Technology Transition Workshop| *Kenneth K. Kidd*

Genetics of SNP Markers

Why SNPs?

- **Plentiful – millions exist in the human genome**
- **Genetically simple – di-allelic and co-dominant**
- **Very low mutation rates – genetically stable**
- **Robust to DNA damage – small amplicons**
- **Multiple typing methods – easy to type**
- **Typing automatable – fast results**
- **Genotype calling automatable – interpretation easy, qualitative calls**

Technology Transition Workshop

Requirements for DNA Markers in Forensics

- **The genetic nature of the polymorphism must be well understood** [OK for SNPs]
- **The molecular methodology for testing the marker must be reliable** [OK for SNPs]
- **The markers should be usable in mixtures** [**NOT ± OK** for SNPs]
- **The statistical methods for evaluating the data must be sound** [OK for SNPs]
- **The data for use in the statistics must be sufficient** [**NOT OK** for SNPs **yet**]

Technology Transition Workshop NIJ National Institute of Justice

Problems with SNPs in Forensics

1. **Few SNPs have extensive population database support**
2. **No SNPs have forensic databases accumulated (i.e., offender and crime scene SNP data)**
3. **SNPs are problematic with sample mixtures**
4. **Few forensic labs have experience with SNPs**
5. **No agreed upon common set of SNPs to consider**
6. **SNPs vary widely in their population genetic characteristics**
7. **Different SNPs are needed for different purposes**

Technology Transition Workshop

NIJ National Institute of Justice

Progress on Problems with SNPs in Forensics

1. **Few SNPs have extensive population database support**
   - **Increasing numbers of SNPs are being tested on multiple populations**
   - **Though sample sizes per population are often less than usually considered adequate, there are multiple populations from each geographic area in the accumulating body of knowledge**
   - **Our studies include work to collect more SNP data on many population samples**
   - **We are also attempting to accumulate in one place the allele frequency data being collected and published by many research groups**
   - **We are using the NSF-supported database ALFRED: http://alfred.med.yale.edu**

Technology Transition Workshop NIJ National Institute of Justice

Progress on Problems with SNPs in Forensics

2. **No SNPs have forensic databases accumulated (i.e., offender and crime scene SNP data)**
   - ***If*** **SNP panels can be agreed upon, a parallel processing can be done relatively cheaply**
   - **In many cases, SNPs would be sufficient for a local crime with clear suspect**
3. **SNPs are problematic with sample mixtures**
   - **Procedures and statistics are being developed to use SNPs to detect mixtures, but I am not sure what the number of SNPs needed will be or how technically robust the methods will be**

Technology Transition Workshop NIJ National Institute of Justice

Progress on Problems with SNPs in Forensics

4. **Few forensic labs have experience with SNPs**
   - **I may be wrong in this statement but existence of this workshop supports the belief**
5. **No agreed upon common set of SNPs to consider**
   - **The European SNPforID Consortium has two SNP panels now recognized and used by forensic labs in several countries**
   - **My research is attempting to find better panels, to be discussed later in this talk**

Technology Transition Workshop NIJ National Institute of Justice

Progress on Problems with SNPs in Forensics

6. **SNPs vary widely in their population genetic characteristics**
7. **Different SNPs are needed for different purposes**

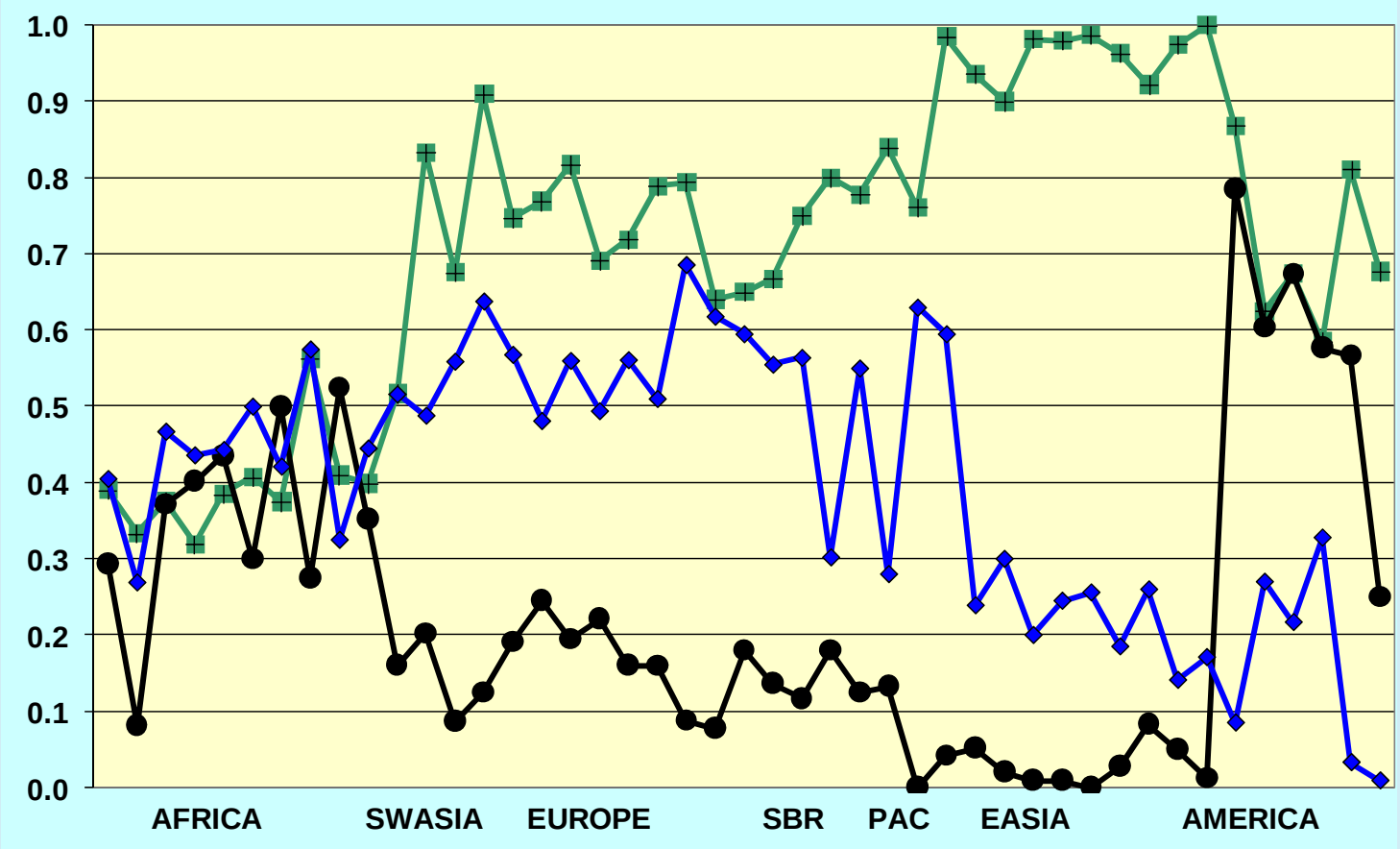
The following slides illustrate how allele frequencies among populations can vary greatly among SNPs.

That variation, or lack thereof, can be used for different forensic purposes.

Technology Transition Workshop

NIJ National Institute of Justice

Allele Frequencies at Some SNPs Vary Greatly Among Populations



Technology Transition Workshop

NIJ National Institute of Justice

Types of SNP Panels

- **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.**

Technology Transition Workshop

NIJ National Institute of Justice

Reproduced in Butler et al., 2008

Requirements for IISNPs

- **Individual Identification SNPs (IISNPs):**
  - **SNPs that collectively give very low probabilities of two individuals having the same multisite genotype**
    - **We have added the additional criterion that ethnicity should not be an issue in determining the match probability**
    - **Therefore, we are requiring all such SNPs to be close to maximally informative all around the world**
    - **That translates in practice to SNPs that have globally average heterozygosities > 0.4 and global Fst values < 0.06**
    - **The work we have done is covered in detail in the following slides**

Technology Transition Workshop NIJ National Institute of Justice

Procedures for Identifying IISNPs

- **Identify likely candidate polymorphisms**
- **Screen on a few populations**
- **Retain the "best"**
- **Test on many populations**
- **Retain the "best"**
  - **Reliability of typing**
  - **Hardy-Weinberg criterion**
- **Test for LD and linkage**

Technology Transition Workshop

NIJ National Institute of Justice

Population Basis of Final IISNP Panel

Data on 2358 individuals from 44 populations

	Number of Individuals Sampled	Number of Populations Sampled
Africa	503	10
South West Asia	273	4
Europe	567	9
Siberia	148	3
South Central Asia	30	1
East Asia	481	9
Pacific Islands	60	2
Americas	296	6

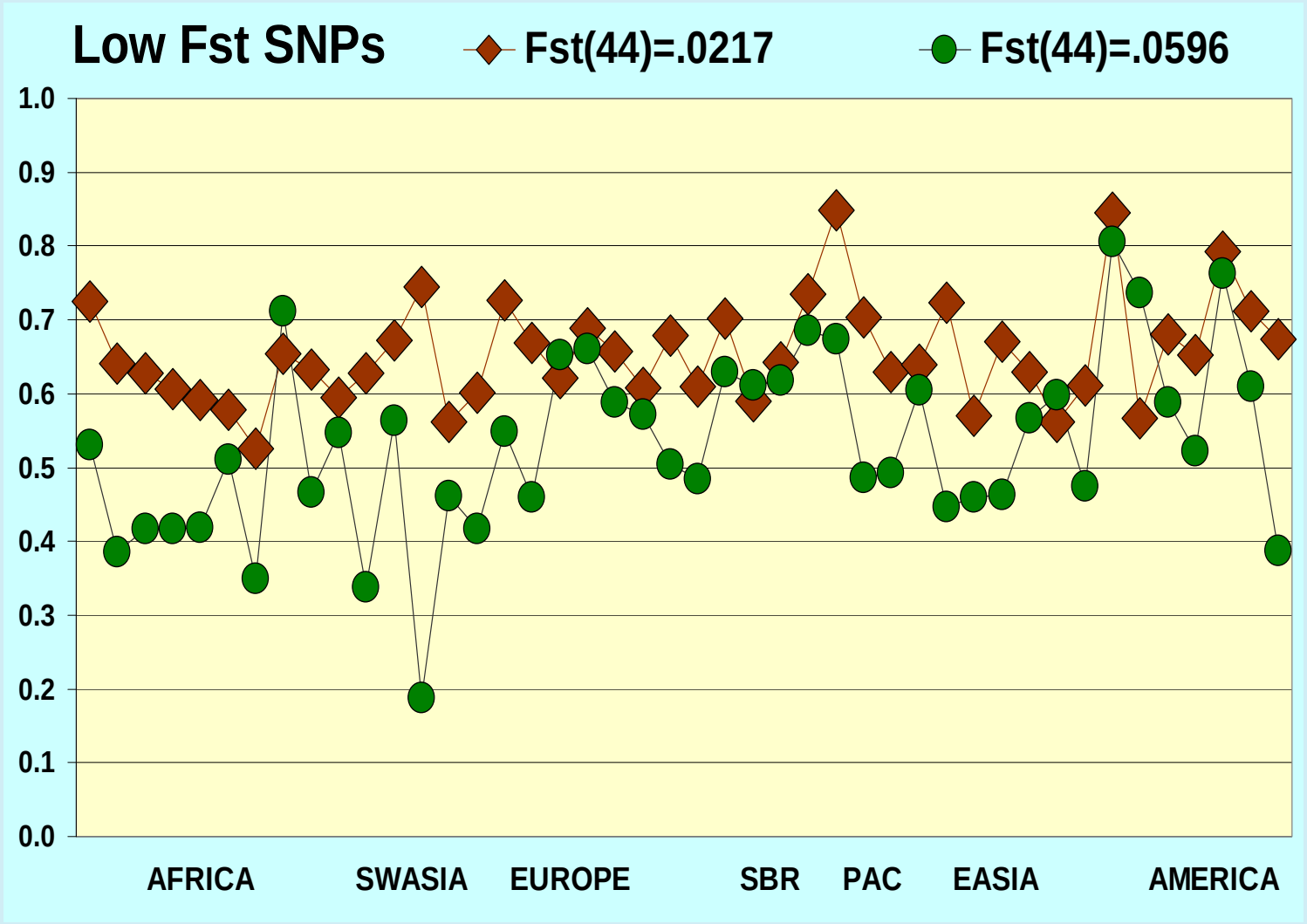
Technology Transition Workshop

NIJ National Institute of Justice

Examples of candidate IISNPs — high heterozygosity and little allele frequency variation among populations

Technology Transition Workshop

NIJ National Institute of Justice

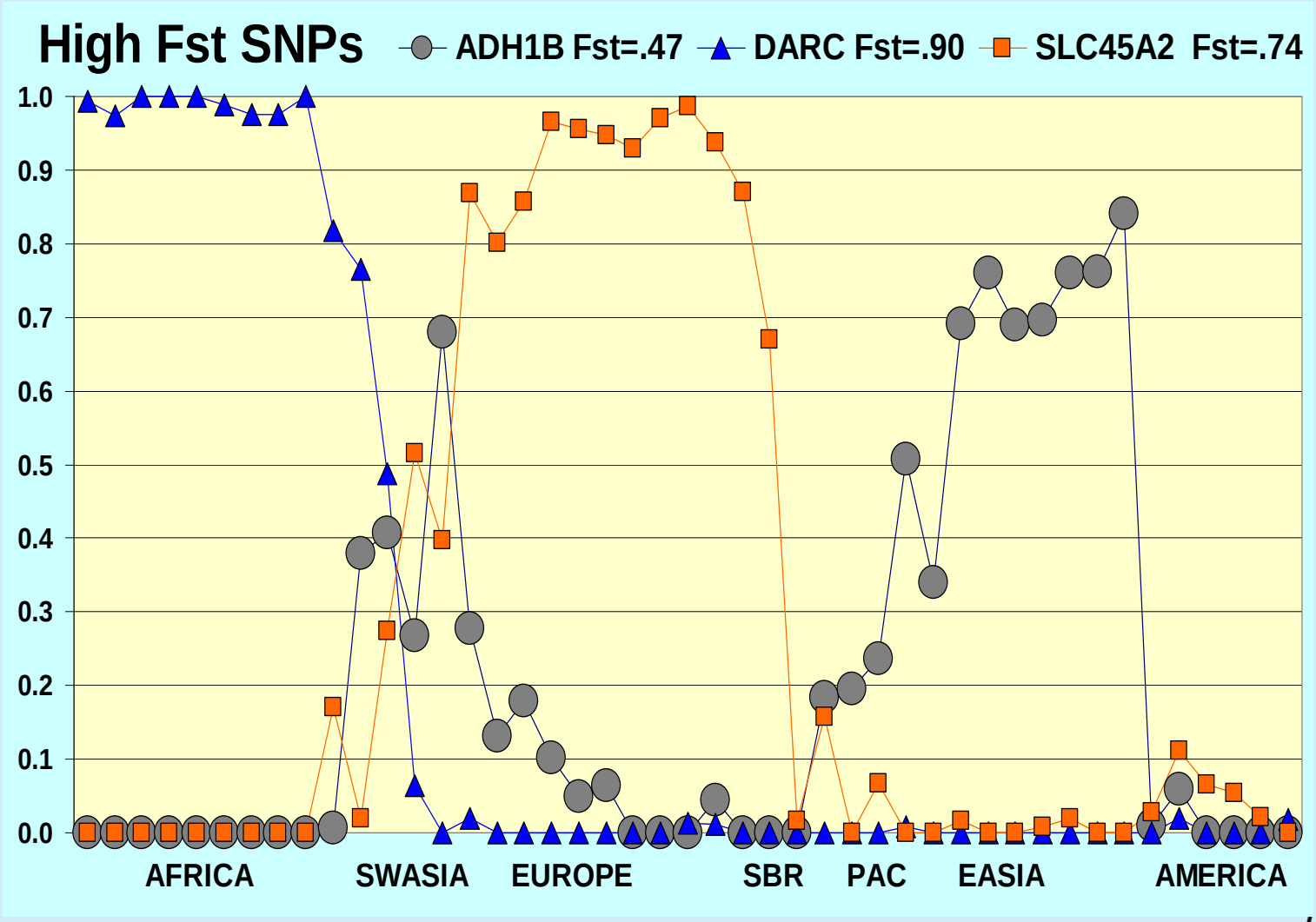


Technology Transition Workshop

NIJ National Institute of Justice

If you think those show high variation, consider examples of candidate AISNPs

Technology Transition Workshop



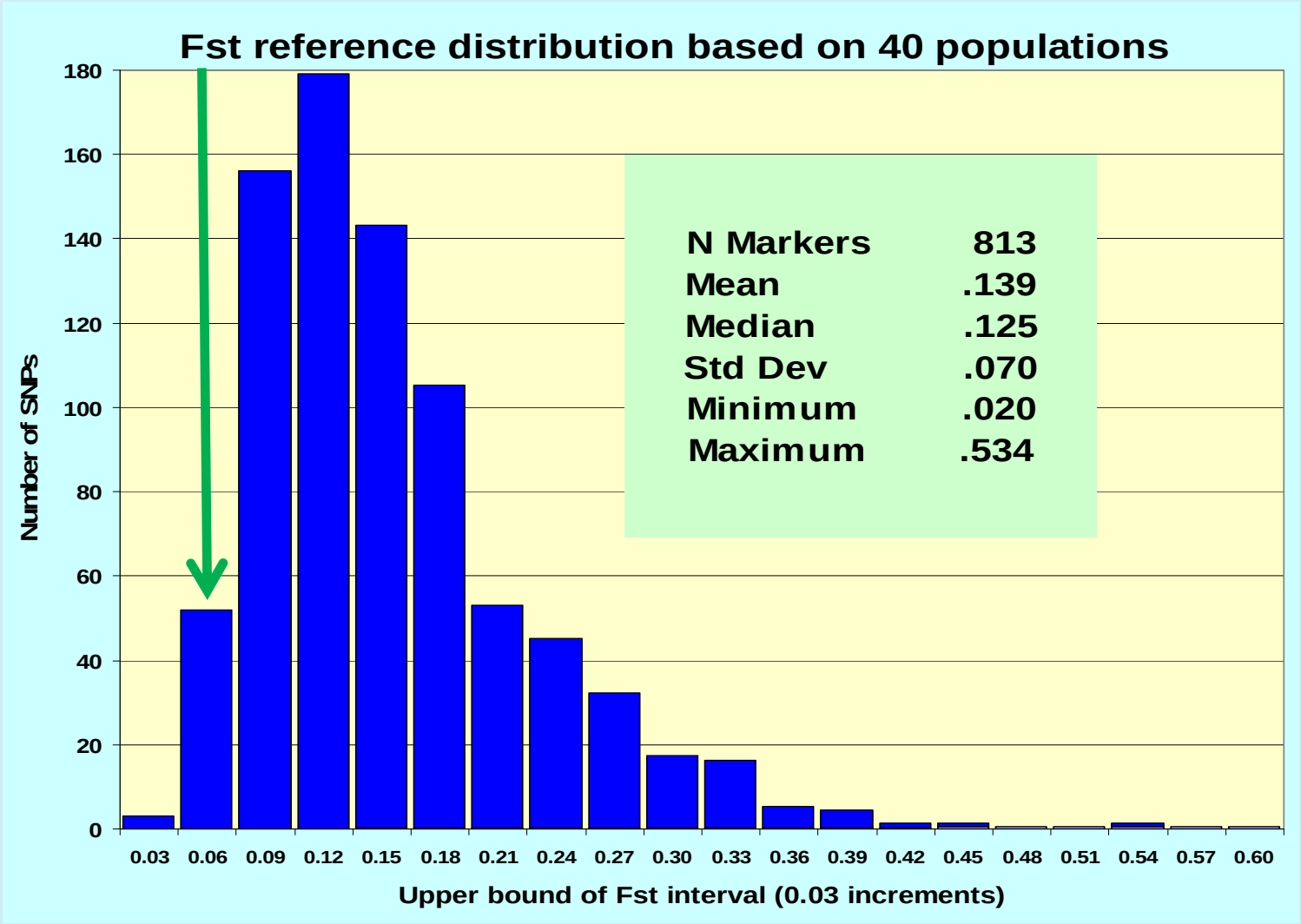
Technology Transition Workshop

NIJ National Institute of Justice

Summary of Screening for IISNPs

- We screened several data sets with allele frequencies for multiple SNPs on 4 to 50 diverse populations
- We typed > 500 of these on our 44 populations
- 92 SNPs (~20% of those screened) met criteria of average heterozygosity > 0.4 and Fst < 0.06
- 86 of those 92 showed no significant pairwise LD in the (86 x 85) / 2 = 3,655 tests
- **45 of those 86 have no or very loose linkage: our final IISNP panel**

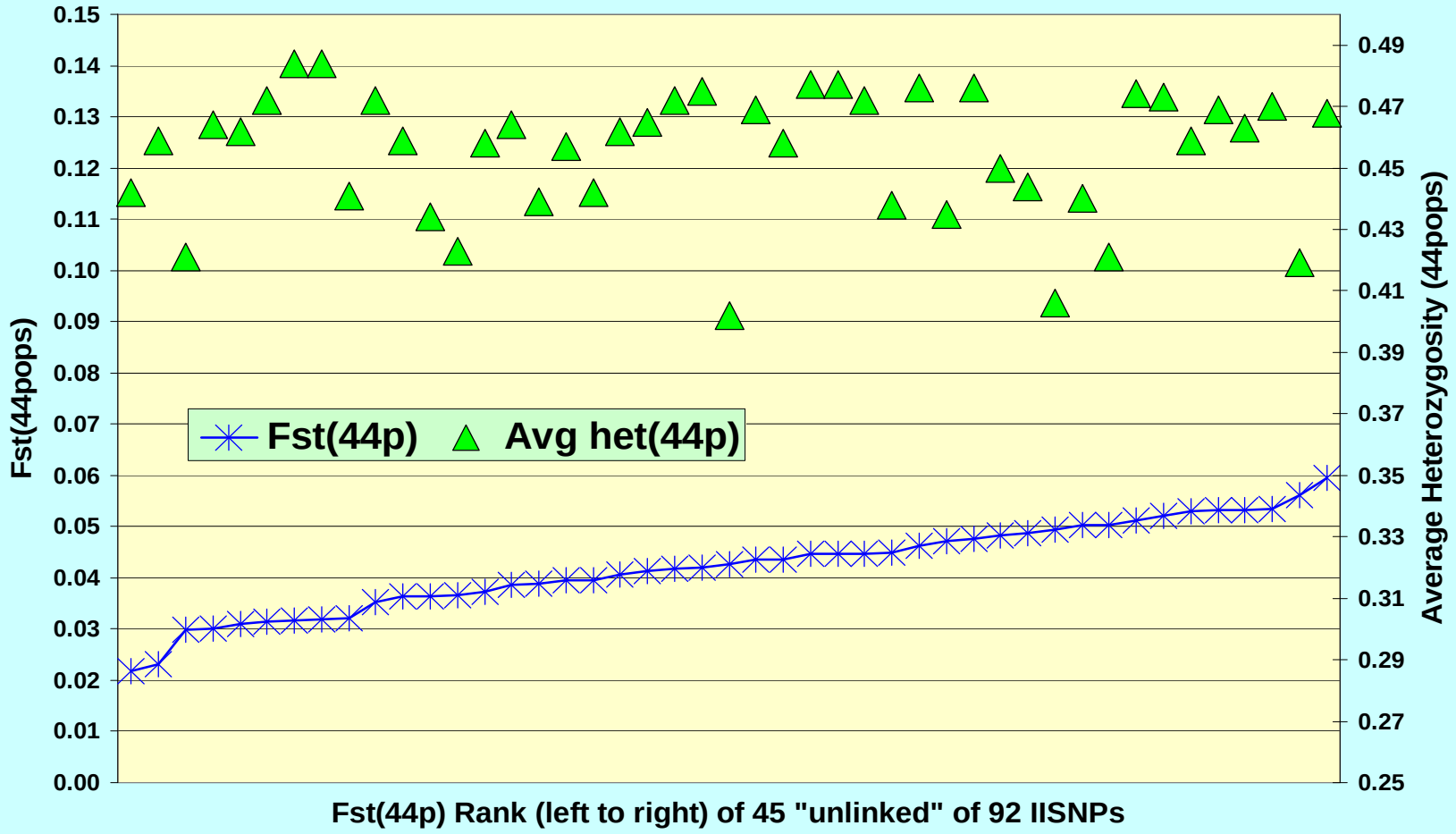
Technology Transition Workshop NIJ National Institute of Justice



Technology Transition Workshop

NIJ National Institute of Justice

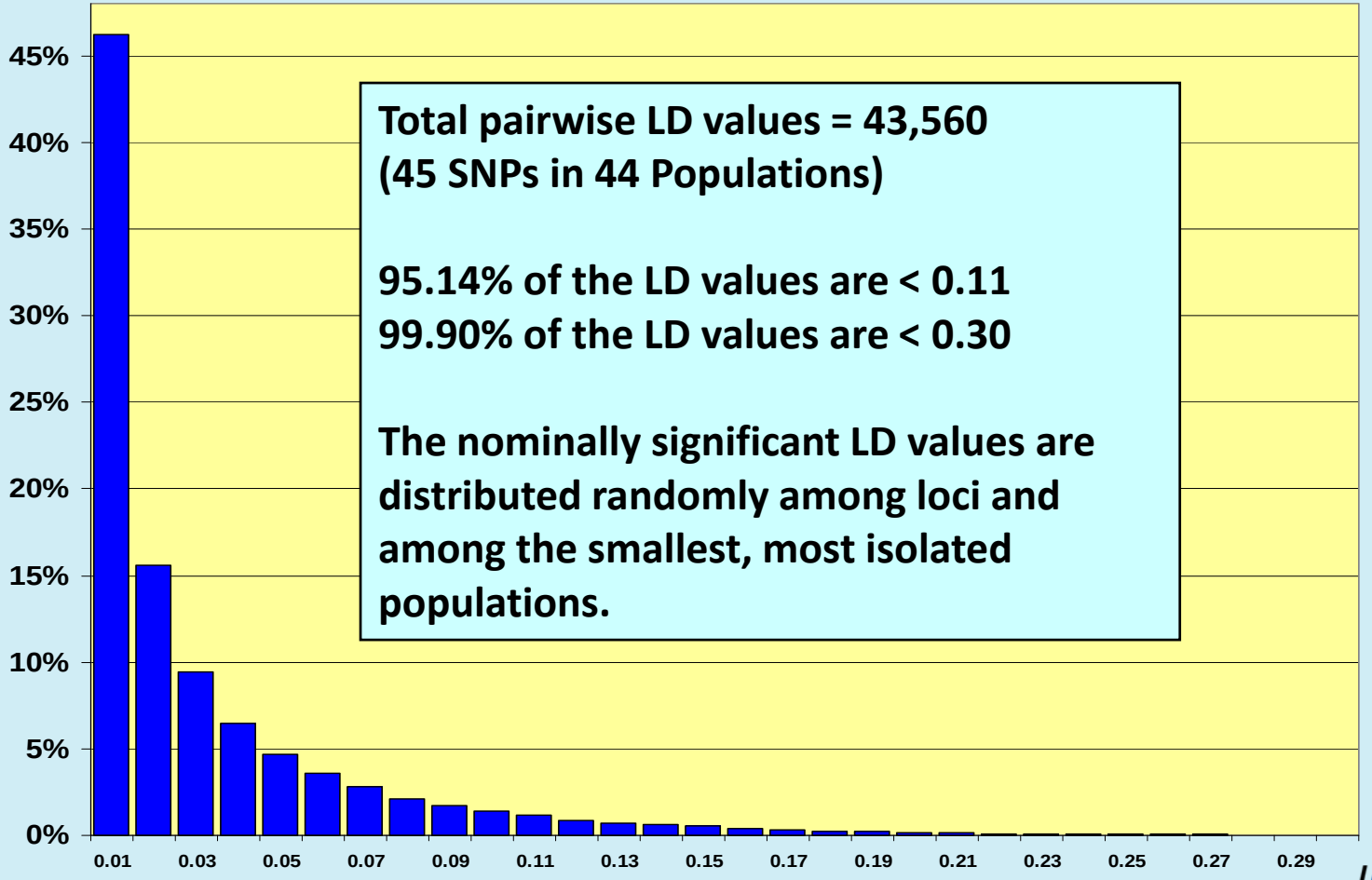
Fst and Average Heterozygosity for the 45-SNP Panel



Technology Transition Workshop

NIJ National Institute of Justice

Distribution of All Pairwise LD Values (r^2)



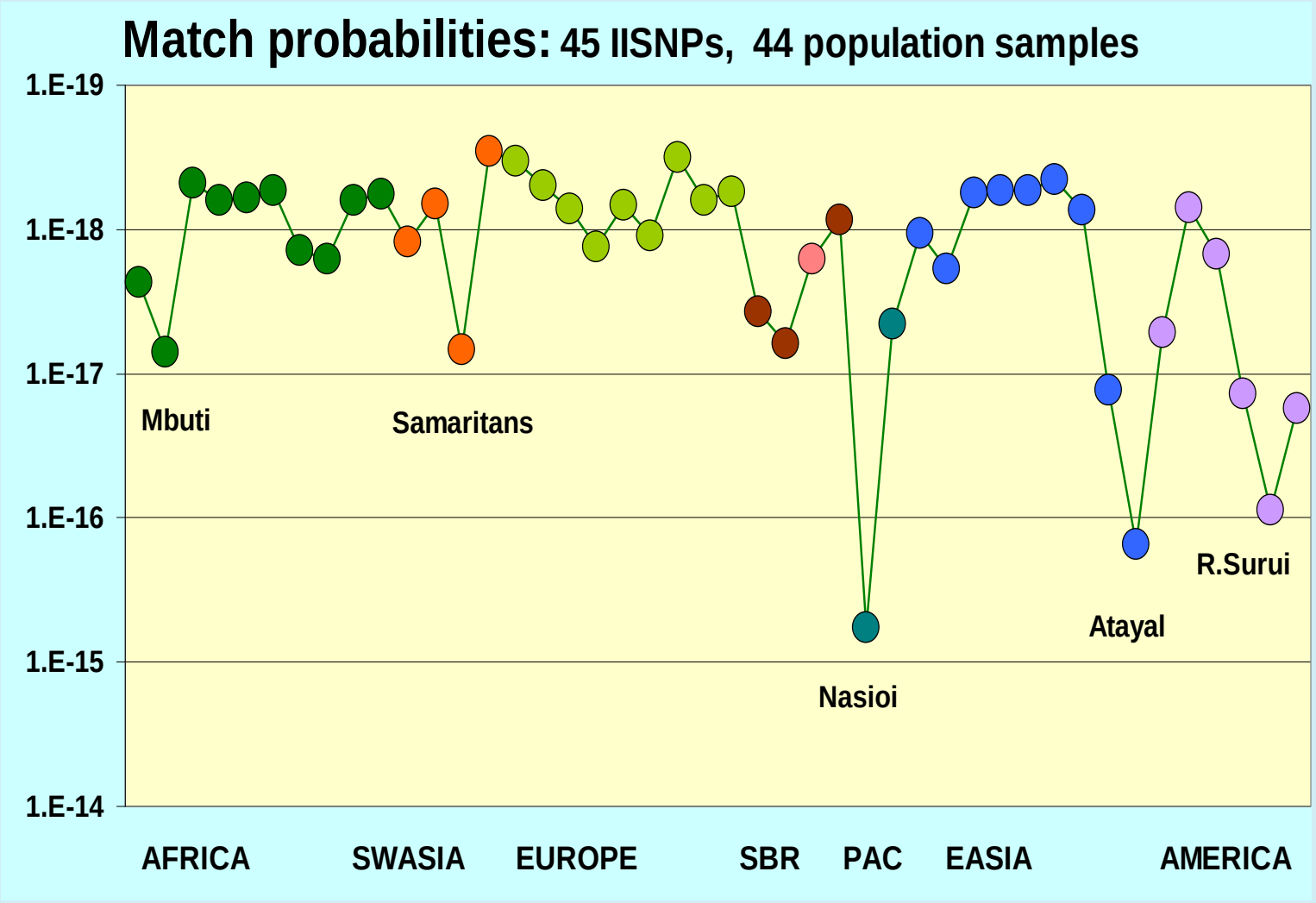
Total pairwise LD values = 43,560
(45 SNPs in 44 Populations)

95.14% of the LD values are < 0.11
99.90% of the LD values are < 0.30

The nominally significant LD values are distributed randomly among loci and among the smallest, most isolated populations.

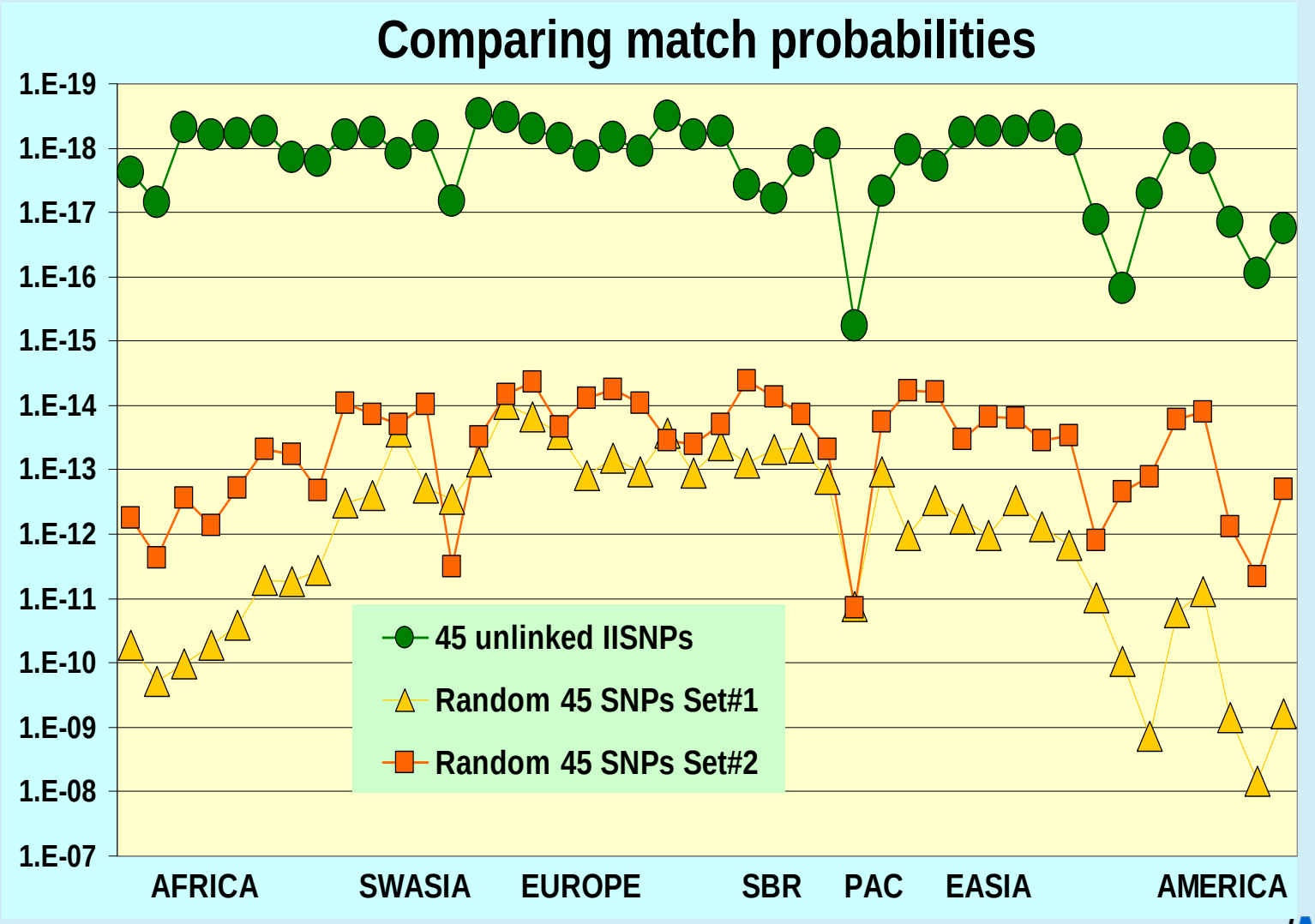
Technology Transition Workshop

NIJ National Institute of Justice



Technology Transition Workshop

NIJ National Institute of Justice



Technology Transition Workshop

NIJ National Institute of Justice

Points To Note On Previous Slide

- **The "random" SNPs are from our in-lab database of SNPs we study because they show high heterozygosity in one region of the world**
- **Truly random SNPs would have much lower average heterozygosity and be much less informative**
- **There is a strong European bias because until recently most SNPs were discovered because they had a high heterozygosity in Europe**

An Empiric Evaluation of Matches

- **We have compared genotypes between all possible pairs of individuals who had data on all 45 IISNPs**
- **The numbers of genotypes out of 45 that match have been tabulated separately for pairs within a population and pairs between populations since some of the populations, especially the small tribal ones, contain closely related individuals**

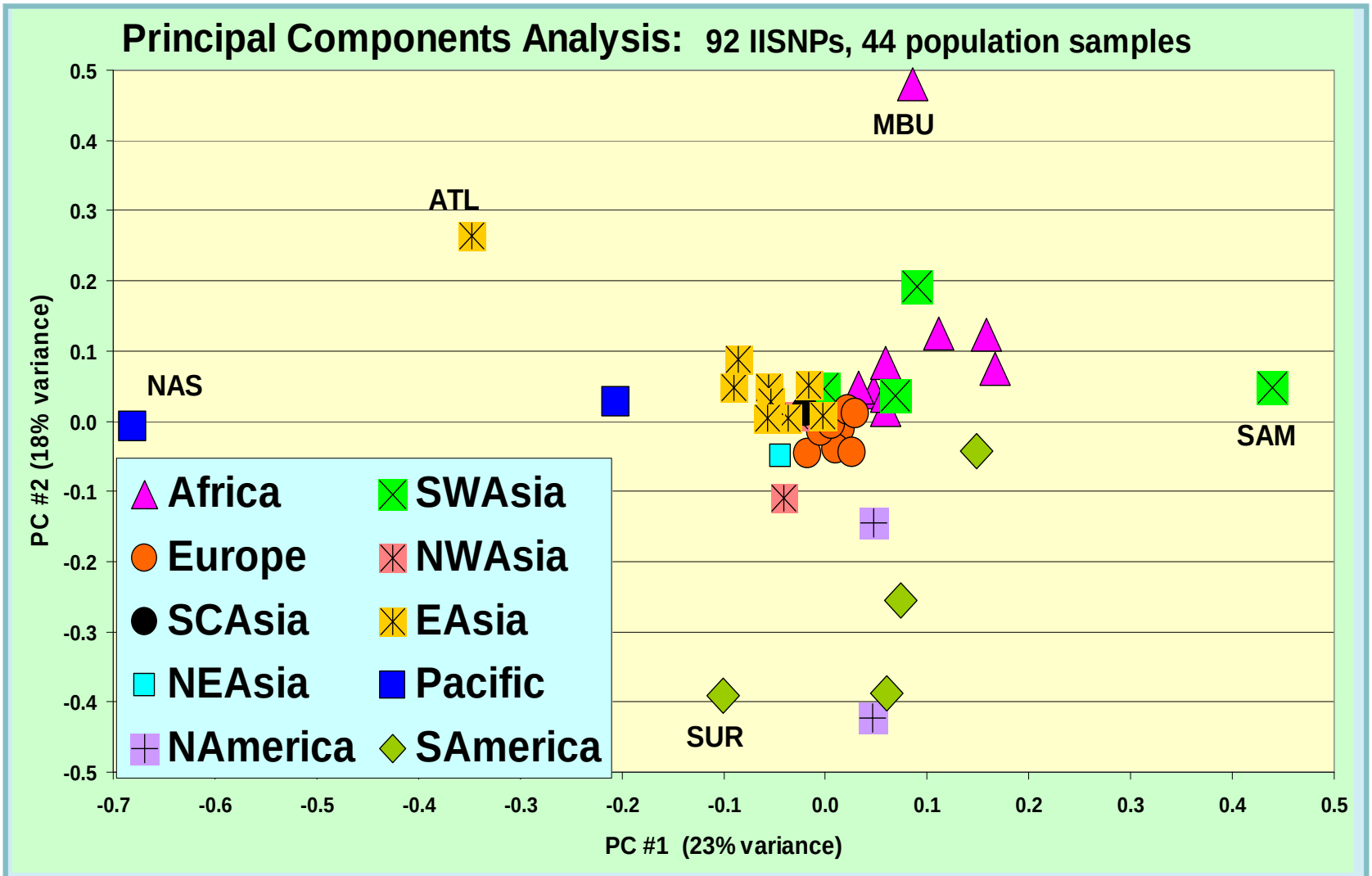
Match	Within	Between	Combined
0	0	0	0
1 or 2	0	0	0
3 or 4	0	18	18
5 or 6	7	348	355
7 or 8	82	4,150	4,232
9 or 10	514	25,346	25,860
11 or 12	1,974	94,245	96,219
13 or 14	5,040	225,443	230,483
15 or 16	8,933	362,366	371,299
17 or 18	10,873	398,947	409,820
19 or 20	9,307	308,707	318,014
21 or 22	5,770	168,386	174,156
23 or 24	2,731	64,779	67,510
25 or 26	929	18,030	18,959
27 or 28	342	3,484	3,826
29 or 30	121	477	598
31 or 32	39	31	70
33 or 34	12	4	16
35 or 36	3	1	4
37 or 38	1	0	1
39 or 40	0	0	0
41 or 42	0	0	0
43 or 44	0	0	0
45	0	0	0
Totals	**46,678**	**1,674,762**	**1,721,440**

Numbers of genotypes matching in all possible pairwise comparisons of 1856 individuals (in all 44 populations) that were fully typed for all 45 IISNPs

← The highest number of loci matching

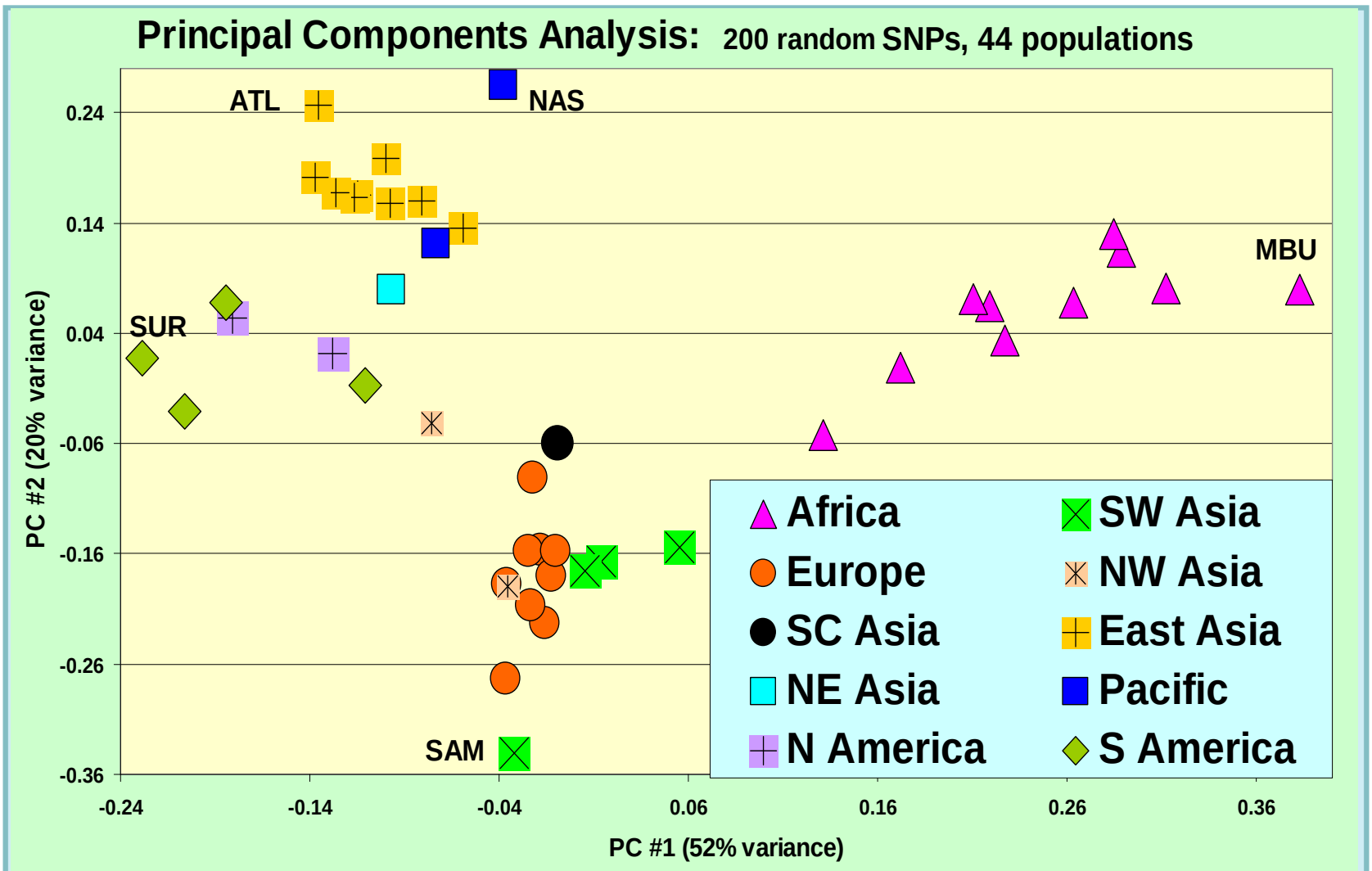
Technology Transition Workshop

NIJ National Institute of Justice



Technology Transition Workshop

NIJ National Institute of Justice



Technology Transition Workshop

NIJ National Institute of Justice

Some Obvious Observations

- **Fst values can change significantly as the number of populations considered in the calculations increases but Fst stabilizes with global coverage**
- **A few populations are 'outliers' and often have significantly different allele frequencies**
  - **Isolated populations?**
  - **Bottleneck?**
  - **Founder effect?**
  - **Small sample size?**
- **Any other measure of allele frequency variation should be highly correlated with Fst, so the set of IISNPs identified should be quite generally valid, though the rank-order might change**

Technology Transition Workshop

NIJ National Institute of Justice

The Next Steps

- **Test on other typing platforms – who will do this?**
- **Develop multiplex assays – already being done by AB**
- **Test in more populations – who will do this?**
- **Test in forensic practice – being initiated**
- **Adopt a standard panel**

Technology Transition Workshop NIJ National Institute of Justice

Requirements for AISNPs

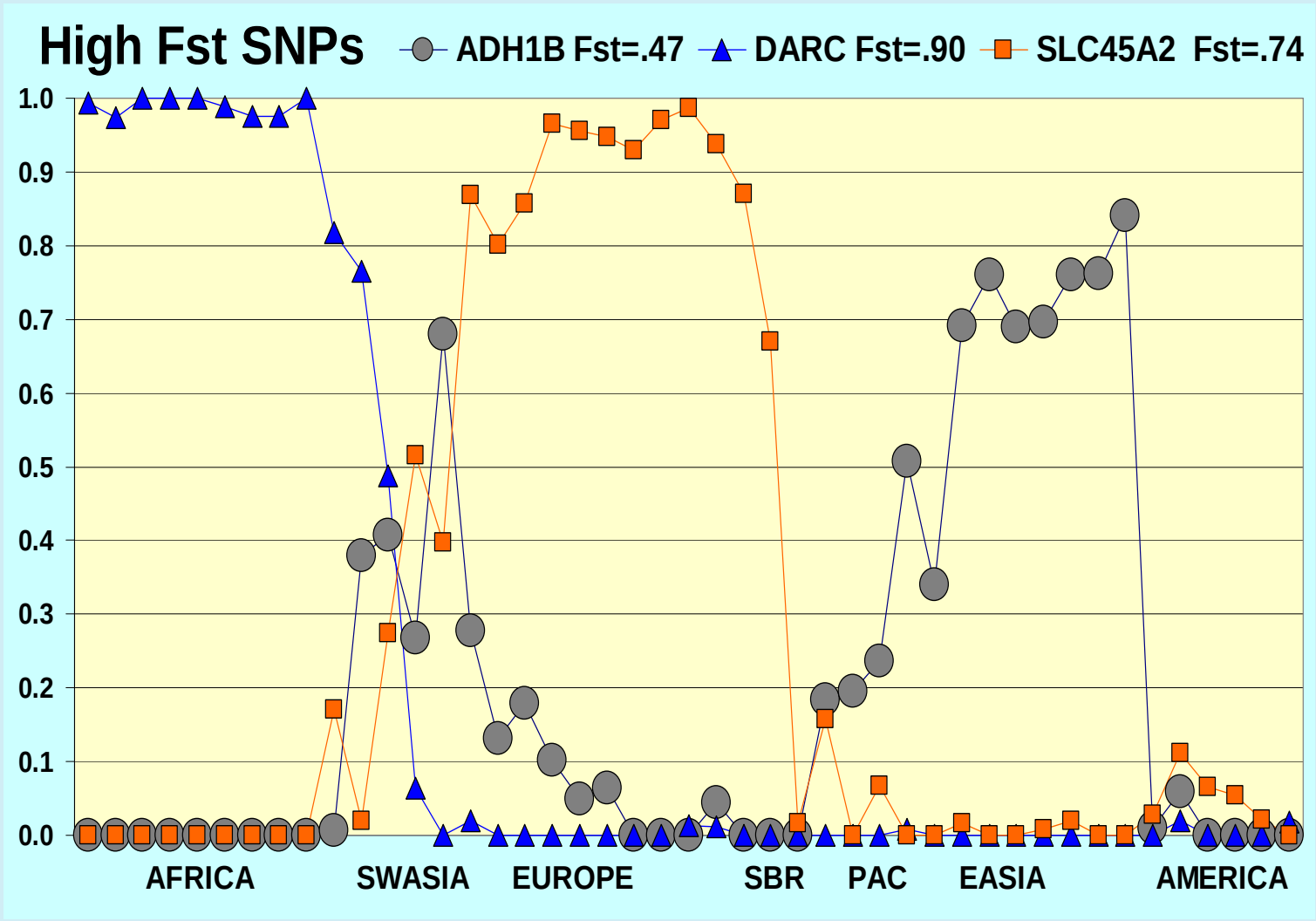
- **Ancestry Informative SNPs (AISNPs):**
  - **SNPs that collectively give a high probability of an individual's ancestry being from one part of the world or being derived from two or more areas of the world**
    - **We are currently accumulating data on SNPs that show very high allele frequency variation among populations**
    - **It is very easy to find SNPs that will differentiate ancestry entirely from indigenous peoples of West Africa, Western Europe, Far East Asia, or the Americas**
    - **It is far more difficult to differentiate ancestry from geographically "intermediate" regions**
    - **The components of admixed ancestry are also very difficult to determine**
    - **In my opinion the companies that sell such services are not sufficiently accurate for forensic purposes**

Technology Transition Workshop NIJ National Institute of Justice

Preliminary Results

- **As part of our general research we have studied many markers on our populations and have been evaluating haplotypes for their utility in distinguishing among populations**
- **We have begun identifying individual SNPs that are highly informative in variation among populations**

Technology Transition Workshop



Technology Transition Workshop

NIJ National Institute of Justice

Results for 506 haplotypes based on 2556 SNPs in 45 populations, a total of 6.22 million genotypes

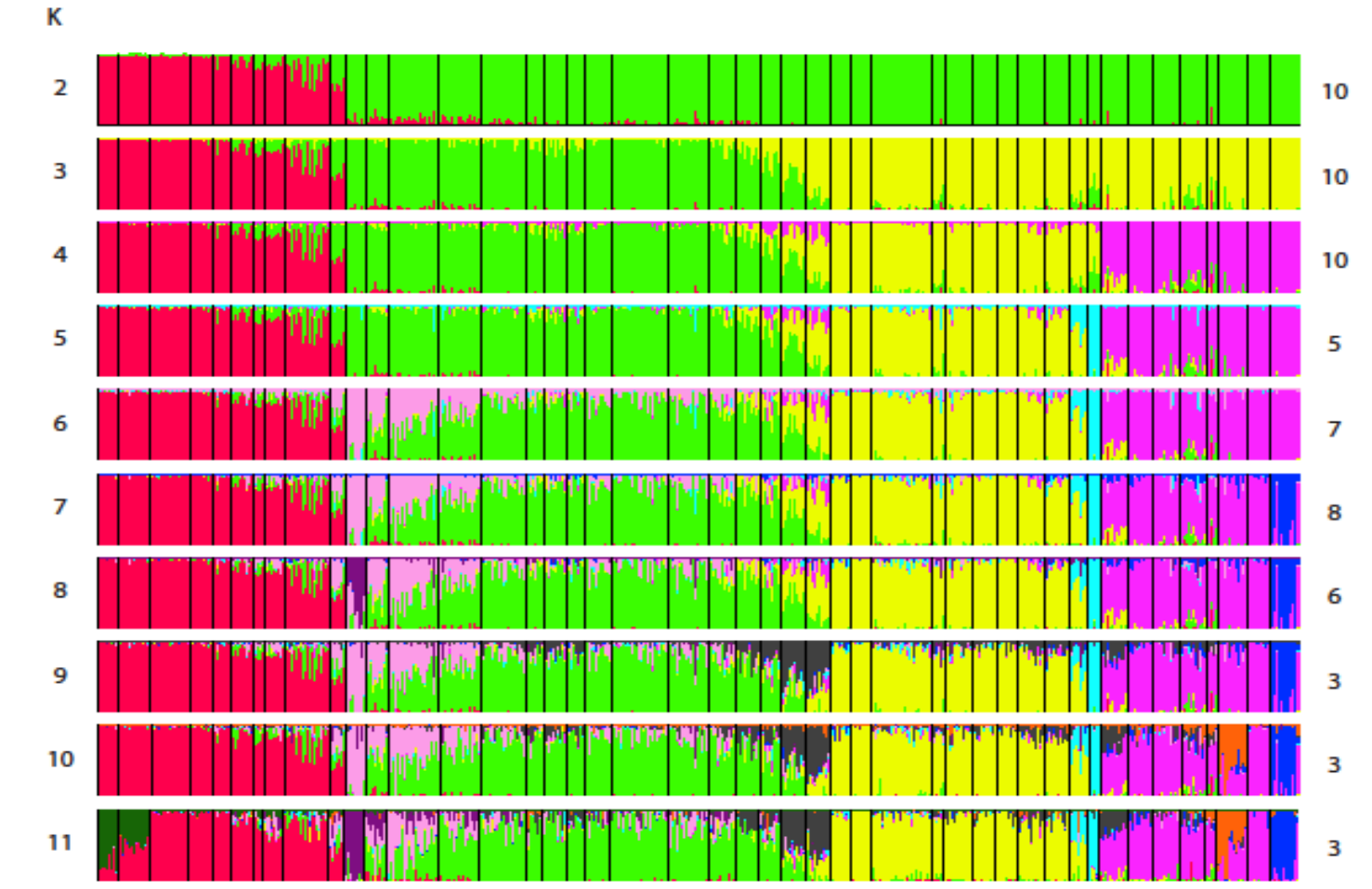
followed by

results for 128 high-Fst SNPs in 71 populations

Technology Transition Workshop

NIJ National Institute of Justice

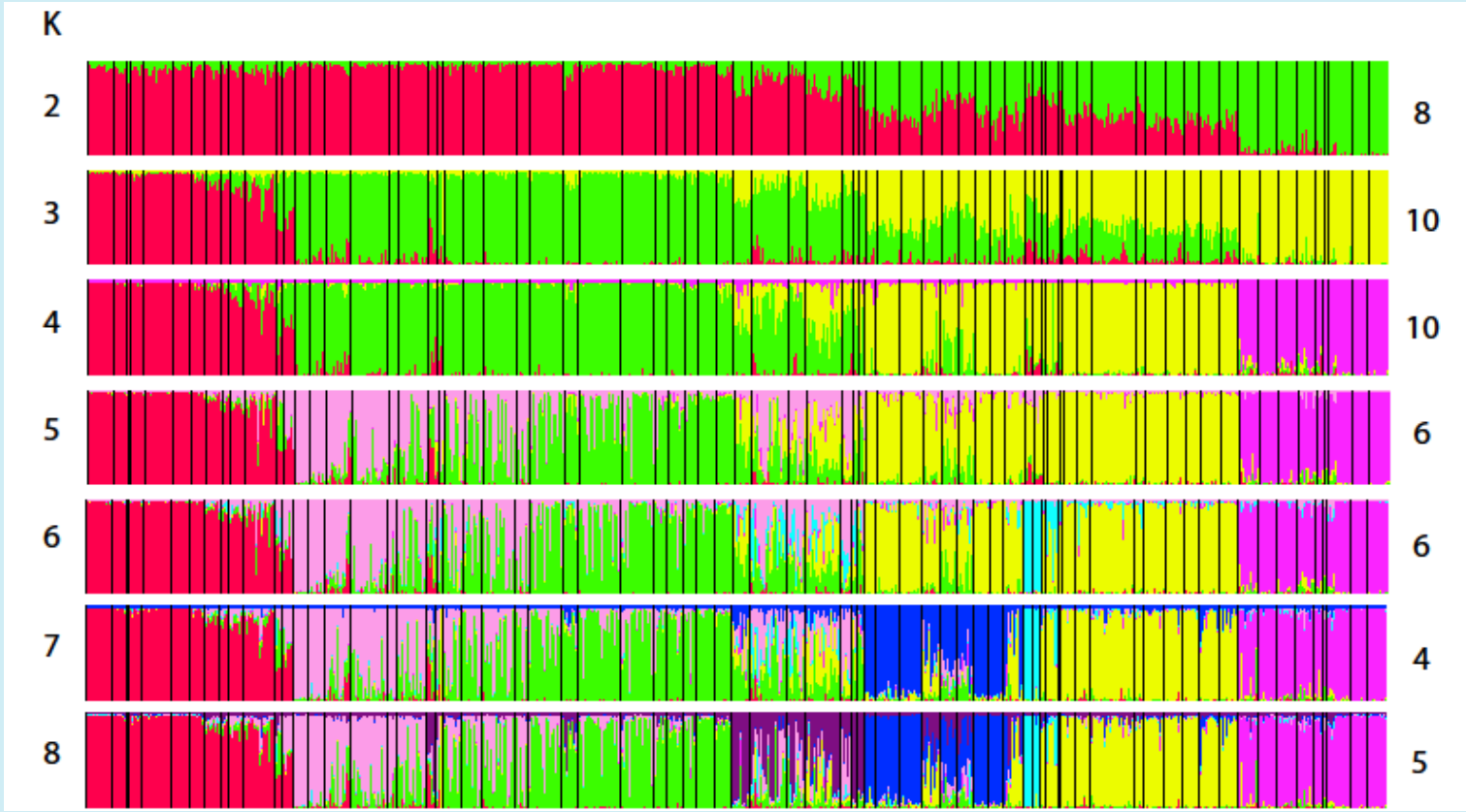
STRUCTURE Analyses of 506 Haplotypes



Technology Transition Workshop

NIJ National Institute of Justice

STRUCTURE Analysis of 128 High-Fst SNPs in 71 Populations



Technology Transition Workshop

NIJ National Institute of Justice

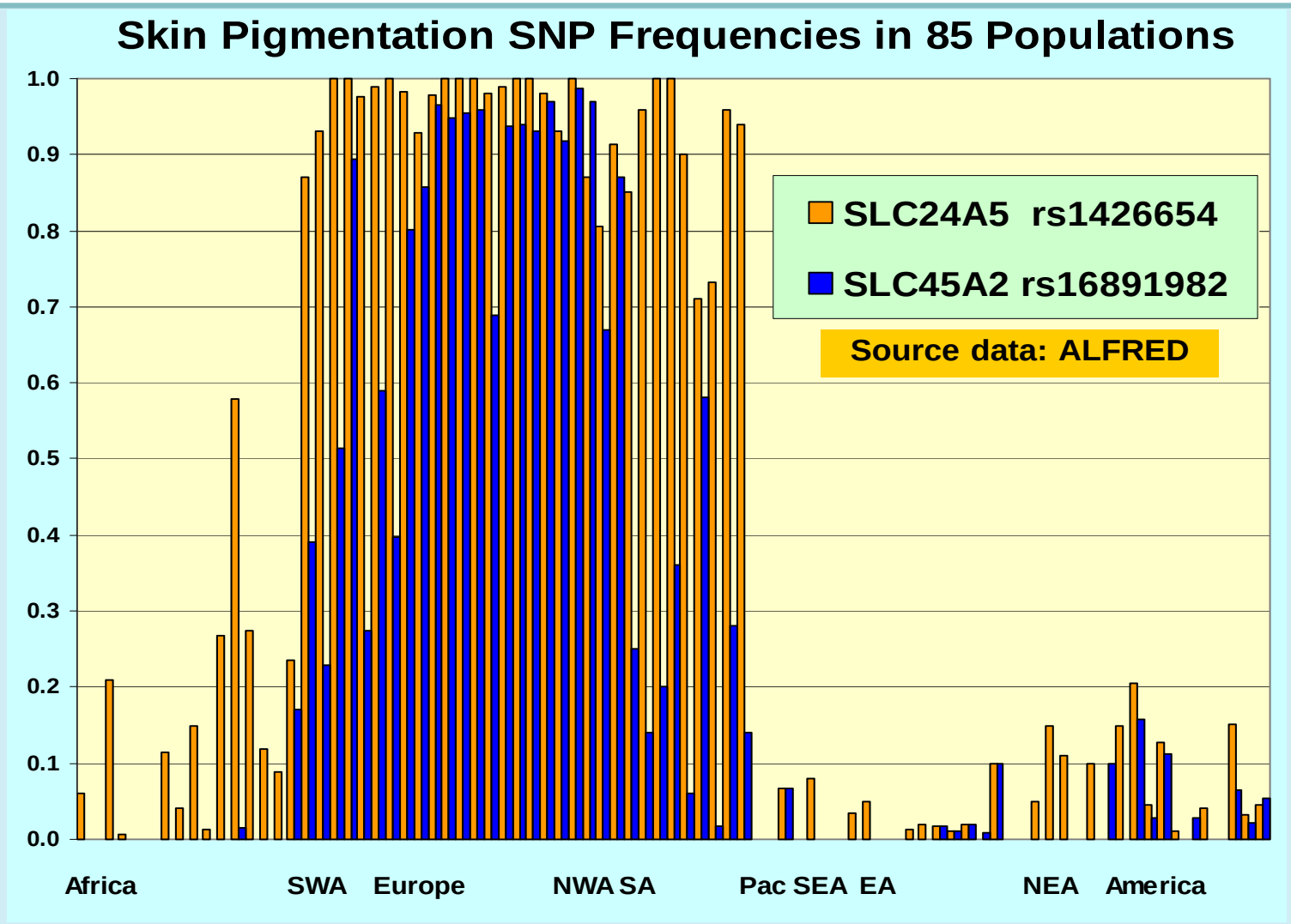
Requirements for LISNPs

- **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**
    - **Many of the haplotypes in the previous figure will be useful for this purpose**
    - **Each will need to be evaluated for its heterozygosity, lack of frequent recombination, etc.**

Technology Transition Workshop NIJ National Institute of Justice

Requirements for PISNPs

- **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.**
    - **This is a very problematic area because phenotype is complex and most existing data are correlational, not biologically definitive**
    - **The four loci believed most responsible for light skin color in Europeans have very different allelic distributions in the regions flanking Western Europe (ALFRED and unpublished results from Kidd Lab)**



Technology Transition Workshop

NIJ National Institute of Justice

Requirements for PISNPs

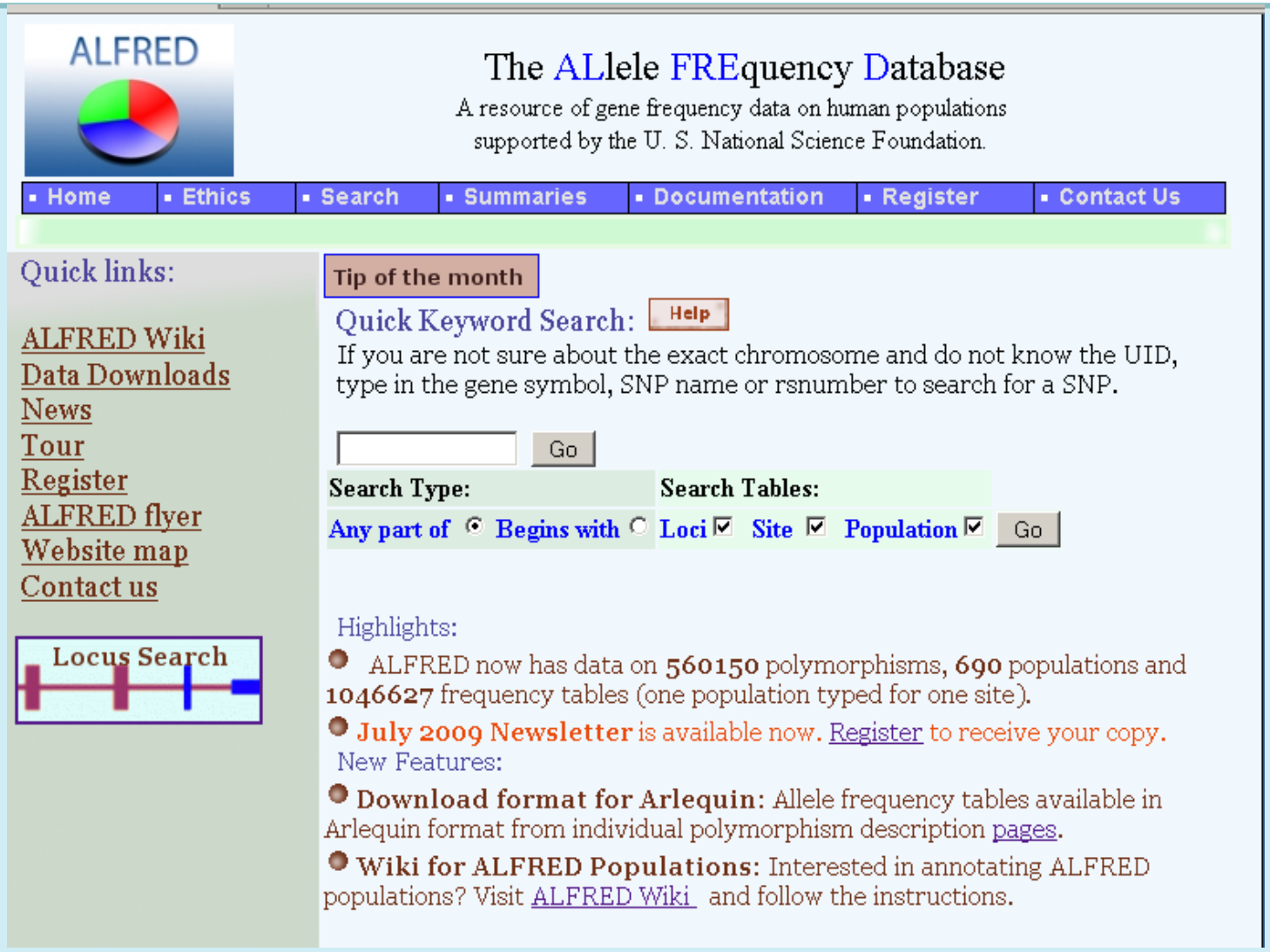
- **A biological understanding of the relationship between the three genotypes at a SNP and the phenotype variation**
- **A biological understanding of the relationships among the genotypes at the several loci and the phenotype variation**
- **A population genetic understanding of how the genotype frequencies vary among populations**
- **PISNPs are not ready for "prime time"; a simple correlation at the population level is generally not sufficient**

Technology Transition Workshop NIJ National Institute of Justice

Data Availability

- **As we are accumulating data on additional markers and additional populations the data are being made public through two sources:**
  - **(1) the Kidd Lab website with relevant forensic annotation and**
    - **Kidd Lab: http://info.med.yale.edu/genetics/kkidd**
  - **(2) ALFRED with the links to genetic, population, and molecular descriptions**
    - **ALFRED: http://alfred.med.yale.edu**

Technology Transition Workshop NIJ National Institute of Justice



Technology Transition Workshop

NIJ National Institute of Justice

Acknowledgements

- **This work is currently funded by grant 2007-DN-BX-K197 from the NIJ**
- **NIH Grants AA009379 and GM057672 fund the ongoing general work on population genetics of DNA markers providing resources on which the forensic studies rely**
- **NSF Grant BCS-0938633 funds the maintenance of ALFRED and helps us make our forensic data publically available**

Technology Transition Workshop

Acknowledgements

- **The data presented here are the result of work by many individuals:**
  - **Andrew J. Pakstis, Ph.D.**
  - **Judith R. Kidd, Ph.D.**
  - **William C. Speed**
  - **Eva Straka**
- **We also thank the many hundreds of anonymous individuals for their participation in these studies**
- **These studies would not be possible without their voluntary consent to give blood samples for studies of genetic variation**

Technology Transition Workshop

NIJ National Institute of Justice

Questions?

Technology Transition Workshop

Contact Information

Kenneth K. Kidd

Department of Genetics

Yale University School of Medicine

P.O. Box 208005

New Haven, CT 06520-8005

203-785-2654

kenneth.kidd@yale.edu

Note: *All images are courtesy of Dr. Kenneth K. Kidd.*

Technology Transition Workshop NIJ National Institute of Justice

Appended Information for Reference or to Address Questions

Technology Transition Workshop

NIJ National Institute of Justice

Publications of Kidd Laboratory Research to Date

- **449. Kidd K.K., A.J.Pakstis, W.C. Speed, E.L. Grigorenko, S.L.B. Kajuna, N.J. Karoma, S. Kungulilo, J.-J. Kim, R.-B. Lu A. Odunsi, F. Okonofua, J. Parnas, L.O. Schulz, O.V. Zhukova, and J.R. Kidd, 2006. Developing a SNP panel for forensic identification of individuals. *Forensic Science International* 164 :20-32**
- **461. Pakstis A. J., W. C. Speed, J. R. Kidd, and K. K. Kidd, 2007. Candidate SNPs for a Universal Individual Identification Panel. *Human Genetics* 121: 305-317**
- **467. Pakstis, A. J., W. C. Speed, J. R. Kidd, and K. K. Kidd, 2008. SNPs for Individual Identification. *Progress in Forensic Genetics : Genetics Supplement Series* 1: 479–481**
- **468. Butler, J. M., B. Budowle, P. Gill, K. K. Kidd, C. Phillips, P. M. Schneider, P. M. Vallone, and N. Morling, 2008. *Report on ISFG SNP Panel Discussion. Progress in Forensic Genetics: Genetics Supplement Series* 1: 471–472**

Technology Transition Workshop

NIJ National Institute of Justice

Additional Information on Our Forensic Research to Date

- **See the Microsoft® PowerPoint® versions of relevant posters and talks in the "Library" section of the Kidd Lab website:**
  - **http://info.med.yale.edu/genetics/kkidd/**
- **Other publications from our laboratory can be found under "Publications" on that website**

Technology Transition Workshop NIJ National Institute of Justice

Populations Studied (Sample Sizes)

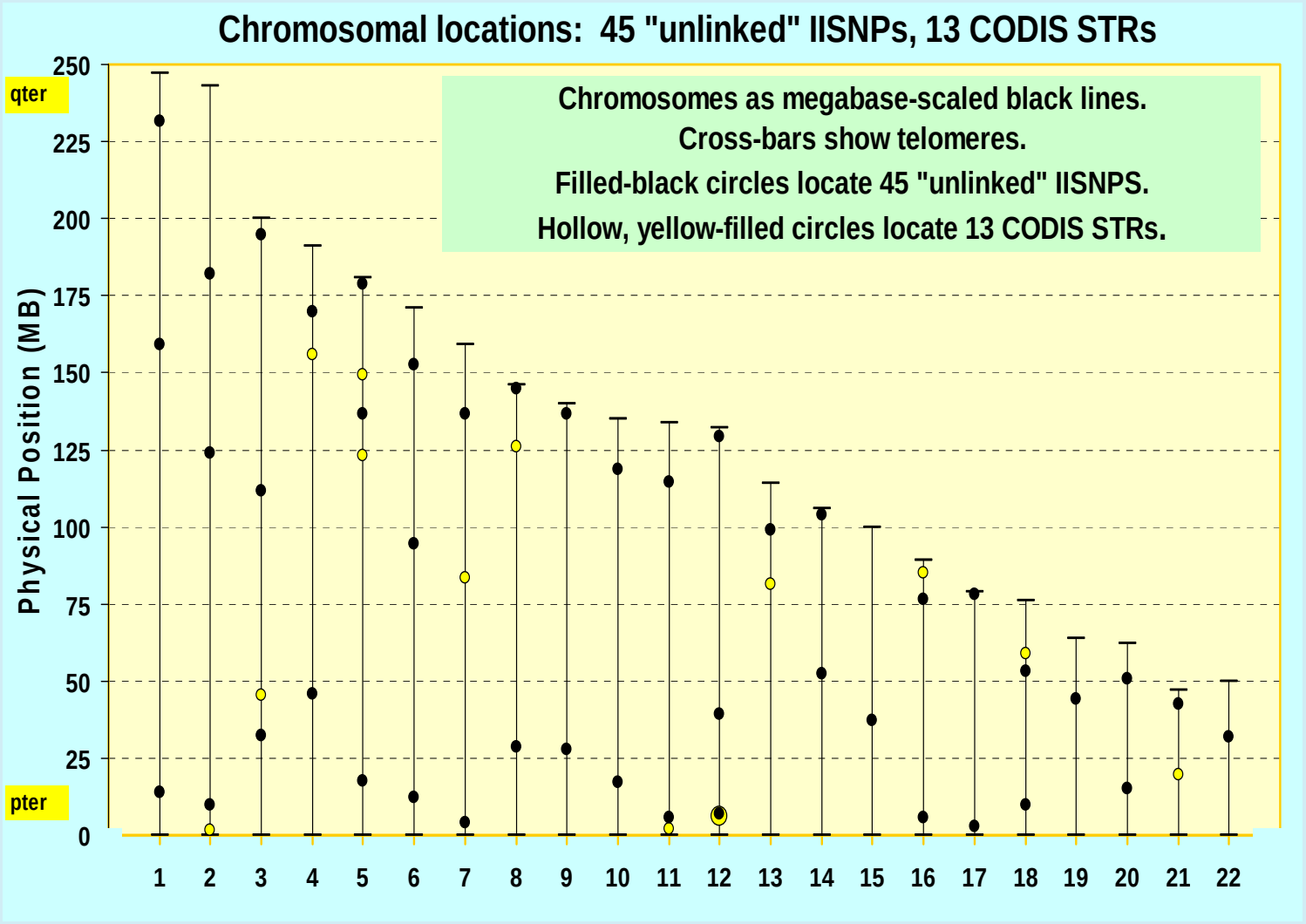
Africa	S.W. Asia	Eurpoe	N. Asia
Biaka (70)	Yemenites (43)	Adygei (54)	Komi Zyrian (47)
Mbuti (39)	Druze (106)	Chuvash (42)	Kyanty (50)
Yoruba (78)	Samaritans (41)	Russians	Yakut (51)
Ibo (48)	Ashkenazi (83)	Archangelsk (34)	
Hausa (39)		Vologda (48)	**S.C. Asia**
Masai (22)		Hungarians (92)	Keralites (30)
Chagga (45)		Finns (36)	
Sandawe (40)		Danes (51)	
Ethiopians (32)		Irish (118)	
African Americans (90)		EuroAmericans (92)	

Technology Transition Workshop

Populations Studied (Sample Sizes)

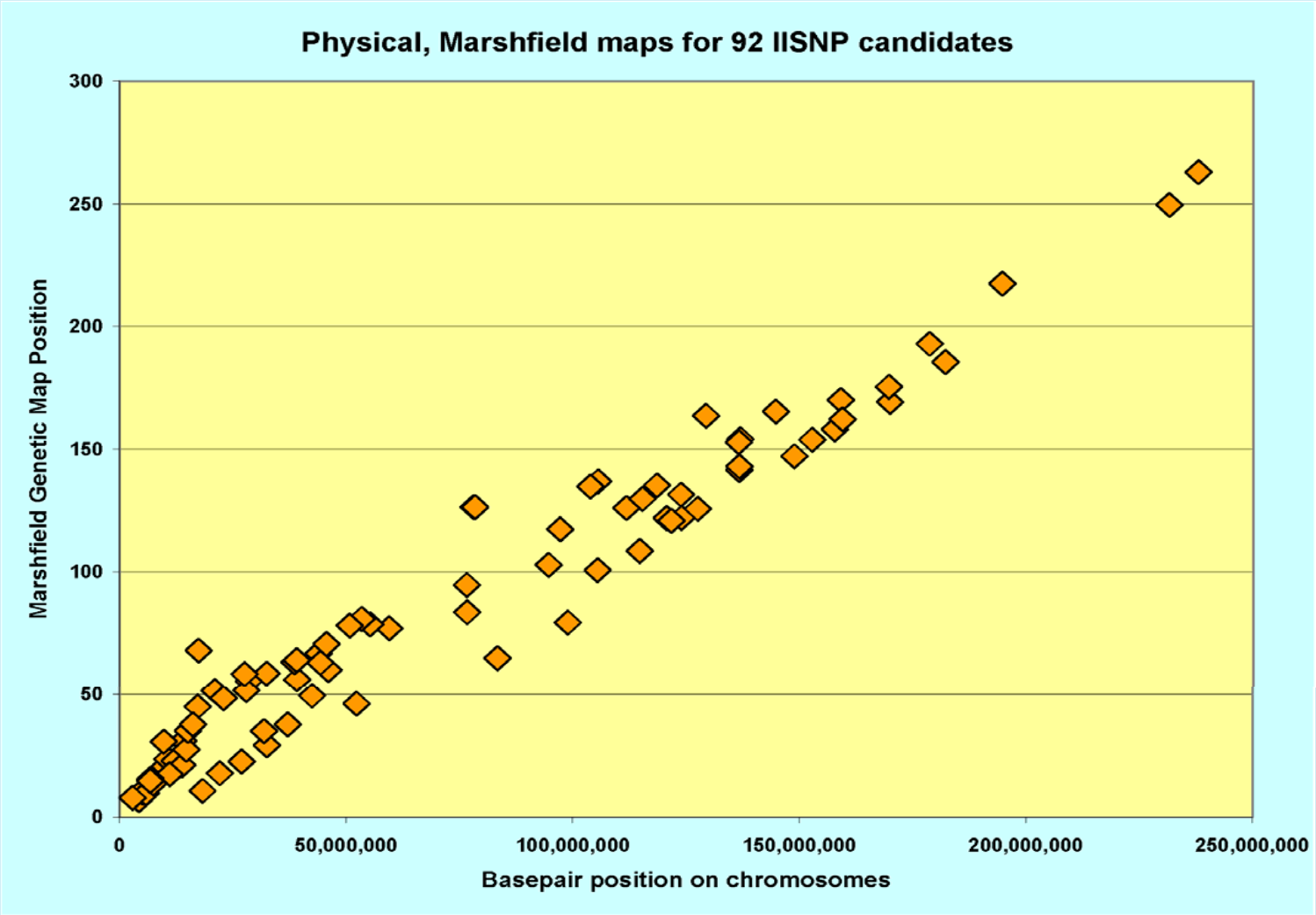
Pacific Islands	East Asia	Americas
Nasioi (23)	Chinese, SF (60)	Pima, Mexico (53)
Micronesians (37)	Chinese, TW (49)	Maya (52)
	Hakka (41)	Quechua (22)
	Koreans (54)	Ticuna (65)
	Japanese (51)	Rondonian Surui (47)
	Ami (40)	Karitiana (57)
	Atayal (42)	
	Cambodians (25)	
	Laotians (119)	

Technology Transition Workshop



Technology Transition Workshop

NIJ National Institute of Justice



Technology Transition Workshop

NIJ National Institute of Justice