Introduction to Population Genetics

Lynn B. Jorde
Department of Human Genetics
University of Utah School of Medicine
7 March 2012

JOHNS HOPKINS MEDICINE
CONTINUING MEDICAL EDUCATION

Current Topics in Genome Analysis 2012

Lynn Jorde

No Relevant Financial Relationships with Commercial Interests

Overview

- Patterns of human genetic variation
  - Among populations
  - Among individuals
- "Race" and its biomedical implications
- Linkage disequilibrium, the HapMap project, and the 1000 Genomes project

Human Genetic Variation: Applications

- Deciphering human history
- Inferring individual ancestry
- Forensics
- Finding and understanding disease-causing genes

Mutation and Genetic Variation

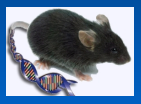
Human mutation rate is 1.0 – 2.5 x 10-8 per bp per generation: we transmit 30-75 new DNA variants with each gamete (J. Roach *et al*., 2010, *Science;* D. Conrad *et al*., 2011, *Nature Genetics*)

"*The capacity to blunder slightly is the real marvel of DNA. Without this special attribute, we would still be anaerobic bacteria and there would be no music.*"

- Lewis Thomas

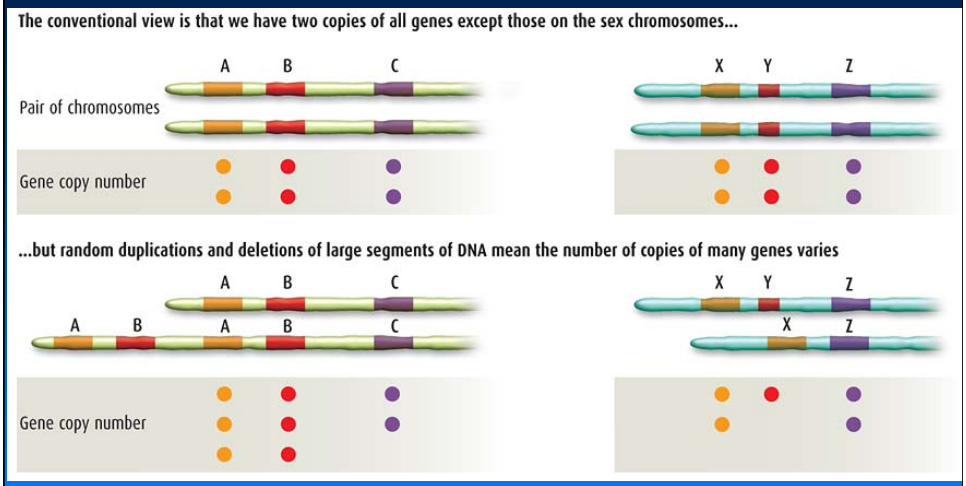
How much do we differ?

(number of aligned DNA base differences)

- Identical twins  0
- Unrelated humans  1/1,000
- Human vs. chimp  1/100
- Human vs. mouse  1/6 - 1/3

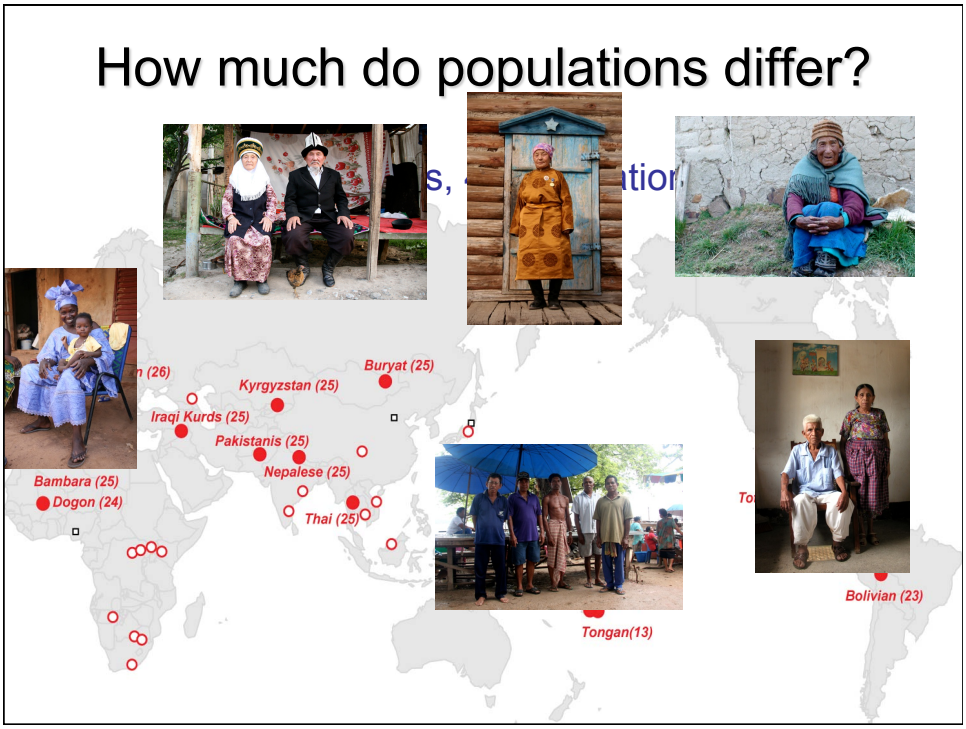
- 3 billion DNA bases → 3 million differences (single nucleotide polymorphisms; SNPs) between each pair of haploid human DNA sequences

Copy number variants (deletions/duplications > 1000 bp) account for as much variation as do single-nucleotide variants



Each human is heterozygous for at least 100 CNVs (3 Mb).

How much do populations differ?



Allele frequencies in populations

Population	SNP 1	SNP 2	SNP 3
1	0.588	0.890	0.880
2	0.671	0.559	0.528
3	0.792	0.790	0.828

1/1000 bp varies between a pair of individuals: how is this variation distributed between continents?

$$F_{ST} = \frac{H_T - \bar{H}_S}{H_T}$$

F_{ST} is the amount of genetic variation that is due to population differences

H_T is the total heterozygosity (variation) in the sample

H_S is the average heterozygosity within each population (continent)

F_{ST} = 0: All variation exists within populations; none exists between

F_{ST} = 1: All variation exists between populations

1/1000 bp varies between individuals: how is this variation distributed among continents?

	60 STRs	30 RSPs	100 *Alus*	75 L1s	250K SNP	
Between individuals, within continents	90%	87%	86%	88%	88%	
Between continents (F_{ST})	10%	13%	14%	12%	12%	

Jorde *et al.*, 2000, *Am. J. Hum. Genet.*
Xing *et al.*, 2009, *Genome Res.*

1/1000 bp varies between individuals: how is this variation distributed among continents?

	60 STRs	30 RSPs	100 *Alus*	75 L1s	250K SNP	Skin Color
Between individuals, within continents	90%	87%	86%	88%	88%	10%
Between continents (F_{ST})	10%	13%	14%	12%	12%	90%

Jorde *et al.*, 2000, *Am. J. Hum. Genet.*
J. Xing *et al.*, 2009, *Genome Res.*

% SNP minor alleles shared among four major regions (Africa, Europe, E. Asia, India): 250K chip results

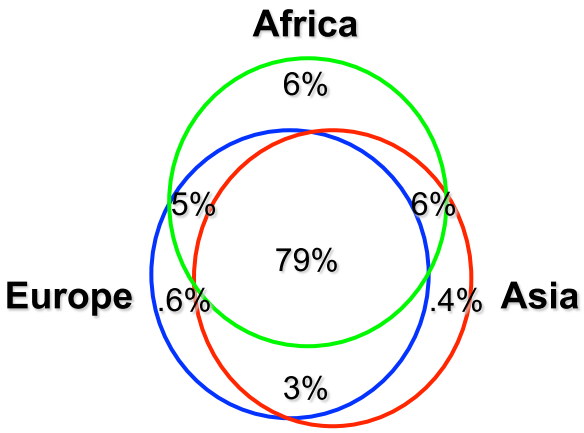
Minor allele present in:	
All 4 groups	78.6%
At least 3 groups	88.0%
At least 2 groups	92.1%
Africa only	7.4%
Any non-African group	0.5%

No SNPs were fixed present in one population, fixed absent in another

J. Xing *et al.*, 2010, *Genome Res.*

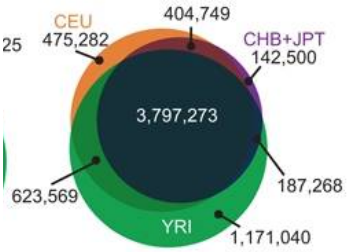
Most common genetic variants are shared among populations:

7,742 SNPs >.05 in ENCODE database



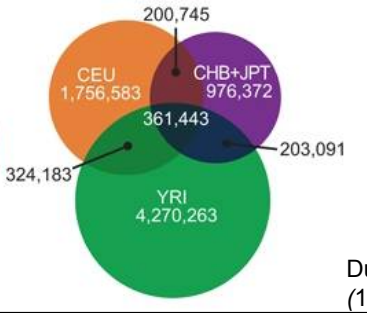
Rare SNPs are much more likely to be population-specific

Common SNPs previously identified in dbSNP



Average allele frequency difference between populations: 15%

New rarer SNPs identified by sequencing



Rare alleles (frequency < 5%) don't affect F_{ST} very much.

Durbin *et al.*, 2010, *Nature* (1000 Genomes Project)

A simple genetic distance measure

$$D_{ij} = |p_i - p_j|$$

D_{ij} is the genetic distance between populations i and j; p_i and p_j are the allele frequencies of a SNP in populations i and j.

Pop.	SNP 1	SNP 2	SNP 3
1	0.588	0.890	0.880
2	0.671	0.559	0.528
3	0.792	0.790	0.828

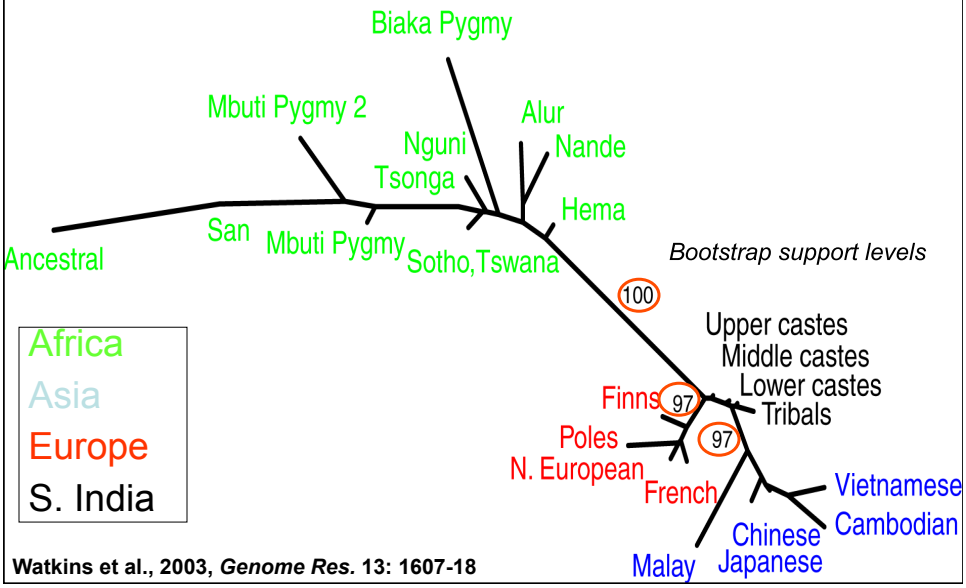
$D_{12} = |0.588 - 0.671| = 0.083$ (avg. over all SNPs)

Building a population network

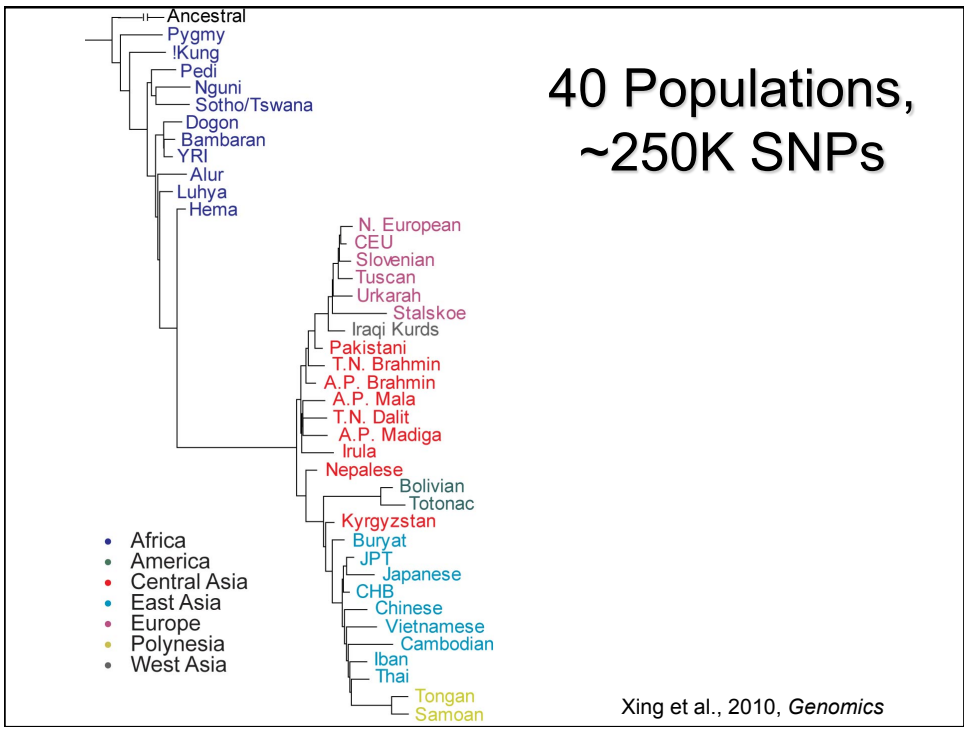
Pop.	SNP 1
1	0.588
2	0.671
3	0.792

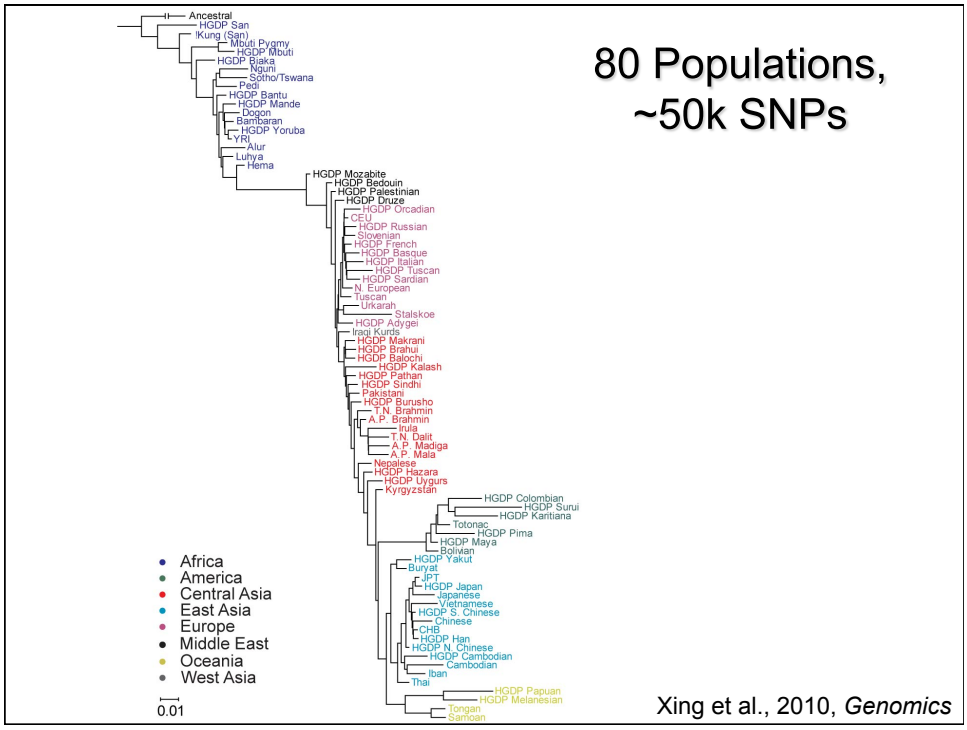
$|p_1 - p_2|$ $| p_3 - (p_1 + p_2)/2 |$

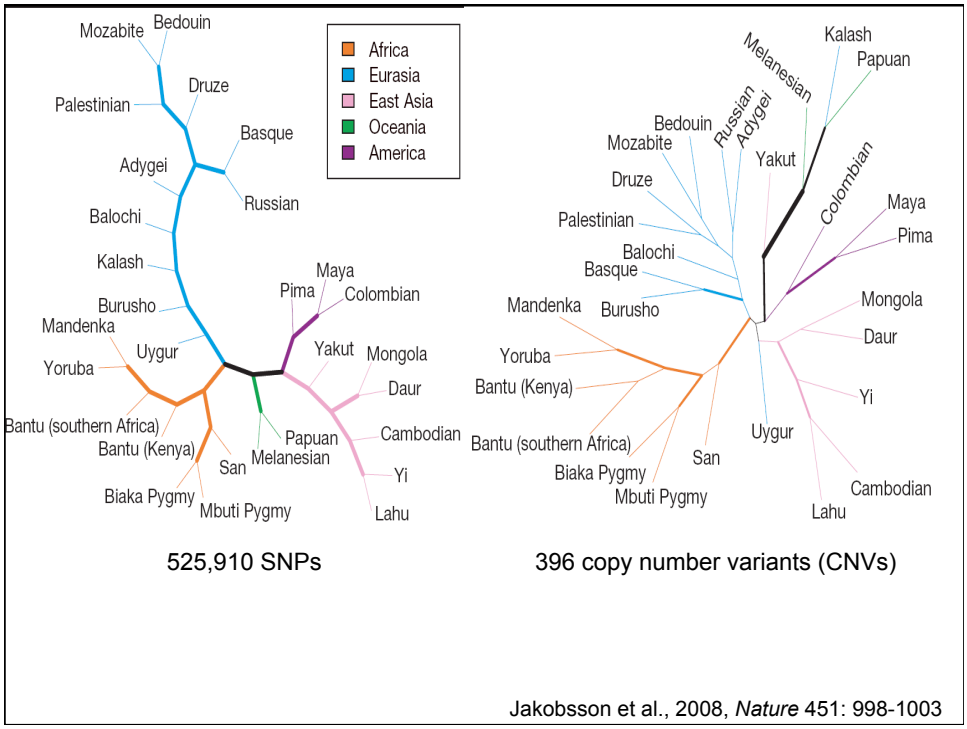
Genetic relationships based on 100 autosomal *Alu* polymorphisms



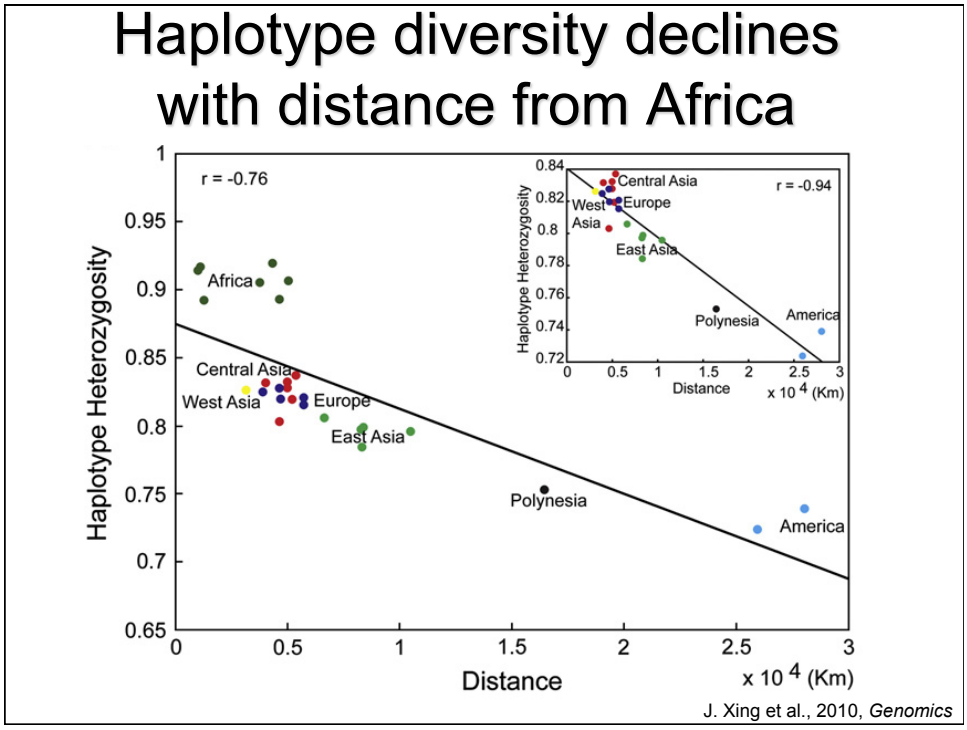
Watkins et al., 2003, *Genome Res.* 13: 1607-18


Ancestral
Pygmy
!Kung
Pedi
Nguni
Sotho/Tswana
Dogon
Bambaran
YRI
Alur
Luhya
Hema
N. European
CEU
Slovenian
Tuscan
Urkarah
Stalskoe
Iraqi Kurds
Pakistani
T.N. Brahmin
A.P. Brahmin
A.P. Mala
T.N. Dalit
A.P. Madiga
Irula
Nepalese
Bolivian
Totonac
Kyrgyzstan
Buryat
JPT
Japanese
CHB
Chinese
Vietnamese
Cambodian
Iban
Thai
Tongan
Samoan
Africa
America
Central Asia
East Asia
Europe
Polynesia
West Asia
40 Populations,
~250K SNPs
Xing et al., 2010, *Genomics*


Ancestral
HGDP San
!Kung (San)
Mbuti Pygmy
HGDP Mbuti
HGDP Biaka
Nguni
Sotho/Tswana
Pedi
HGDP Bantu
HGDP Mande
Dogon
Bambaran
HGDP Yoruba
YRI
Alur
Luhya
Hema
HGDP Mozabite
HGDP Bedouin
HGDP Palestinian
HGDP Druze
HGDP Orcadian
CEU
HGDP Russian
Slovenian
HGDP French
HGDP Basque
HGDP Italian
HGDP Tuscan
HGDP Sardian
N. European
Tuscan
Urkarah
Stalskoe
HGDP Adygei
Iraqi Kurds
HGDP Makrani
HGDP Brahui
HGDP Balochi
HGDP Kalash
HGDP Pathan
HGDP Sindhi
Pakistani
HGDP Burusho
T.N. Brahmin
A.P. Brahmin
Irula
T.N. Dalit
A.P. Madiga
A.P. Mala
Nepalese
HGDP Hazara
HGDP Uygurs
Kyrgyzstan
HGDP Colombian
HGDP Surui
HGDP Karitiana
Totonac
HGDP Pima
HGDP Maya
Bolivian
HGDP Yakut
Buryat
JPT
HGDP Japan
Japanese
Vietnamese
HGDP S. Chinese
Chinese
CHB
HGDP Han
HGDP N. Chinese
HGDP Cambodian
Cambodian
Iban
Thai
HGDP Papuan
HGDP Melanesian
Tongan
Samoan
Africa
America
Central Asia
East Asia
Europe
Middle East
Oceania
West Asia
0.01
80 Populations,
~50k SNPs
Xing et al., 2010, *Genomics*



Jakobsson et al., 2008, *Nature* 451: 998-1003

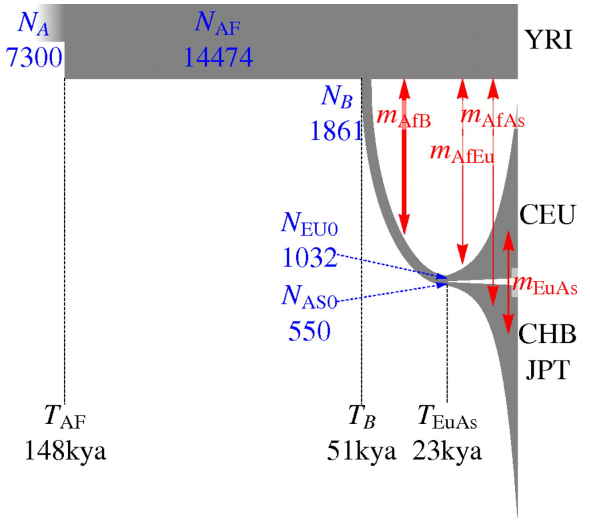


J. Xing et al., 2010, *Genomics*

Sequence data permit more accurate inferences about population history

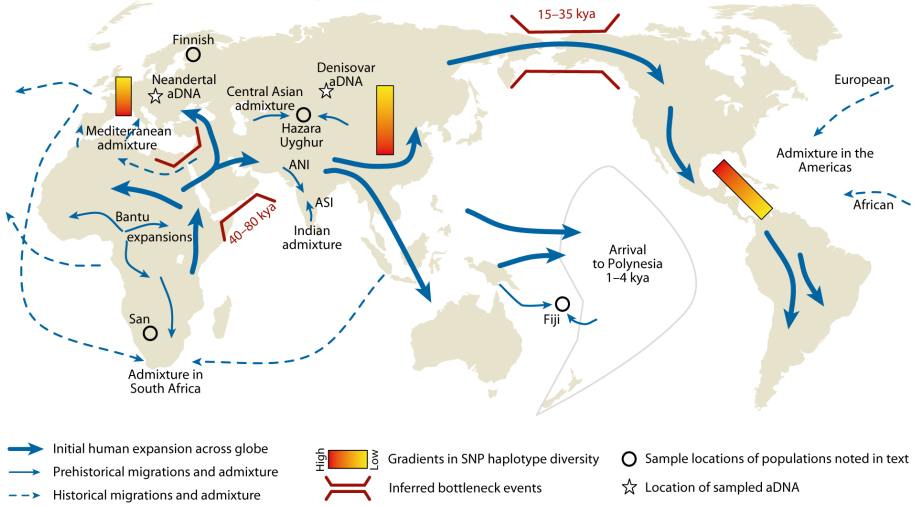
- Microarray SNPs are selected for higher frequency and diversity in Europeans
- Complete DNA sequences are unbiased and include information about rare variants

An inferred demographic model, with line width corresponding to population size and time flowing from left to right.



Gravel S et al. PNAS 2011;108:11983-11988

Recent African origin of anatomically modern humans



Novembre J, Ramachandran S. 2011.
Annu. Rev. Genomics Hum. Genet. 12:245–74

Evidence for mixture between Neanderthals and modern humans

- mtDNA sequences from 12 Neanderthal skeletons show no evidence for recent shared polymorphisms
- Evidence for mixture from nuclear sequence (1.3x coverage): 1-4% of modern human DNA has Neanderthal origins (Green et al., 2010, *Science*)
  - Only non-Africans share DNA with Neanderthals
  - Neanderthal DNA sharing is seen in all non-African populations
  - Could some of the shared sequences have adaptive significance?

What can genetics tell us about "race"?

"'Race' is biologically meaningless"
-- Schwartz, 2001, *N. Engl. J. Med.*

"I am a racially profiling doctor"
-- Satel, May 5, 2002, *New York Times*

"These [genetic] data also show that any two individuals within a particular population are as different genetically as any two people selected from any two populations in the world."
-- American Anthropological Association, 1997

Bamshad and Olson, 2003

Tabulation of DNA sequence differences among individuals

TTGCAGCTCTCC
TTGCAGCTCTCC

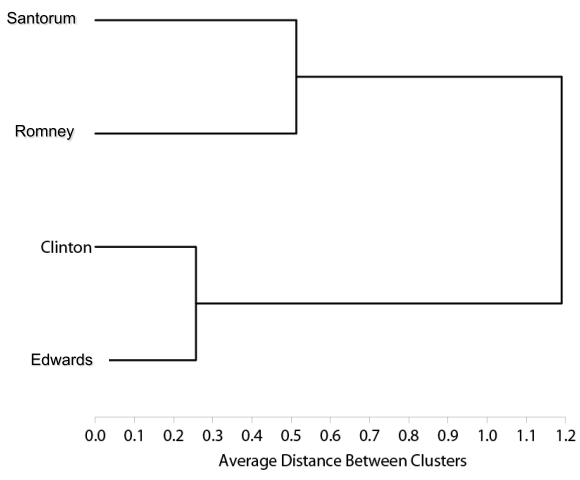
TTGCAGCTCTCC
ATGCAGCTCTCG

ATGCAGCTCTCG
ATGCTGCTCTCG

ATGCTGCTCTCG
ATGCTGCTCTCG

	Santorum	Romney	Clinton	Edwards
Santorum	0	.	.	.
Romney	2	0	.	.
Clinton	5	3	0	.
Edwards	6	4	1	0

DNA differences can be summarized in a "tree"

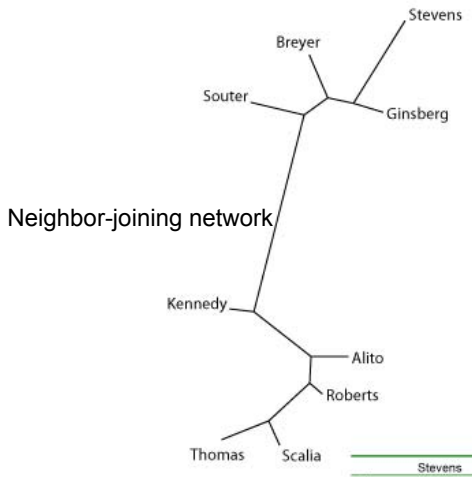


A distance matrix based on Supreme Court decisions

Distance matrix: % disagreement

	Stevens	Ginsberg	Souter	Breyer	Kennedy	Alito	Roberts	Scalia	Thomas
Stevens	0								
Ginsberg	15	0							
Souter	26	15	0						
Breyer	19	13	15	0					
Kennedy	45	36	34	35	0				
Alito	56	48	44	45	13	0			
Roberts	55	49	40	48	19	8	0		
Scalia	59	52	50	58	28	19	11	0	
Thomas	64	55	53	60	29	21	15	9	0

Thanks to: Steve Guthery, MD

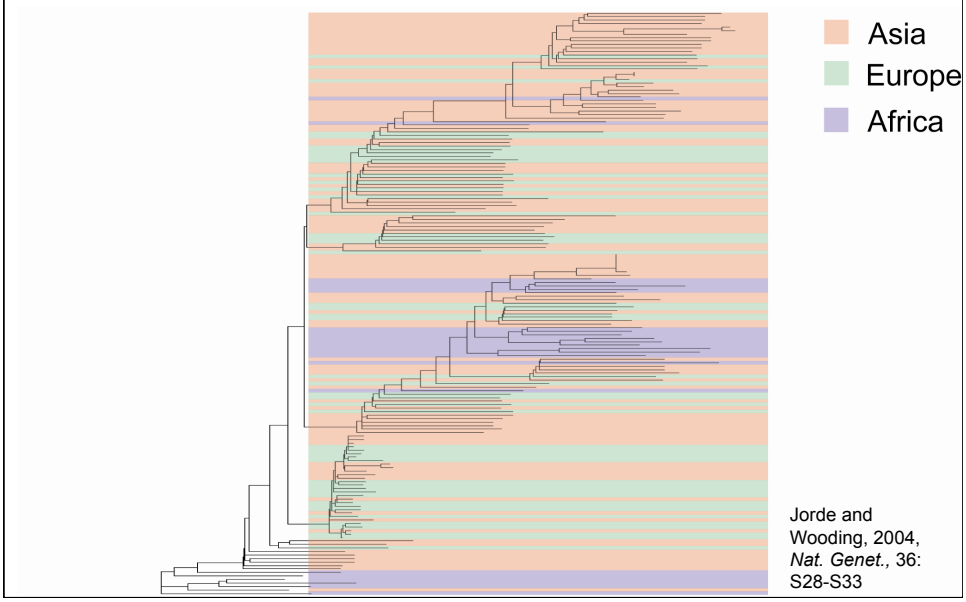


Distance matrix: % disagreement

	Stevens	Ginsberg	Souter	Breyer	Kennedy	Alito	Roberts	Scalia	Thomas
Stevens	0								
Ginsberg	15	0							
Souter	26	15	0						
Breyer	19	13	15	0					
Kennedy	45	36	34	35	0				
Alito	56	48	44	45	13	0			
Roberts	55	49	40	48	19	8	0		
Scalia	59	52	50	58	28	19	11	0	
Thomas	64	55	53	60	29	21	15	9	0

Thanks to: Steve Guthery, MD

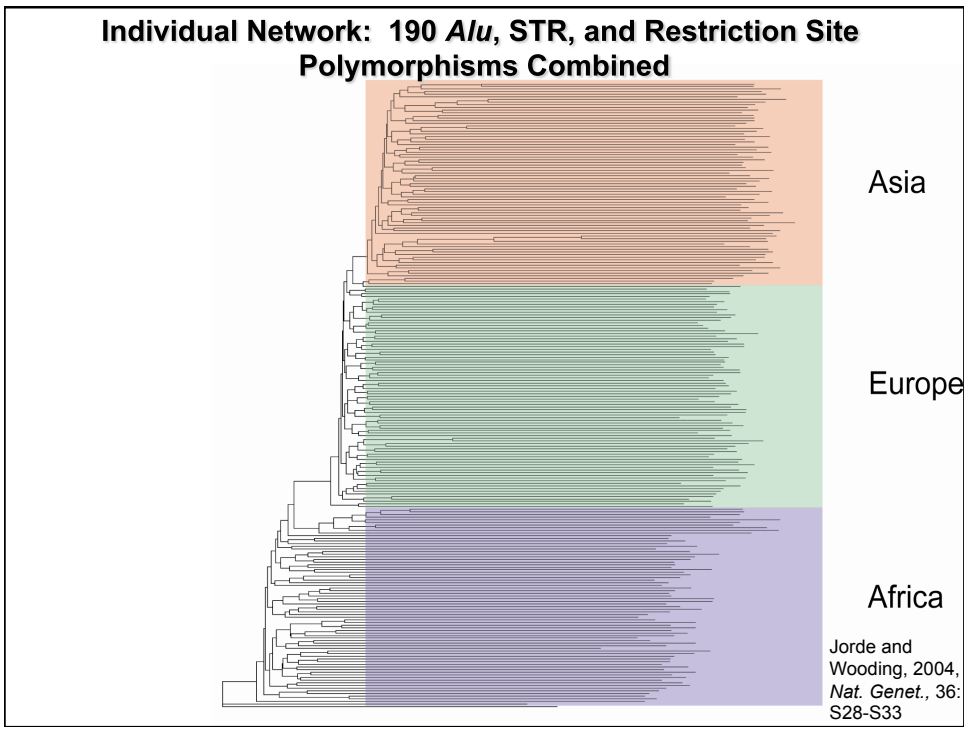
Individual network: 14 kb sequence in angiotensinogen gene

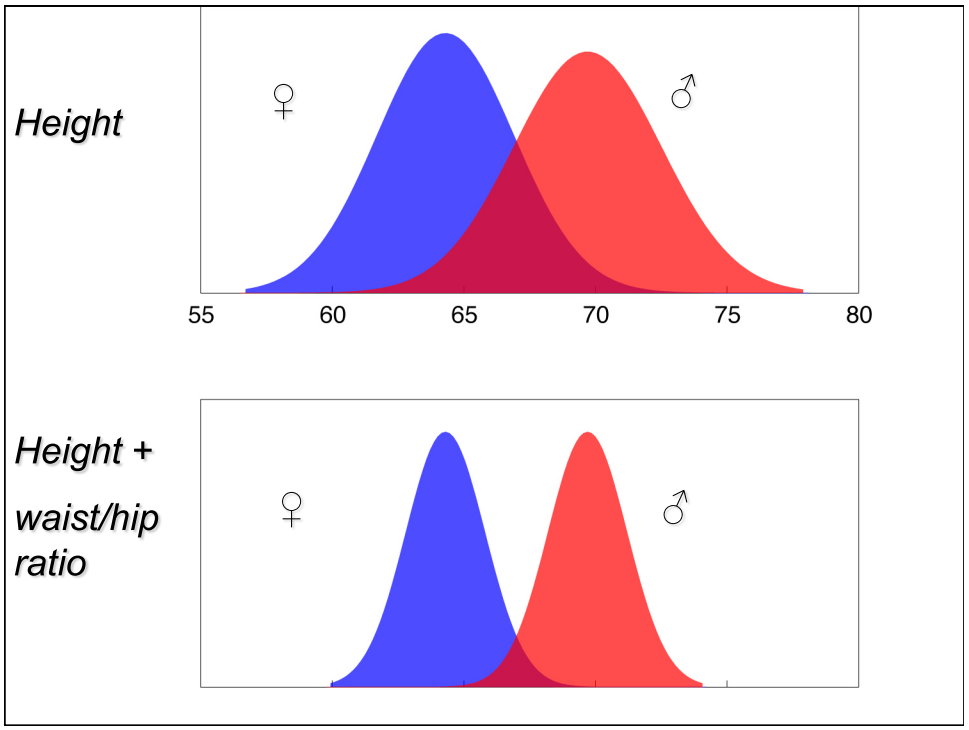


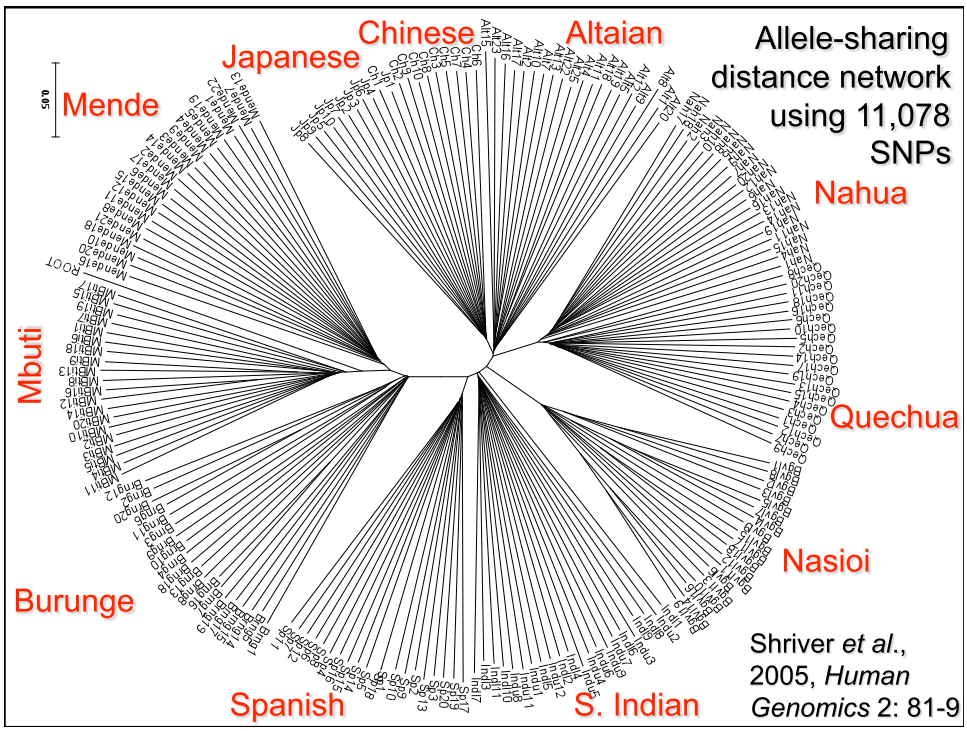
Jorde and Wooding, 2004, *Nat. Genet.*, 36: S28-S33

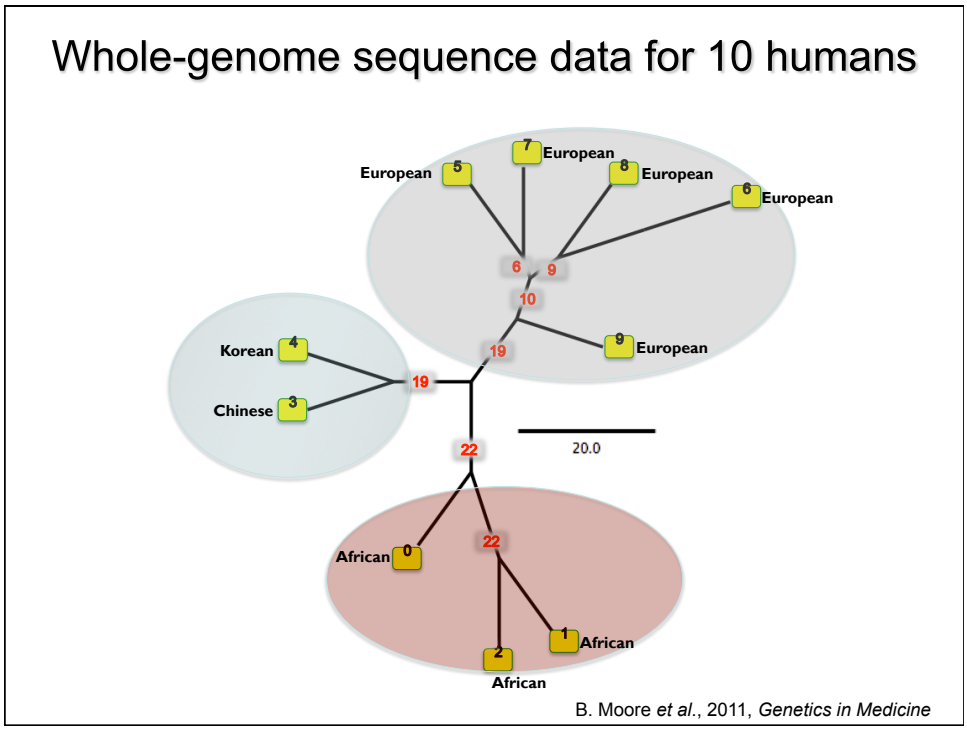
It may be doubted whether any character can be named which is distinctive of a race and is constant."

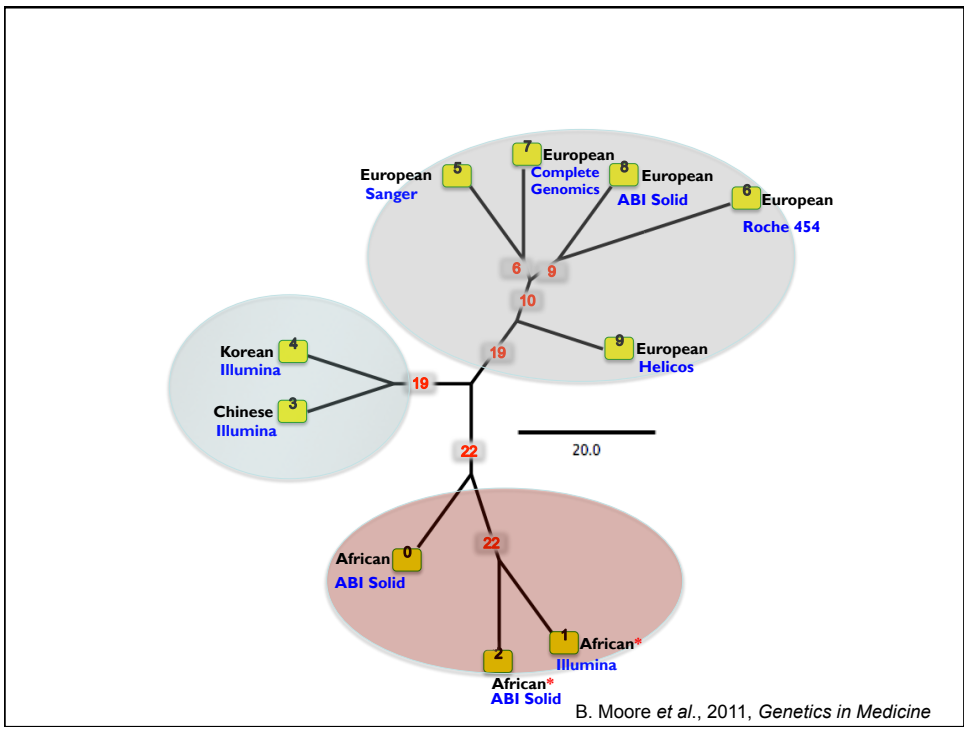
-- Charles Darwin, 1871, *The Descent of Man, and Selection in Relation to Sex*


Individual Network: 190 Alu, STR, and Restriction Site Polymorphisms Combined
Asia
Europe
Africa
Jorde and Wooding, 2004, Nat. Genet., 36: S28-S33


Height
♀
♂
55
60
65
70
75
80
Height + waist/hip ratio
♀
♂

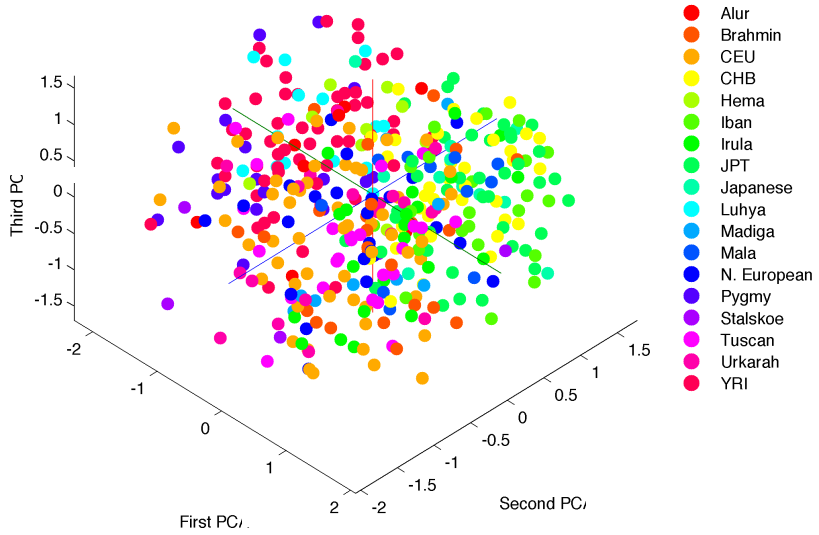

Allele-sharing distance network using 11,078 SNPs
Chinese
Altaian
Japanese
Mende
Nahua
Mbuti
Quechua
Nasioi
Burunge
Spanish
S. Indian
0.05
Shriver *et al.*, 2005, *Human Genomics* 2: 81-9


Whole-genome sequence data for 10 humans
European
European
European
European
Korean
Chinese
African
African
African
20.0
B. Moore *et al.*, 2011, *Genetics in Medicine*

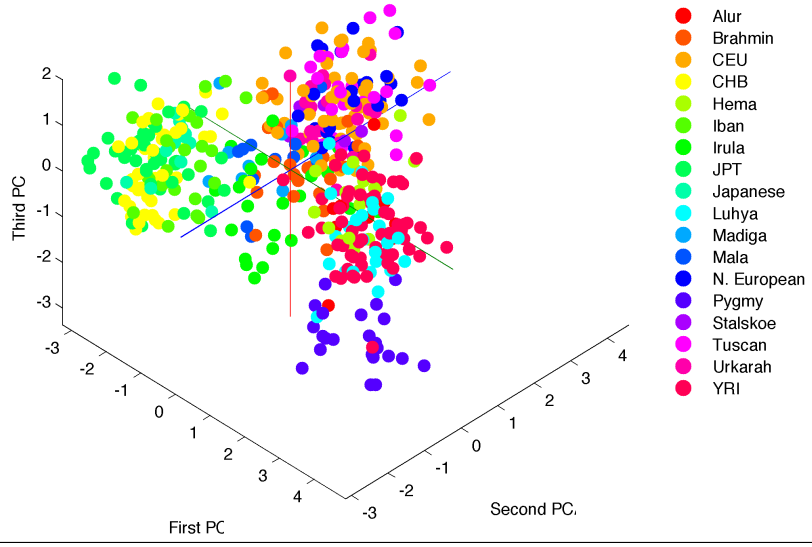


B. Moore *et al.*, 2011, *Genetics in Medicine*

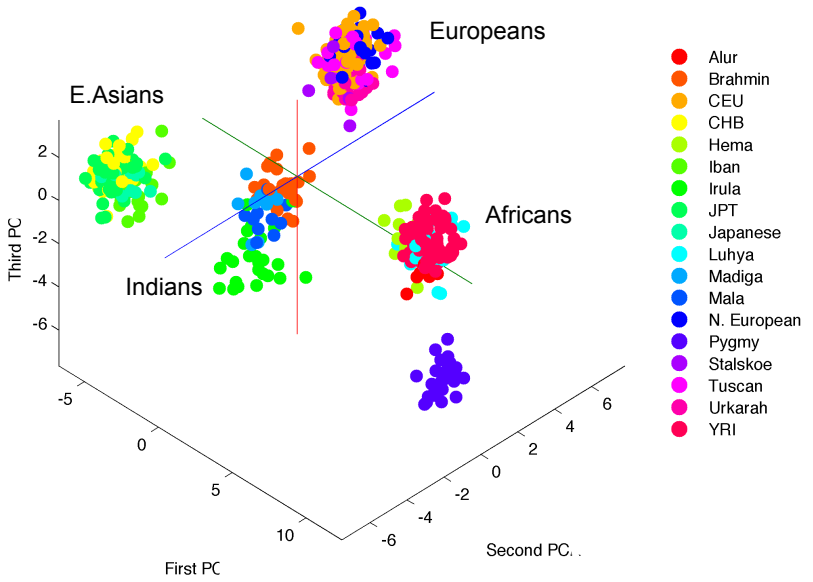
Genetic distances (principal components analysis) among 467 individuals: 10 SNPs



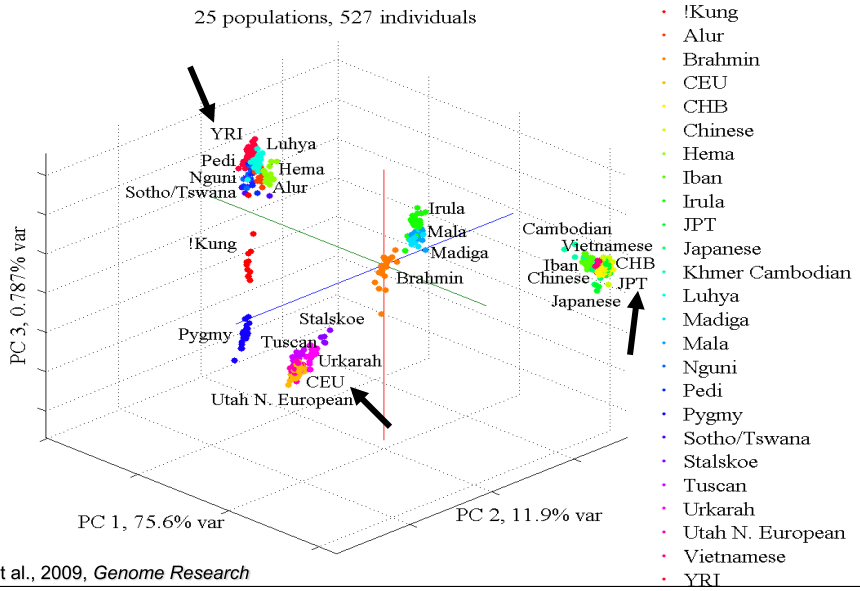
Genetic distances among 467 individuals: 100 SNPs



Genetic distances among 467 individuals: 1000 SNPs

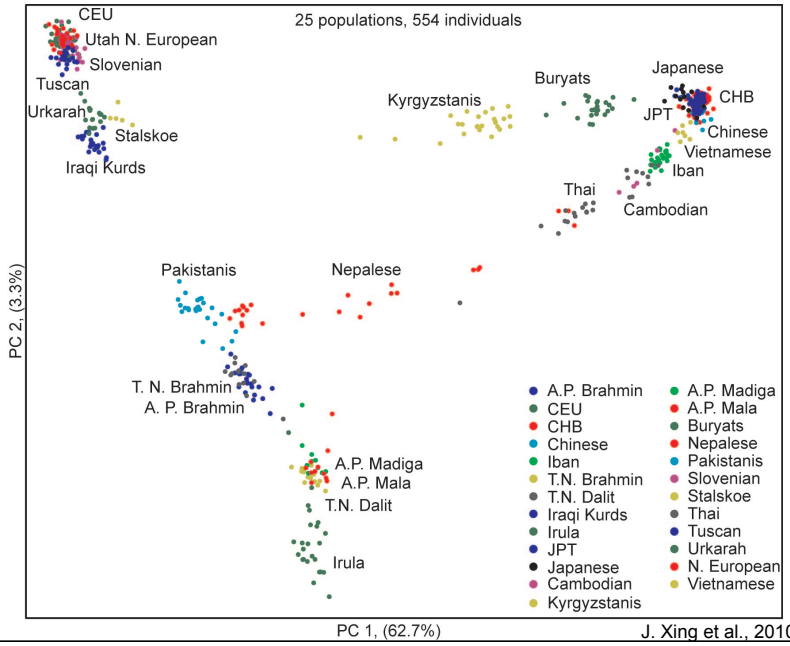


Genetic distances: 250K SNPs



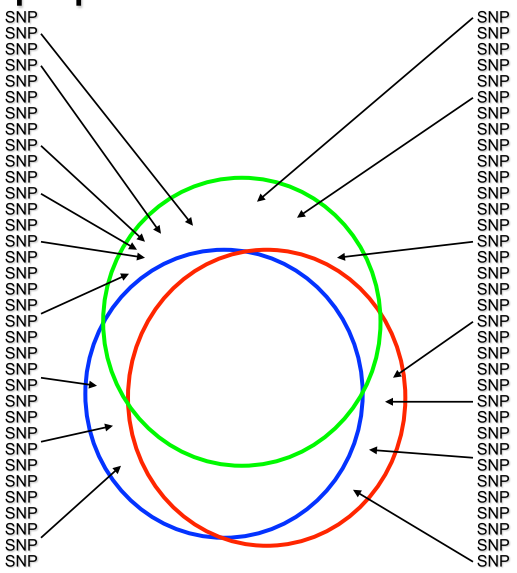
J. Xing et al., 2009, *Genome Research*

Eurasian Populations



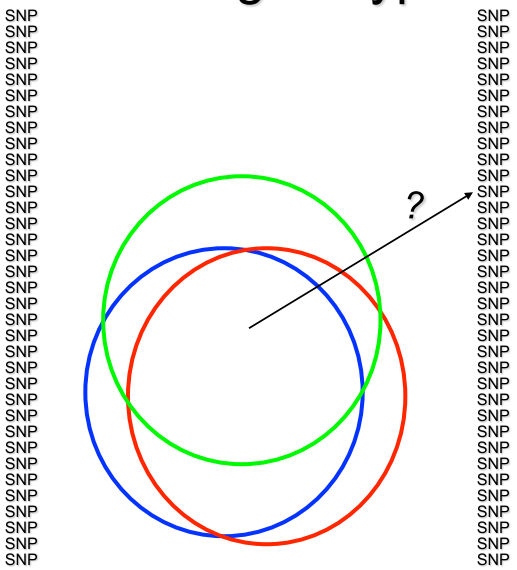
J. Xing et al., 2010, *Genomics*

Multiple polymorphisms can predict population affiliation

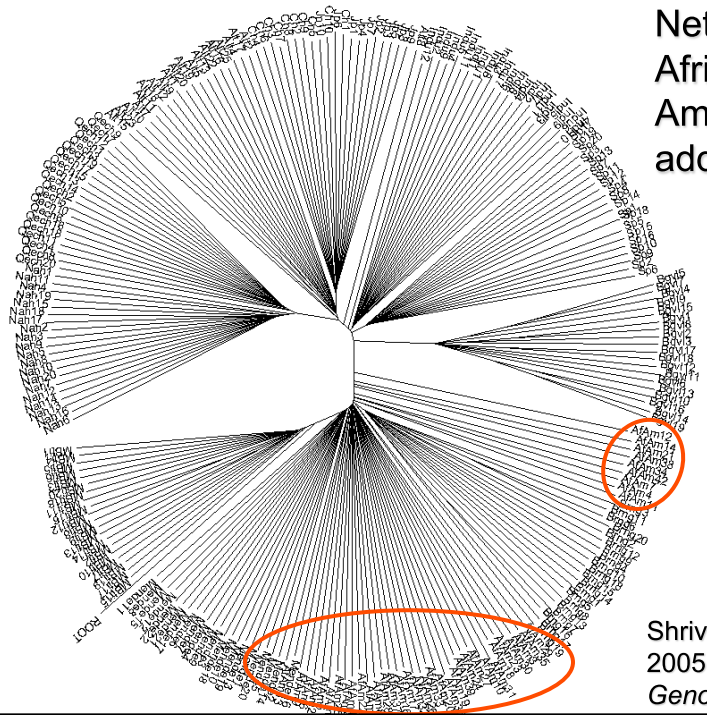


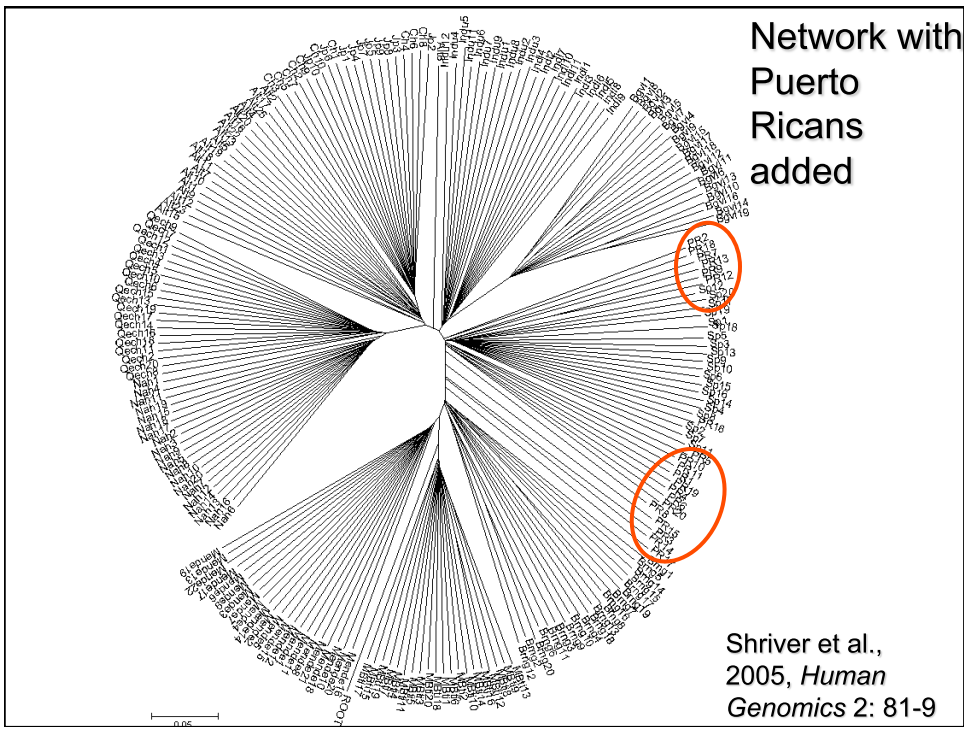
Population affiliation cannot accurately predict individual genotypes or traits

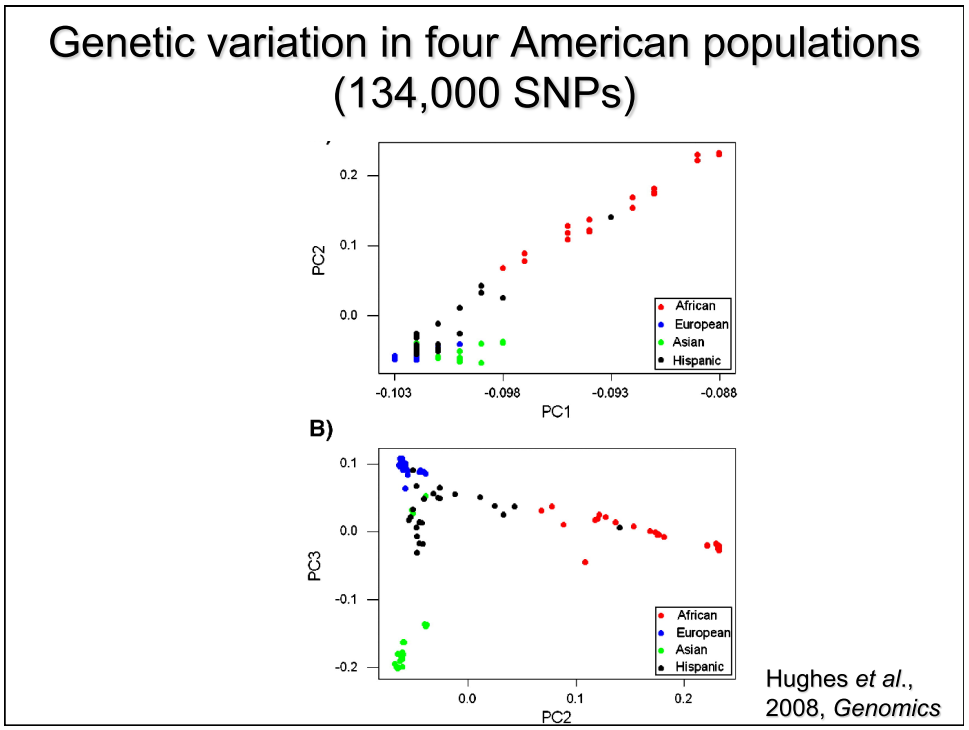
Individual traits or genotypes are shared across populations and differ only in their frequencies



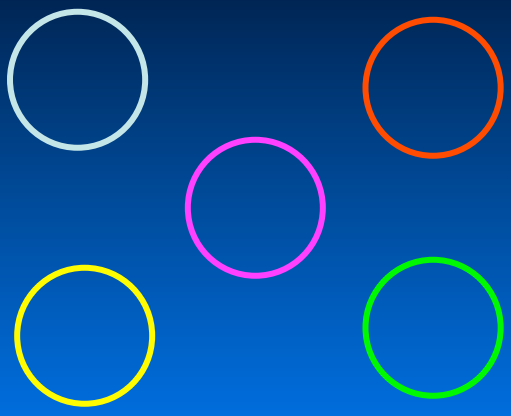
Can we classify everybody?




Network with Puerto Ricans added
Shriver et al., 2005, *Human Genomics* 2: 81-9


Genetic variation in four American populations (134,000 SNPs)
PC2
PC1
African
European
Asian
Hispanic
B)
PC3
PC2
Hughes *et al.*, 2008, *Genomics*

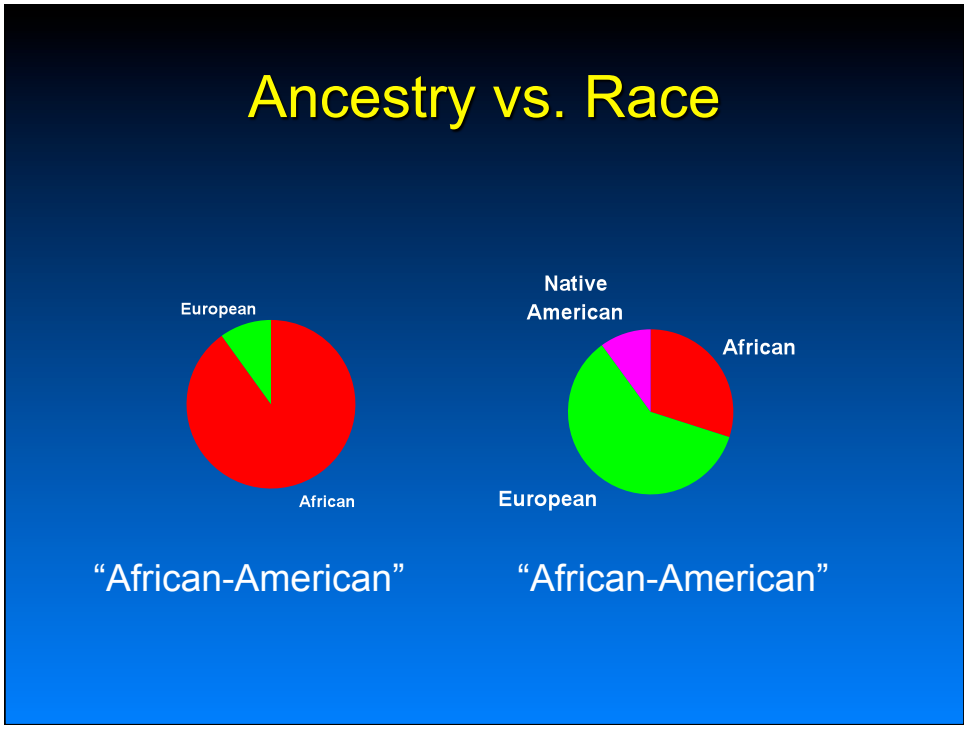
The Fallacy of Typological Thinking

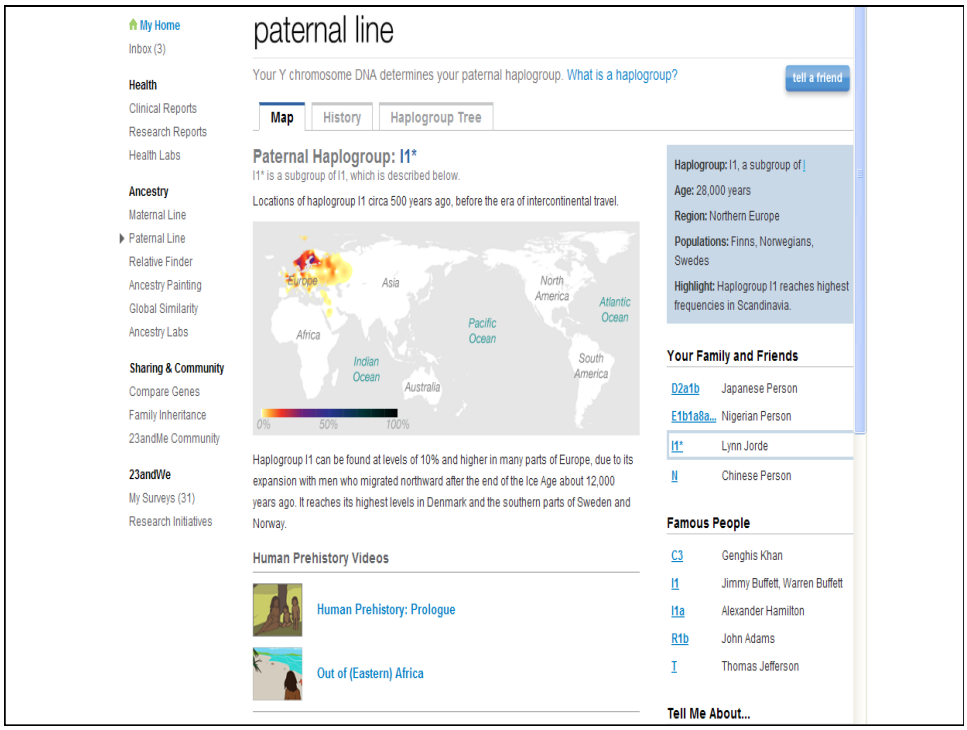


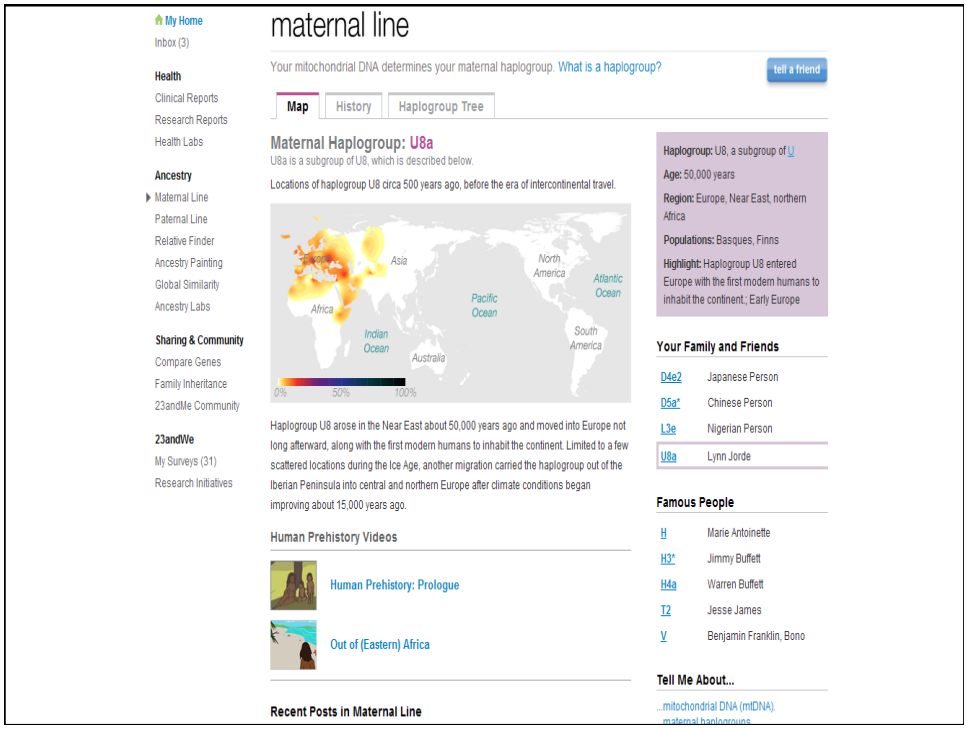
Race as a predictor of ancestry proportions

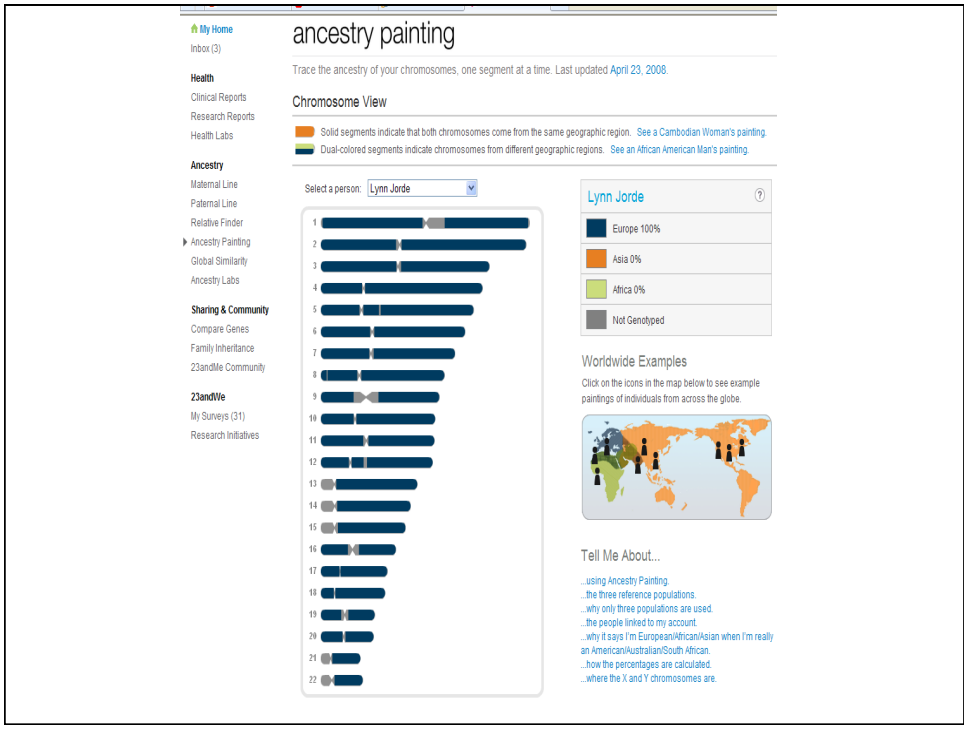
Wayne Joseph

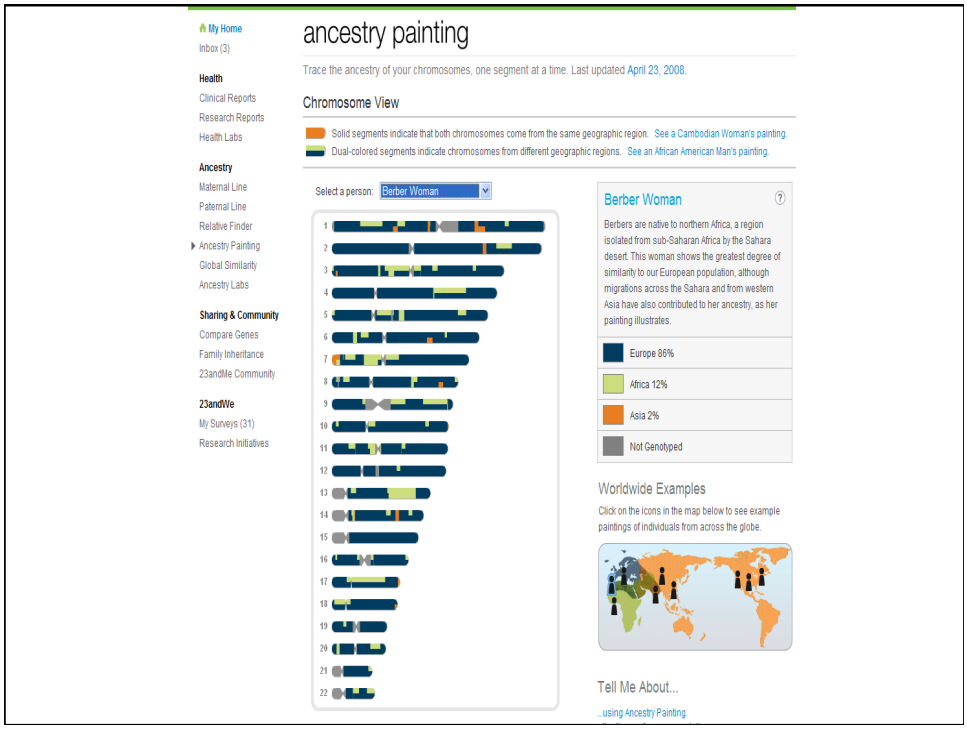
57% European
39% Native American
4% East Asian

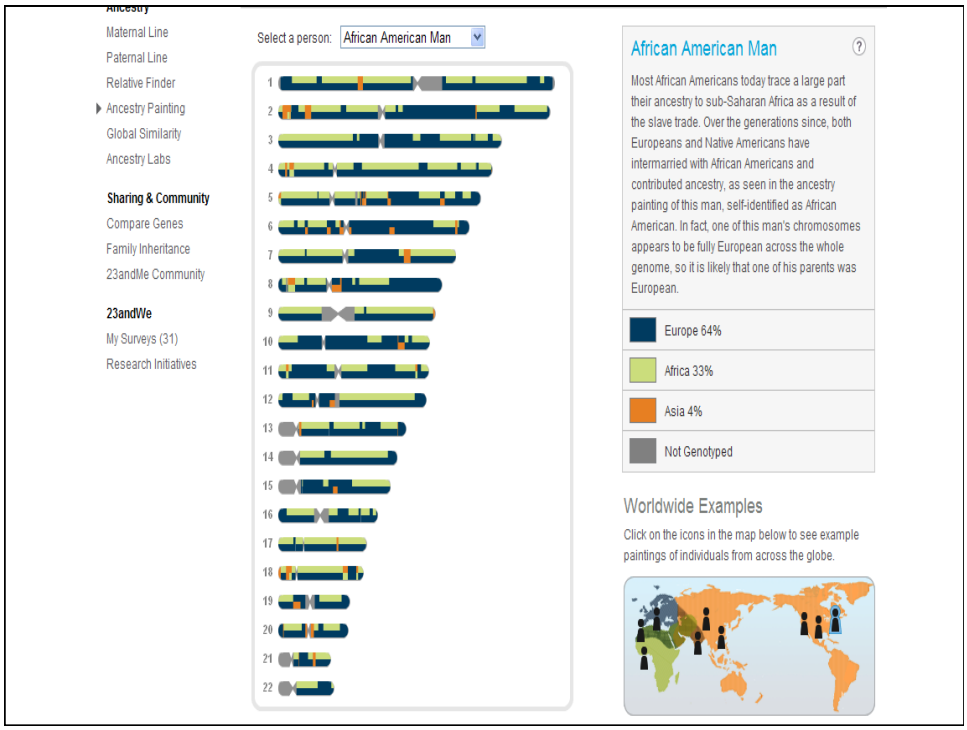

Ancestry vs. Race
European
African
Native American
African
European
"African-American"
"African-American"


My Home
Inbox (3)
Health
Clinical Reports
Research Reports
Health Labs
Ancestry
Maternal Line
Paternal Line
Relative Finder
Ancestry Painting
Global Similarity
Ancestry Labs
Sharing & Community
Compare Genes
Family Inheritance
23andMe Community
23andWe
My Surveys (31)
Research Initiatives
paternal line
Your Y chromosome DNA determines your paternal haplogroup. What is a haplogroup?
tell a friend
Map
History
Haplogroup Tree
Paternal Haplogroup: I1*
I1* is a subgroup of I1, which is described below.
Locations of haplogroup I1 circa 500 years ago, before the era of intercontinental travel.
Europe
Asia
Africa
Indian Ocean
Australia
Pacific Ocean
North America
Atlantic Ocean
South America
0%
50%
100%
Haplogroup I1 can be found at levels of 10% and higher in many parts of Europe, due to its expansion with men who migrated northward after the end of the Ice Age about 12,000 years ago. It reaches its highest levels in Denmark and the southern parts of Sweden and Norway.
Human Prehistory Videos
Human Prehistory: Prologue
Out of (Eastern) Africa
Haplogroup: I1, a subgroup of I
Age: 28,000 years
Region: Northern Europe
Populations: Finns, Norwegians, Swedes
Highlight: Haplogroup I1 reaches highest frequencies in Scandinavia.
Your Family and Friends
D2a1b Japanese Person
E1b1a8a... Nigerian Person
I1* Lynn Jorde
N Chinese Person
Famous People
C3 Genghis Khan
I1 Jimmy Buffett, Warren Buffett
I1a Alexander Hamilton
R1b John Adams
T Thomas Jefferson
Tell Me About...


My Home
Inbox (3)
Health
Clinical Reports
Research Reports
Health Labs
Ancestry
Maternal Line
Paternal Line
Relative Finder
Ancestry Painting
Global Similarity
Ancestry Labs
Sharing & Community
Compare Genes
Family Inheritance
23andMe Community
23andWe
My Surveys (31)
Research Initiatives
maternal line
Your mitochondrial DNA determines your maternal haplogroup. What is a haplogroup?
tell a friend
Map
History
Haplogroup Tree
Maternal Haplogroup: U8a
U8a is a subgroup of U8, which is described below.
Locations of haplogroup U8 circa 500 years ago, before the era of intercontinental travel.
Asia
North America
Atlantic Ocean
Africa
Pacific Ocean
Indian Ocean
South America
Australia
0%
50%
100%
Haplogroup U8 arose in the Near East about 50,000 years ago and moved into Europe not long afterward, along with the first modern humans to inhabit the continent. Limited to a few scattered locations during the Ice Age, another migration carried the haplogroup out of the Iberian Peninsula into central and northern Europe after climate conditions began improving about 15,000 years ago.
Human Prehistory Videos
Human Prehistory: Prologue
Out of (Eastern) Africa
Recent Posts in Maternal Line
Haplogroup: U8, a subgroup of U
Age: 50,000 years
Region: Europe, Near East, northern Africa
Populations: Basques, Finns
Highlight: Haplogroup U8 entered Europe with the first modern humans to inhabit the continent.; Early Europe
Your Family and Friends
D4e2 Japanese Person
D5a* Chinese Person
L3e Nigerian Person
U8a Lynn Jorde
Famous People
H Marie Antoinette
H3* Jimmy Buffett
H4a Warren Buffett
T2 Jesse James
V Benjamin Franklin, Bono
Tell Me About...
...mitochondrial DNA (mtDNA).
...maternal haplogroups.


My Home
Inbox (3)
Health
Clinical Reports
Research Reports
Health Labs
Ancestry
Maternal Line
Paternal Line
Relative Finder
Ancestry Painting
Global Similarity
Ancestry Labs
Sharing & Community
Compare Genes
Family Inheritance
23andMe Community
23andWe
My Surveys (31)
Research Initiatives
ancestry painting
Trace the ancestry of your chromosomes, one segment at a time. Last updated April 23, 2008.
Chromosome View
Solid segments indicate that both chromosomes come from the same geographic region. See a Cambodian Woman's painting.
Dual-colored segments indicate chromosomes from different geographic regions. See an African American Man's painting.
Select a person: Lynn Jorde
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
Lynn Jorde
Europe 100%
Asia 0%
Africa 0%
Not Genotyped
Worldwide Examples
Click on the icons in the map below to see example paintings of individuals from across the globe.
Tell Me About...
...using Ancestry Painting.
...the three reference populations.
...why only three populations are used.
...the people linked to my account.
...why it says I'm European/African/Asian when I'm really an American/Australian/South African.
...how the percentages are calculated.
...where the X and Y chromosomes are.


My Home
Inbox (3)
Health
Clinical Reports
Research Reports
Health Labs
Ancestry
Maternal Line
Paternal Line
Relative Finder
Ancestry Painting
Global Similarity
Ancestry Labs
Sharing & Community
Compare Genes
Family Inheritance
23andMe Community
23andWe
My Surveys (31)
Research Initiatives
ancestry painting
Trace the ancestry of your chromosomes, one segment at a time. Last updated April 23, 2008.
Chromosome View
Solid segments indicate that both chromosomes come from the same geographic region. See a Cambodian Woman's painting.
Dual-colored segments indicate chromosomes from different geographic regions. See an African American Man's painting.
Select a person: Berber Woman
Berber Woman
Berbers are native to northern Africa, a region isolated from sub-Saharan Africa by the Sahara desert. This woman shows the greatest degree of similarity to our European population, although migrations across the Sahara and from western Asia have also contributed to her ancestry, as her painting illustrates.
Europe 86%
Africa 12%
Asia 2%
Not Genotyped
Worldwide Examples
Click on the icons in the map below to see example paintings of individuals from across the globe.
Tell Me About...
...using Ancestry Painting.

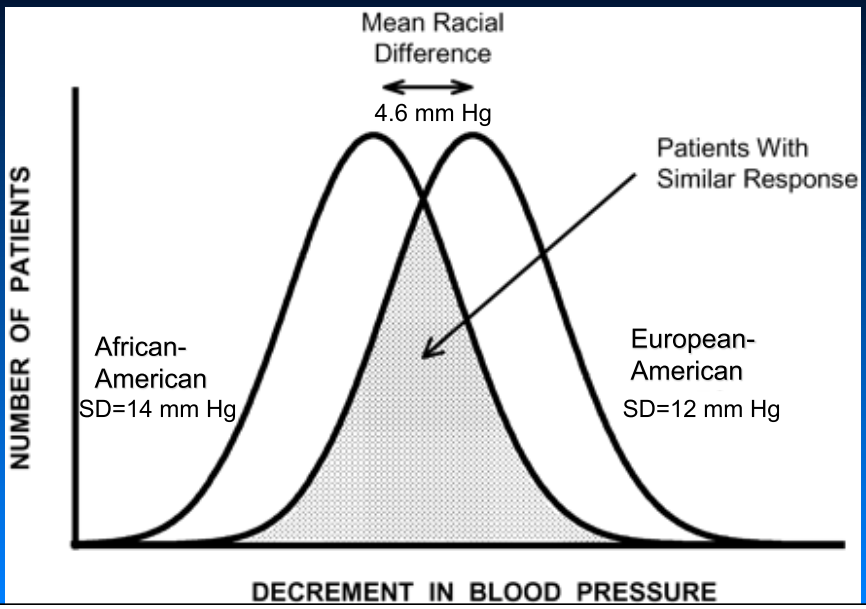

Maternal Line
Paternal Line
Relative Finder
Ancestry Painting
Global Similarity
Ancestry Labs
Sharing & Community
Compare Genes
Family Inheritance
23andMe Community
23andWe
My Surveys (31)
Research Initiatives
Select a person: African American Man
African American Man
Most African Americans today trace a large part their ancestry to sub-Saharan Africa as a result of the slave trade. Over the generations since, both Europeans and Native Americans have intermarried with African Americans and contributed ancestry, as seen in the ancestry painting of this man, self-identified as African American. In fact, one of this man's chromosomes appears to be fully European across the whole genome, so it is likely that one of his parents was European.
Europe 64%
Africa 33%
Asia 4%
Not Genotyped
Worldwide Examples
Click on the icons in the map below to see example paintings of individuals from across the globe.

What do these findings imply for biomedicine?

- Large numbers of independent DNA polymorphisms can inform us about ancestry and population history
- Responses to many therapeutic drugs may involve variation in just a few genes (along with environmental variation)
- These variants typically differ between populations only in their *frequency* and imply substantial overlap between populations

Blood pressure response to ACE inhibitors

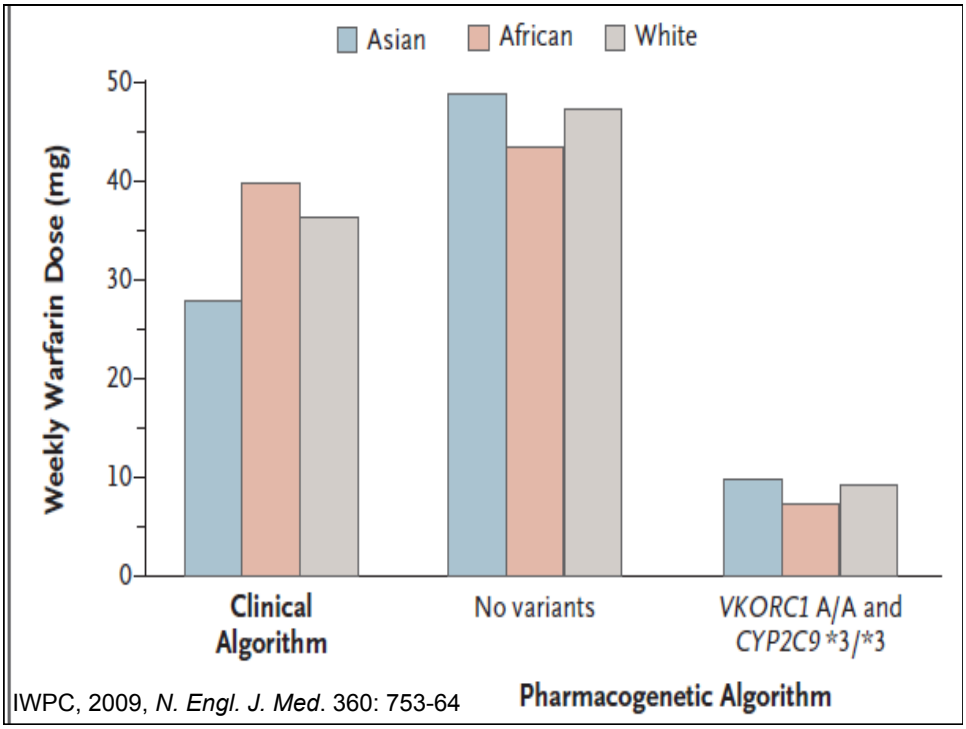
(Sehgal, 2004, *Hypertension* 43: 566-72)



EGFR inhibitors and non-small cell lung cancer

- Gefitinib and erlotinib inhibit epidermal growth factor receptor (EGFR) tyrosine kinase activity
- Effective in 10% of Europeans, 30% of Asians (Japanese, Chinese, Koreans)
- Somatic mutations in *EGFR* found in 10% of Europeans, 30% of Japanese
- 70-80% of those with mutations respond to gefitinib; <10% of those without mutations respond

Johnson, 2005, *Cancer Res.* 65: 7525-9; McDermott et al., *N. Engl. J. Med.* 364: 340-50

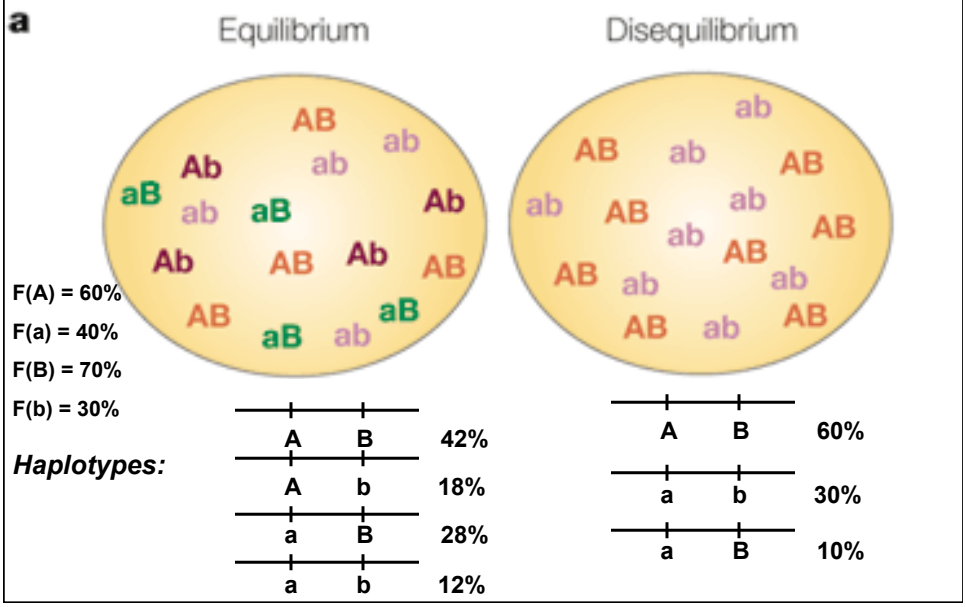


IWPC, 2009, *N. Engl. J. Med.* 360: 753-64

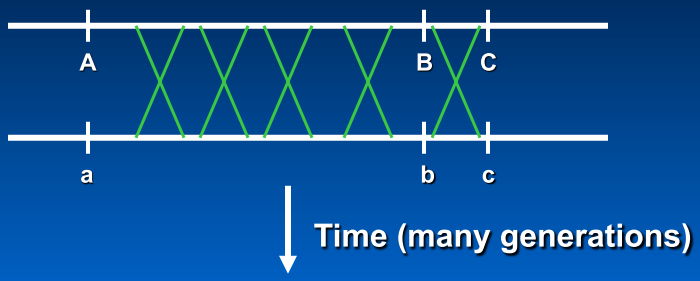
Genetic Variation and "Race"

- Genetic variation is correlated with geography and tends to be distributed continuously across geographic space
- "Race" may not be biologically meaningless, but it is biologically imprecise
- Individual ancestry provides more medically useful information

Linkage disequilibrium: nonrandom association of alleles at linked loci



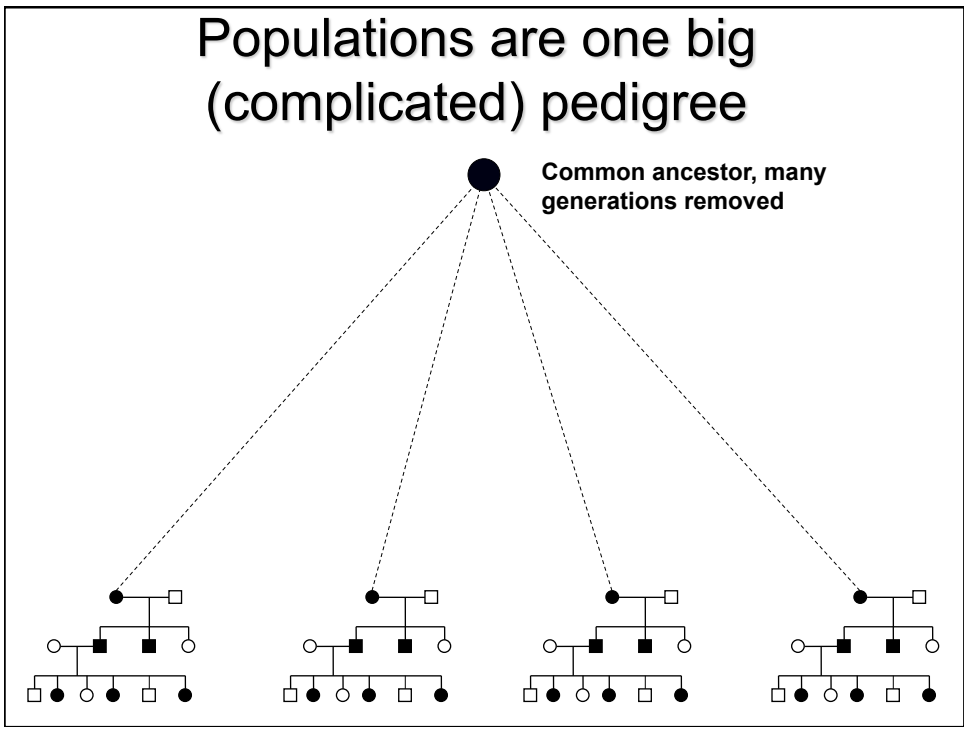
Over time, more crossovers will occur between loci located further apart

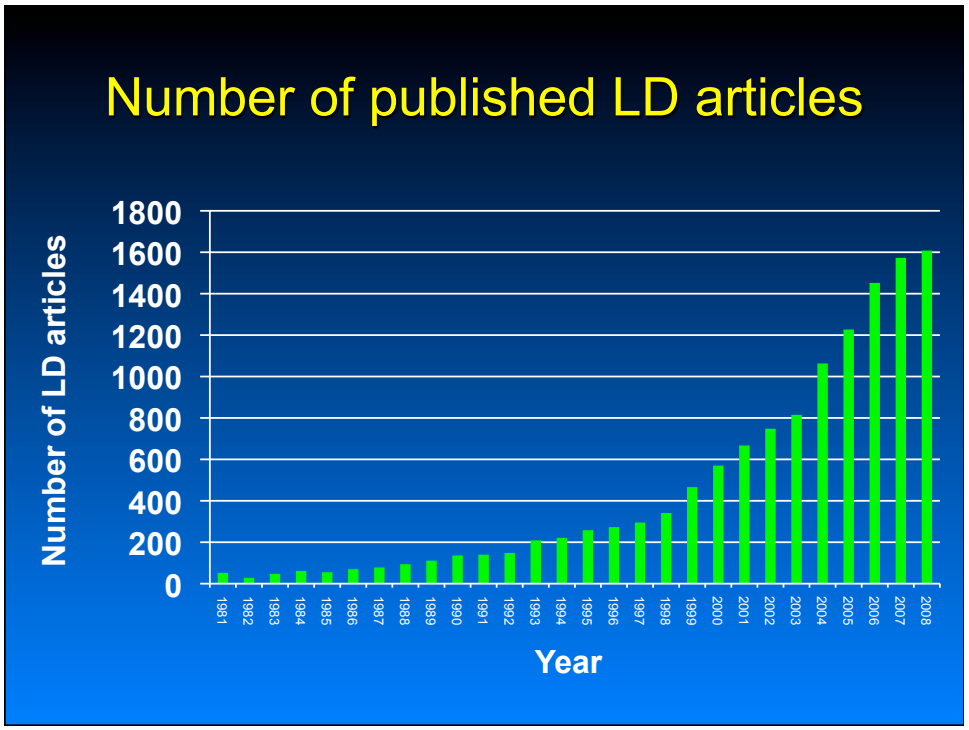


B and C will be found together on the same haplotype more often than A and B: there is more *linkage disequilibrium* between B and C than A and B

Potential advantages of linkage disequilibrium (LD)

- Family data are *not* necessarily needed
- Microarray technology now exists that allows dense genotype assays (SNPs every 3 kb)
- Association studies (linkage disequilibrium) can incorporate many past generations of recombination to narrow the candidate region

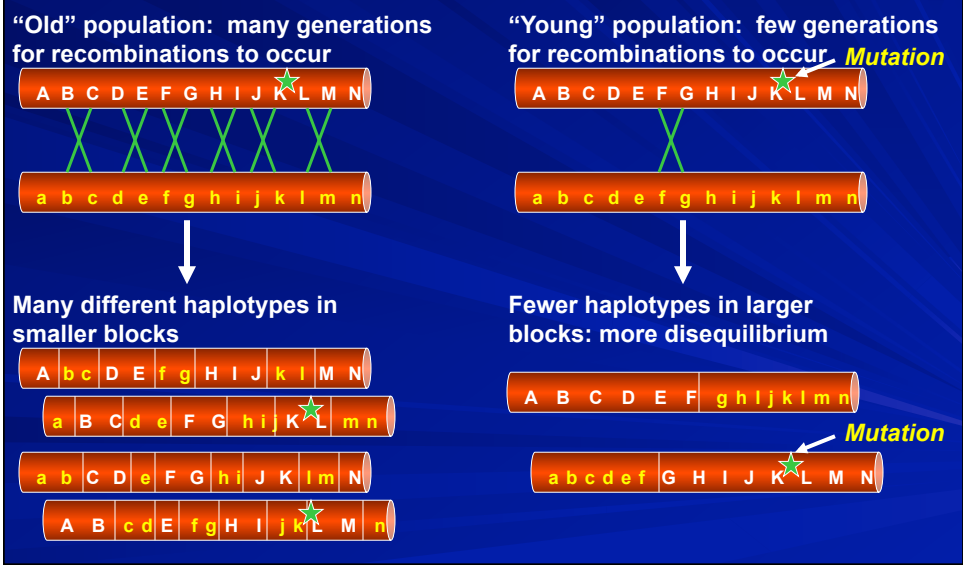

Populations are one big (complicated) pedigree
Common ancestor, many generations removed


Number of published LD articles
Number of LD articles
1800
1600
1400
1200
1000
800
600
400
200
0
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
Year

Factors that May Affect Linkage Disequilibrium Patterns

- Chromosome location
  - Telomeric vs. centromeric
  - Intragenic vs. extragenic
- DNA sequence patterns (GC content; presence of *Alu* elements)
- Recombination hotspots (1 every 50-100 kb)
  - 13-mer bound by *PRDM9* associated with 40% of hotspots
- Evolutionary factors: LD varies among populations
  - Natural selection
  - Gene flow
  - Mutation, gene conversion
  - Genetic drift

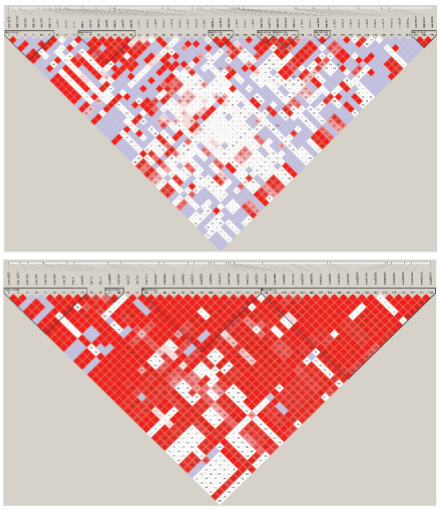
Population "age" can affect haplotype structure



Pairwise LD at the Angiotensinogen locus

Africans

Eurasians



Haploview; red = high LD

How general are these patterns?

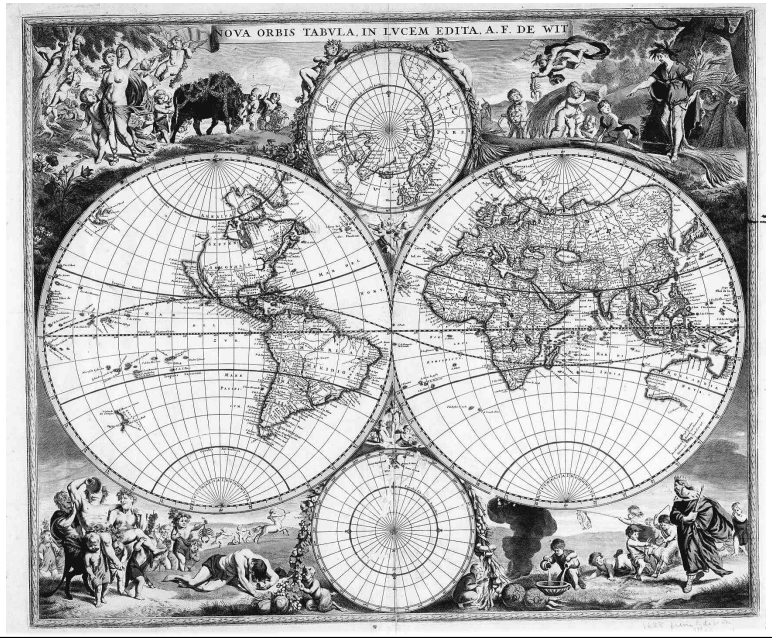
To what extent does LD vary with genomic location and population?

A Map of the World, 1544

In search of a better map: The International Haplotype Map Project

- 600,000 SNPs (1 per 5 kb) genotyped in 270 individuals
  - 90 CEPH Utah individuals (30 trios)
  - 90 Yoruban from Nigeria (30 trios)
  - 90 East Asians (45 Chinese, 45 Japanese)
- Evaluate patterns of linkage disequilibrium and haplotype structure
  - Variation in different genomic regions
  - Variation in different populations

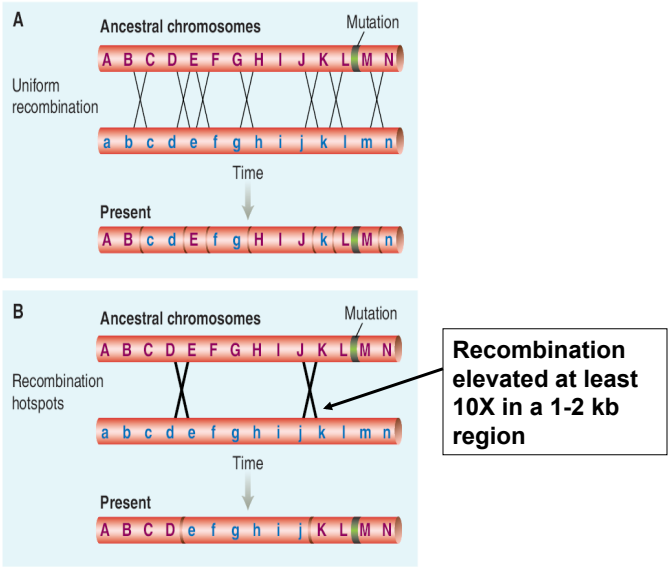
A Map of the World, 1688

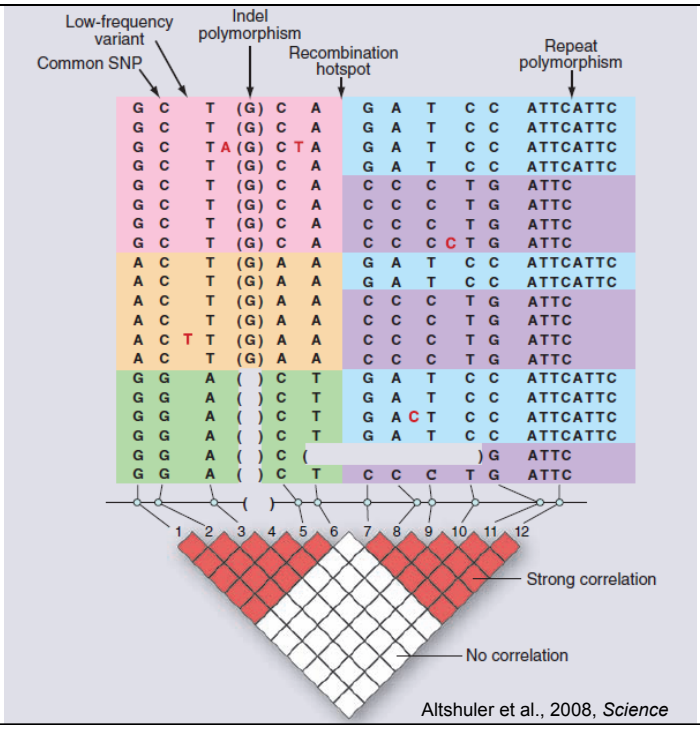


Genetic applications of HapMap

- Understanding human genome-wide haplotype diversity
- Detection of recombination hotspots
- Detection of genes that have experienced strong natural selection
- Detection of disease-causing mutations

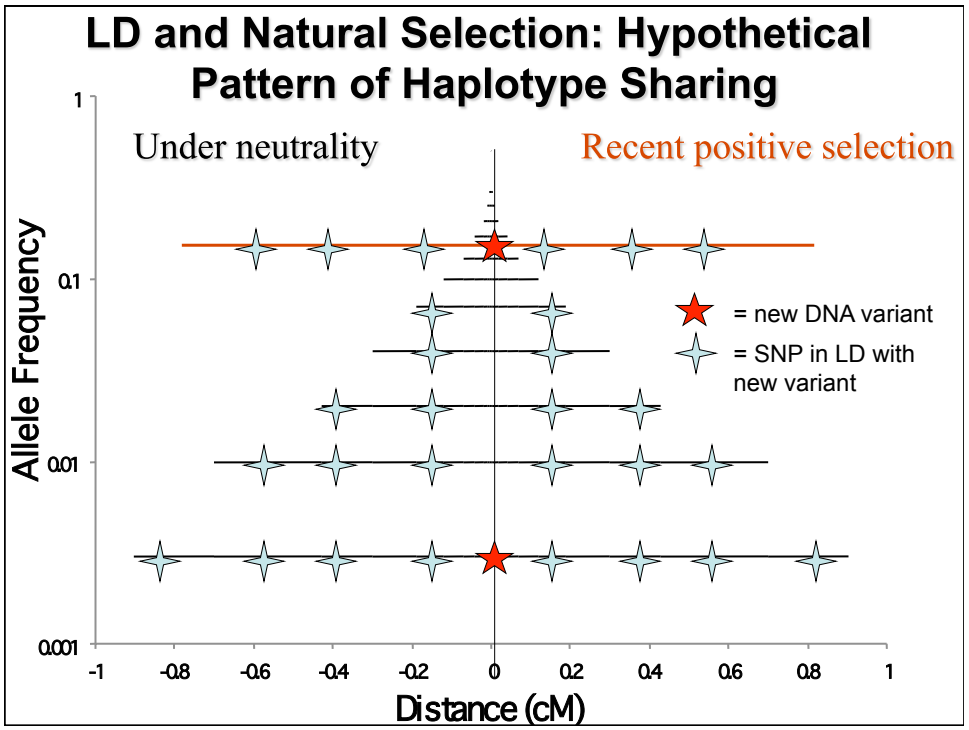
Recombination hotspots and haplotype blocks





Recombination hotspots

- LD patterns indicate 25,000 - 50,000 hotspots in human genome (1 every 50 – 100 kb) (Myers et al., 2005, *Science)*
- 60% of crossovers occur in only 6% of the genome (Coop et al., 2008, *Science* 319: 1395-8)
- Hotspots are not congruent in human and chimpanzee, despite 99% sequence identity: suggests hotspots evolve rapidly and may not be sequence-dependent



Examples of genes in which elevated LD indicates recent natural selection

Gene	Phenotype
G6PD	Malaria protection
HFE (hemochromatosis)	Iron absorption
CYP3A5	Sodium retention
*LCT (*lactase enhancer)	Lactase persistence
SLC24A5	Skin pigmentation
Alcohol dehydrogenase	Ethanol metabolism
EPAS1, EGLN1	Hypoxia response

Voight et al., 2006, *PLOS Biology* 4: 446-458

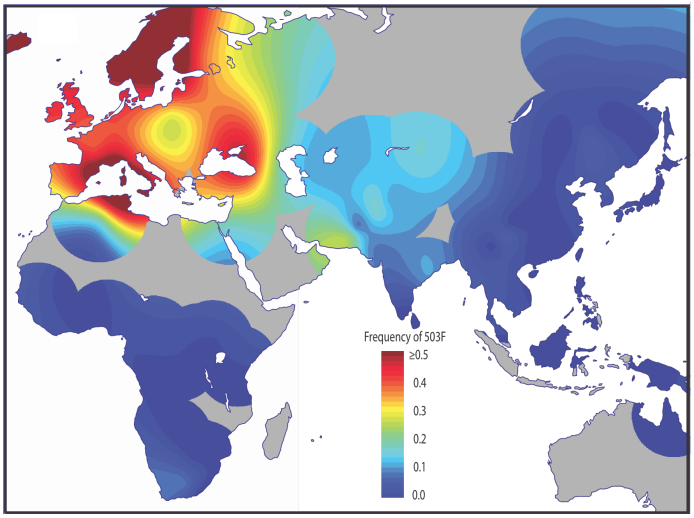
Simonson et al., 2010, *Science*

Recent Positive Selection, IBD5, and Crohn Disease

- Crohn disease is a chronic disorder of the intestinal tract with a prevalence of approximately 0.2%.
- IBD5 is a 250 kb multi-gene haplotype that is strongly associated with Crohn disease in Europeans.
- Multiple GWAS have reported association between Crohn disease and SNPs in the *OCTN1* gene in the IBD5 region.

Huff et al., 2011, *Mol. Biol. Evol.*

503F Variant of *OCTN1*



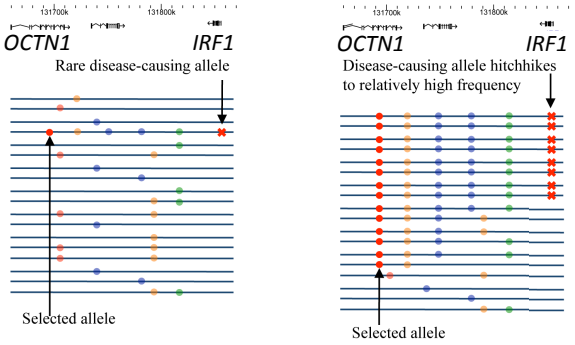
•Arose approximately 12,000 years ago; freq. 30-50% in Europe.
•503F is a gain-of-function mutation that increases ergothioneine substrate efficiency by 300%.

Recent Positive Selection at IBD5

Sample	iHS	p-value
HapMap CEU	-3.1	0.0007
HGDP Russian	-2.75	0.0044
HGDP Sardinian	-2.76	0.0075
HGDP French	-2.64	0.0076
HGDP Basque	-2.37	0.0128

Huff et al., 2011, *Mol. Biol. Evol.*

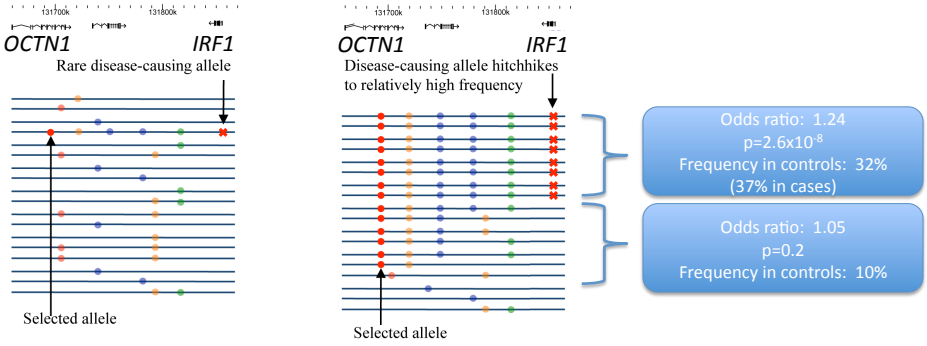
The *OCTN1* association could be explained by genetic hitchhiking



IRF1 is involved in innate immunity and clearance of intracelluar bacteria

Time

Disease association in 1868 cases and 5540 controls



IRF1 is expressed 72% more highly in Crohn disease intestinal tissue than in control tissue; no other gene in *IBD5* region shows expression differences

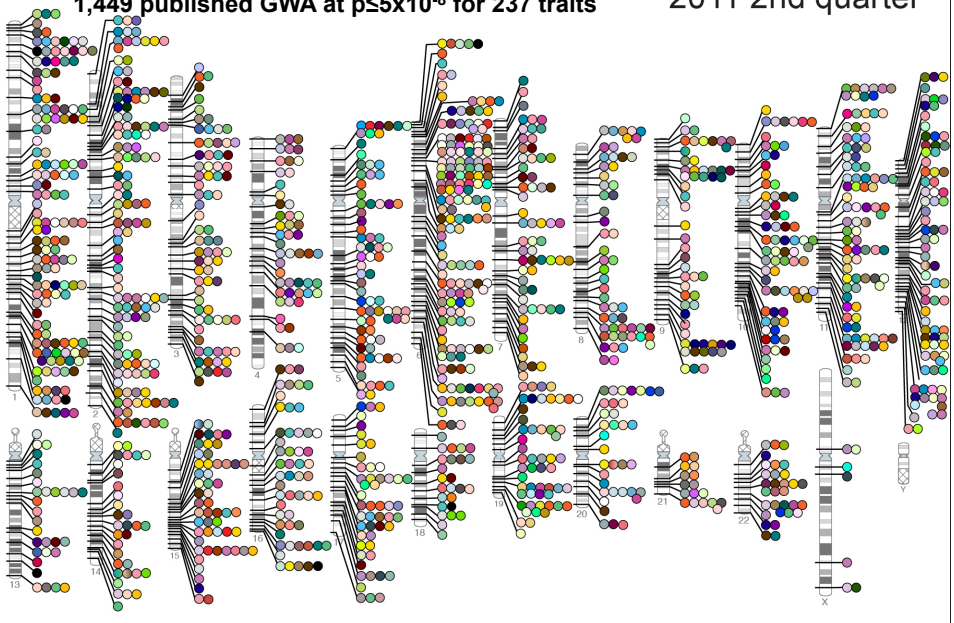
Huff et al., 2011, *Mol. Biol. Evol.*

SNPs in disequilibrium are redundant: we don't need to type all of them

For genome-wide association studies, "complete" coverage is given by about 1.6 million SNPs for African populations, 600,000 to 1M SNPs for non-African populations

Published Genome-Wide Associations through 06/2011,
1,449 published GWA at p≤5x10^{-8} for 237 traits

2011 2nd quarter



NHGRI GWA Catalog
www.genome.gov/GWAStudies

Population genetics is guiding development of new sequence analysis resources

- 1000 Genomes Project
  - Provides "control sequences" for variant analysis
  - Most rare alleles are not shared among populations (Gravel *et al*., 2011, *PNAS*)
- When is a variant functionally significant?
  - Evidence of purifying selection (elimination of deleterious variants) [Yandell *et al.,* 2011, *Genome Res.*]
  - Evolutionary conservation among species

Population genetics and genome analysis

- Genetic variation contains useful information about population history
- Genetic variation provides a more informed view of “race” and its relevance to medicine
- Population genetic analysis has been critical in understanding linkage disequilibrium and its application in disease-gene mapping
- Population genetics is *fun*!