# *Microhaplotypes Are Coming of Age*

Kenneth K. Kidd, Ph.D.
Professor Emeritus and Senior Research Scientist
and the Kidd Lab in the Department of Genetics,
Yale University
on behalf of many collaborators at
the Podini Lab at The George Washington University
and the Human Identification Group at Thermo Fisher

**Green Mountain DNA Meeting**
**Burlington, Vermont**
**July 23, 2017**

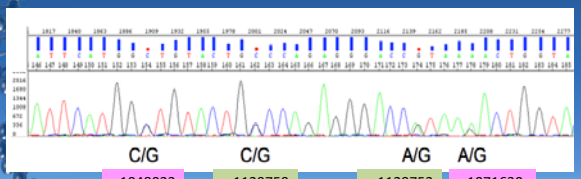
## Forensic Uses of DNA

- Individualization—random match probabilities
- Ancestry inference—biogeographic origins
- Familial relationships—types of biologic relationships
- Phenotype inference
- Mixture deconvolution—identifying individual components of a mixture

**Microhaplotypes can be good for all five**

## What is a microhaplotype?

Microhaplotypes (MHs) are loci of two or more SNPs within a short distance from each other (<300 nucleotides i.e. 'micro') with three or more allelic combinations ('haplotypes'). The short distance allows a single MPS read to cover the entire distance resolving phase, i.e., separating the maternal and paternal chromosomes.




**A small coding segment of HLA-DQA1**

## What is a microhaplotype?

A microhaplotype locus is a SNP-based multiallelic locus:

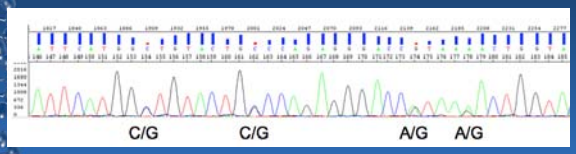
Small enough amplicon for a single read

No stutter

All alleles at a locus are the same size

Low mutation rate

## Sanger Sequencing Involves Both Chromosomes at a Microhaplotype Locus



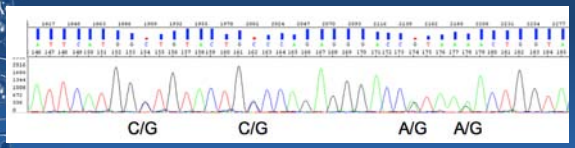

There are 16 different haplotypes (alleles) that can exist for these four SNP sites. A total of (N*(N+1))/2 = 136 different genotypes can exist. Eight of those different genotypes are heterozygous at all four sites. Which is the genotype of this individual?

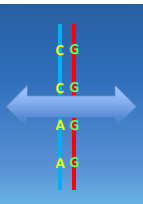
## The 16 Alleles (Haplotypes) possible for These Four Sites

- CCAA
- CGAA
- GCAA
- GGAA
- CCAG
- CGAG
- GCAG
- GGAG
- CCGA
- CGGA
- GCGA
- GGGA
- CCGG
- CGGG
- GCGG
- GGGG

Four genotypes have the same unphased phenotype: heterozygous at all four sites

## The Functional Effects of MPS




MPS methods allow clonal sequencing of individual strands, thereby distinguishing the parental haplotypes at a locus

## A Focus on Microhaplotype Loci Since 2012 The Exploratory & Discovery Phases

- Kidd et al. 2013. FSI:Genetics Supplement Series 4:e123-e124
- Kidd et al. 2014. FSI:Genetics 12:215-224
- Kidd and Speed 2015. Investigative Genetics 6:1
- Kidd et al. 2015. FSI:Genetics Supplement Series 5:e677-e679
- Kidd et al. 2017. FSI:Genetics


Forensic Science International: Genetics 29 (2017) 29–37

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Research paper

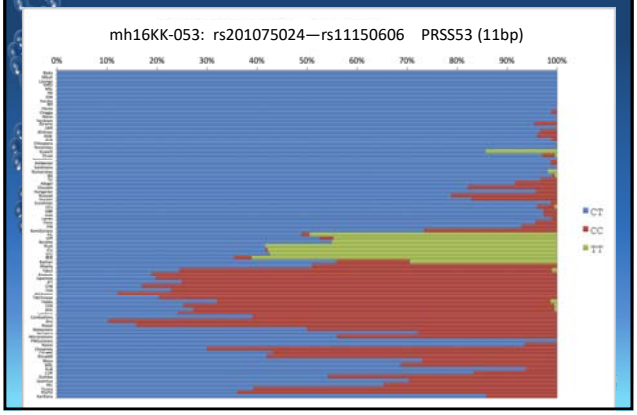
Evaluating 130 microhaplotypes across a global set of 83 populations

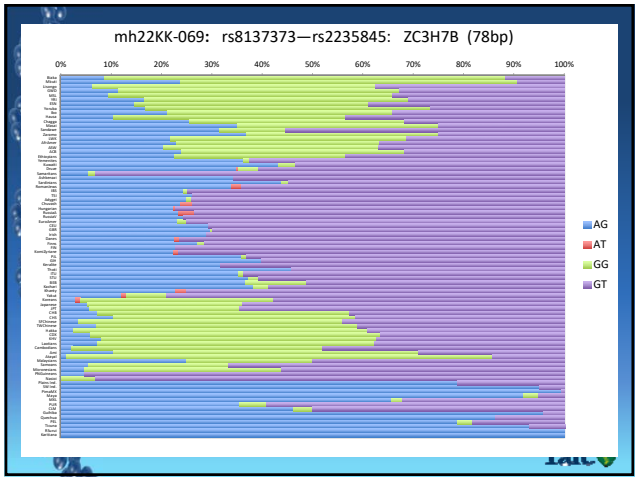
Kenneth K. Kidd[a,*], William C. Speed[a], Andrew J. Pakstis[a], Daniele S. Podini[b], Robert Lagacé[c], Joseph Chang[c], Sharon Wootton[c], Eva Haigh[a], Usha Soundararajan[a]

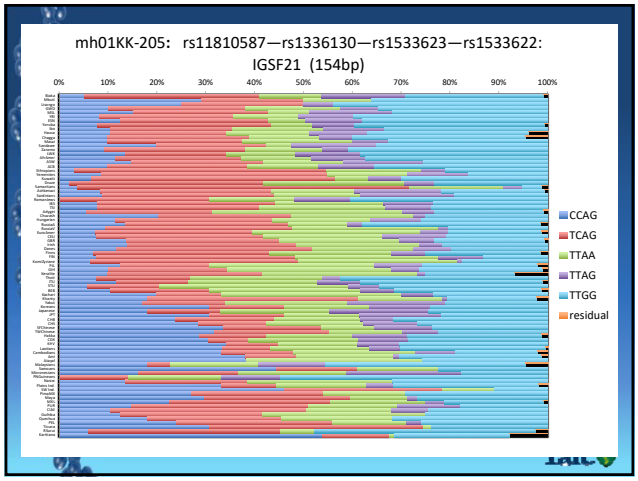
## Discovery Phase Has Reached 182 Microhap Loci Evaluated in 83 Populations

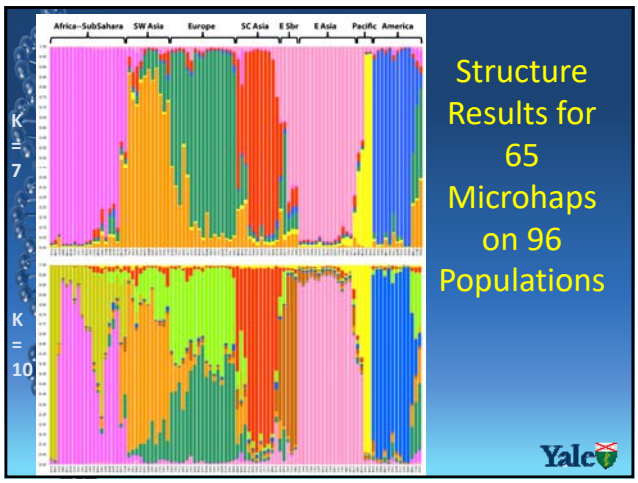
- All are < 300bp in extent, nearly half are < 100bp
- Typed as individual SNPs by TaqMan and PHASEd
- All have at least three alleles in almost all of the 83- 96 populations studied
- The allele (haplotype) frequencies of all microhaplotype loci and for all individual SNPs are in ALFRED <alfred.med.yale.edu> for each of the 130 loci and each of the 83 populations under the keyword *microhap.*

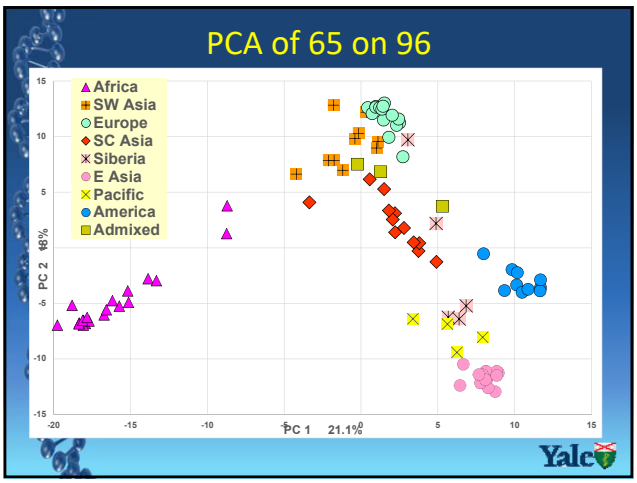
## Examples of microhaplotypes

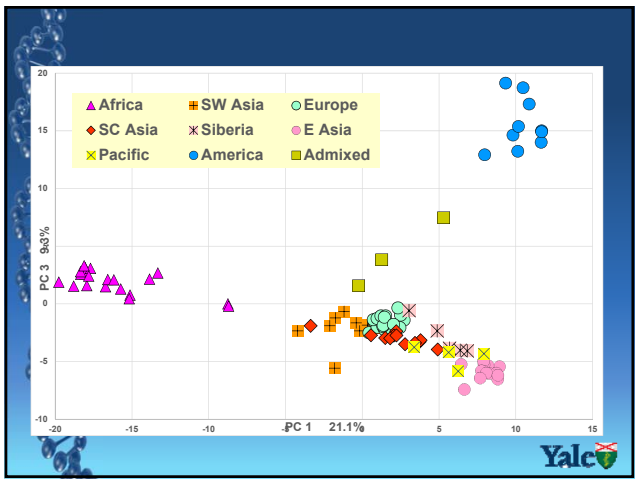


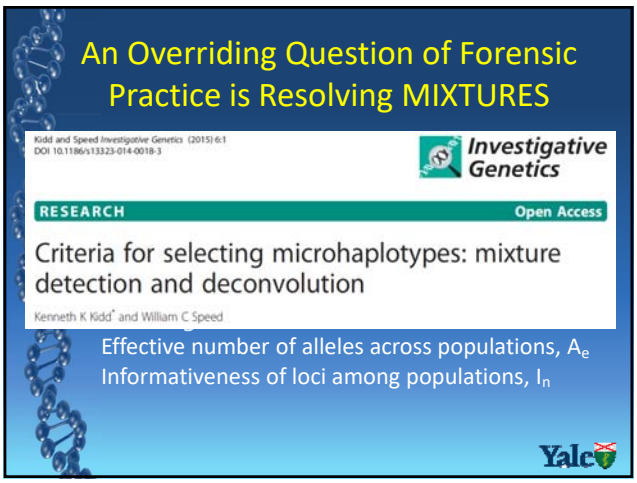



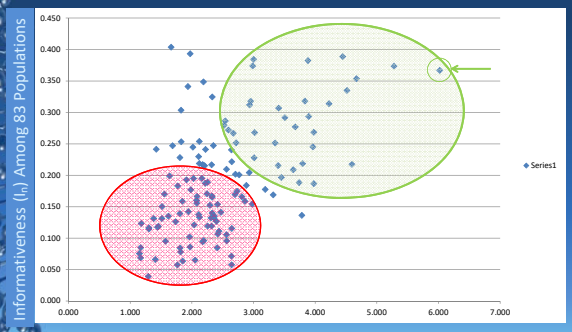

Africa–SubSahara
SW Asia
Europe
SC Asia
E Sbr
E Asia
Pacific
America
K = 7
K = 10
Structure Results for 65 Microhaps on 96 Populations


PCA of 65 on 96
Africa
SW Asia
Europe
SC Asia
Siberia
E Asia
Pacific
America
Admixed
PC 2 18%
PC 1 21.1%


Africa
SW Asia
Europe
SC Asia
Siberia
E Asia
Pacific
America
Admixed
PC 3 9.3%
PC 1 21.1%


An Overriding Question of Forensic Practice is Resolving MIXTURES
Kidd and Speed Investigative Genetics (2015) 6:1
DOI 10.1186/s13323-014-0018-3
Investigative Genetics
RESEARCH
Open Access
Criteria for selecting microhaplotypes: mixture detection and deconvolution
Kenneth K Kidd* and William C Speed
Effective number of alleles across populations, $A_e$
Informativeness of loci among populations, $I_n$

## Some loci are good for both mixtures and ancestry
## Some loci are not good for either purpose




Global Effective Number of Alleles ($A_e$) across 83 populations

## How Do Microhaps Work in the Lab?
## The Evaluation Phase

- Collaborators and co-authors at Thermo Fisher have designed a multiplex reaction for 74 of these microhap loci
- Collaborators and co-authors at The George Washington University are using these assays with Ion Torrent technology

**ThermoFisher**
- Sharon Wootton
- Robert Lagace
- Joe Chang
- Ryo Hasegawa

**GW University**
- Daniele Podini
- Lindsay Bennett
- Kelly Long
- Rebecca Walter
- Katrina Maddela
- Fabio Oldoni

## Developing the Panel from the 130 Microhaplotypes

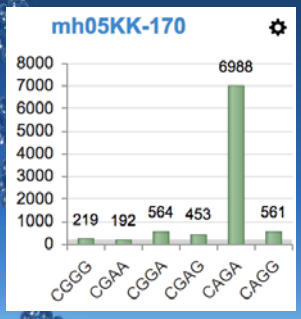
Selection of loci with high $A_e$
Loci with clear allele calls
Confirmation of alleles in discovery phase
Locus balance demonstrated
Allelic balance demonstrated
Artifacts shown to be minimal
Sensitivity at less than 1 ng
Mixture detection at 1:40 and better

→ A Panel of 74 Microhaplotypes

## Mixture Example

When looking at all the loci the profile is consistent with having one major contributor

## Mixture Example




At this locus there are 5 alleles that belong to minor contributors indicating the presence of at least three individuals (plus one major).

Thus the mixture is consistent with at least 4 individuals

## 4 person Mixture ~ 10:1:1:1


Single Source Profiles

Mixture

## The advantages of microhaps over STRs when processing casework mixture samples

- Microhaplotypes are better for mixture cases where:
  - STR data for the minor contributor are compromised such that a useable STR profile is not obtained
    - Low level of the minor contributor (undetected)
    - Minor contributor peaks masked by multiple donors
  - Stutter peaks confound interpretation

## MPS STR profile of the same mixture sample

## Stutter or minor contributor?

## Theoretical Probability of Detecting a Mixture using 28 Microhaps with $A_e > 3.0$

| | | | Number of Loci | | | | |
|---|---|---|---|---|---|---|---|
| $A_e$ | 1 | 2 | 3 | 4 | 5 | 6 | 21 |
| 3 | 0.4444 | 0.6913 | 0.8285 | 0.9047 | 0.9471 | 0.9706 | **0.999995** |
| 4 | 0.6563 | 0.8818 | 0.9594 | 0.9860 | **0.9952** | 0.9984 | |
| 5 | 0.7680 | **0.9462** | 0.9875 | 0.9971 | 0.9993 | 0.9998 | |

Using these 28 loci the chance of missing a second contributor is $< 10^{-7}$ but how many contributors are there?

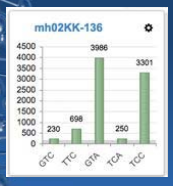
## Mixture Example

When looking at all the loci the profile is consistent with having one major contributor

## Given the data on multiple loci, how many contributors are there?

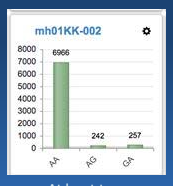
- One of 18 loci indicates at least four individuals
- Two of 18 loci indicate at least three individuals
- Fifteen of 18 indicate at least two individuals
- Are the three anomalous or do all 18 loci agree with four individuals in the mixture?

## How Many Individuals?

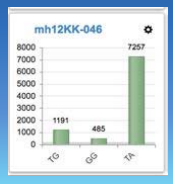



At lease three


At least three




At least two



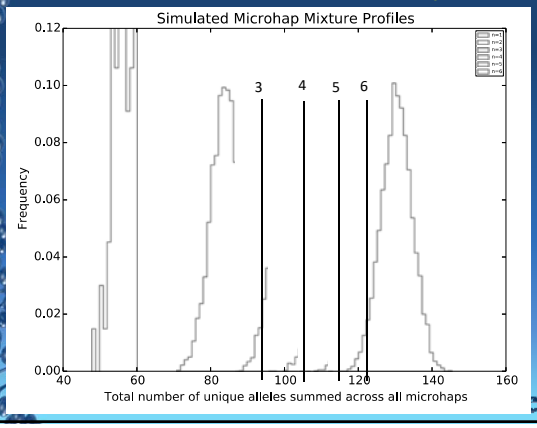

At least two, maybe 3

## Sharon Wooton's Idea: Simply sum the numbers of alleles detected across all loci

The logic is that the more individuals in the mix the more likely that more of the possible alleles will appear.

For example, a locus with only three possible alleles is more likely to have all three if there are more than two individuals in the mix

## MPS-MH Mixture study

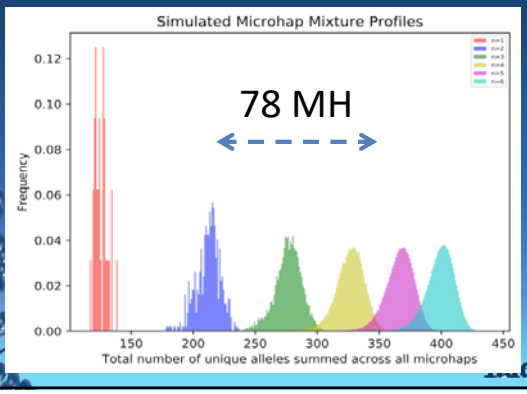



## MPS-MH Mixture study

33 Microhaps were typed on a set of mixtures

| | |
|---|---|
| 3 person 10: 1:1 | Total individual alleles: 106 |
| 4 person 10:1:1:1 | Total individual alleles: 116 |
| 5 person 10:1:1:1:1 | Total individual alleles: 128 |
| 6 person 1:1:1:1:1:1 | Total individual alleles: 137 |

## With more loci the total number of alleles seen in a mixture increases




## Acceptance in the Courts?

Daubert and Frye Criteria
re New Types of Scientific Evidence

- Whether the theory or technique employed by the expert is generally accepted in the scientific community;
- Whether it has been subjected to peer review and publication;
- Whether it can be and has been tested, is reliable;
- Whether the known or potential rate of error is acceptable; and
- Whether the research was conducted independent of the particular litigation or dependent on an intention to provide the proposed testimony.

## Laboratory Methodology and Statistical Interpretation

- These are the two different components of DNA in the Courtroom.
- Both need to be documented sufficiently to meet the Daubert/Frye standards
- A problem is that, to date, the Kidd Lab and collaborators are the only ones who have published on microhaplotypes; that should change soon.

## Precedents?

- DNA polymorphisms have been accepted
- Multiallelic genetic systems have been accepted
- Haplotypes, e.g. HLA, have been accepted
- Allele frequencies and population differences have been accepted
- SNPs have been accepted?
- Sequencing of DNA has been accepted

## Conclusions

- Microhaplotypes (MHs) *are* "The Next Generation Forensic DNA Marker" lacking only reference databases of "criminals"
- MHs fulfill all of the requirements for a multi-locus forensic panel: individualization, ancestry inference, relationship identification, mixture deconvolution
- MHs are better than STRs for several reasons

## Conclusions

- Microhaplotype (MH) heterozygote allele read ratios (RR) with MPS are comparable to STR peak height ratios (PHRs) with CE and vary less since all alleles at a locus are the same size
- With MHs contributor read ratios in a mixture are consistent throughout loci
- MHs allow the identification of the presence of multiple contributors in a mixture
- MHs have been shown to be effective on synthetic mixtures and forensic samples
- **NO STUTTER!**

## Acknowledgements

Yale University (Kidd Lab)
Andrew J. Pakstis
William C. Speed
Judith R. Kidd
Haseena Rajeevan
Usha Soundararajan
Randall Branchcomb

Thermo Fisher HID Group
Sharon Wooton
Robert Lagace
Joseph Chang
Ryo Hasagawa

The George Washington University
Daniele Podini
Lindsay Bennett
Kelly Long
Rebecca Walter
Katrina Maddela
Fabio Oldoni

## Acknowledgements

Special thanks are due to the many hundreds of individuals who volunteered to give DNA samples for studies of gene frequency variation.

## Acknowledgements

The discovery aspects of this work were funded primarily by NIJ Grants 2015-DN-BX-K023, 2013-DN-BX-K023, and 2010-DN-BX-K225 to KKK awarded by the National Institute of Justice, Office of Justice Programs, US Department of Justice. Points of view in this presentation are those of the authors and do not necessarily represent the official position or policies of the U.S. Department of Justice.

## My colleagues and I thank you for your attention

References for recent papers can be found under publications on the Kidd Lab web site

https://medicine.yale.edu/lab/kidd

alfred.med.yale.edu
frog.med.yale.edu