NIJ National Institute of Justice

Technology Transition Workshop| *Kenneth K. Kidd, Ph.D.*

How Basic Population Genetics Informs on Ancestry and Phenotype

Components of the Human Genome

- **Mitochondrial DNA (mtDNA): 16,569 bp**
- **Nuclear DNA: 3.3 billion bp**
  - **Autosomes: 22 chromosome pairs**
  - **Sex chromosomes: 1 pair of unmatched chromosomes**
    - **X chromosome**
    - **Y chromosome (male determining)**
      - **NRY: non-recombining part of the Y**
      - **Pseudoautosomal DNA : homologous to and recombines with the X, 2.6Mb**

Technology Transition Workshop

NIJ National Institute of Justice

The Concept of Polymorphism

A polymorphism is a site or locus in the genome that exists in "many forms" in the population. Usually, the most common allele must be less frequent than 99%.

Note, it is the site that is the polymorphism, not one allele at the site. Also, note that the term is often used, incorrectly, for ANY variation, even the single instance of a varying DNA sequence, as is the case for many so called SNPs (Single Nucleotide Polymorphisms), without knowing the frequency of the variant. Many sites with such instances of a single instance of a variant have been tested on many individuals and no variants found. This has given rise to the absurdity of a monomorphic SNP.

A rare variant is NOT a polymorphism.

Technology Transition Workshop NIJ National Institute of Justice

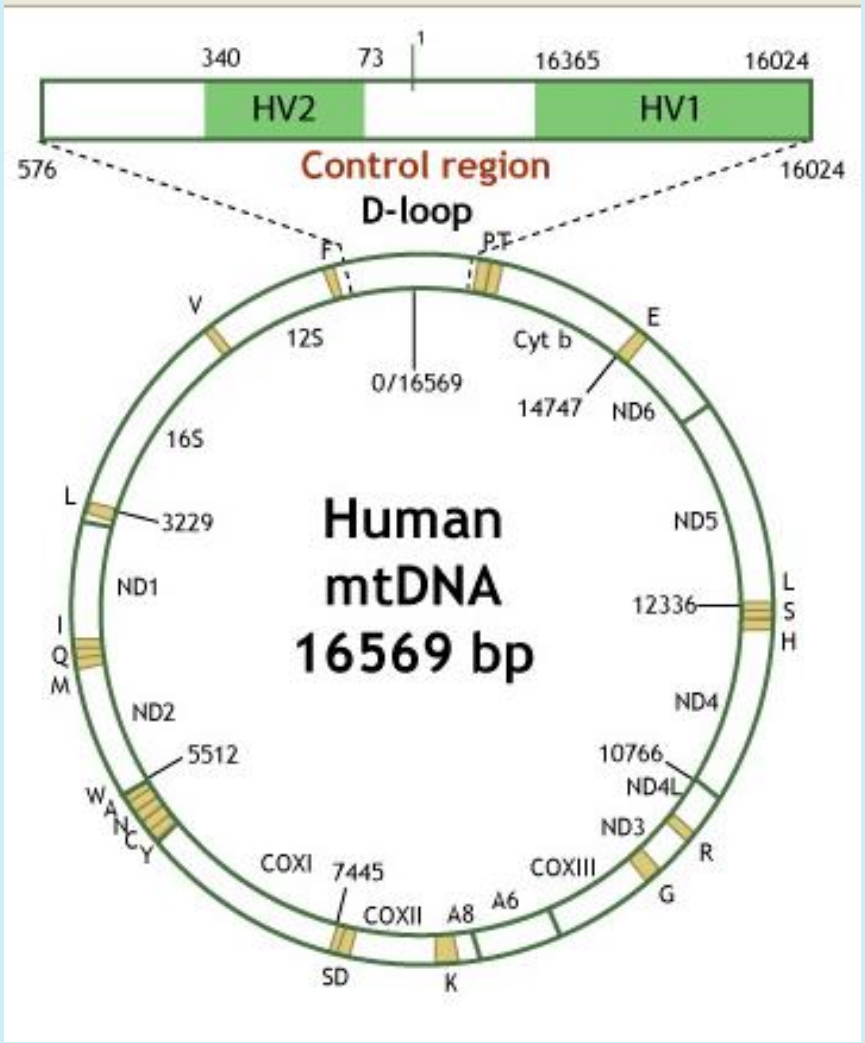
The Types of DNA Sequence Polymorphisms

- **RFLP: Restriction Fragment Length Polymorphism***
- **STRP: Short Tandem Repeat Polymorphism**
- **VNTR: Variable Number of Tandem Repeats**
- **InDel: Insertion/Deletion**
- **SNP: Single Nucleotide Polymorphism**
- **CNV: Copy Number Variation/Variant**

Technology Transition Workshop NIJ National Institute of Justice

Human mtDNA

The mitochondrial DNA is almost all functional. It codes for several proteins that are involved in the energy generating aspect of metabolism. It also codes for the transfer RNAs that are used to translate that code into proteins. The control region (also called the D-loop) is highly variable in sequence. This circular DNA molecule is inherited only through females and exists in thousands of copies in each cell.



http://www.nfstc.org/pdi/Subject09/pdi_s09_m02_01_a.htm

Technology Transition Workshop

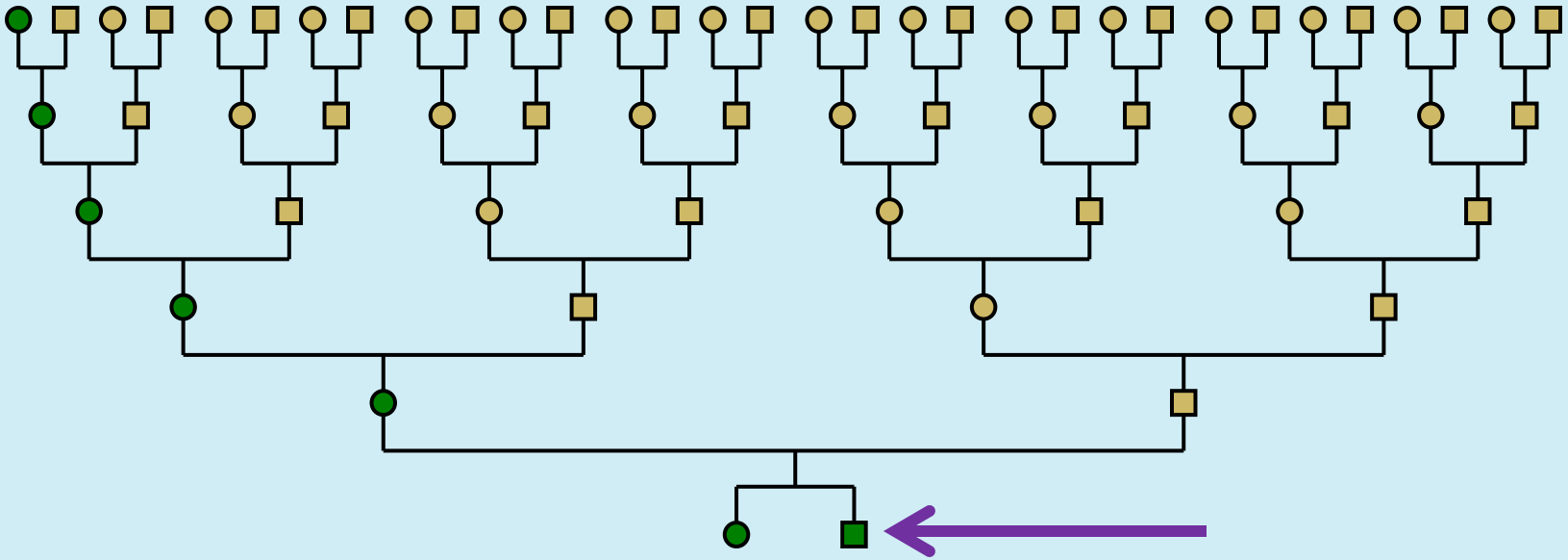
NIJ National Institute of Justice

Types of Variation in mtDNA

- **The control region segments HVI and HVII that show many different sequences**
- **Individual nucleotide differences in the coding regions of the molecule**

Globally there are many different mtDNA types, but within any single population there are not a large number of different types.

Gene Lineages: mtDNA

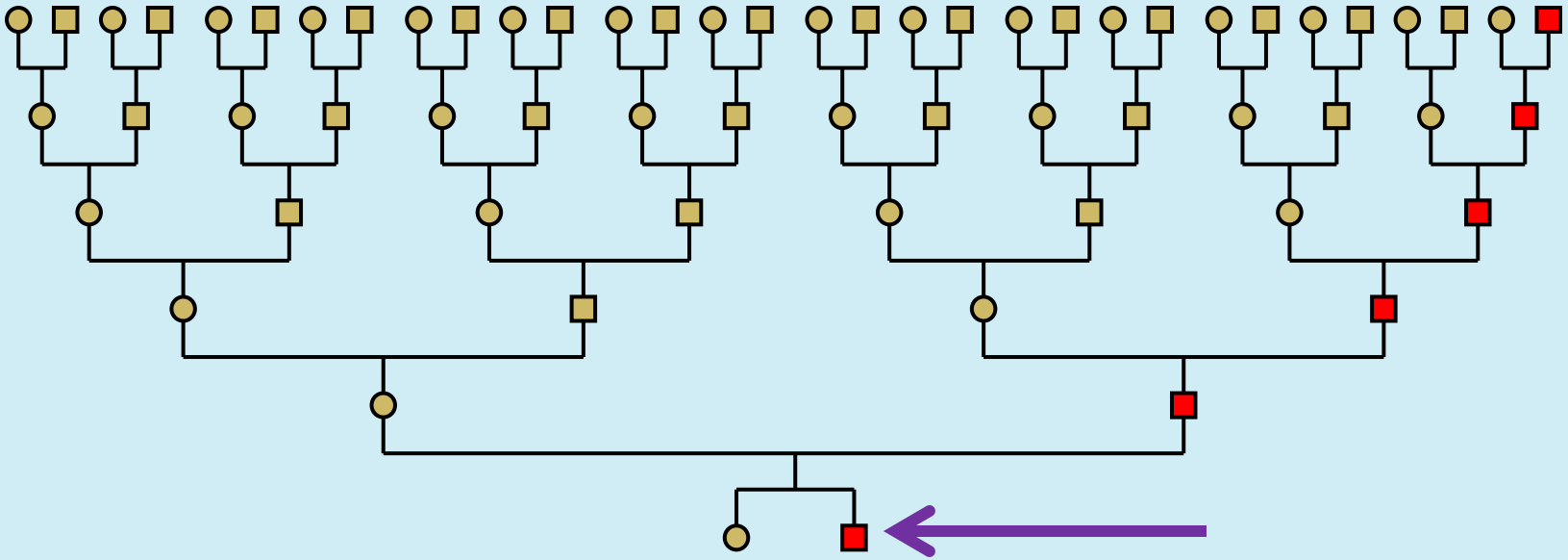


With respect to ancestry of an individual, mtDNA variation provides information on the maternal lineage. This is one ancestor out of 32 ancestors just 5 generations ago.

Technology Transition Workshop

NIJ National Institute of Justice

Gene Lineages: NRY



With respect to ancestry of an individual, NRY variation provides information on the paternal lineage. This is one ancestor out of 32 ancestors just 5 generations ago.

Technology Transition Workshop

NIJ National Institute of Justice

Types of Variation in the Non-Recombining Y Chromosome DNA (NRY)

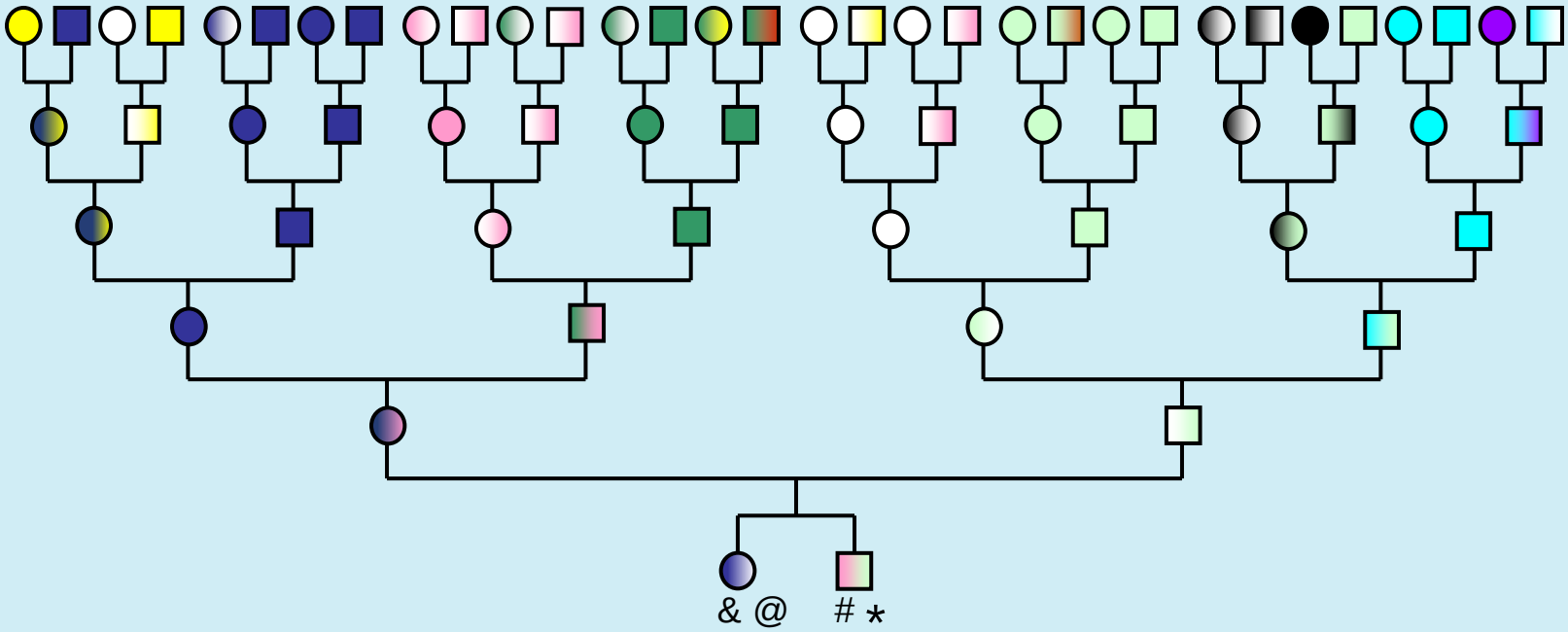
- **Many STRPs occur within the non-recombining regions of the chromosome**
- **Individual nucleotide differences (SNPs) occur in the coding and non-coding regions of the chromosome**

Globally there are many different NRY types, but within any single population there are not a large number of different types based on SNPs. However, within each of the haplogroups, there can be many STRP variants because of their higher mutation rate.

Technology Transition Workshop

NIJ National Institute of Justice

Gene Lineages: Autosomal

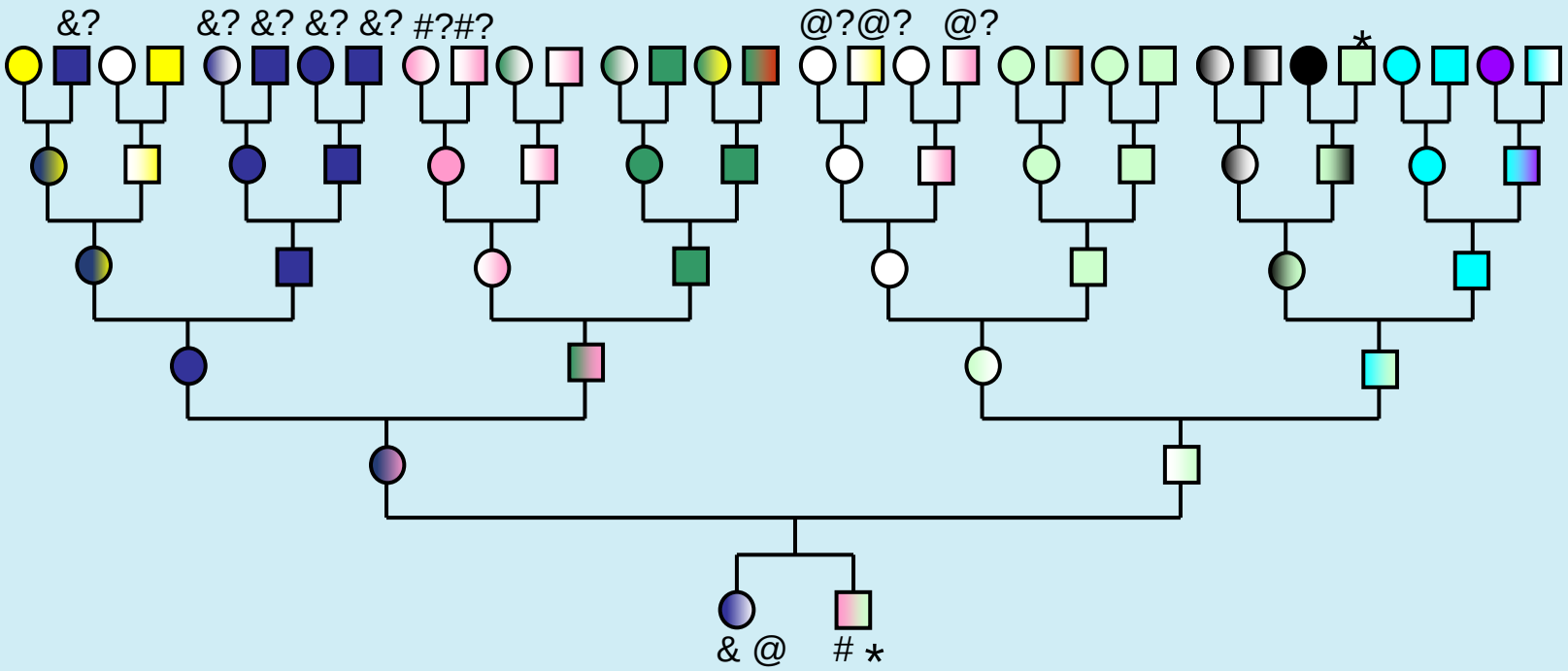


Each chromosome has a different genealogy or ancestral origin. With recombination different segments of each chromosome have a different genealogy. What are the origins of the four chromosomes (&, @, #, and *) in the two siblings at the bottom?

Technology Transition Workshop

NIJ National Institute of Justice

Gene Lineages: Autosomal



Only allele * has an unambiguous ancestral origin. Tracing identity by descent breaks down when an ancestor is homozygous for the allele.

Technology Transition Workshop

NIJ National Institute of Justice

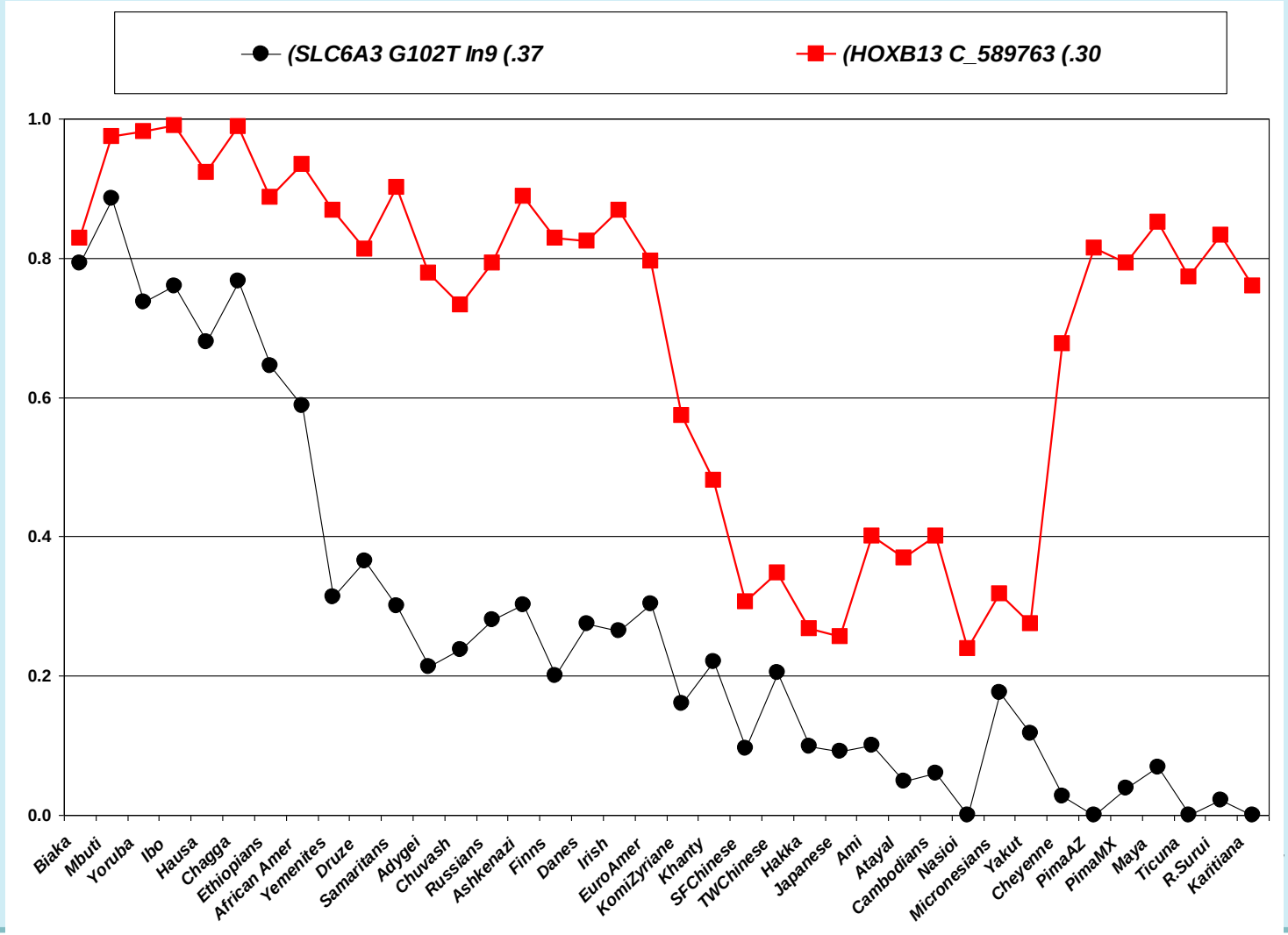
Allele Frequency Variation Among Populations is Measured Using F_{st}

F_{st} is a standardized variance of allele frequencies among populations. In theory it is related to random genetic drift with variance accumulating as a function of time in generations divided by twice the effective population size. For our purposes it is simplest to consider it one possible measure of allele frequency variation among populations.

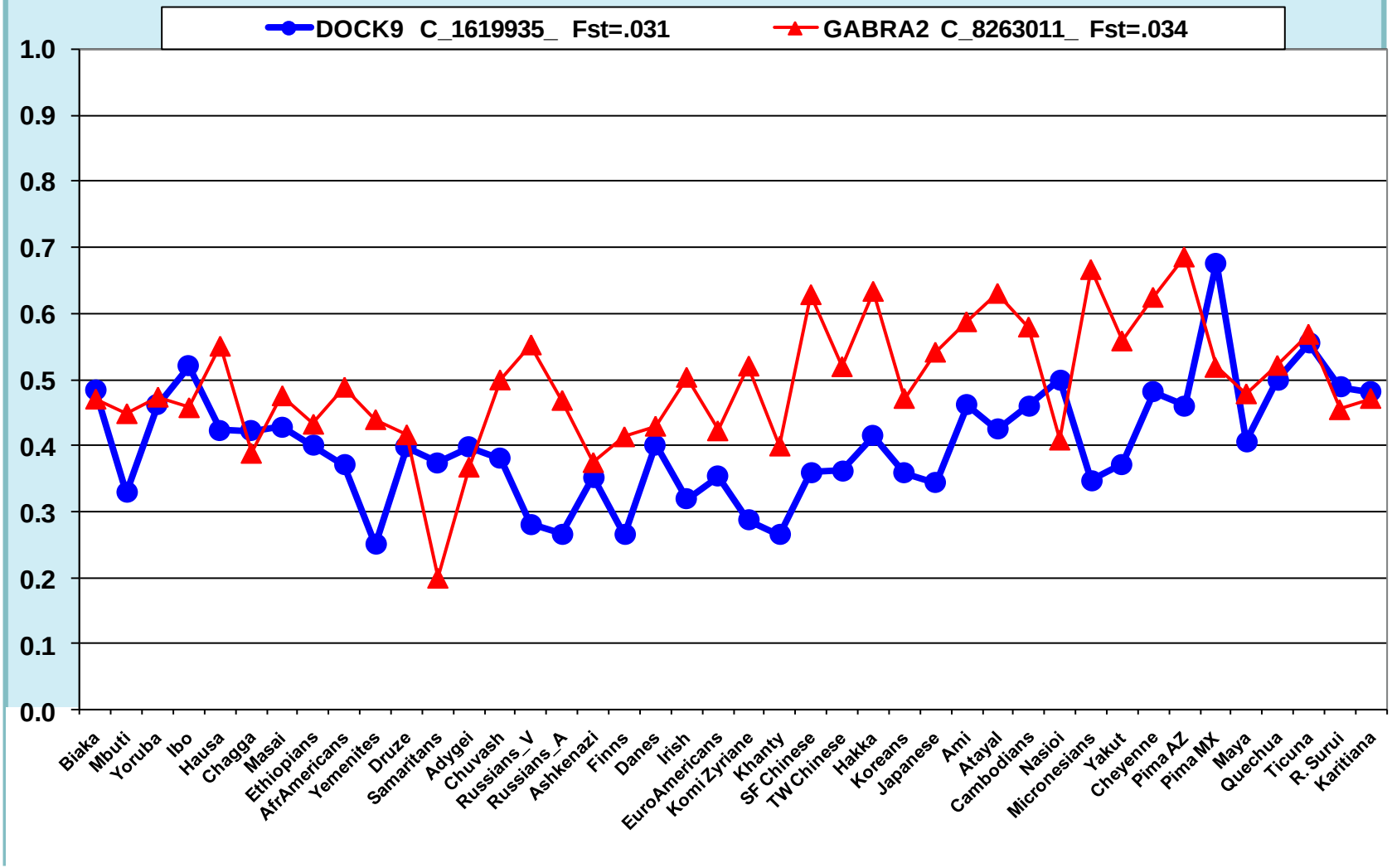
Technology Transition Workshop

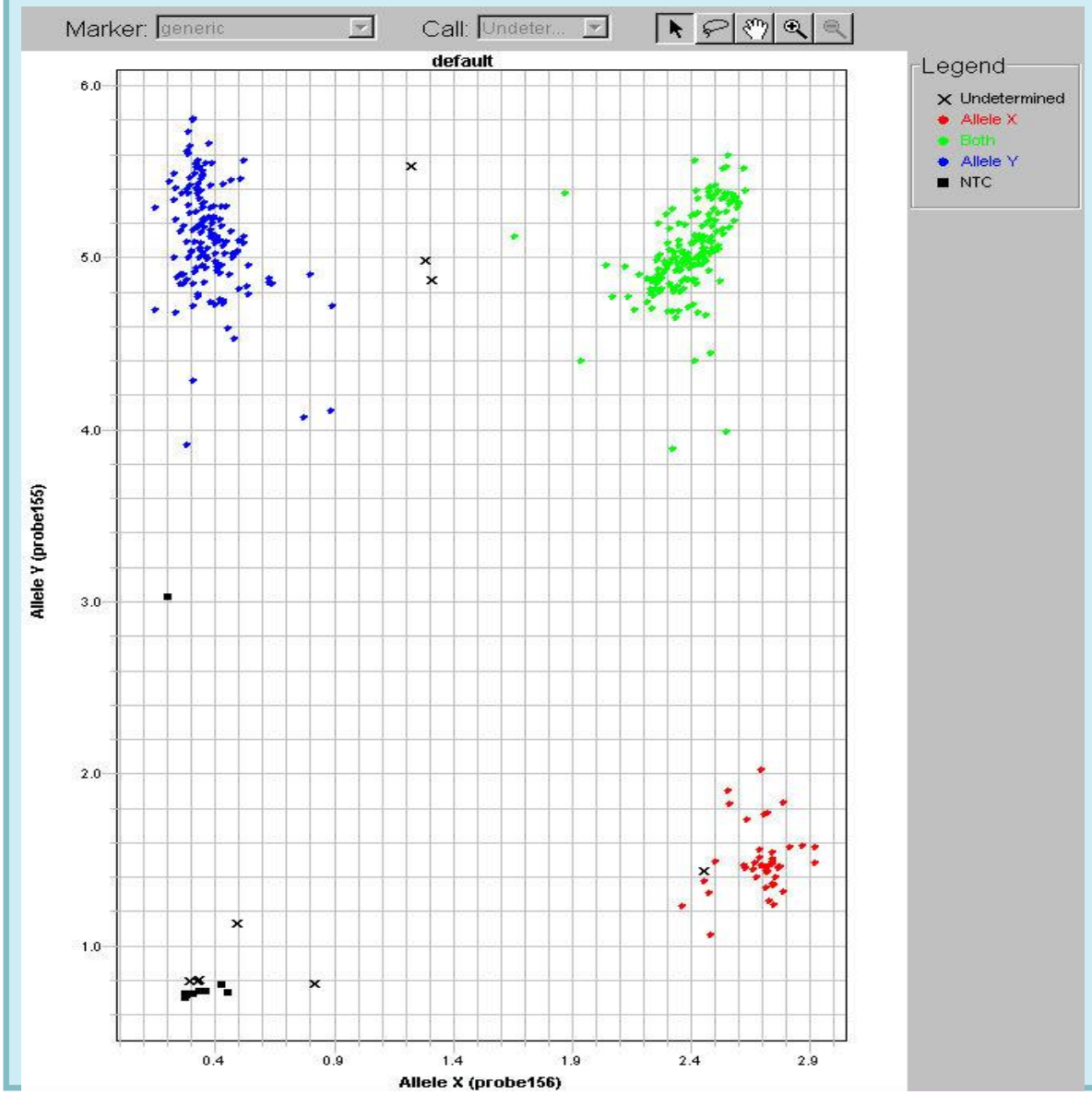
NIJ National Institute of Justice

SNPs with High F_{st}



SNPs with Very Low F_{st}





SNP Marker Typed by TaqMan

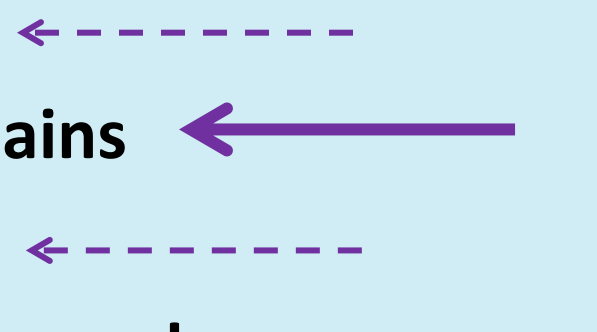
Each point is an individual and the cluster gives the genotype.

Technology Transition Workshop

NIJ National Institute of Justice

DNA is Used in Forensics to:

- **Identify a criminal**
- **Identify human remains**
- **Confirm parentage**
- **Exonerate innocent people**



Technology Transition Workshop

NIJ National Institute of Justice

DNA is Used in Forensics to:

- Identify a criminal
- **Identify human remains**
  - **Match to a known DNA profile**
  - **Determine phenotypic characteristics of the unknown**
  - **Determine ancestry (ethnic/geographic origins) of the unknown**

Technology Transition Workshop

Questions in Matching to a Known Person

- **What are the DNA patterns?**
- **Are the patterns the same? (Do they match?)**
- **What is the chance of two unrelated people having that pattern? (How unique is that DNA pattern?)**

Those questions are best answered by an individual identification panel with extensive population data documenting rarity of all genotypes. The CODIS markers or a panel of IISNPs would be appropriate.

Technology Transition Workshop

NIJ National Institute of Justice

Types of Panels of SNPs for Forensic Applications

Individual Identification SNPs (IISNPs): SNPs that collectively give very low probabilities of two individuals having the same multisite genotype.

Ancestry Informative SNPs (AISNPs): SNPs that collectively give a high probability of an individual's ancestry being from one part of the world or being derived from two or more areas of the world.

Lineage Informative SNPs (LISNPs): Sets of tightly linked SNPs that function as multiallelic markers that can serve to identify relatives with higher probabilities than simple di-allelic SNPs.

Phenotype Informative SNPs (PISNPs): SNPs that provide high probability that the individual has particular phenotypes, such as a particular skin color, hair color, eye color, etc.

Reproduced in Butler et al. 2008. *Progress in Forensic Genetics 12*

Technology Transition Workshop NIJ National Institute of Justice

Different Criteria for Different Panels

- **SNPs that meet population genetics criteria for the first three types of panels are a very small fraction of the millions of available SNPs**
- **Finding candidate SNPs requires an initial population genetics approach and adequate population resources**
- **Phenotype informative SNPs are also uncommon; many candidates are as yet poorly documented functionally and population-specific association appears to occur**

Technology Transition Workshop NIJ National Institute of Justice

General Criteria for Use of a SNP in Forensics

1. **An easily and reliably typable unique locus**
2. **Highly informative for the stated purpose**
3. **Well documented relevant characteristics such as allele frequencies**

Technology Transition Workshop

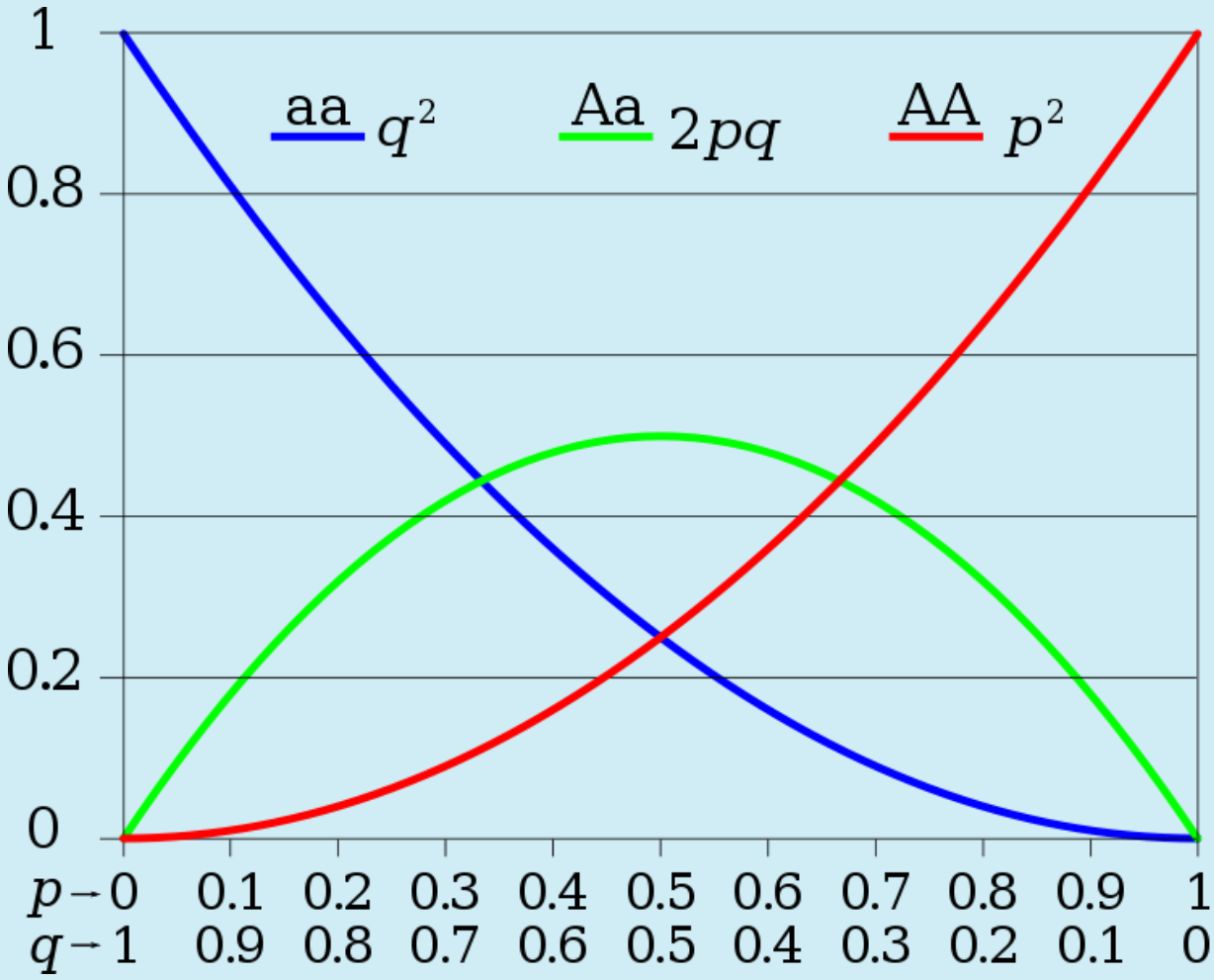
What SNPs Will Be Best?

1. **Those SNPs that will provide the maximum amount of information per SNP. What do we mean by information?**

2. **SNPs that are not subject to typing difficulties. What kind of "typing difficulties" can exist?**

Technology Transition Workshop NIJ National Institute of Justice

Hardy-Weinberg Ratios



H-W ratios are robust to the assumptions and can be used to relate allele frequencies in a population to the genotype frequencies in that population.

http://en.wikipedia.org/wiki/File:Hardy-Weinberg.svg

Technology Transition Workshop

NIJ National Institute of Justice

The Out-of-Africa Theory

The overall pattern of modern human variation is explained by the Out-of-Africa theory.

Modern humans evolved in Africa about 200,000 years ago. Considerable genetic variation accumulated.

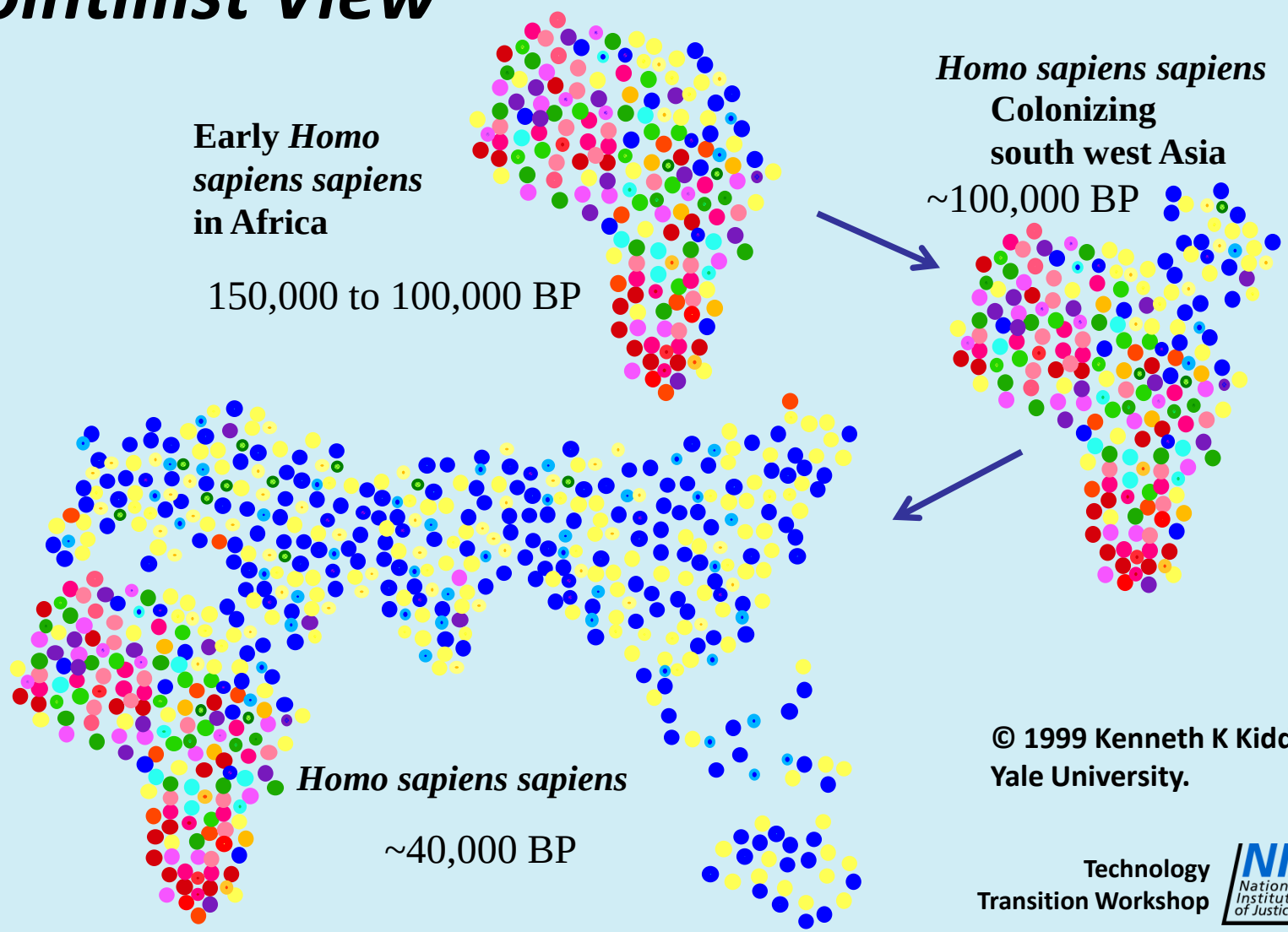
About 100,000 years ago (possibly as recently as 80,000 years ago) SOME individuals left Africa into Southwest Asia.

That SINGLE population had only a SMALL fraction of the genetic variation present in Africa. That population expanded to occupy the rest of the world.

Technology Transition Workshop

NIJ National Institute of Justice

Pointillist View



Signs of Recent Selection

A significantly higher F_{st} than the average suggests selection may have operated in one region of the world.

One common haplotype extending a much longer molecular distance than others indicates that it rapidly increased in frequency recently. This suggests that selection may have been the cause. (What might be another cause?)

Other indicators of selection are more theoretically based and involve less obvious assumptions, e.g., allele frequency distributions at a multi-allelic locus.

Genes Showing Evidence of Recent Selection

Many loci have been hypothesized to show evidence of selection but relatively few show strong evidence. The loci associated with resistance to malaria are among the best known: alpha and beta hemoglobin variants, G6PD variants, and the Duffy blood group. Others clearly have undergone selection but the nature of the selective factor is not clear. These include lactase (LCT) and alcohol dehydrogenase (ADH1B). The 900 kb inversion polymorphism on chromosome 17 has also been hypothesized to show evidence of selection. Many candidates are being studied and reported in the literature.

Technology Transition Workshop

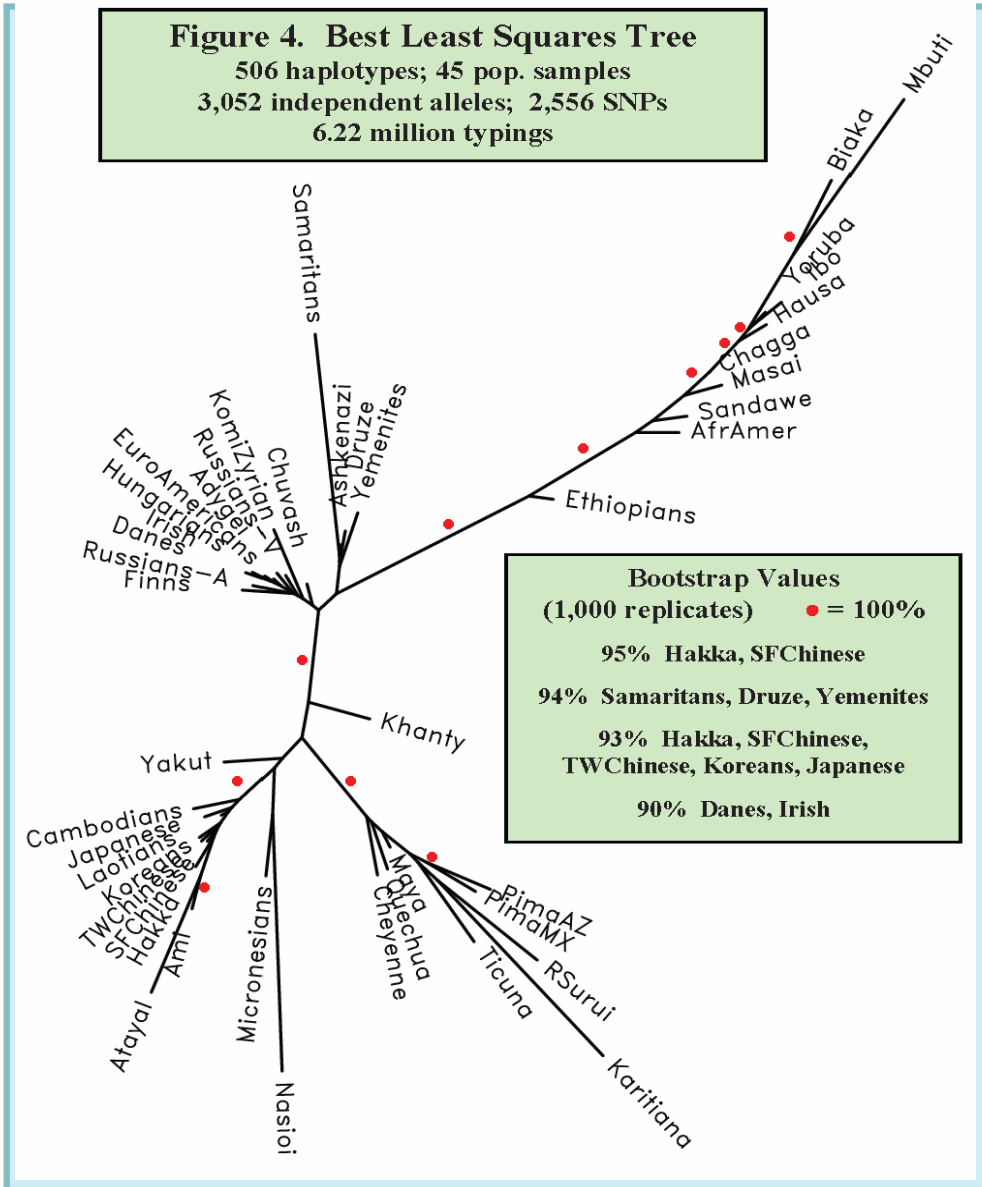
NIJ National Institute of Justice

46 Population Samples Studied Routinely in Kidd Lab

Attention: this is not a uniformly distributed sample!

Technology Transition Workshop

NIJ National Institute of Justice



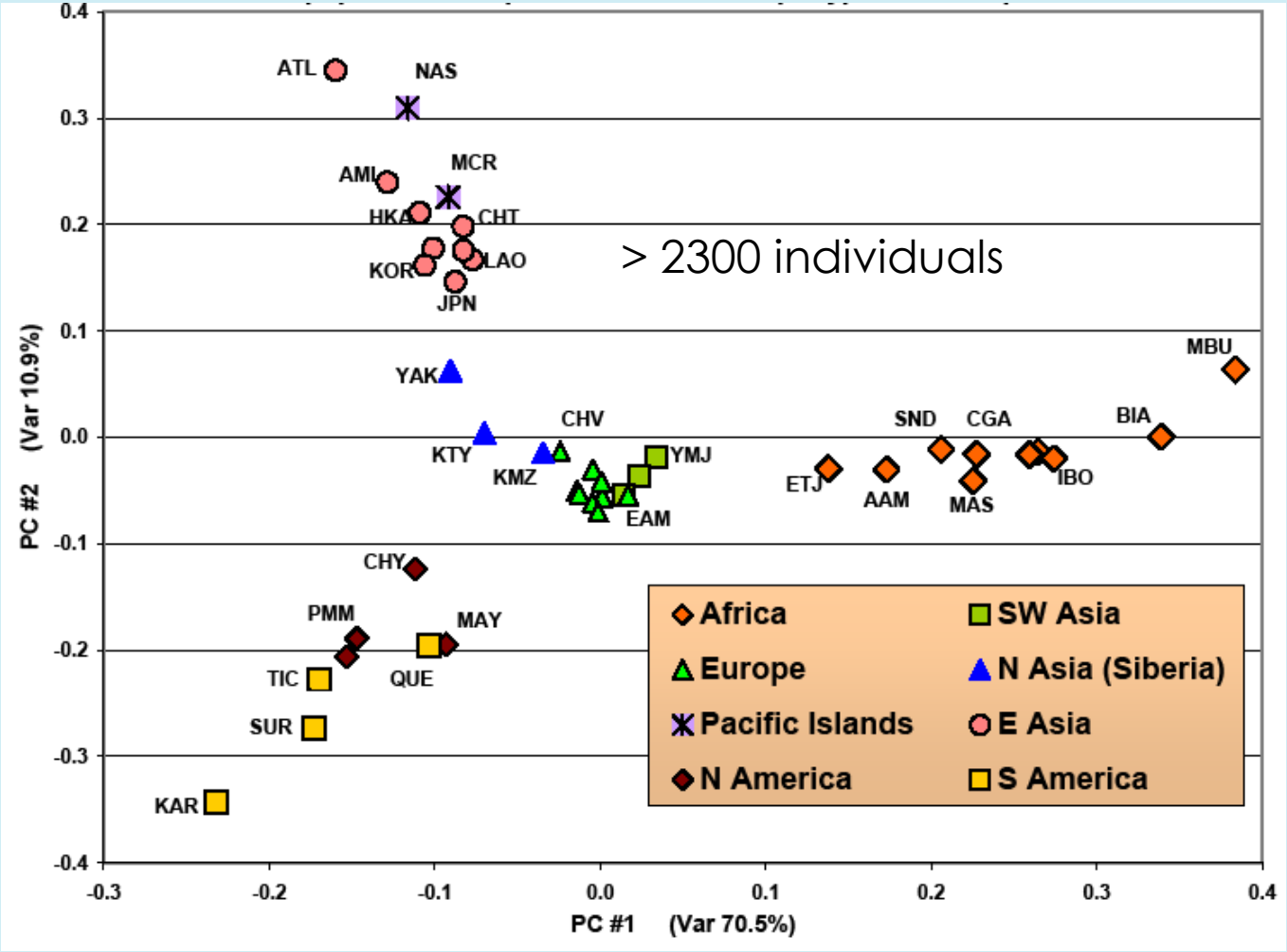
A Haplotype-based Tree with High Bootstrap Support

A redrawn version of this tree is being published in Kidd et al., *Amer. J. Phys. Anthro.* 2011, in press.

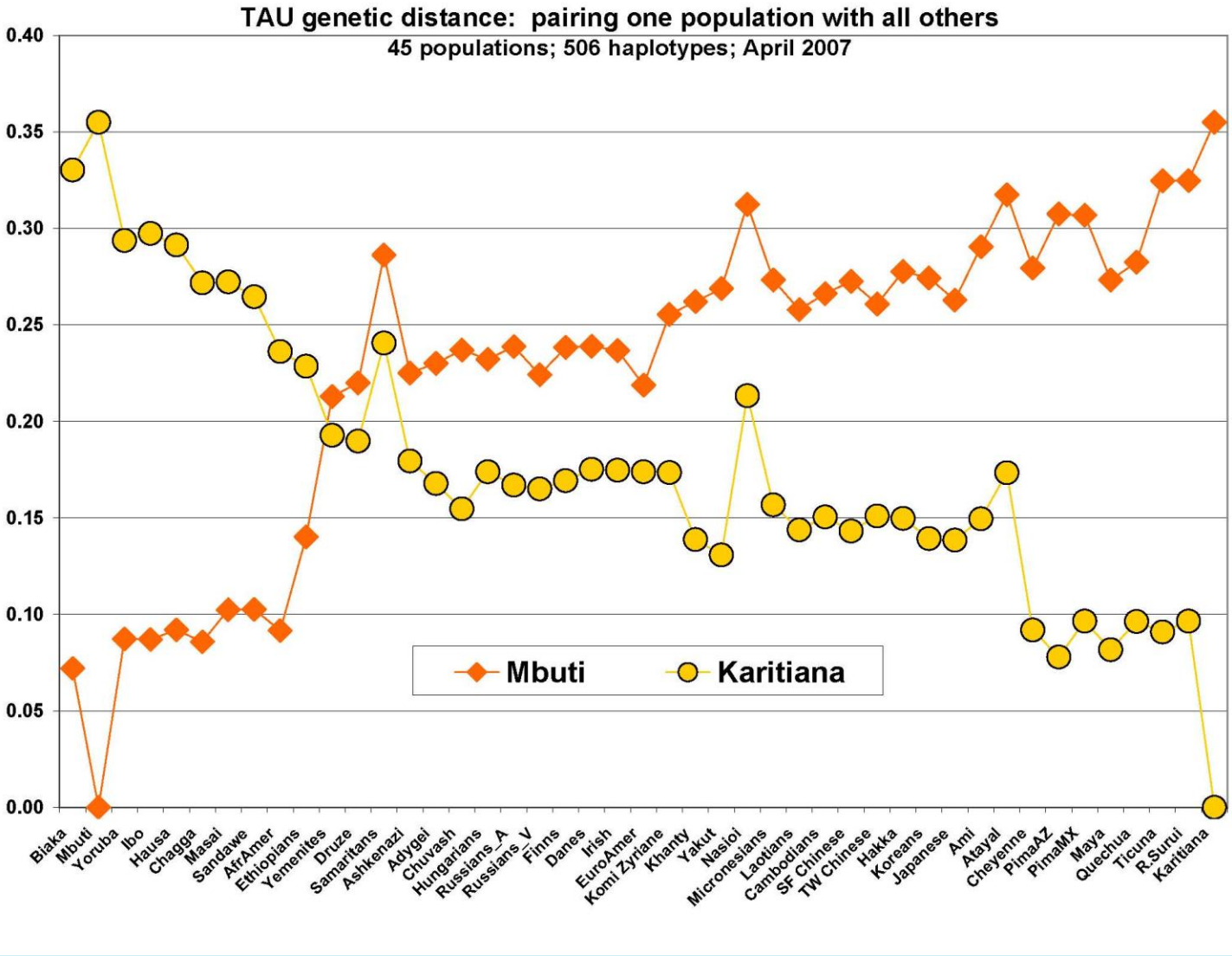
Technology Transition Workshop

NIJ National Institute of Justice

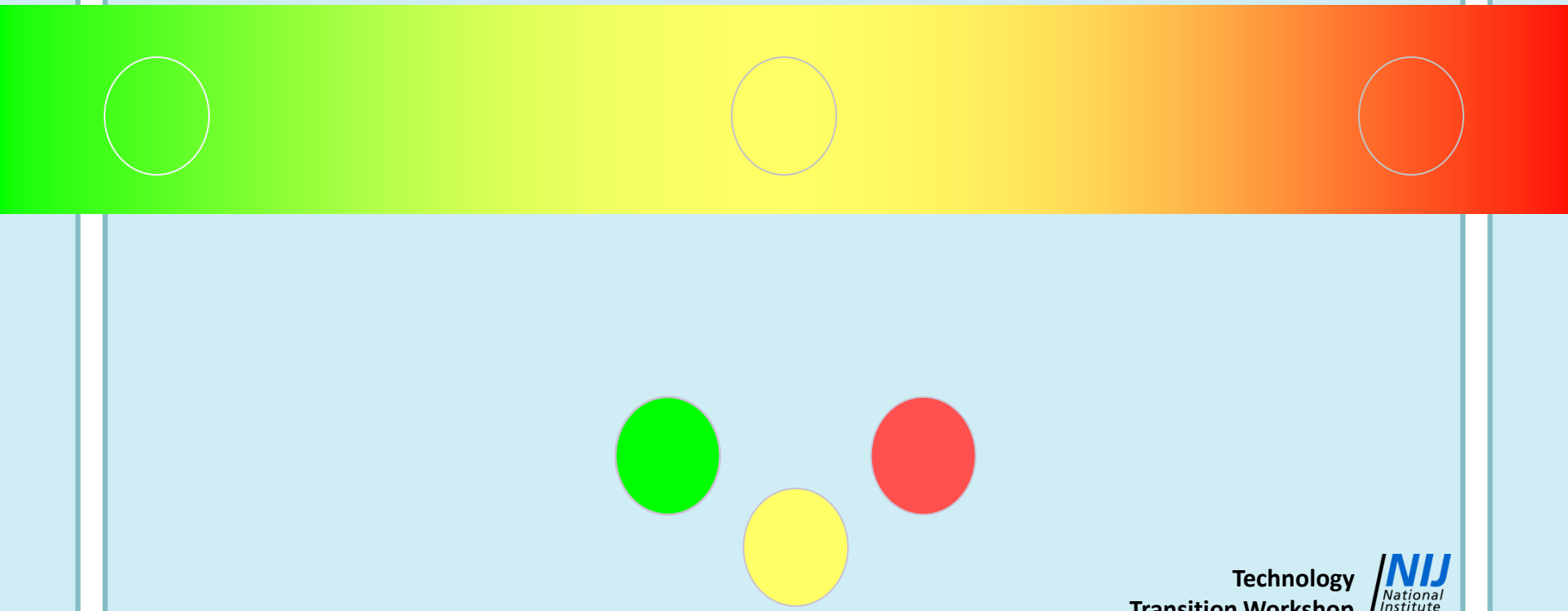
PCA of the 506 Haplotype Dataset



Technology Transition Workshop


TAU genetic distance: pairing one population with all others
45 populations; 506 haplotypes; April 2007
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
Mbuti
Karitiana
Biaka
Mbuti
Yoruba
Ibo
Hausa
Chagga
Masai
Sandawe
AfrAmer
Ethiopians
Yemenites
Druze
Samaritans
Ashkenazi
Adygei
Chuvash
Hungarians
Russians_A
Russians_V
Finns
Danes
Irish
EuroAmer
Komi Zyriane
Khanty
Yakut
Nasioi
Micronesians
Laotians
Cambodians
SF Chinese
TW Chinese
Hakka
Koreans
Japanese
Ami
Atayal
Cheyenne
PimaAZ
PimaMX
Maya
Quechua
Ticuna
R.Surui
Karitiana

Species-wide vs. Specific-population Comparisons



Technology Transition Workshop

Conclusions on the Overall Pattern of Genetic Diversity and the Causes

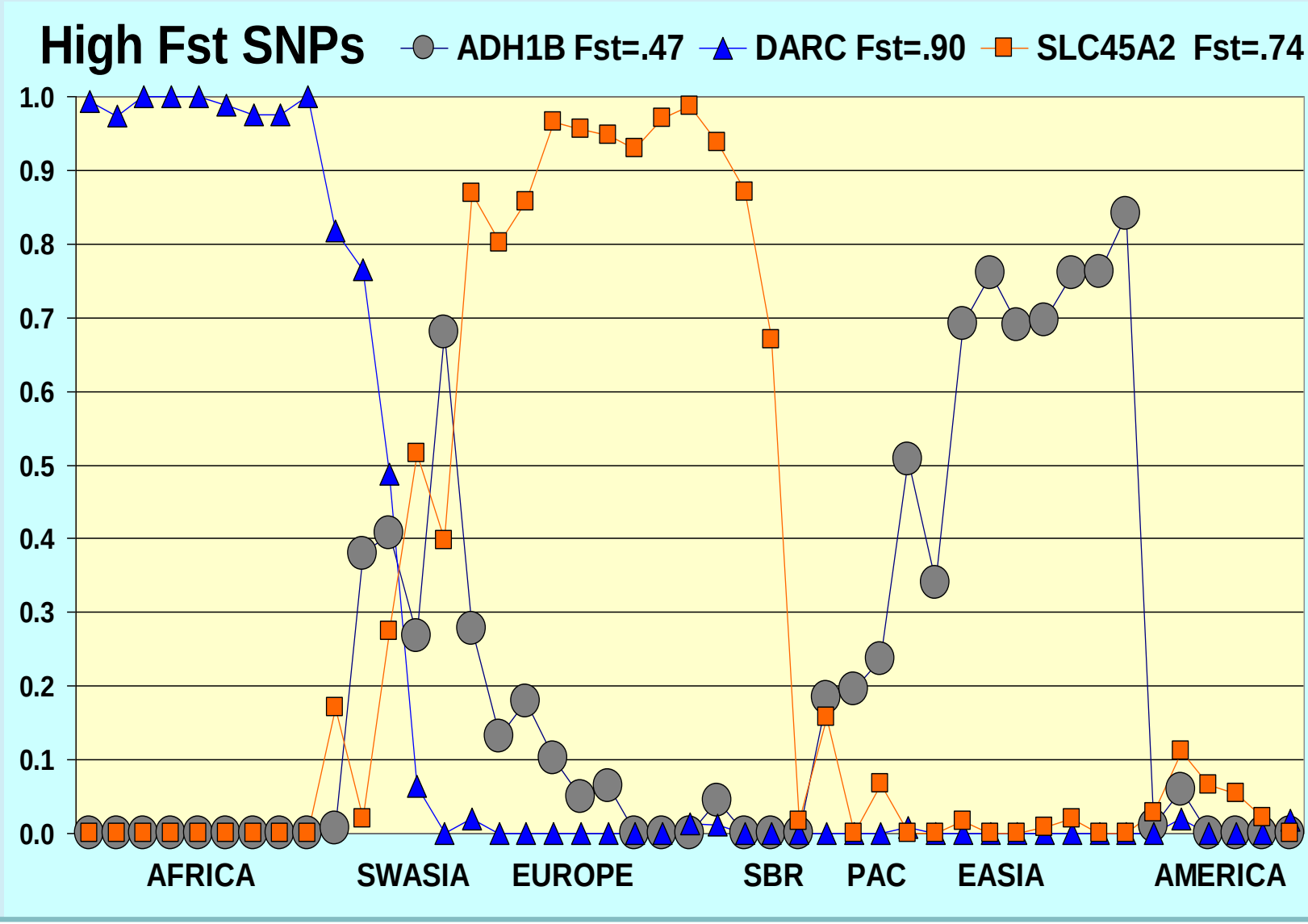
- **The expectation is that allele frequencies will be different in different populations**
- **The basic pattern of overall genetic diversity is primarily the result of random chance (genetic drift) in conjunction with the history of the spread of modern humans around the world**
- **However, there is evidence of natural selection affecting specific genes differentially in different parts of the world**

Variation in function can be, usually is, normal. This is the variation that makes every human unique.

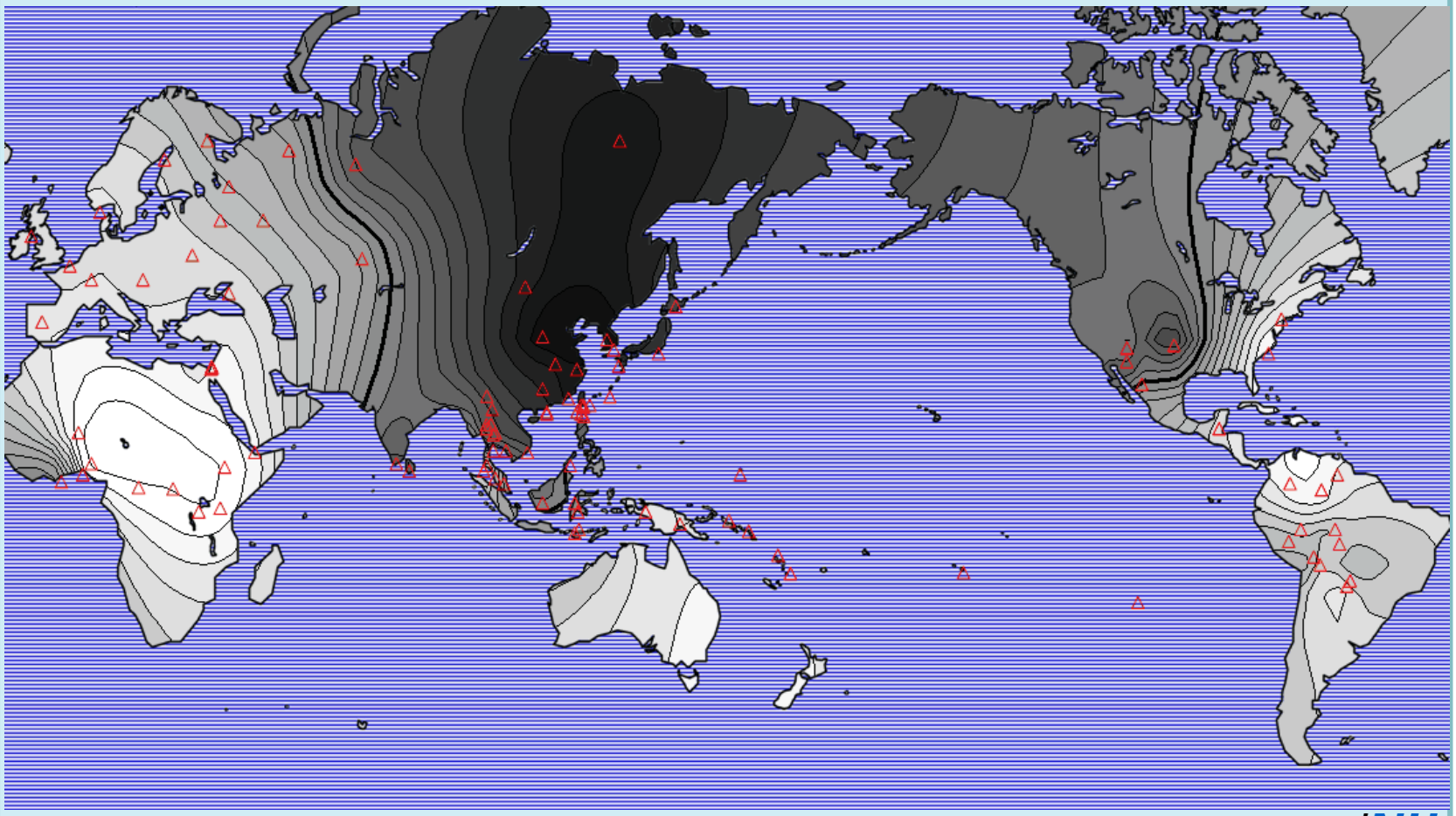
Many genes have common variation that gives, for example, different enzyme activities. Some of that variation is associated with different levels of risk for different disorders. However, a common allele that gives slightly increased risk for a disorder, say hypertension, is not abnormal. Nor is it necessarily selected for or against.

Technology Transition Workshop

NIJ National Institute of Justice


High Fst SNPs
ADH1B Fst=.47
DARC Fst=.90
SLC45A2 Fst=.74
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
AFRICA
SWASIA
EUROPE
SBR
PAC
EASIA
AMERICA

Wet vs. Dry Earwax Allele Frequency

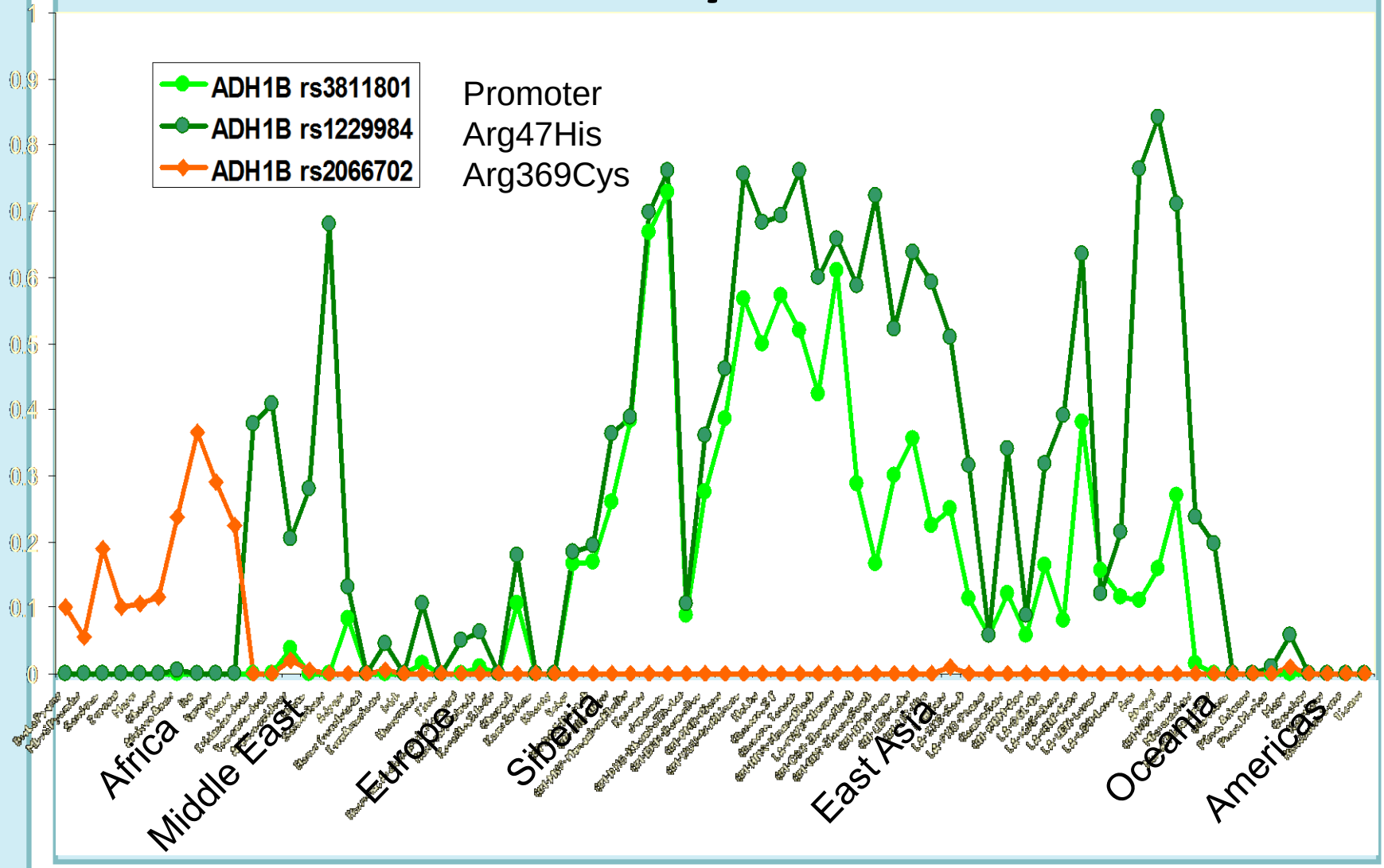


SURFER Plot of Data in ALFRED

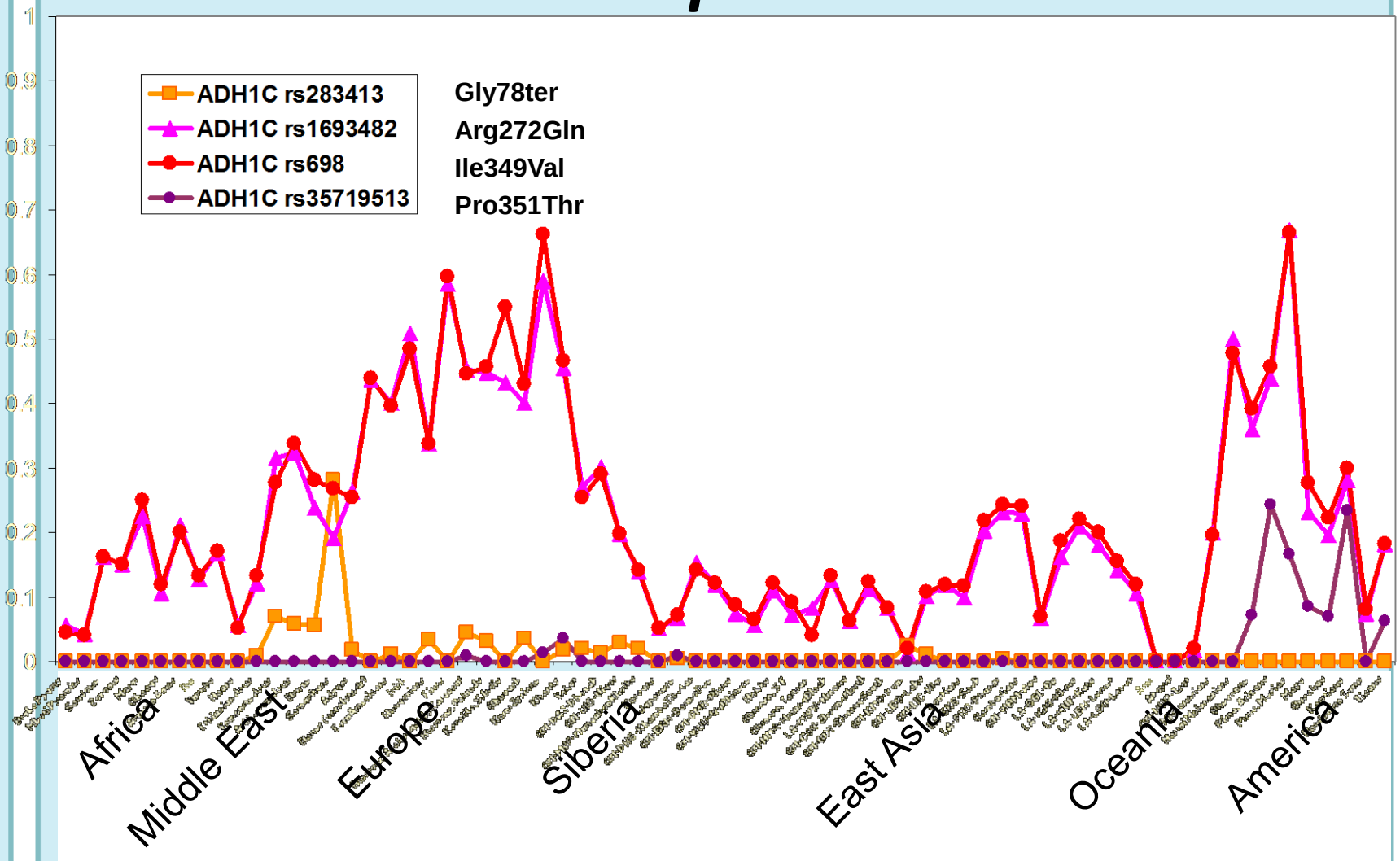
Technology Transition Workshop

NIJ National Institute of Justice

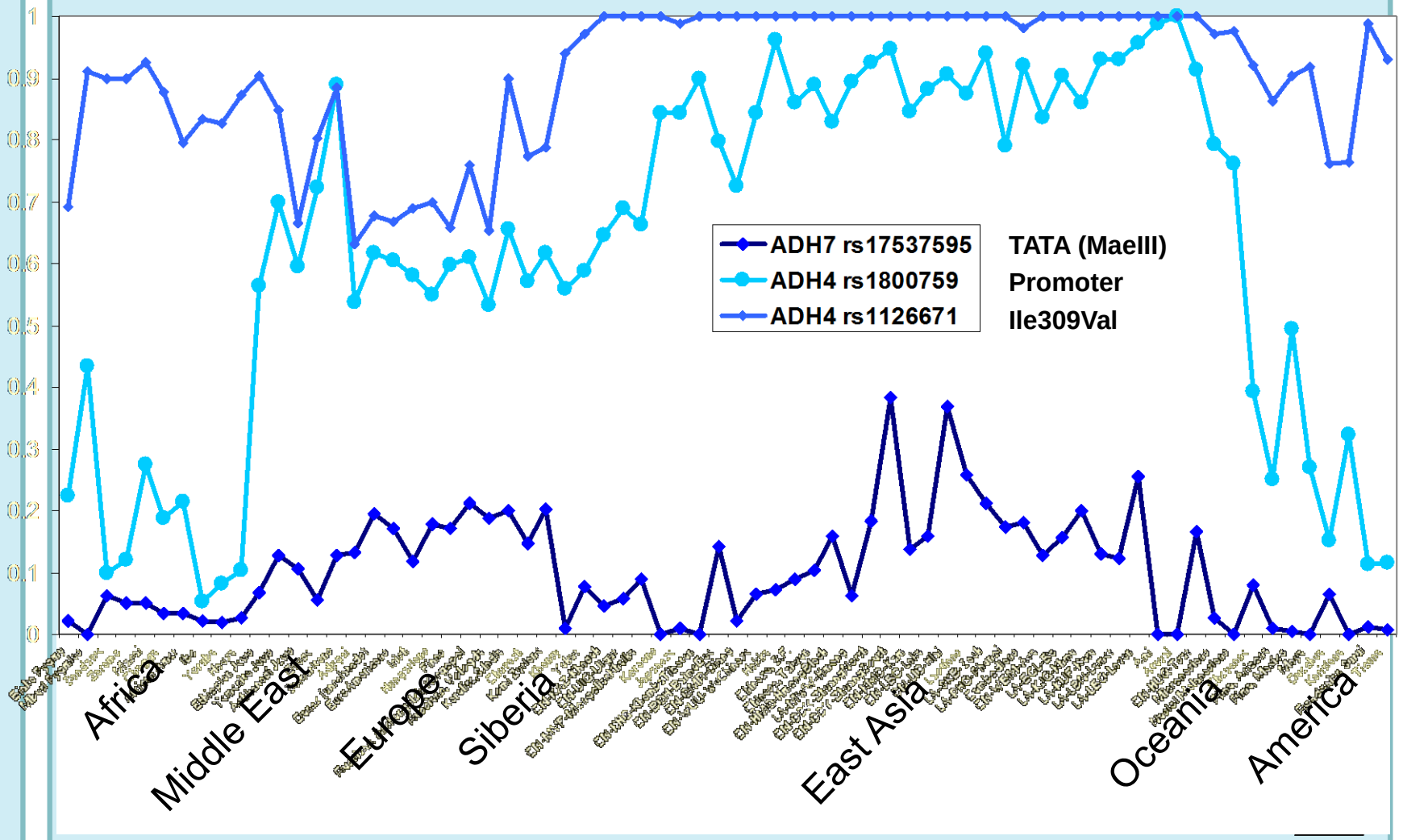
ADH1B Variant Frequencies



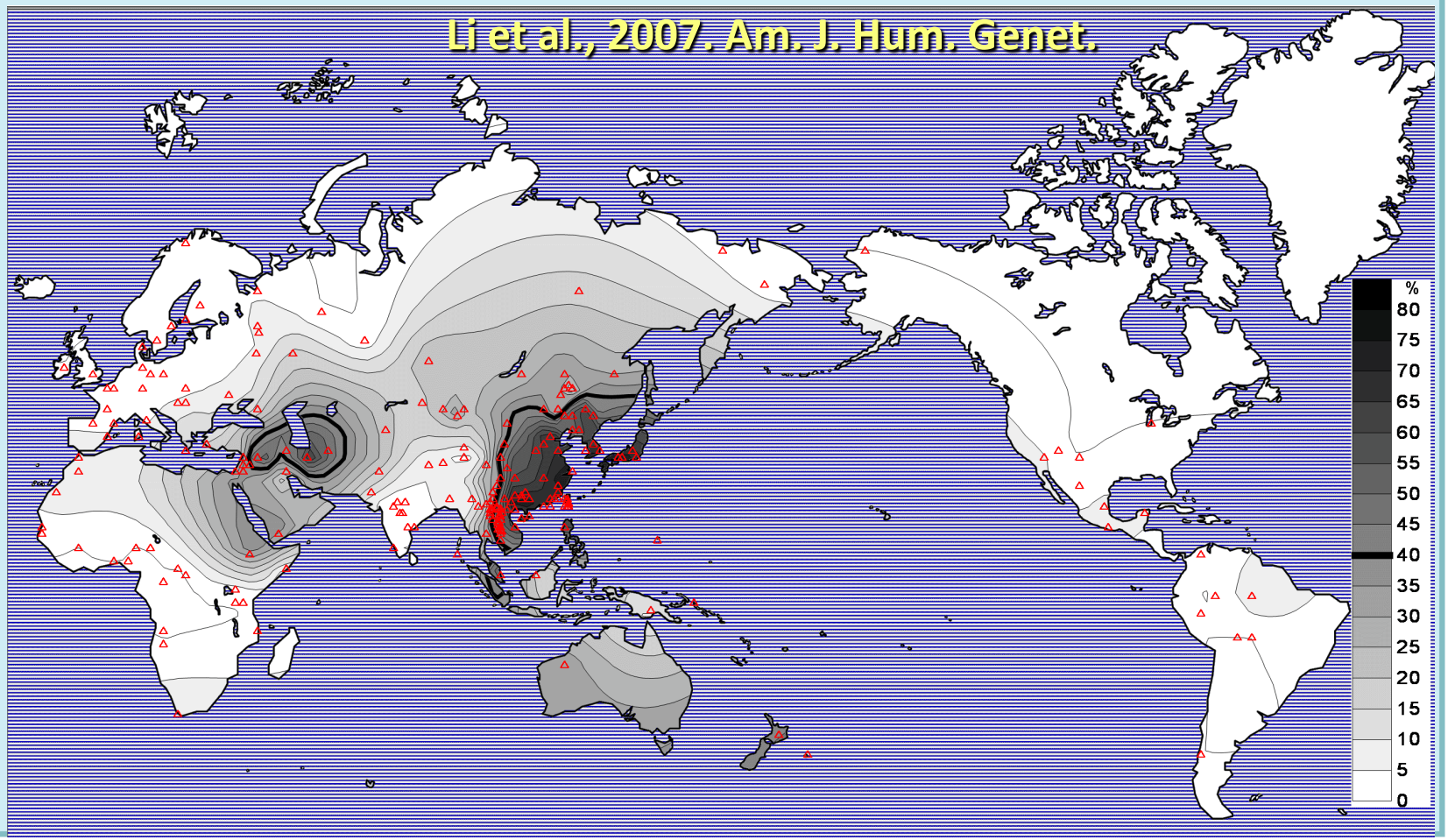
ADH1C Variant Frequencies



ADH4 & ADH7 Variant Frequencies

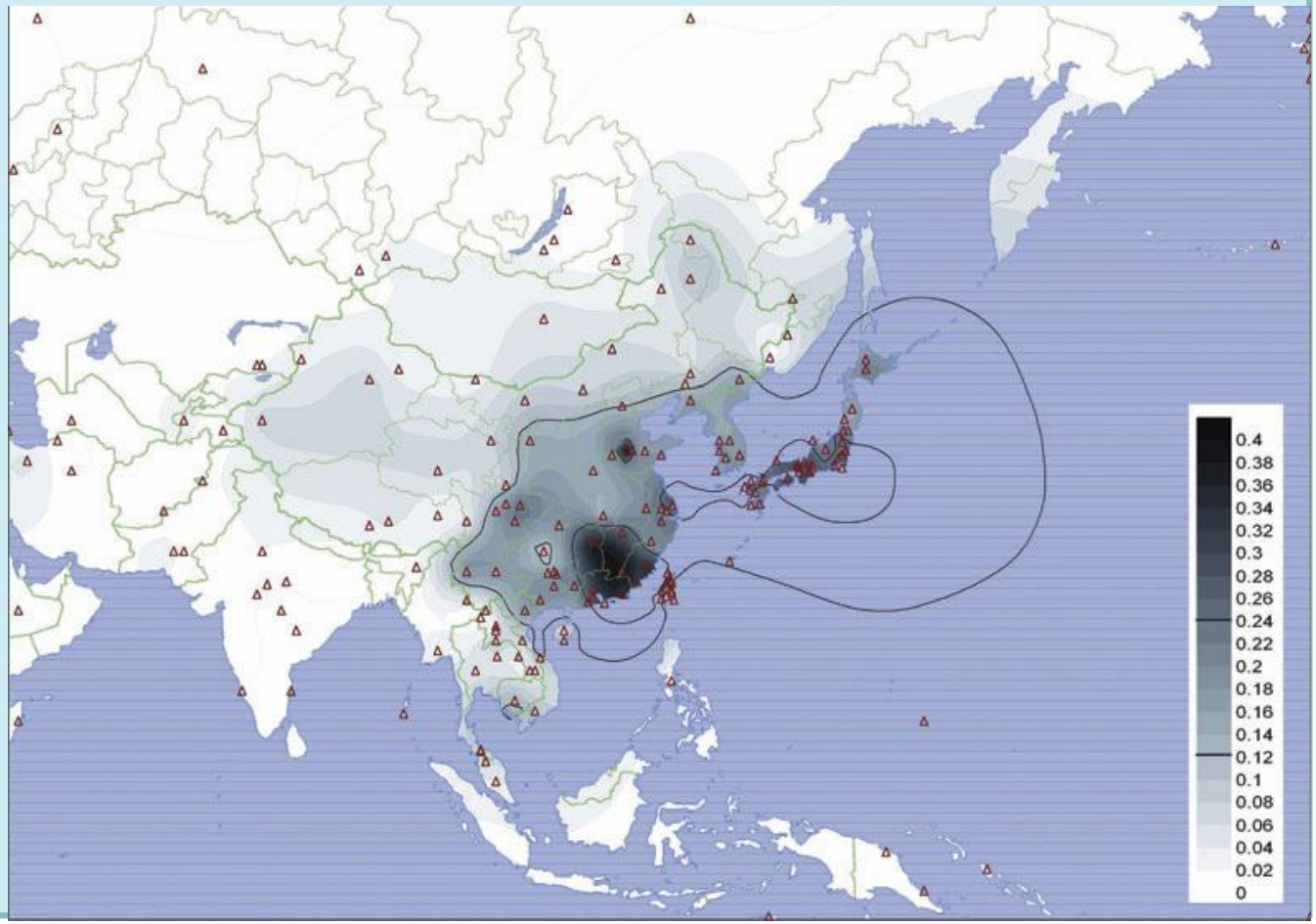


*Allele Frequency of ADH1B*47His, Updated, March 2008*



*Distribution of the ALDH2*504Lys Allele*

The *ALDH2*504Lys* allele is a "dominant negative": it disrupts the heteroduplex enzyme greatly reducing activity in heterozygotes and has no enzyme activity when homozygous.



Question: Can DNA Tell What Someone Looks Like?

Some SNPs are clearly related to appearance. In some cases the basic biology is understood. In most cases the biology is known to be relevant but the exact ways in which the different alleles alter the appearance are not fully understood.

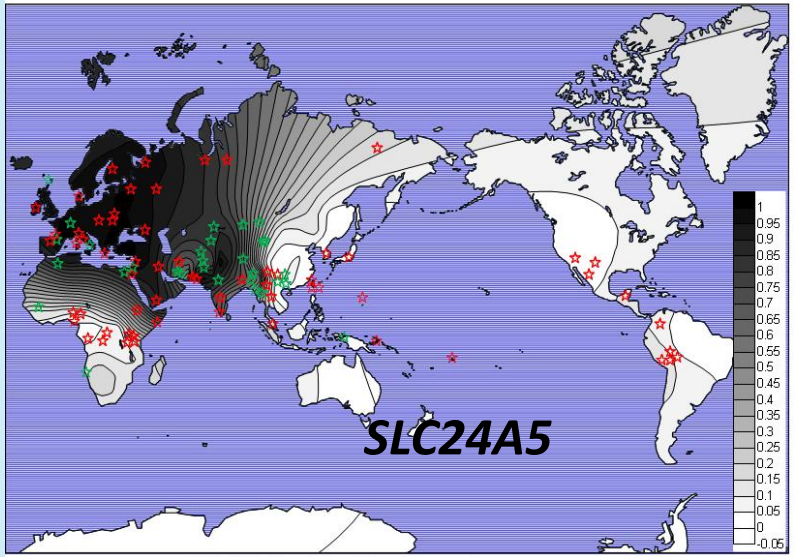
Technology Transition Workshop

Four "Skin Color" and "Eye Color " Genes

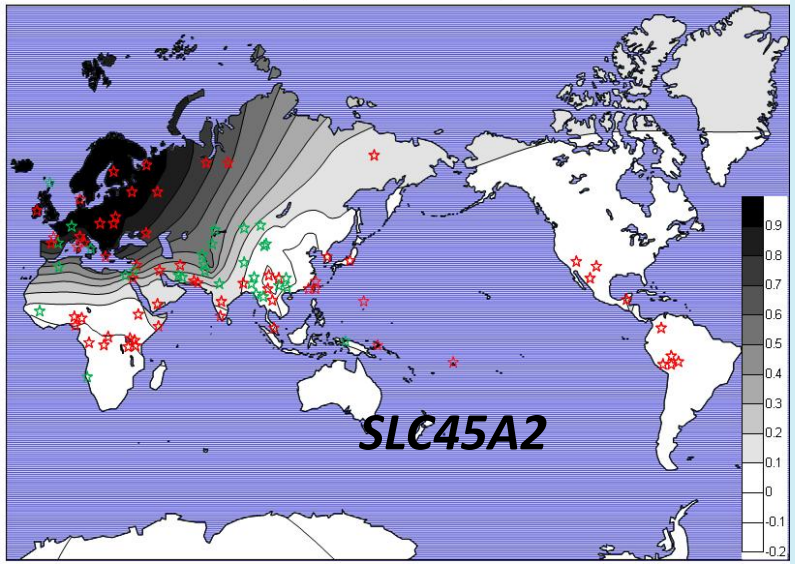
Golden (SLC24A5), MATP (SLC45A2), OCA2, and MC1R are genes that are expressed in skin and affect the processing of melanin and melanin granules. In other organisms individuals homozygous for non-functional forms of SLC24A5 have lighter pigmentation, but in humans the variant form gives only slightly lighter pigmentation.

Technology Transition Workshop NIJ National Institute of Justice

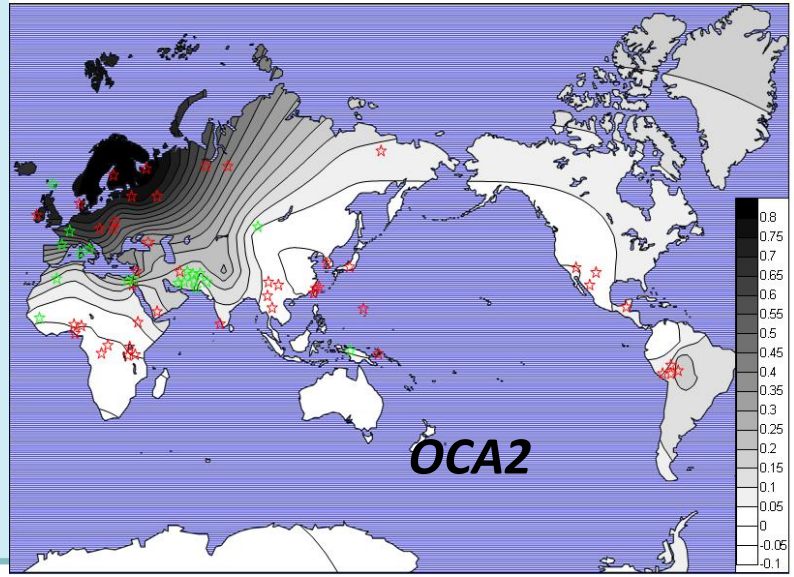
SLC24A5 Light Skin Allele Frequencies



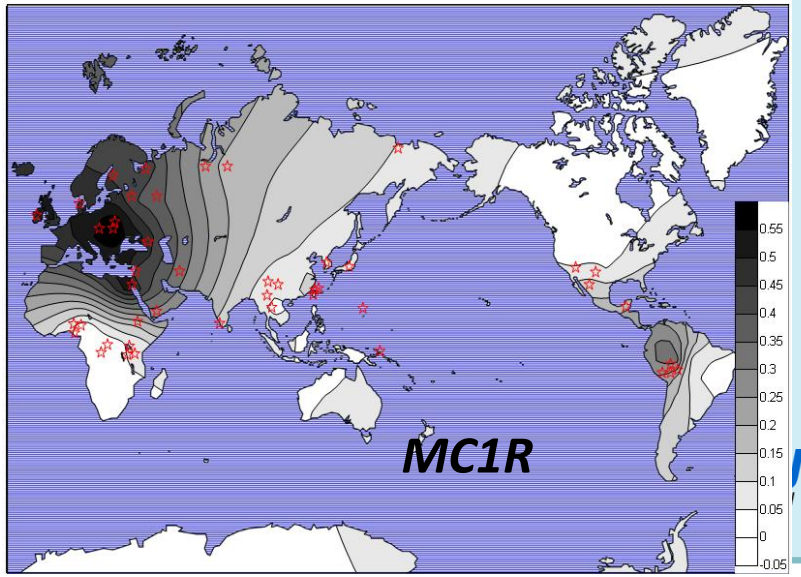
SLC45A2 Light Skin Allele Frequencies



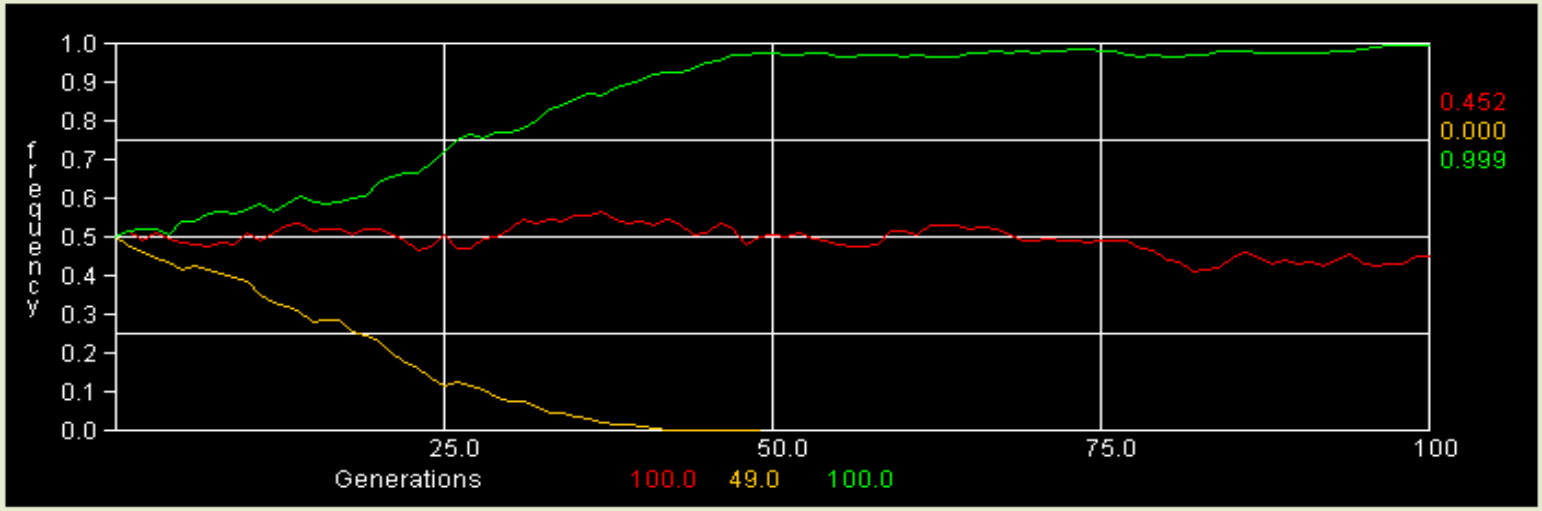
OCA2 Blue Eye Haplotype Frequencies



MC1R "European" Haplotype Frequencies



PopGen Simulation: Selection for "A" Allele in Green



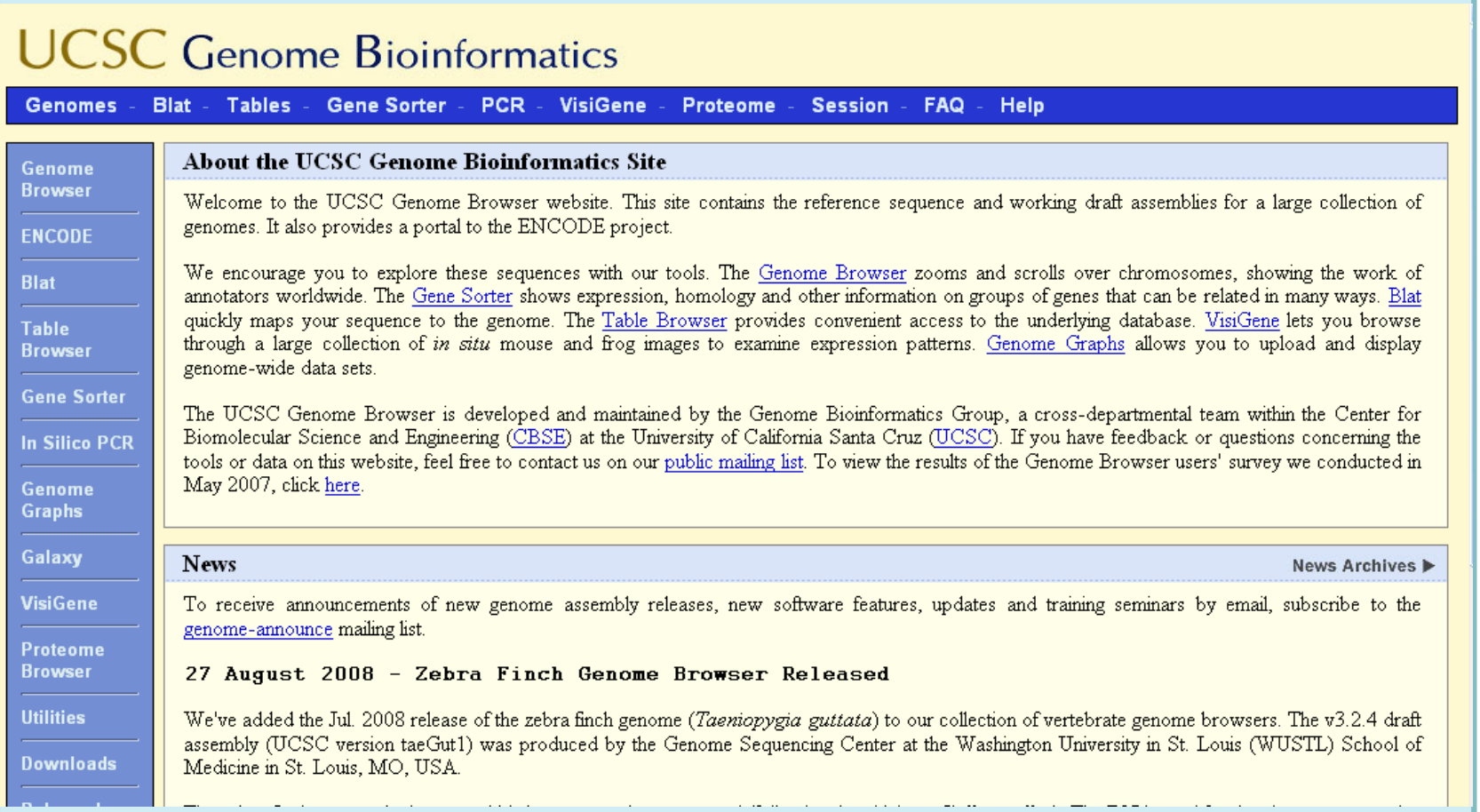
Intial allele frequency: 0.5
Population size: 500
Generations: 100
Continue last simulation
Submit
Clear Simulation
Printable View

Genotypes fitness
W_{AA}: 1
W_{Aa}: 0.9
W_{aa}: 0.9

Technology Transition Workshop

NIJ National Institute of Justice

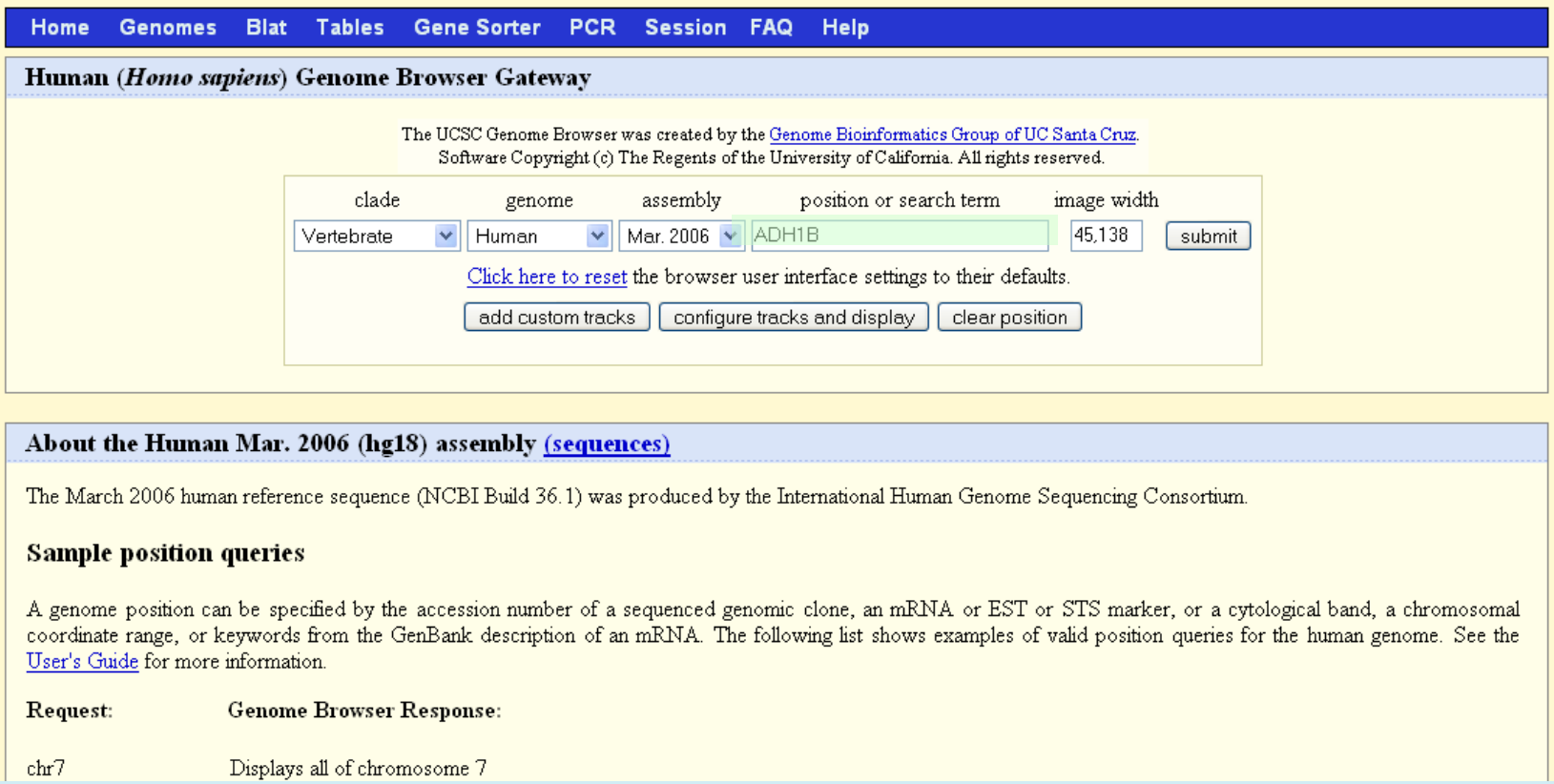
UCSC Genome Browser http:genome.ucsc.edu



Technology Transition Workshop

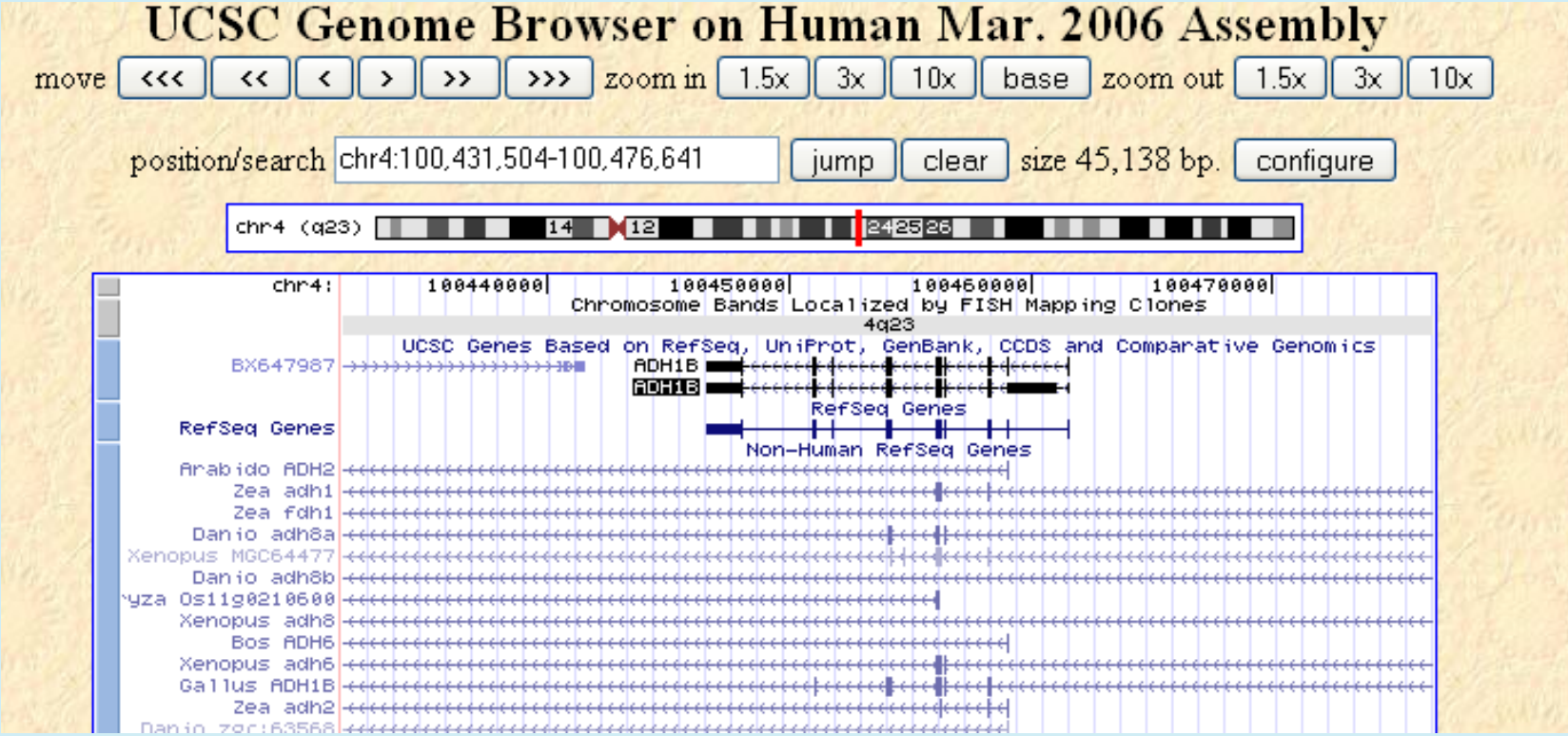
NIJ National Institute of Justice

A Genome Browser Query



Technology Transition Workshop

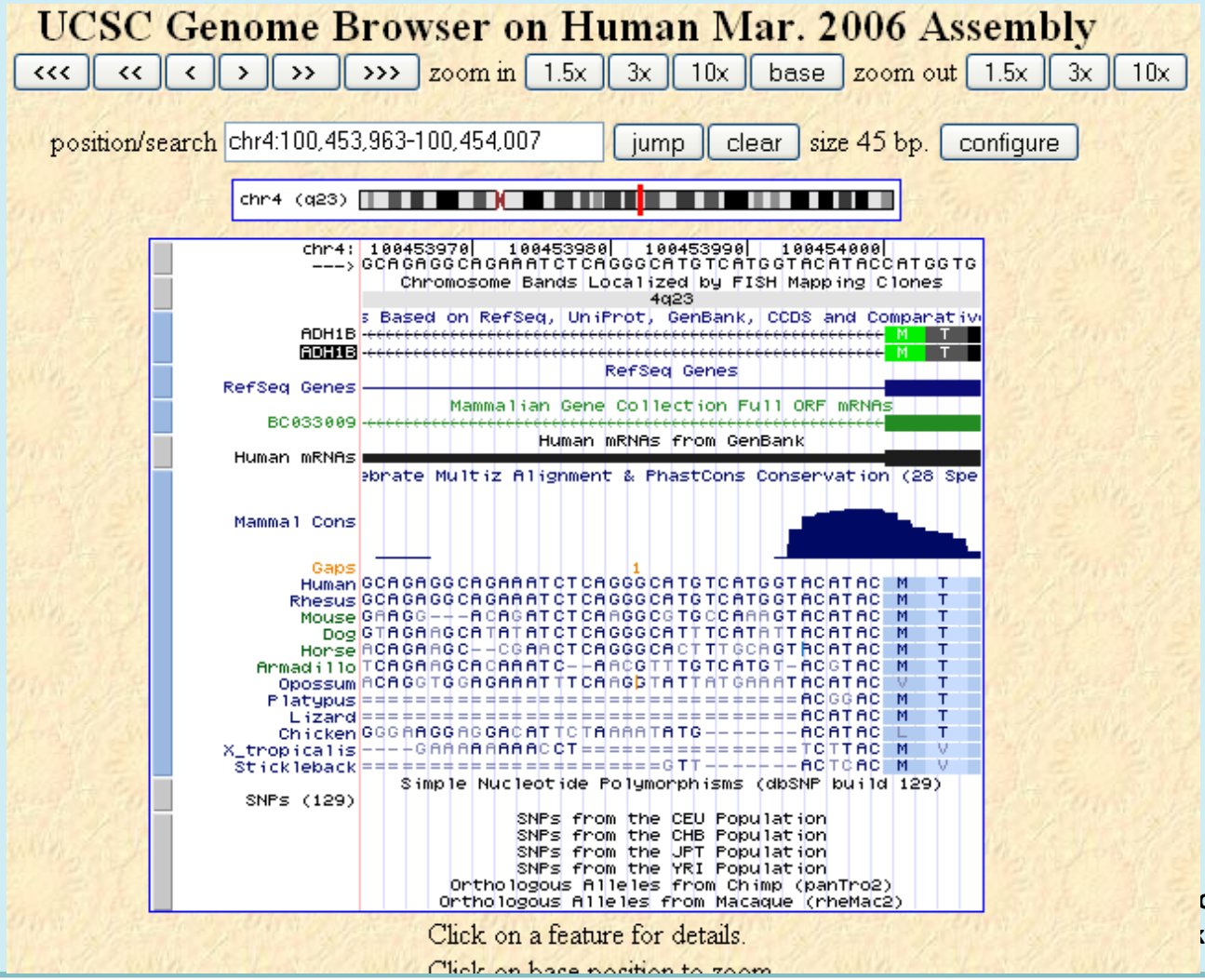
Part of a Genome Browser Result



Technology Transition Workshop

NIJ National Institute of Justice

Sequence at an Intron-Exon Junction



Questions?

Technology
Transition Workshop

NIJ National Institute of Justice

Basic Probability Rules

- **The probability of occurrence of any one of mutually exclusive events is the sum of the probabilities of the individual events**

- **The probability of occurrence of all of more than one independent events is the product of the probabilities of those events**

Technology Transition Workshop

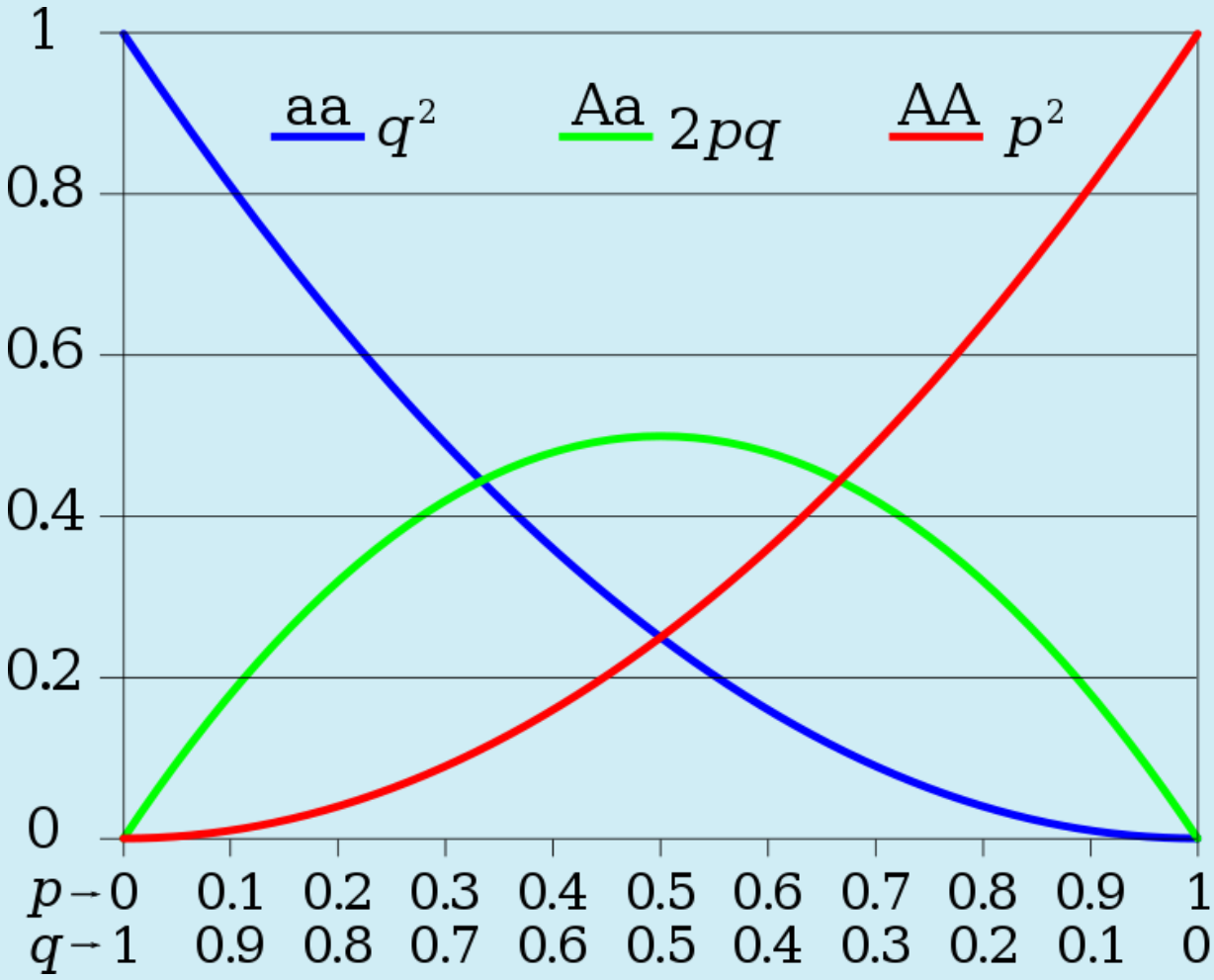
Basic Probability Rules

- **What is the probability of rolling snake-eyes with a pair of dice?**
  - **$(1/6) \cdot (1/6) = 1/36$**
- **What is the probability of rolling either snake-eyes or boxcars with a pair of dice?**
  - **$(1/6) \cdot (1/6) + (1/6) \cdot (1/6) = 2/36 = 1/18$**

Technology Transition Workshop

NIJ National Institute of Justice

Hardy-Weinberg Ratios



H-W ratios are robust to the assumptions and can be used to relate allele frequencies in a population to the genotype frequencies in that population.

http://en.wikipedia.org/wiki/File:Hardy-Weinberg.svg

Technology Transition Workshop NIJ National Institute of Justice

At what allele frequency will a SNP most likely show a different genotype for two unrelated individuals?

The probability of different genotypes at one SNP between two unrelated individuals in the population:

$$p^2(1-p^2)+2pq(1-2pq)+q^2(1-q^2)$$

$$=1-(p^4+(2pq)^2+q^4)$$

This is the discriminatory power of a marker.

Technology Transition Workshop

NIJ National Institute of Justice

What Rules of Probability Did We Just Use?

1. **The probability of both of two independent events is the product of the probabilities of each of the events.**

2. **The probability of just one of two (or more) mutually exclusive events is the sum of the probabilities of the individual events.**

Technology Transition Workshop NIJ National Institute of Justice

The Probability of Two Individuals Having the Same Genotype at One SNP:

$$p^2(p^2) + 2pq(2pq) + q^2(q^2)$$

$$= p^4 + (2pq)^2 + q^4$$

This is the average match probability of a marker. At what allele frequency is this smallest?

Technology Transition Workshop

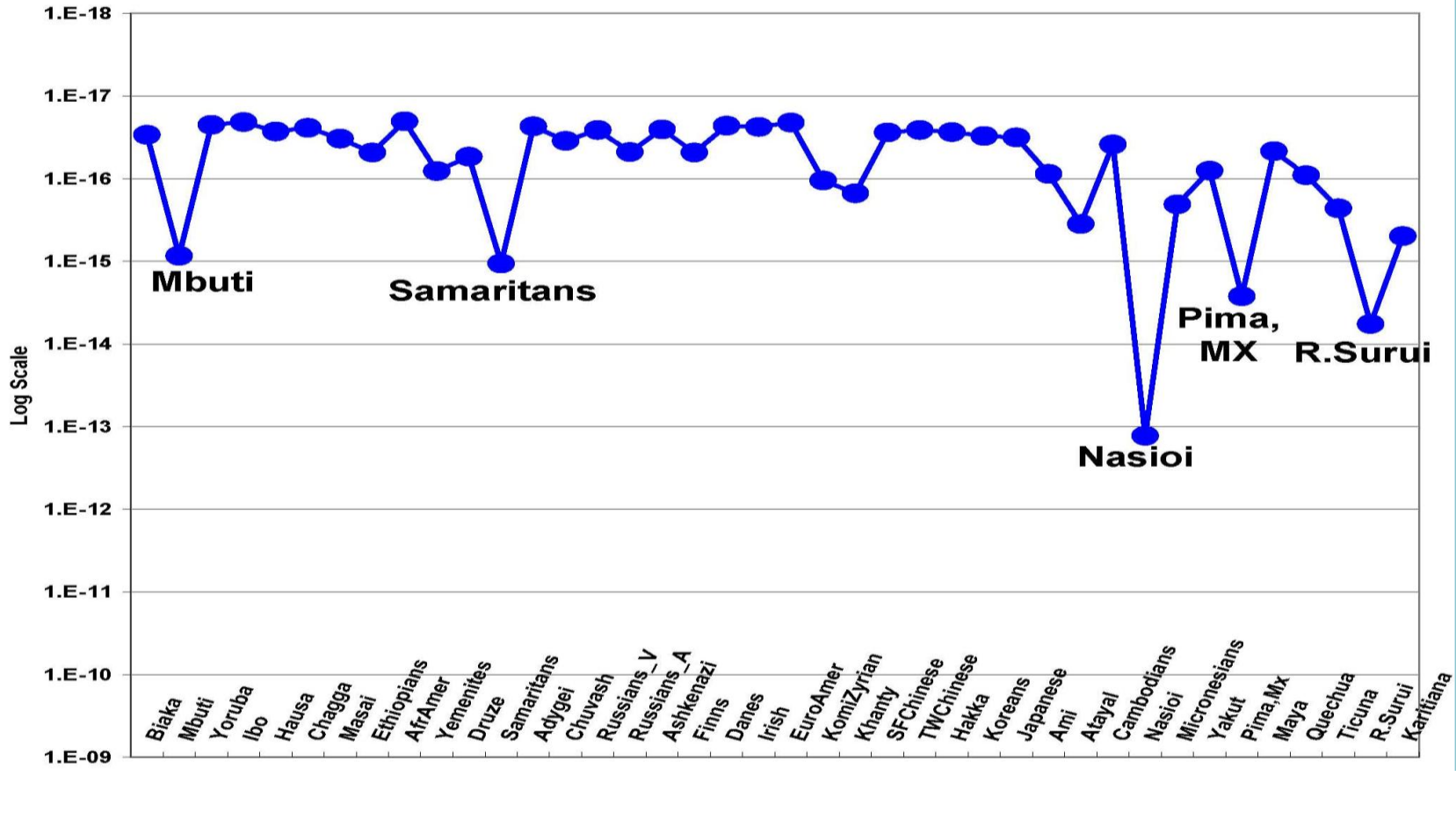
SNPs for Individual Identification: IISNPs

- **Find SNPs with the highest heterozygosity to minimize the probability that two unrelated individuals have the same genotype**
- **Find SNPs that have the smallest allele frequency variation so that the probabilities have minimal dependencies on ethnicity**

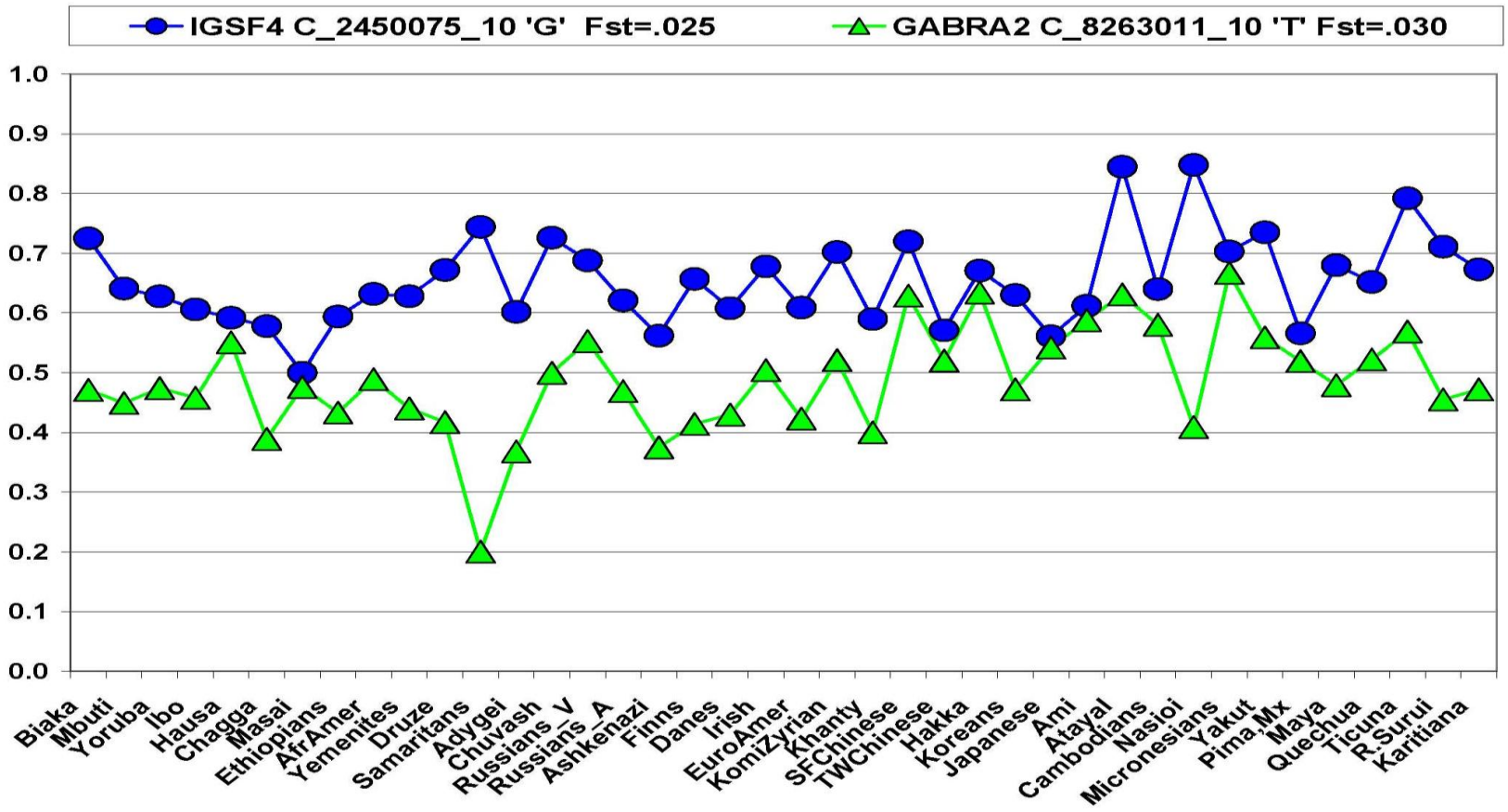
Technology Transition Workshop NIJ National Institute of Justice

Match Probabilities

40 Best SNPs with F_{st} <0.06



Allele Frequencies for 2 SNPs with Very Low F_{st}



Technology Transition Workshop

NIJ National Institute of Justice

The Population Genetics of Ancestry/Ethnicity Inference

- **Empirically allele frequencies vary around the world**
- **Statistically, it is well accepted that the population in which the observed genotype of the unknown is most frequent is the population the unknown is most likely to come from**
- **The concept of relative likelihood must be used to interpret the data**

Technology Transition Workshop NIJ National Institute of Justice

Likelihood is Related to Probability

- **One considers the Probability of an outcome given a particular hypothesis: Pr(event|hypothesis)**
- **One considers the Likelihood of a hypothesis given a set of data, i.e., an outcome**
- **Likelihood is considered proportional to the Probability of the event given the hypothesis**

$$L(\text{hypothesis \#1} \mid \text{dataset \#1}) \propto Pr(\text{dataset \#1} \mid \text{hypothesis \#1})$$

$$L(\text{hypothesis \#2} \mid \text{dataset \#1}) \propto Pr(\text{dataset \#1} \mid \text{hypothesis \#2})$$

Technology Transition Workshop

NIJ National Institute of Justice

Because of the Unknown Constant of Proportionality, One Uses Relative Likelihoods

- **The relative likelihood of hypothesis #1 relative to hypothesis #2 is the ratio of the probabilities given those hypotheses**
- **L (hypothesis #1)/L (hypothesis #2) =**

$$\frac{\Pr(\text{dataset \#1} \mid \text{hypothesis \#1})}{\Pr(\text{dataset \#1} \mid \text{hypothesis \#2})}$$

Technology Transition Workshop NIJ National Institute of Justice

"Scientific" Issues in Forensic DNA Use

- **Should labs be accredited and, if so, how?**
- **How can DNA be used in missing person cases?**
- **When does a "close" match implicate a relative?**
- **Forensic use of DNA is not fool proof.**

While the underlying science may be rigorous, its implementation may not always be so rigorous.

Issues include chain of custody of samples, stringent laboratory procedures, careful training of technicians.

Technology Transition Workshop NIJ National Institute of Justice

SNP Database Resource

- **SNP allele frequencies are essential for forensic applications of SNP panels**
- **ALFRED is making those data accessible for for the scientific and forensic communities**

ALFRED: the ALlele FREquency Database
http://alfred.med.yale.edu

Technology Transition Workshop NIJ National Institute of Justice

The ALlele FREquency Database

ALFRED is a resource of gene frequency data on human populations
supported by the U. S. National Science Foundation.

▪ Home ▪ Ethics ▪ Search ▪ Summaries ▪ Documentation ▪ Register ▪ Contact Us

ALFRED is designed to make allele frequency data on human population samples readily available for use by the scientific and educational communities.

ALFRED Wiki

Tour ALFRED

ALFRED FAQ

Data Downloads

Register

ALFRED flyer

Contact us

ALFRED now has data on **663,412** polymorphisms, **710** populations and **35,227,290** frequency tables (one population typed for one site).

Quick Keyword Search:

Help

If you are not sure about the exact chromosome and do not know the UID, type in the gene symbol, SNP name or rsnumber to search for a SNP.

Go

Search Type:	Search Tables:	
Any part of ◉	Loci ☑ Sites ☑	Go
Begins with ○	Populations ☑	

Suggestions or comments Kidd Lab Home

New to ALFRED? Click Here

Latest Newsletter? Register

ALFRED News

NIJ National Institute of Justice

SNP Sets in ALFRED: Both IISNPs and AISNPs

SNP Sets

Set	Citation
Interim Panel of 40 IISNPs	- Pakstis AJ, Speed WC, Kidd JR, Kidd KK. "Candidate SNPs for a Universal Individual Identification Panel". *Human Genetics* **121**:305-317. (2007) Online citation.
45 Unlinked IISNPs	- Pakstis AJ, Speed WC, Fang R, Hyland FCL, Furtado MR, Kidd JR, Kidd KK. "SNPs for a universal individual identification panel". *Human Genetics* **127**:315-24. (2010) Online citation.
Final List of 86 IISNPs	- Pakstis AJ, Speed WC, Fang R, Hyland FCL, Furtado MR, Kidd JR, Kidd KK. "SNPs for a universal individual identification panel". *Human Genetics* **127**:315-24. (2010) Online citation.
SNPforID 52-plex	- Sanchez JJ, Phillips C, Borsting C, Balogh K, Bogus M, Fondevila M, Harrison CD, Musgrave-Brown E, Salas A, Syndercombe-Court D, Schneider PM, Carracedo A, Morling N. "A multiplex assay with 52 single nucleotide polymorphisms for human identification". *Electrophoresis.* **27**:1713-1724. (2006) Online citation.
SNPforID 34-plex	- Phillips C, Salas A, Sánchez JJ, Fondevila M, Gómez-Tato A, Álvarez-Dios J, Calaza M, Casares de Cal M , Ballard D, Lareu MV, Carracedo A - The SNPforID Consortium "Inferring ancestral origin using a single multiplex assay of ancestry-informative marker SNPs". *Forensic Science International: Genetics* **1**:273-280. (2007) Online citation.
CODIS Set	- Budowle B, Moretti TR, Niezgoda SJ, Brown BL "CODIS and PCR-based short tandem repeat loci: law enforcement tools, in: Proceedings of the Second European Symposium on Human Identification ". *Proceedings of the Second European Symposium on Human Identification,Promega Corporation, Madison, WI,* 73-88. (1998) Online citation.

SNP Sets in ALFRED: IISNPs

Reference-

Pakstis AJ, Speed WC, Fang R, Hyland FCL, Furtado MR, Kidd JR, Kidd KK. "SNPs for a universal individual identification panel". *Human Genetics* **127**:315-24. (2010) Online citation.

Shows some linkage	Unlinked	No Info

Sort by : Locus | FST | AVG_HET | #Pop

Sorted by Locus

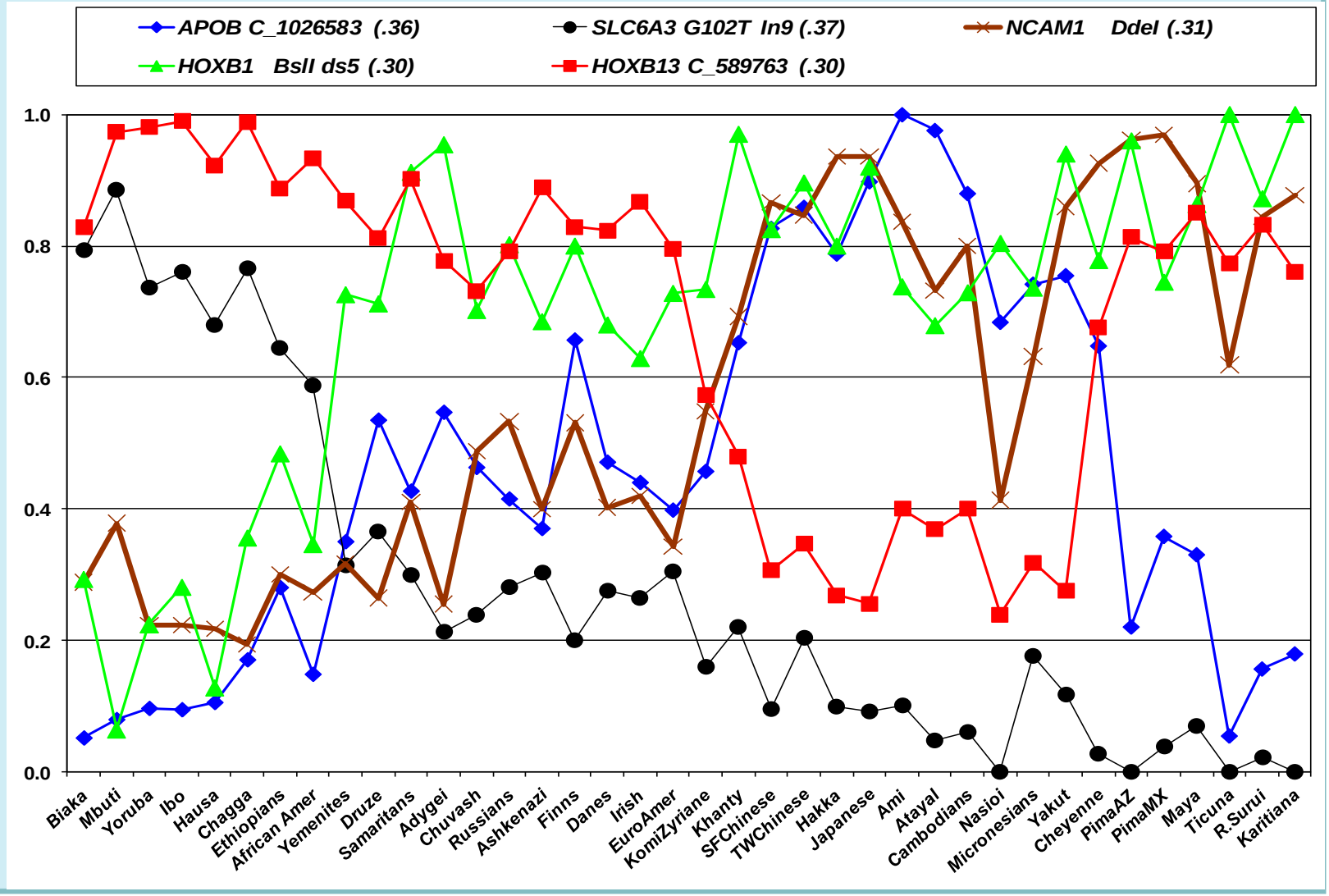
Locus	Site	dbSNP rs#	Fst	Avg Het	# Pops
A2BP1	rs2342747	rs2342747	0.048	0.423	83 GoogleMap
ABCG1	rs221956	rs221956	0.034	0.457	84 GoogleMap
ADAMTS2	C 3153696 10	rs338882	0.077	0.459	82 GoogleMap
ATP13A4	C 25749280 10	rs6444724	0.049	0.467	83 GoogleMap

Consider Ancestry Inference SNPs

- **Considerable allele frequency differences exist among the different populations/regions around the world**
- **A reference database of those allele frequencies is essential**
- **The problem is identifying a relatively small number of SNPs with considerable information, i.e., large amounts of allele frequency variation around the world**
- **This is the opposite of IISNPs**

Technology Transition Workshop NIJ National Institute of Justice

Highly Variable Diallelic Single Nucleotide Polymorphisms



Multiple Panels of AISNPs Exist

- **No single panel is necessarily going to be perfect or optimal for all questions of ancestry**
- **Most published panels have been tested on only a few populations and their broad applicability is uncertain**
- **One panel has been tested on a large number of populations: our study of the Seldin group's panel of 128 SNPs on 4871 individuals from 119 different population samples**

Kosoy et al., 2009. *Human Mutation* 30:69-78
Kidd et al., 2011. *Investigative Genetics* 2:1

Technology Transition Workshop NIJ National Institute of Justice

SNP Sets in ALFRED: AISNPs

Reference-

Phillips C, Salas A, Sánchez JJ, Fondevila M, Gómez-Tato A, Álvarez-Dios J, Calaza M, Casares de Cal M , Ballard D, Lareu MV, Carracedo A - The SNPforID Consortium "Inferring ancestral origin using a single multiplex assay of ancestry-informative marker SNPs". *Forensic Science International: Genetics* **1**:273-280. (2007) Online citation.

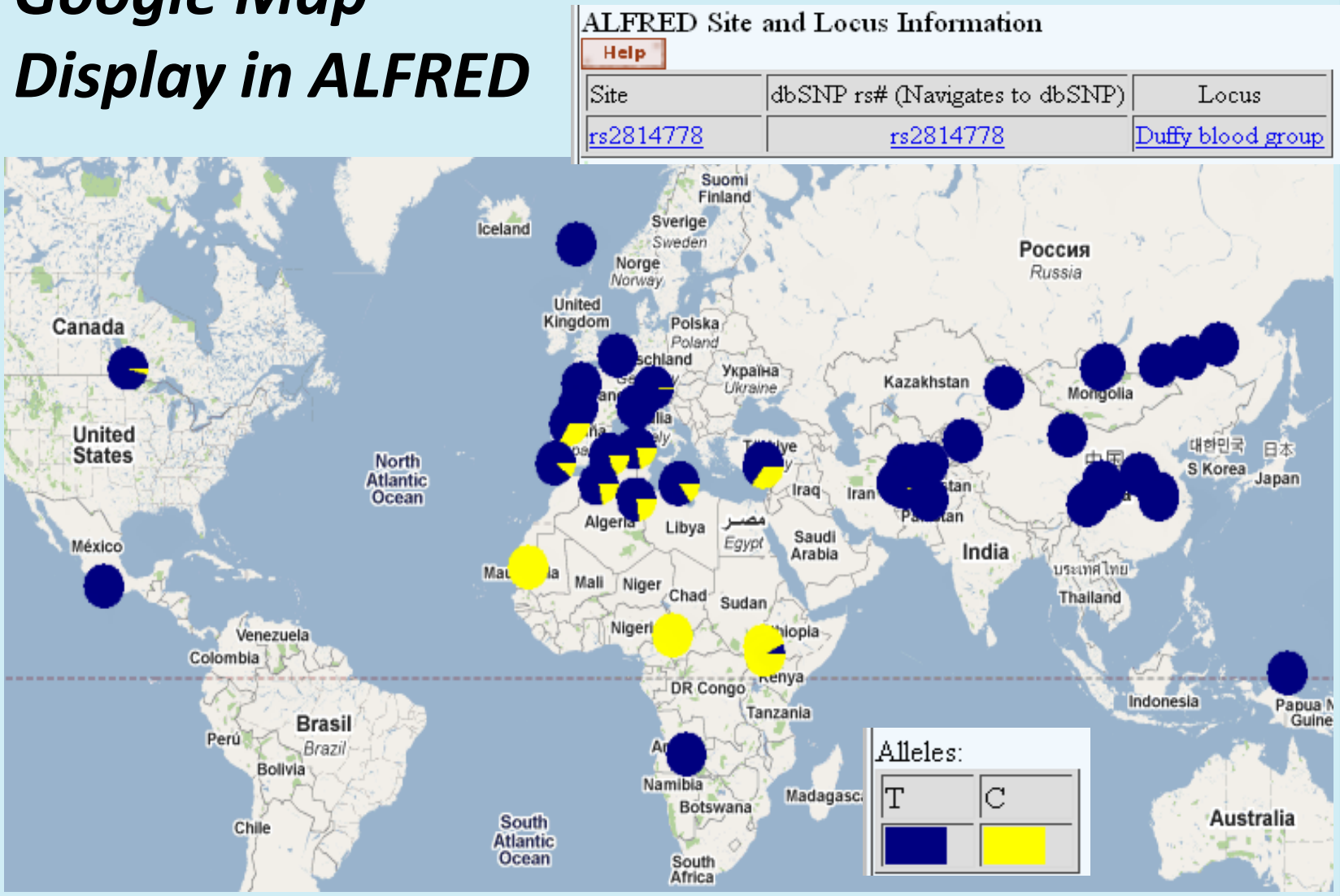
Shows some linkage	Unlinked	No Info

Sort by : Locus | FST | AVG_HET | #Pop

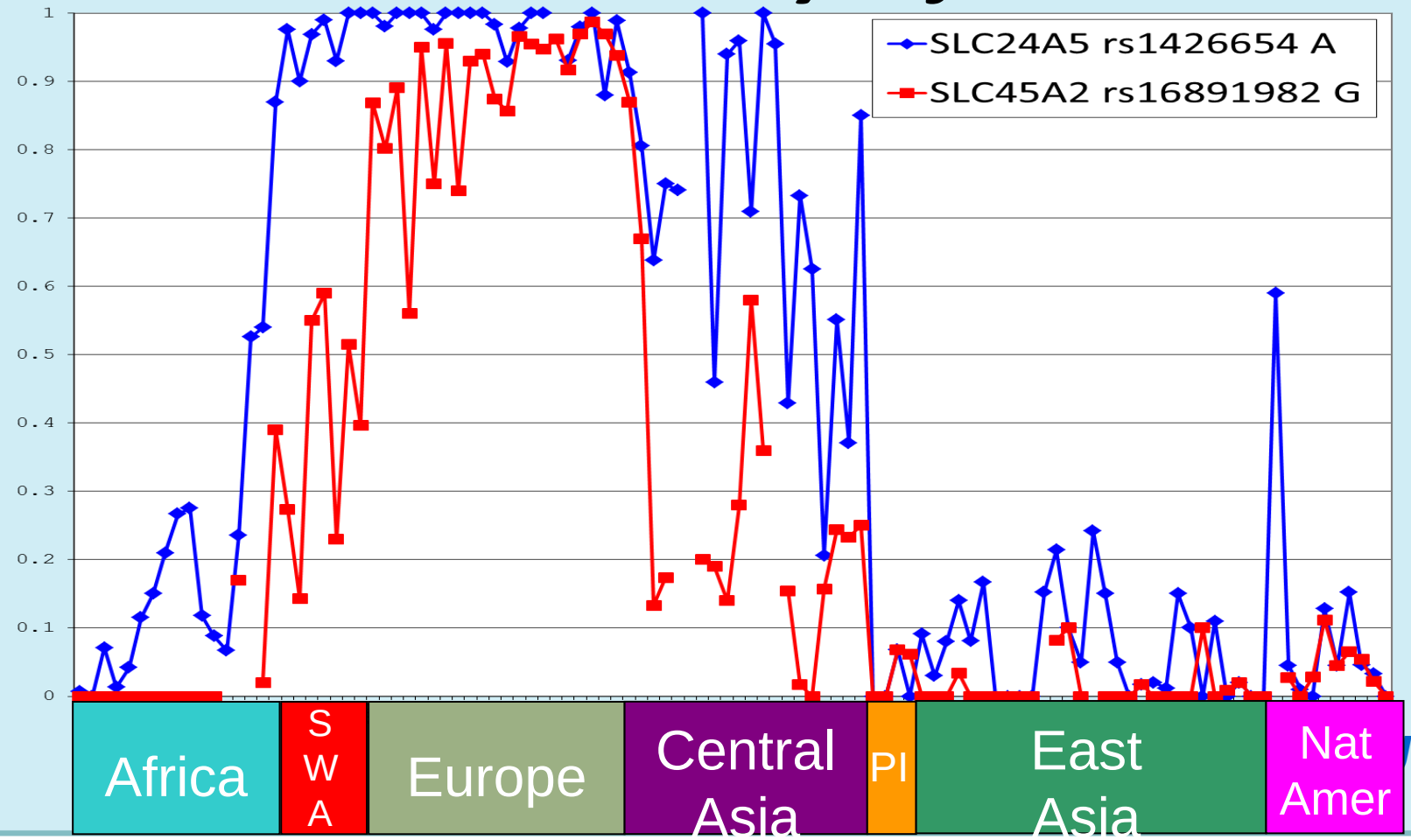
Sorted by FST

Locus	Site	dbSNP rs#	Fst	Avg Het	# Pops
WT1	rs5030240	rs5030240	0	0	0 GoogleMap
TLR1	rs4540055	rs4540055	0	0	0 GoogleMap
DARC	rs2814778	rs2814778	0.798	0.061	88 GoogleMap
SLC24A5	Ala111Thr	rs1426654	0.767	0.116	90 GoogleMap

Google Map Display in ALFRED



Phenotype Informative SNPs Can Also Be Ancestry Informative



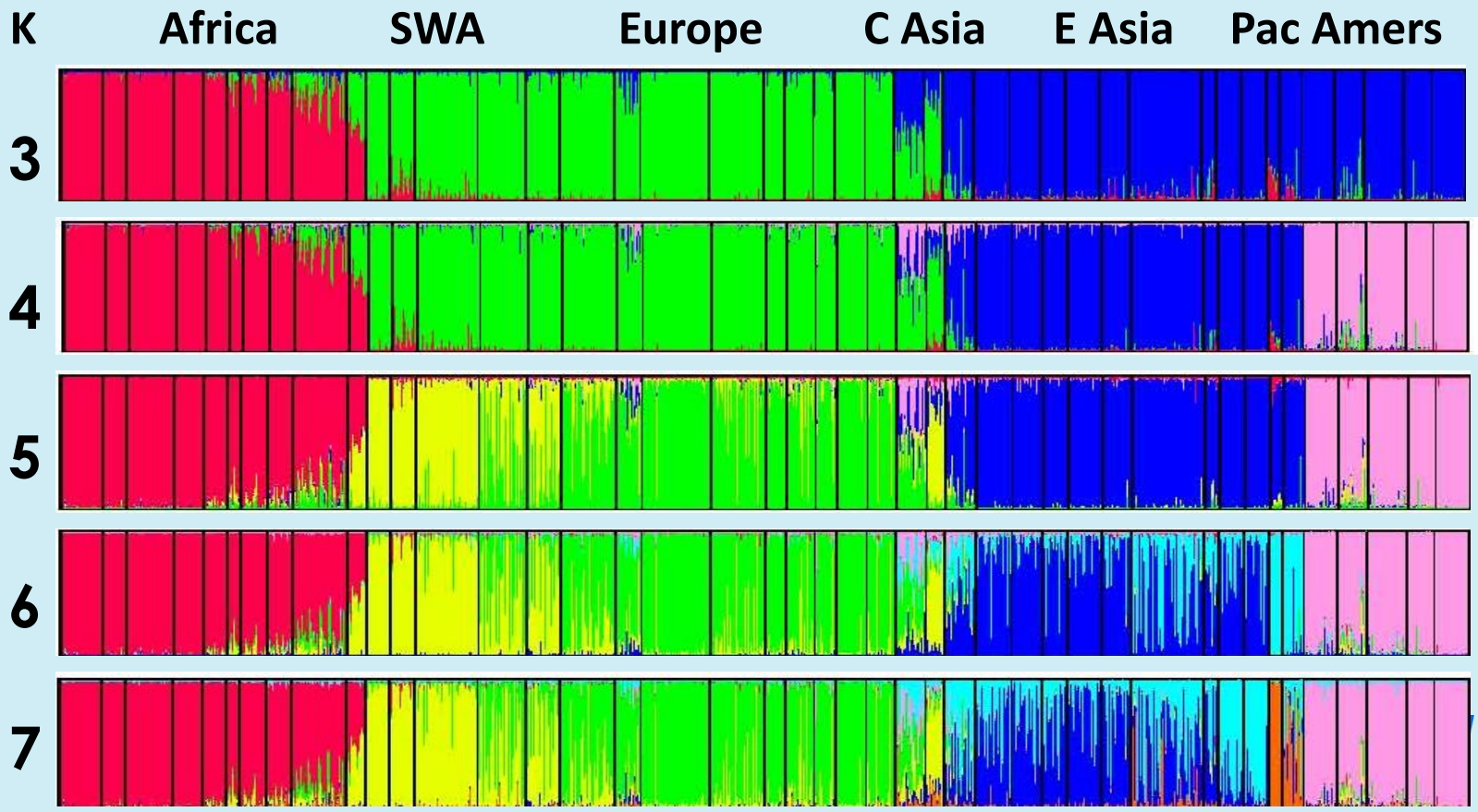
Our Ongoing Direction to Identify and Characterize Additional AISNPs

Work to obtain data for multiple potential AISNPs with data uniformly available on a large number diverse populations.

Include Phenotype Informative SNPs whenever possible.

Technology Transition Workshop

STRUCTURE Analyses of 2278 Individuals from 43 Populations Using 39 Provisional AISNPs



Interpreting STRUCTURE Analyses

By a search varying the parameters (allele frequencies), STRUCTURE estimates the allele frequencies of the K assumed populations that have the best "fit" to the genotypes of all individuals by placing them in one or another of the K populations. STRUCTURE does not consider prior knowledge of the origins of the individuals, but their placement into the K populations can be colored and origin displayed by grouping by known origin.

The optimal number of underlying populations (K) and pattern of assignment are determined by likelihood. The results are a function of the numbers and similarities of the individuals in the dataset analyzed.

Forensic **R**esource **O**n **G**enetics

knowledge base

A pilot version of our new web site/database is online. To date it has only a pilot (pre-beta) version of a function to calculate probabilities of genotypes by population for certain IISNP and AISNP panels.

http://frog.med.yale.edu

Technology Transition Workshop

NIJ National Institute of Justice

Data Input and Output Screens for IISNPs

Data input for a Mexican Pima individual

rsnumber	chromosome	chr_position	AA	AB	BB	Unknown
rs7520386	1	14027989	○ AA	◉ AG	○ GG	○ NN
rs560681	1	159053294	◉ AA	○ AG	○ GG	○ NN
rs1294331	1	231515036	○ AA	◉ AG	○ GG	○ NN
rs3780962	10	17233352	○ CC	◉ CT	○ TT	○ NN
rs740598	10	118496889	○ AA	◉ AG	○ GG	○ NN
rs1498553	11	5665604	○ CC	◉ CT	○ TT	○ NN
rs10488710	11	114712386	○ CC	◉ CG	○ GG	○ NN

Set all unselected to unknown

Compile

Results

Population	Match Probability
Pima_ Mexico	1.295e-17
Hakka	1.564e-18
Sandawe	7.608e-19
Jews_ Ethiopian	4.206e-19
Ibo	3.649e-19
Lao Loum	3.561e-19
Biaka	3.327e-19
Japanese	2.67e-19
Yoruba	2.523e-19

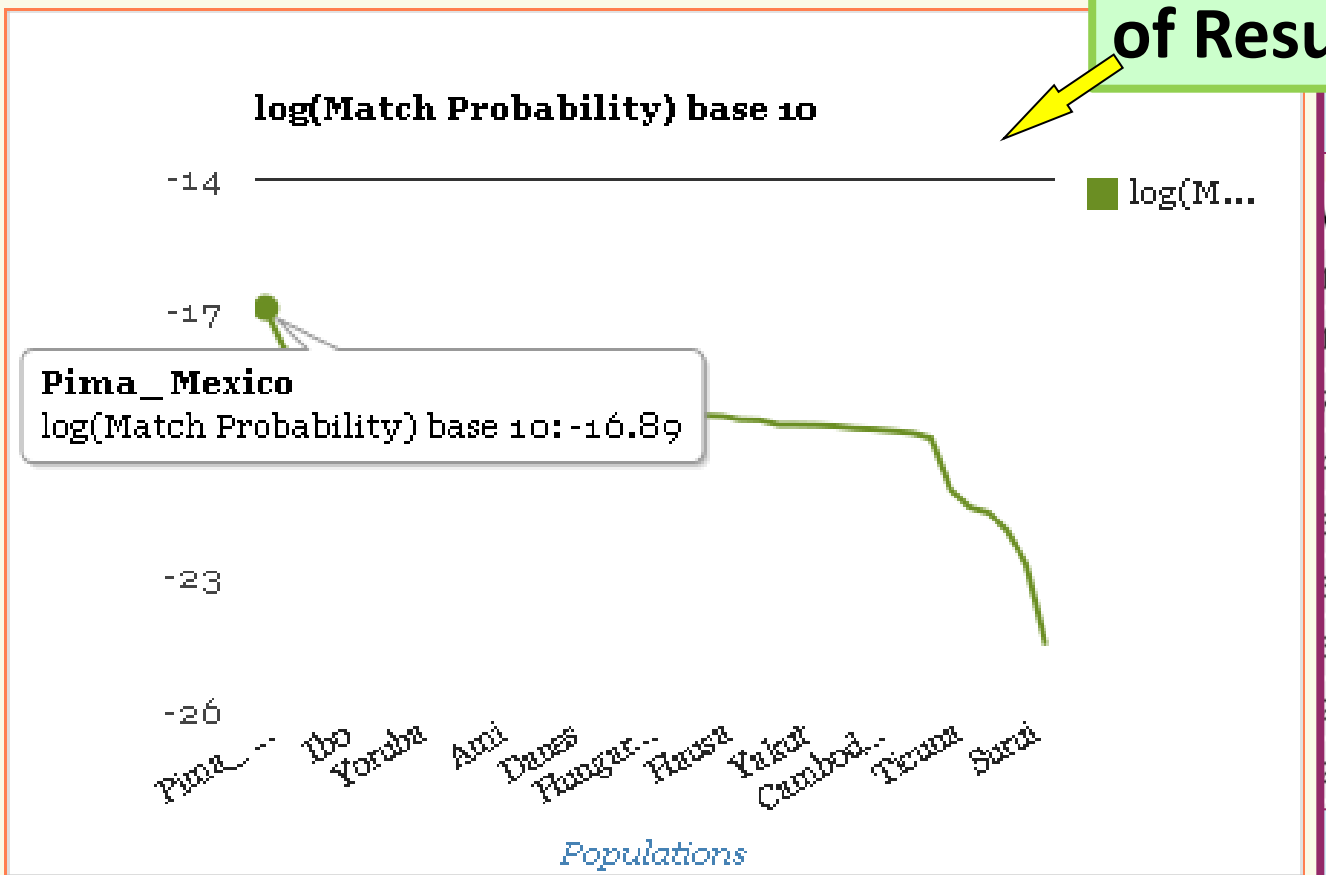
Technology Transition Workshop

NIJ National Institute of Justice

Data Input and Output Screens for IISNPs

Data input for a Mexican Pima individual

Graphic Version of Results



Match Probability
1.295e-17
1.564e-18
7.608e-19
4.206e-19
3.649e-19
3.561e-19
3.327e-19
2.67e-19
2.523e-19

Technology Transition Workshop

NIJ National Institute of Justice

Data Input and Output Screens for AISNPs

Data input for a Korean individual

rsnumber	chromosome	chr_position	AA	AB	BB	Unknown
rs3737576	1	101482151	◉ AA	○ AG	○ GG	○ NN
rs7554936	1	149389113	○ CC	○ CT	◉ TT	○ NN
rs2814778	1	157441307	○ CC	○ CT	◉ TT	○ NN
rs4918664	10	94911055	○ AA	○ AG	◉ GG	○ NN
rs174570	11	61353788	◉ CC	○ CT	○ TT	○ NN
rs1079597	11	112801496	○ AA	◉ AG	○ GG	○ NN
rs2238151	12	110696216	◉ CC	○ CT	○ TT	○ NN
rs7997709	13	33745737	○ CC	○ CT	◉ TT	○ NN
rs1572018	13	40613282	○ AA	○ AG	◉ GG	○ NN

Set all unselected to unknown

Compile

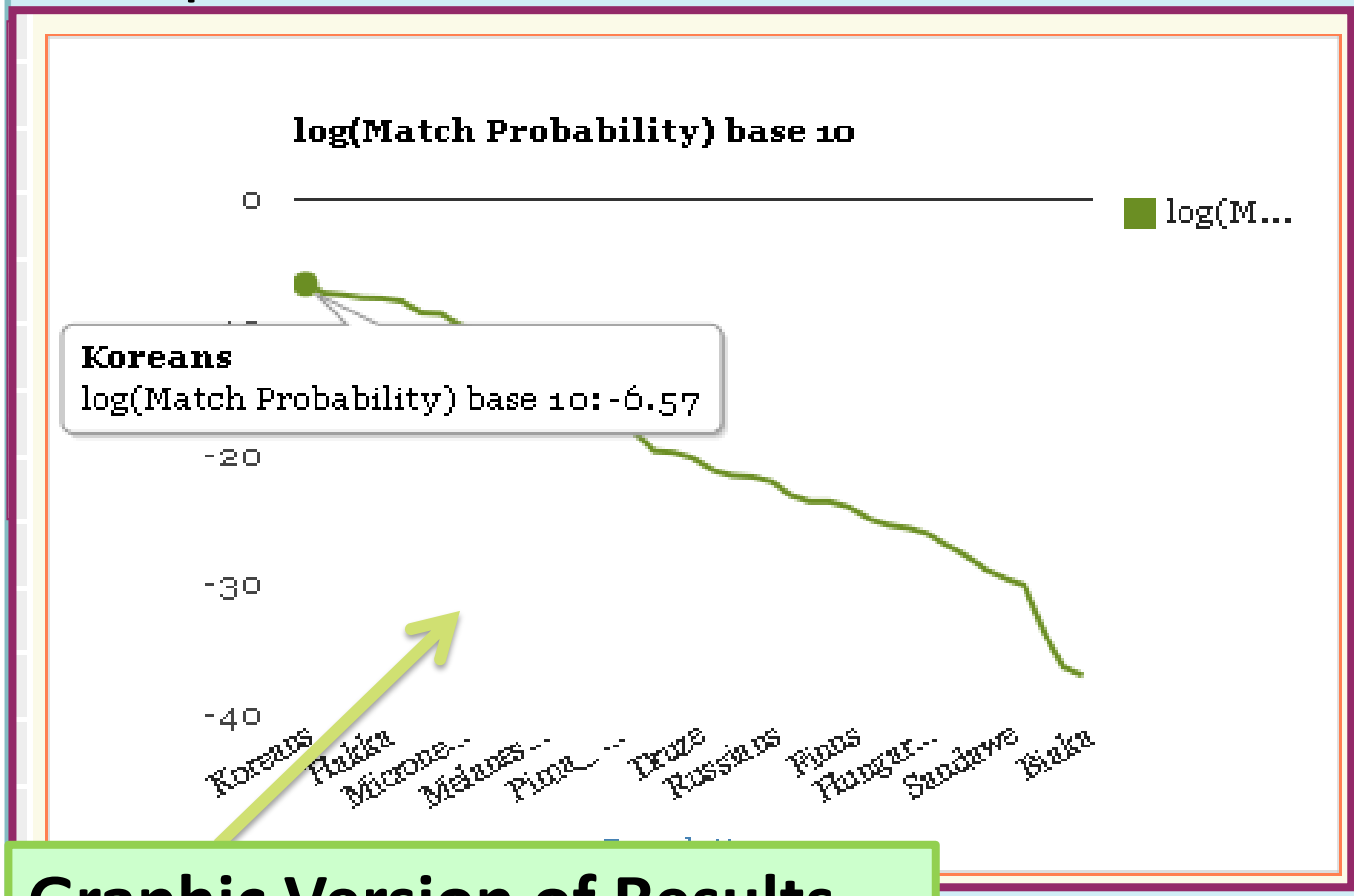
Results

Population	Match Probability
Koreans	2.683e-7
Cambodians_ Khmer	4.22e-8
Japanese	3.018e-8
Lao Loum	1.87e-8
Hakka	1.705e-8
Han	9.769e-9
Yakut	1.274e-9
Atayal	1.092e-9
Micronesians	1.509e-10
Ami	4.936e-11
Keralite	6.026e-13
Samaritans	4.266e-13
Melanesian_ Nasioi	2.237e-13
Maya_ Yucatan	1.19e-14
Khanty	3.066e-15
Ticuna	1.623e-15
• • •	
Chagga	1.577e-29
Sandawe	3.187e-30
Ibo	9.407e-31
Mbuti	2.12e-34
Yoruba	4.127e-37
Biaka	9.549e-38

Data Input and Output Screens for AISNPs

Data input for a Korean individual

Results



n	Match Probability
	2.683e-7
hmer	4.22e-8
	3.018e-8
	1.87e-8
	1.705e-8
	9.769e-9
	1.274e-9
	1.092e-9
	1.509e-10
	4.936e-11
	6.026e-13
	4.266e-13
sioi	2.237e-13
	1.19e-14
	3.066e-15
	1.623e-15
	1.577e-29
	3.187e-30
	9.407e-31
	2.12e-34
	4.127e-37
	9.549e-38

Graphic Version of Results

Technology Transition Workshop

NIJ National Institute of Justice

Interpreting the FROG Results

- **The probabilities given are the probability of the input genotype given the allele frequencies for those SNPs in each specific population**
- **For IISNPs this is interpretable as the probability of another occurrence of the specified genotype**
- **For AISNPs this is interpretable as the likelihood of the specified genotype occurring in each population. Only the relative likelihoods are meaningful.**

Technology Transition Workshop NIJ National Institute of Justice

AISNPs and PISNPs Will Be Important for Forensic Anthropology

- **STRPs provide no information on ancestry or phenotype**
- **Routine use of well documented panels of AISNPs and PISNPs will require that forensic labs incorporate the equipment and have the technicians trained in the protocols for SNPs**
- **This will overcome one of the barriers to more routine use of SNPs in individual identification**

Technology Transition Workshop NIJ National Institute of Justice

Cited Scientific References

- **Butler, J.M.; Budowle, B.; Gill, P.; Kidd, K.K.; Phillips, C.; Schneider, P.M.; Vallone, P.M.; Morling, N. Report on ISFG SNP panel discussion. *Forensic Science International: Genetics Supplement Series (Progress in Forensic Genetics 12) 2008, 1(1): 471-472*.**
- **Kidd, J.R.; Friedlaender, F.R.; Speed, W.C.; Pakstis, A.J.; De La Vega, F.M.; Kidd, K.K. Analysis of a Set of 128 Ancestry Informative Single-nucleotide Plolymorphisms in a Global Set of 119 Population Samples. *Investigative Genetics* 2011, 2(1); http://www.investigativegenetics.com/content/2/1/1 (Accessed Jul 21, 2011).**
- **Kidd, J.R.; Friedlaender, F.R.; Pakstis, A.J.; Furtado, M.; Fang, R.; Wang, X.; Nievergelt, C.M.; Kidd, K.K. SNPs and Haplotypes in Native American Populations. *American Journal of Physical Anthropology* April 2011, In press.**
- **Kosoy, R.; Nassir, R.; Tian, C.; White, P.A.; Butler, L.M.; Silva, G.; Kittles, R.; Alarcon-Riquelme, M.E.; Gregersen, P.K.; Belmont, J.W.; De La Vega, F.M., Seldin, M.F. Ancestry Informative Marker Sets for Determining Continental Origin and Admixure Proportions in Common Populations in America. *Human Mutation* 2009, 30(1): 69-78.**

Technology Transition Workshop

NIJ National Institute of Justice

Cited Scientific References (Continued)

- **Li, H.; Mukherjee, N.; Soundararajan, U.; Tárnok, Z.; Barta, C.; Khaliq, S.; Mohyuddin, A.; Kajuna, S.L.B.; Mehdi, S.Q.; Kidd, J.R.; Kidd, K.K. Geographically Separate Increases in the Frequency of the Derived ADH1B*47His Allele in Eastern and Western Asia. *American Journal of Human Genetics* 2007, 81(4): 842-846; http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2227934/ (Accessed Jul 21, 2011).**
- **Phillips, C.; Salas, A.; Sánchez, J.J.; Fondevila, M.; Gómez-Tato, A.; Álvarez-Dios, J.; Calaza, M.; Casares de Cal, M.; Ballard, D.; Lareu, M.V.; Carracedo, A. – The SNPforID Consortium Inferring Ancestral Origin Using a Single Multiplex Assay of Ancestry-informative Marker SMPs. *Forensic Science International: Genetics* 2007, 1(3): 273-280.**
- **Pritchard, J.K.; Stephens, M.; Donnelly, P. Inference of Population Structure Using Multilocus Genotype Data. *Genetics Society of America* 2000, 155: 945-959; http://pritch.bsd.uchicago.edu/publications/structure.pdf (Accessed Jul 21, 2011).**

Technology Transition Workshop

NIJ National Institute of Justice

Questions?

Technology Transition Workshops are a project of NIJ's Forensic Technology Center of Excellence, operated by the National Forensic Science Technology Center (www.nfstc.org), funded through cooperative agreement #2010-DN-BX-K210.

Technology Transition Workshop

NIJ National Institute of Justice

Contact Information

Kenneth K. Kidd, Ph.D.

Genetics Department

Yale University School of Medicine

New Haven, CT 06520-8005

Kenneth.kidd@yale.edu

Note: All images are courtesy of Dr. Kenneth K. Kidd, unless otherwise indicated.

Technology Transition Workshop

NIJ National Institute of Justice