

# **African Integrative Genomics: Implications for Human Origins and Disease**

**Sarah Tishkoff**

Departments of Genetics and Biology  
University of Pennsylvania

# Key Challenges in Human Genomics Research

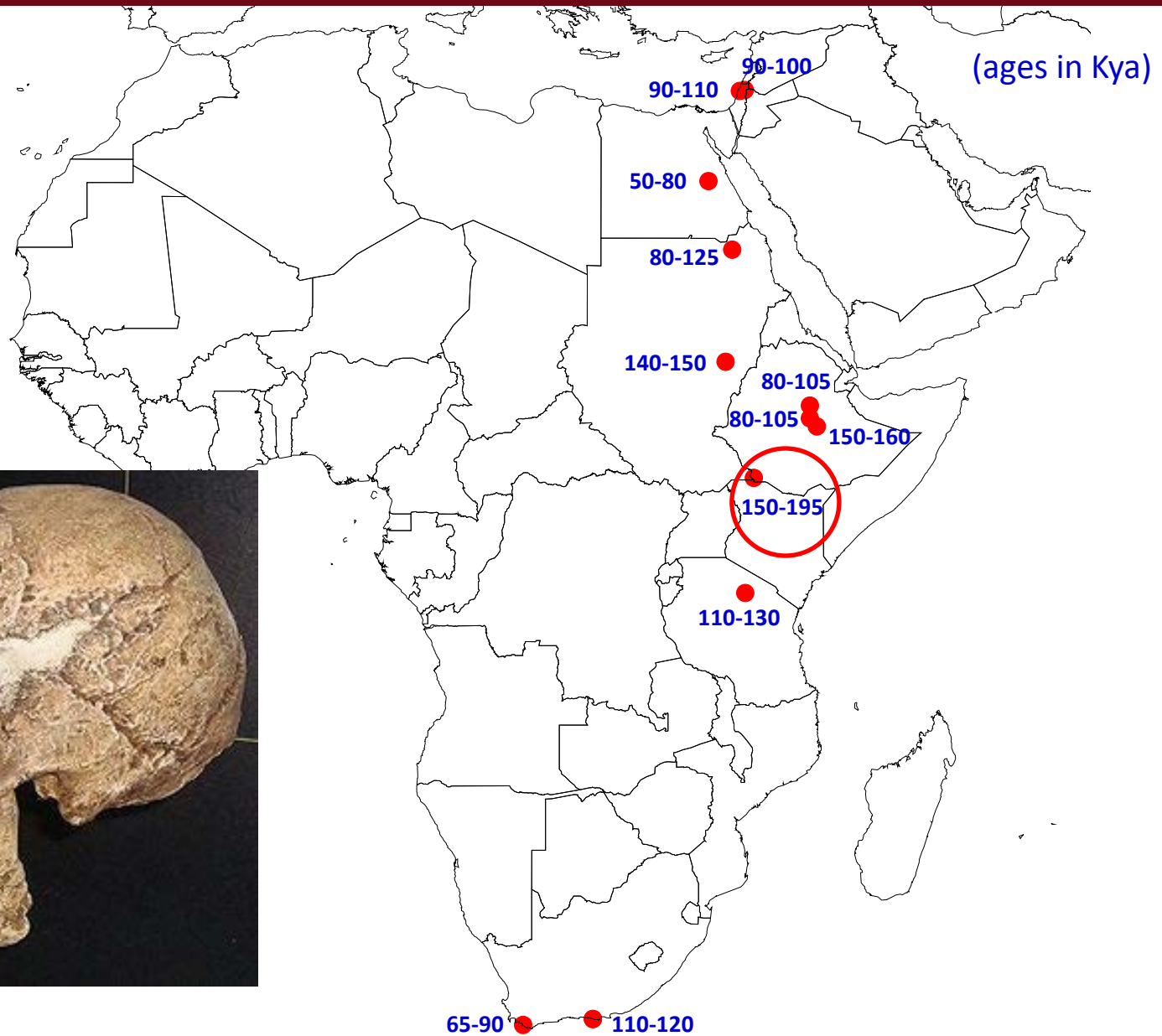


How do ethnically diverse humans differ in regard to genomic and phenotypic variation?

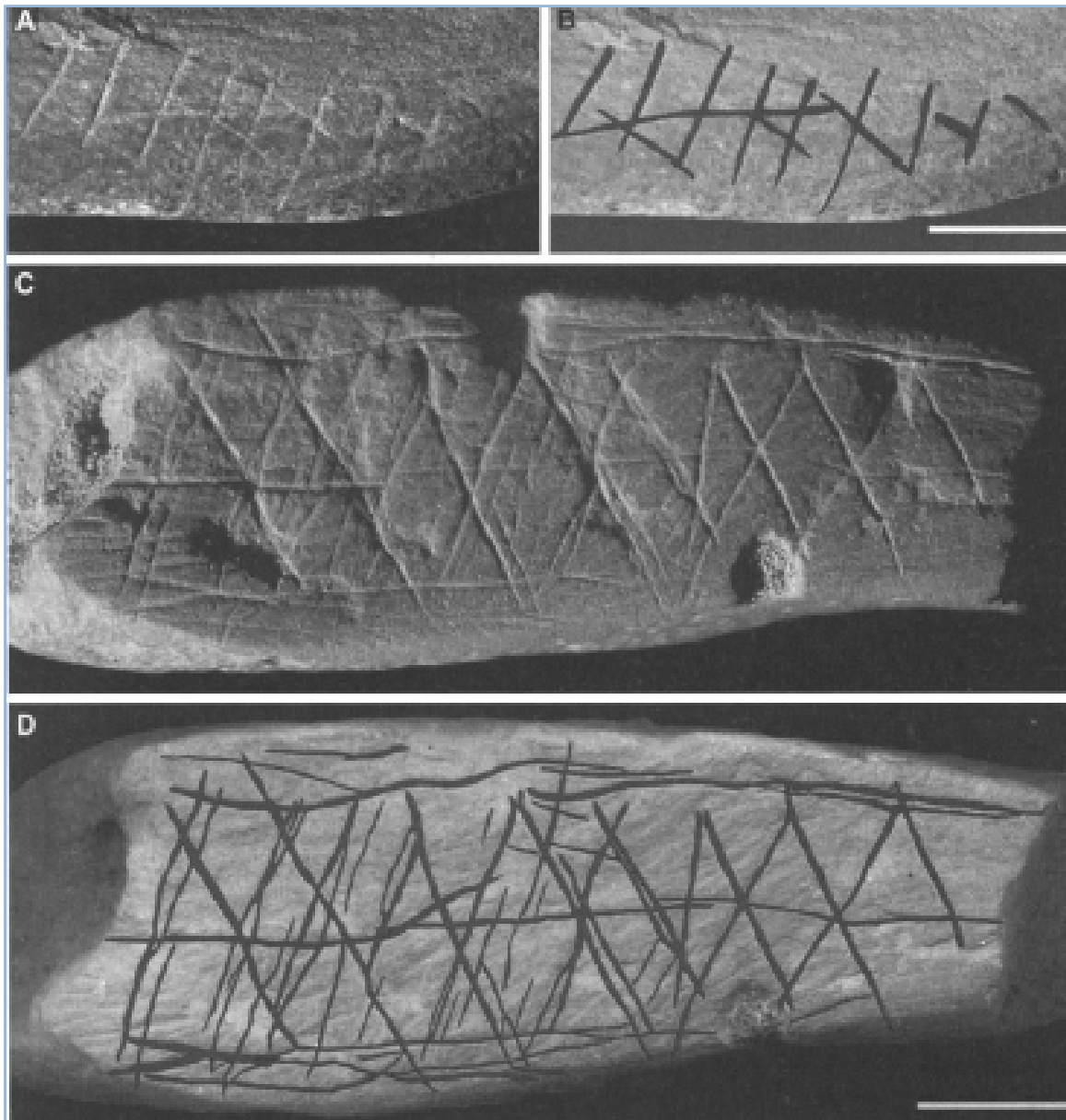
What are the evolutionary processes generating and maintaining that variation?

How do gene-gene, gene-protein, and gene-environment interactions contribute to phenotypic variability in humans?

# Modern Human Fossil Record



# Carved Ochre, Blombos Cave, South Africa ~70,000 ybp





# Major Migration Routes



Human Molecular Genetics, 3/e (©Garland Science 2004)

# How much do we differ?



Identical twins:

Number of DNA base differences

0



Unrelated humans: 1/1,000



Human vs. chimp: 1/100

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

*(slide courtesy of Lynn Jorde)*

# Levels and Distribution of Genetic Diversity in Humans

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%)

# Why Should We Study African Genetic Diversity?

To reconstruct human demographic and evolutionary history

To study the African Diaspora and African American ancestry

To study the genetic basis of susceptibility to communicable and non-communicable diseases

To understand differences in drug response











# Training and Capacity Building in Africa



# Returning Results to Participants





# Measuring Phenotypic Diversity

## Detailed anthropometric phenotypes

height, weight, waist circumference, %body fat, limb length, arm circumference, grip strength, skin pigmentation

## Cardiovascular, lung, blood phenotypes

blood pressure, oxygen level, lung capacity, hemoglobin, red cell and white cell counts

## Metabolic function / Adaptation to diet

lactose, glucose, protein, and lipid challenges  
amylase function and activity  
taste perception tests

## Infectious disease status

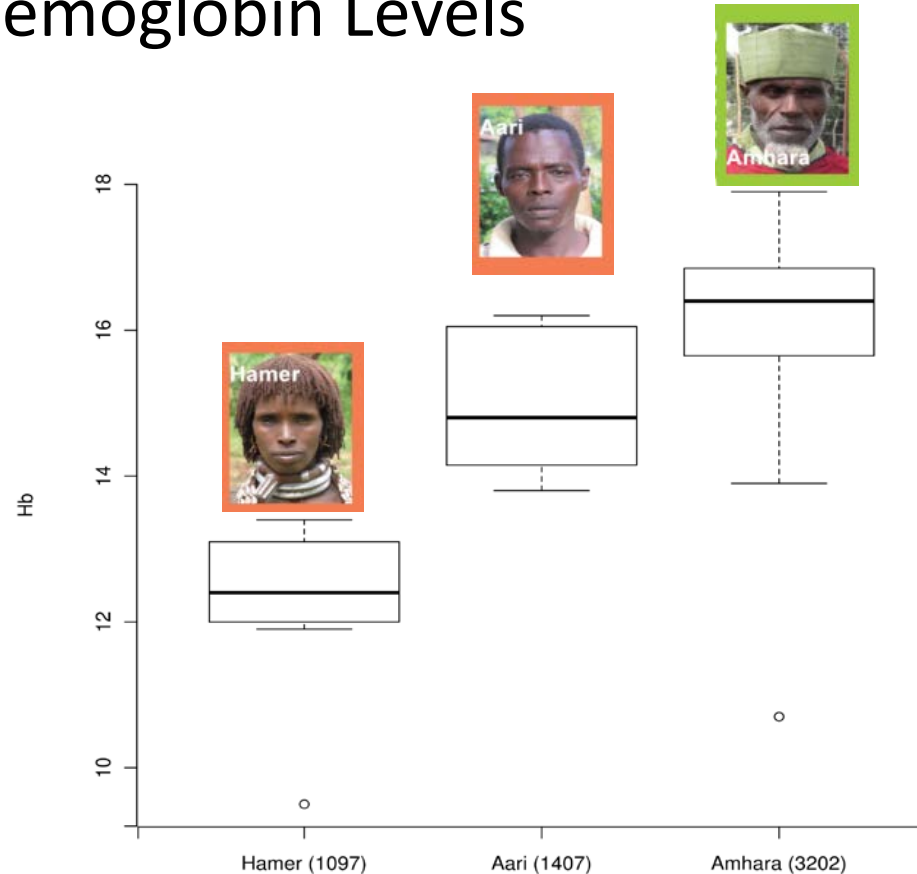
malaria, TB, HIV, immune response



# Phenotypic variation in Africa

## Adaptation to High Altitude Hemoglobin Levels

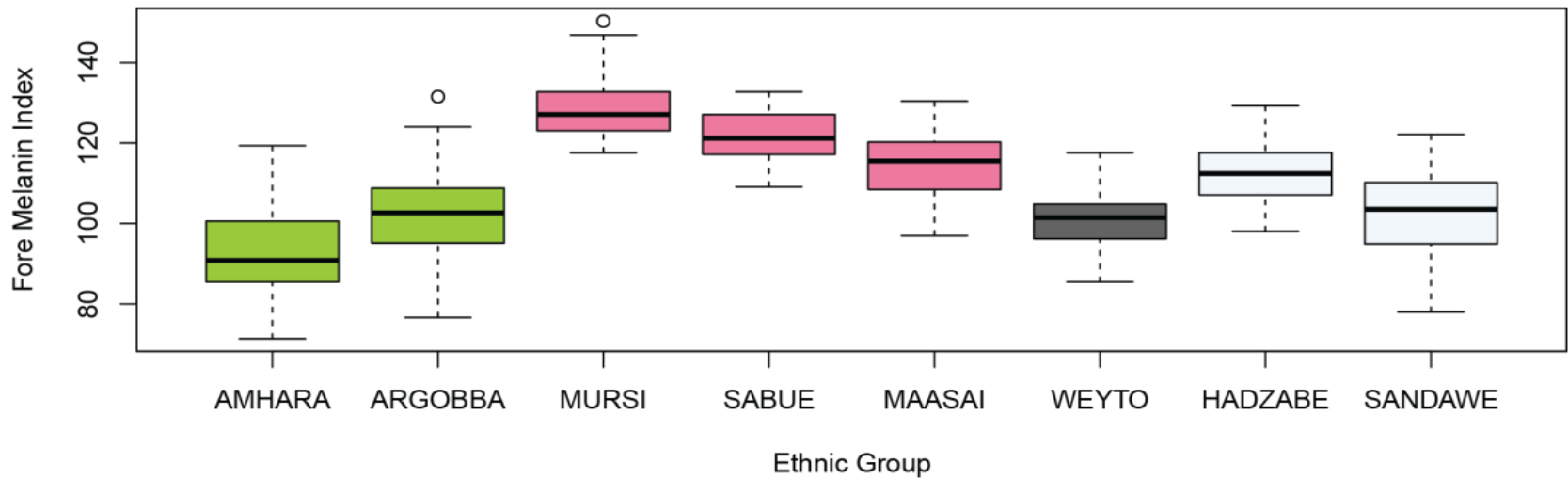
- Hb levels are significantly higher in high altitude Ethiopian males
- *THRB* and *ARNT2* are targets of selection only in high altitude Amhara and are associated with Hb levels
- Important role in the HIF-1 pathway critical for response to hypoxia *in utero*





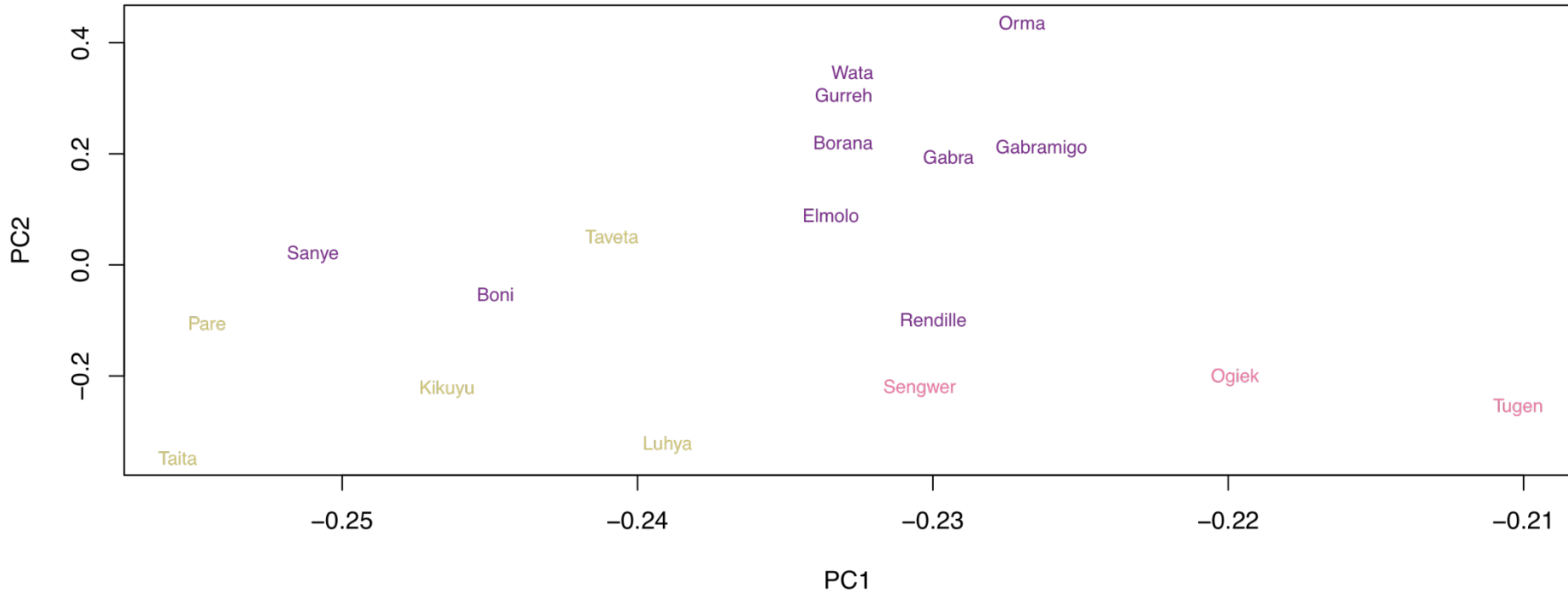
# Phenotypic variation in Africa

## Skin Pigmentation Melanin Levels



# Phenotypic Variation in Africa

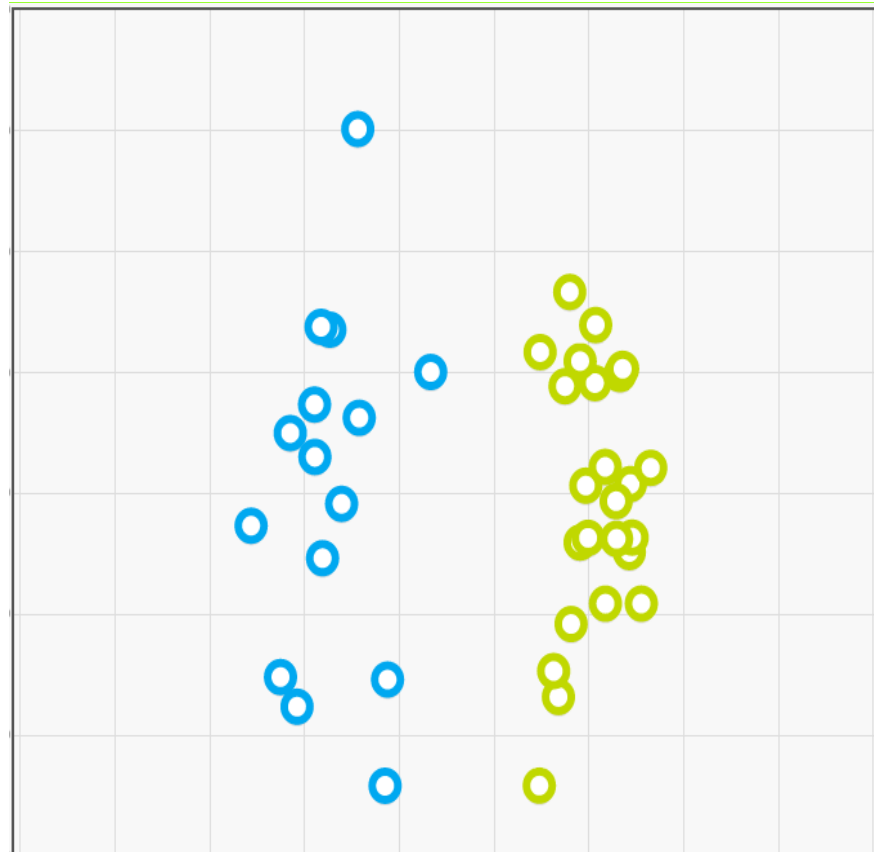
## Cardiovascular PCA



# Phenotypic Variation in Africa

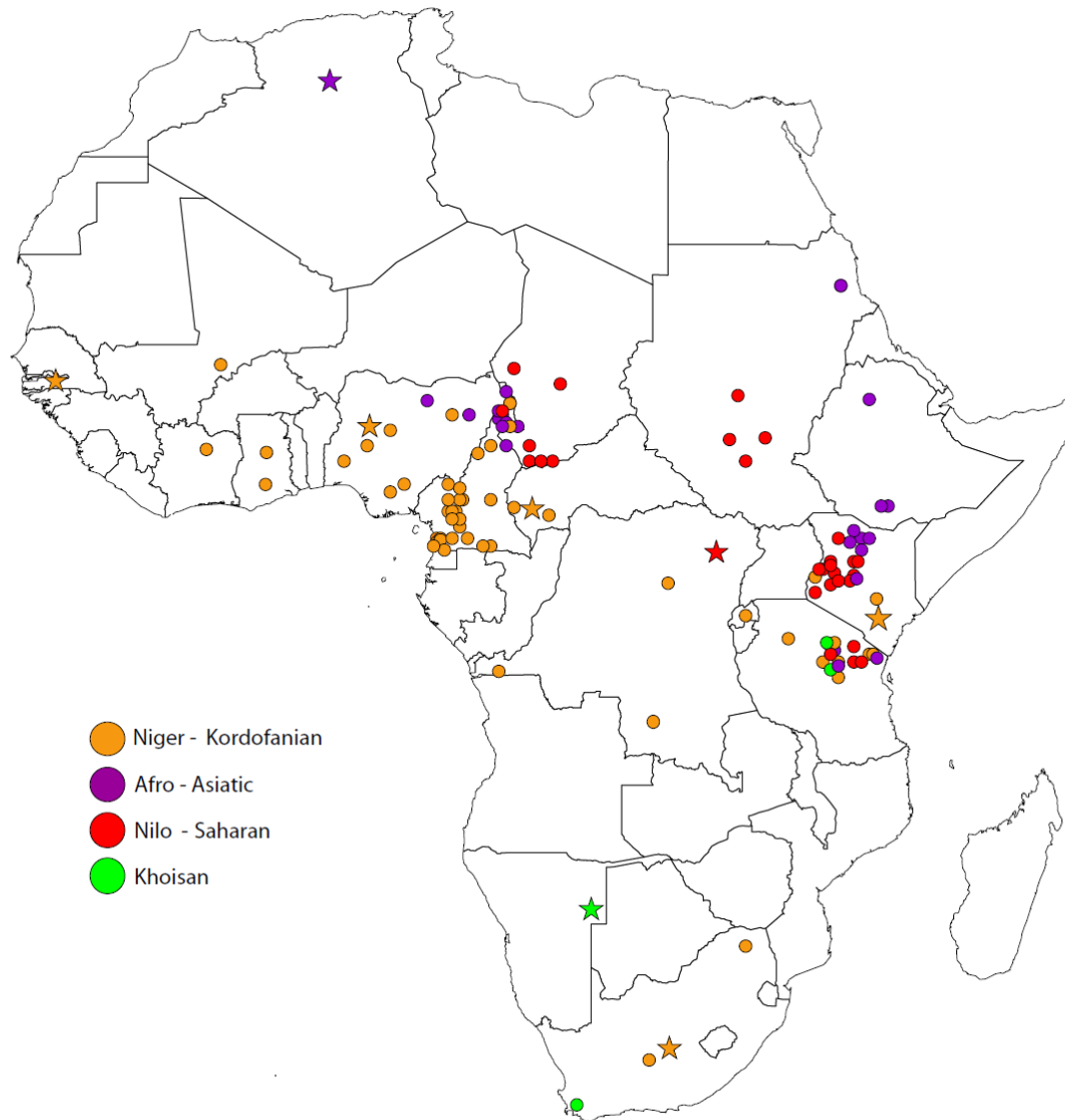
## Metabolic traits PCA of Metabolites

- Hunter-gatherers
- Pastoralists



# The Genetic Structure and History of Africans and African Americans

Tishkoff et al. *Science* May 22;324(5930):1035-44. 2009



1165 Microsatellite and In/Del Polymorphisms

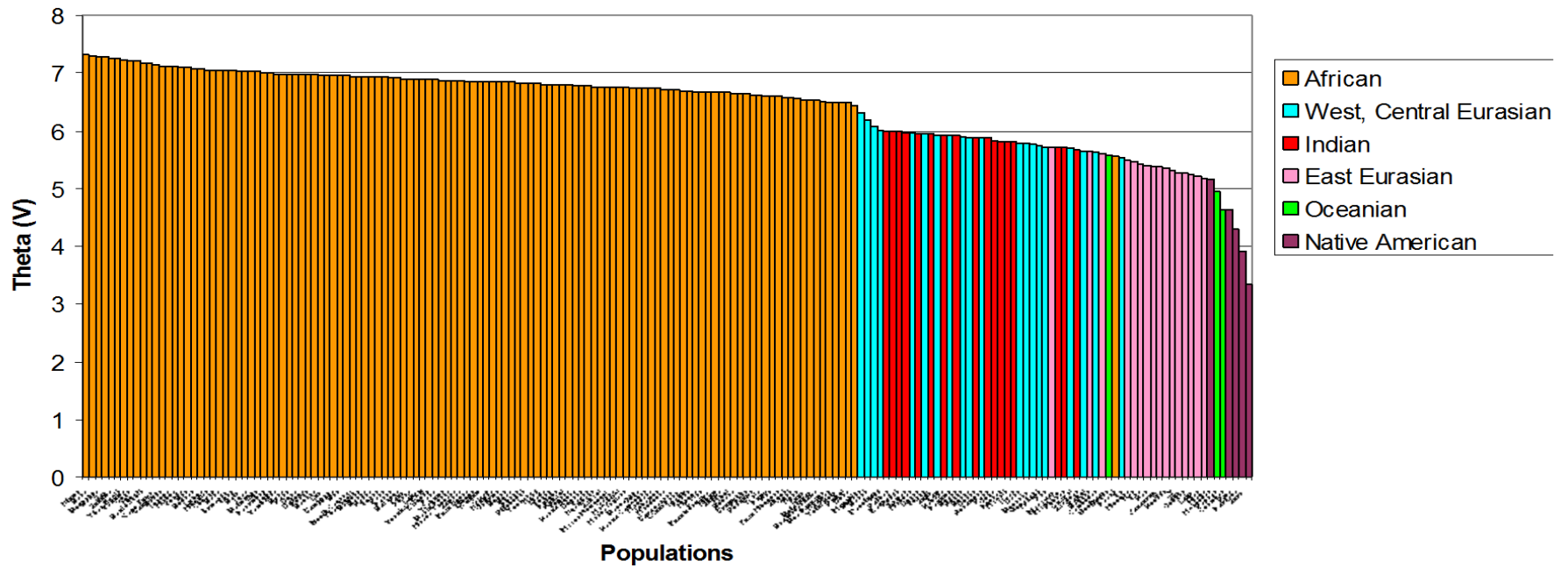
Genotyped in >2,500 Africans from 121 ethnic populations

98 African Americans from four regions in the US

>1,500 comparative non-African individuals

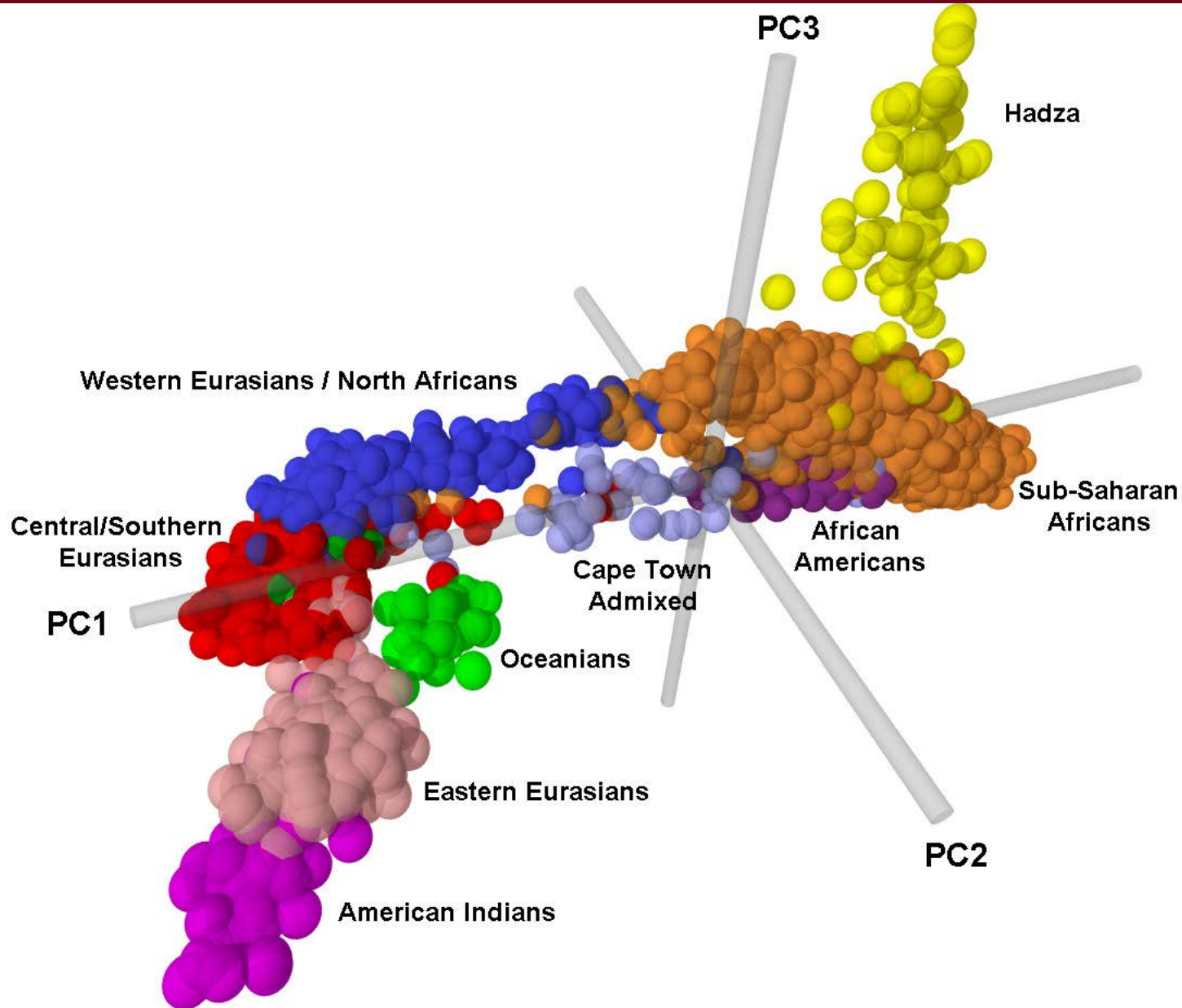


# Diversity Levels are Highest in Africa

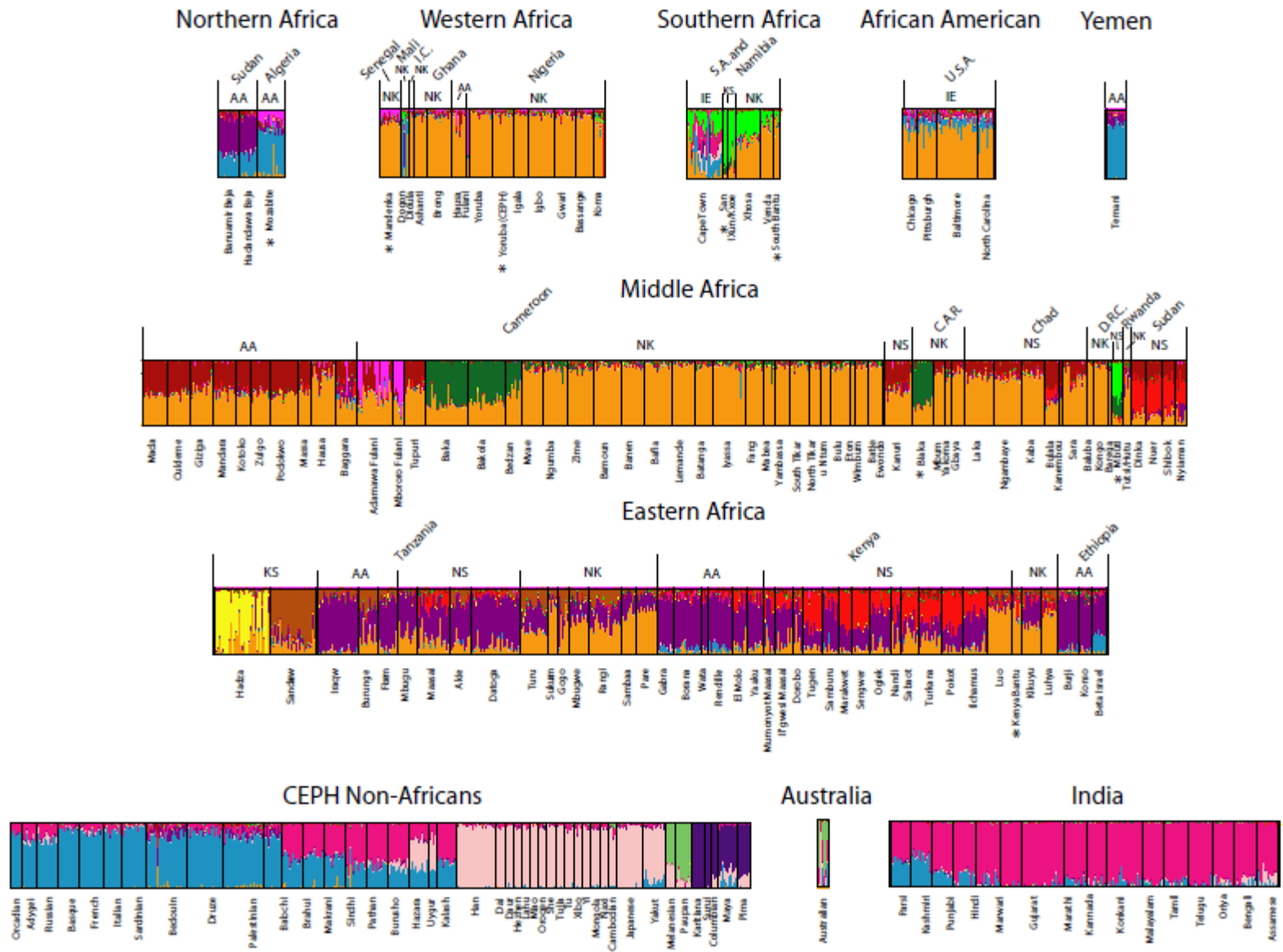




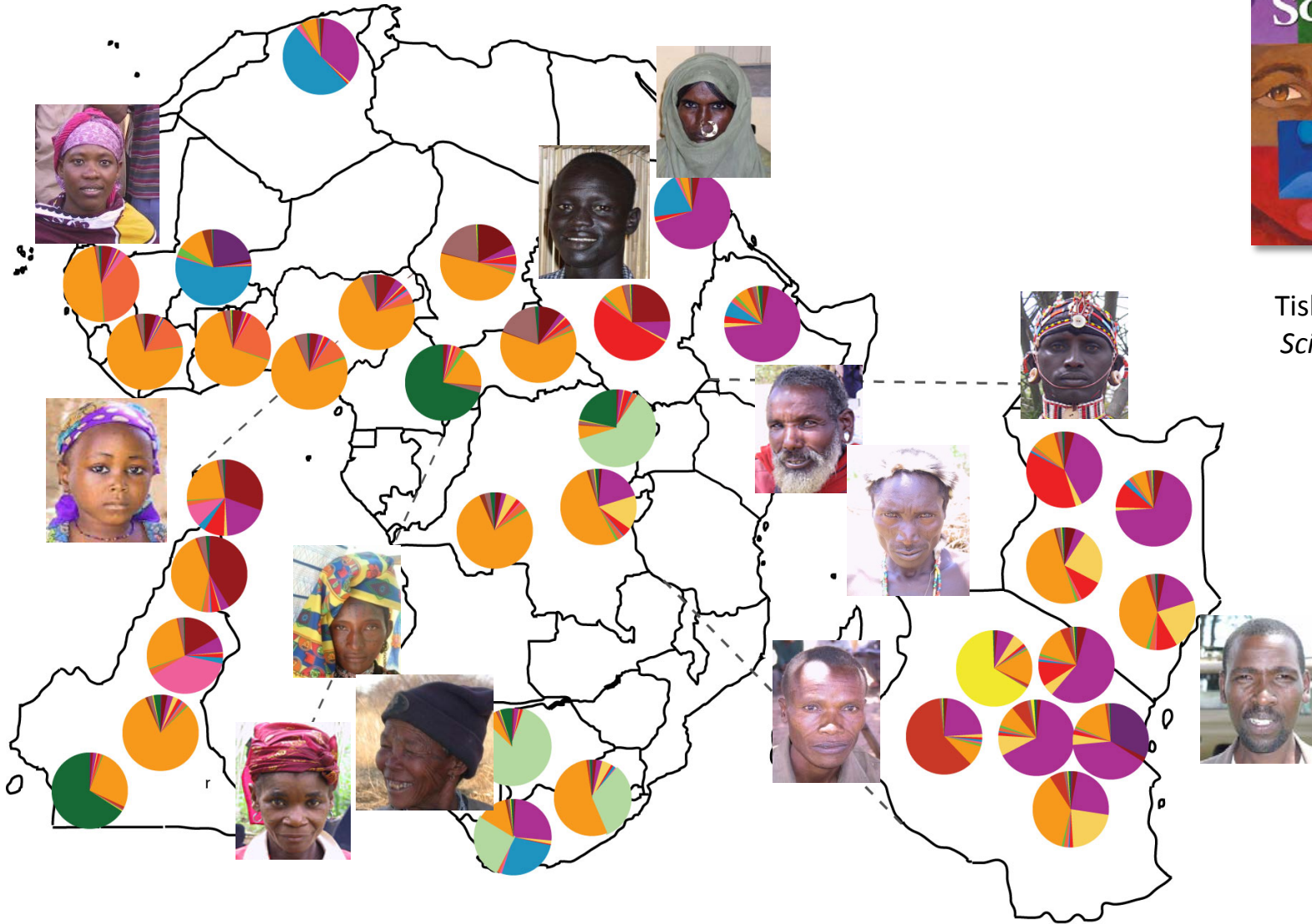
# PCA of Individual Genotypes Indicates Geographic Clustering



# Global Patterns of Substructure and Ancestry

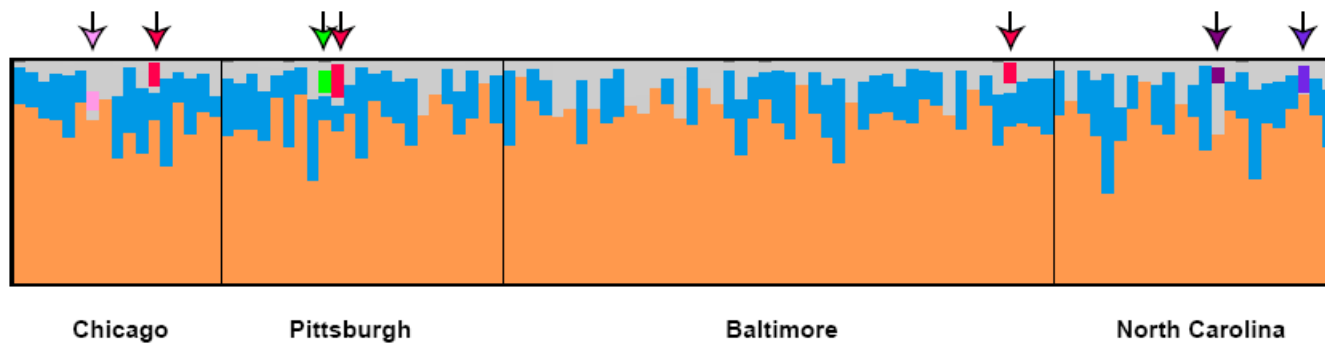
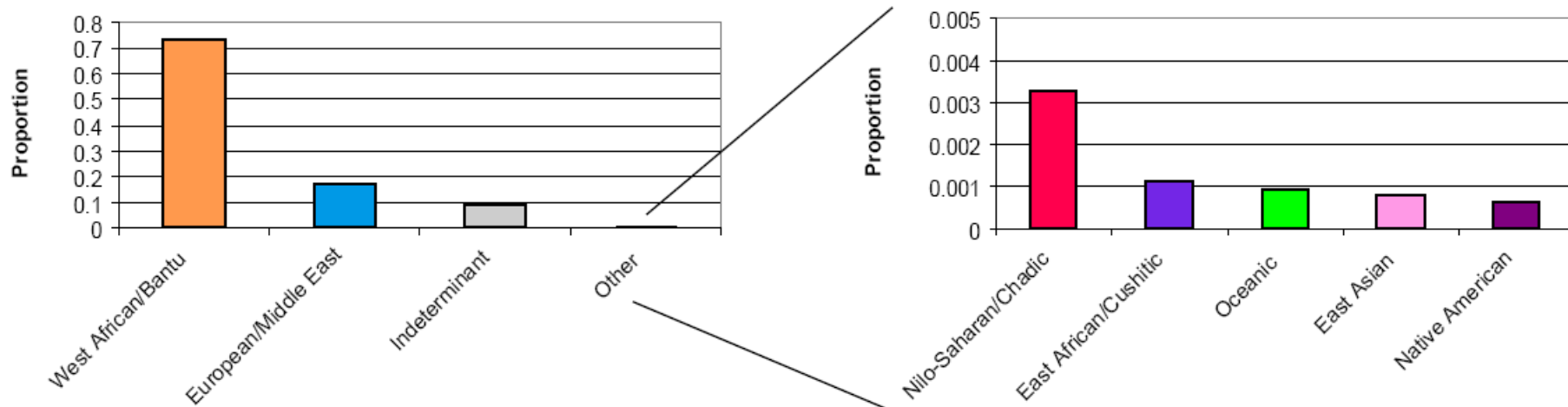


# Genetic Variation and Structure in Africa



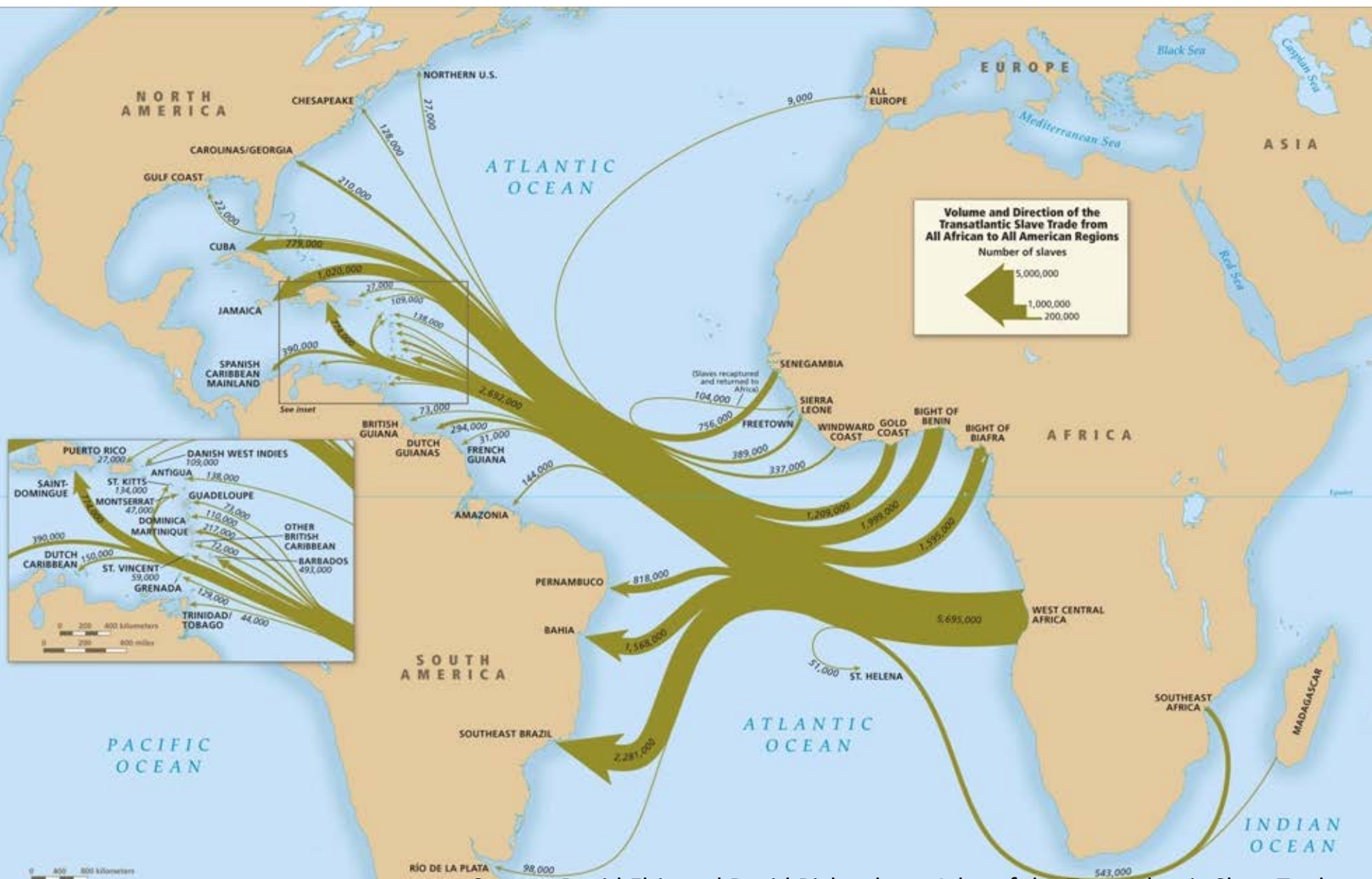
Tishkoff, *et al*  
*Science* 2009

# 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



# Identification of genetic variants that play an important role in adaptation and disease

## Identification of **genetic signatures of natural selection**

Mutations associated with disease in modern populations may have been selectively advantageous (e.g. adaptive) in the past (i.e. hypertension, diabetes, obesity, asthma).



**Pygmy**

Hunter-gatherer  
*Tropical*

**Samburu**

Pastoralist  
*Savannah*

**Bantu**

Agriculturalist  
*Tropical*

**Hadza**

Hunter-gatherer  
*Savannah*

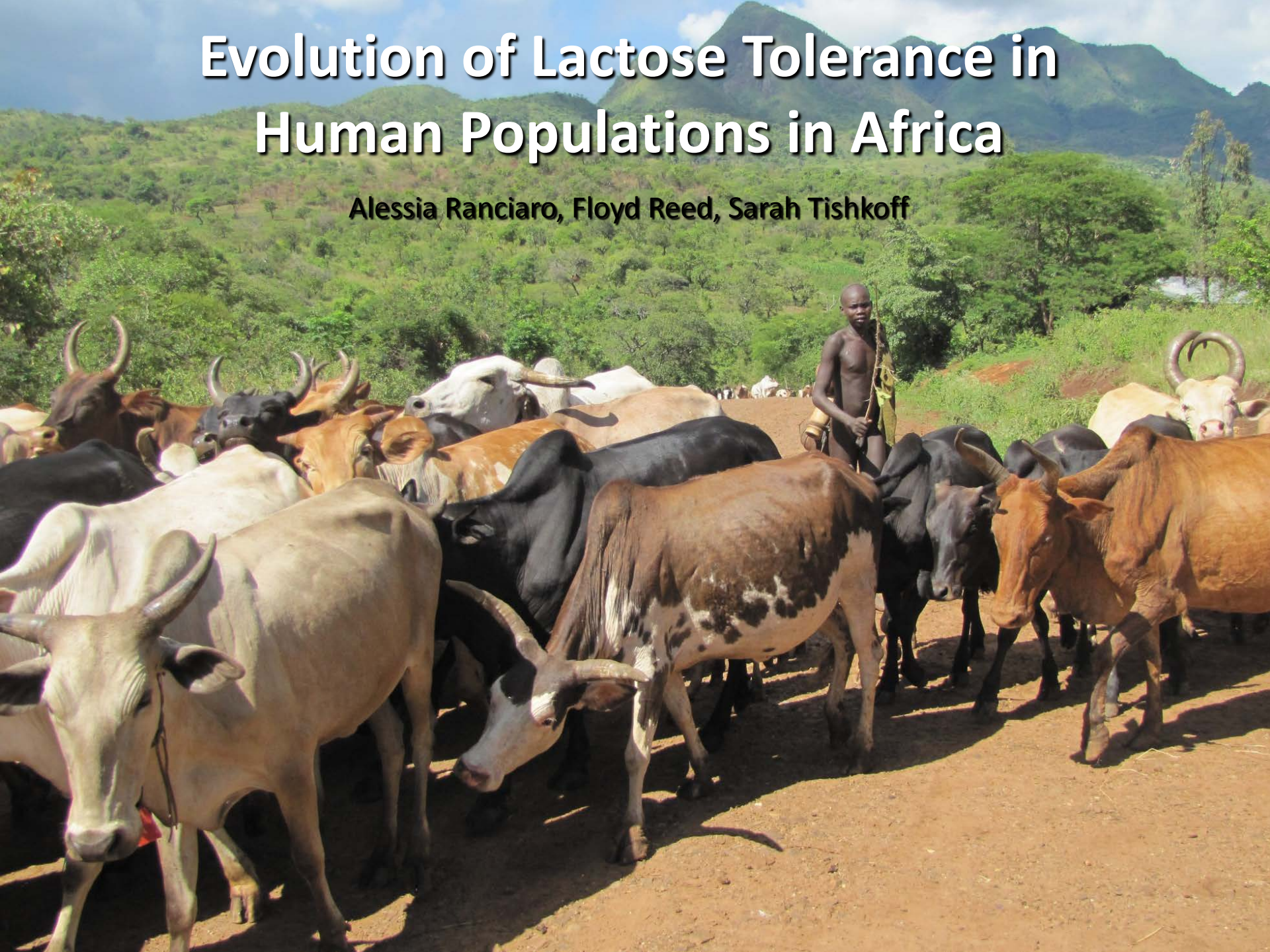
**Amhara**

Agriculturalist  
*High Altitude*



# Evolution of Lactose Tolerance in Human Populations in Africa

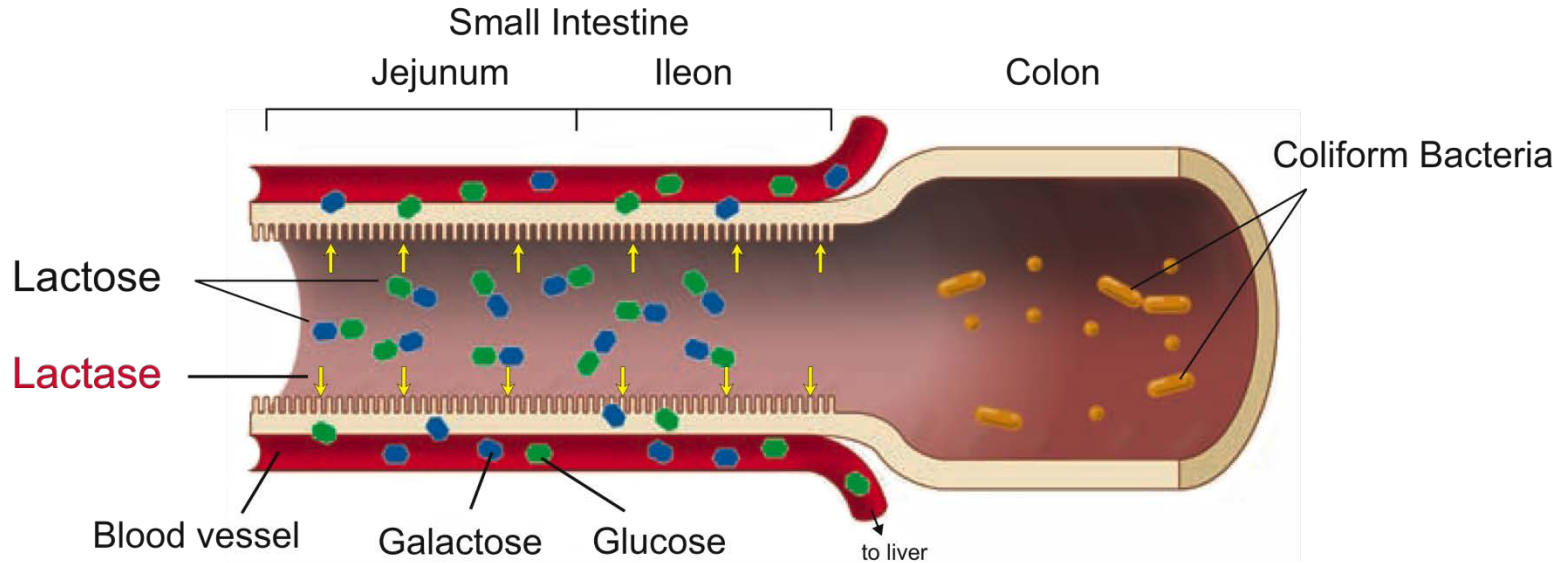
Alessia Ranciaro, Floyd Reed, Sarah Tishkoff



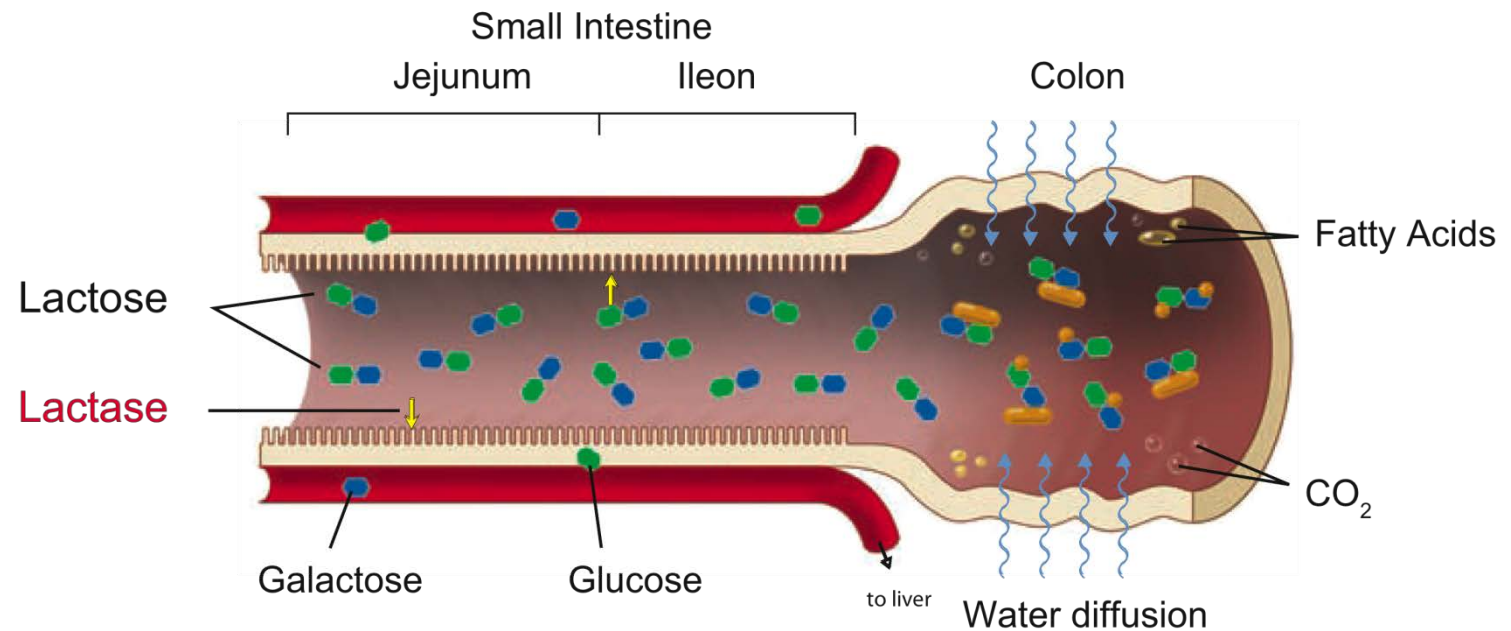


# Lactase-phlorizine hydrolase (LPH) hydrolyzes lactose into glucose and galactose

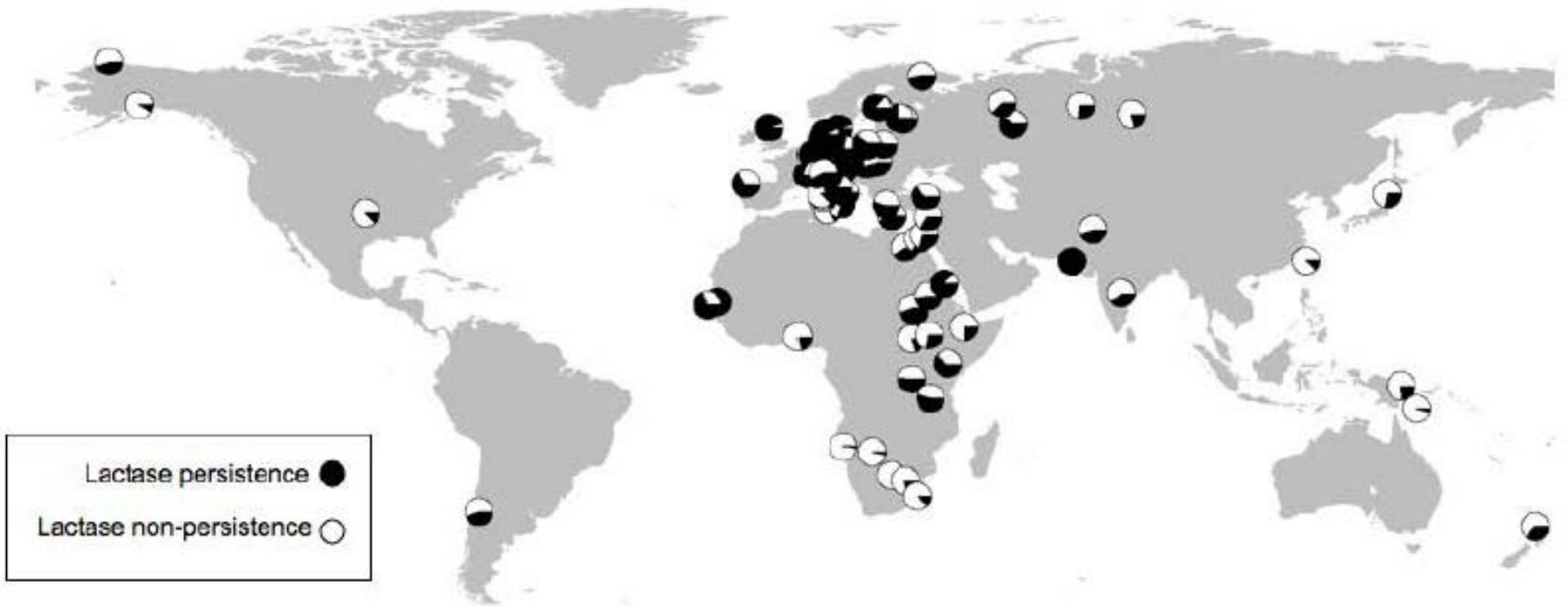
## Lactase Persistence



## Lactase Non-Persistence



# Global Distribution of Lactose Tolerance







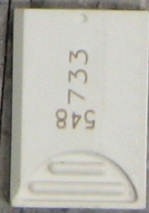










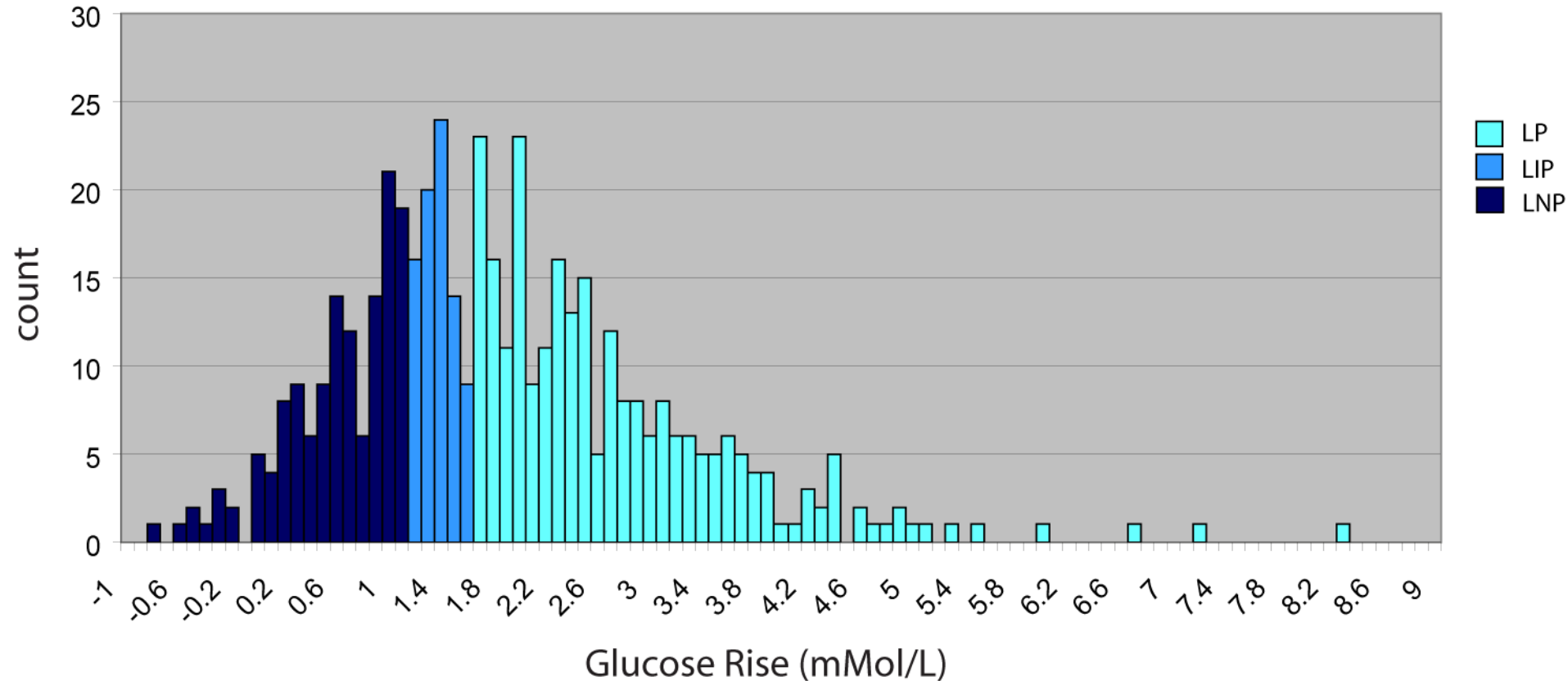








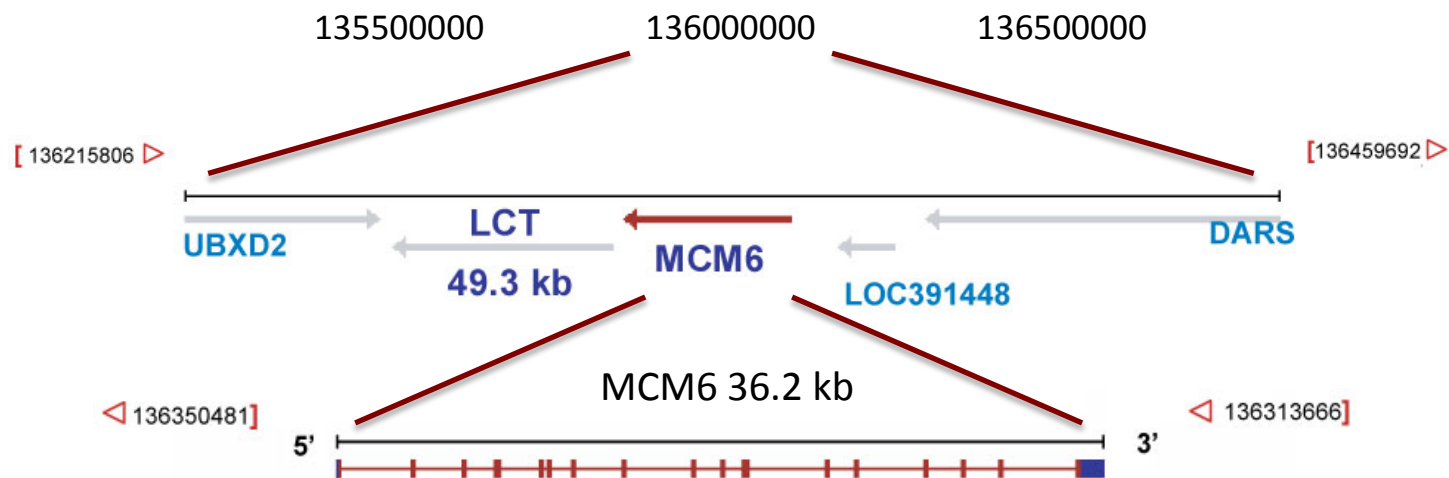
# LTT Phenotype Distribution



LP > 1.7 mM glucose rise, 1.7 mM > LIP > 1.1 mM, LNP < 1.1 mM

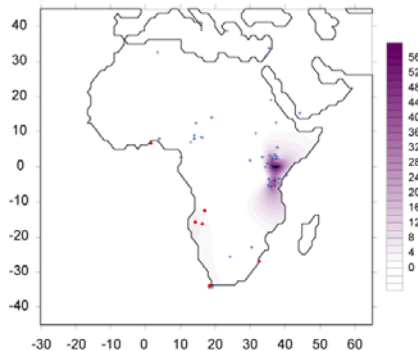
N = 470 individuals from Tanzania, Kenya, and the Sudan

# Identification of novel genetic variants associated with lactose tolerance and lactase gene expression in African pastoralist populations



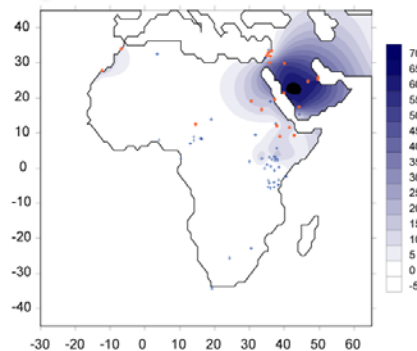
...**C/G**TAAGTTACCA.....  
-14010 bp

**C -14010**

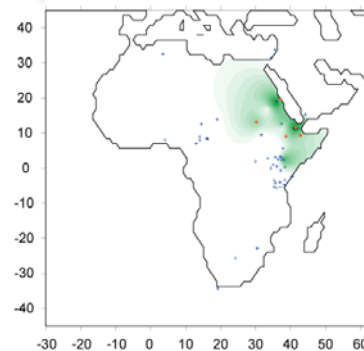


...AAGATAA**T/G**GTAG**C/T****C/G**TG....  
-13915 bp    -13910 bp    -13907 bp

**G -13915**



**C -13907**



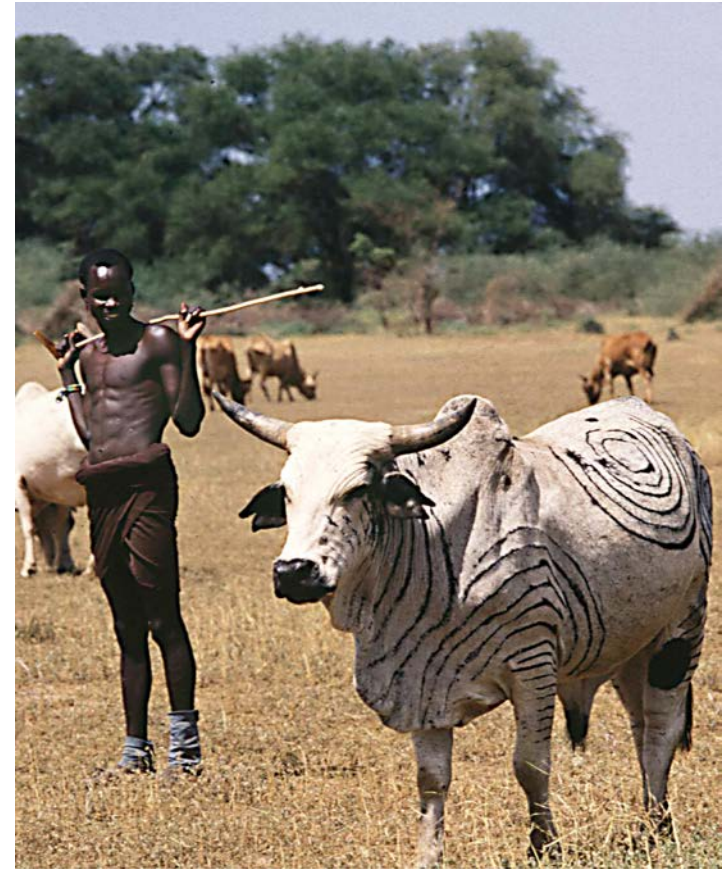
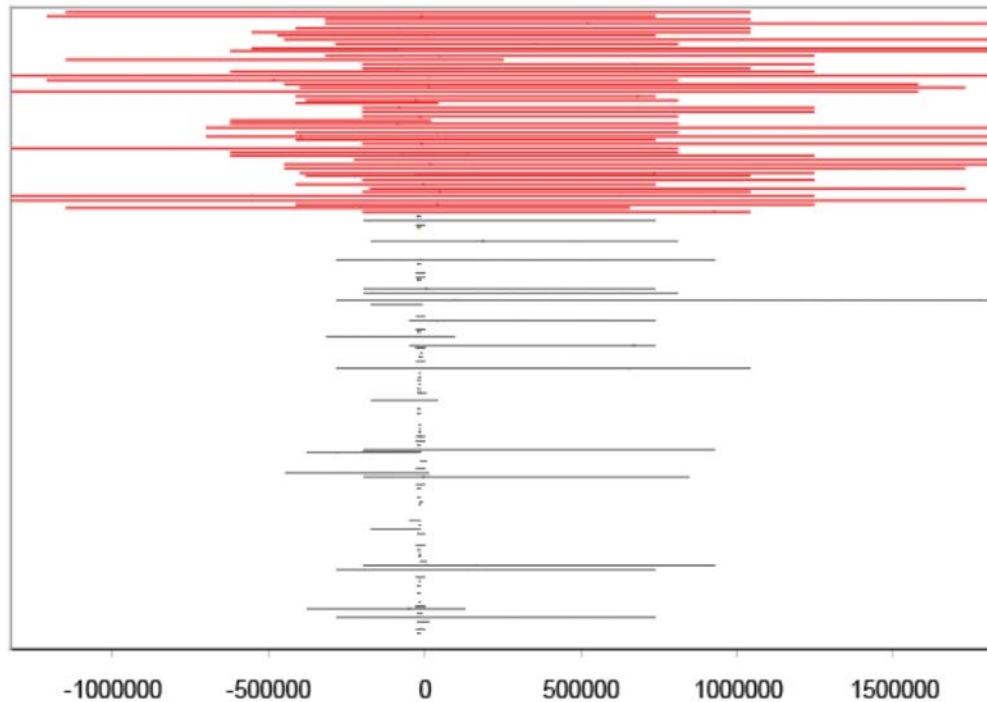
Intron 13

Tishkoff et al,  
*Nature Genetics*,  
2007

Ranciaro et al.,  
*AJHG*, in press

# Origins of Lactose Tolerance in East Africa

African G/C - 14010



Estimated Age ~3,000 – 7,000 years

**Mendelian traits are the low hanging fruit**





# Height is a “Classic” Complex Trait

- Height is highly heritable (Heritability = .80 - .90)
- GWAS in tens of thousands of Europeans has identified >180 loci, each of very small effect
- Explains ~10% of the variation in height
- Most are **not** part of the growth hormone (GH)/IGF1 pathway

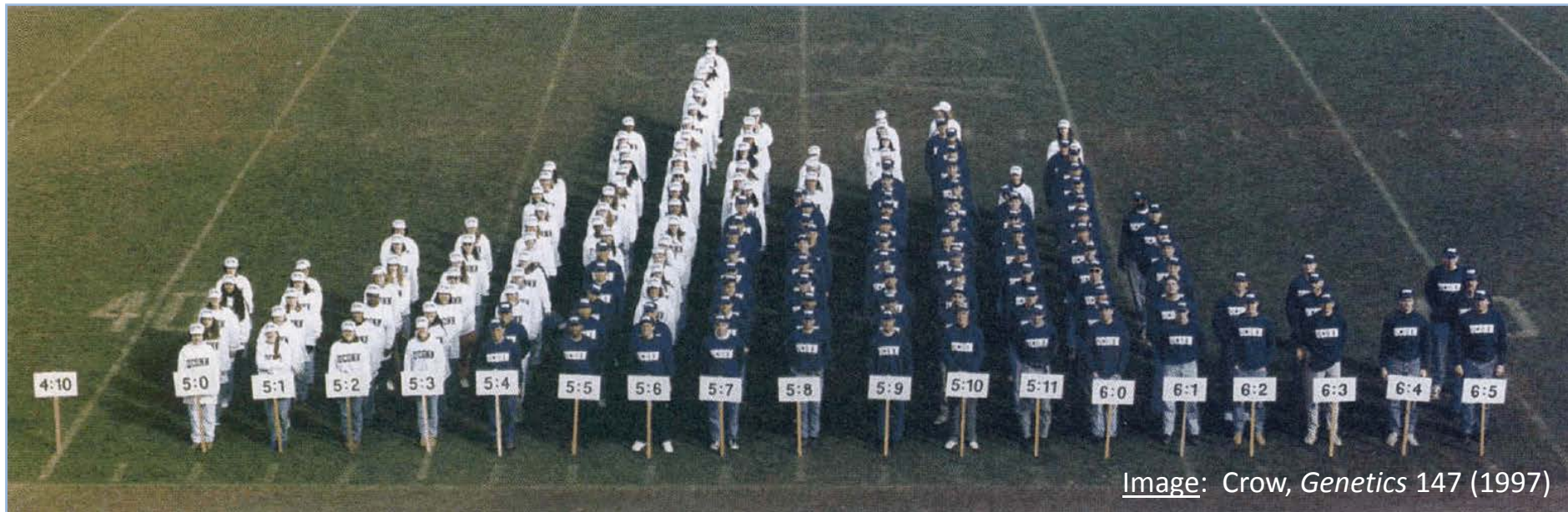


Image: Crow, *Genetics* 147 (1997)

# Height is highly variable among African populations

**Bakola Pygmy**  
Cameroon



152/146  
(M/F)

**Hadza**  
Tanzania



164/153  
(M/F)

**Samburu**  
Kenya



171/161  
(M/F)

**Dinka**  
Sudan



190/182  
(M/F)

Average height in cm



















# The Genetic Basis of Short Stature in Pygmies

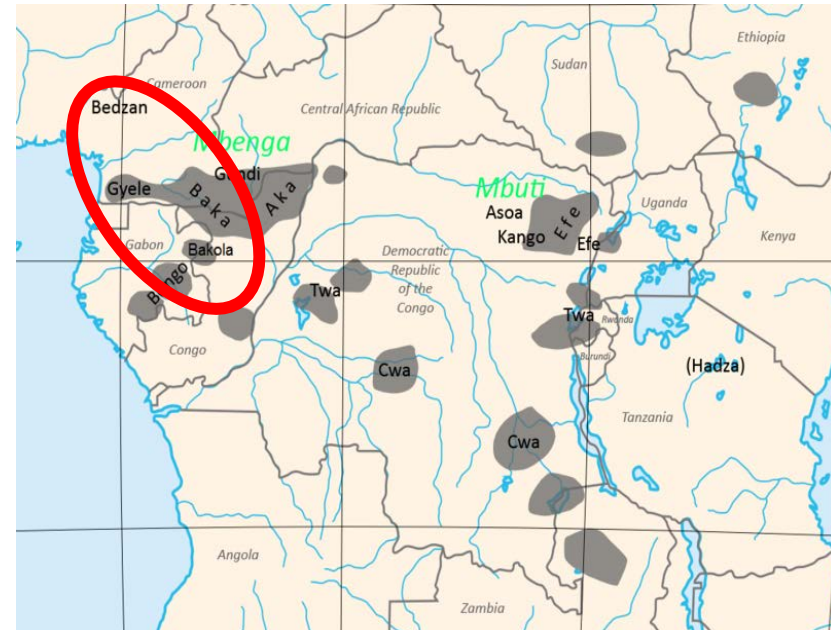
## Contrast in height between Pygmies and neighboring Bantu-speaking populations

- mean height in Pygmies is 152 cm
- mean height in Bantu is 170 cm
- **difference in height of ~18 cm**

## Adaptive hypotheses regarding short stature/small body size

- thermoregulation
- limited food resources/energetics
- locomotion
- **life-history tradeoff**

Physiologic and metabolic studies implicate **disruptions of the GH/IGF1 pathway** to play a role in short stature



# Pygmy height is correlated with Bantu ancestry

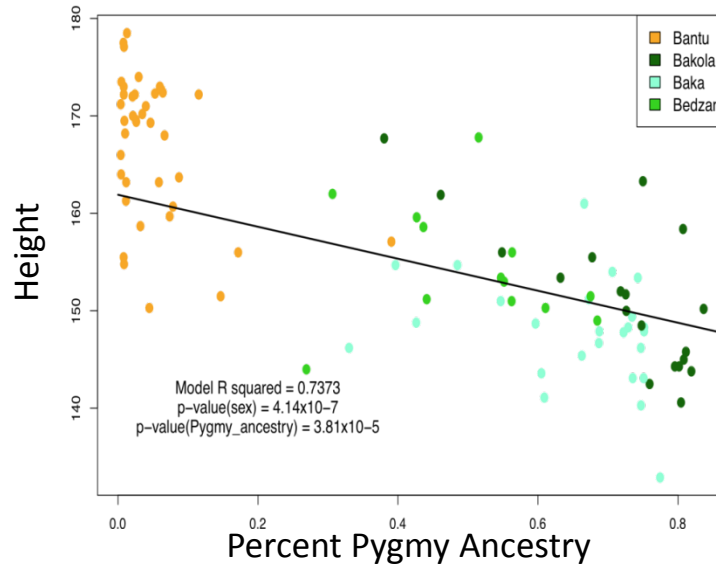
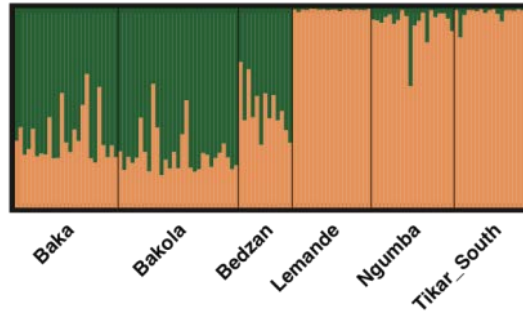
Pygmy



Bantu



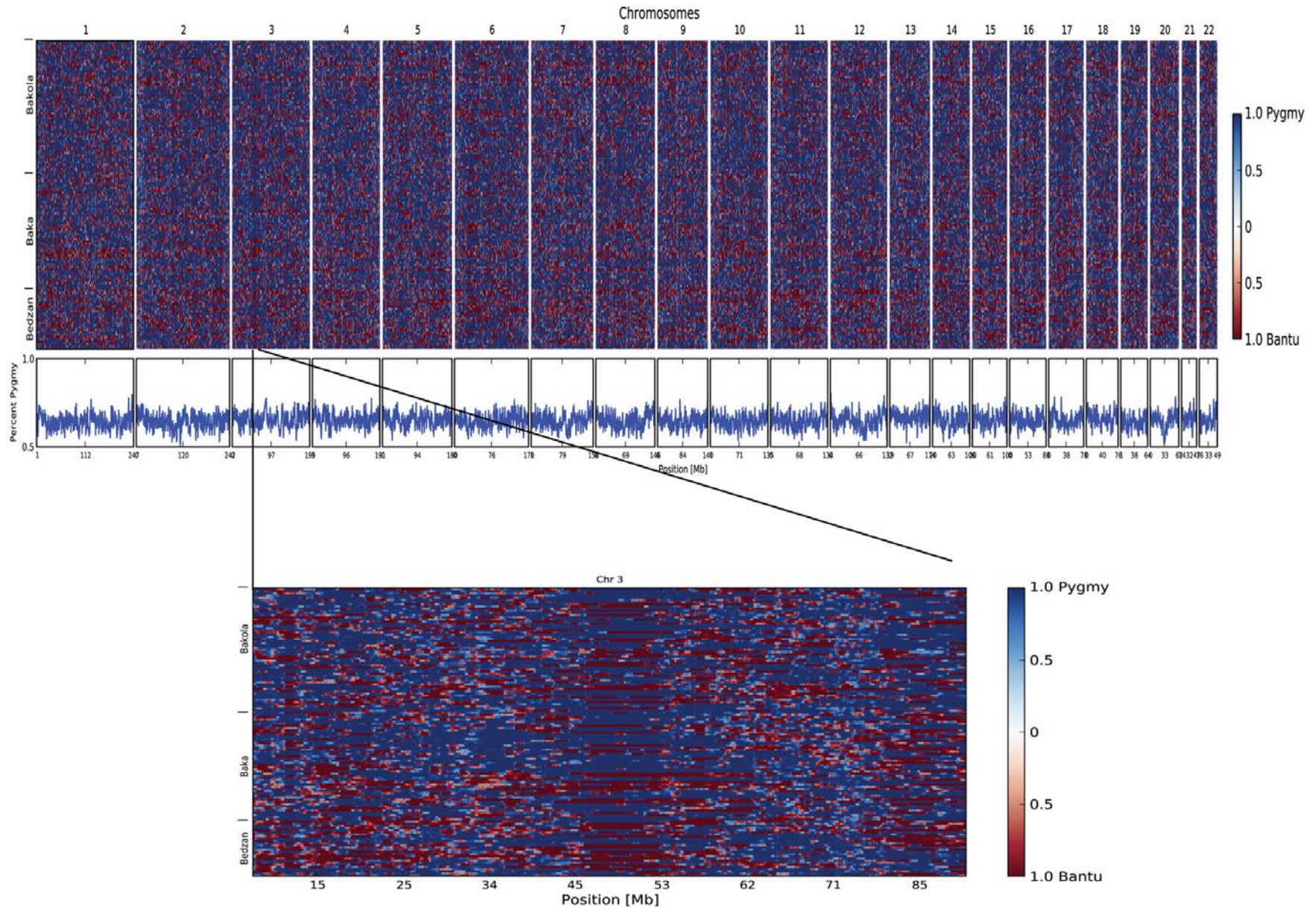
K = 2



Illumina 1M SNP array  
genotyped in 67 Pygmy and 58  
neighboring Bantu individuals  
Jarvis *et al.* *PLoS Gen.* 2012



# Support Mix uses a support vector machine training approach





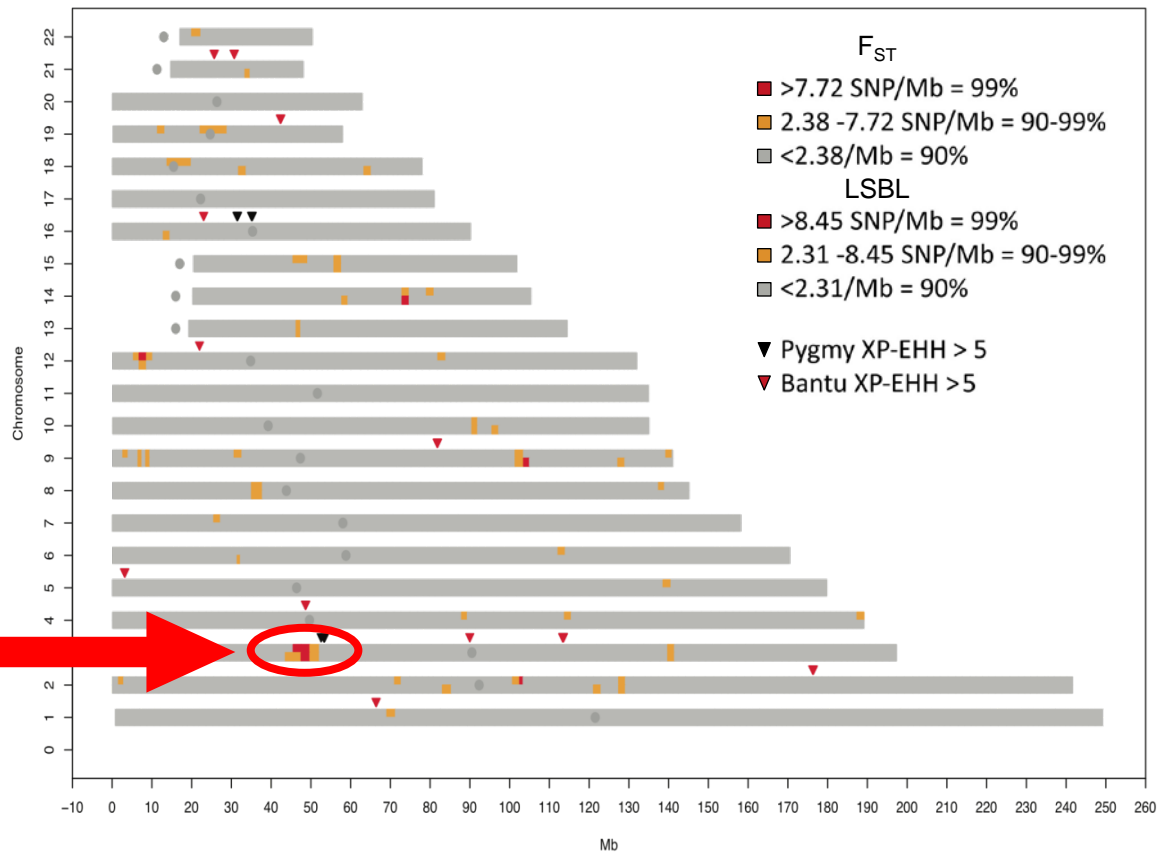
# How do the genomes of Pygmies differ from the genomes of other populations?



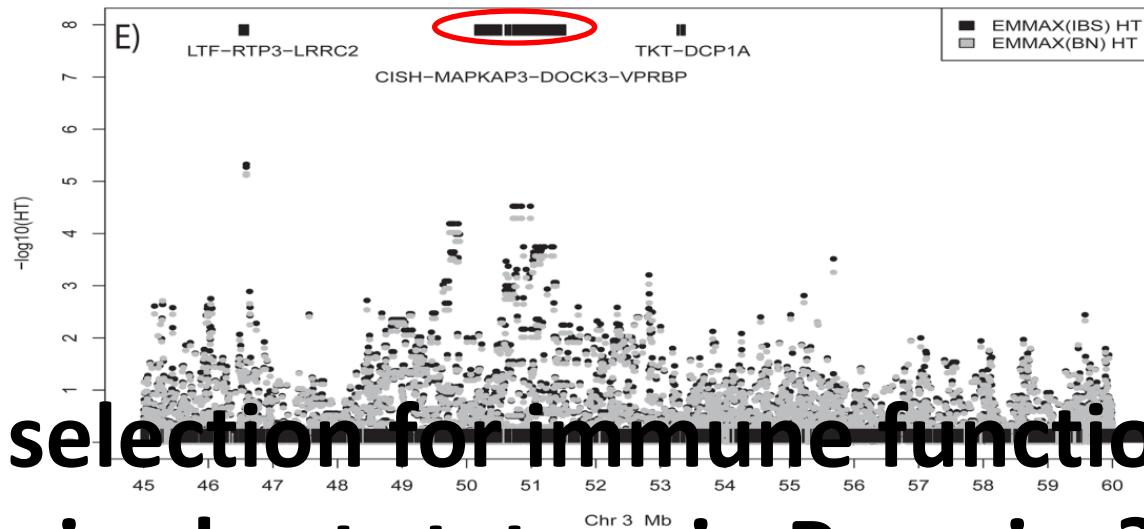
vs.



and



# Identification of a candidate region associated with short stature in Pygmies



## Could selection for immune function result in short stature in Pygmies?

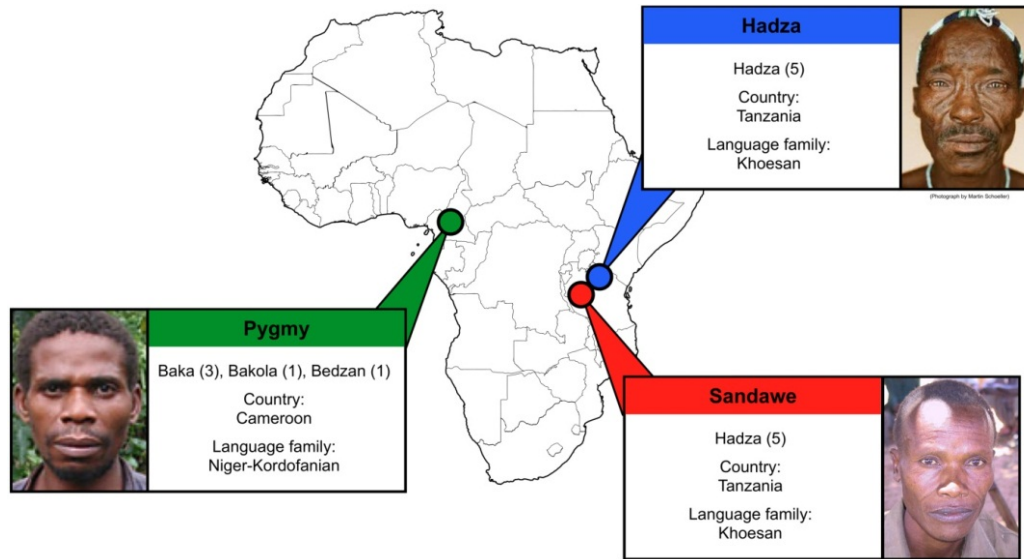
**DOCK3**

**CISH**

- also associated with height variation in non-African populations
- member of cytokine signaling (*SOCS*) family of proteins
- up-regulated by interleukin-2 (*IL-2*) and plays a critical role in cytokine signaling
- associated with resistance to bacteremia, malaria, and tuberculosis in Africa

**CISH directly inhibits HGHR action** by blocking the STAT5 phosphorylation pathway

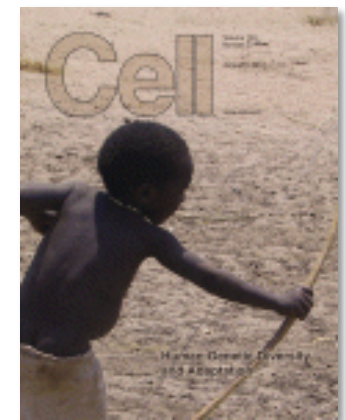
# High Coverage Genome Sequencing



Genomes from 15 hunter-gatherers contain  
**13,420,318** variants

**3,062,541** variants are novel

Many novel variants are in known  
regulatory sites



Lachance, *et al.*  
*Cell*, 2012

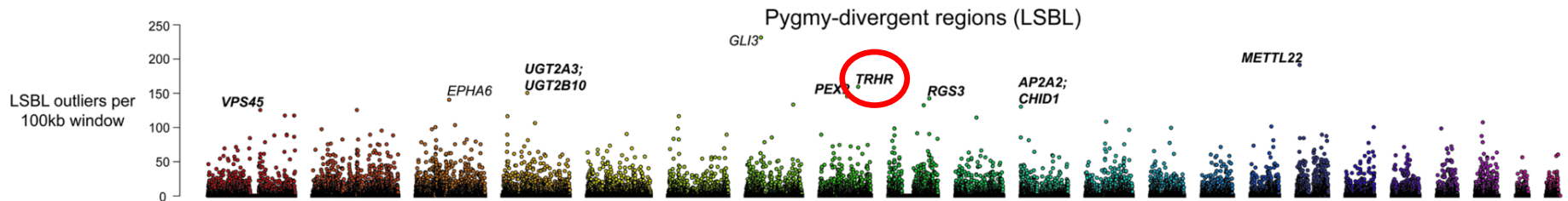


# Which pathways are enriched for genes near targets of selection in Pygmies?

## 100 kb windows near targets of selection

(Lachance *et al. Cell* 2012)

(Jarvis *et al. PLoS Gen.* 2012)



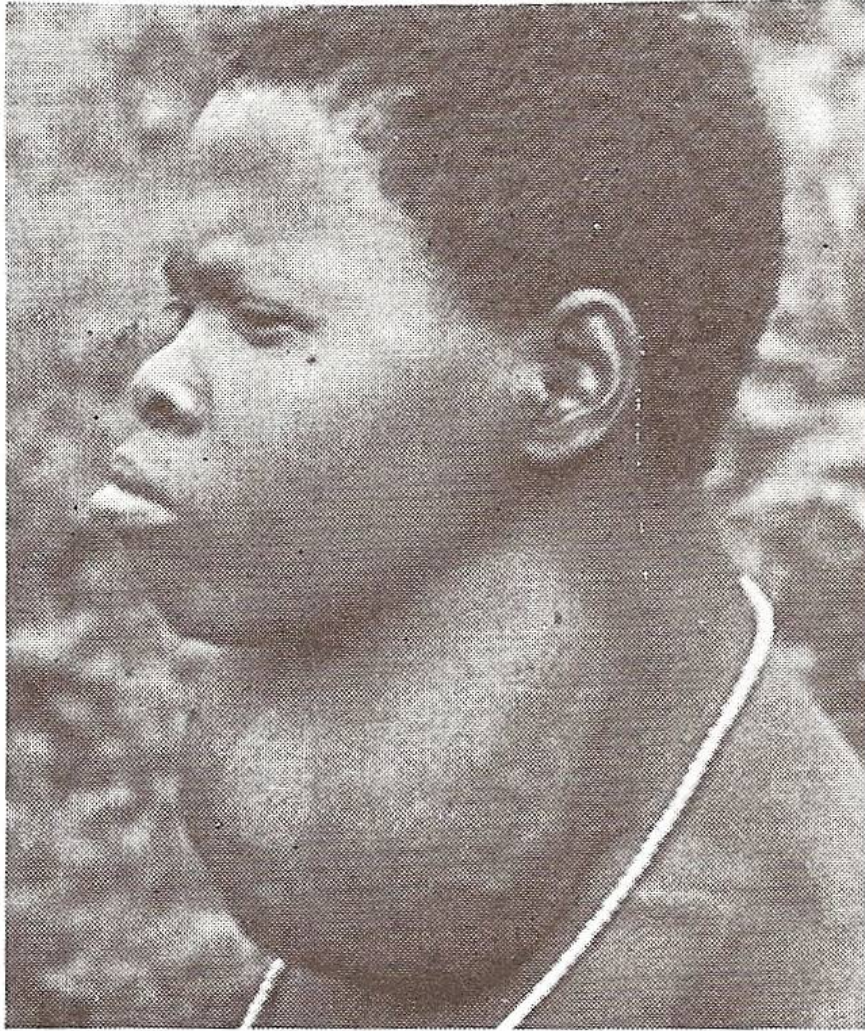
Enriched pathways include genes involved in:

- neuro-endocrine signaling
- reproduction
- metabolism
- immune function

We observe a significant enrichment for genes that play a role in **pituitary function** in Pygmies:

- FSHR
- GHR
- HESX1
- TRHR

# TRHR signaling pathway



Bantu individual with goiter

Regulation of the hypothalamic-pituitary–thyroid axis

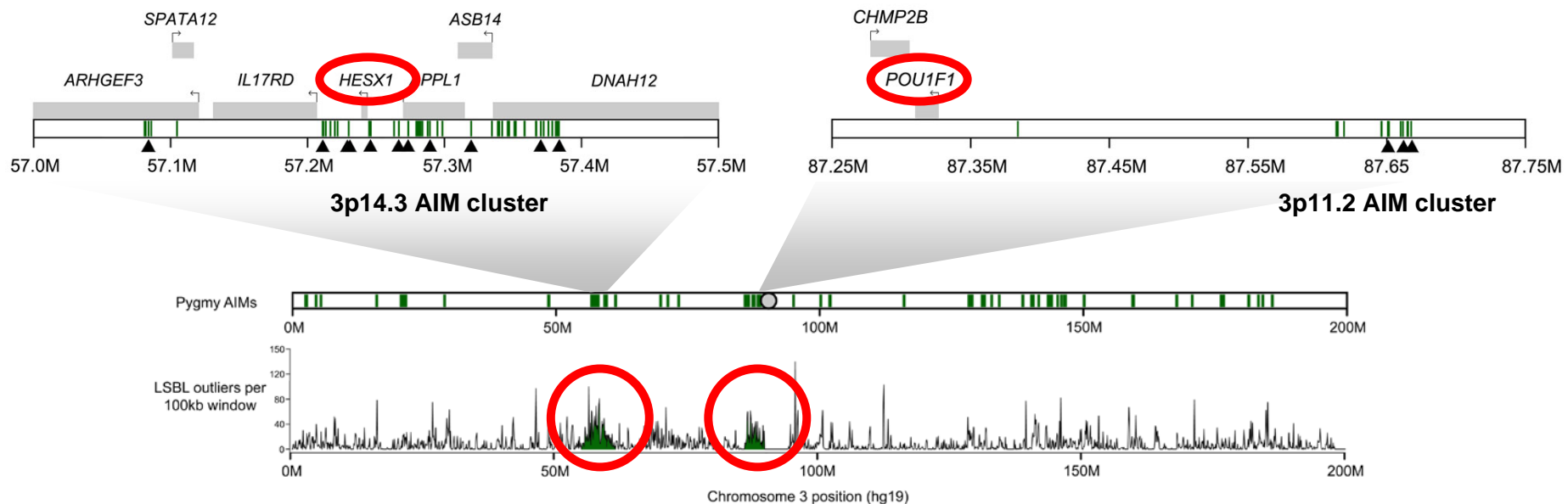
Influences growth, thermo-regulation, reproduction, and immune response

Pygmies have low levels of Goiter compared to neighboring Bantu populations

Pygmies may possess a biological adaptation to a low iodine environment

# Pygmy Ancestry Informative Markers (AIMs) on Ch. 3

Lachance *et al. Cell* 2012



**HESX1:** 44 SNPs in 100% LD over 170 kb, missense mutation in *HESX1* (Asn125Ser)

**POU1F1:** 12 SNPs in 100% LD ~200 kb upstream

The HESX1 and POU1F1 AIM cluster regions are the most differentiated in the genome compared to other African populations

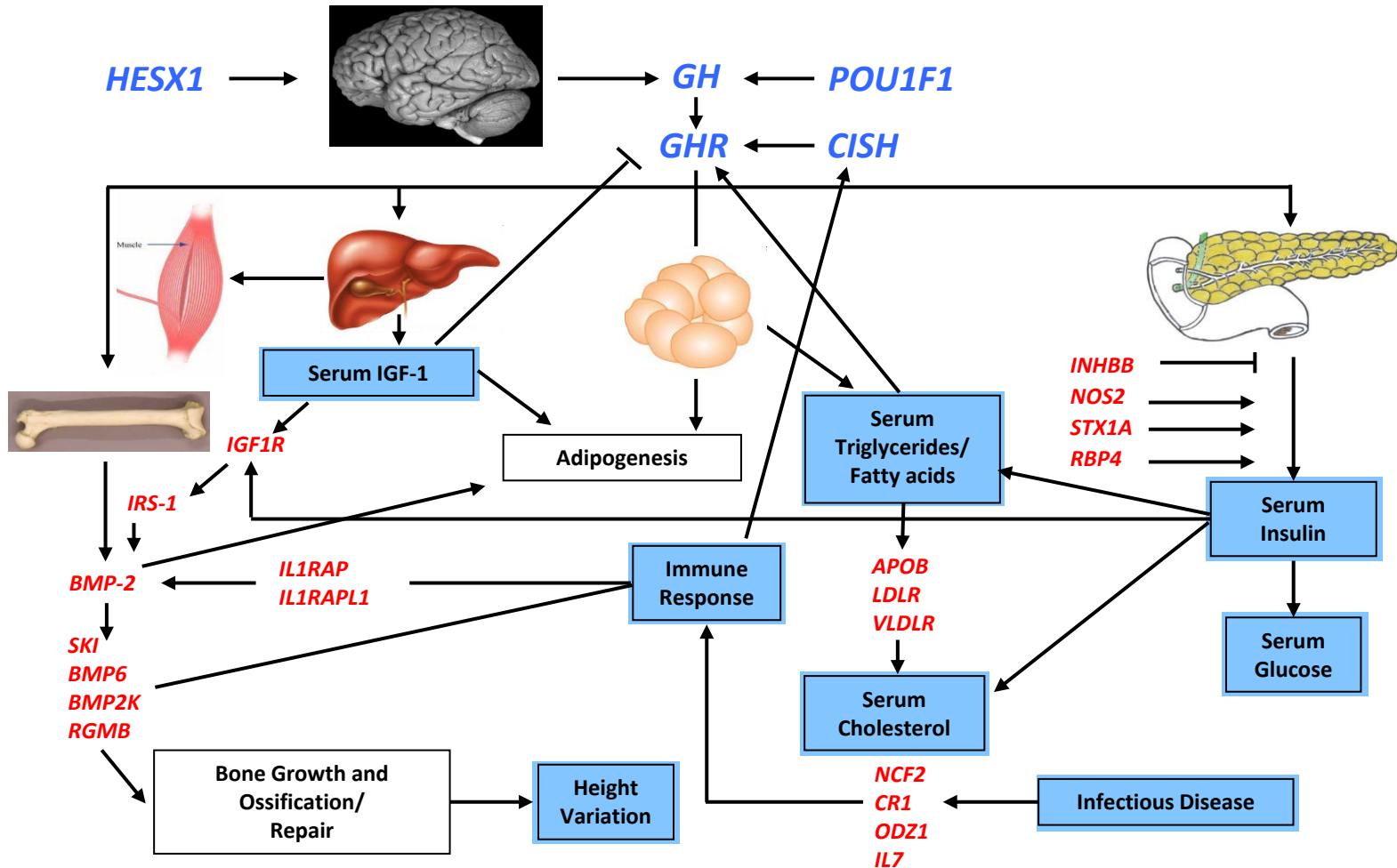
**HESX1** and **POU1F1** AIM SNPs are common in western and eastern Pygmies

**HESX1** and **POU1F1** AIM SNPs are statistically associated with short stature in Pygmies

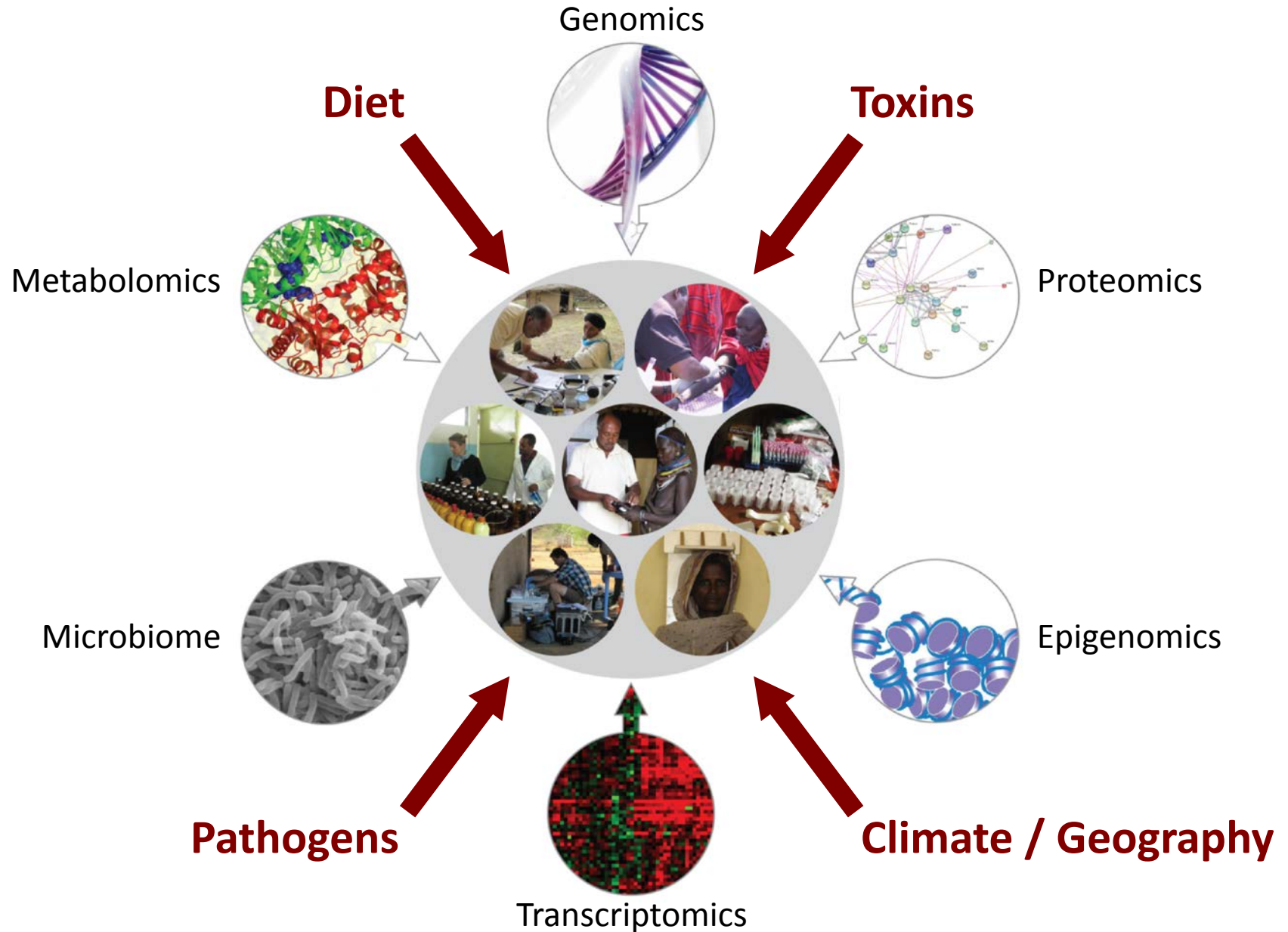


# Hypotheses

Development of the **GH/IGF1 pathway** may be produced by the **metabolic** or **genetic** may show a **central role** in the **regulation of metabolism**, acting on **pleiotropic** number of traits, including **growth, reproduction, metabolism, and immunity**.



# Future Challenges: Integrative Evolutionary Genomic Analyses of Complex Traits





# What is the role of environment on physiologic traits?

## Urban

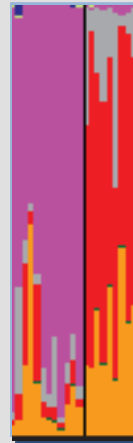


## Rural



# Distinct Genetic Ancestry, Different Disease Susceptibility

Fulani, Cameroon, pastoralist

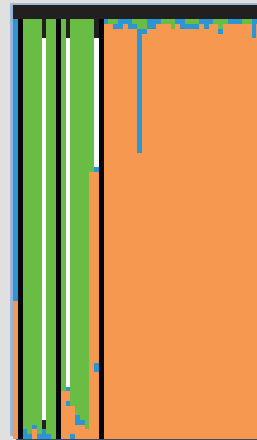


Fulani Tupuri

Tupuri, Cameroon, agro-pastoralist



San, northern Botswana,  
hunter-gatherer



San Bantu



Bantu Speaker, northern Botswana,  
agro/pastoralist/fisherman



# Conclusions



Africans have the highest levels of genetic diversity within and among populations.



The demographic history of Africans and local adaptation to diverse environments has resulted in population or region specific genetic variation.

We need to include ethnically diverse Africans in genomic studies to identify both unique rare and common variants in Africa which are of functional importance, including those associated with disease risk

# Thanks



## Tishkoff lab, University of Pennsylvania

Joseph Lachance / Joe Jarvis / Laura Scheinfeldt /  
Sameer Soi / Clara Elbers / Bart Ferwerda / Bill Beggs /  
Simon Thompson / Alessia Ranciaro / Jibril Hirbo

## Sequencing Study

Josh Akey / Benjamin Vernot / Wenqing Fu / Kun Zhang  
/ Timothy Rebbeck / Alain Froment / Jean-Marie Bodo /  
Godfrey Lema / Thomas Nyambo / Larson Omberg /  
Jason Mezey

## Marshfield Study

Jim Weber / Francoise Friedlander / Mike Smith /  
Scott Williams / Sabah Omar / Muntaser Ibrahim /  
Thomas Nyambo / Godfrey Lema / Charles Wambebe  
/ Ogabara Doumbo / Mohamadu Thera / Alain  
Froment

## Lactase Study

BF Voight / CC Babbit / JS Silverman / K Powell /  
H Mortensen / JB Hirbo / M Osman / M Ibrahim /  
SA Omar / G Lema / TB Nyambo / J Ghorri /  
S Bumpstead / JK Pritchard / GA Wray / P Deloukas







ank