 (2018) 191–199

- Avoid 'blind swabbing' that might create mixtures
- A more directed sampling approach ("Smart")
"sensitive, measurable, attainable, relevant, targeted"
- Optimized direct PCR (lysis + PCR)

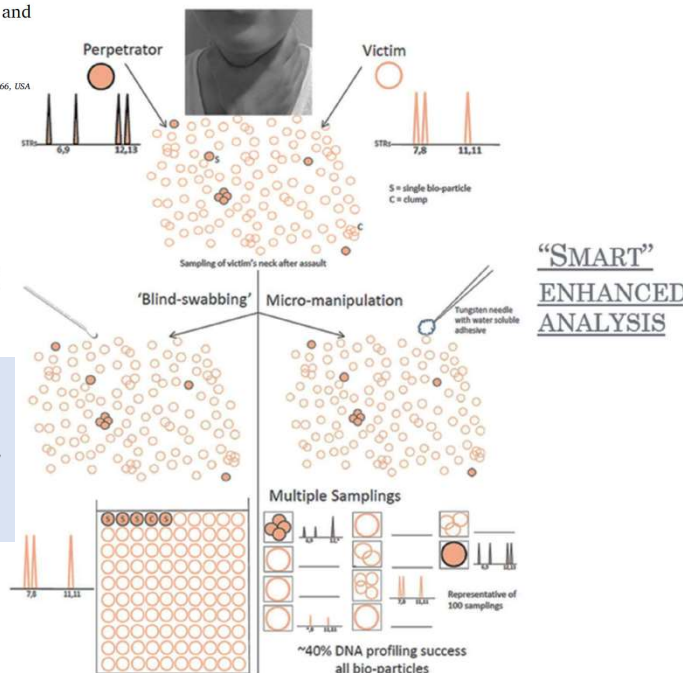
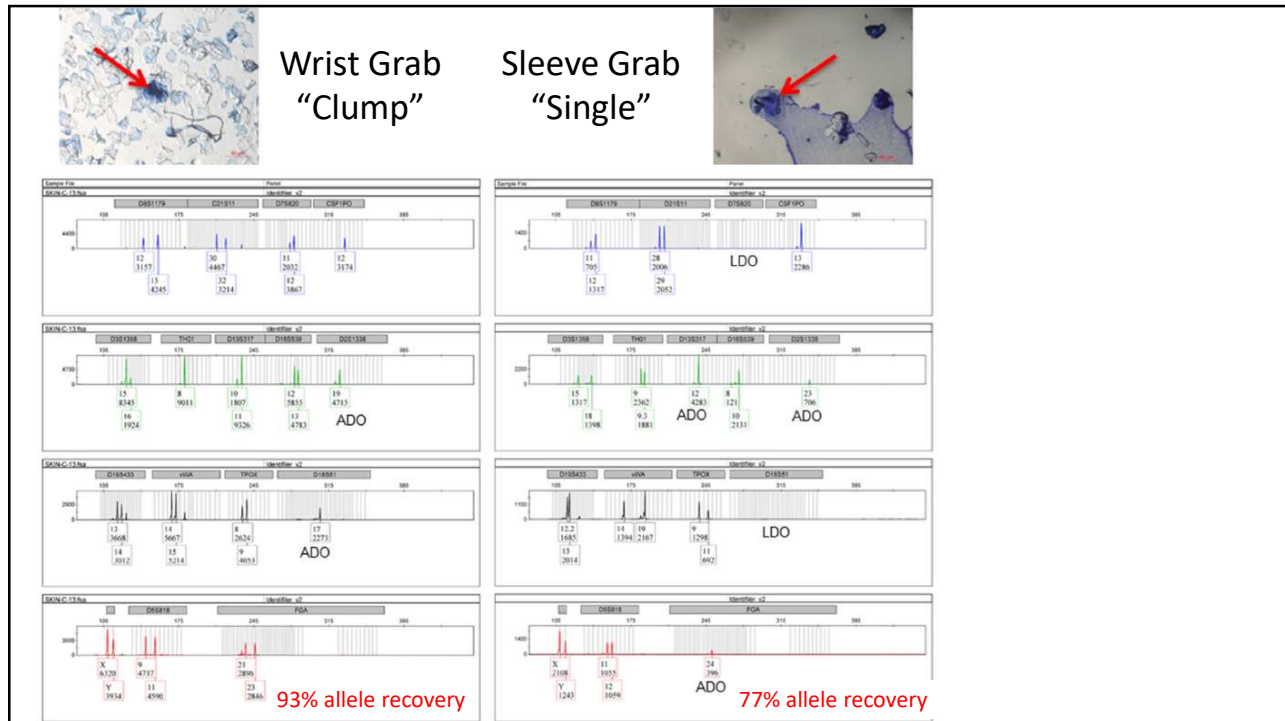


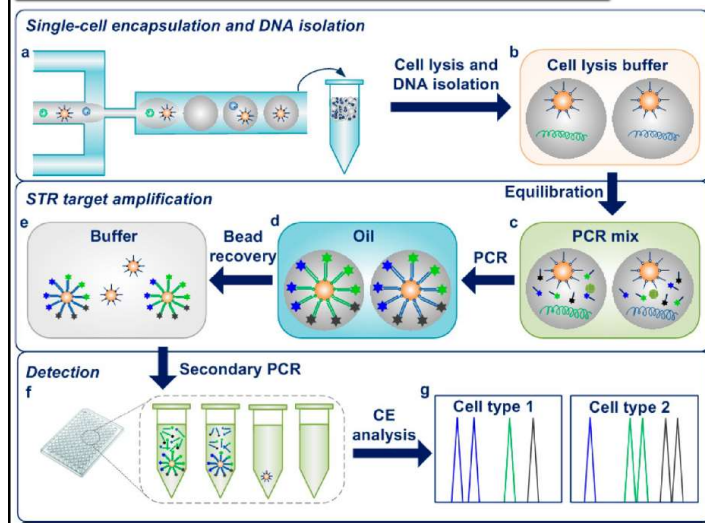
Fig. 1. Schema of standard versus "smart" enhanced analysis of touch DNA. Cells from the victim and perpetrator are indicated as unfilled and filled circles respectively.



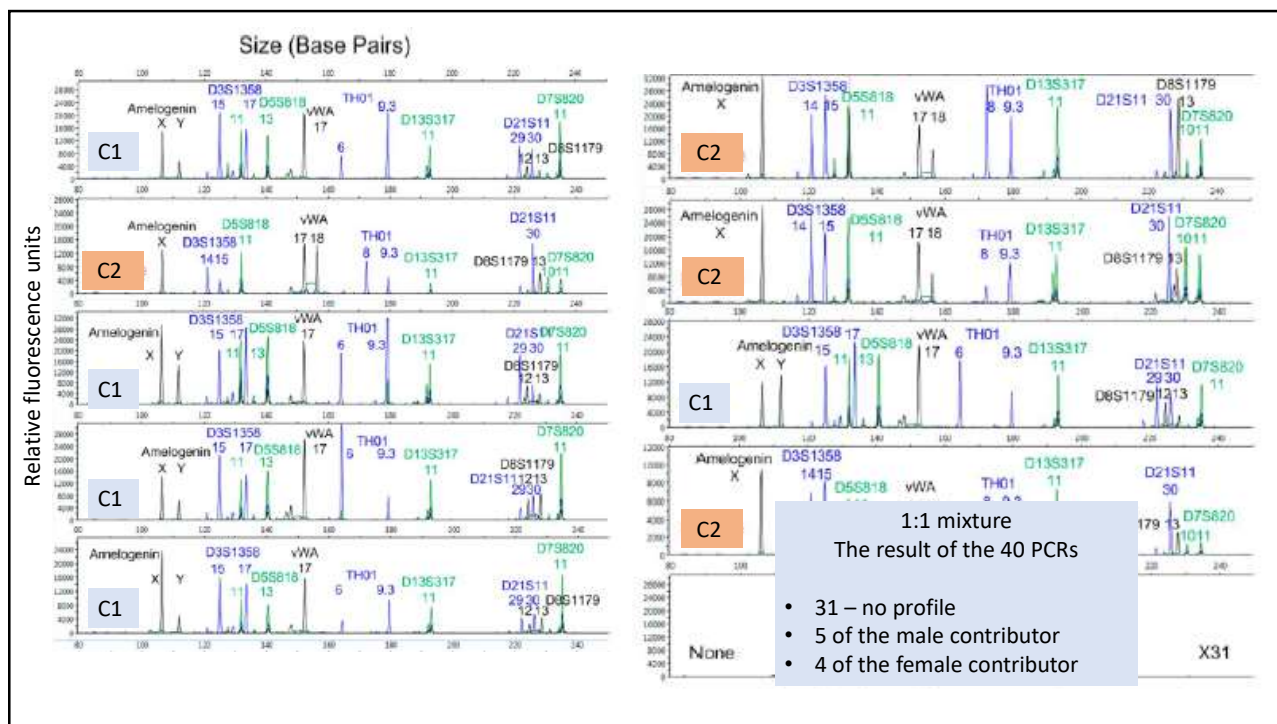
analytical chemistry
pubs.acs.org/ac

Single-Cell Forensic Short Tandem Repeat Typing within Microfluidic Droplets
 Tao Geng,¹ Richard Novak,² and Richard A. Mathies^{1,2}

¹Department of Chemistry, University of California, Berkeley, California 94720, United States
²UCSF/UC Berkeley Graduate Program in Bioengineering, University of California, Berkeley, California 94720, United States



- A cell and a primer-coated bead are encapsulated in agarose droplets
- Cell lysis and PCR take place in the droplet
- Droplets are dissolved and beads containing the PCR amplicons are diluted further for another round of PCR
- Products are separated and detected by CE
- CE runs will show: either no profile or a single source profile from the contributors



Thoughts and considerations

- Define and understand the problem
- Understand the potential of the new technology to address the problem
- Consider the cost of implementation as a whole
- There is always something new on the horizon

Thank you for your attention! Questions?



Contact: Peter.Vallone@nist.gov

Thanks to

- **Dr. Sarah Riman (NIST)**
- **Members of the NIST mixture review team (John, Hari, Rich, Melissa, and Sheila)**
- **Funding**
 - NIST Special Programs Office: *Forensic DNA*