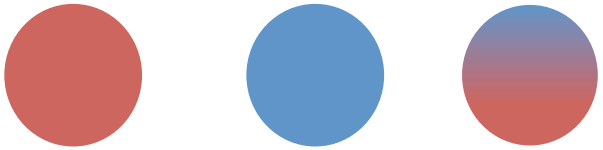
1000 Genomes and Population Genomics

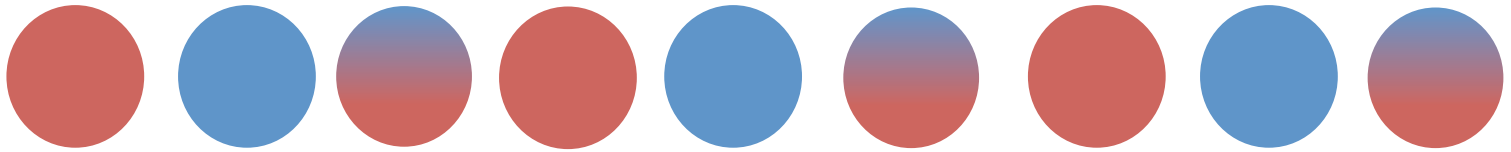
As small groups separate, they evolve in isolation

How do the frequencies of alleles change over time?

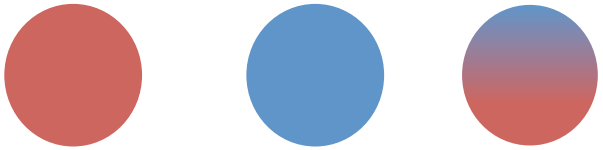
- Selection
- Genetic Drift
- Hardy-Weinberg



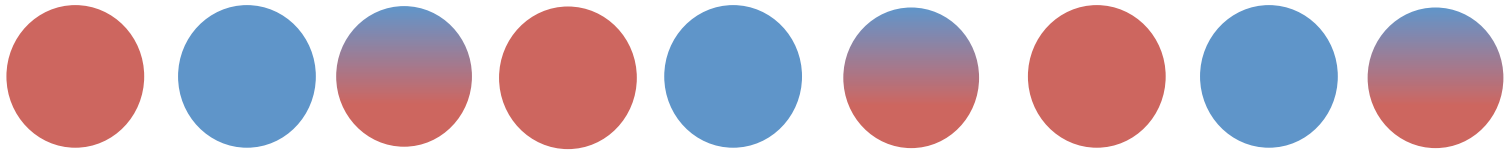
Two alleles, Red (A) and Blue (a)
Creates 3 genotypes AA, aa, and Aa



Gen 0



Two alleles, Red (A) and Blue (a)
Creates 3 genotypes AA, aa, and Aa

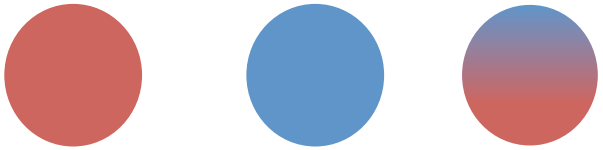


Gen 0

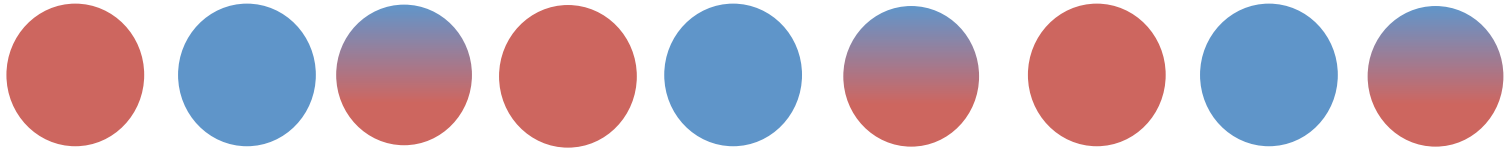
Frequency of allele A; p
Frequency of allele a: q
p+q=1

Frequency of AA: $D=p^2$
Frequency of Aa: $H=2pq$
Frequency of aa: $R=q^2$

Assume no selection and no mutation
Assume no migration and random mating
Assume infinite population sizes (no restrictions on numbers)



Two alleles, Red (A) and Blue (a)
Creates 3 genotypes AA, aa, and Aa



Gen 0

Mating	Probability	Offspring Genotype frequencies		
		AA	Aa	aa
AA x AA	D^2	1	0	0
AA x Aa	2DH	½	½	0
AA x aa	2DR	0	1	0
Aa x Aa	H^2	¼	½	¼
Aa x aa	2HR	0	½	½
aa x aa	R^2	0	0	1
Total for Gen 1		D’	H’	R’

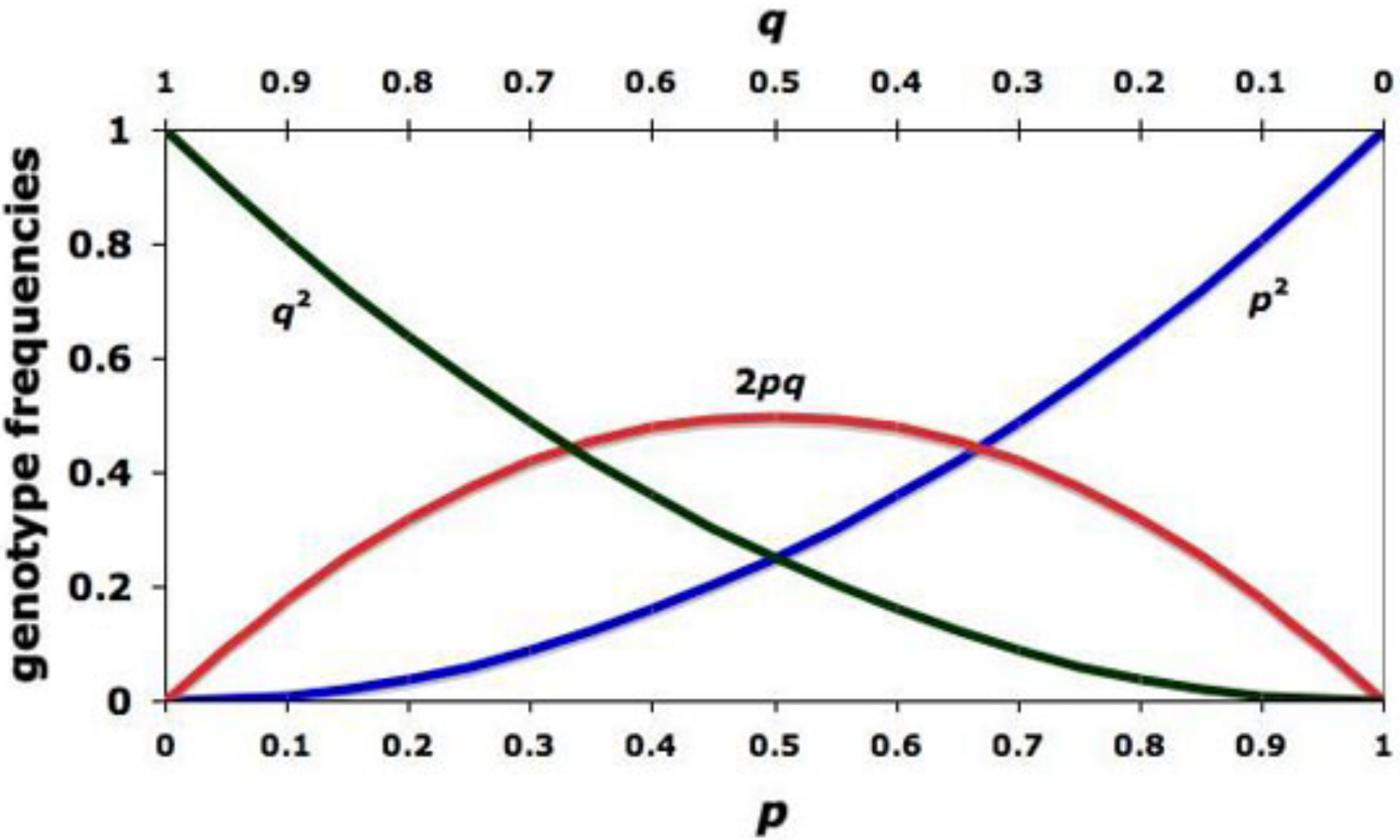
Mating	Probability	Offspring Genotype frequencies		
		AA	Aa	aa
AA x AA	D^2	1	0	0
AA x Aa	2DH	½	½	0
AA x aa	2DR	0	1	0
Aa x Aa	H^2	¼	½	¼
Aa x aa	2HR	0	½	½
aa x aa	R^2	0	0	1
Total for Gen 1		D’	H’	R’

$$D' = D^2 + \frac{2DH}{2} + \frac{H^2}{4} = \left(D + \frac{H}{2}\right)^2 = \left(p^2 + p(1-p)\right)^2 = \left(p^2 + p - p^2\right) = p^2$$

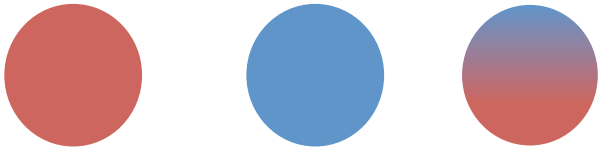
$$R' = \frac{H^2}{4} + \frac{2HR}{2} + R^2 = \left(R + \frac{H}{2}\right)^2 = \left(q^2 + q(1-q)\right)^2 = (q^2 + q - q^2) = q^2$$

Genotypes are stable over time

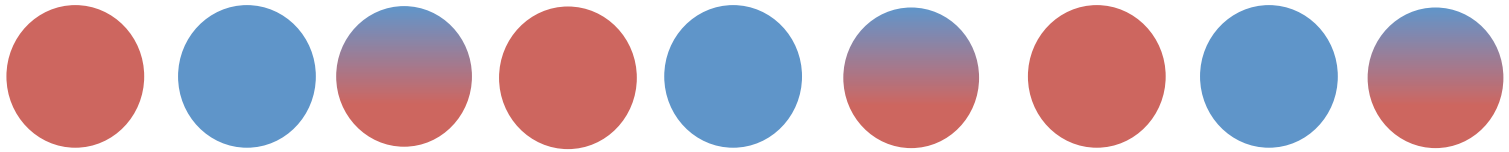
$$H' = \frac{2DH}{2} + 2DR + \frac{H^2}{2} + \frac{2HR}{2} = 2\left(D + \frac{H}{2}\right)\left(R + \frac{H}{2}\right) = 2pq$$



Andrews, C. (2010) The Hardy-Weinberg Principle. *Nature Education Knowledge* **3**(10):65



Two alleles, Red (A) and Blue (a)
Creates 3 genotypes AA, aa, and Aa

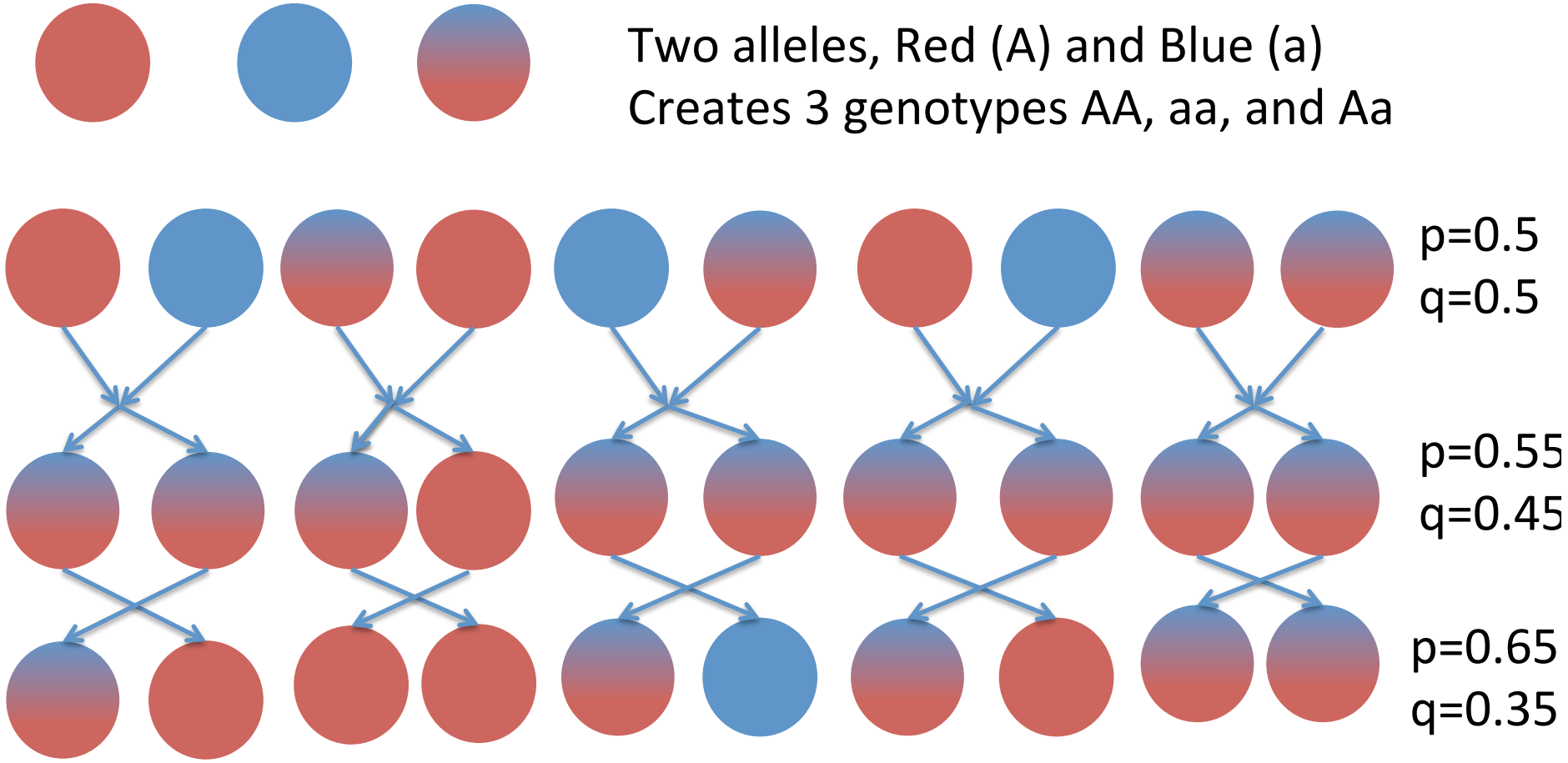


Gen 0

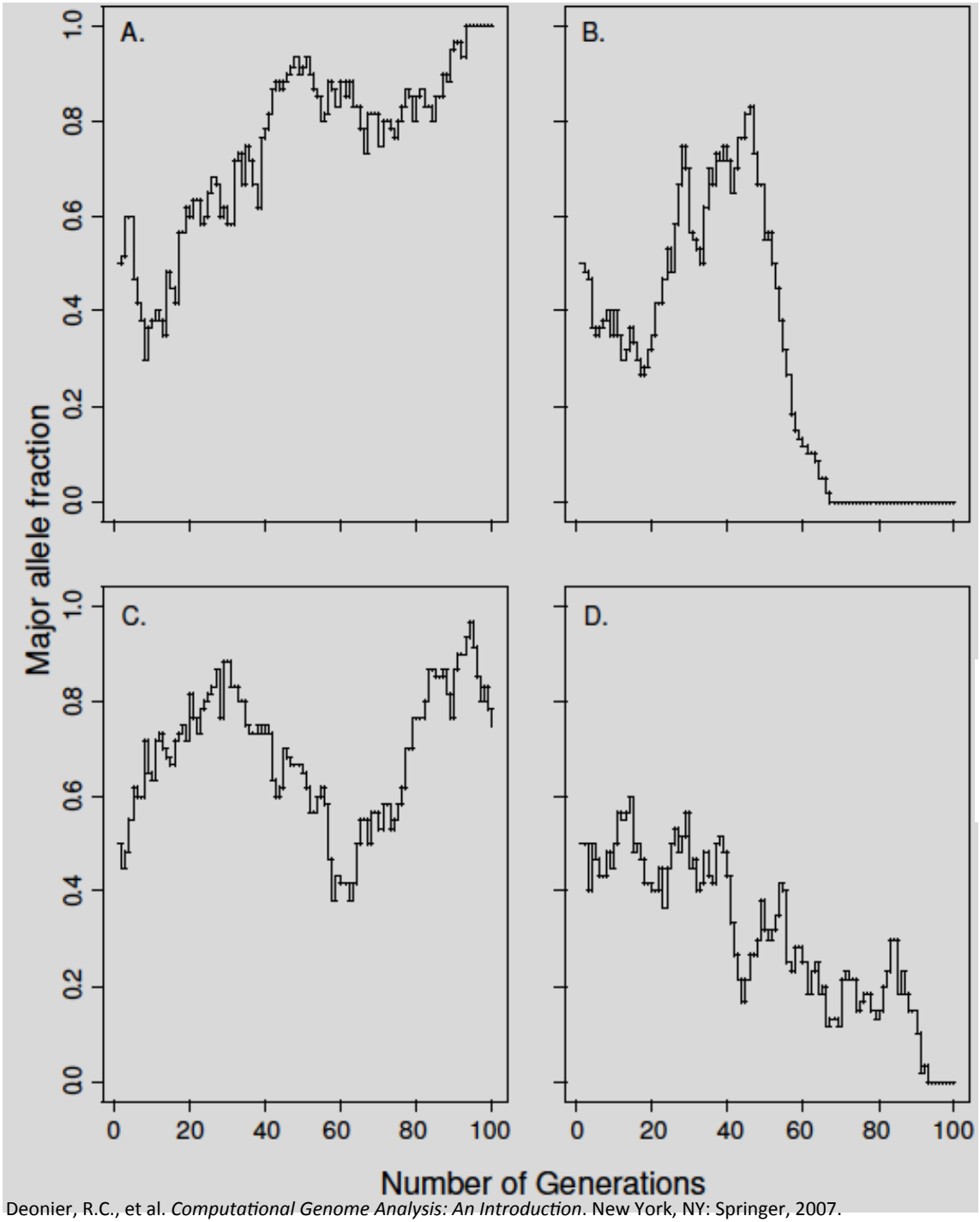
Frequency of allele A; p
Frequency of allele a: q
p+q=1

Frequency of AA: $D=p^2$
Frequency of Aa: $H=2pq$
Frequency of aa: $R=q^2$

Assume no selection and no mutation
Assume no migration and random mating
~~Assume infinite population sizes (no restrictions on numbers)~~



Genetic Drift



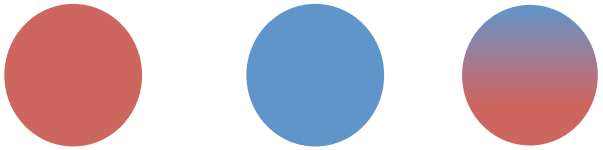
Deonier, R.C., et al. *Computational Genome Analysis: An Introduction*. New York, NY: Springer, 2007.

If not infinite number of offspring, then random fluctuations in number of alleles can cause dramatic changes in allele frequency

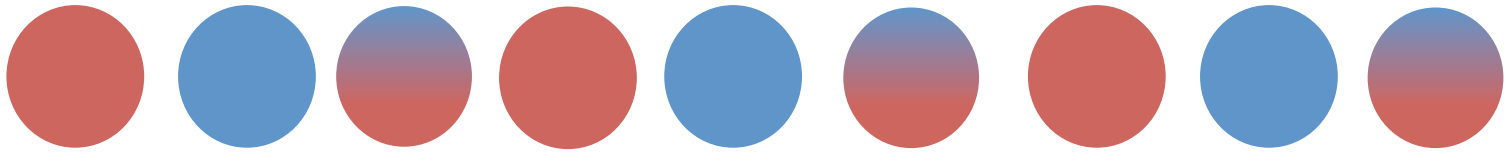
These drifts are more pronounced in a smaller population

$$h(n) = \left(1 - \frac{1}{2N}\right)^n h(0).$$

Heterozygosity is lost at a geometric rate with the number of generations (n) that depends on the number of individuals (N)



Two alleles, Red (A) and Blue (a)
Creates 3 genotypes AA, aa, and Aa

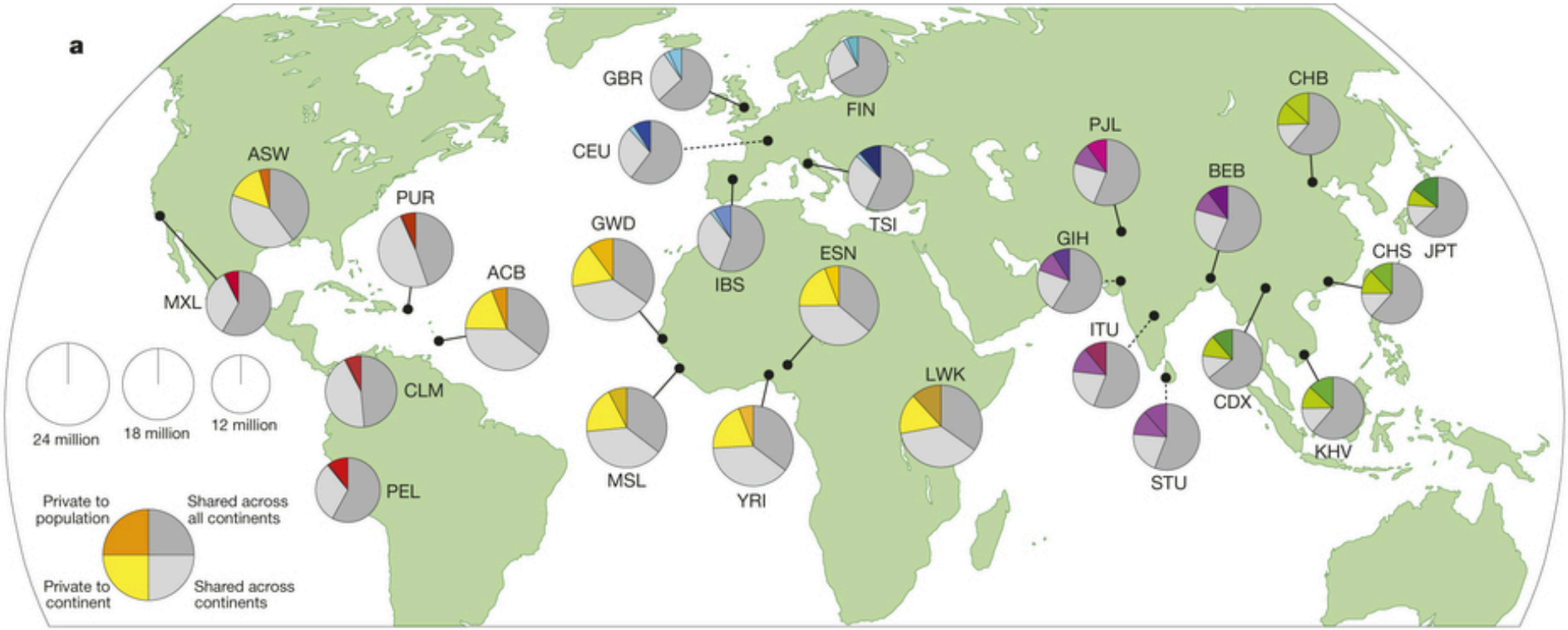


Gen 0

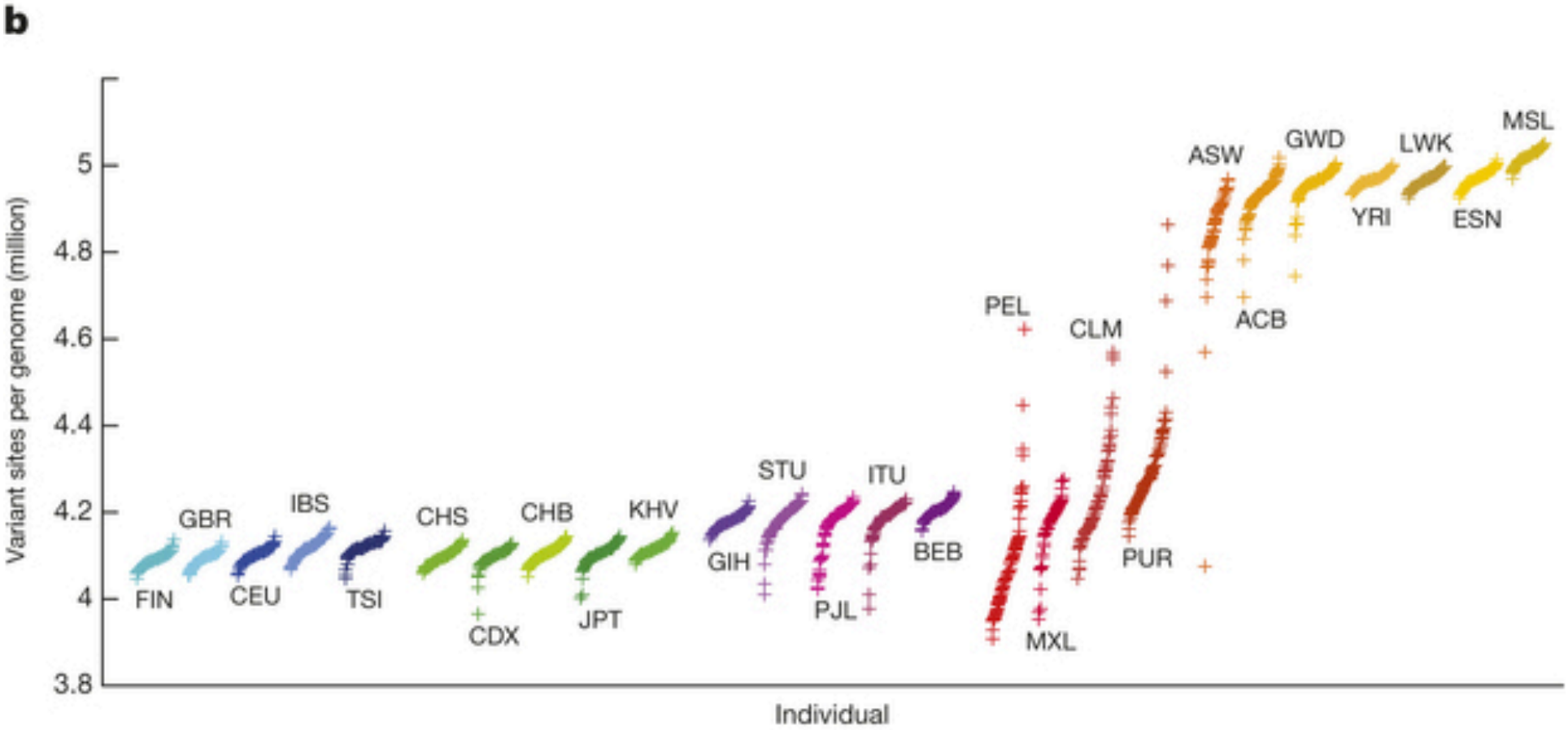
Frequency of allele A; p
Frequency of allele a: q
p+q=1

Frequency of AA: $D=p^2$
Frequency of Aa: $H=2pq$
Frequency of aa: $R=q^2$

Assume no selection and no mutation
Assume no migration and random mating
~~Assume infinite population sizes (no restrictions on numbers)~~

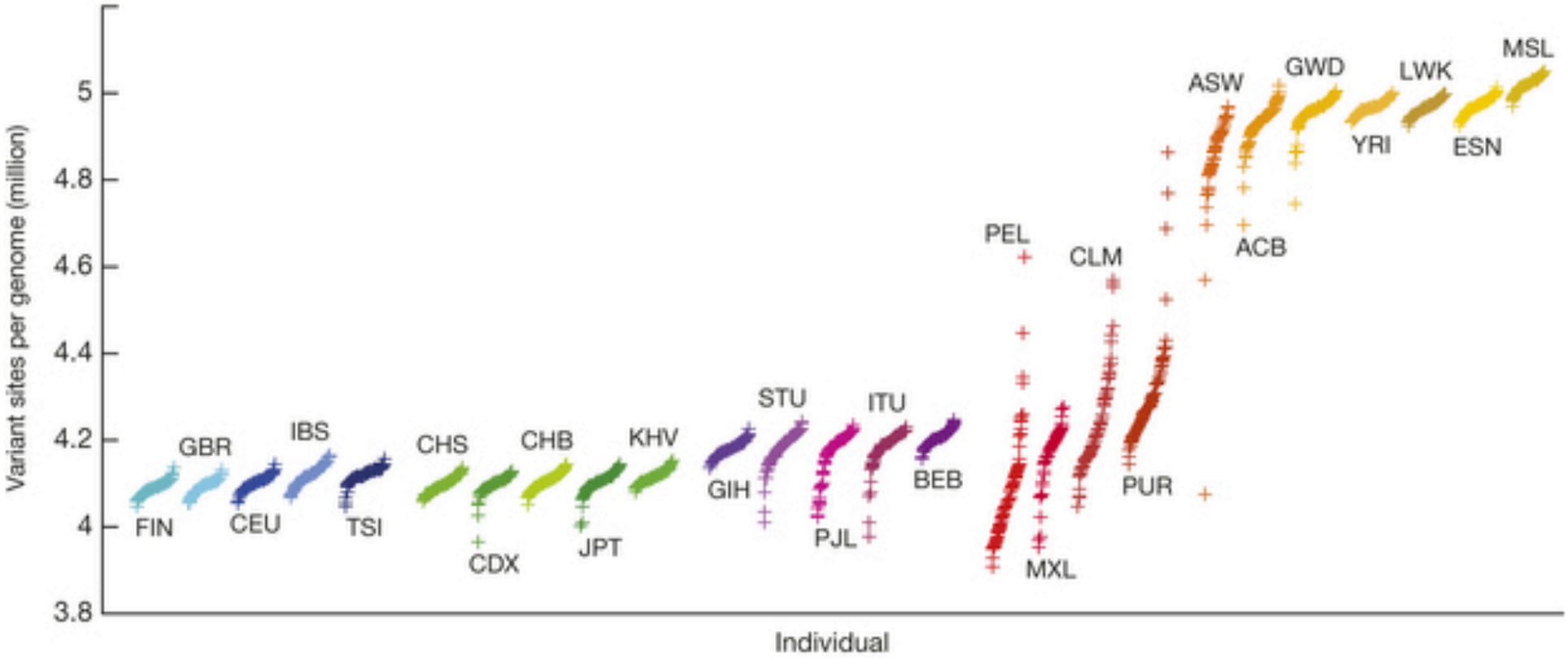


The 1000 Genomes Project Consortium; *Nature* **526**, 68–74 (01 October 2015) doi:10.1038/nature15393



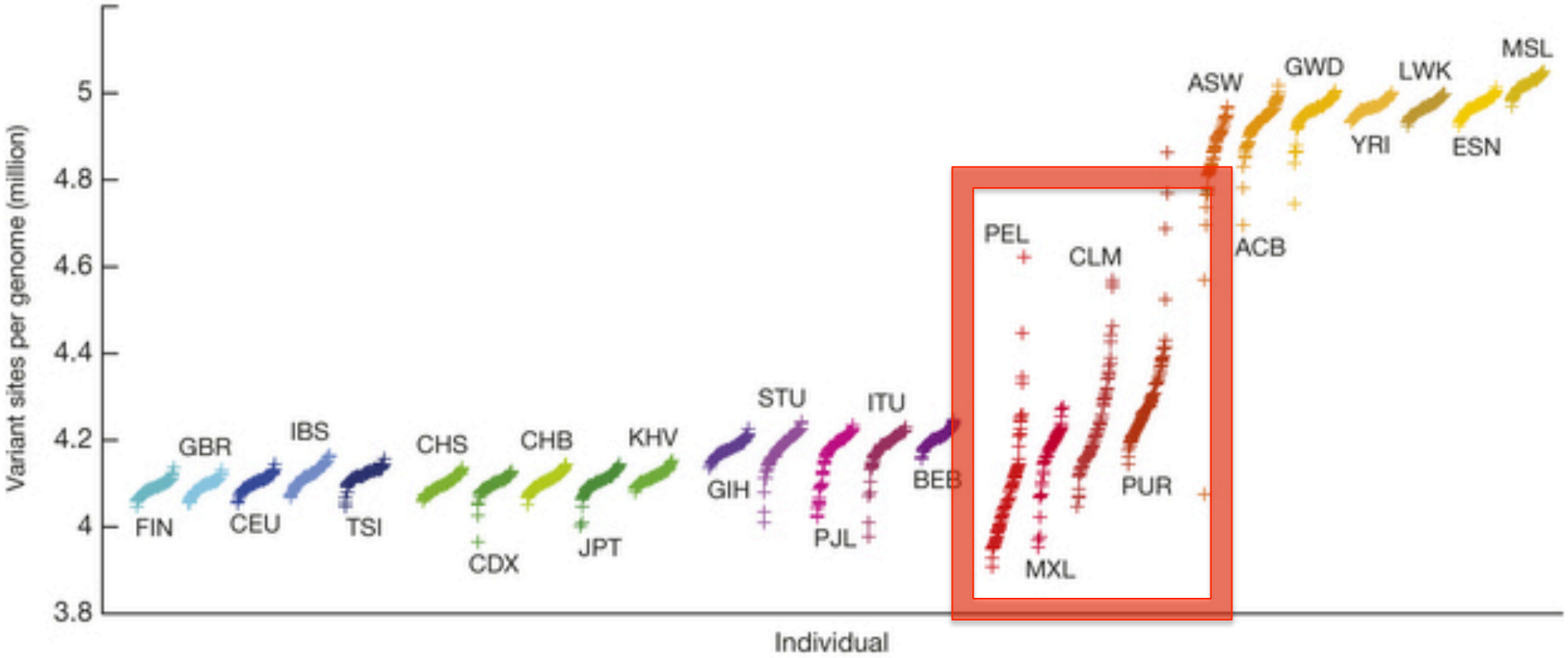
The 1000 Genomes Project Consortium; *Nature* **526**, 68–74 (01 October 2015) doi:10.1038/nature15393

Why does the higher level of variants per genome support the out of Africa hypothesis of human evolution and migration?



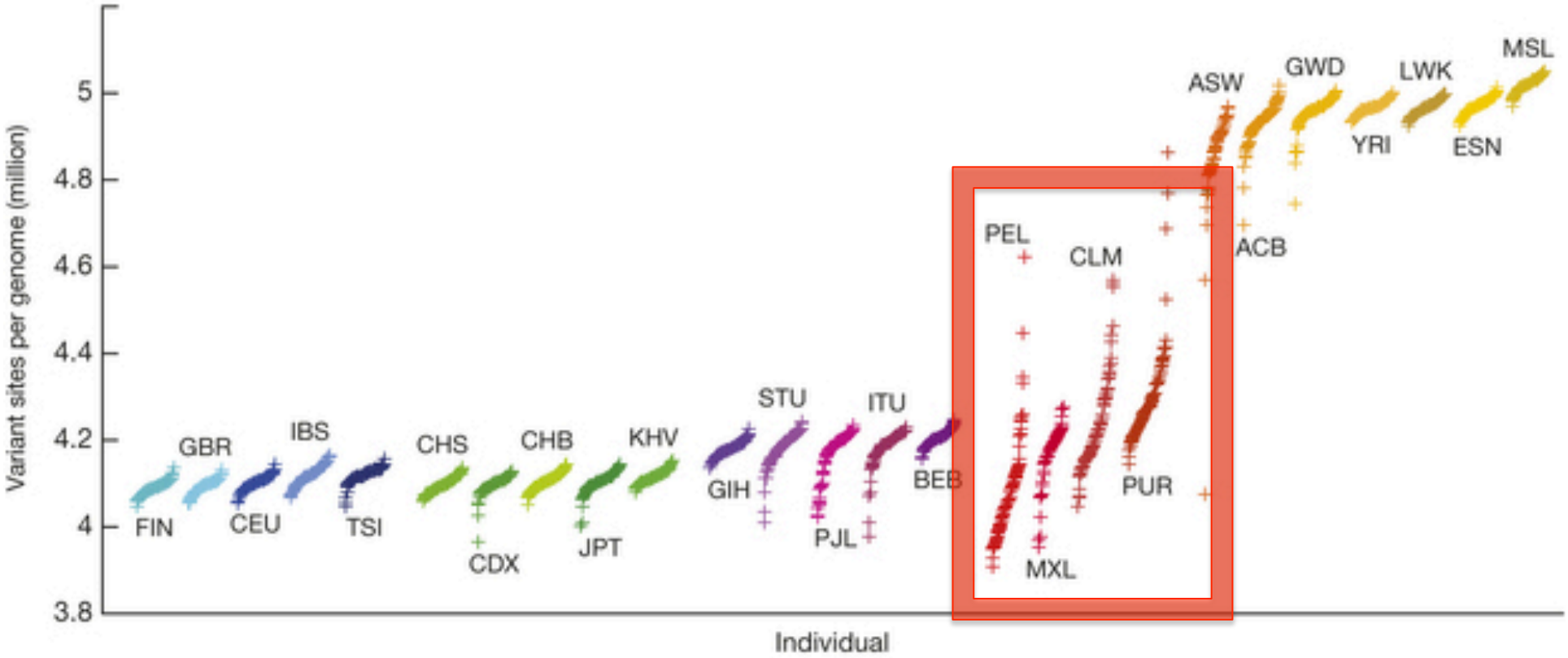
The 1000 Genomes Project Consortium; *Nature* **526**, 68–74 (01 October 2015) doi:10.1038/nature15393

The populations not in Africa are more similar to each other than those within Africa, indicating a common set of ancestors from Africa (founder population), while within Africa, the least common ancestor is further back, evolutionarily.



The 1000 Genomes Project Consortium; *Nature* **526**, 68–74 (01 October 2015) doi:10.1038/nature15393

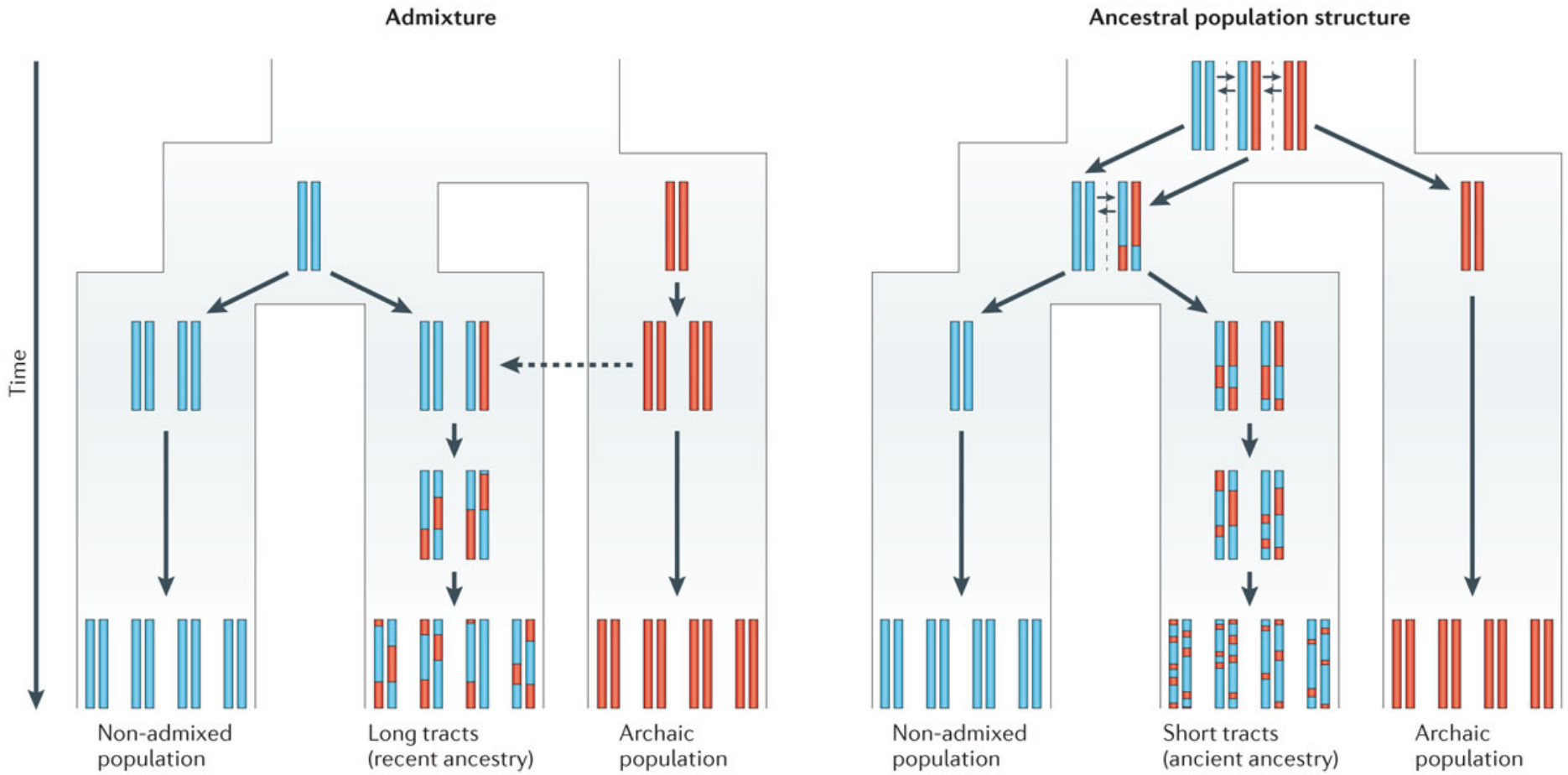
What's going on with these populations? Why do they seem to split the non-African and African patterns?



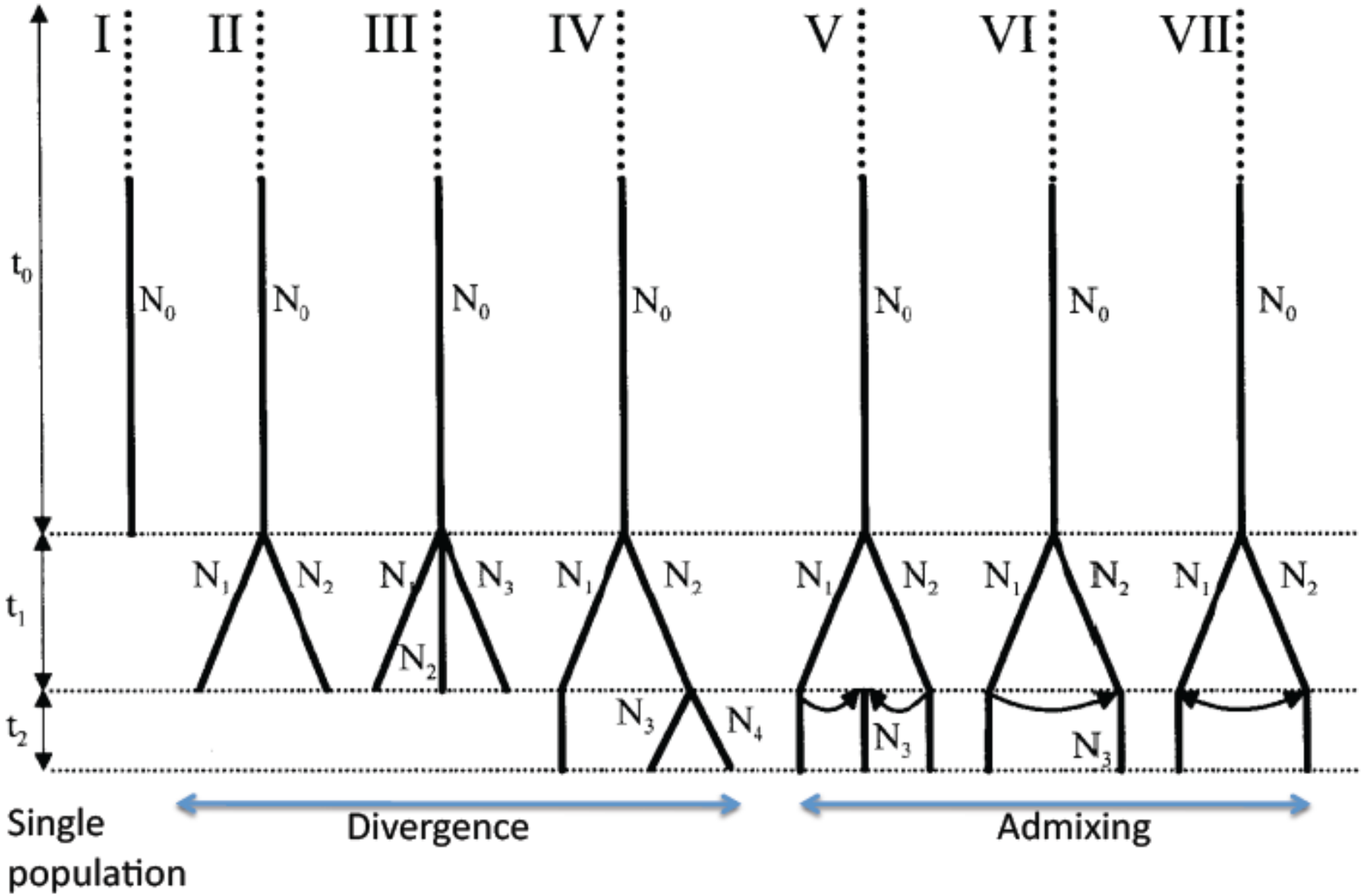
The 1000 Genomes Project Consortium; *Nature* **526**, 68–74 (01 October 2015) doi:10.1038/nature15393

What's going on with these populations? Why do they seem to split the non-African and African patterns?

They are the populations that have the highest admixture of African with non-African. Mostly places that were central to the African slave trade.



Nature Reviews Genetics 16, 359–371 (2015) doi:10.1038/nrg3936



Adapted from Seyoung Kim

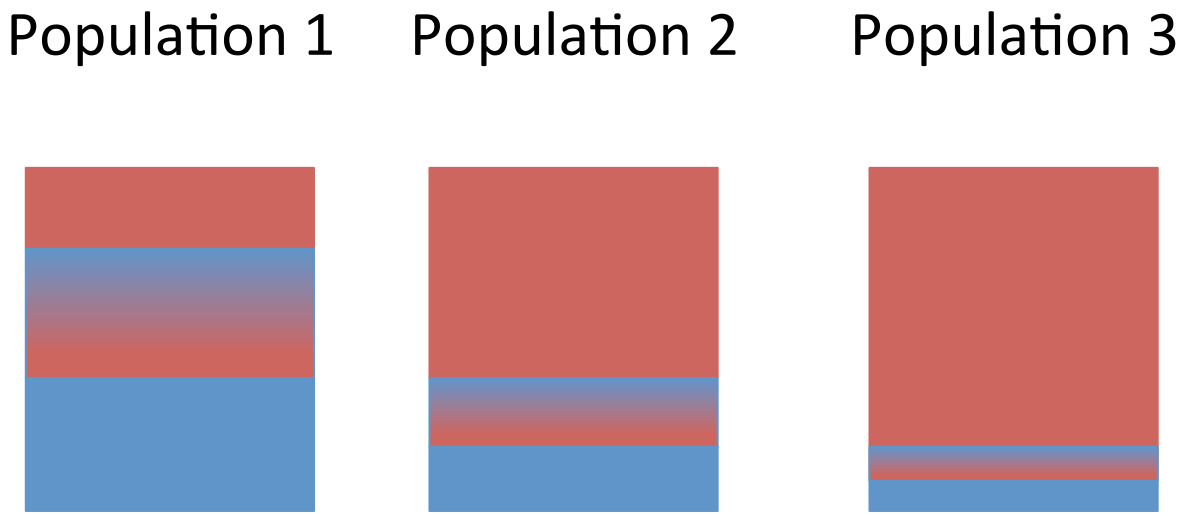
Which chromosome is not admixed or recombined? Is there other DNA in a cell that also transmits without change from generation to generation?

Which chromosome is not admixed or recombined? Is there other DNA in a cell that also transmits without change from generation to generation?

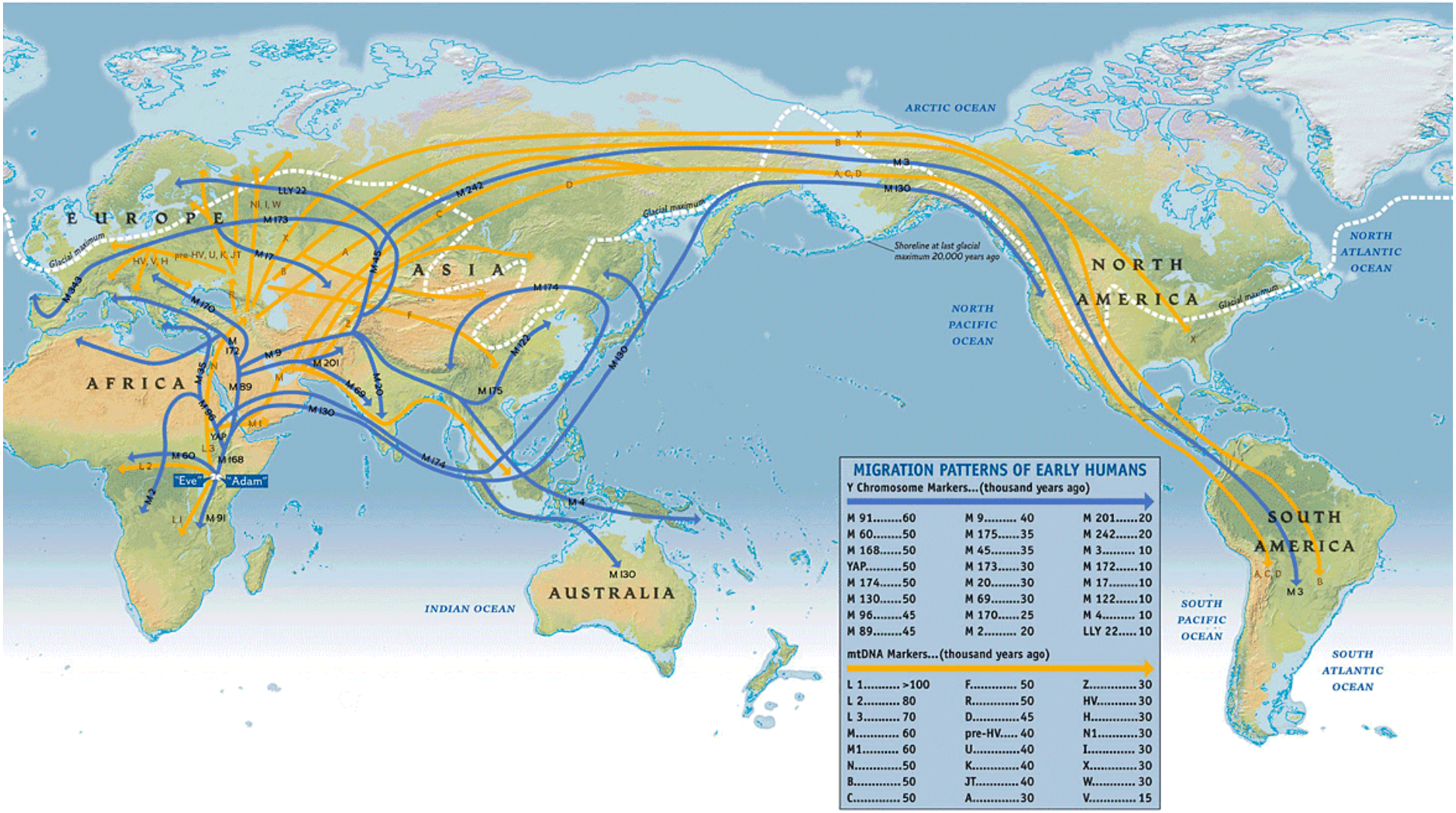
The Y chromosome doesn't have a pair, so can not recombine, and therefore doesn't get admixed.

Mitochondrial DNA is transmitted intact from the maternal line

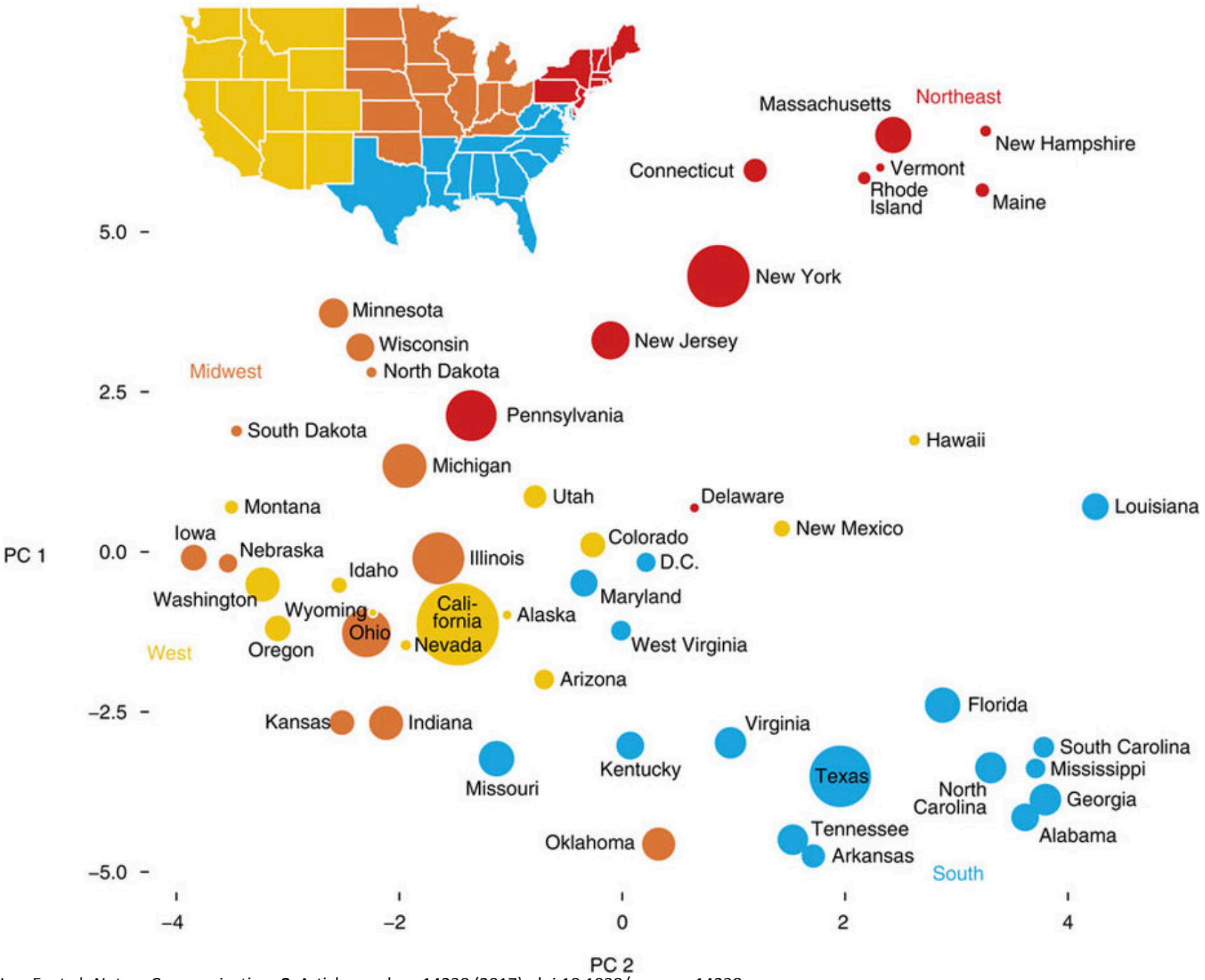
Because the allele probabilities change during evolution/migration/drift, different populations will have different probabilities of a given allele.



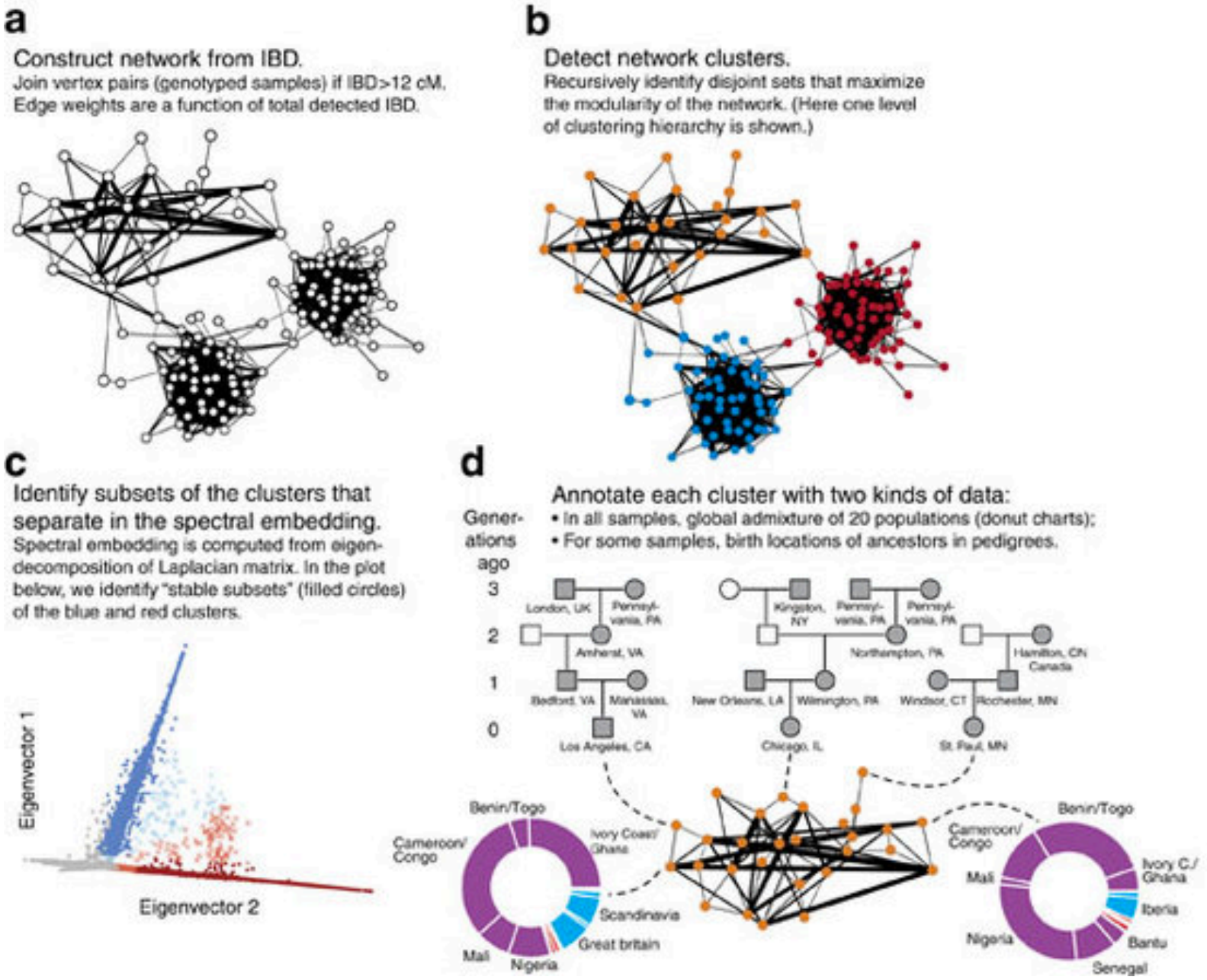
Allows us to predict ancestry, but also gets in the way of determining what alleles associate with a phenotype



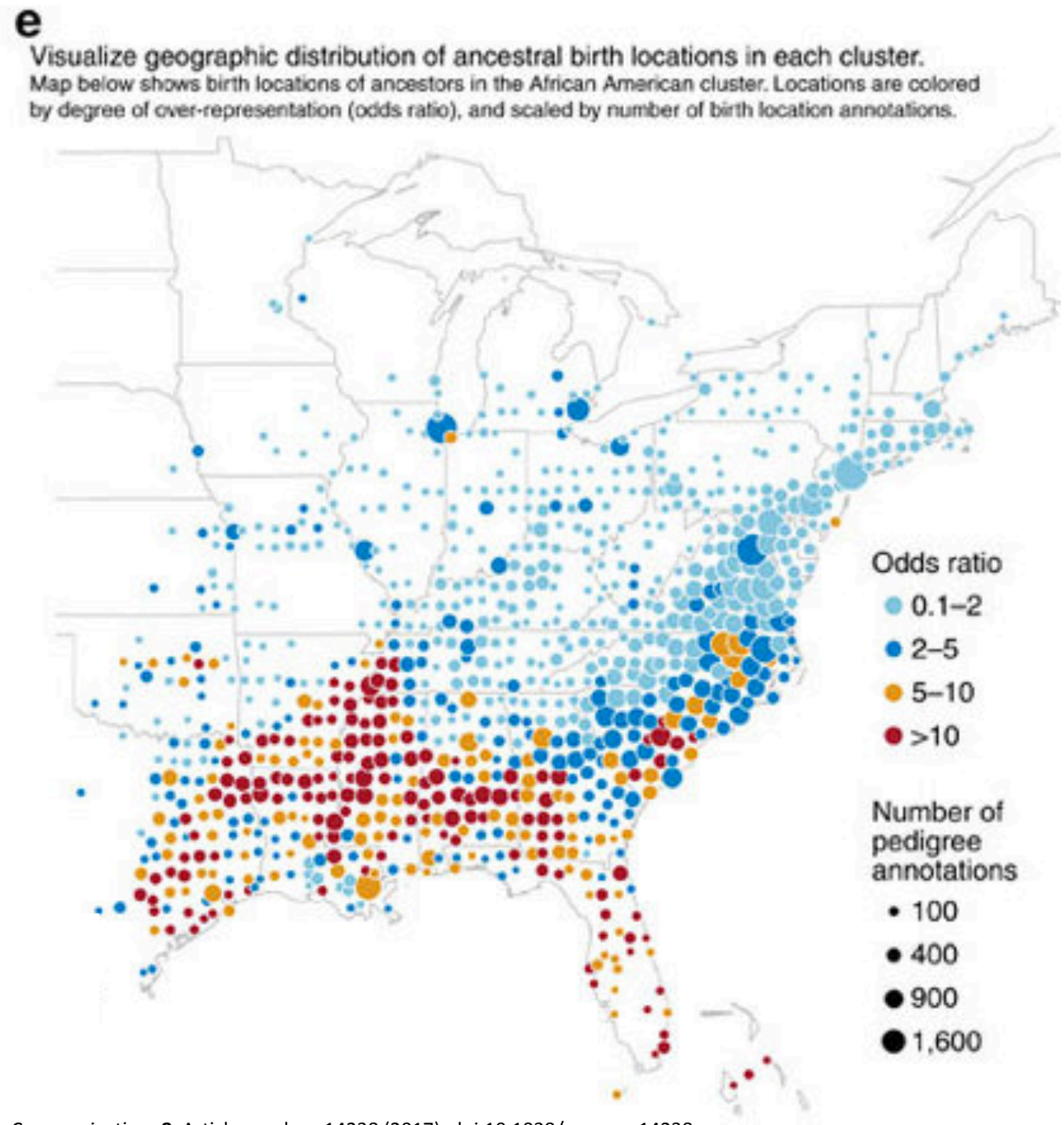
https://genographic.nationalgeographic.com/



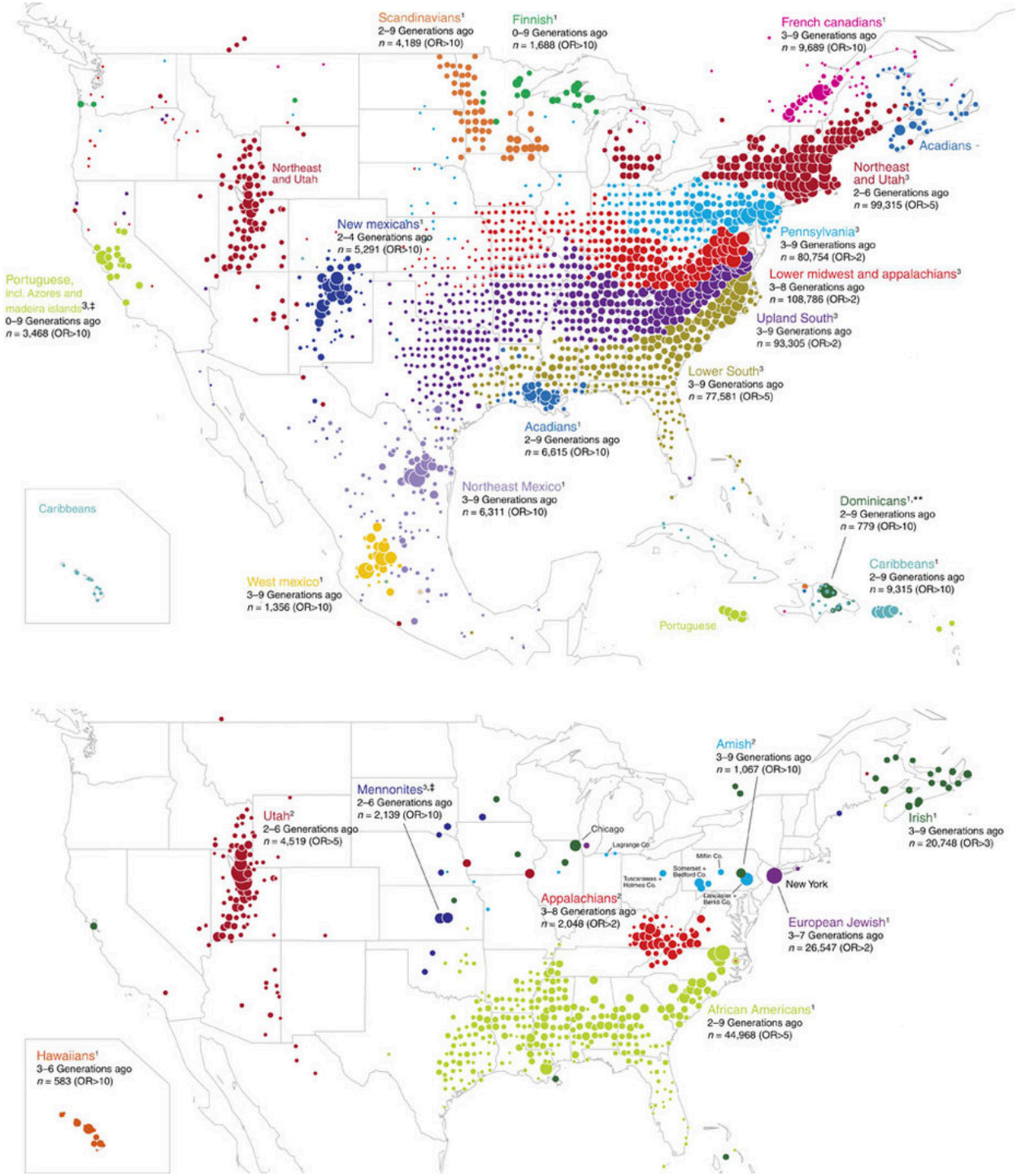
Han, E, et al; *Nature Communications* **8**, Article number: 14238 (2017); doi:10.1038/ncomms14238



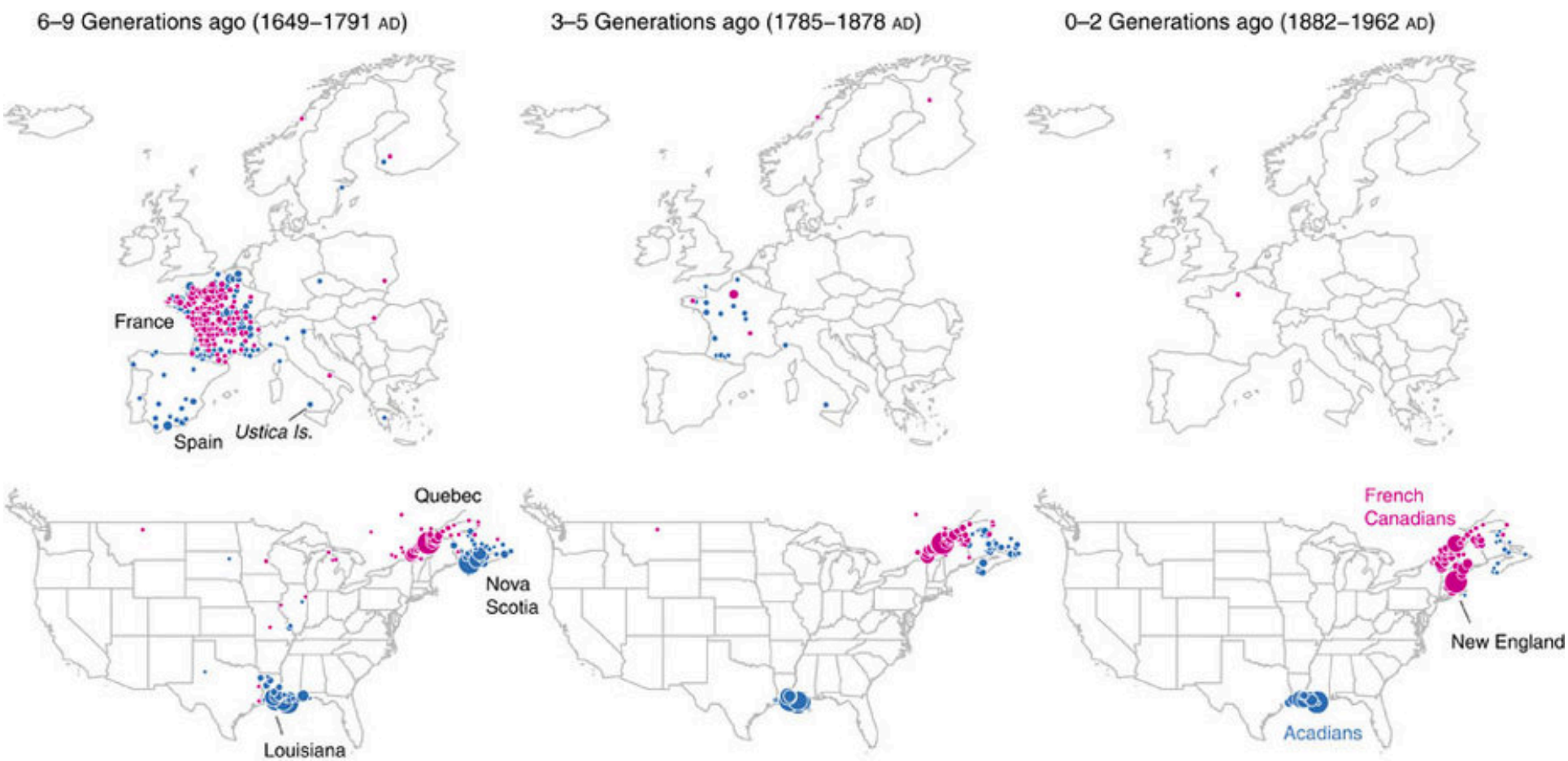
Han, E, et al; *Nature Communications* **8**, Article number: 14238 (2017); doi:10.1038/ncomms14238



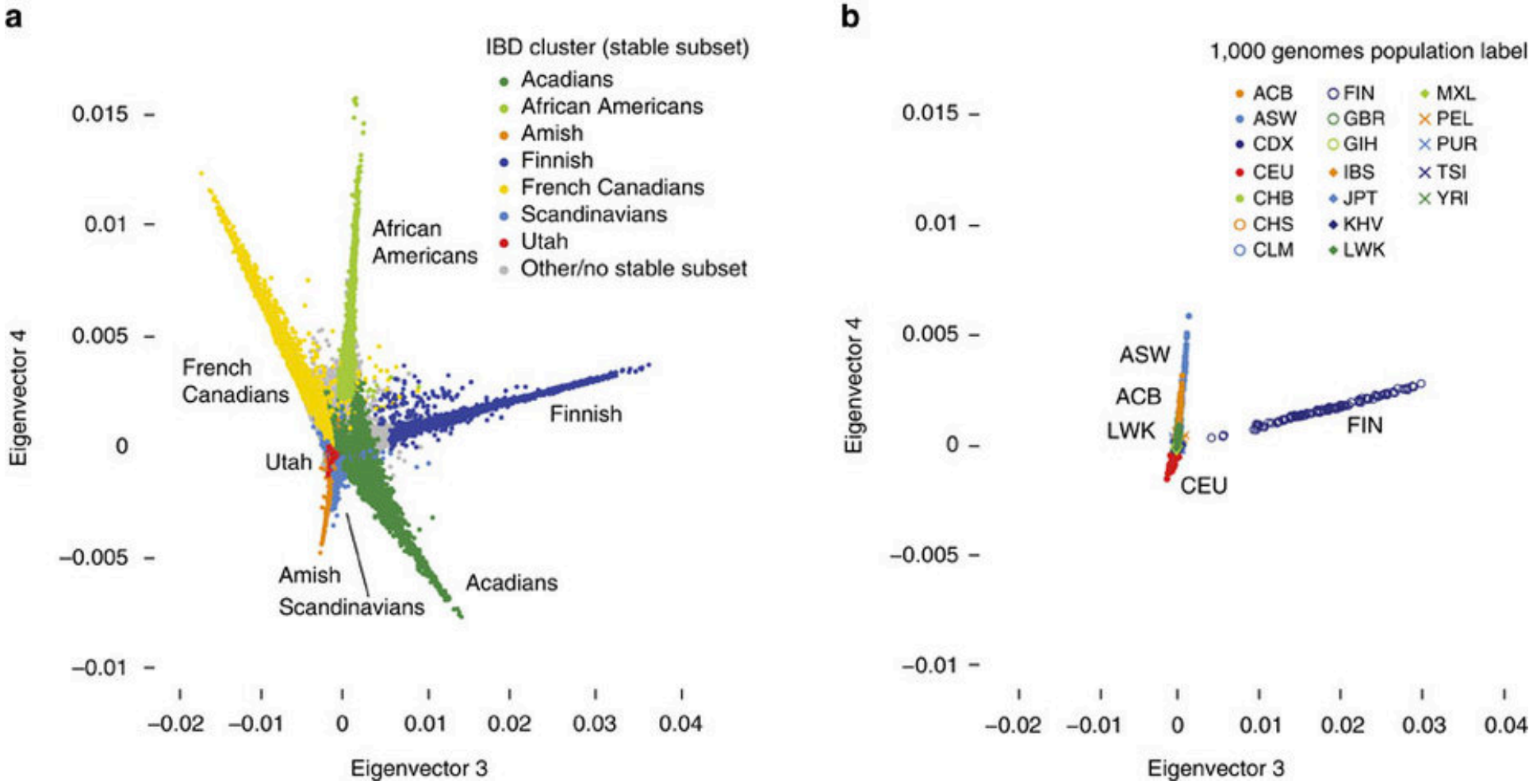
Han, E, et al; *Nature Communications* **8**, Article number: 14238 (2017); doi:10.1038/ncomms14238



Han, E, et al; *Nature Communications* **8**, Article number: 14238 (2017); doi:10.1038/ncomms14238



Han, E, et al; *Nature Communications* **8**, Article number: 14238 (2017); doi:10.1038/ncomms14238



Han, E, et al; *Nature Communications* **8**, Article number: 14238 (2017); doi:10.1038/ncomms14238

Far more information than in 1000 Genome Project! Greater diversity in some ancestries

Difficulties in their data

- Some populations had poor ancestry data
  - African–Americans
  - Ashkenazi Jews
- Data still very biased toward European ancestry
- The way they biased the data to increase resolution in recent ancestry, omits populations that are less distincitve/more admixed
  - African-Americans communities in Northern cities for generations

Implications of this study

- Exciting that we can trace migrations so tightly
- Private data set
  - Ancestry.com is not a public repository
    - Company grew out of LDS genealogy projects
  - They got their data by having people pay for their own data

Summary

Next Class

- Guest speaker on genomics of breast and ovarian cancer
- Please come prepared to ask questions