Measuring relatedness in human populations from DNA

Alejandro Ochoa, John D. Storey Lab, Princeton University

DrAlexOchoa   viiia.org/research/   ochoa@princeton.edu
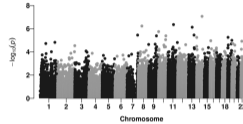
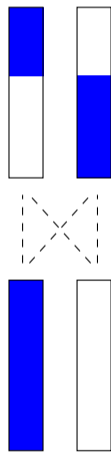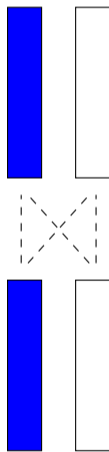Princeton Research Day 2017

Why study relatedness?

Human genetics is fascinating!

Search for disease-causing genetic variants
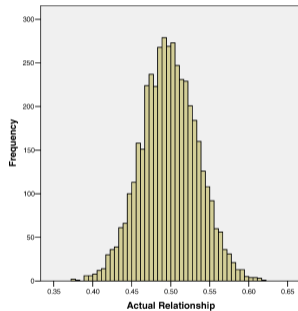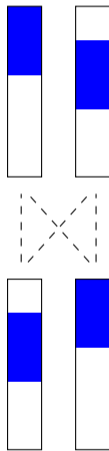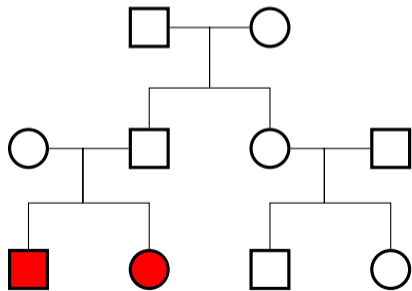

Heritability of complex traits

Animal and plant breeding

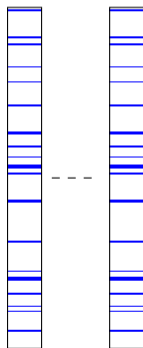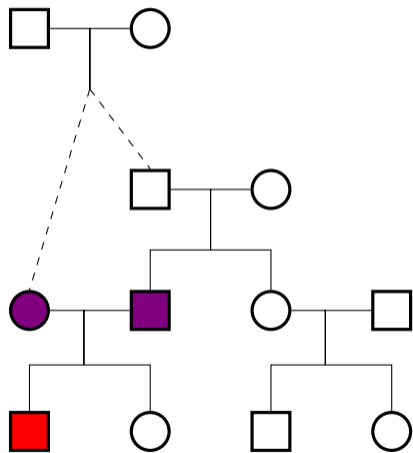# The kinship coefficient for parent-child: $\frac{1}{4}$
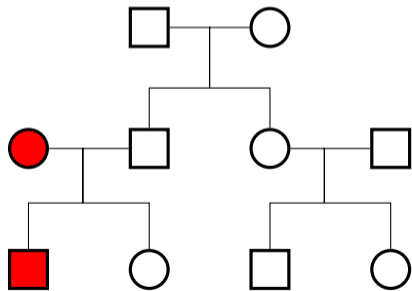
The kinship coefficient for siblings: $\frac{1}{4}$ on average

Visscher *et al.* (2006)

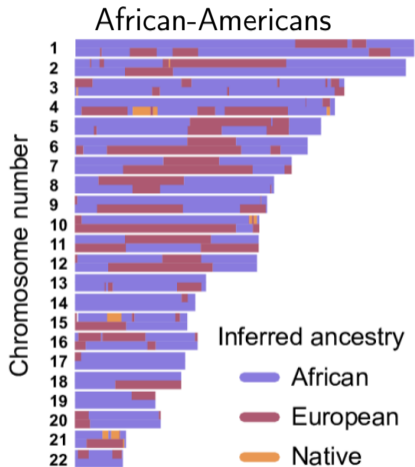# The inbreeding coefficient in populations

Measurements relative to a reference pop.:

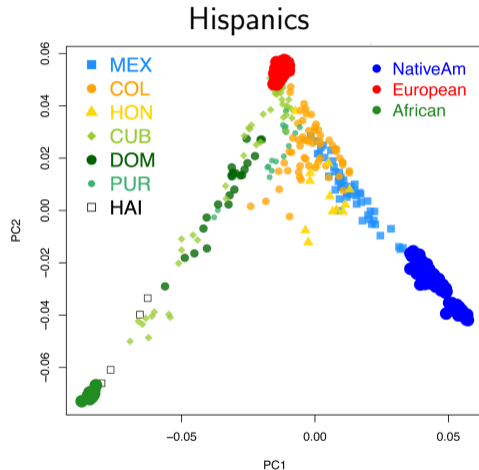Inbreeding $= 0$ in the local population

Inbreeding ≥ 0 relative to a distant ancestral population

Better measured using covariance

Recently admixed populations

African-Americans

Baharian *et al.* (2016)

Hispanics

Moreno-Estrada *et al.* (2013)

Kinship model for genotypes

symbol	meaning
T	ref ancestral population
i	locus index
j, k	individual indexes
p_i^T	ref allele frequency
x_{ij}	genotype (num ref alleles)
φ_{jk}^T	kinship of j, k
f_j^T	inbreeding of j

(Wright 1921, 1951; Malécot 1948; Jacquard 1970).

Statistical model:

$$E[x_{ij}|T] = 2p_i^T,$$
$$\text{Var}(x_{ij}|T) = 2p_i^T \left(1 - p_i^T\right)\left(1 + f_j^T\right),$$
$$\text{Cov}(x_{ij}, x_{ik}|T) = 4p_i^T \left(1 - p_i^T\right)\varphi_{jk}^T.$$

We developed a new kinship estimation framework that works for arbitrary population structures!
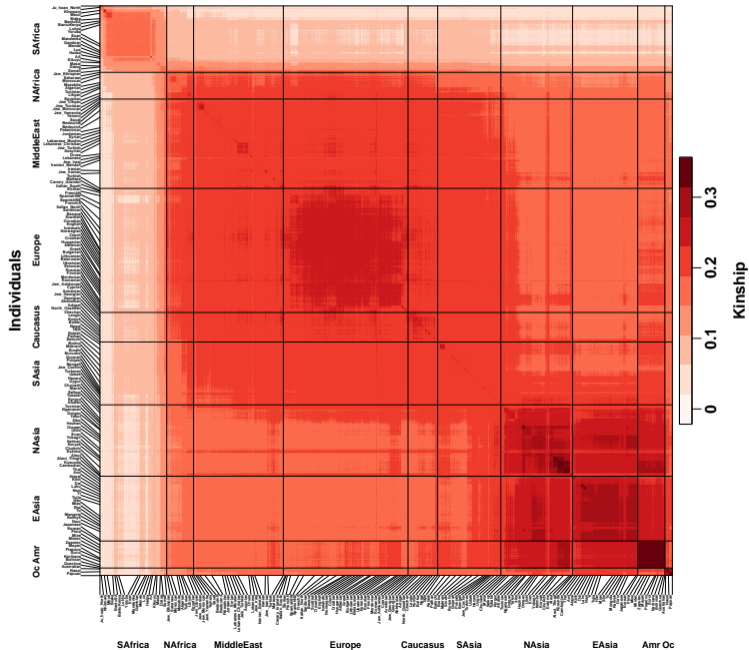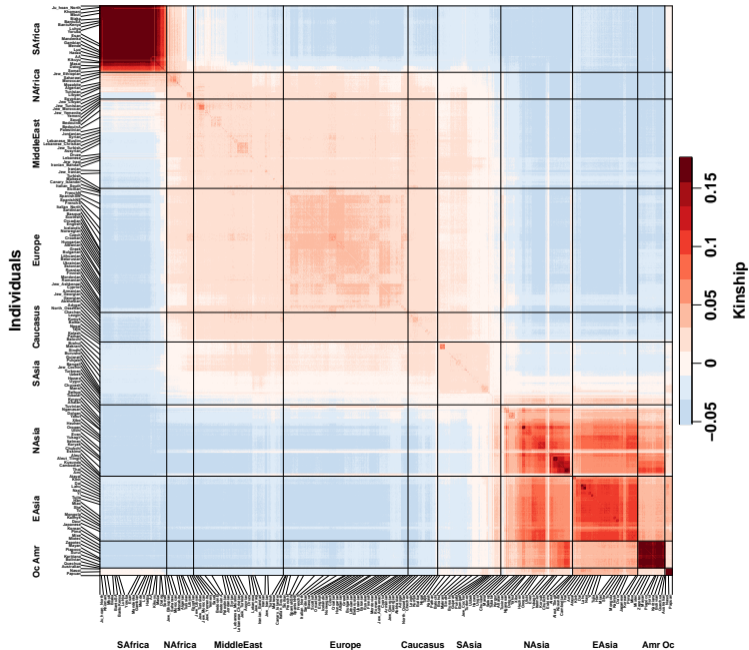
New kinship estimates

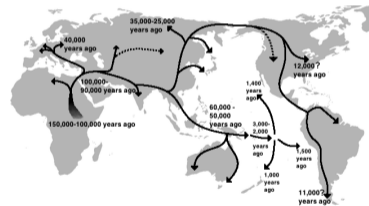Genotypes from "Human Origins" (Lazaridis et al. 2014, 2016)
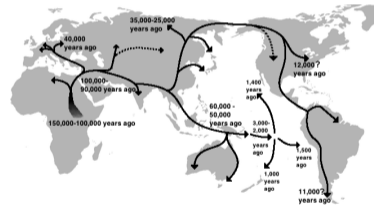
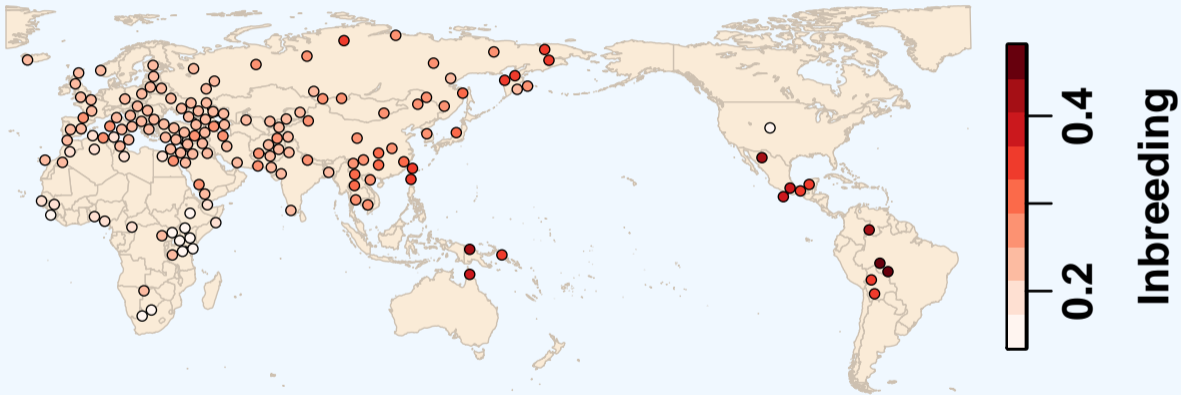By Ephert [CC BY-SA 3.0], via Wikimedia Commons

Standard kinship estimates

Genotypes from "Human Origins" (Lazaridis et al. 2014, 2016)

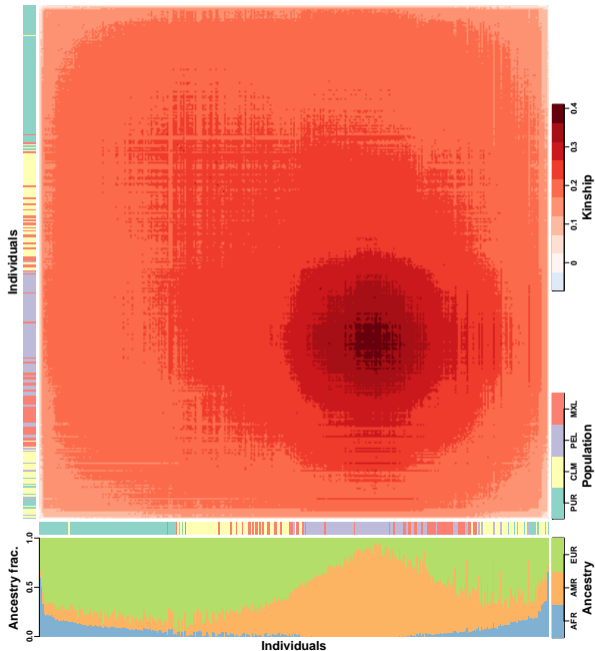By Ephert [CC BY-SA 3.0], via Wikimedia Commons

Population-level inbreeding

Kinship driven by admixture in Hispanics

New kinship estimates

Genotypes from the 1000 Genomes Project (2012)

Acknowledgments

John D. Storey
Andrew Bass
Irineo Cabreros
Chee Chen
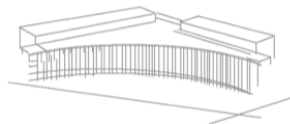**Wei Hao**
Emily Nelson
Riley Skeen-Gaar

Neo Christopher Chung
Institute of Informatics
University of Warsaw

PRINCETON
UNIVERSITY

Lewis-Sigler Institute for Integrative Genomics

CENTER FOR
STATISTICS AND
MACHINE LEARNING