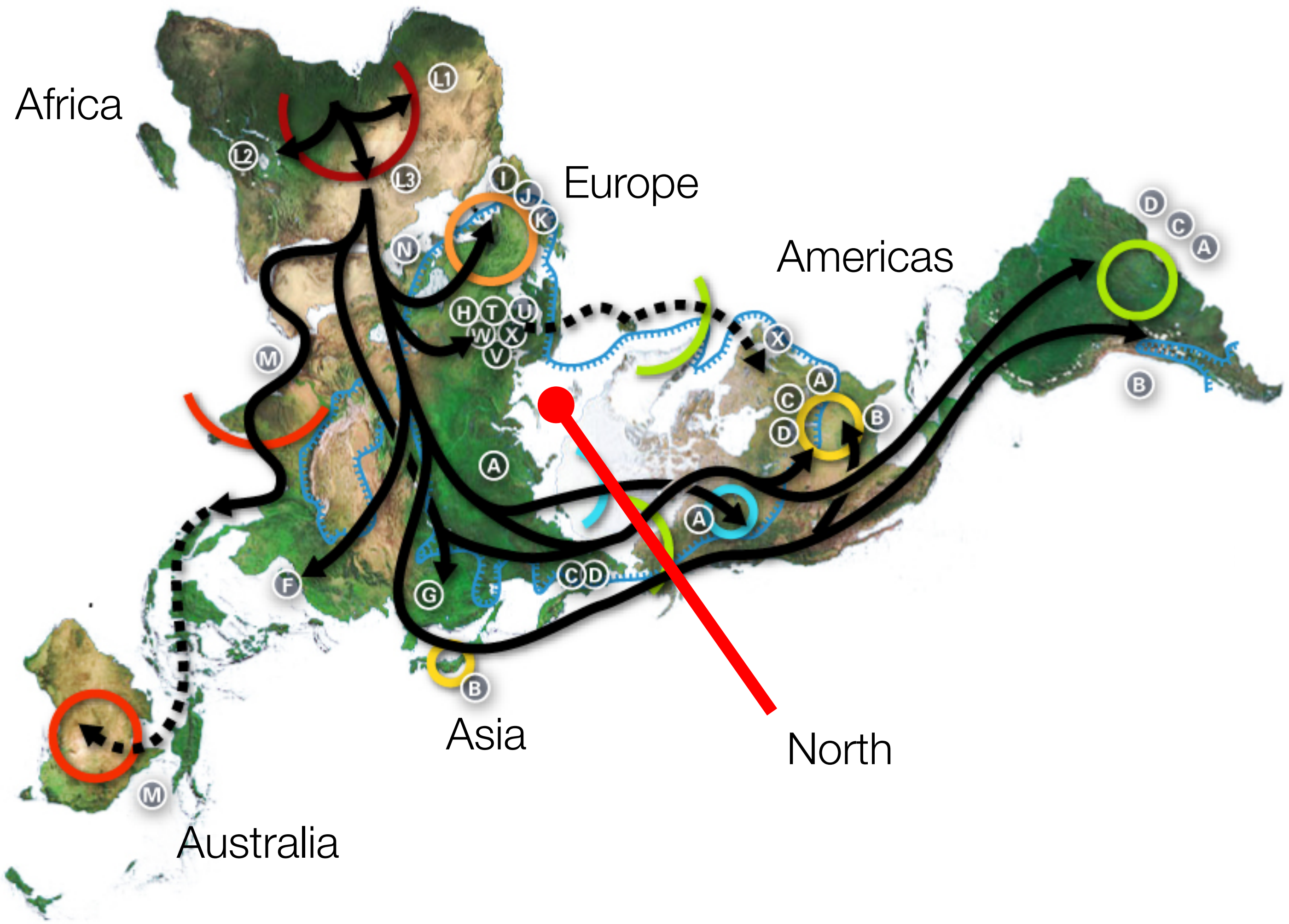
Introduction to Advanced Population Genetics

Learning Objectives

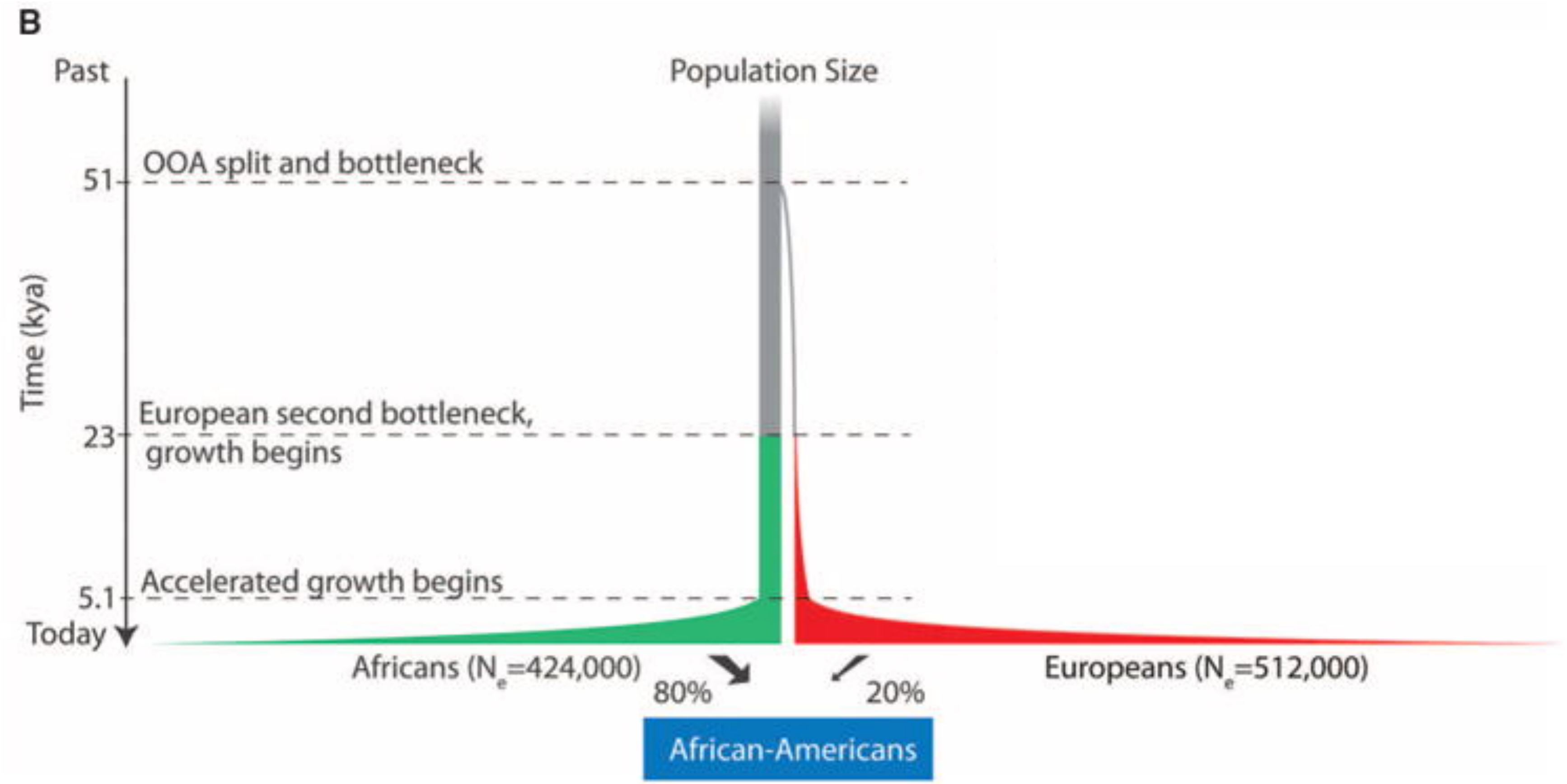
- Describe the basic model of human evolutionary history
- Describe the key evolutionary forces
- How demography can influence the site frequency spectrum
  - Be able to interpret a site frequency spectrum
  - Understand how the SFS is affected by evolutionary forces

Out of Africa Model!

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?



Tennessen et al. (2012) Science

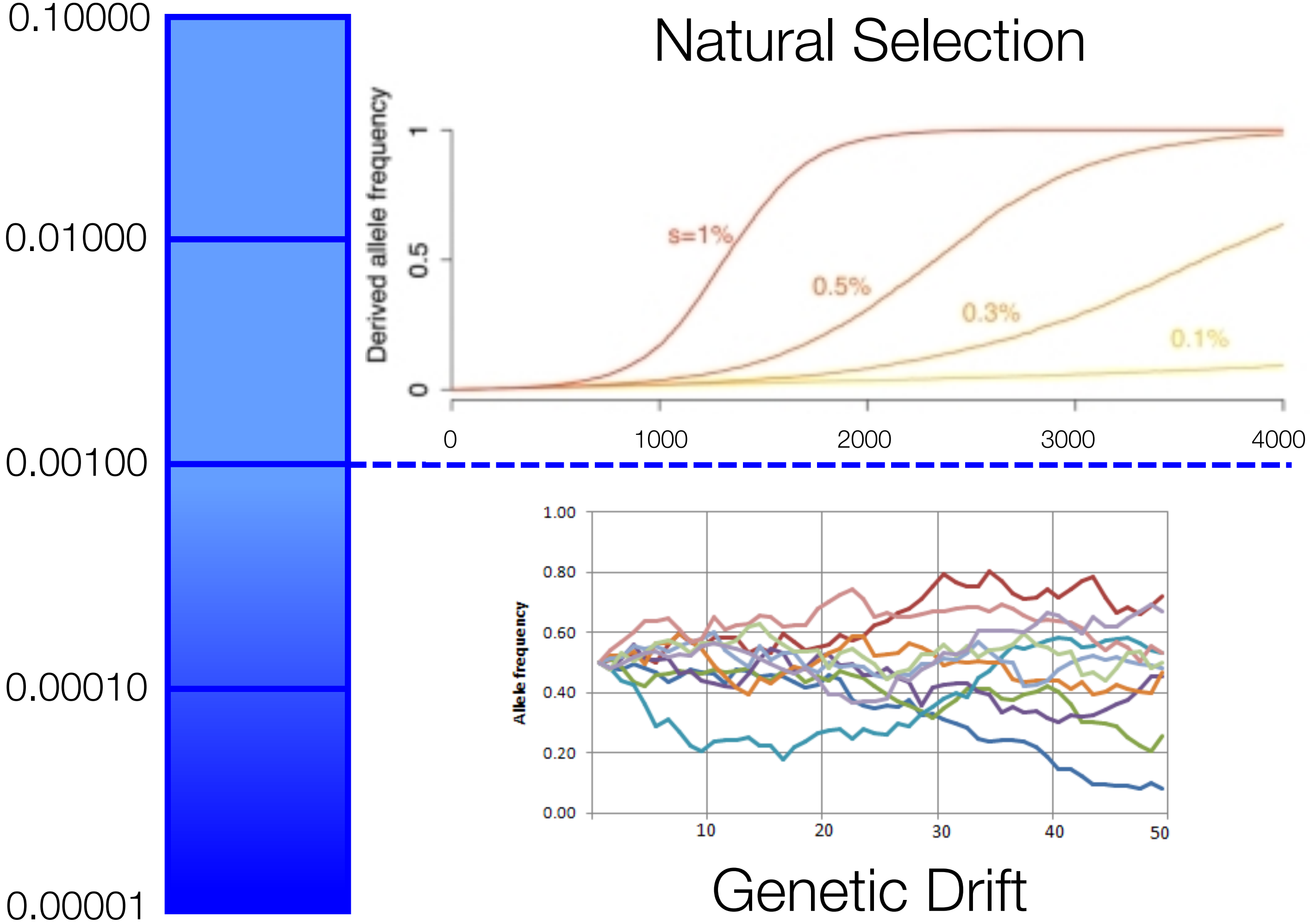
Review: What are the assumptions of Hardy-Weinberg?

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

Selection Coefficient



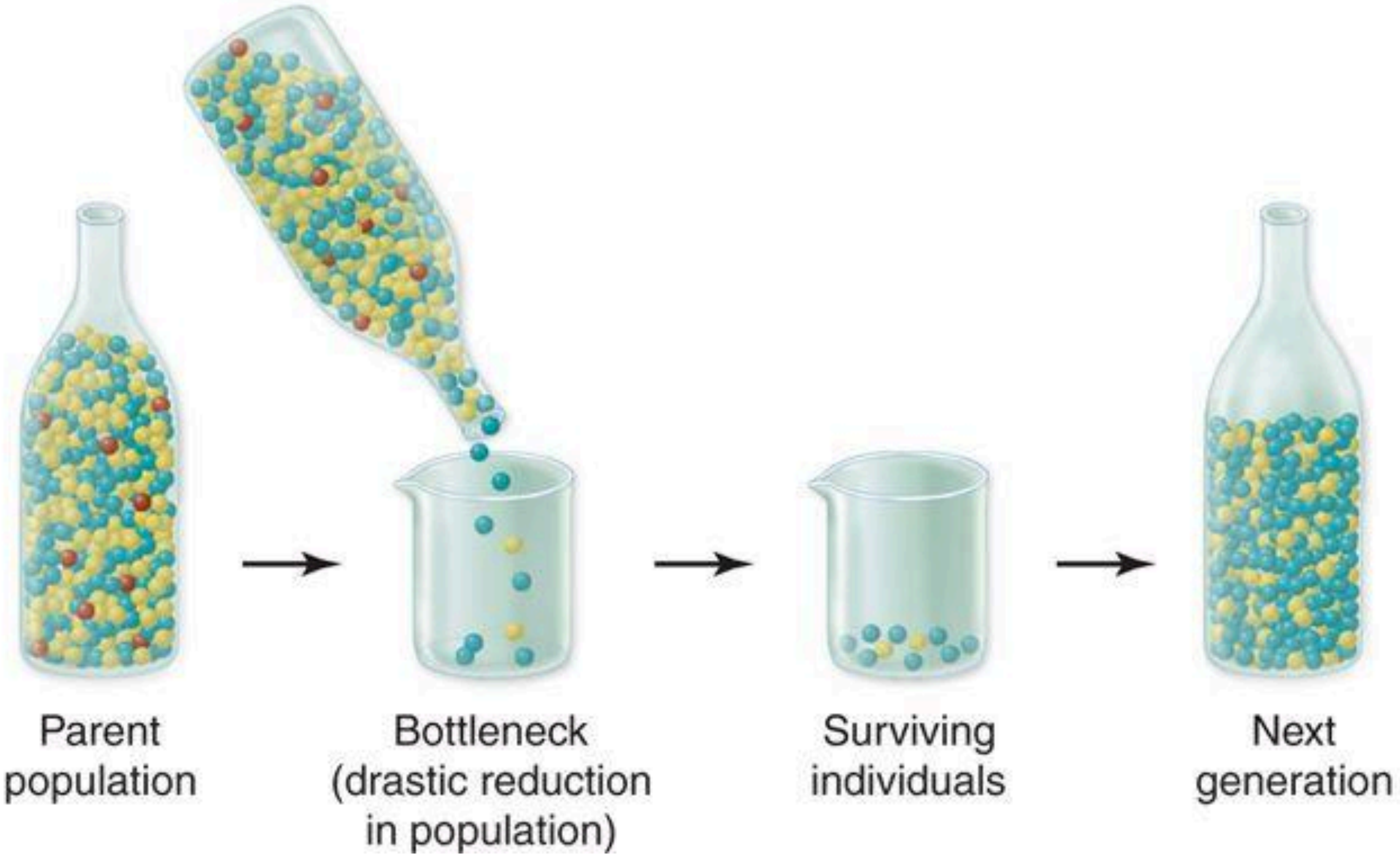
Natural Selection

Genetic Drift

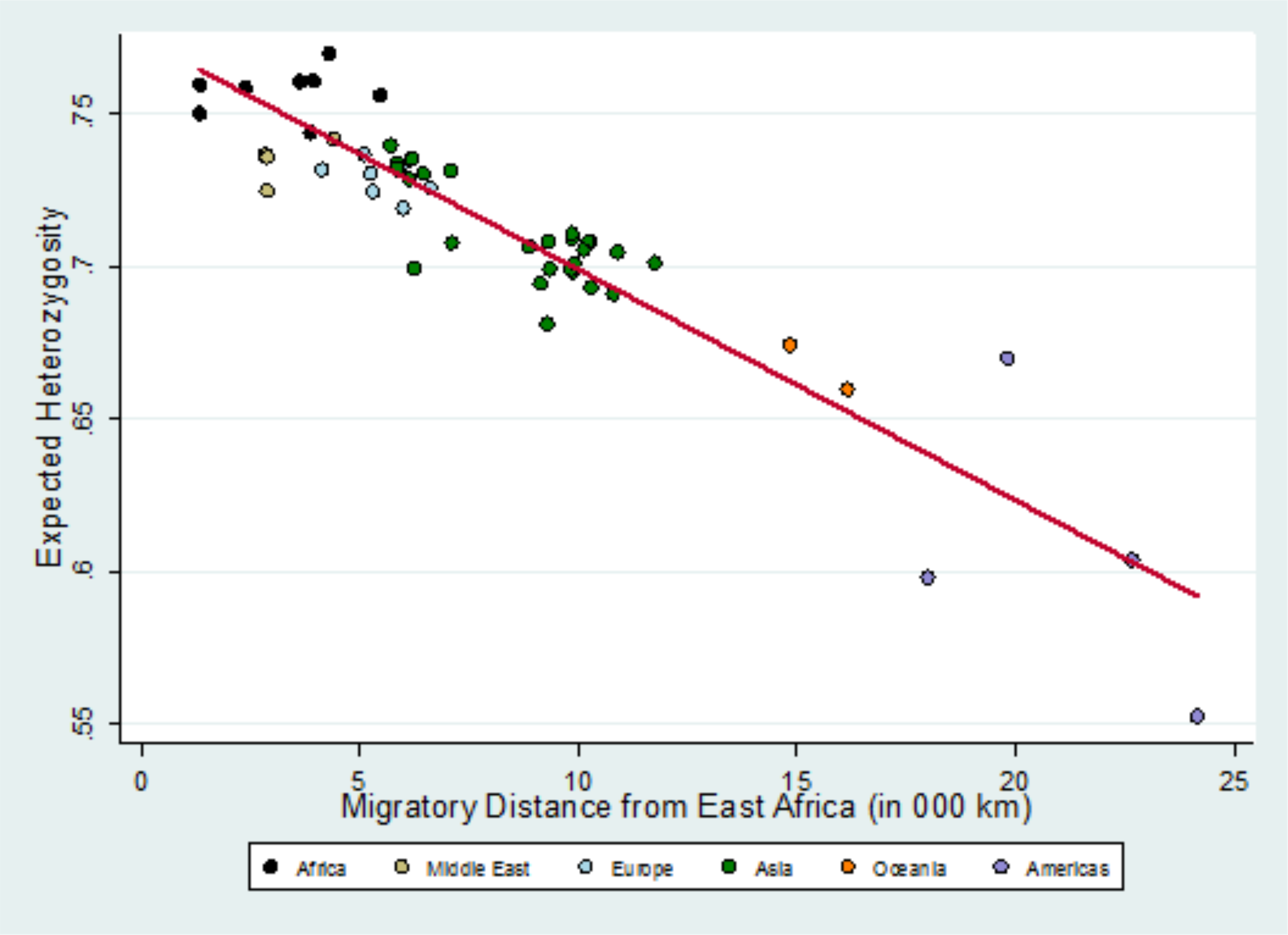
$1/(2N_e)$

$N_e = 500$

Genetic drift: Serial founder effect

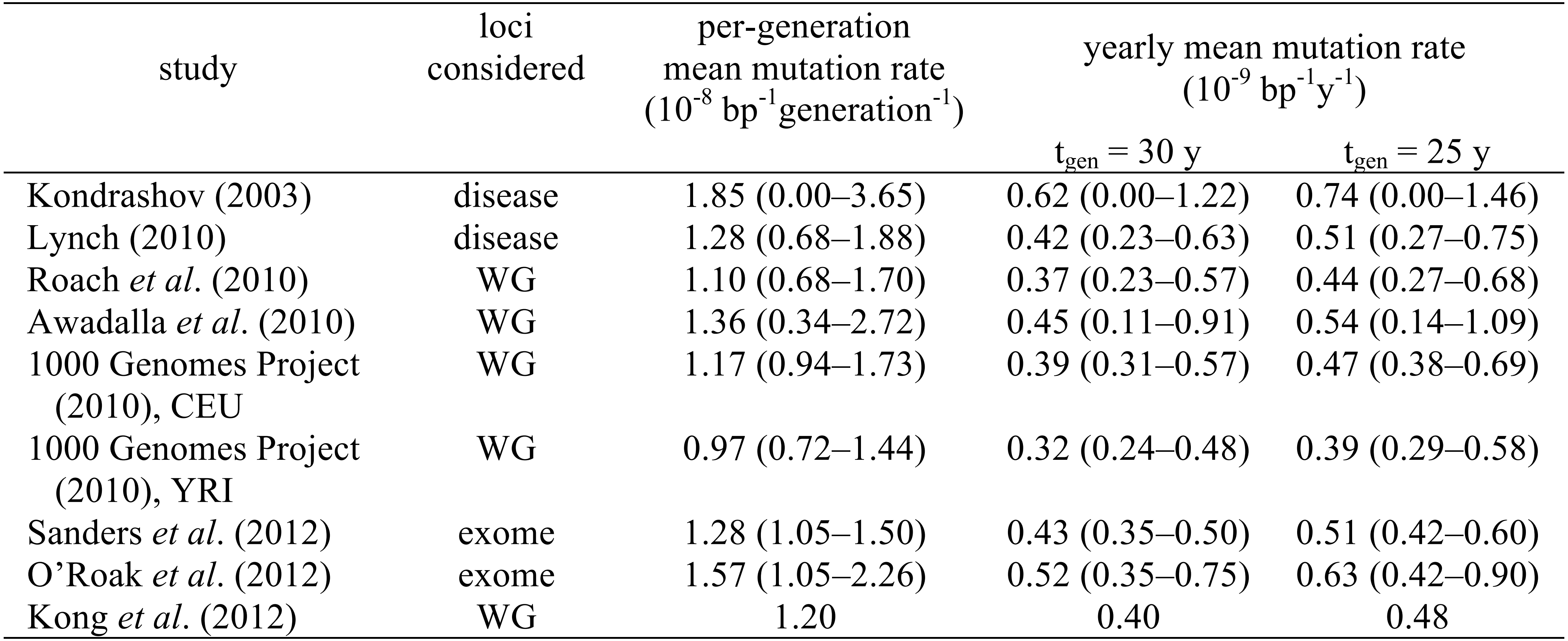


Heterozygosity is correlated with distance from East Africa



Ramachandran et al. 2005 PNAS

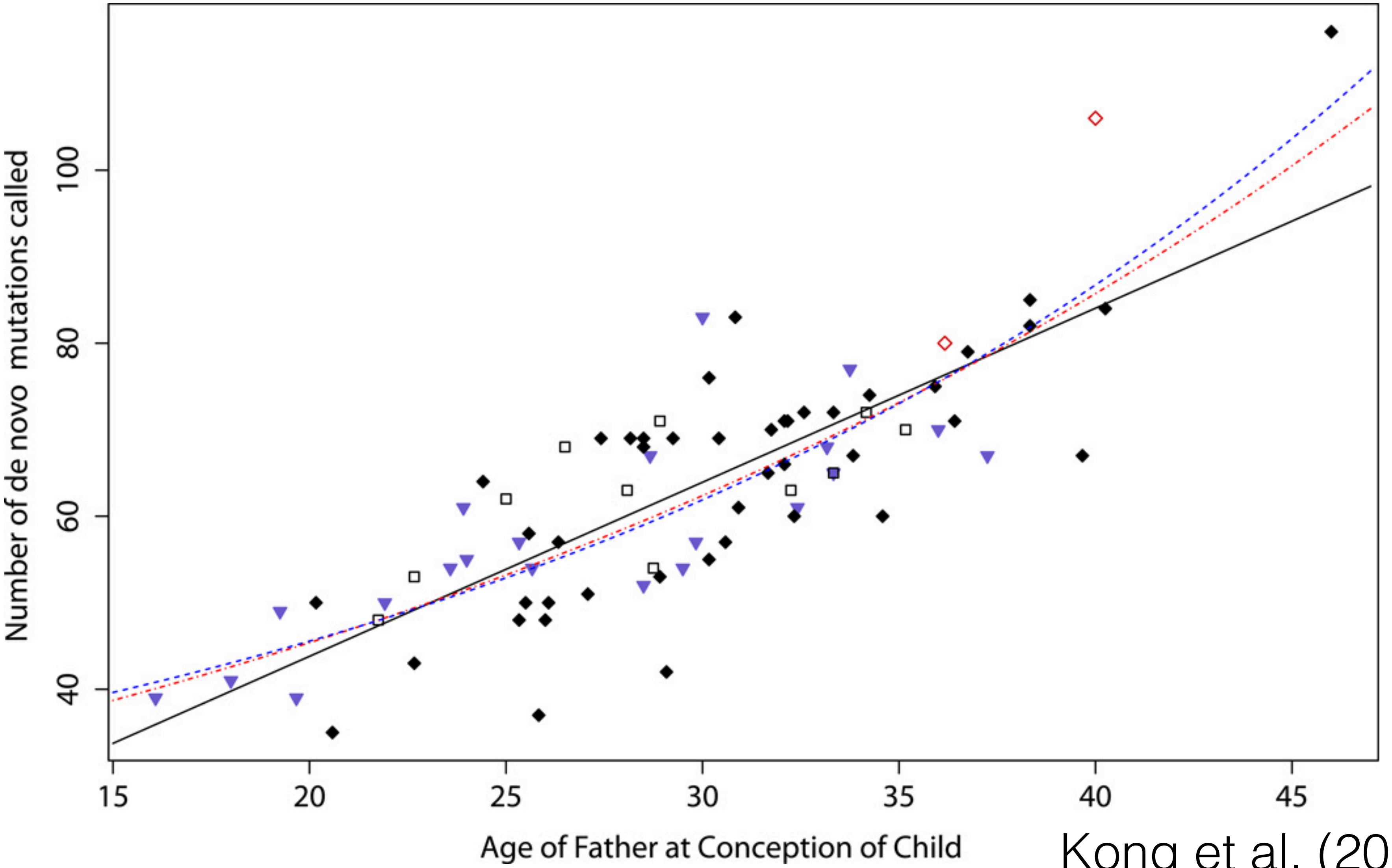
Mutation: How often do mutations arise?



study	loci considered	per-generation mean mutation rate (10^{-8} $bp^{-1}generation^{-1}$)	yearly mean mutation rate (10^{-9} $bp^{-1}y^{-1}$)	
			t_{gen} = 30 y	t_{gen} = 25 y
Kondrashov (2003)	disease	1.85 (0.00–3.65)	0.62 (0.00–1.22)	0.74 (0.00–1.46)
Lynch (2010)	disease	1.28 (0.68–1.88)	0.42 (0.23–0.63)	0.51 (0.27–0.75)
Roach *et al*. (2010)	WG	1.10 (0.68–1.70)	0.37 (0.23–0.57)	0.44 (0.27–0.68)
Awadalla *et al*. (2010)	WG	1.36 (0.34–2.72)	0.45 (0.11–0.91)	0.54 (0.14–1.09)
1000 Genomes Project (2010), CEU	WG	1.17 (0.94–1.73)	0.39 (0.31–0.57)	0.47 (0.38–0.69)
1000 Genomes Project (2010), YRI	WG	0.97 (0.72–1.44)	0.32 (0.24–0.48)	0.39 (0.29–0.58)
Sanders *et al*. (2012)	exome	1.28 (1.05–1.50)	0.43 (0.35–0.50)	0.51 (0.42–0.60)
O'Roak *et al*. (2012)	exome	1.57 (1.05–2.26)	0.52 (0.35–0.75)	0.63 (0.42–0.90)
Kong *et al*. (2012)	WG	1.20	0.40	0.48

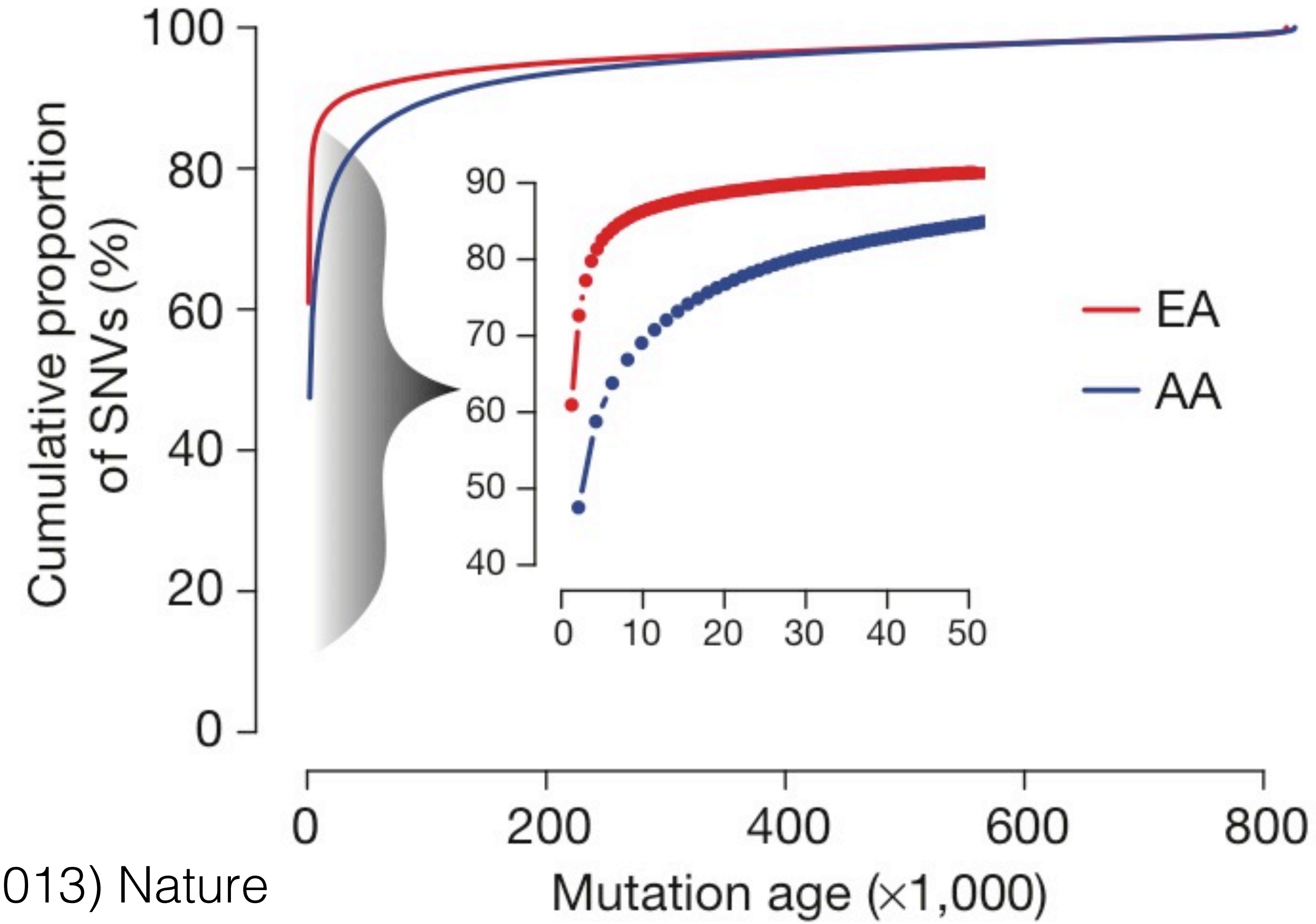
Scally and Durbin (2012) Nature Rev. Genet.

What are the effects of paternal age on mutation rate?



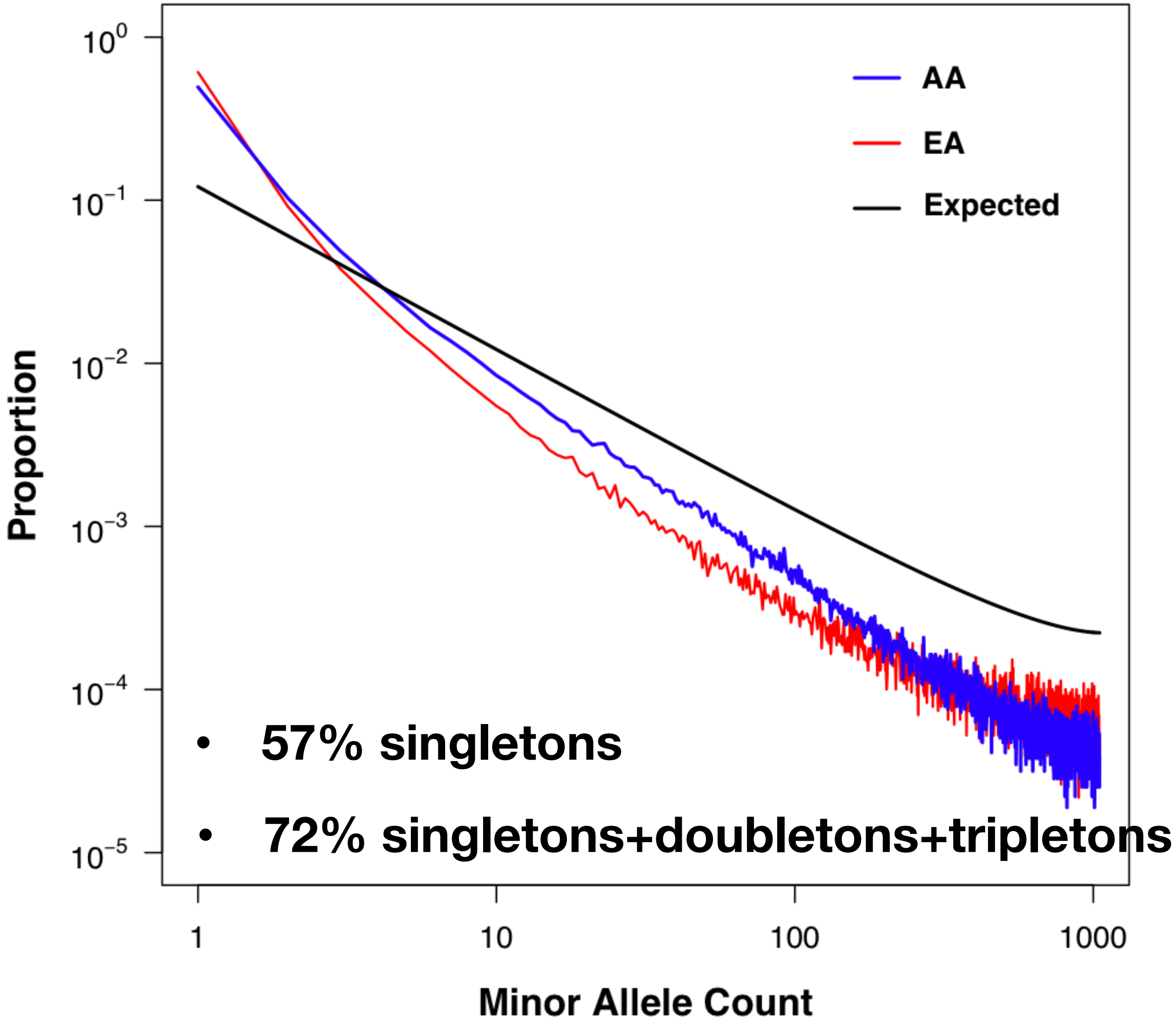
Kong et al. (2012) Nature

When did most variation arise?



Fu et al. (2013) Nature

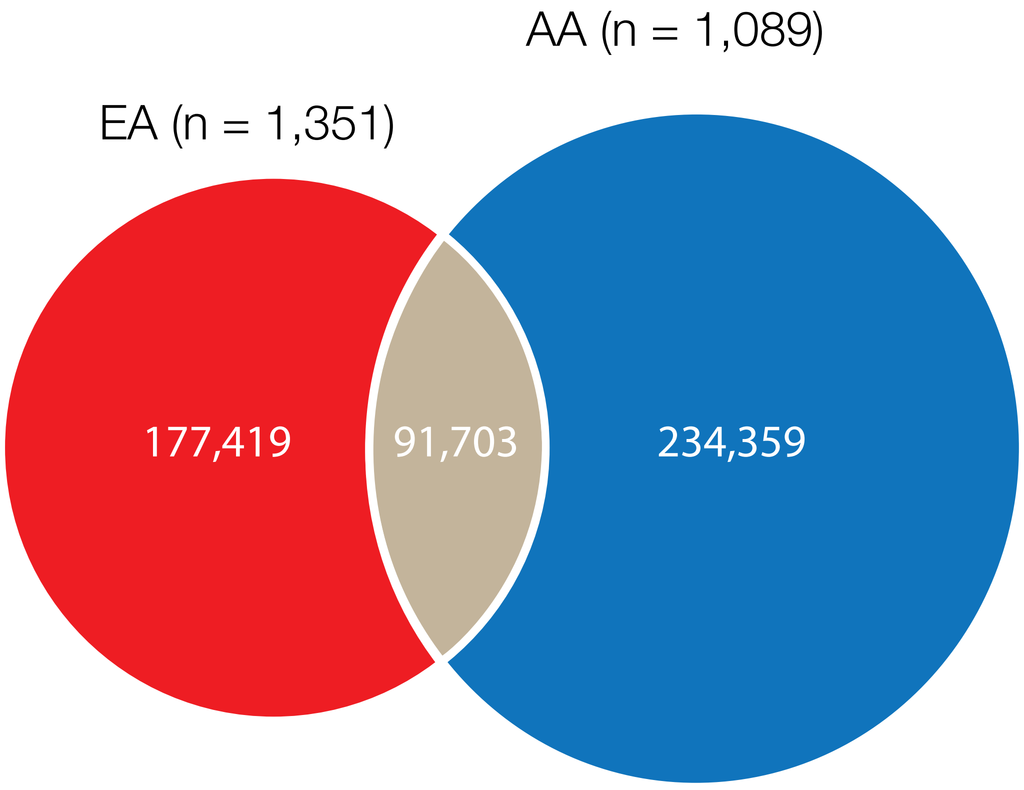
Most SNVs are very rare



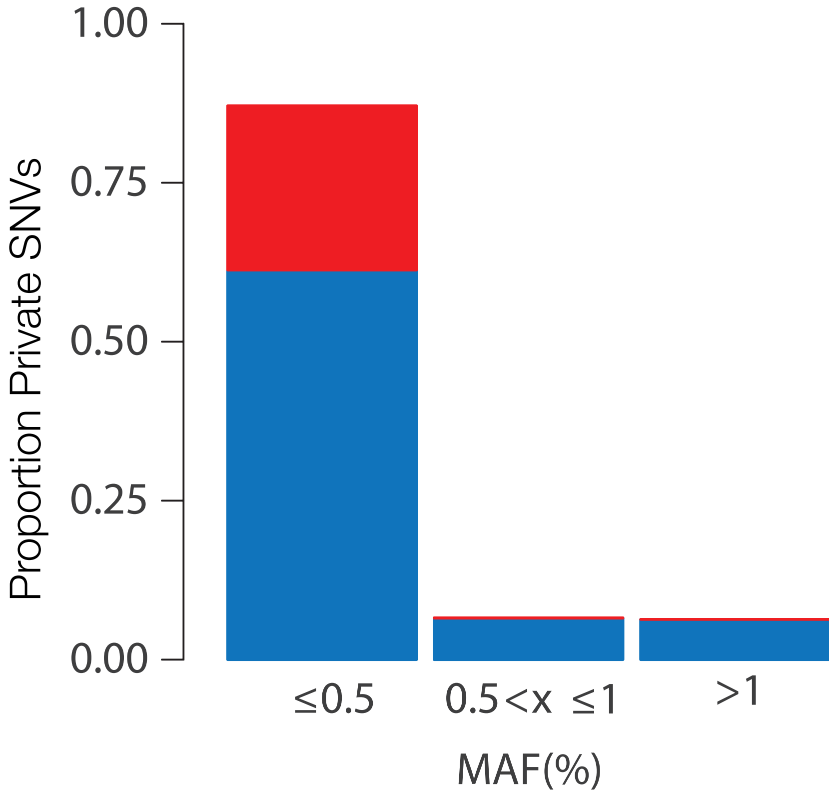
- **57% singletons**
- **72% singletons+doubletons+tripletons**

Tennessen et al. (2012) Science

Most SNVs are population specific

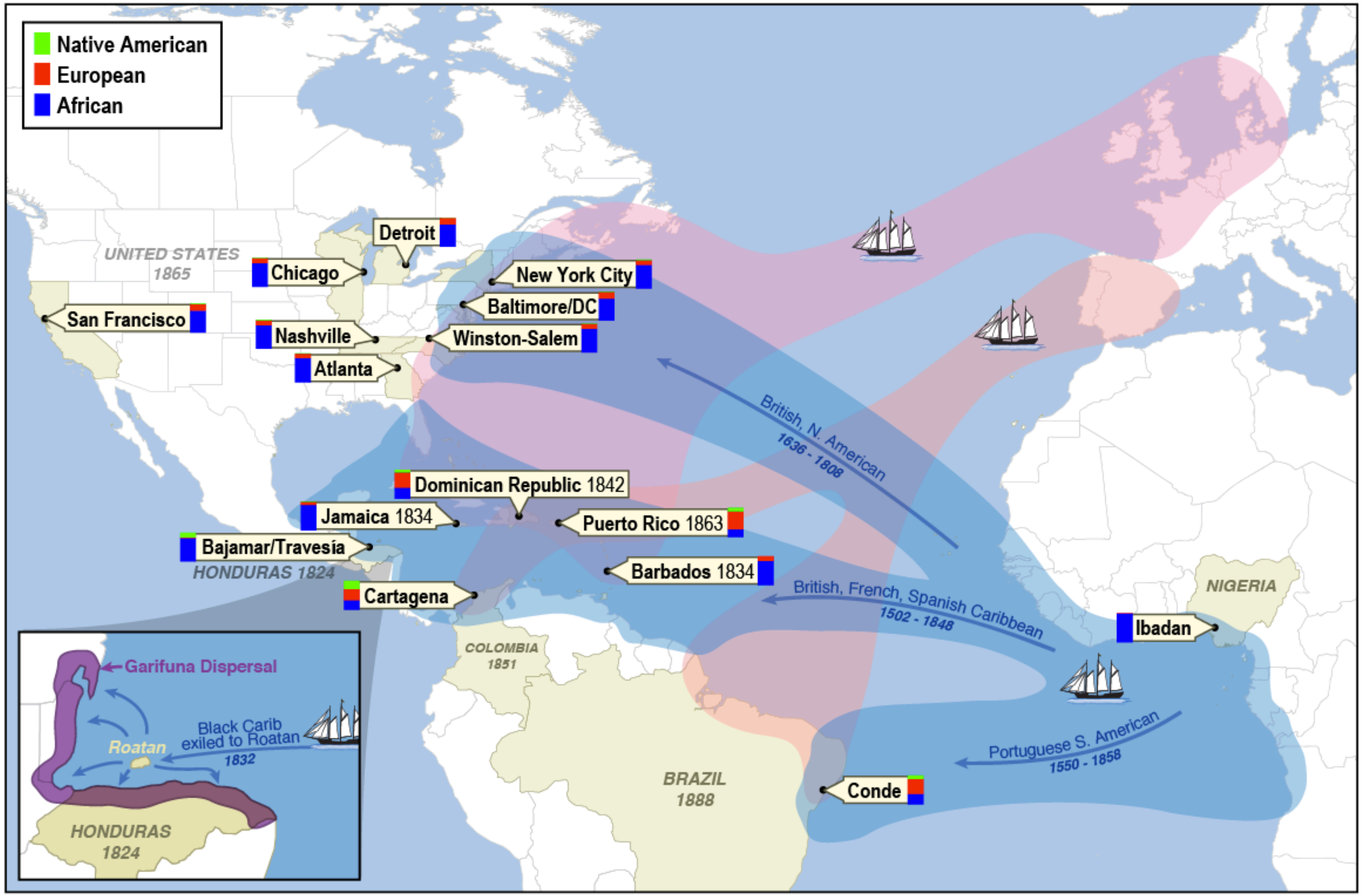


503,481 SNVs



Tennessen et al. (2012) Science

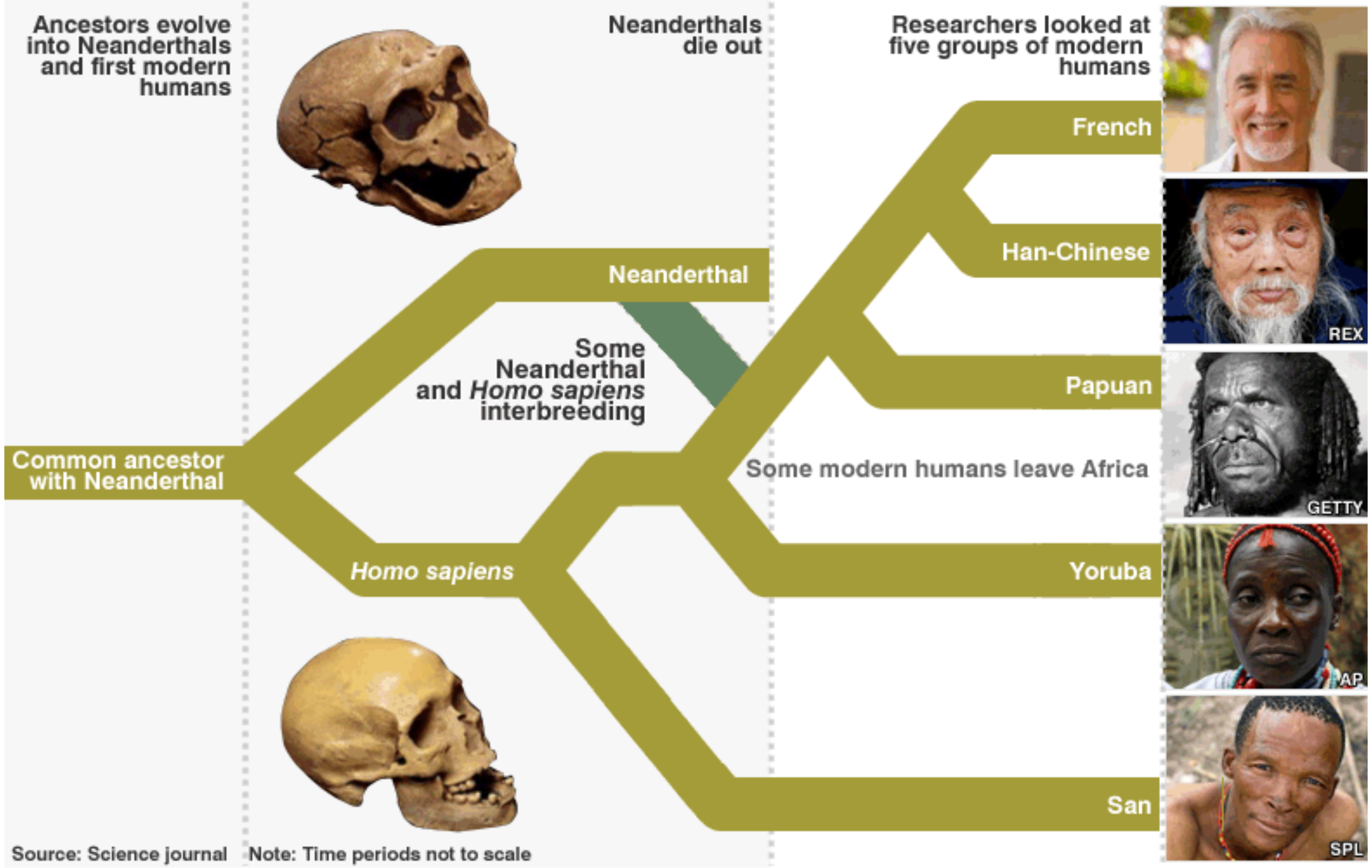
Recent admixture: Migrations can have a profound effect on genetics



Mathias et al. (2016)
Nature Comm.

Ancient admixture: Neanderthals are still among us

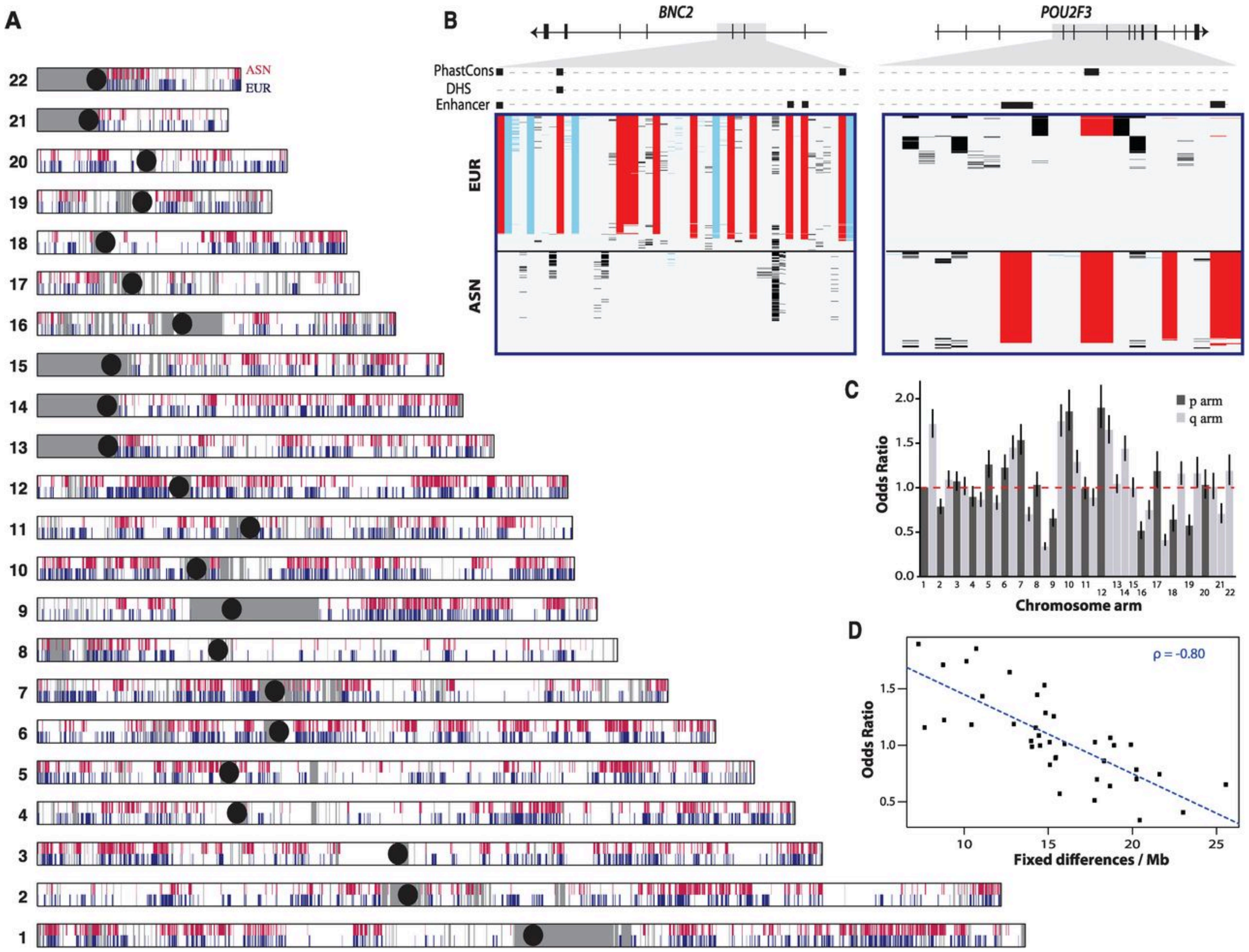
- Recent genetic data suggests that 1-4% of non-African genomes are derived from Neanderthals



Source: Science journal

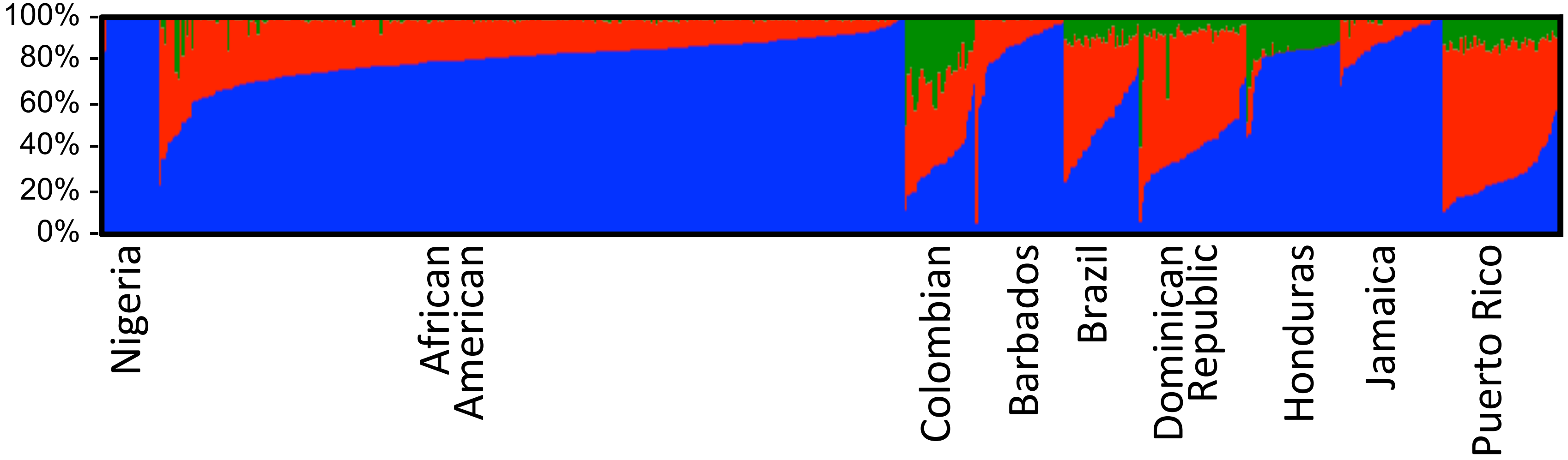
Note: Time periods not to scale

Neanderthals are still among us

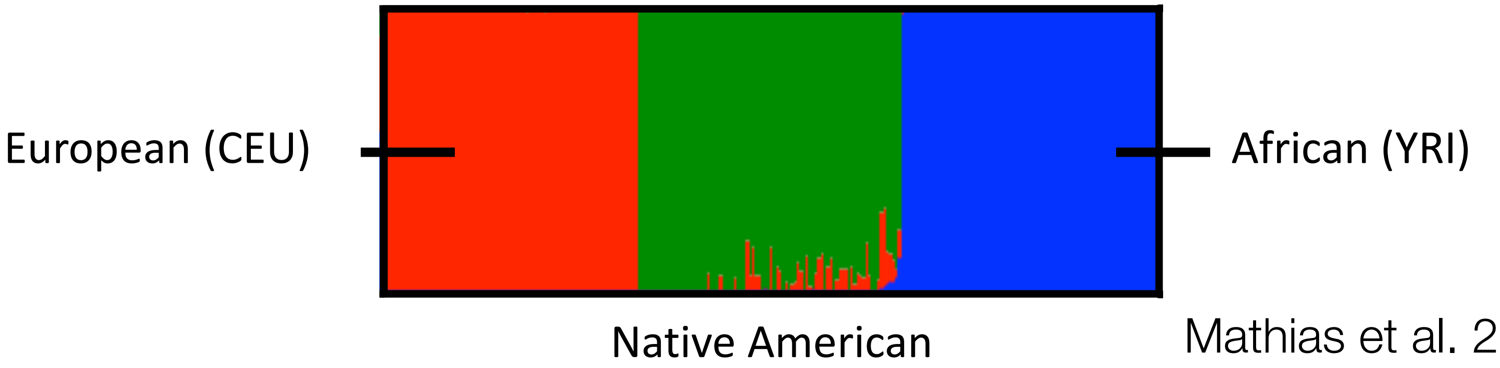


Estimates of global ancestry

CAAPA

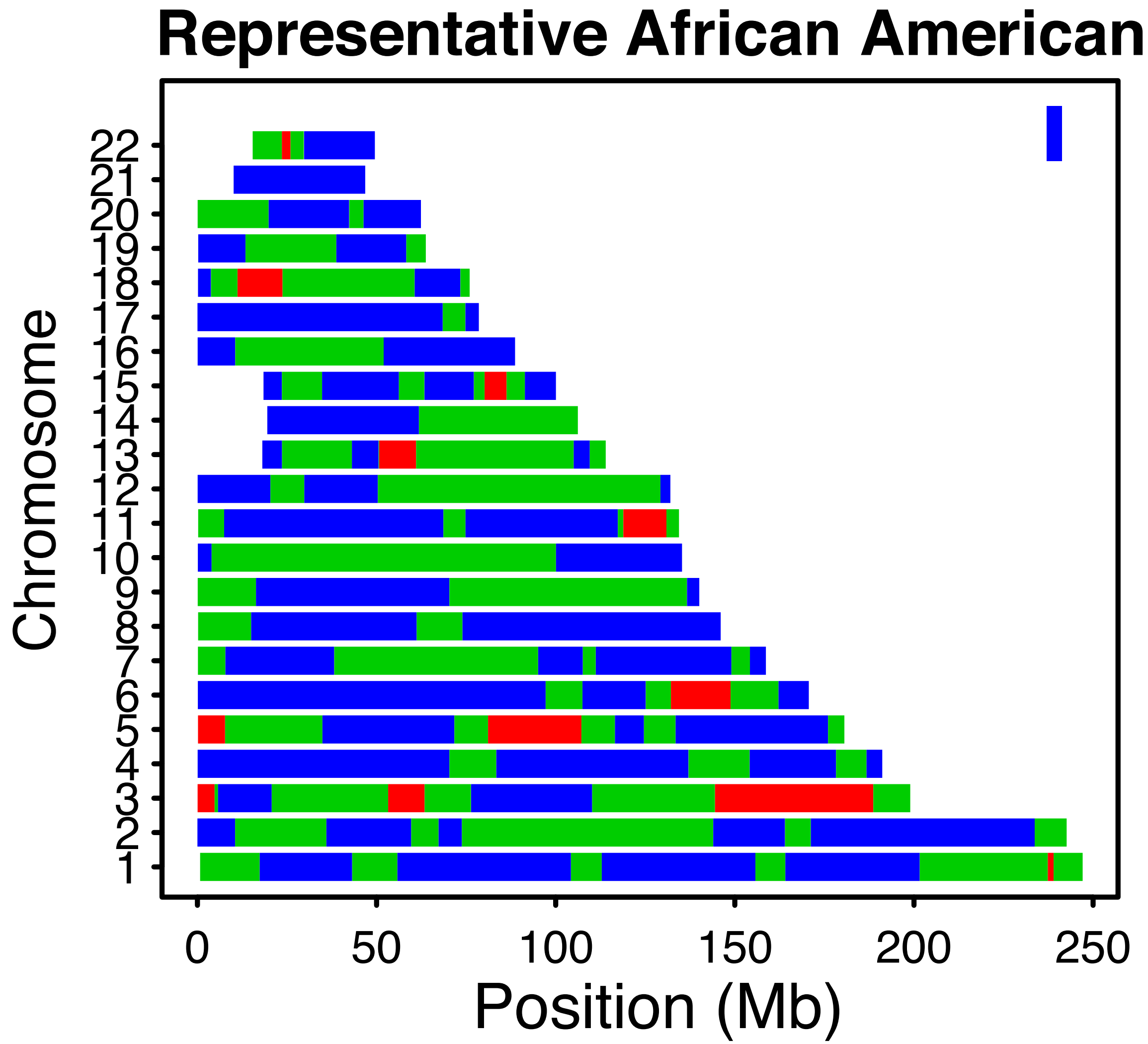


Reference



Mathias et al. 2016 Nature Comm.

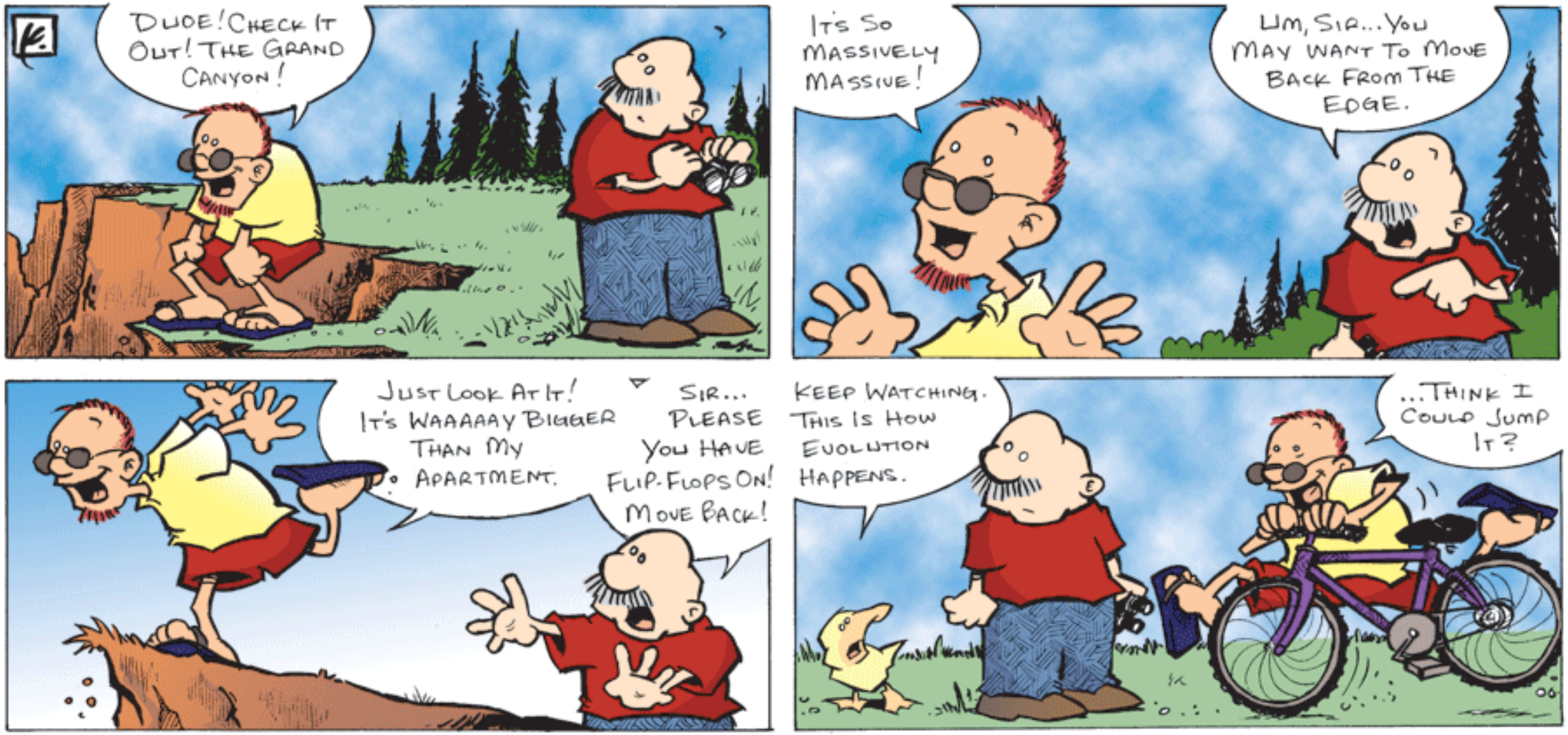
Local ancestry of a single individual



Bryc et al. 2009 PNAS

Adaptive (Darwinian) Selection

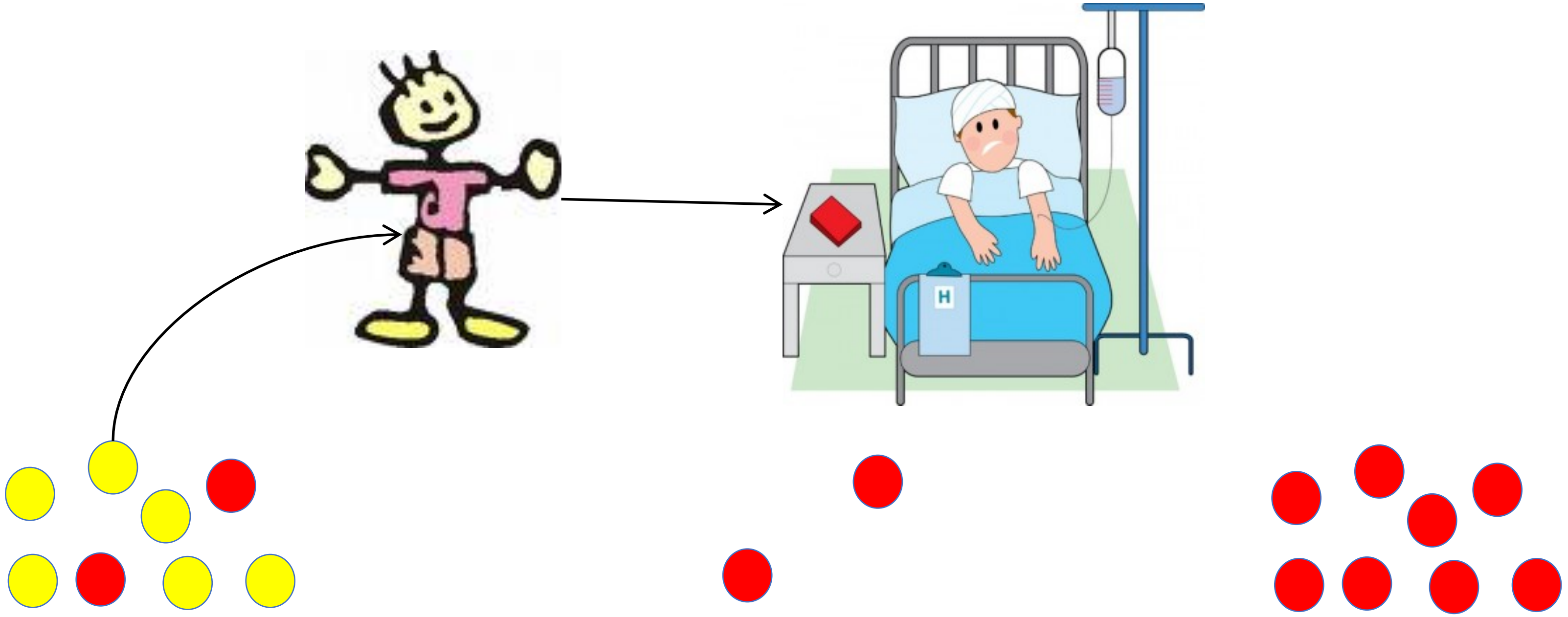
“I have called this principle, by which each slight variation, if useful, is preserved, by the term Natural Selection.”—Charles Darwin from "The Origin of Species”, 1859



SHELDONCOMICS.COM

© DAVE KELLETT

Antibiotic resistance is an example of adaptive evolution



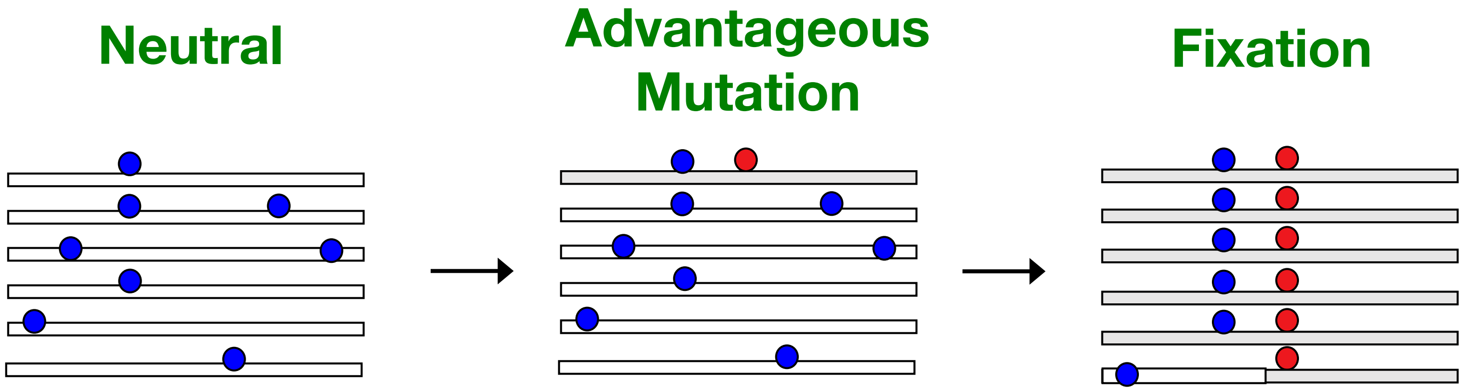
Initial Population

After Antibiotic

Final Population

G	T
Not Resistant	Resistant

Reading the genome for signatures of positive selection



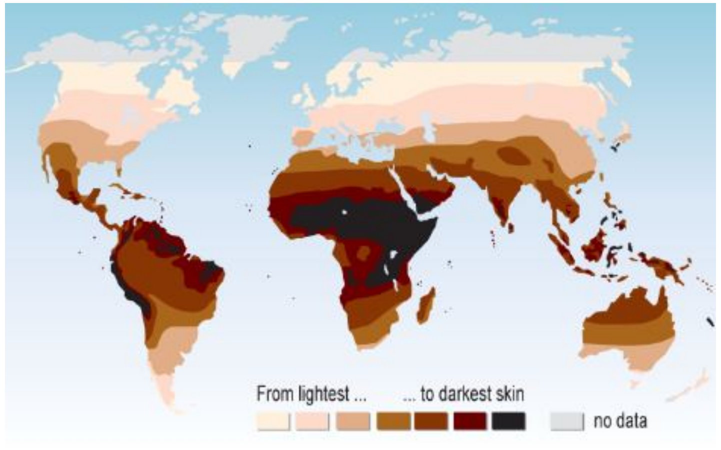
- This process imparts “signatures” on patterns of genetic variation that we can use to find adaptively evolving genes

Genes that influence physical traits have been targets of recent selection

Eye Color

HERC2

Skin Pigmentation



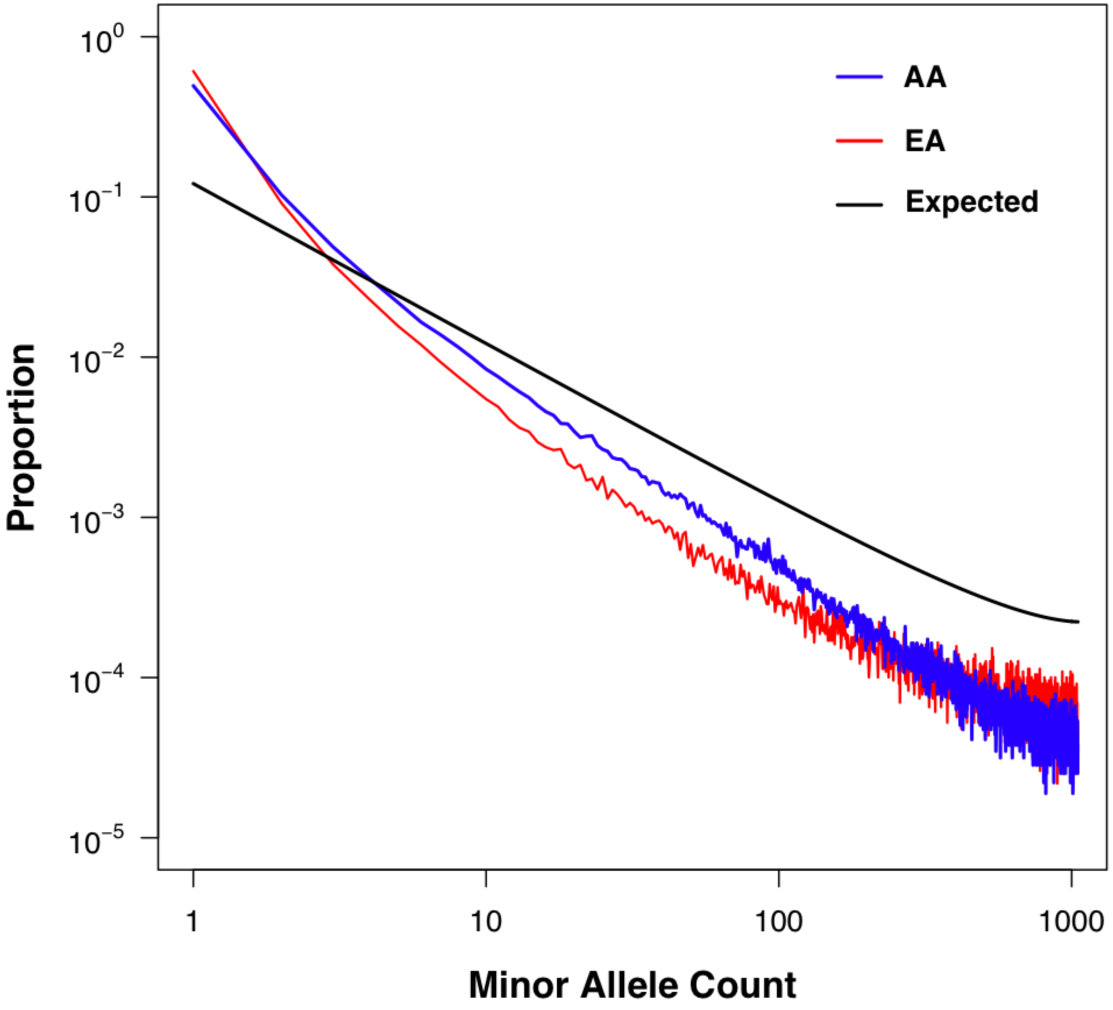
Source: Chaplin G.©, *Geographic Distribution of Environmental Factors Influencing Human Skin Coloration*, American Journal of Physical Anthropology 125:292–302, 2004; map updated in 2007.

SLC24A5, OCA2, TYRP1

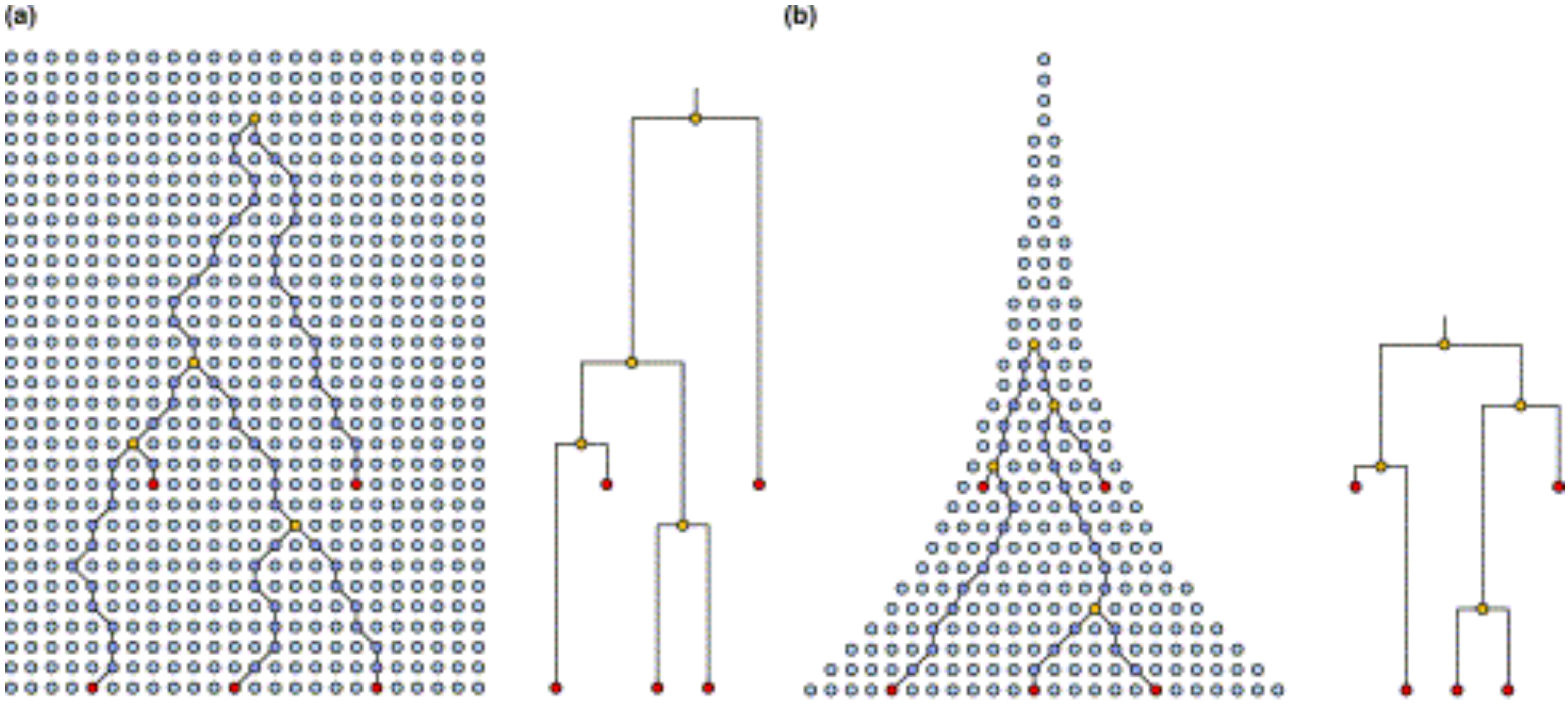
Hair Texture

EDAR

These forces all affect the Site Frequency Spectrum (SFS)



Primer on coalescent



TRENDS in Ecology & Evolution

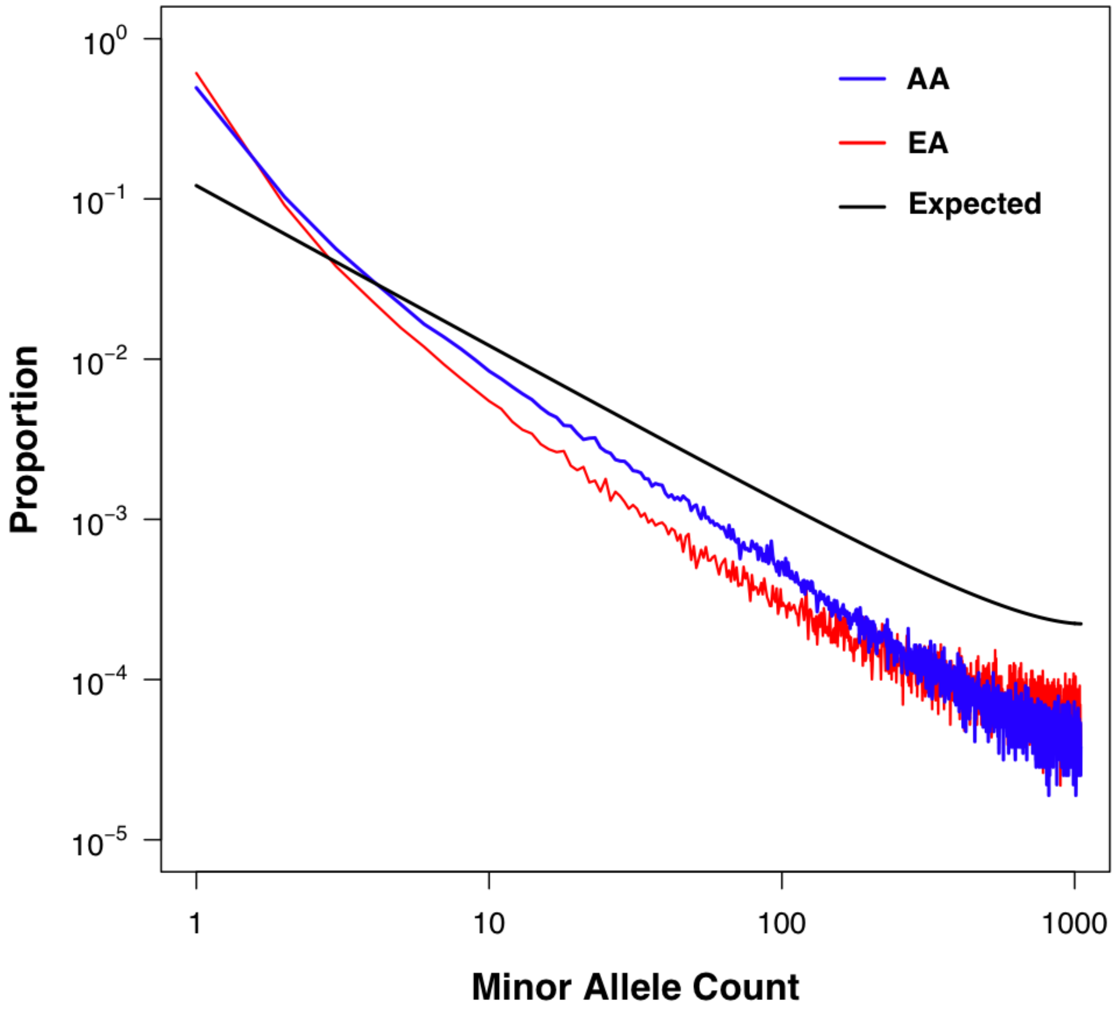
Primer on coalescent

$$E(T_i) = \frac{2}{i(i-1)} \qquad Var(T_i) = \left(\frac{2}{i(i-1)}\right)^2$$

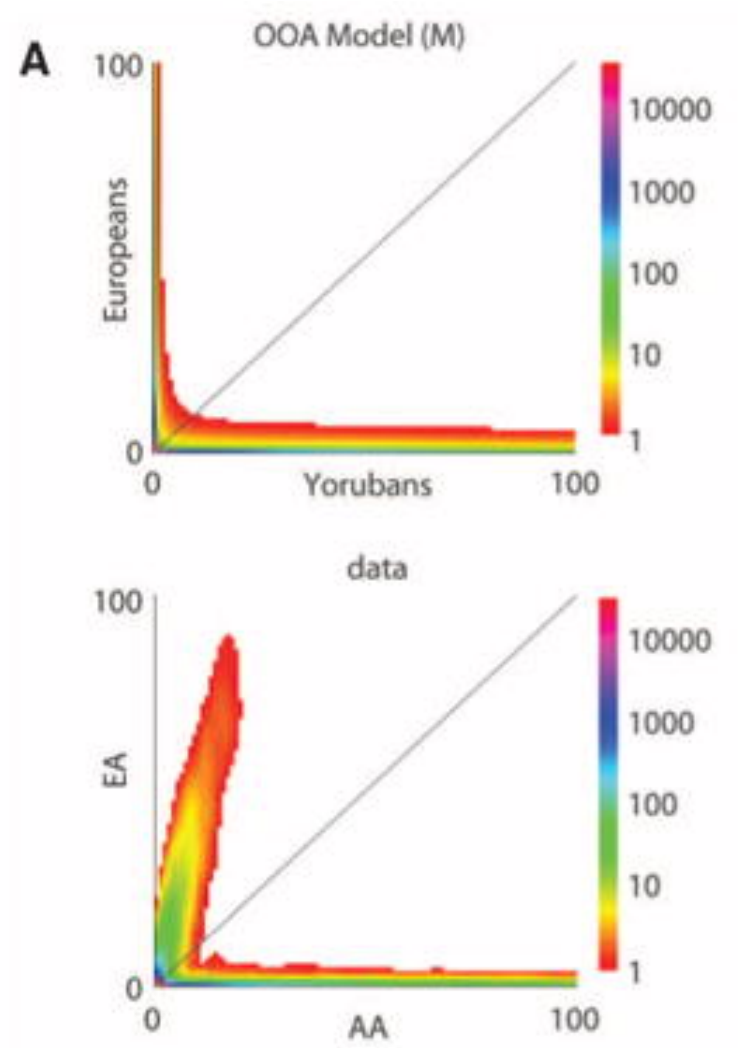
To generate a genealogy of i genes under Kingman's coalescent:

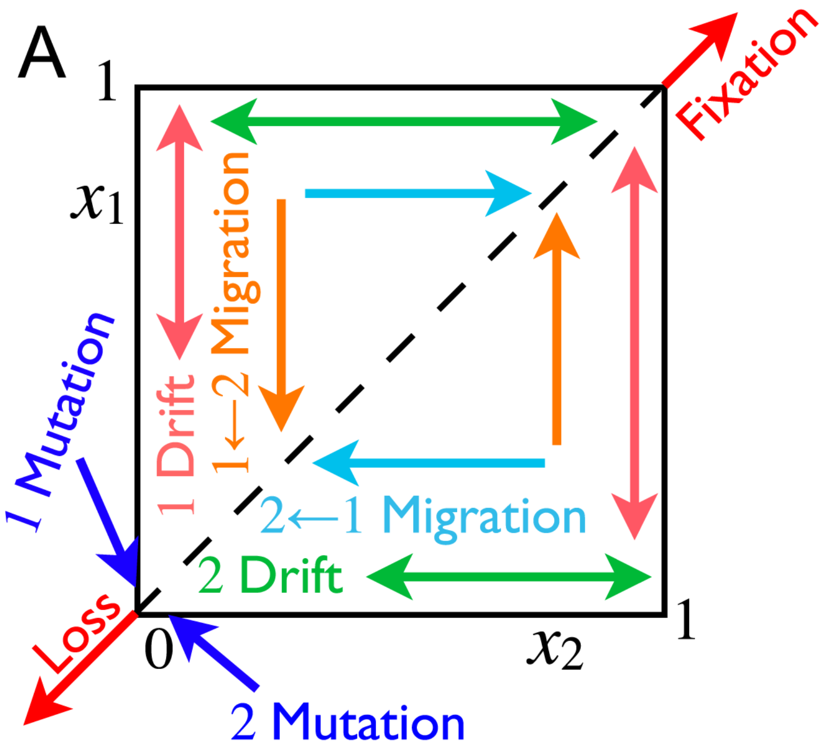
- Draw an observation from an exponential distribution with mean $\mu = 2/(i(i-1))$. This will be the time of the first coalescent event (looking from the present backwards in time).
- Pick two lineages at random to coalescence.
- Decrease i by 1.
- If $i = 1$, stop. Otherwise, repeat these steps [8, 9].

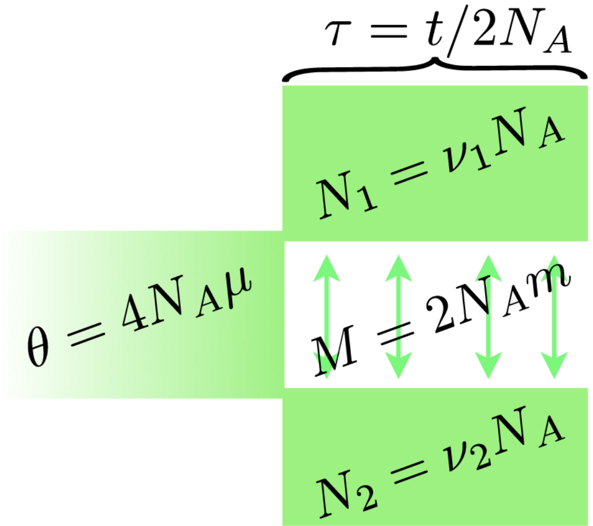
Site Frequency Spectrum (SFS)

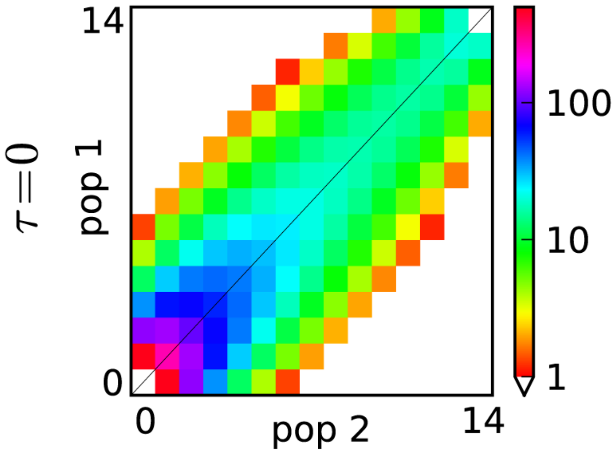


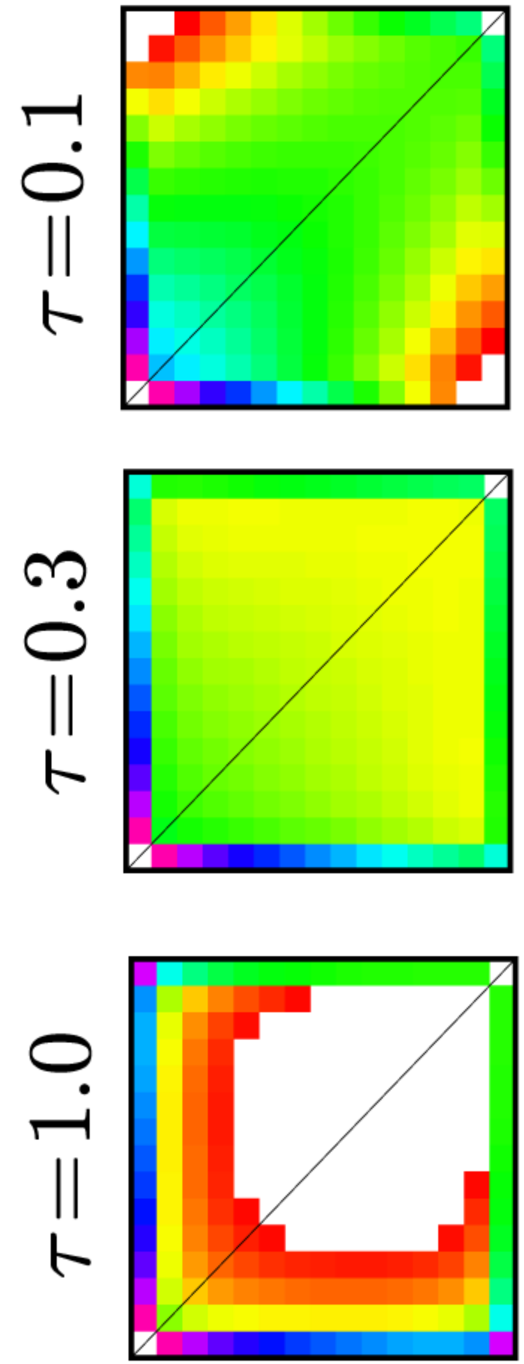
Joint Site Frequency Spectrum (JSFS)











Gutenkunst et al. 2009 PLoS Genet.

Useful equations

Time: $t = T/(4*N_{ref}*Gen)$

- N_{ref} = reference or ancestral population size
- Gen = number of years per generation
- T = chronological years

$\theta = 4*N_{ref}*\mu*Length$;

- μ = mutation rate
- Length is the bp of the segment simulated (aka nsites for recombination)

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

Recombination: $\rho = 4N_{ref}r$

- r is the recombination rate between the ends of a unit length sequence

Migration: $M_{ij} = 4N_{ref}m_{ij}$

- m_{ij} is 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.