Impact of Demographic History and Natural Selection on Human Genomic Diversity

Sarah Tishkoff
Departments of Genetics and Biology
University of Pennsylvania
Evolutionary Forces Influencing Genetic Variation

• Mutation
• Genetic Drift
• Migration
• Selection
How much do we differ?

- Identical twins: 0
- Unrelated humans: 1/1,000
- Human vs. chimp: 1/100
- Human vs. mouse: 1/30
- Human vs. broccoli: 2/3

3 billion DNA bases → 3 million differences between each pair of individuals

(slide courtesy of Lynn Jorde)
<0.1% divergence

http://dariaburke.wordpress.com/2011/03/21/ethic-diversity-at-estee-lauder/
There is also considerable structural variation across human genomes (insertions/deletions, gene duplications, and inversions)

More genetic variation within (~85%) relative to between populations (~15%)
1165 Microsatellite and In/Del Polymorphisms

Genotyped in >2,500 Africans from 121 ethnic populations

98 African Americans from four regions in the US

>1,500 comparative non-African individuals
Diversity Levels are Highest in Africa
Global Patterns of Substructure and Ancestry
Genetic Variation and Structure in Africa

Tishkoff, et al
Science 2009
African American Ancestry
Volume and Direction of the Transatlantic Slave Trade from Africa to the Americas

Source: David Eltis and David Richardson, Atlas of the Transatlantic Slave Trade
Genetic variation correlates with geography

Bryc et al. PNAS, 2010
African Ancestry = \frac{b}{(a + b)}

a and b are the chord distances from the European and African centroids.
Local Genomic Ancestry in African Americans

- African ancestry
- Shared African and European ancestry
- European ancestry

Representative African American

Recent admixture

Recent European ancestry

High degree of European ancestry
Variation in Skin Pigmentation

- Under selection
  - Vitamin D
  - Folate
  - Melanoma
Variation in Skin Pigmentation

$N = 2200$ for melanin

San
Anthropometric trait variation in Africa
Cardiovascular trait variation in Africa

![Graph showing systolic blood pressure and pulse rate variation across different African populations.](image-url)
Lipid biomarker variation in Africa

[Graph showing variation in triglycerides and LDL levels across different populations in Africa]

- Hunter-Gatherer
- Agriculturalist
- Pastoralist
- Agropastoralist
- Mixed
Identification of novel genetic variants associated with lactose tolerance and lactase gene expression in African pastoralist populations

Tishkoff et al., *Nature Genetics*, 2007

Ranciaro et al., *AJHG*, 2014
Historically, biological classification of races has been associated with:

- hierarchical ranking of races
- biological determinism
- Eugenics
- justification for genocide (e.g. the Nazi-led holocaust)
- colonialism, slavery, and other social inequities
• Problem with using race as an identifier is the lack of a clear definition of race

• Historically, race has been classified based on both socio-cultural and biological characteristics

Race does not correlate with patterns of population structure inferred from genetic data
• Ethnic groups differ for risk to disease.

• People whose ancestors lived in distinct environments have been subject to different selection forces and have differing demographic histories.

• Some diseases may be influenced by geographically restricted susceptibility alleles.

• Information about individual ancestry could provide important medical information for diagnosis and treatment.
However….one must be wary of racial profiling and ignorance of the continuous nature of genetic variation and high levels of admixture in modern populations, which can result in mis-classification and mis-diagnosis.

Cannot make generalizations based on continental region of origin because there can be tremendous variation amongst populations within a region (i.e. Africa).

Populations/ethnic groups differ in prevalence of disease due to both genetic and environmental factors and a key issue will be to disentangle these factors and to study their interaction.
• The goal should be personalized medicine—treating for the specific etiology in the individual patient.

• Information about “Race” could be important if trying to distinguish socio-cultural and environmental risk factors.

• However, for many studies race will be an insufficient identifier and detailed information about ethnicity, geographic origin, language, religion and/or culture will be important.

• “Ancestry” should supplant “race” in discussions where the group identity of the patient is at issue.
Thanks

Tishkoff lab, University of Pennsylvania
Matt Hansen/ Nick Crawford/Joseph Lachance / Joe Jarvis / Laura Scheinfeldt / Sameer Soi / Clara Elbers / Bart Ferwerda / Bill Beggs / Simon Thompson / Alessia Ranciaro / Jibril Hirbo

Illumina 5 M GWAS study
Stephen Chanock/Meredith Yeager (NCI)

African Collaborators

Population Studies Center at Penn