Introduction to Advanced Population Genetics
Learning Objectives

• Describe the basic model of human evolutionary history
• Describe the key evolutionary forces
• How demography can influence the site frequency spectrum
  • Be able to interpret a site frequency spectrum
  • Understand how the SFS is affected by evolutionary forces
We now have an excellent “road map” of how humans evolved in Africa and migrated to populate the rest of the earth.
How has our population size grown?

Tennesen et al. (2012) Science
Review: What are the assumptions of Hardy-Weinberg?

1) There must be no mutation
2) There must be no migration
3) Individuals must mate at random with respect to genotype
4) There must be no selection
5) The population must be infinitely large

How do these affect allele frequencies?
Drift, mutation, migration, and selection

**Natural Selection**

- Selection Coefficient
- Derived allele frequency
- Allele frequency

**Genetic Drift**

- $1/(2N_e)$
- $N_e = 500$
Genetic drift: Serial founder effect
Heterozygosity is correlated with distance from East Africa

Ramachandran et al. 2005 PNAS
## Mutation: How often do mutations arise?

<table>
<thead>
<tr>
<th>study</th>
<th>loci considered</th>
<th>per-generation mean mutation rate (10^{-8} \text{bp}^{-1} \text{generation}^{-1})</th>
<th>yearly mean mutation rate (10^{-9} \text{bp}^{-1} \text{y}^{-1})</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>(t_{\text{gen}} = 30 \text{y})</td>
</tr>
<tr>
<td>Kondrashov (2003)</td>
<td>disease</td>
<td>1.85 (0.00–3.65)</td>
<td>0.62 (0.00–1.22)</td>
</tr>
<tr>
<td>Lynch (2010)</td>
<td>disease</td>
<td>1.28 (0.68–1.88)</td>
<td>0.42 (0.23–0.63)</td>
</tr>
<tr>
<td>Roach et al. (2010)</td>
<td>WG</td>
<td>1.10 (0.68–1.70)</td>
<td>0.37 (0.23–0.57)</td>
</tr>
<tr>
<td>Awadalla et al. (2010)</td>
<td>WG</td>
<td>1.36 (0.34–2.72)</td>
<td>0.45 (0.11–0.91)</td>
</tr>
<tr>
<td>1000 Genomes Project (2010), CEU</td>
<td>WG</td>
<td>1.17 (0.94–1.73)</td>
<td>0.39 (0.31–0.57)</td>
</tr>
<tr>
<td>1000 Genomes Project (2010), YRI</td>
<td>WG</td>
<td>0.97 (0.72–1.44)</td>
<td>0.32 (0.24–0.48)</td>
</tr>
<tr>
<td>Sanders et al. (2012)</td>
<td>exome</td>
<td>1.28 (1.05–1.50)</td>
<td>0.43 (0.35–0.50)</td>
</tr>
<tr>
<td>O’Roak et al. (2012)</td>
<td>exome</td>
<td>1.57 (1.05–2.26)</td>
<td>0.52 (0.35–0.75)</td>
</tr>
<tr>
<td>Kong et al. (2012)</td>
<td>WG</td>
<td>1.20</td>
<td>0.40</td>
</tr>
</tbody>
</table>

What are the effects of paternal age on mutation rate?

Kong et al. (2012) Nature
When did most variation arise?

Fu et al. (2013) Nature
Most SNVs are very rare

- 57% singletons
- 72% singletons + doubletons + tripletons

Tennessen et al. (2012) Science
Most SNVs are population specific

503,481 SNVs

AA (n = 1,089)

EA (n = 1,351)

177,419

91,703

234,359

Proportion of Population-specific SNVs

Minor Allele Count

Expected Proportion

Sample Size

Tajima's D

PolyPhen2

SFS-Del

GERP

48943

13537

5683

93

2069

7241

0.00

0.25

0.50

0.75

1.00

≤0.5

0.5<x≤1

>1

MAF(\%)

0

125

250

375

500

All

Missense

Synonymous

Nonsense

Splice

Known / Novel

Tenessen et al. (2012) Science
Recent admixture: Migrations can have a profound effect on genetics

Ancient admixture: Neanderthals are still among us

- Recent genetic data suggests that 1-4% of non-African genomes are derived from Neanderthals.
Neanderthals are still among us
Estimates of global ancestry

CAAPA

Reference

European (CEU)  African (YRI)

Native American

Local ancestry of a single individual

The Bantu expansion occurred ∼4,000 years ago, originating in Cameroon or Nigeria and expanding throughout sub-Saharan Africa (40, 41). The clustering of the Xhosa, Fang, Bamoun, and Kongo populations, all of which are Bantu Niger-Kordofanian-speaking populations, likely reflect a Bantu migration from Nigeria/Cameroon expanding toward the south. Although we have limited sample sizes (with three of our populations having sample sizes of less than 10), the relative order of clustering (the East-West axis, followed by the North-South axis) suggests that the strongest differentiating axis among the African populations is linguistic classification corresponding to Chadic and Nilo-Saharan vs. Niger-Kordofanian ancestry. The relatively weaker North-South axis may result from the genetic similarity among the Niger-Kordofanian linguistic groups because of their recent common ancestry. Although sampled in Nigeria, the very distinct Fulani are part of a nomadic pastoralist population that occupies a broad geographical range across Central and Western Africa. Analyses of microsatellite and insertion/deletion polymorphisms indicate that they share ancestry with Niger-Kordofanian, North African, and Central African Nilo-Saharan populations, as well as low levels of European and/or Middle Eastern ancestry (2).

Exempting the Fulani, our LD analyses show no large differences in rates of LD decay among our sampled African populations, with all populations exhibiting a faster decay of LD (i.e., larger inferred effective population size) than previously characterized populations of European ancestry (see SI Text).

Interestingly, the Kongo population does not follow the overall trend of East-West and North-South clustering. The Kongo population’s genetic proximity to geographically distant Bantu populations from Cameroon could be explained by the genetic similarity of Bantu-speaking populations in the region, as seen in the FRAPPE analyses (Fig. 1). Alternatively, although these individuals self-identified as Kongo and were refugees from locations within the Democratic Republic of Congo, the samples were collected in Cameroon; therefore, self-identified ancestry might poorly represent the long-term geographical origins or may reflect recent admixture.
Adaptive (Darwinian) Selection

“I have called this principle, by which each slight variation, if useful, is preserved, by the term Natural Selection.” —Charles Darwin from “The Origin of Species”, 1859
Antibiotic resistance is an example of adaptive evolution.
Reading the genome for signatures of positive selection

- This process imparts “signatures” on patterns of genetic variation that we can use to find adaptively evolving genes
Genes that influence physical traits have been targets of recent selection.

- **Eye Color**: HERC2
- **Skin Pigmentation**: SLC24A5, OCA2, TYRP1
- **Hair Texture**: EDAR

These forces all affect the Site Frequency Spectrum (SFS)
Primer on coalescent
Primer on coalescent

\[ E(T_i) = \frac{2}{i(i - 1)} \quad V ar(T_i) = \left( \frac{2}{i(i - 1)} \right)^2 \]

To generate a genealogy of \( i \) genes under Kingman’s coalescent:

- Draw an observation from an exponential distribution with mean \( \mu = 2/(i(i - 1)) \). This will be the time of the first coalescent event (looking from the present backwards in time).

- Pick two lineages at random to coalesce.

- Decrease \( i \) by 1.

- If \( i = 1 \), stop. Otherwise, repeat these steps [8, 9].
Site Frequency Spectrum (SFS)
Joint Site Frequency Spectrum (JSFS)
Genetic Drift can profoundly shape allele frequencies.

\[ \theta = 4N_A \mu \]
\[ M = 2N_A m \]
\[ \tilde{N}_1 = \nu_1 N_A \]
\[ \tilde{N}_2 = \nu_2 N_A \]

\[ \tau = t/2N_A \]

---

Useful equations

Time: \( t = \frac{T}{4N_{\text{ref}}*\text{Gen}} \)
- \( N_{\text{ref}} = \) reference or ancestral population size
- \( \text{Gen} = \) number of years per generation
- \( T = \) chronological years

\[ \theta = 4\times N_{\text{ref}} \times \mu \times \text{Length}; \]
- \( \mu = \) mutation rate
- Length is the bp of the segment simulated (aka nsites for recombination)

Growth: \( N(t) = N(0)e^{-t\alpha} \)

Recombination: \( \rho = 4N_{\text{ref}}r \)
- \( r = \) the recombination rate between the ends of a unit length sequence

Migration: \( M_{ij} = 4N_{\text{ref}}m_{ij} \)
- \( m_{ij} = \) the fraction of subpopulation \( i \) that is made up of migrants from subpopulation \( j \) in forward time.
Concluding Summary

• OOA model is an isolation by distance model leading to modern day peopling of the globe with subsequent recontact in the last 500 years.

• Four main evolutionary forces are: Mutation, migration, selection, and drift.

• These forces change the site frequency spectrum in informative ways that we can use for both demographic analysis and simulation.