Week 3: 1000 Genomes

General Stuff:
- Week 4 reading?
- Updates to unanswered questions: Piazza
- Domain OH: **CSB 106**
- Today: 1000 Genomes!

Discussion Questions

- What is the 1000 Genomes Project?
- What **technologies** were used for data collection? Generally, how do these technologies work?
- How does **human variation differ across the world**?
- What is a reference allele? A non-reference allele?
- What is a **haplotype**? What is **linkage disequilibrium**? (will likely have to search beyond the paper for this one)
- What are three (3) lessons you learned from reading the 1000G paper?

HapMap Project

- First release: 2005
- HapMap3: 1301 samples (incl. 270 samples from Phase I and Phase 2)
- 11 populations (originally)
  - ASW: African Ancestry (Southwest, USA)
  - CEU: Northern/Western Europeans (Utah)
  - CHB: Han (Beijing, China)
  - CHD: Chinese (Metro Denver, CO)
  - GIH: Gujarati (Houston, TX)
  - JPT: Japanese (Tokyo)
  - LWK: Luhya (Kenya)
  - MEX: Mexican (LA, USA)
  - MKK: Maasai (Kenya)
  - TSI: Toscans (Italy)
  - YRI: Yoruba (Nigeria)

https://www.broadinstitute.org/medical-and-population-genetics/hapmap-3

1000 Genomes Project: Pilot (2010)

- **Samples**: HapMap samples (volunteer donors; presumably healthy; provided informed consent)
- **Experiments**:
  - low-coverage whole-genome sequencing: 179 individuals
  - high-coverage sequencing: two mother–father–child trios
  - exon-targeted sequencing: 697 individuals from seven populations

"aim of the 1000 Genomes Project is to discover, genotype and provide accurate haplotype information on all forms of human DNA polymorphism in multiple human populations"

1000 Genomes:

Pilot: 2010 (179 genomes; 3.5 populations)

Phase I: 2012 (1,092 human genomes; 14 populations)
- British from England/Scotland (GBR)
- Colombians (CLM)
- Han Chinese from South China (CHS)
- Iberian Populations from Spain (IBS)
- Puerto Ricans (PUR)

Phase III: 2015 (2,054 human genomes; 26 populations)
- Whole-genome sequencing (7.4x)
- Exome-sequencing (65.7x)
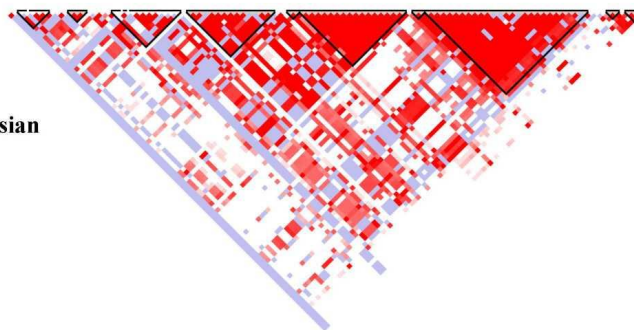- First degree relatives were genotyped using SNP microarray
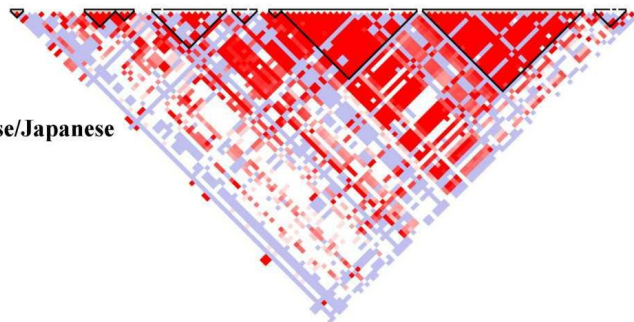
Linkage Disequilibrium

non-random association of alleles at different loci in a given population

Caucasian

Chinese/Japanese

Yoruba

Discussion: Why do individuals of African descent (here: Yorubans) have smaller LD blocks?

Europe
45,000
years ago

Levant and
Arabian peninsula
120,000 to 90,000
years ago

North Asia
20,000
years ago

Americas
15,000 years ago

Homo sapiens in Africa
150,000 to
200,000 years ago

South Asia, Indonesia
and Australia
50,000 years ago

https://www.sciencenewsforstudents.org/article/globes-non-africans-all-descend-single-move-out-africa