Selecting microhaplotypes optimized for different purposes

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 31, 2018

Forensic Uses of Microhaplotypes

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

An Overriding Question of Forensic Practice is Resolving MIXTURES


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, \( I_n \)

Global Effective Number of Alleles (\( A_e \)) across 83 populations
Populations Analyzed for Ancestry with 65 microhaps

Structure analysis of 96 populations using 65 microhaplotype loci: K=10

Selecting the “best” loci for ancestry inference
- The best loci for ancestry inference are those with high variation in allele frequencies among populations
- Different measures can be used: $F_{st}$, $I_n$, etc.
- We have used Rosenberg’s Informativeness, $I_n$

Selecting the “best” loci for individualization
- The best loci for individual identification are those with high variation in alleles within each population
- Different measures can be used: Het, $A_e$, etc.
- Each of those is population specific
- We have used the global average effective number of alleles, $A_e$
Ancestry with the top 50 $A_e$ and $I_n$

PCA using top $I_n$ microhaps

K=9

PCA using top $A_e$ microhaps

Random Match Probabilities
Frequencies of Most Common Genotype

Selecting the “best” loci for mixture deconvolution

- The best loci for mixture deconvolution are those with a large number of alleles within each population
- Different measures can be used: average number of alleles, Het, $A_e$, etc.
- Each of those measures is population specific
- We have used the global average effective number of alleles, $A_e$

Illustrations re mixtures

Examples were given yesterday in the talk by Dr. Fabio Oldoni

Selecting the “best” loci for determining relationships

- The best loci for relationship determination are those with a large number of alleles within each population
- Different measures can be used: variance in frequencies of alleles, Het, $A_e$, etc.
- Each of those measures is population specific
- We have used the global average effective number of alleles, $A_e$
Bioinformatics Searches

- The 1000 Genomes data and the CEPH-HGDP datasets allow searches for any type of microhaplotype.
- The problems are global coverage for the 1000 Genomes data and genome coverage for the HGDP data.
- A “need” now is small amplicons that will allow use for any purpose with degraded DNA.

Target Capture Enrichment

Works best with fragments <75 bp

Global variation in a 50bp microhap

33 loci < 50 bp

How small is small enough?
Conclusions

- Microhaplotype loci (Microhaps) are “The Next Generation Forensic DNA Marker” lacking only reference databases of “criminals”
- Microhaps fulfill all of the requirements for a multi-locus forensic panel: individualization, ancestry inference, relationship identification, mixture deconvolution
- Sets of microhaps can be optimized for different purposes but many sets can fulfill many purposes
- More microhaps with high $A_e$ and short extent are needed
Acknowledgements

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

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

The George Washington University
Daniele Podini
Fabio Oldoni
Aishwaryaa Subramanian
Sathya Prakash Harihar
Drew Bader
Leena Yoon

Special thanks are due to the many
thousands of anonymous individuals who
volunteered to give DNA samples for
studies of genetic variation around the
world.

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