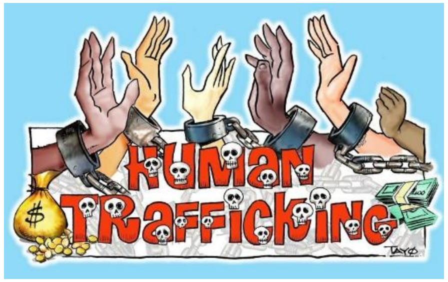
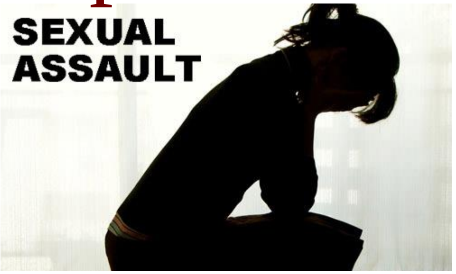
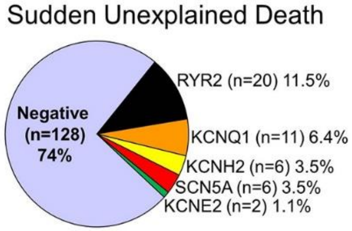
UNT | HEALTH SCIENCE CENTER

Forensic Genomics Innovations and Their Impact on Public Safety







© MAYO FOUNDATION FOR MEDICAL EDUCATION AND RESEARCH. ALL RIGHTS RESERVED.

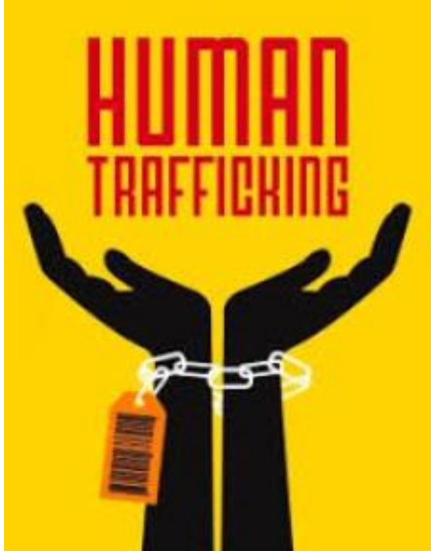
Bruce Budowle
Center for Human Identifications
Department of Microbiology, Immunology, and Genetics,
University of North Texas Health Science Center
Fort Worth, Texas USA

Disclaimer

Speaker was provided travel and hotel support by Thermo Fisher Scientific for this presentation, but no remuneration. When used for purposes other than Human Identification or Paternity Testing the instruments and software modules cited are for Research Use Only. Not for use in diagnostic procedures. Thermo Fisher Scientific and its affiliates are not endorsing, recommending, or promoting any use or application of Thermo Fisher Scientific products presented by third parties during this seminar. Information and materials presented or provided by third parties are provided as-is and without warranty of any kind, including regarding intellectual property rights and reported results. Parties presenting images, text and material represent they have the rights.

Human Trafficking

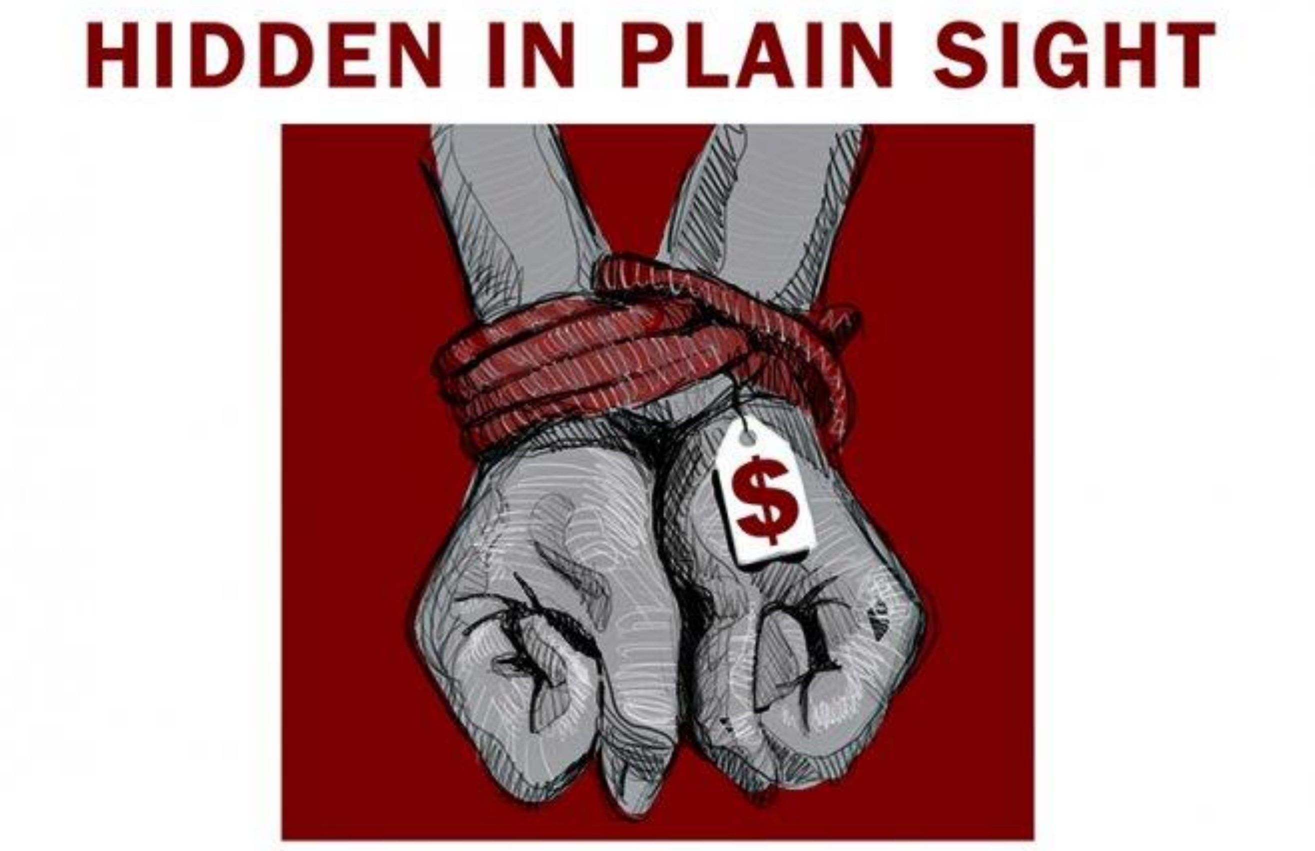
- The Trafficking Victims Protection Act
- Severe Forms of trafficking
  - Sex trafficking in which a commercial sex act is induced by force, fraud, or coercion, or in which the person induced to perform such an act has not attained 18 years of age; or,
  - Recruitment, harboring, transportation, provision, or obtaining of a person for labor or services, through the use of force, fraud, or coercion for the purpose of subjection to involuntary servitude, peonage, debt bondage, or slavery.
- A victim need not be physically transported from one location to another in order for the crime to fall within these definitions.



Forms of Human Trafficking

- Sex Trade/Tourism
- Child Prostitution
- Forced / Bonded Labor
- Debt Bondage / Migrant Workers
- Involuntary Domestic Servitude
- Child Labor
- Child Soldiers
- Forced Marriage

Magnitude of the Problem



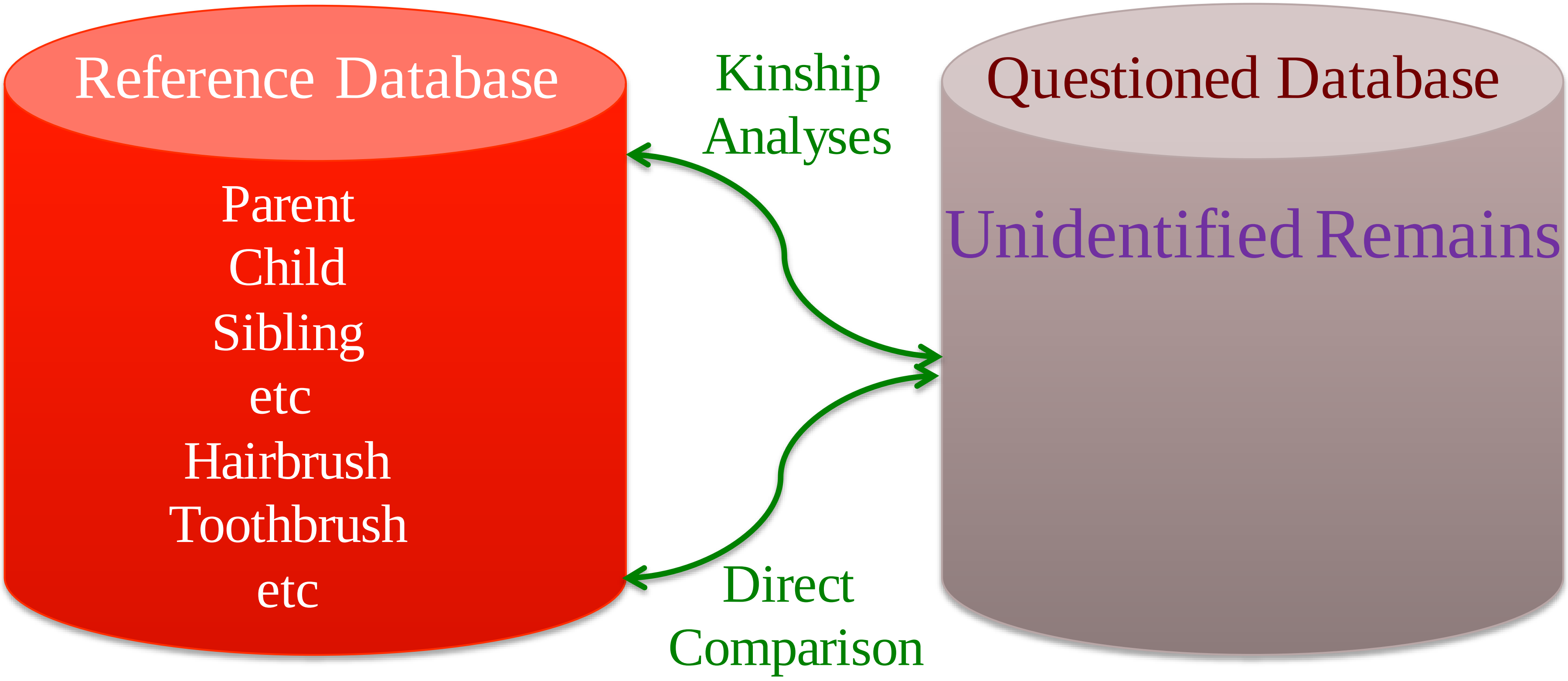
12.3 million victims at any given time
49,105 victims identified in 2009
4,166 prosecuted cases in 2009

Adoption Fraud

- The kidnapping or unlawful buying/selling of an infant or child for the purpose of offering that child for adoption
- Illegal adoptions do not necessarily involve the use of force, fraud or coercion

This image cannot curre…

Missing Persons Database



Reference Database

Parent
Child
Sibling
etc
Hairbrush
Toothbrush
etc

Kinship Analyses

Direct Comparison

Questioned Database

Unidentified Remains

Program Issues

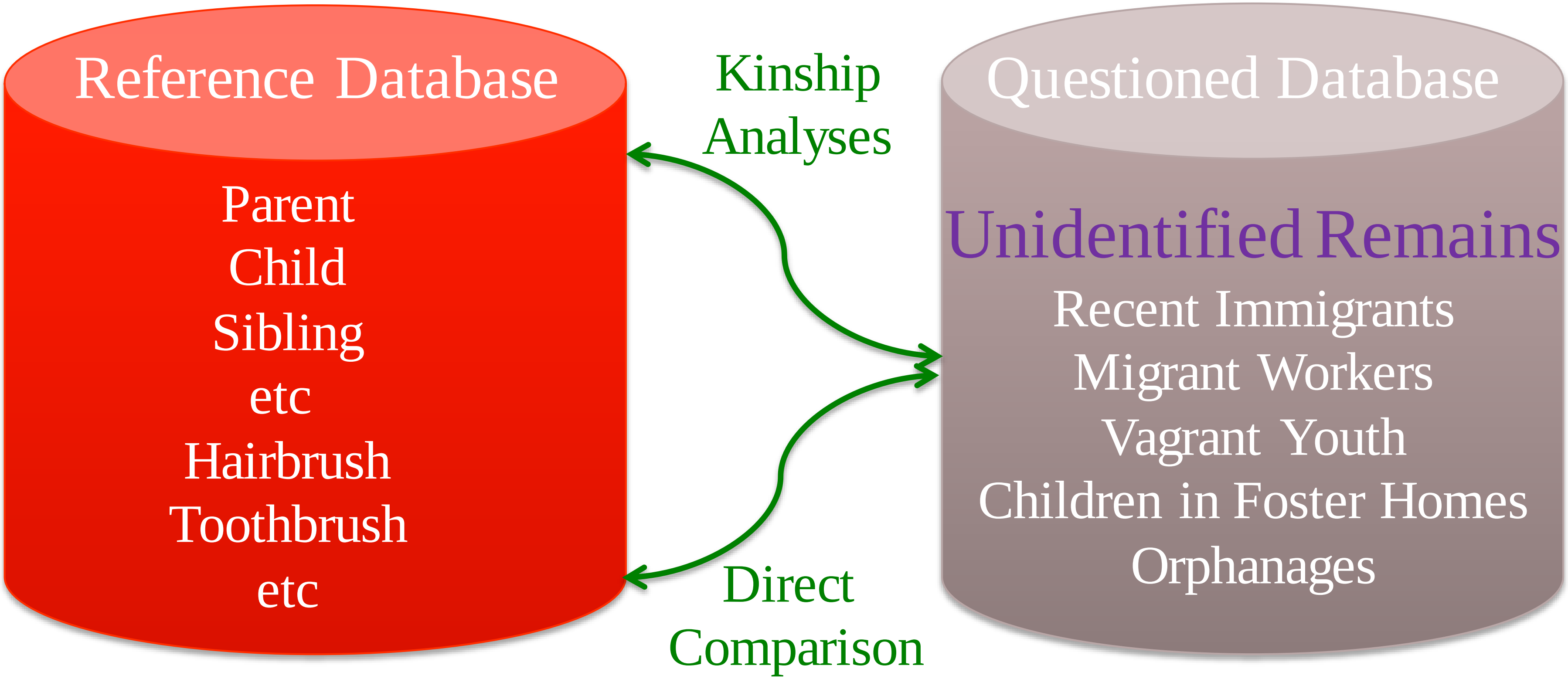
Human remains identification

- Gives family members resolution
- Provides law enforcement with critical tools to develop intelligence and investigations
- However, only protects, for example, sex workers after they become victims of homicide

Vulnerable population

- Need to understand
  - Sex workers
  - Privacy concerns surrounding DNA profiling
  - Potential benefits and risks of using a database to prevent and investigate crime
- Develop policies
  - How law enforcement may use the database
  - When and how profiles may be expunged
  - Which indices are searched in a Potential Victims Database
  - Define vulnerable segments of population

Vulnerable Population/Human Trafficking Model Database



Reference Database

Parent
Child
Sibling
etc
Hairbrush
Toothbrush
etc

Kinship Analyses

Direct Comparison

Questioned Database

Unidentified Remains

Recent Immigrants
Migrant Workers
Vagrant Youth
Children in Foster Homes
Orphanages

This image cannot curre…

Humanitarian DNA Identification Database

- Center for Human Identification (UNTCHI)
- Missing Persons Database
- Stand alone database
- CODIS Software
- Identify missing persons (border crossings)
- Exchange with Central America and Mexico
- Another way to bring some resolution to families from a tragic event

Goals of Human Trafficking Project

(Award No. S-INLEC-GR-1013)

- Decrease instances of human trafficking within and from Central America
- Decrease the effects of human trafficking on the populations of the three Northern Triangle countries (Guatemala, El Salvador, and Honduras)

DEPARTMENT OF STATE
UNITED STATES OF AMERICA

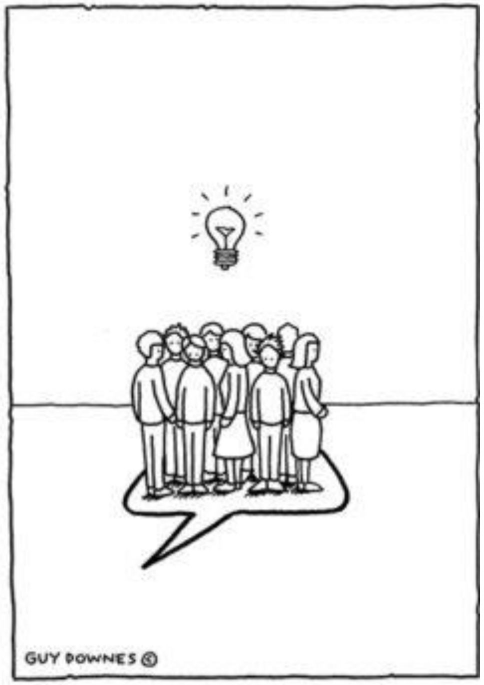
CENTER FOR HUMAN IDENTIFICATION

INTERNATIONAL NARCOTICS & LAW ENFORCEMENT
JUSTICE & SECURITY
UNITED STATES DEPARTMENT OF STATE

GORDON THOMAS HONEYWELL
GOVERNMENTAL AFFAIRS

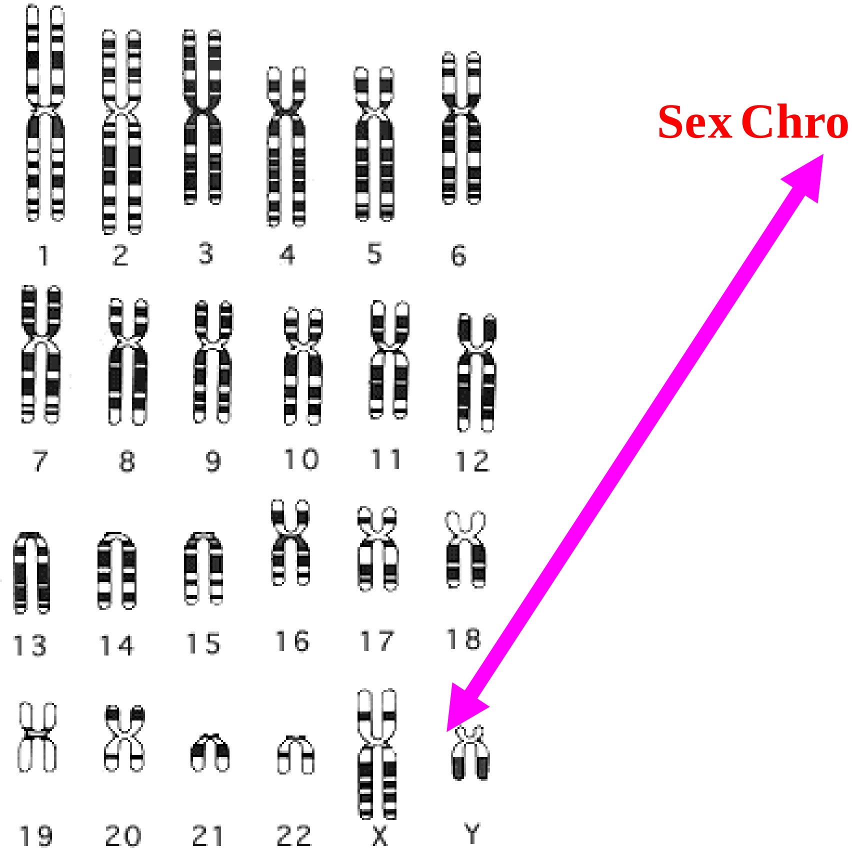
Objectives

- Create legal framework for use of forensic data to combat human trafficking
- Enhance Technical Capacity
- Create a Database
- Public Outreach
- Increased Coordination



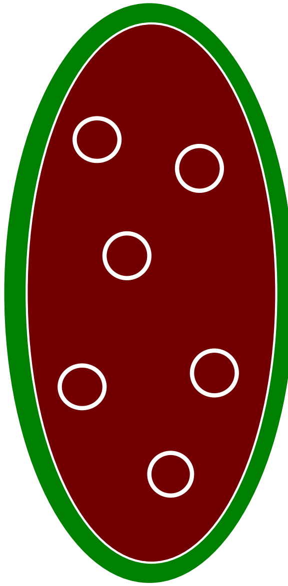
Kinship Analyses
Exploit the Genome

Cell Nucleus – 3.2 billion bp



Sex Chromosomes (XX or XY)

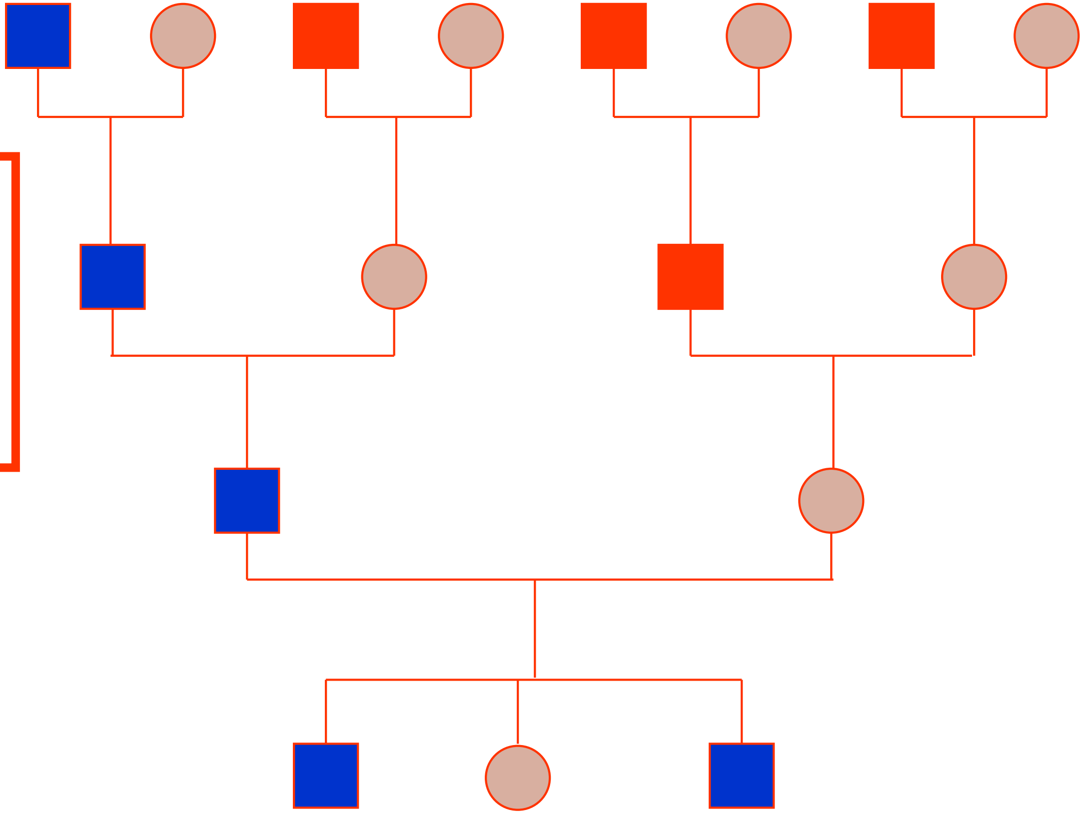
mitochondria
100s of mtDNA
copies per cell



This image cannot curr

Lineage Markers

Y Chromosome



Y chromosome passed along paternal lineage

100% Similar

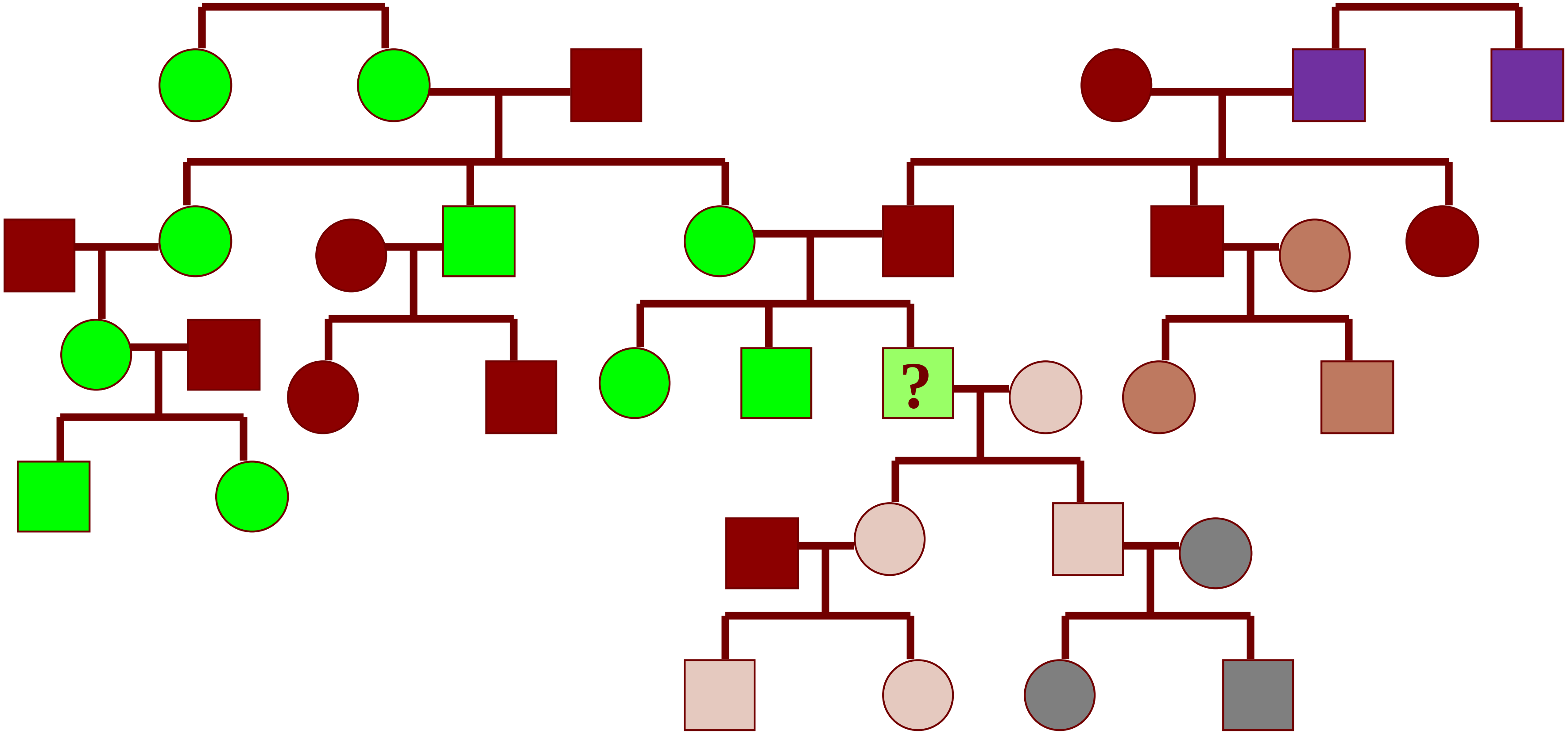
This image cannot curre...

Autosomal DNA
1/8 from Great-grandparents

Maternal Lineage

Mitochondrial DNA

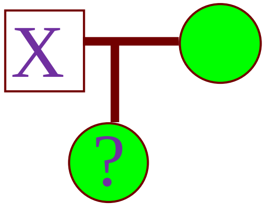
Transferred from mother to all children

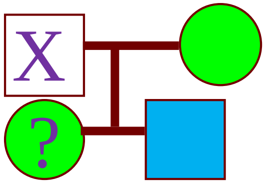


Real Case

- Skull
  - Only a mtDNA profile was generated
  - CODIS search has a hit with a pedigree containing a single individual, a sister of a missing person.
  - LR =163.
- A few years later lower half of a skeleton) was found in the same county
  - Same mtDNA profile
  - Full STR profile
  - LR = 0.031.
- Combined LR ~5.

Kinship Testing

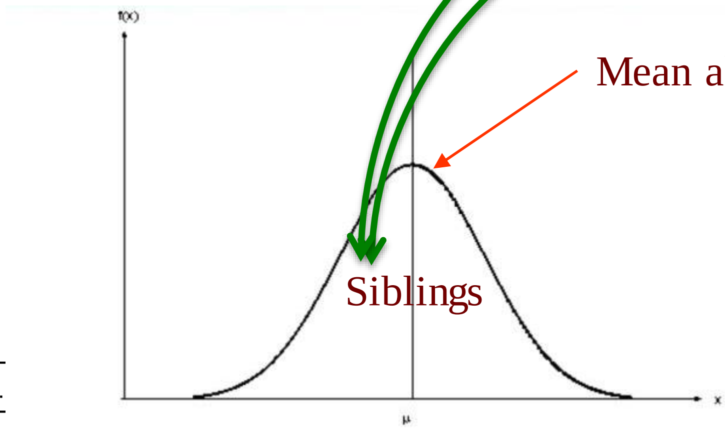




LR=1000

LR= 250

Is there support for the relationship?



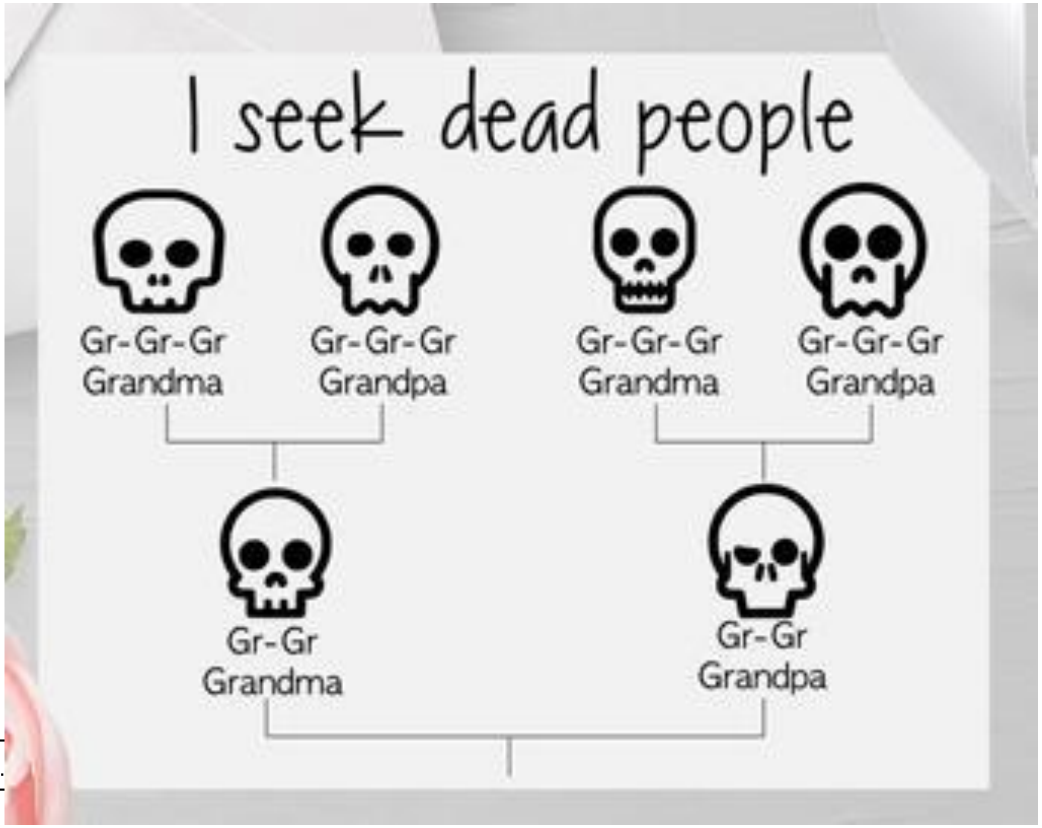
allele sharing distribution

Improve Kinship Resolution

- More relatives – limited
- More DNA – sample dependent
- More markers – alternate technology
  - STRs are not enough

Improve Kinship Resolution

- Tools of Genetic Genealogy
- Dense SNP data (chip technology or MPS/NGS)



This image cannot curre..

Golden State Killer

- One of the most notorious serial killers in California
- Years active 1974 – 1986

This image cannot curre...

The Washington Post

Public Safety

To find alleged Golden State Killer, investigators first found his great-great-great-grandparents

- Sacramento detective used DNA recovered from a crime scene to find the killer’s great-great-great grandparents, who lived in the early 1800s.

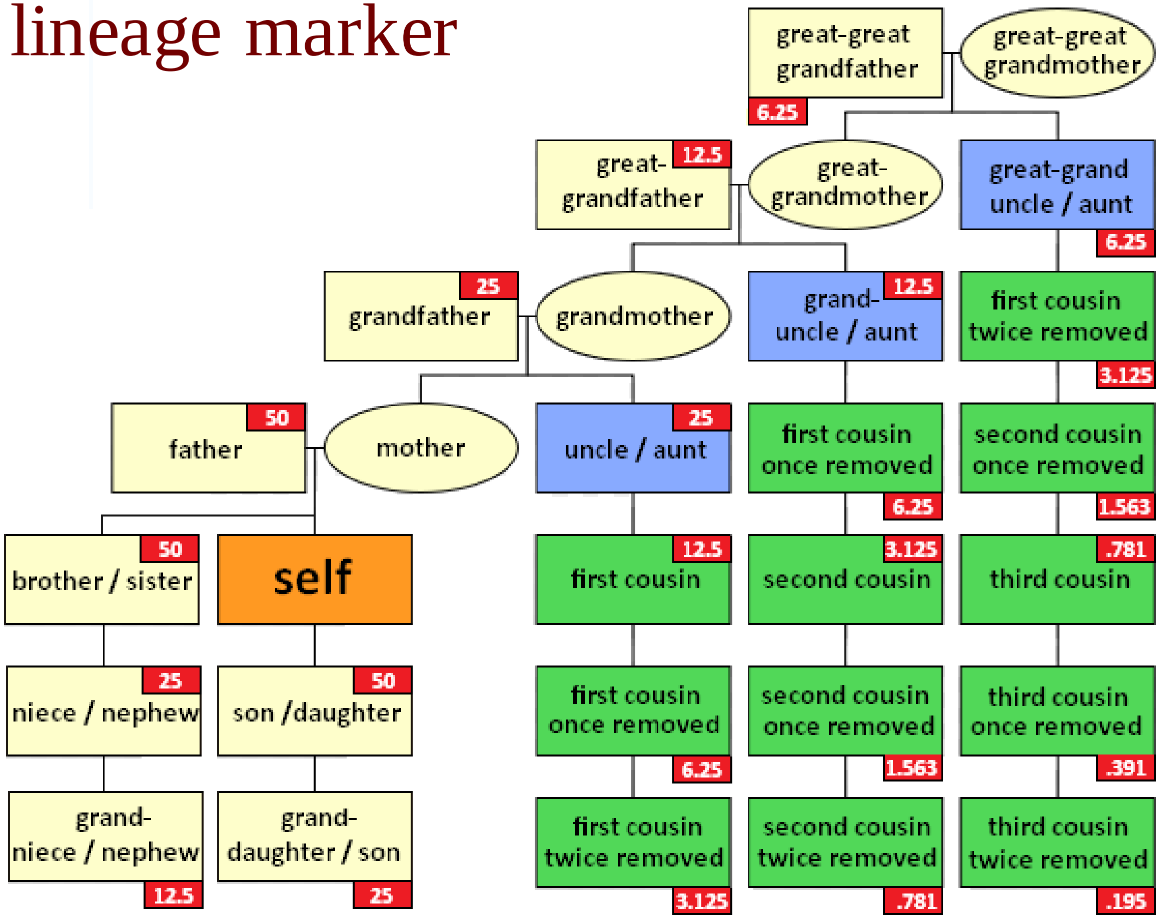
- Branch by painstaking branch, he and a team created about 25 family trees containing thousands of relatives down to the present day.

- Sharing of the DNA in this case was to the level of 3rd cousins (10-20 total)

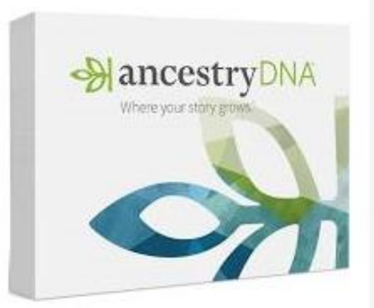
Genetic Genealogy

- %DNA shared
- The shared DNA is a lineage marker

ISOGG wiki statistics:
Chromosome sharing ---
Parent/child: 3539-3748 cMs
1st cousins: 548-1034 cMs
1st cousins 1R: 248-638 cMs
2nd cousins: 101-378 cMs
2nd cousins 2R: 43-191 cMs
3rd cousins: 43-ca 150 cMs
3rd cousins 1R: 11.5-99 cMs
More distant cousins: 5-ca 50 cMs



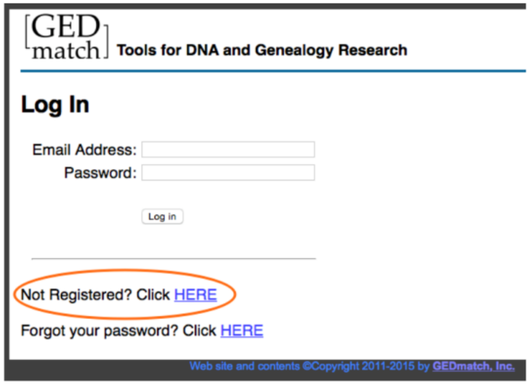
This image cannot curre...



23andMe

Family Tree DNA

Ancestry.com



[GED match] Tools for DNA and Genealogy Research

Log In

Email Address:
Password:

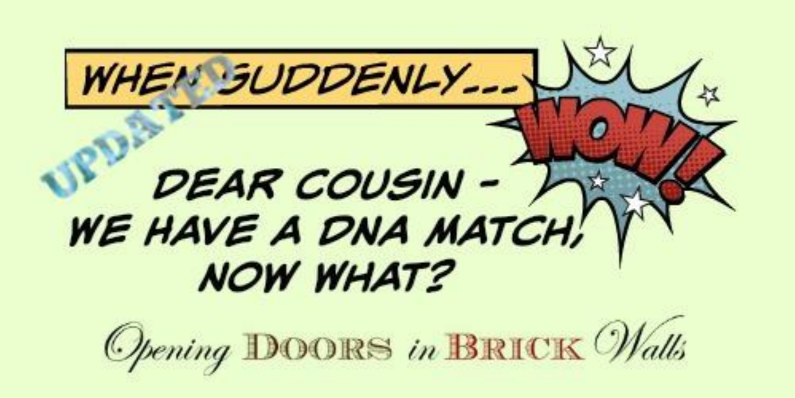
Not Registered? Click HERE

Forgot your password? Click HERE

Web site and contents ©Copyright 2011-2015 by GEDmatch, Inc.

929,000 genetic profiles
(May 2018)

GED match --- open-source personal genomics database and genealogy website



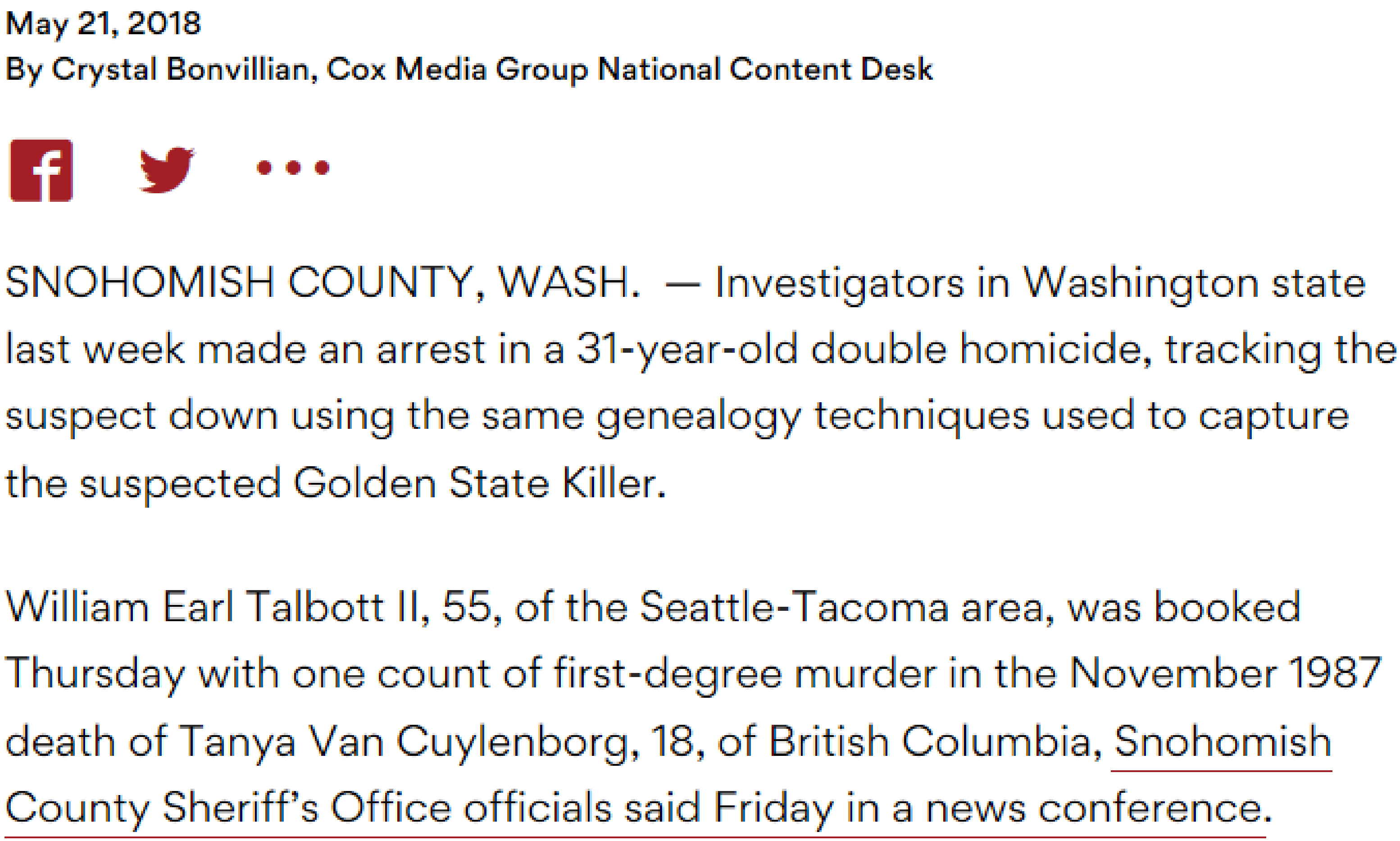
This image cannot curre...

Several Cases Solved
Murder of Marcia King

- Buckskin Girl identified

This image cannot curre…

More Cold Cases Solved…



May 21, 2018

By Crystal Bonvillian, Cox Media Group National Content Desk

SNOHOMISH COUNTY, WASH. — Investigators in Washington state last week made an arrest in a 31-year-old double homicide, tracking the suspect down using the same genealogy techniques used to capture the suspected Golden State Killer.

William Earl Talbott II, 55, of the Seattle-Tacoma area, was booked Thursday with one count of first-degree murder in the November 1987 death of Tanya Van Cuylenborg, 18, of British Columbia, Snohomish County Sheriff's Office officials said Friday in a news conference.

Limitations

- Quantity of DNA is about 20-30 times higher than that of traditional DNA testing.
- Requires “enrollment” of family members in the GEDmatch database.
- Biased to European population.
- Genetic data about individual may impact privacy
    - However, chromosome sharing does not disclose genetic information
- Lack of defined protocols

Benefits

- Can extend the genetic search beyond the traditional paternity trio/first order relatives.
- Reduce need for lineage-based reference samples
- Generate leads for difficult, old, and cold cases
- Chance of error is small
  - Verify person of interest with traditional forensic marker typing.
- UNTCHI is developing novel algorithms to work with low quality samples

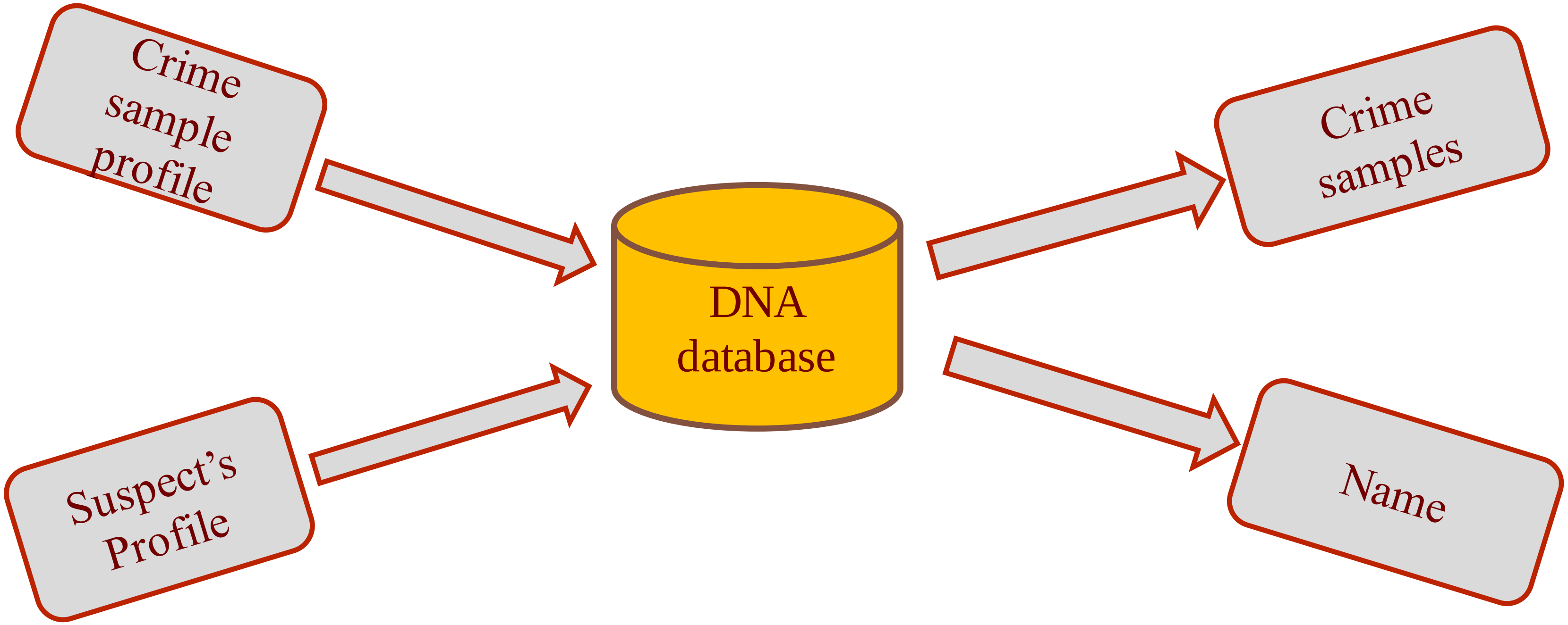
Potential

- Using dense SNP data can improve kinship analysis
- Using dense SNP data may improve missing persons identifications

This image cannot curre…

Timely Response

- DNA database of known offenders and forensic samples
- Timeliness of the match can be critical and can reduce the resources required to solve the crime



Rapid DNA Analysis

- A fully automated (hands free) process of developing a DNA profile from a reference buccal (cheek) swab <u>without human intervention</u>
- CODIS Core Loci STR profile generated
- Uploaded to National DNA Index System (NDIS)
- “Swab in- profile out” process

Modified Rapid DNA Analysis

- An automated (hands free) process of developing a CODIS Core Loci STR profile from a known reference sample *requiring some human interpretation and technical review*

Five/Four Lab Processes in One "Field" Device

- Disposable microfluidic biochip technology integrates and automates five laboratory processes into one field device operated by **non-expert users**

DNA extraction/ purification
3 hours

DNA Quantification
2 hours

DNA STR Marker Copying
3 hours

Separation/ Detection
2 hours

Human Analyst Profile Match
Hours - Weeks

90 Min

Note: These time frames presented by RAPID DNA Companies. Times will vary depending on processes.

Rapid DNA

- Main concepts/features
  - Short turn around timeframe
  - Ease of use
  - Robust
    - Dependent on application
- Speed and convenience

Potential Rapid DNA Applications

Traditional DNA Forensic Lab

In the forensic lab, personnel can process urgent samples more efficiently and enable those working in the lab who have limited forensic DNA experience

Law Enforcement

In the police station, law enforcement personnel can generate investigative leads and identify suspects faster

Military and Border Control

In the field, personnel can conduct rapid identification of unaccompanied minors, limit human trafficking, and reunited familes by conducting familial relationship verification

Disaster Victim Identification

In the field, personnel can conduct rapid identification of disaster victims, reunite families, or support other field deployment needs

Rapid DNA for Law Enforcement

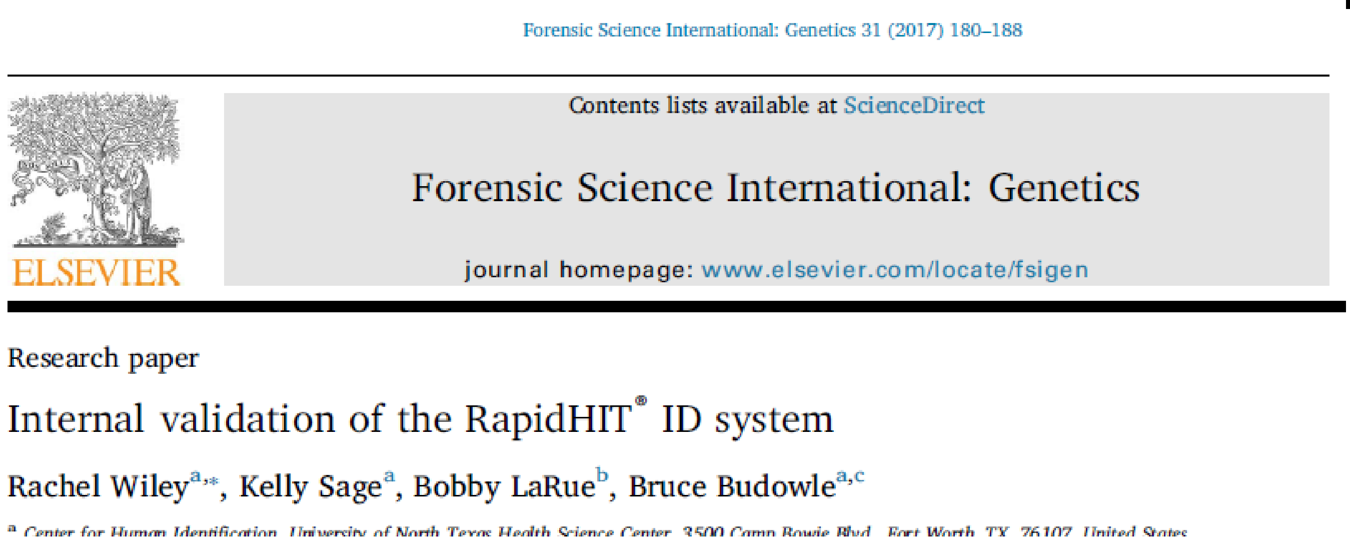
- Moves DNA testing closer to the time the crime was committed
  - Backlogs at forensic labs can result in delayed processing of reference samples for searching against the crime scene evidence
- Reduces cost and time of an investigation
  - Generates investigative leads more quickly
  - Eliminates persons of interest early on
  - Conserves investigators' time and focuses their efforts
- Complementary to standard forensic lab work
  - Allows labs to focus on more complex casework samples

RapidHIT instruments

- Applied Biosystems RapidHIT™ ID System
  - Generates CODIS-eligible profiles with confirmed sample types
  - Uses NDIS-approved chemistry with the Applied Biosystems GlobalFiler™ Express PCR Amplification Kit
  - Offers compatibility with established DNA databases
  - 1 sample in 90 minutes

- RapidHIT™ 200 System
  - Bead-based cell lysis and DNA purification,
  - Re-useable CE capillaries
  - 5 samples + 3 controls, or
  - 8 samples in < 2 hours

Publications



Forensic Science International: Genetics 31 (2017) 180–188

Contents lists available at ScienceDirect

Forensic Science International: Genetics

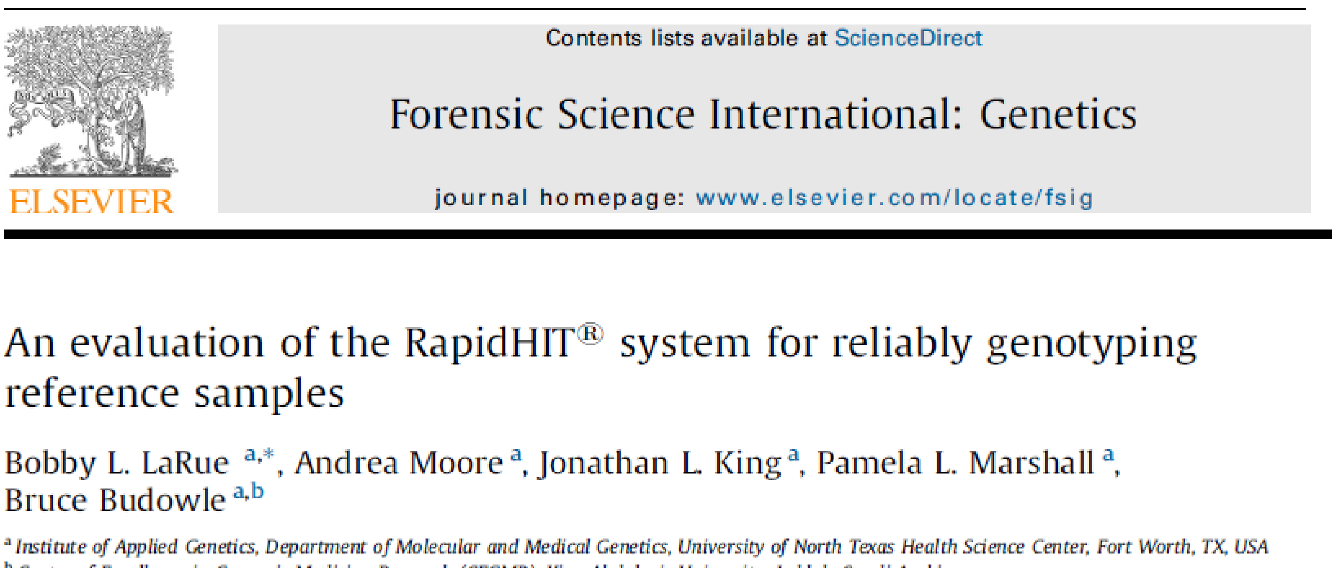
journal homepage: www.elsevier.com/locate/fsigen

Research paper

Internal validation of the RapidHIT® ID system

Rachel Wiley[a,*], Kelly Sage[a], Bobby LaRue[b], Bruce Budowle[a,c]

[a] Center for Human Identification, University of North Texas Health Science Center, 3500 Camp Bowie Blvd., Fort Worth, TX, 76107, United States



Contents lists available at ScienceDirect

Forensic Science International: Genetics

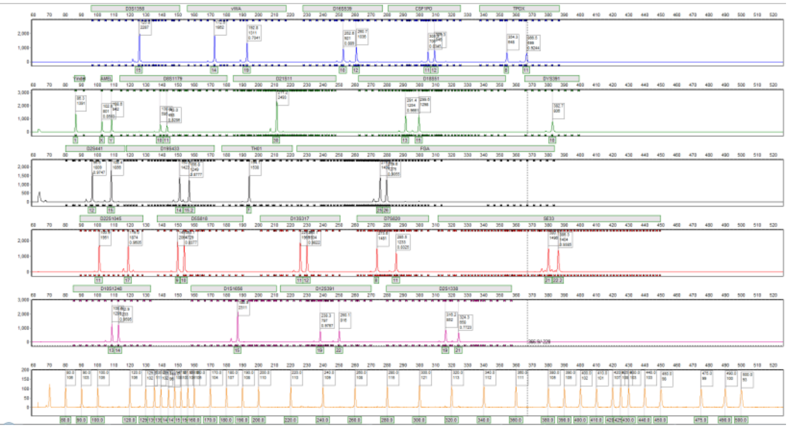
journal homepage: www.elsevier.com/locate/fsig

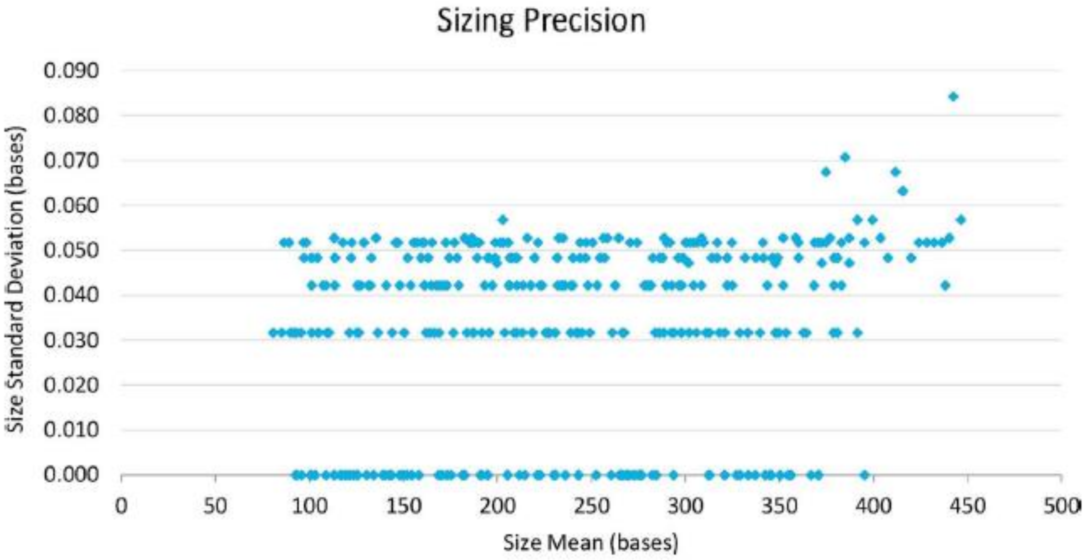
An evaluation of the RapidHIT® system for reliably genotyping reference samples

Bobby L. LaRue [a,*], Andrea Moore [a], Jonathan L. King [a], Pamela L. Marshall [a], Bruce Budowle [a,b]

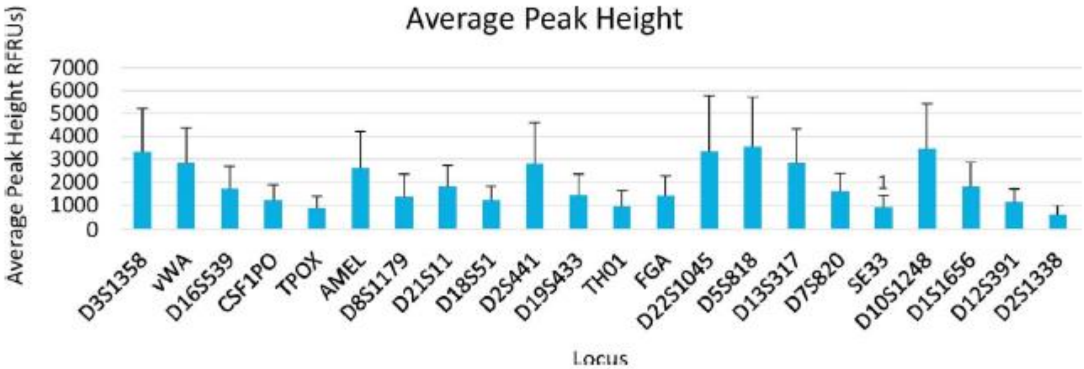
[a] Institute of Applied Genetics, Department of Molecular and Medical Genetics, University of North Texas Health Science Center, Fort Worth, TX, USA

Performance Criteria





Sizing Precision



Average Peak Height

Summary of Rapid DNA Features

- Simple operation
- Use by non-scientists
- But interpretation may still require some expertise
- Cartridge-based systems
- Preloaded reagents
- Sample to profile capability on bench top systems
- Portable (actually transportable)
- Low throughput
- Buccal swabs and high template DNA evidence samples

Massively Parallel Sequencing

Precision ID System for human identification

Sequencing and Forensic Applications

- Human ID
  - mtDNA
  - STRs
  - SNPs
  - Direct
  - Mixtures
  - Kinship
  - Ancestry
  - Phenotype

- Pharmacogenetics
  - Molecular autopsy

- Microbial Forensics

MPS Technologies

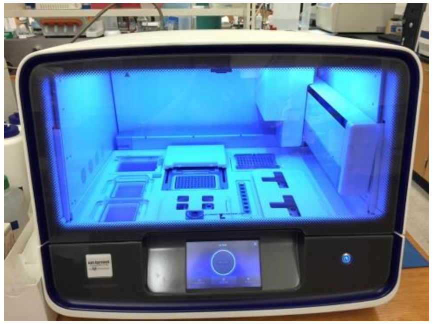
- Increasingly user-friendly
- Highly automated workflow
- Compatible with numerous genetic marker types
- Available data analysis options
- Accurate, reliable data

Ion Chef™ System

- Simplifies workflow
- Helps reduce hands-on time
- Reduces user-introduced variability
- Enables more consistent results
  - Similar impact that capillary electrophoresis had with STR typing

General MPS Workflow

Automated



DNA → Library Preparation → Template Preparation → Sequencing → Analysis

Library Preparation ↓

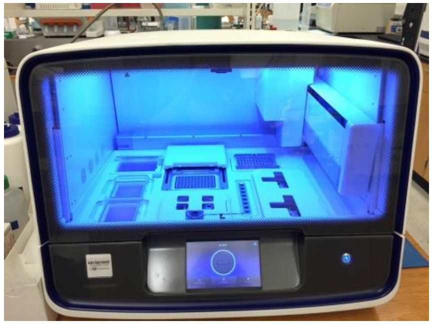
Amplify Targets → Digest Primers → Ligate Adapters → Purify Libraries → Equalize and Pool Libraries

Two pipetting steps to prepare library

One pipetting step to load chip

General MPS Workflow

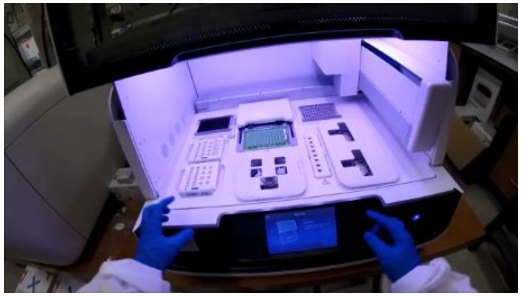
Automated



DNA → **Library Preparation** → **Template Preparation** → **Sequencing** → **Analysis**

Template Preparation ↓

Library Amplification → **ISP Recovery and Enrichment** → **Chip Loading**

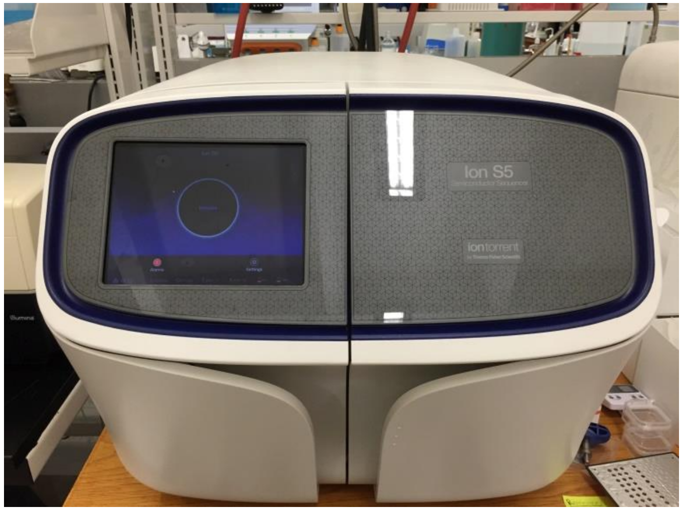


One instrument/one pipetting step!

General MPS Workflow

DNA → **Library Preparation** → **Template Preparation** → **Sequencing** → **Analysis**

Sequencing ↓ **Load Sequencer**

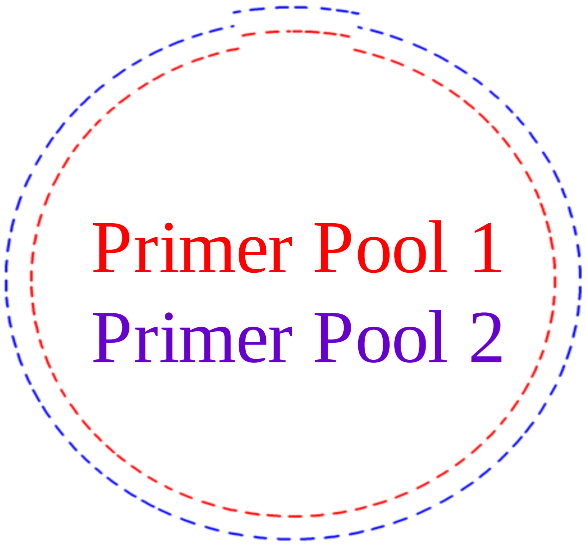


Analysis of Difficult Samples

mtDNA is the most successful marker

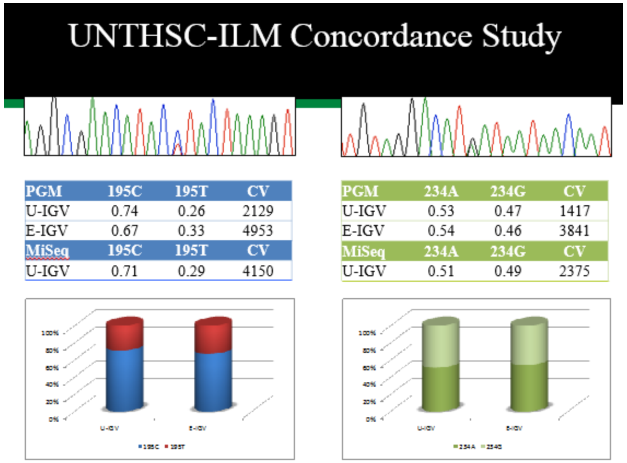
Mitochondrial Genome Panel

- Multiplex short amplicon system
    - Applied Biosystems™ Precision ID mtDNA Whole Genome Panel
- Spans entire mitochondrial genome
    - Two multiplex panels
    - Each panel contains 81 primer pairs (plus degenerate primers)
    - Tiled, overlapping pattern
    - Amplicons are ≤ 175 bps in length



Validation Studies

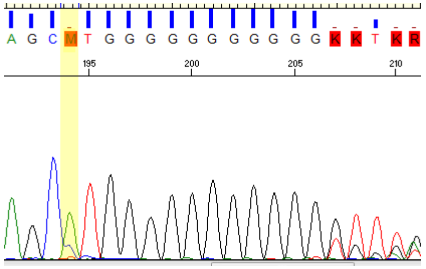
- Population studies
- Concordance
  - Manual and Ion Chef System
  - Orthogonal testing (Illumina and Sanger)
  - Between laboratories
- Reproducibility/Repeatability
  - Accuracy/Precision
  - Amplicon drop out
- Mito Panel v Long PCR
- Sensitivity/Stochastic effects
- Heteroplasmy
- etc

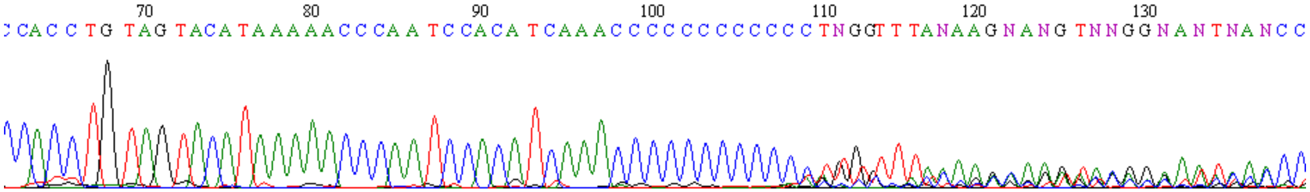


UNTHSC-ILM Concordance Study

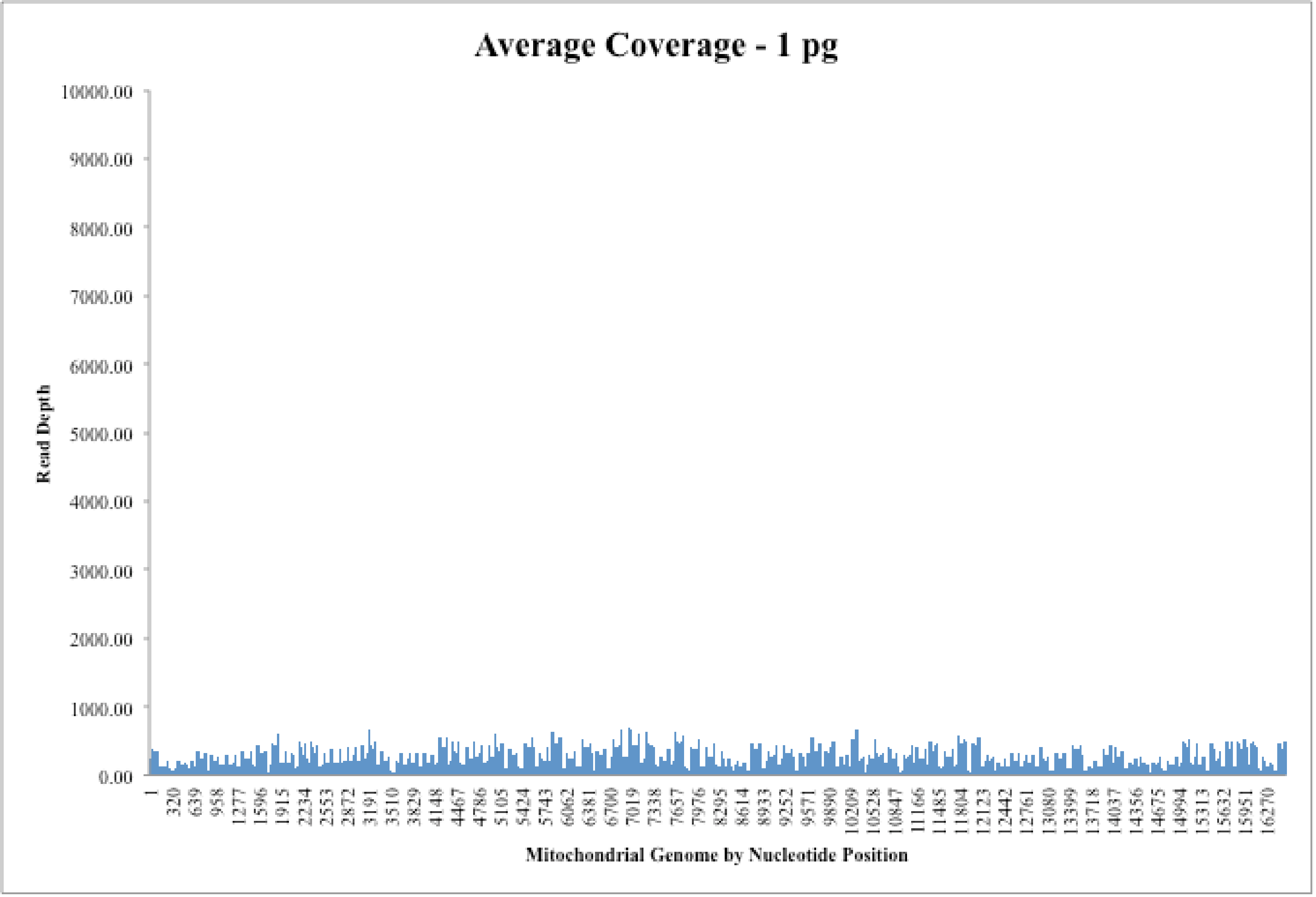
PGM	195C	195T	CV
U-IGV	0.74	0.26	2129
E-IGV	0.67	0.33	4953
MiSeq	**195C**	**195T**	**CV**
U-IGV	0.71	0.29	4150

PGM	234A	234G	CV
U-IGV	0.53	0.47	1417
E-IGV	0.54	0.46	3841
MiSeq	**234A**	**234G**	**CV**
U-IGV	0.51	0.49	2375



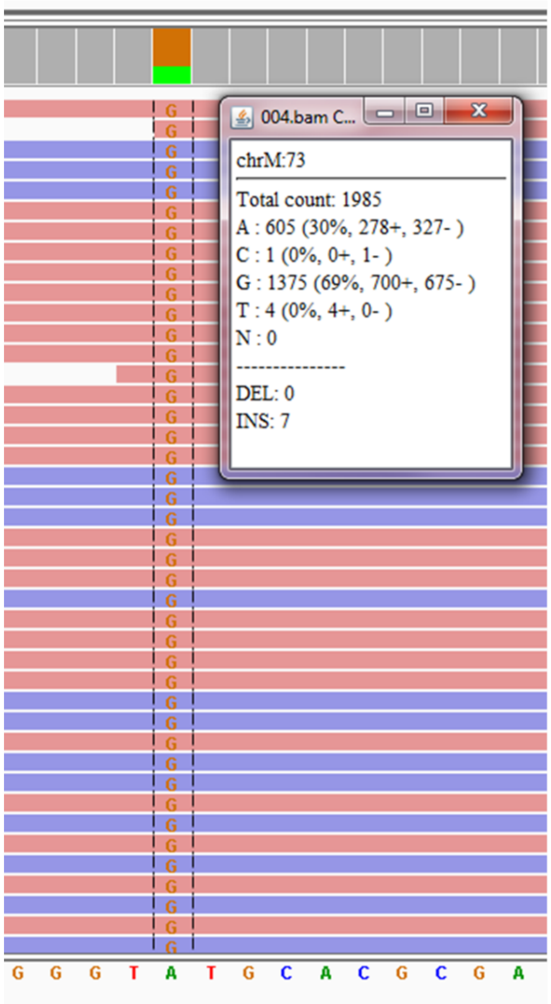


Sensitivity Study



Ranged from 32X to 683X

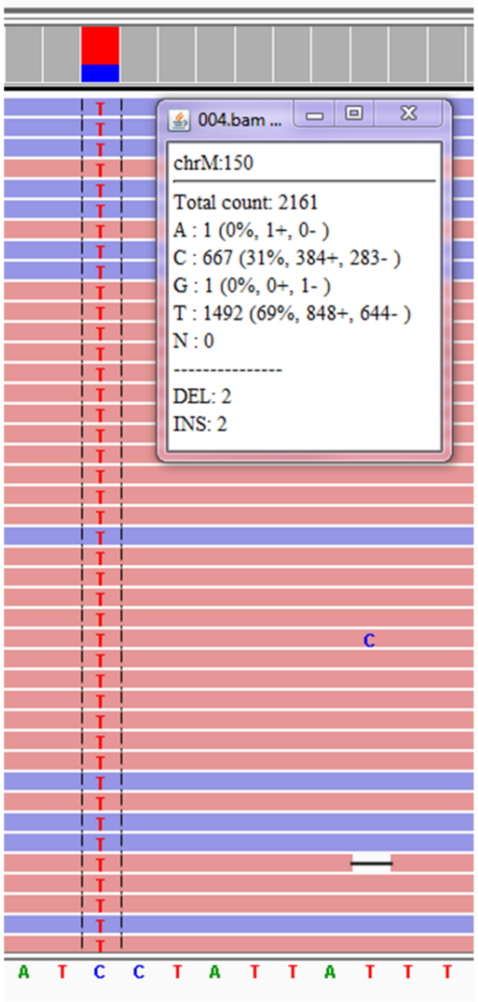
Mixtures



chrM:73

Total count: 1985
A : 605 (30%, 278+, 327-)
C : 1 (0%, 0+, 1-)
G : 1375 (69%, 700+, 675-)
T : 4 (0%, 4+, 0-)
N : 0

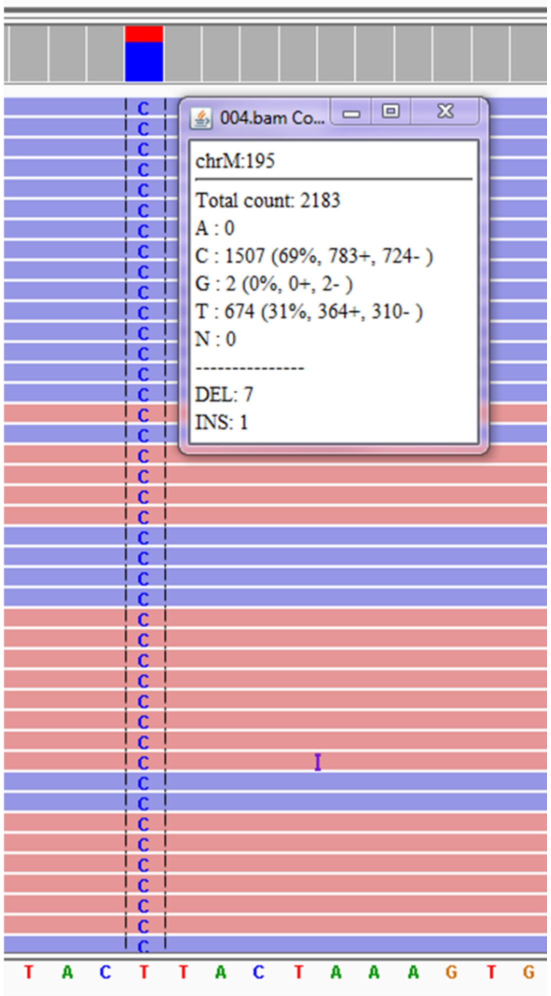
DEL: 0
INS: 7



chrM:150

Total count: 2161
A : 1 (0%, 1+, 0-)
C : 667 (31%, 384+, 283-)
G : 1 (0%, 0+, 1-)
T : 1492 (69%, 848+, 644-)
N : 0

DEL: 2
INS: 2

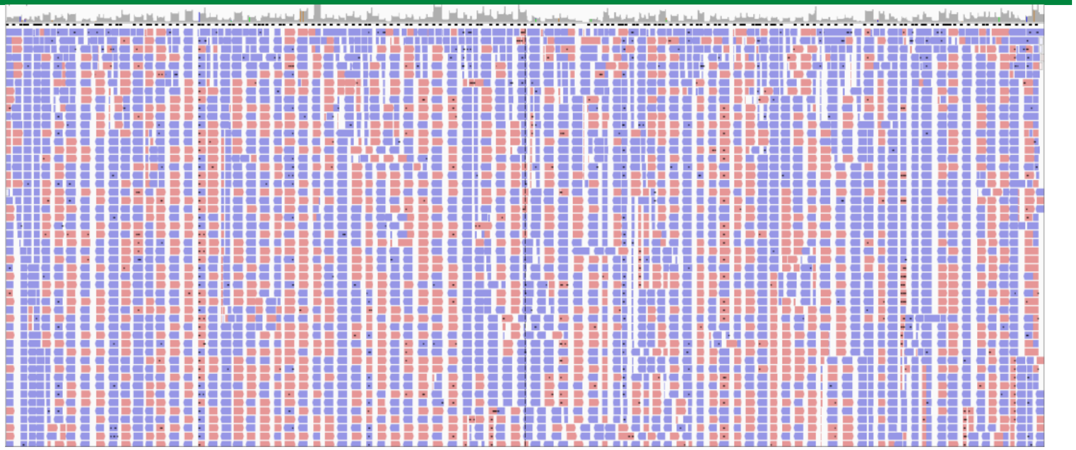


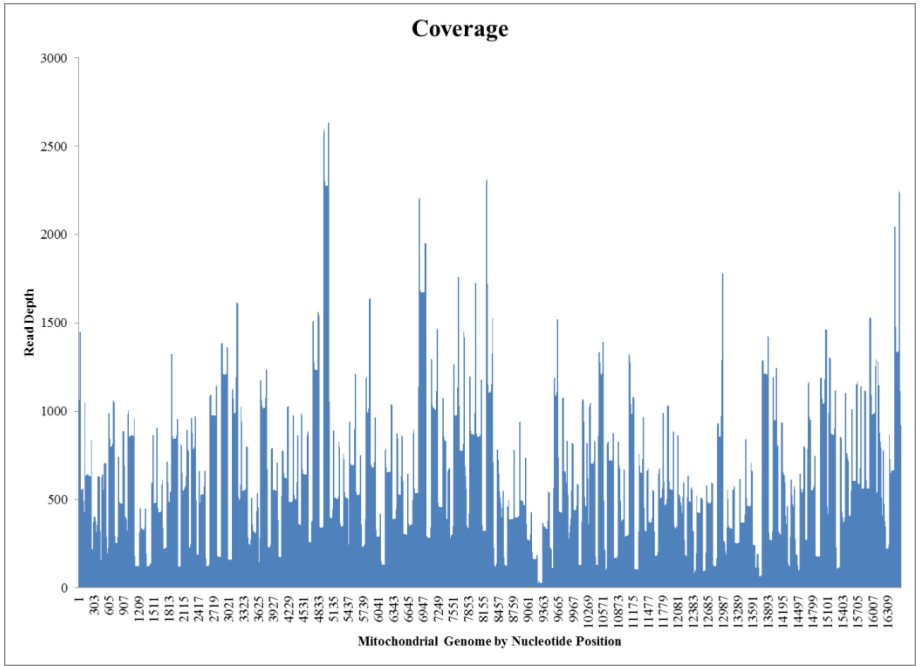
chrM:195

Total count: 2183
A : 0
C : 1507 (69%, 783+, 724-)
G : 2 (0%, 0+, 2-)
T : 674 (31%, 364+, 310-)
N : 0

DEL: 7
INS: 1

Hair Shaft

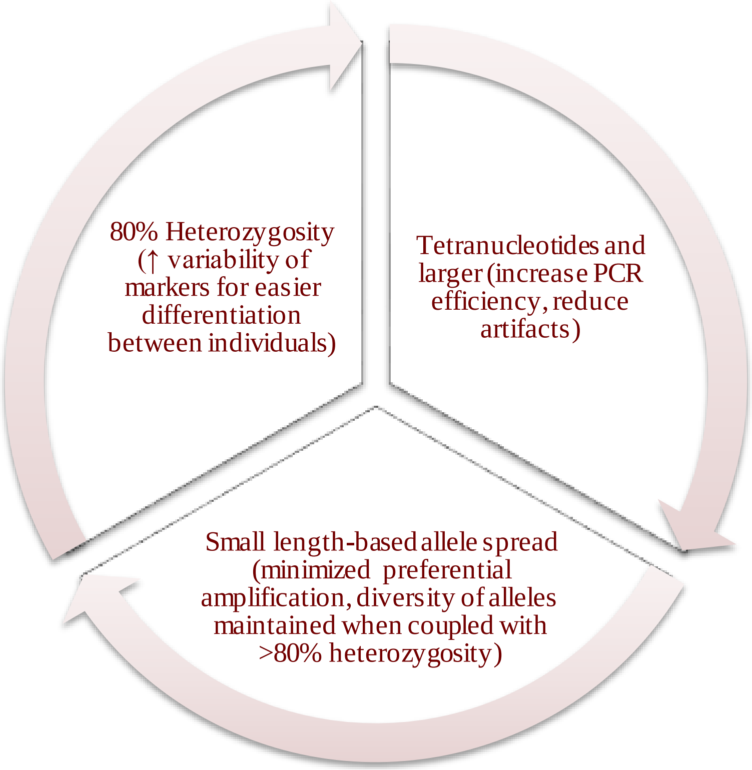




Coverage ranged from 23X to 2,634X across the mitochondrial genome

Search for More Heterozygous STRs to enhance mixture analysis

- 1000 Genomes Project (raw sequences, unsorted)
- STR Catalog Viewer
  - Summary of human STR variation compiled using lobSTR software



80% Heterozygosity (↑ variability of markers for easier differentiation between individuals)

Tetranucleotides and larger (increase PCR efficiency, reduce artifacts)

Small length-based allele spread (minimized preferential amplification, diversity of alleles maintained when coupled with >80% heterozygosity)

Forensic Science International: Genetics 37 (2018) 162–171

Contents lists available at ScienceDirect

Forensic Science International: Genetics

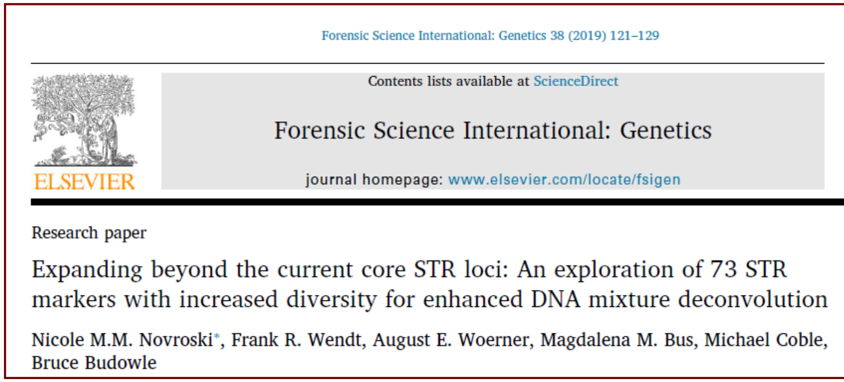
journal homepage: www.elsevier.com/locate/fsigen

ELSEVIER

Research paper

Potential highly polymorphic short tandem repeat markers for enhanced forensic identity testing

Nicole M.M. Novroski[a,b,*], August E. Woerner[a], Bruce Budowle[a]



Forensic Science International: Genetics 38 (2019) 121–129

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsigen

ELSEVIER

Research paper

Expanding beyond the current core STR loci: An exploration of 73 STR markers with increased diversity for enhanced DNA mixture deconvolution

Nicole M.M. Novroski[*], Frank R. Wendt, August E. Woerner, Magdalena M. Bus, Michael Coble, Bruce Budowle

Exploratory Multiplex

- 73 highly heterozygous loci using MPS chemistry
  - 15 loci previously described by Phillips (2016) and others

- 451 unrelated individuals from three U.S. populations
  - Caucasian (CAU; n=155);
  - Hispanic (HIS; n=148);
  - African American (AFA; n=148)

- Each STR locus was characterized and reviewed manually for diversity using in-house Excel workbooks
  - Alleles characterized by length and sequence
  - Population genetics analyses (heterozygosity; Hardy-Weinberg equilibrium (HWE); linkage disequilibrium (LD); random match probabilities (RMP))

In silico Mixtures

- A **subset of 20 loci** selected for comparison to the CODIS core loci
    - The current requirement is **20 CODIS core loci** for upload into the national DNA database
    - high heterozygosity (>90%)
    - Operationally problematic loci (even if heterozygosity >90%) were excluded

- 443 U.S. population samples
    - African American, (AFA; n=140, 8 incomplete profile samples removed)
    - Caucasian, (CAU; n=155)
    - Hispanic, (HIS; n=148)

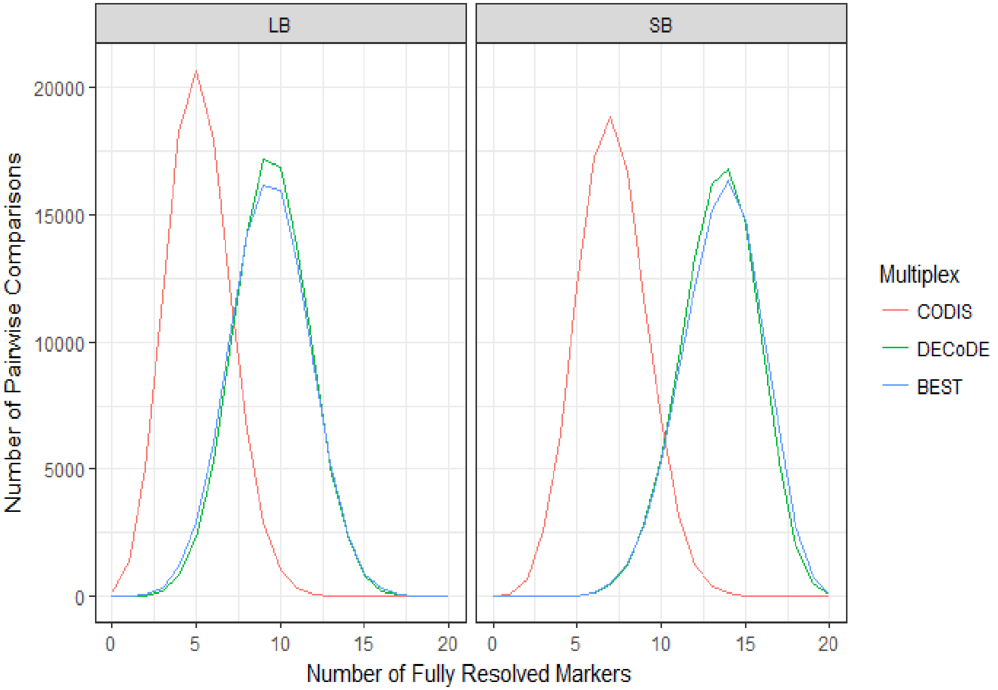
CODIS Panel Loci	Exploratory Panel Loci	BEST Panel Loci
D2S1338	D3S2406	D3S2406
D12S391	D2S1360	D2S1360
D1S1656	D7S3048	D7S3048
D21S11	D8S1132	D8S1132
D8S1179	D11S2368	D11S2368
vWA	D15S822	D15S822
D3S1358	D2N2	D2N2
D18S51	D1N10	D1N10
FGA	D12N15	D12N15
D19S433	D1N16	D1N19
D13S317	D1N19	D1N21
D5S818	D1N21	D8N23
D16S539	D8N23	D15N26
D22S1045	D15N26	D14N56
D7S820	D14N56	D3N61
D2S441	D3N61	D12S1338
CSF1PO	D4N70	D4N70
D10S1248	D11N52	D2S1338
TPOX	D17N32	D1S1656
TH01	D2N43	D11N52

Shaded cells reflect the CODIS core loci.

In silico Mixtures
Top 20 Performance

N = 443 individuals representing three U.S. populations (n=140 African American; n=155 Caucasian; and n=148 Hispanic samples)

Full Resolution	CODIS (LB)	Multiplex (LB)	BEST (LB)	CODIS (SB)	Multiplex (SB)	BEST (SB)
0 Loci	155	0	0	3	0	0
1 Locus	1330	6	7	104	0	0
2 Loci	5001	27	76	663	0	0
3 Loci	11715	211	359	2541	0	1
4 Loci	18200	839	1181	6333	8	4
5 Loci	**20574**	2320	2885	12056	14	41
6 Loci	17881	5242	5971	17179	128	139
7 Loci	12141	9592	10134	**18791**	440	496
8 Loci	6514	14203	14249	16609	1204	1275
9 Loci	2889	**17073**	**16079**	11656	2801	2682
10 Loci	1063	16806	15882	6966	5379	5230
11 Loci	317	13764	13121	3169	9287	8794
12 Loci	98	9305	8937	1253	13194	12117
13 Loci	18	5005	5235	431	16137	15125
14 Loci	5	2398	2475	119	**16747**	**16270**
15 Loci	2	823	919	22	14580	14779
16 Loci	0	226	322	6	10125	10755
17 Loci	0	53	62	1	5266	6560
18 Loci	0	10	8	1	2005	2765
19 Loci	0	0	1	0	539	777
20 Loci (Full Profile)	0	0	0	0	49	93
Total	97,903	97,903	97,903	97,903	97,903	97,903

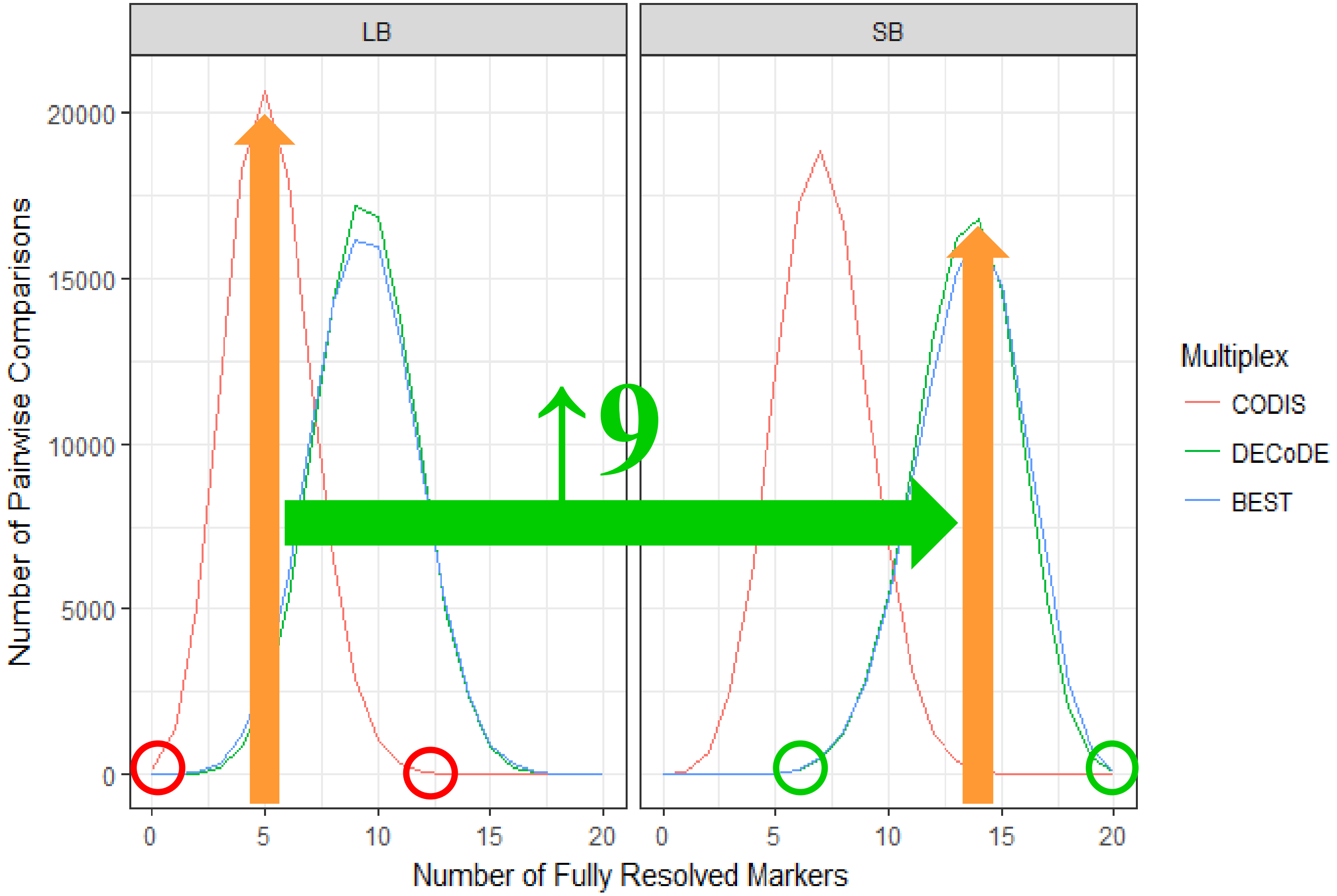


Highlighted in **BOLD** is the mode - the number of loci with the greatest number of counts for which four alleles at each locus were resolved.

LB = length-based; SB = sequence-based

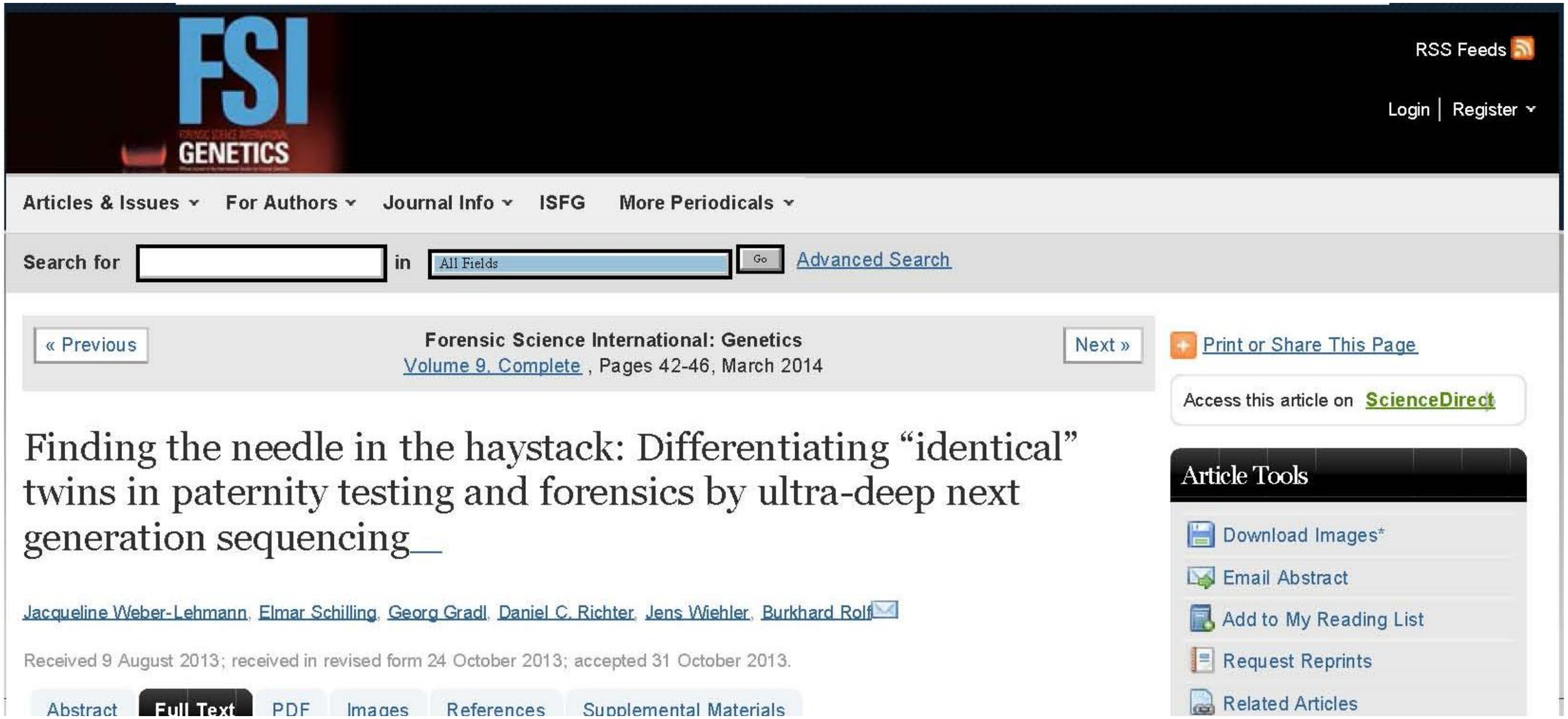
In silico Mixtures
Top 20 Performance

N = 443 individuals representing three U.S. populations (n=140 African American; n=155 Caucasian; and n=148 Hispanic samples)



LB = length-based; SB = sequence-based

Identical Twins



Forensic Science International: Genetics
Volume 9, Complete , Pages 42-46, March 2014

Access this article on ScienceDirect

Finding the needle in the haystack: Differentiating "identical" twins in paternity testing and forensics by ultra-deep next generation sequencing

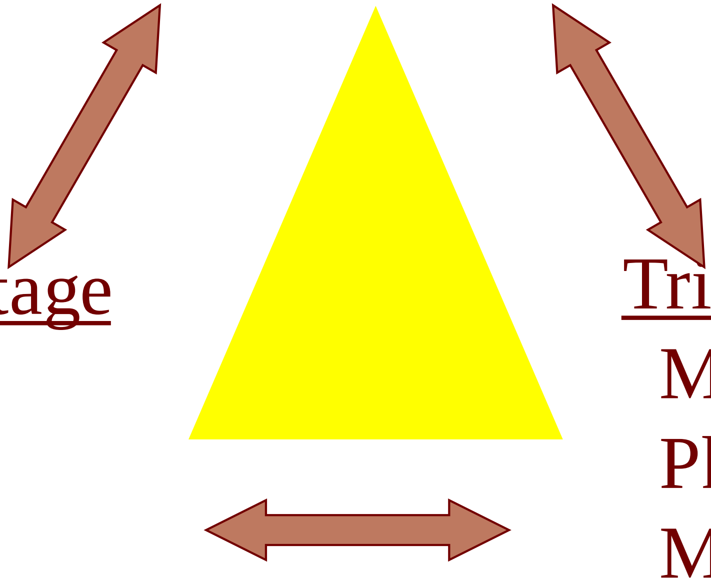
Jacqueline Weber-Lehmann, Elmar Schilling, Georg Gradl, Daniel C. Richter, Jens Wiehler, Burkhard Rolf

Received 9 August 2013; received in revised form 24 October 2013; accepted 31 October 2013.

Genetics of Death

Molecular Autopsy

Predisposition to Sudden Unexpected Death



Genetic variants

Developmental stage

- Neonates
- Infants
- Juveniles
- Adults
- Elderly

Triggering event

- Medication
- Physical stress
- Mental stress
- Climate change
- Nutritional change

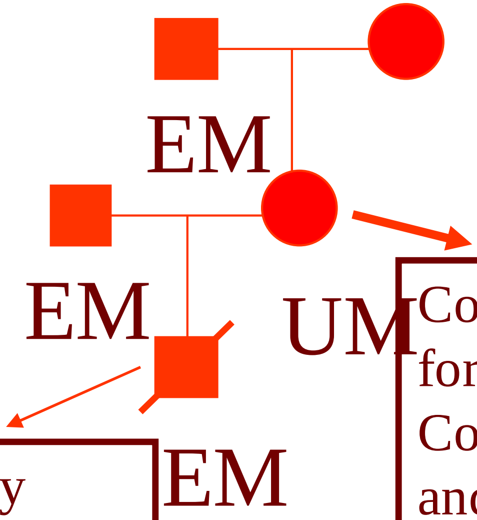
Forensic Example

- Codeine
- Infant died of morphine overdose at 13 days old
  - Mother was prescribed Tylenol #3 (acetaminophen and codeine)
  - Codeine is metabolized into morphine
  - Mother was an ultra rapid metabolizer

http://www.ilike.com/user/codeine0vr

Morphine Poisoning in a Breastfed Neonate

- Genetics
- Mother: CYP2D6*2A / *2x2 (UM)
- Grandfather, father, son: CYP2D6*1/*2 (EM)



Consistent with increased formation of morphine from Codeine and with somnolence and constipation

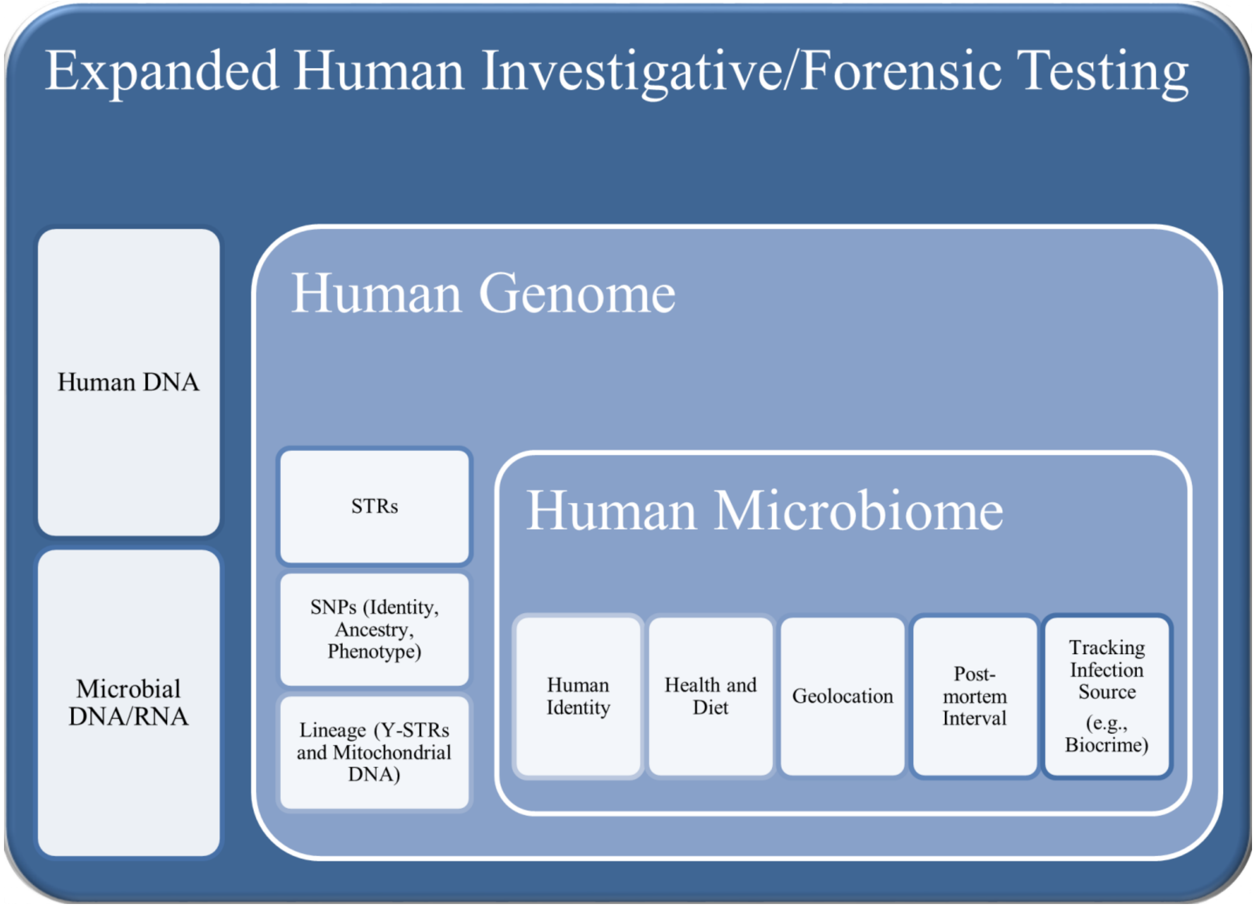
Clinical picture and laboratory tests are consistant with death due to morphine intoxication

poor (PM), intermediate (IM), extensive (EM), or ultrarapid (UM) metabolizer

Koren et al. The Lancet, 2006

Expansion of Microbial Forensics

- Broader Definition
- Today's capabilities enable greater versatility



Expanded Human Investigative/Forensic Testing

Human DNA

Microbial DNA/RNA

Human Genome

STRs

SNPs (Identity, Ancestry, Phenotype)

Lineage (Y-STRs and Mitochondrial DNA)

Human Microbiome

Human Identity

Health and Diet

Geolocation

Post-mortem Interval

Tracking Infection Source

(e.g., Biocrime)

Classification Accuracy

Universal and Nonuniversal Markers

Forensic Science International: Genetics 32 (2018) 50–61

Contents lists available at ScienceDirect

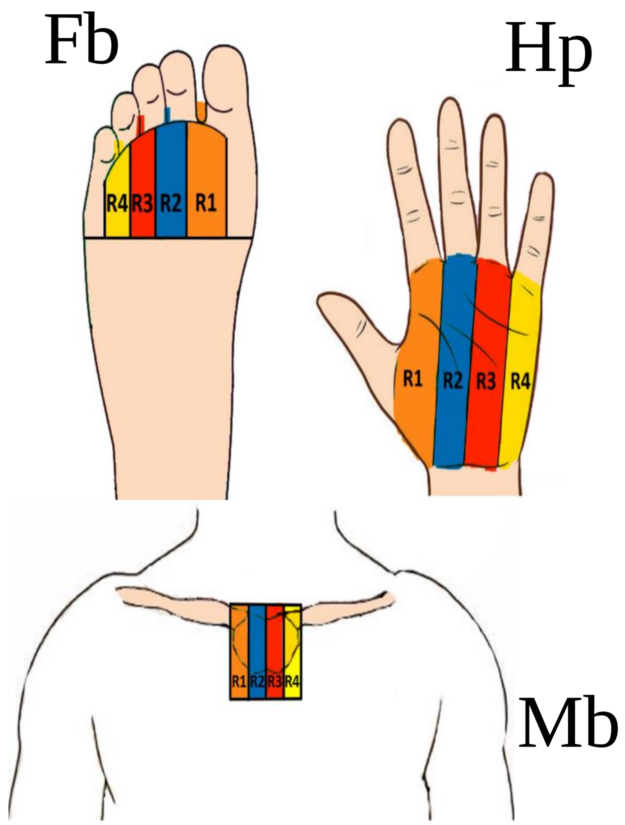
Forensic Science International: Genetics

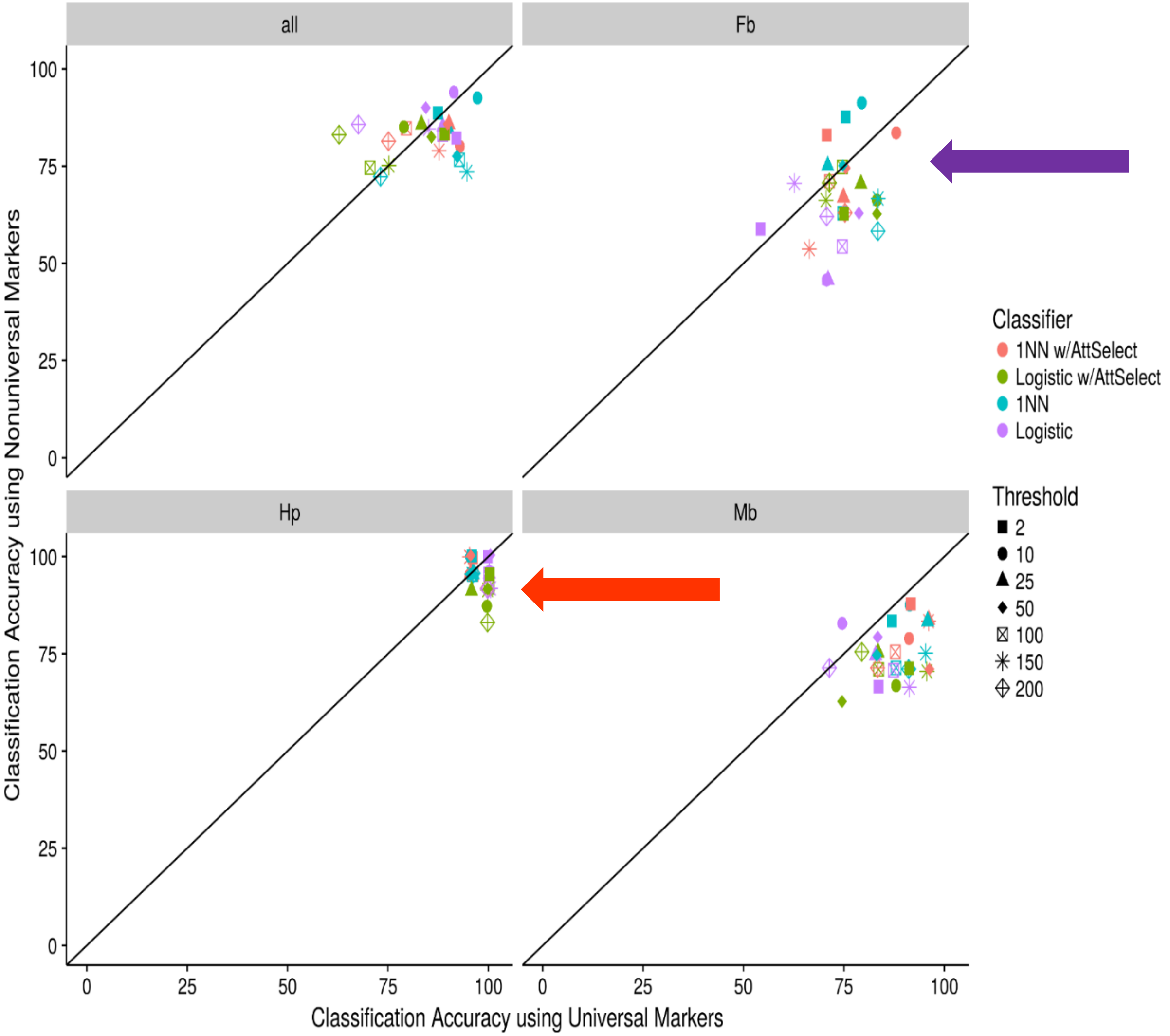
journal homepage: www.elsevier.com/locate/fsigen

Research paper

Targeted sequencing of clade-specific markers from skin microbiomes for forensic human identification

Sarah E. Schmedes[a,b], August E. Woerner[b], Nicole M.M. Novroski[a,b], Frank R. Wendt[a,b], Jonathan L. King[b], Kathryn M. Stephens[c], Bruce Budowle[a,b,d,*]





Guiding Principles of the Forensic DNA Analysis Workflow

Extraction

Automation

Quantitation

Amplification

Detection and Data Analysis

Validation and Quality Assurance

ACKNOWLEDGMENTS

CENTER FOR HUMAN IDENTIFICATION

- Thermo Fisher Scientific
- UNTCHI Research Team
- US State Department (Award No. No. S-INLEC-GR-1013)
- Gordon Thomas Honeywell Governmental Affairs
- Walther Parson and his team
- Sara Katsanis