

IdPrism: Rapid Analysis of Forensic DNA Samples Using MPS SNP Profiles

Darrell O. Ricke, PhD,

James Watkins, Philip Fremont-Smith, & Adam Michaleas

2019 IEEE HPEC

Sept. 25, 2019





Disclaimer

DISTRIBUTION STATEMENT A. Approved for public release. Distribution is unlimited.

This material is based upon work supported under Air Force Contract No. FA8702-15-D-0001 by the Defense Threat Reduction Agency (DTRA). Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the U.S. Air Force or Defense Threat Reduction Agency.

© 2019 Massachusetts Institute of Technology.

Delivered to the U.S. Government with Unlimited Rights, as defined in DFARS Part 252.227-7013 or 7014 (Feb 2014). Notwithstanding any copyright notice, U.S. Government rights in this work are defined by DFARS 252.227-7013 or DFARS 252.227-7014 as detailed above. Use of this work other than as specifically authorized by the U.S. Government may violate any copyrights that exist in this work.

The authors have no competing financial interest to disclose.



Overview

Introduction to DNA Forensics

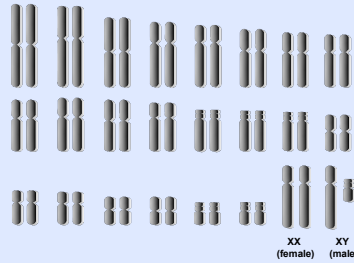
Computational Bottlenecks

- **SNP allele calling**
- **Identification Searching & Mixture Analysis**
- **Statistics – Probability of Random Man Not Excluded**



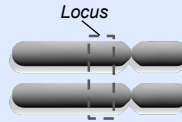
Core DNA Forensics Concepts

Cells contain 23 pairs of chromosomes



Locus:

Specific position on a chromosome



Nucleotide (or base):

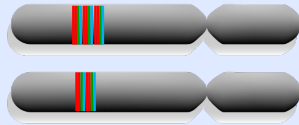
One of 4 residues that make up the DNA polymer

A, C, G or T



Core DNA Forensics Concepts

Allele: One of many DNA sequences that may occupy a locus



Short Tandem Repeat (STR):

Variable number of repeat units

Allele 1 ... CCTA AATGAATGAATGAATGAATGAATG CCGAG ...

Allele 2 ... CCTA AATGAATGAATGAATG CCGAG ...

AATG repeating sequence



Single Nucleotide Polymorphism (SNP):

Difference in one nucleotide

Allele 3 ...CGAGCCTAACGAGCCTA...

Allele 4 ...CGAGCCTAGCGAGCCTA...

A/G Single nucleotide difference

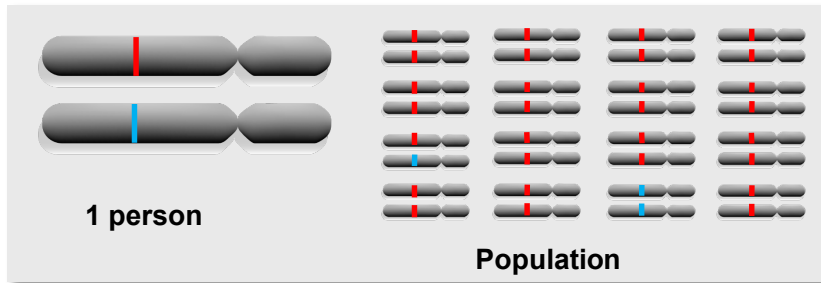


SNP Sequencing To Meet Requirements

Requirements	STR Sizing	SNP Sequencing
Human ID	Green	Green
Multi-contributor samples	Yellow	Green
Extract unknown profiles	Red	Green
Extended kinship	Yellow	Green
Touch samples	Yellow	Green
Biogeographic ancestry	Red	Green
Appearance	Red	Yellow



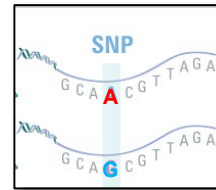
Converting Biological Signatures to Digital 'Barcodes'



Determine Presence of Low Frequency SNPs Across Genome

95%

5%



Visualize as Individual SNP 'barcode'



Presence of minor allele indicated by line (1 bit) at SNP Position

Selection of rare SNPs creates unique minor allele signatures/barcodes for individuals & enables effective differentiation of multiple barcodes in a mixture



Overview

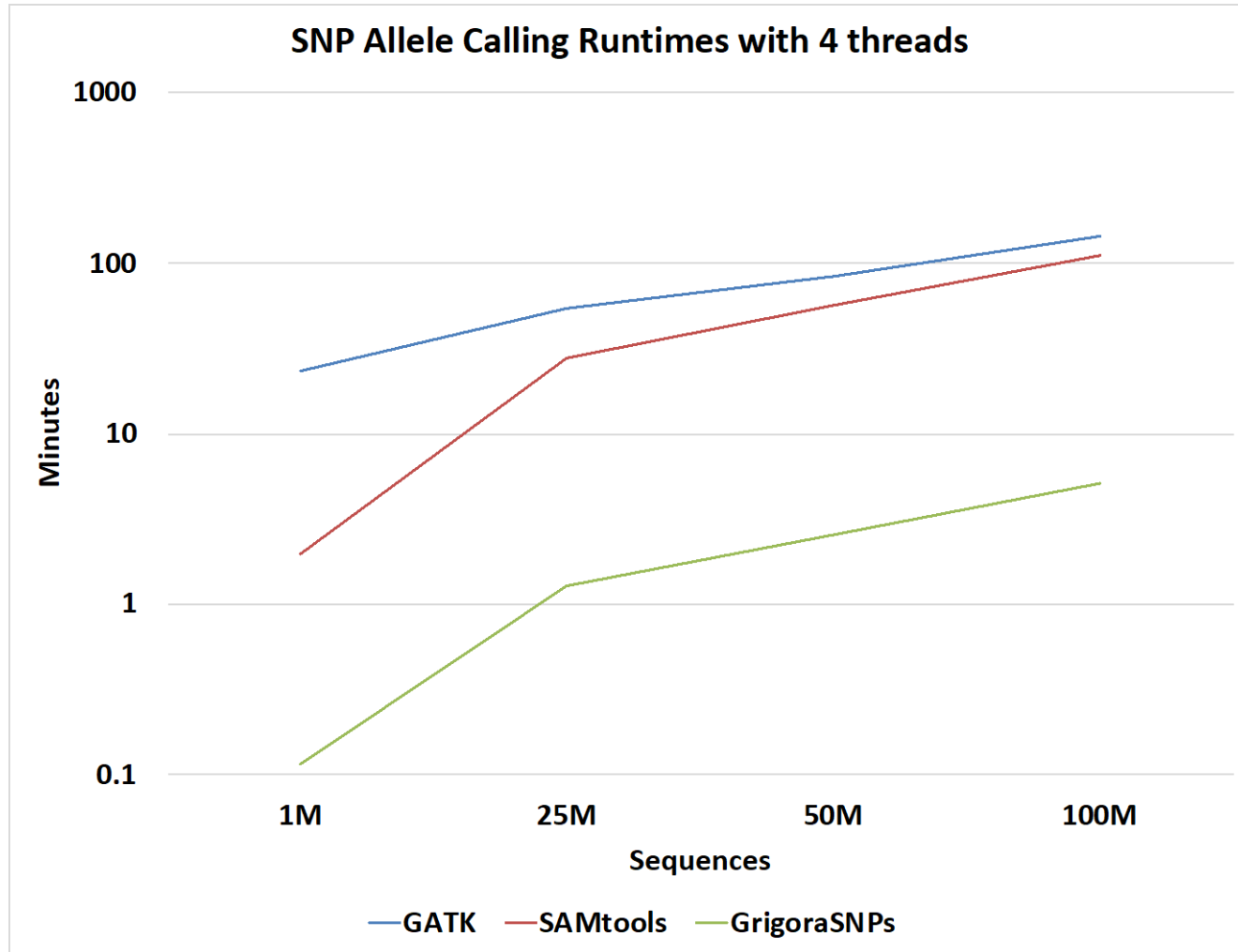
Introduction to DNA Forensics

Computational Bottlenecks

- **SNP allele calling**
- **Identification Searching & Mixture Analysis**
- **Statistics – Probability of Random Man Not Excluded**



SNP Allele Calling Runtime Comparisons





Statistical Power of Large SNP Panel Sequencing

P(False Match) is function of total # of mixture major allele SNPs (N)

Let p be minor allele frequency

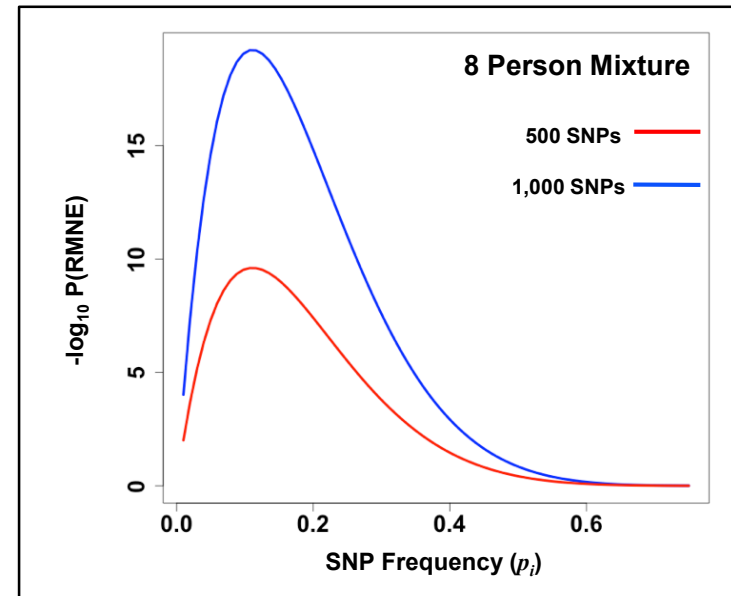
Let q be the major allele frequency ($q = 1 - p$)

Let L be the number of allele mismatches

- enables tolerance for incomplete profiles

$$P_{RMNE}(0 \text{ to } L) = \sum_0^L \underbrace{q^{2N}}_{\substack{\text{Probability of } n \\ \text{contributors} \\ \text{NOT having a minor allele} \\ \text{at SNP } i}} * \underbrace{\frac{N!}{L!(N-L)!}}_{\text{Combination}(N,L)} * \underbrace{\left(\frac{1-q^2}{q^2}\right)^L}_{\substack{\text{Probability of a} \\ \text{minor allele at SNP } i}}$$

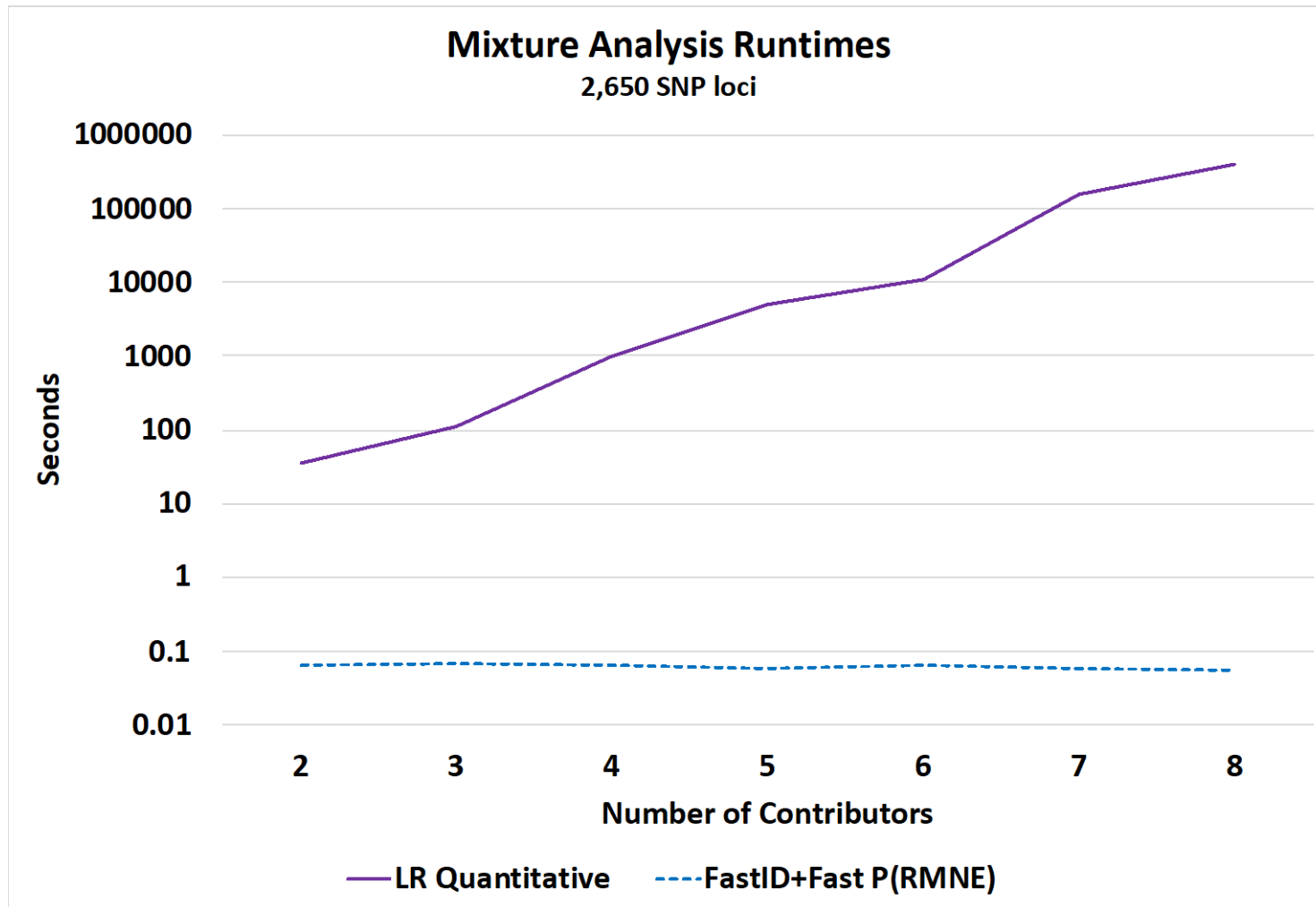
$$P_{RMNE}(L+1) = P_{RMNE}(L) * \frac{(n-L)}{L+1} * K$$



Increased statistical power with more SNPs



Mixture Analysis Runtime Comparisons

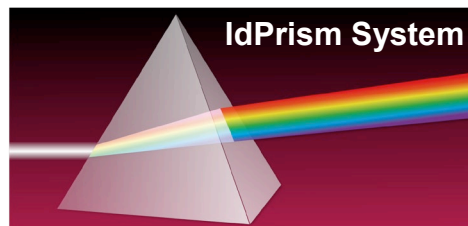
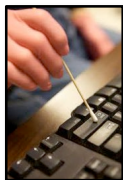


FastID algorithm: $\text{mismatches} = \text{popcount}((\text{reference XOR mixture}) \text{ AND reference})$



UNCLASSIFIED

IdPrism: Advanced DNA Forensics Platform



Best in class for

- Performance
- Speed
- Scalability



Lincoln Bioinformatics

Evidence

Sample Preparation

- Extract DNA
- Screen & normalize samples
- Extract/copy markers

Sequencing Platform

- Proton
- Illumina
- Other...

Profile Generation

SNP Allele Caller

5 minutes

Analysis Modules

- | | |
|------------|------------------|
| 1:1 ID | Mixture Analysis |
| Kinship | Trace |
| Saturation | |

5 minute

User Interface

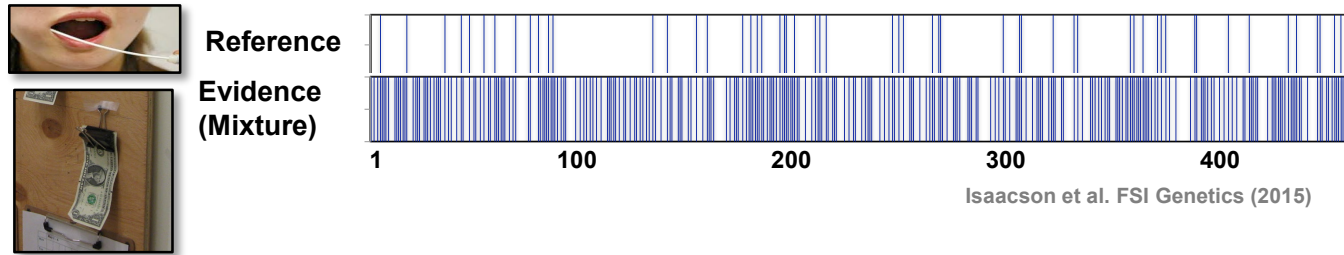
- | | |
|-----------------|------------------------|
| Browse & Search | Annotate |
| Allele Plots | Mixtures Investigation |
| | Network |

The IdPrism Platform architecture for DNA analysis addresses current capability gaps within an extensible and scalable framework

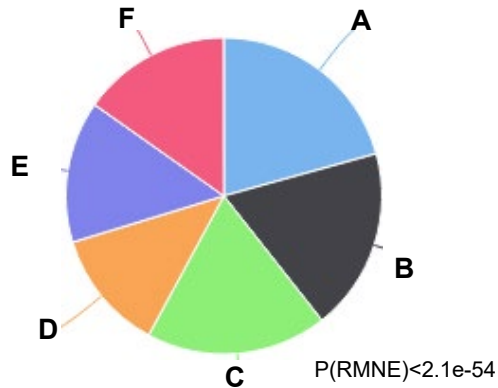


Phase I Results: Finding Known References in DNA Mixture

Mixture Analysis Approach



Lab Equimolar 6-Person Mixture



Individual	P(RMNE)
A	2.8e-54
B	2.1e-54
C	2.1e-54
D	2.1e-54
E	2.1e-54
F	2.7e-54

Demonstrates MIT LL SNP approach can identify 6+ contributors in complex mixtures



Finding Known References in DNA Mixture

Mixture Analysis Approach

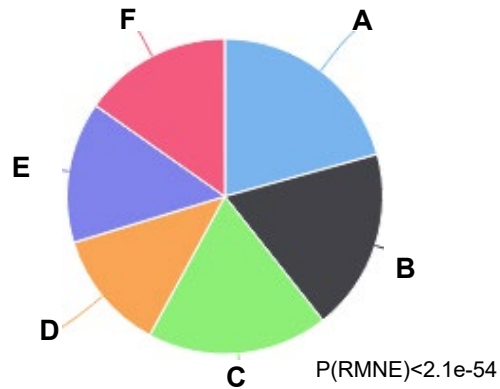


Reference

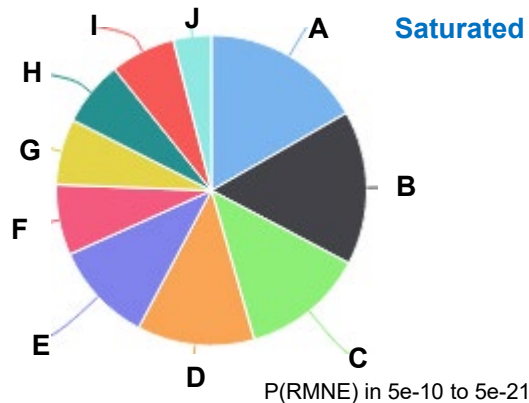
Evidence
(Mixture)

Isaacson et al. FSI Genetics (2015)

Lab 6-Person Mixture
6/6 detected

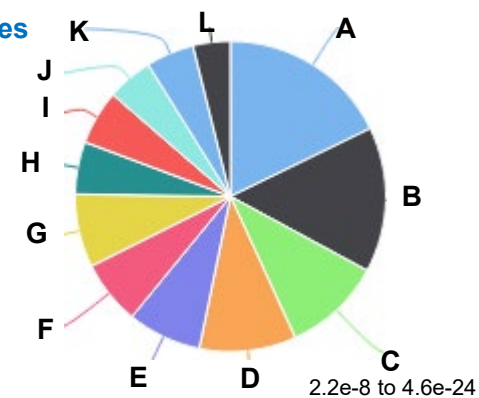


Lab 10-Person Mixture
10/10 detected



Saturated Mixtures

Lab 15-Person Mixture
(12/15 detected)



Demonstrates MIT LL SNP approach can identify 10+ contributors in complex mixtures



References

- Voskoboinik & Darvasi “Forensic Identification of an Individual in Complex DNA Mixtures” *Forensic Sci. Int. Genet.* 5:428-435 (2011)
- Isaacson *et al.* “Robust detection of individual forensics profiles in DNA mixtures. *Forensic Sci. Int. Genet.* 14:31-7 (2015)
- Ricke *et al.* “GrigoraSNPs: Optimized HTS DNA forensic SNP Analysis” *Journal of Forensic Sciences* 63:1841-1845 (2018)
- Ricke “FastID: Extremely Fast Forensic DNA Comparisons” *IEEE HPEC* (2017)
- Ricke & Schwartz “Fast P(RMNE): Fast Forensic DNA Probability of Random Man Not Excluded Calculation” *F1000Research* (2017)
- Ricke *et al.* “Estimating Individual Contributions to Complex DNA SNP Mixtures. *Journal of Forensic Sciences* (2019)
- Patent application: DNA Mixtures from One or More Sources and Methods of Building Individual Profiles Therefrom
 - US 62/534,590 PCT/US2018/041081, W0 2019/010410