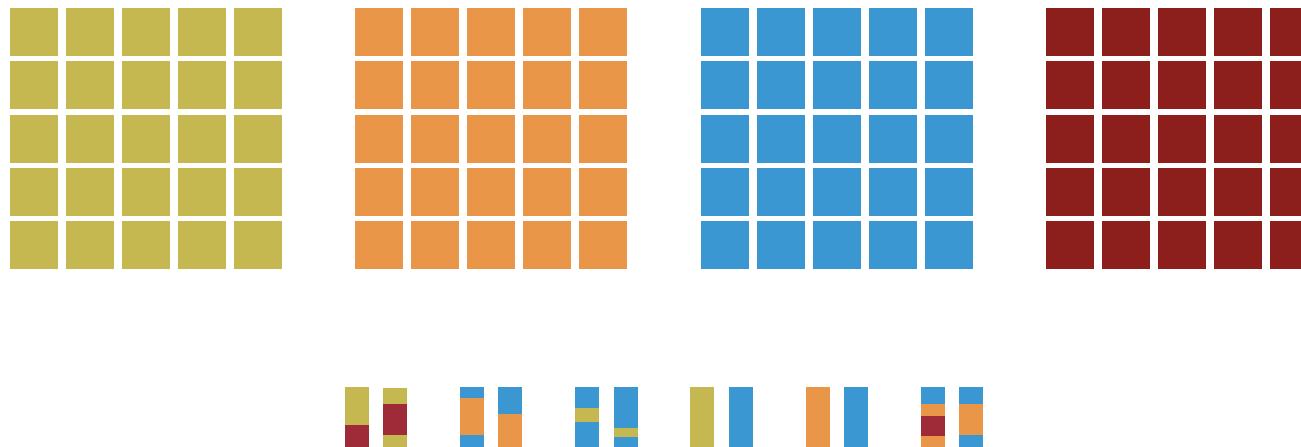


Lecture 6

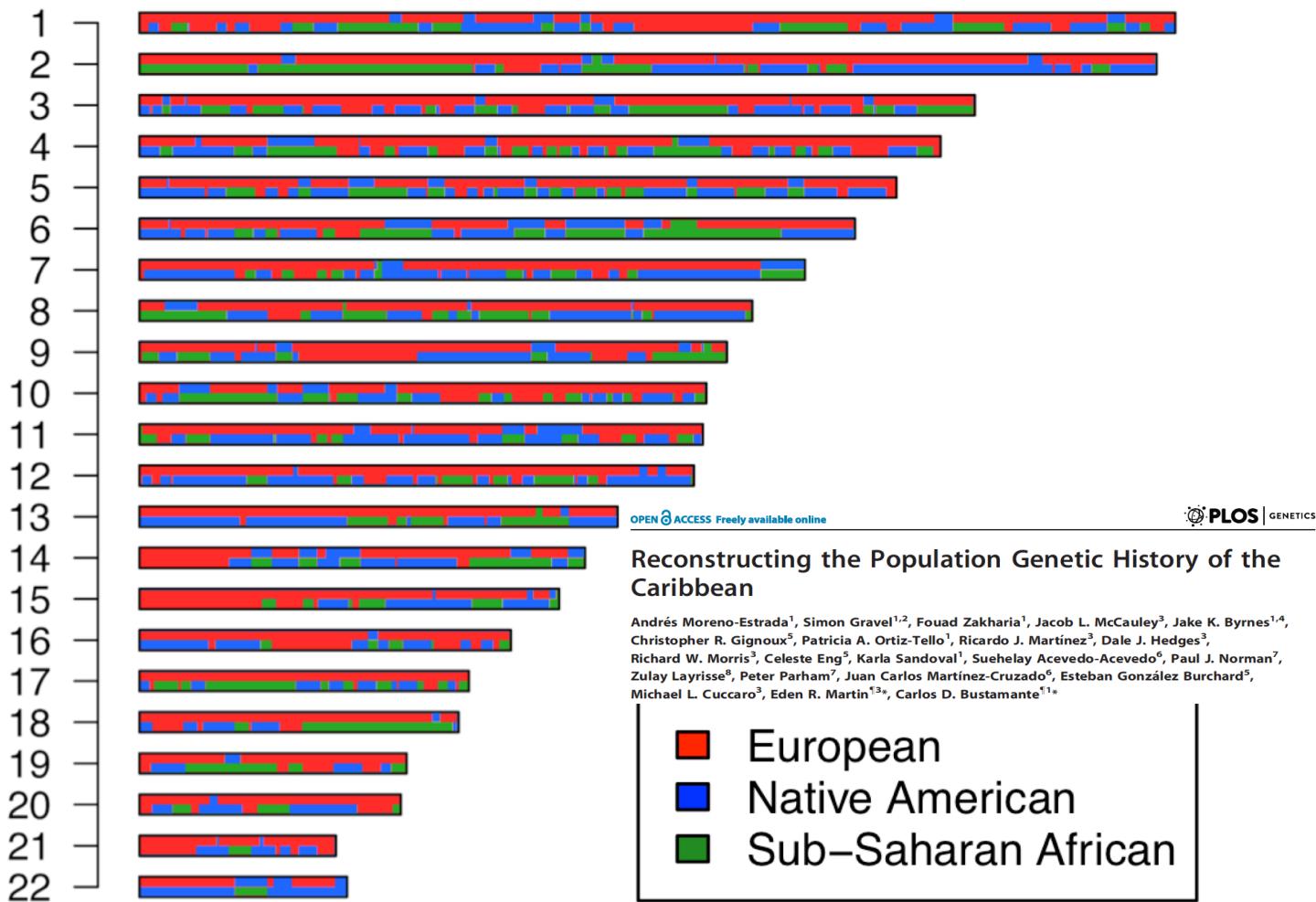
Reference Populations



Identify origins of **chromosomal segments** in individuals of **admixed** ancestry

Approaches: Based on Hidden Markov Models

Local ancestry



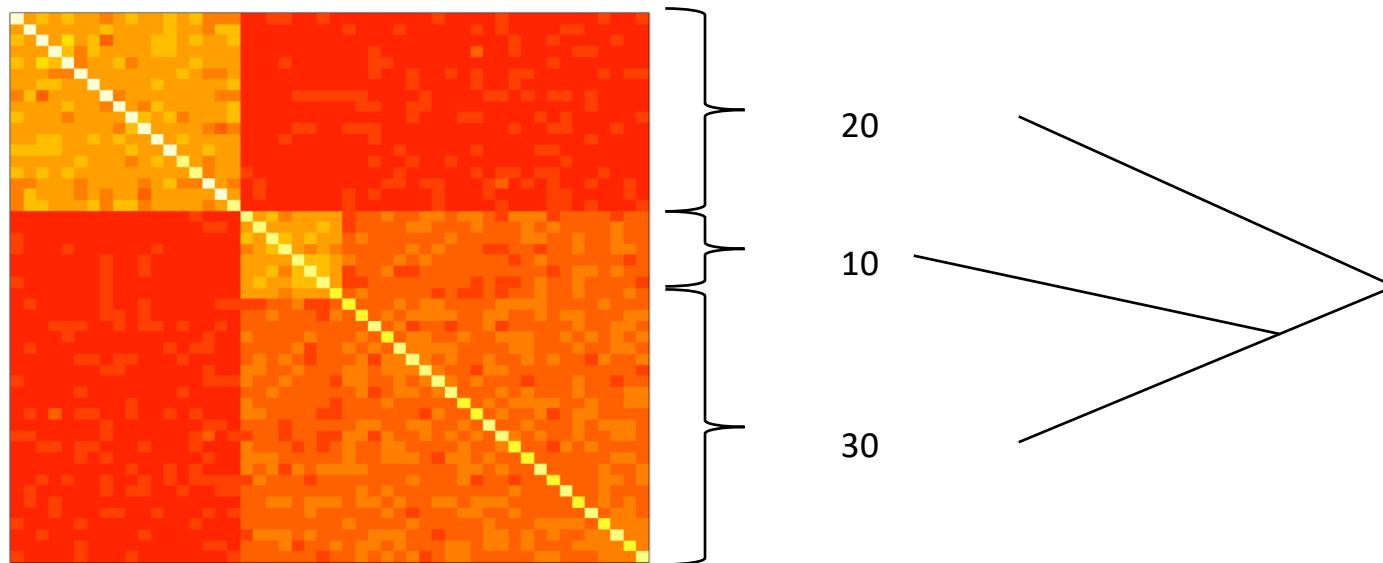
Principal components analysis

Example

Section 2.3.4 of notes

(Simulated data, N=50 individuals, L=1000 SNPs)

Relatedness matrix R



i^{th} and j^{th} entry = average over loci (l) of $(X_{il} - \bar{X}_l)(X_{jl} - \bar{X}_l)$

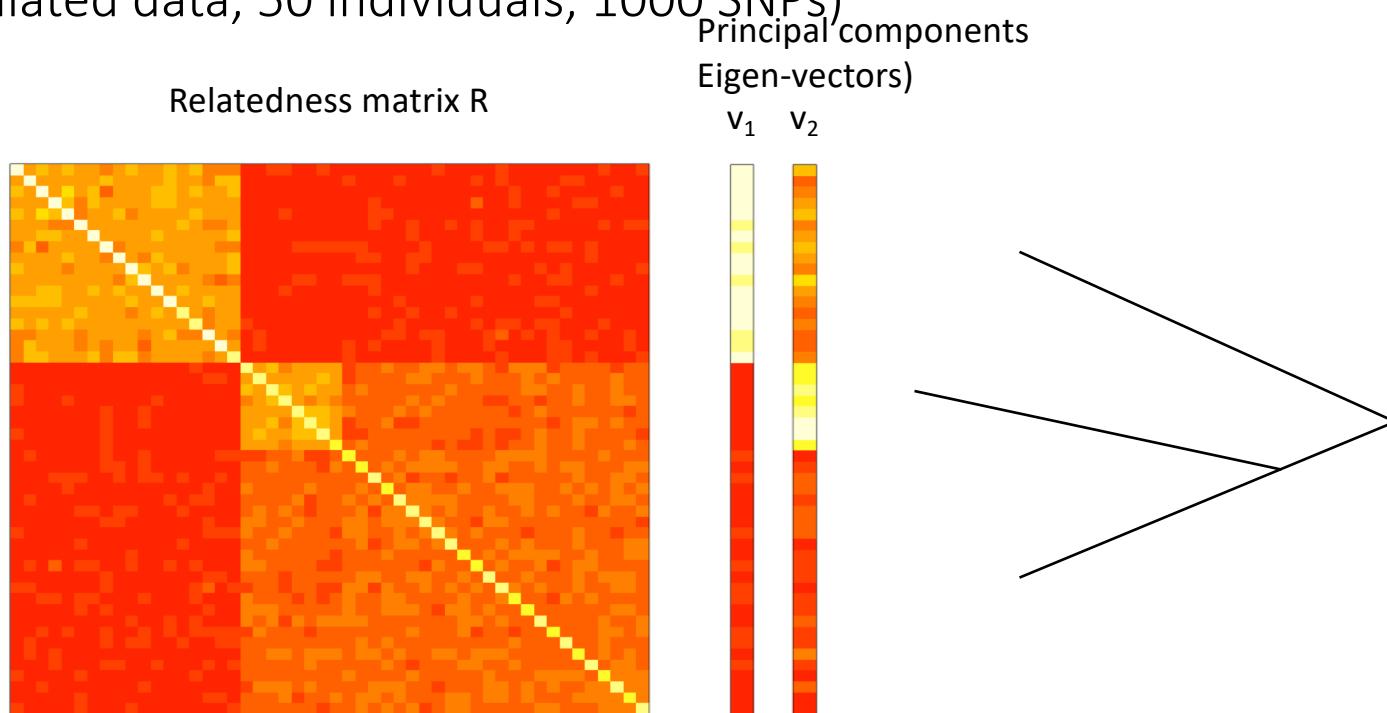
Where X_l is mean freq. of the l^{th} locus.

Modified from slide by Gavin Band

Principal components analysis

Example

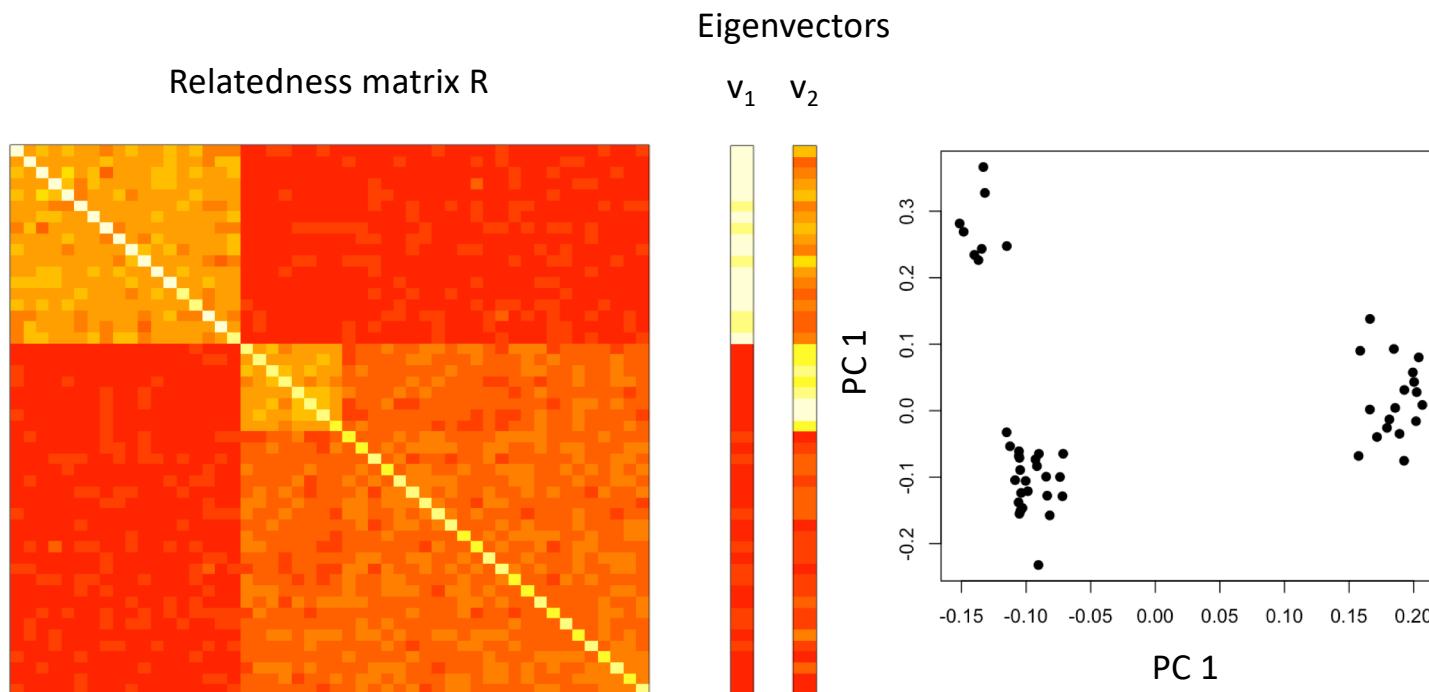
(Simulated data, 50 individuals, 1000 SNPs)



Modified from slide by Gavin Band

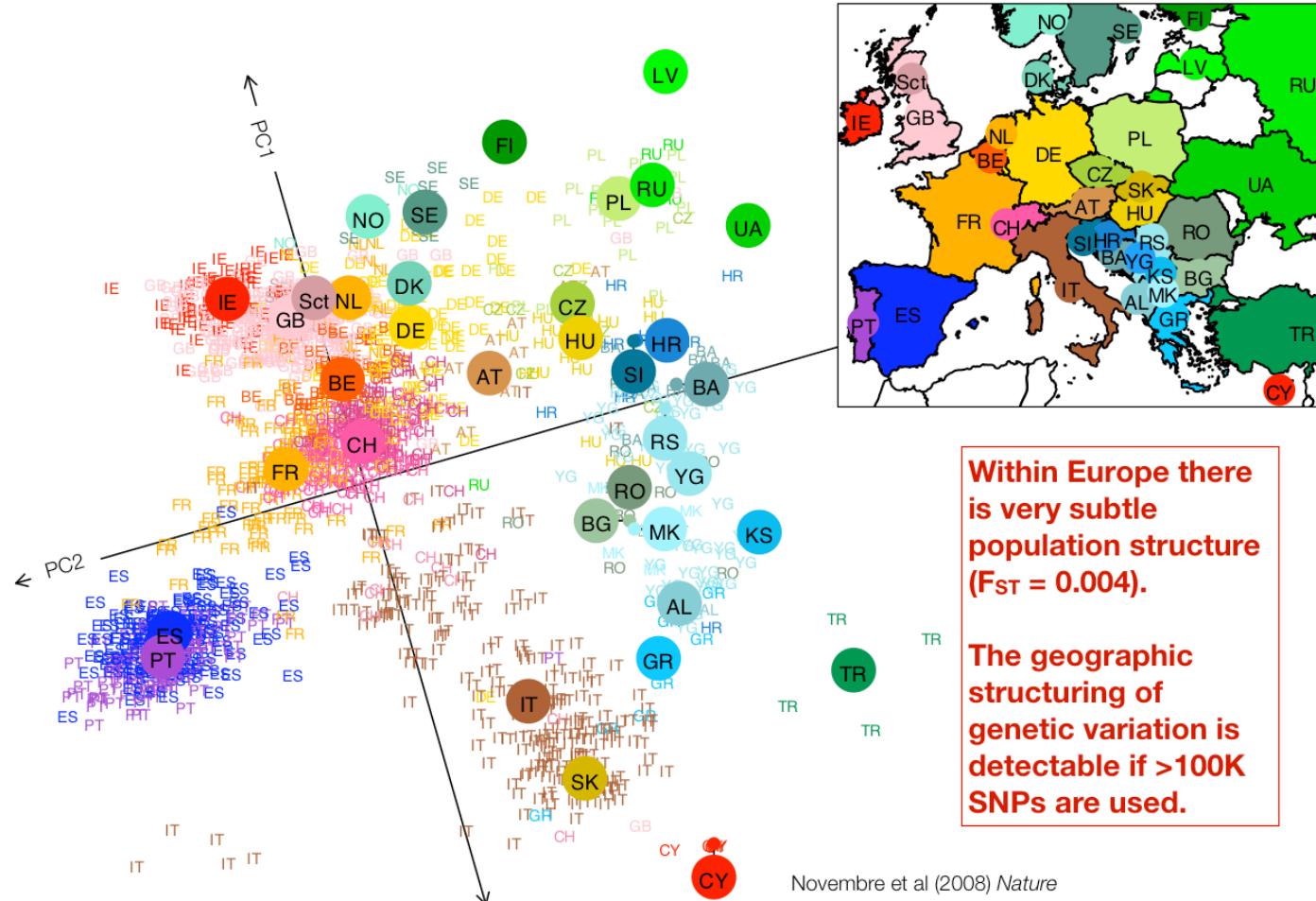
Principal components analysis Example

(Simulated data, 50 individuals, 1000 SNPs)

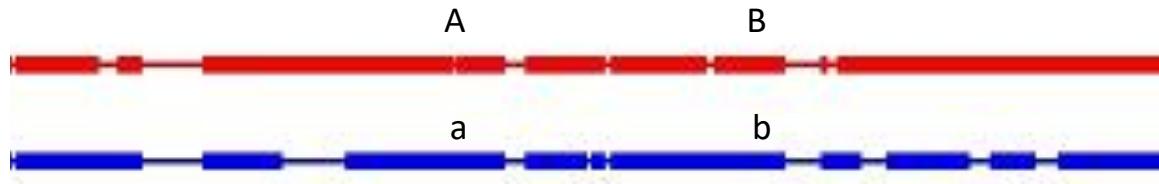


Modified from slide by Gavin Band

Principal Component Analysis of Europeans

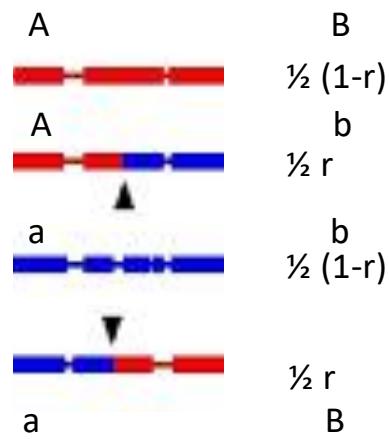


Recombination and Linkage Disequilibrium (LD)



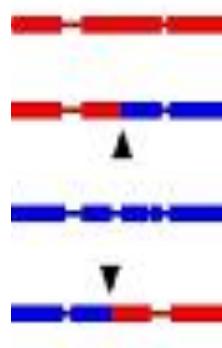
r = recombination fraction
 probability of an odd
 Number of crossovers occur
 Between our markers

$$0 < r < \frac{1}{2}$$



Linkage disequilibrium: The non-random association of alleles at different sites in the genome in a population.

If independent the expected



p_{AB} = frequency of AB frequency of gametes (haplotypes)

$$p_A \times p_B$$

$$p_a \times p_b$$

$$p_A \times p_b$$

p_{Ab} = frequency of Ab

p_{aB} = frequency of aB

$$p_a \times p_B$$

Define “D”

$$D_{AB} = p_{AB} - p_A p_B$$

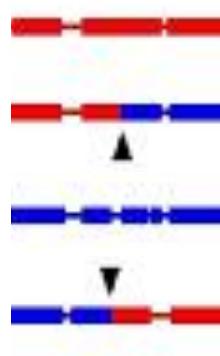
The covariance of A and B.

$$D_{ab} = p_{ab} - p_a p_b$$

$$D_{Ab} = p_{Ab} - p_A p_b$$

$$D_{aB} = p_{aB} - p_a p_B$$

Linkage disequilibrium: The non-random association of alleles at different sites in the genome.



Define “D”

$$D_{AB} = p_{AB} - p_A p_B$$

$$D_{AB} = - D_{Ab}$$

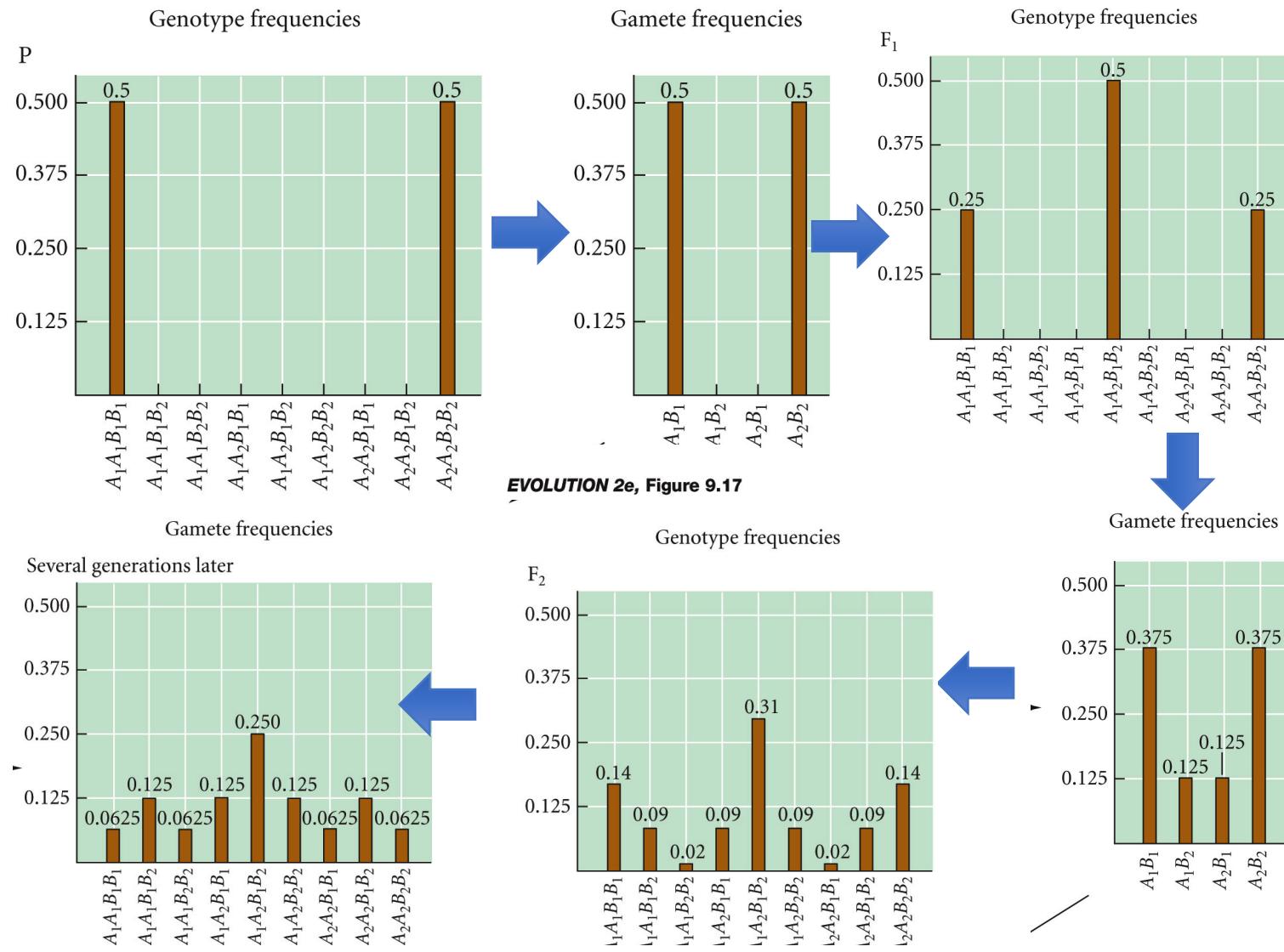
$$D_{AB} = D_{ab} \text{ and } D_{Ab} = D_{aB}$$

(so, knowing D_{AB} is enough - call this “D”)

If $O = E$, then $D = 0$

If $D > 0$ (or $D < 0$) then there is “linkage disequilibrium (LD)”

Note: you can also write $p_{AB} = p_A p_B + D$



Decay of LD in a very large boring
randomly mating population

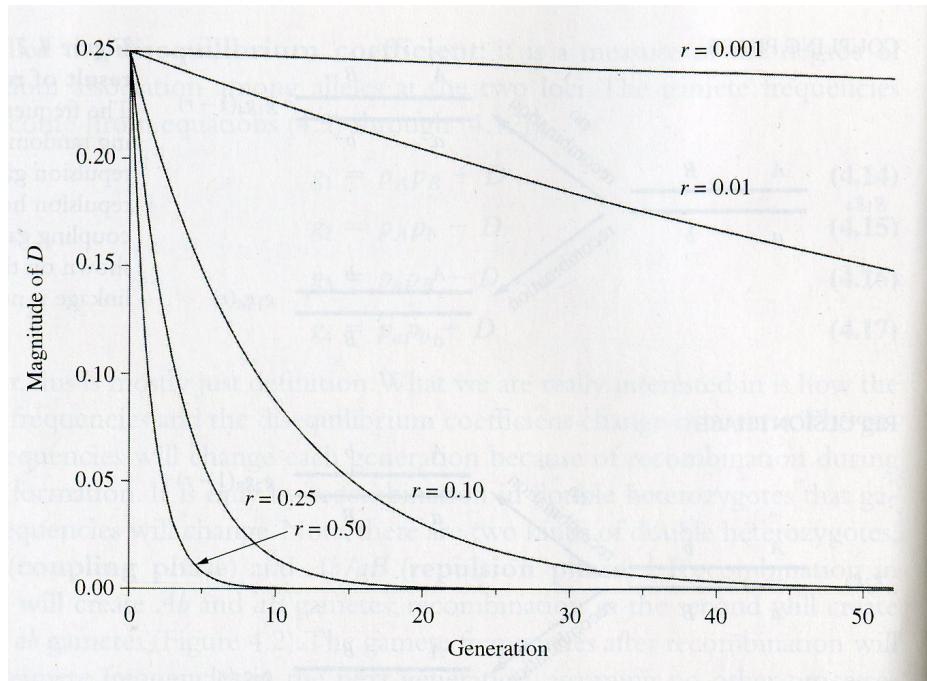
$$D_t = (1 - r)^t D_0$$

With inbreeding coefficient f replace r with $r(1-f)$

linkage disequilibrium

How does LD change over time due to recombination?

$$D_t = (1 - r)^t D_0$$



Note: more distant markers recombine more!

So eventually recombination leads to $D=0$.
Even with free recombination ($r=0.5$), it isn't instantaneous

What creates LD

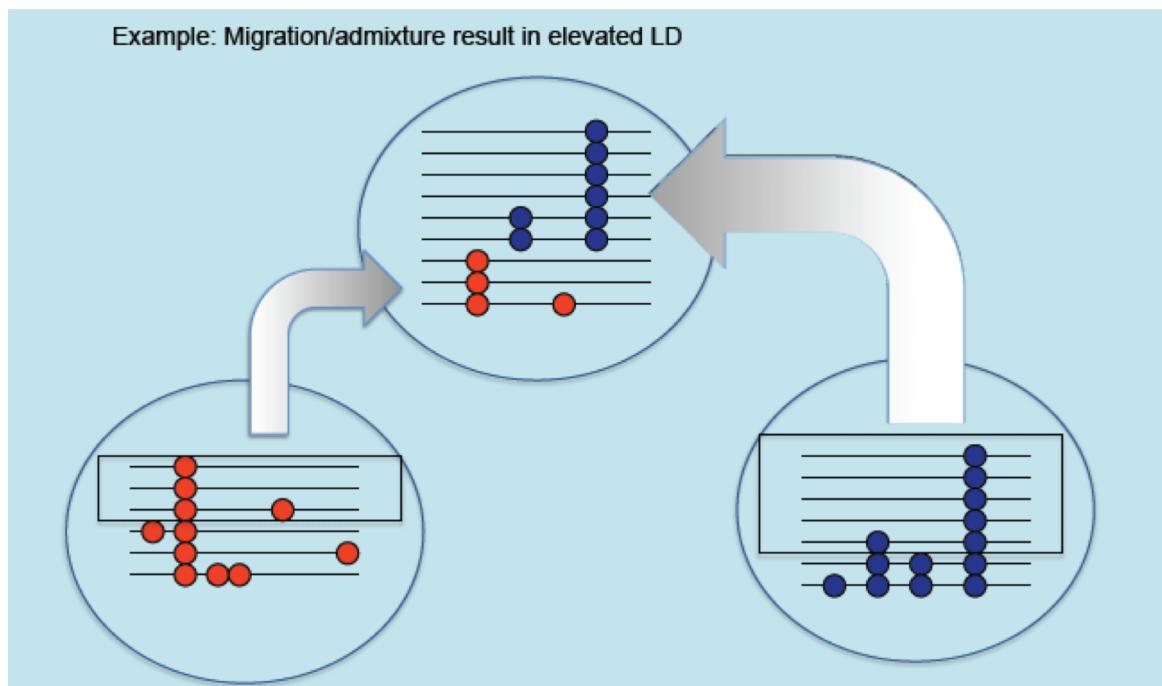
- Mutational origin
- Genetic drift (and Hitchhiking)
- Epistatic selection*
- Assortative mating.
 - Inbreeding
 - Population structure and admixture
 - Assortative mating by phenotype*
- *only for specific markers

What creates LD

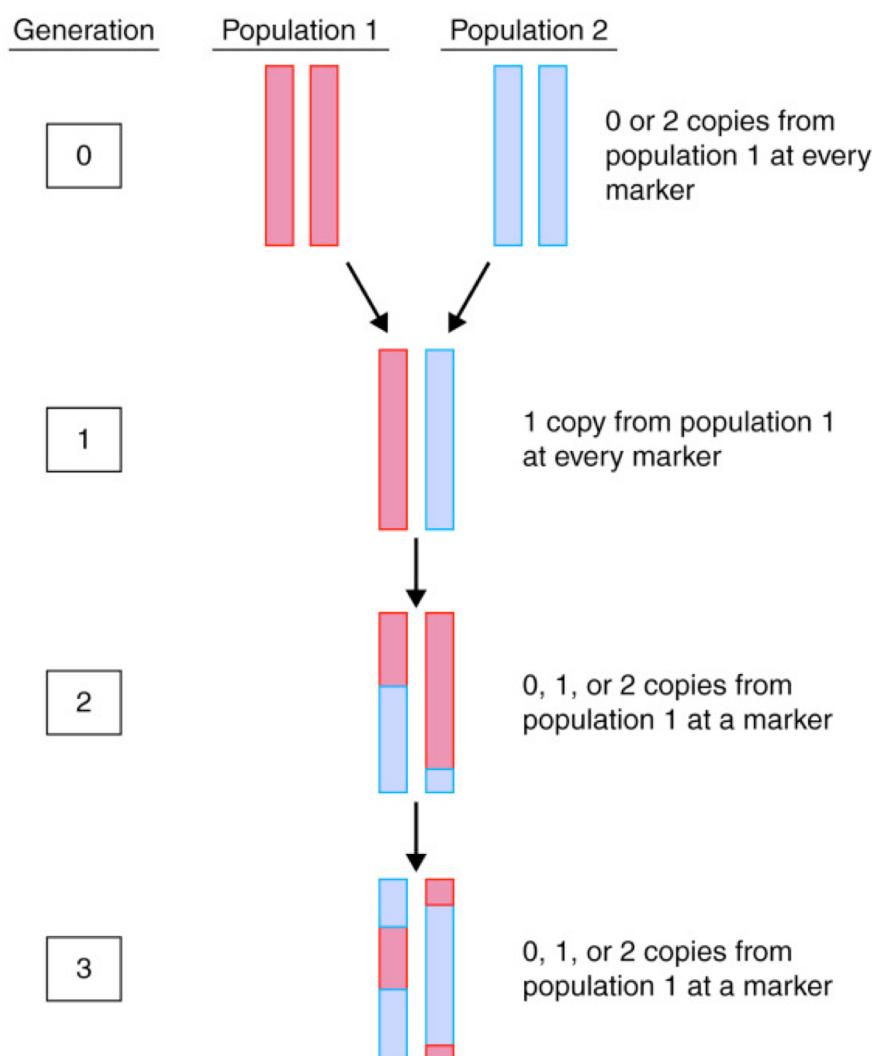
- Mutational origin
- Genetic drift (and Hitchhiking)
- Epistatic selection*
- Assortative mating.
 - Inbreeding
 - Population structure and admixture
 - Assortative mating by phenotype*
- *only for specific markers

Countervailing forces that increase LD

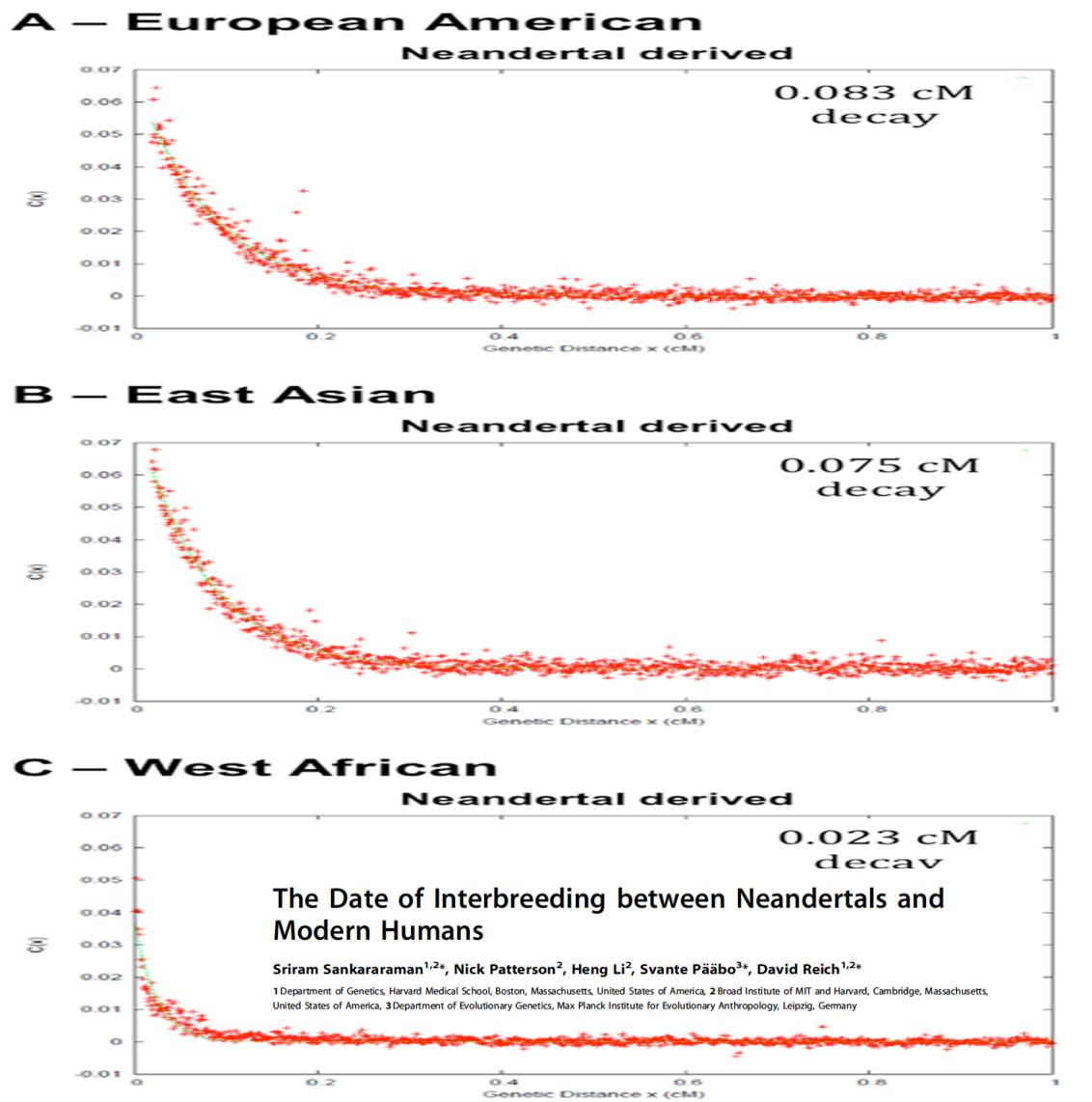
Population structure can increase LD if allele frequencies differ among populations



Pretty pictures courtesy of P. Andolfatto (Princeton)



LD between Neanderthal alleles in modern human populations



Evolution by genetic drift

- **Evolution by Genetic drift:** a change in allele frequency because individuals carry the allele by chance produce more / less offspring in any given generation. *
- *in sexual populations....
 - Genetic drift can affect selected alleles but only if they are very weakly selected (except when they are rare).
 - A neutral allele: An allele with no effect on fitness from other alleles at that locus.

- Neutral polymorphism/alleles:
 - Only 2% of our genome encodes for proteins
 - Changes outside exons may be completely neutral if they do not disrupt regulatory sites.
- Examples of potentially neutral alleles:
 - A synonymous change in a codon.
 - A non-synonymous change that replaces one amino-acid with a functionally similar one.
 - A non-synonymous change which produces a large change in a phenotype on which selection no longer acts.

How much of genetic divergence between species is neutral.

- ~36 million substitutions have occurred since human and chimp last shared a common ancestor.
 - How many of these substitutions fix due to selection?
 - How much of polymorphism is neutral?

(A)	<i>Homo</i>	* * * * C A C A A T A	T	G A G C	T	G A A G A G A T	T	G	T	G A A A A G	T	A
	<i>Pan</i>	* * * *	A	G	C	G	C	G	G	G	G	*
	<i>Gorilla</i>	T A A T	A A	T	T	T	G G	T	G	A A T A T A T A	A A T A T A T A	
	<i>Pongo</i>	T A A T	A A	T	T	T	C G	T	C	A A T A T A T A	A A T A T A T A	
	<i>Macaca</i>	T A A T	C G	T	T	T	C A	T	T	G A A T A T A	G A A T A T A	
	<i>Atelus</i>	T A A T	A C	T	T	T	A T			G A C T A T A	G A C T A T A	
		• • • • •	• • • • •	•	• • • • •	•	• • • • •	•	• • • • •	•	• • • • •	•
		3903	3913	5361	5365	6367	6375	8224	8230	8468	8474	

Why is there so much polymorphism?

- The paradox of variation in population genetics:
Selection quickly fixes alleles that are beneficial so why is there so much genetic polymorphism within natural populations?

Three explanations:

- Balancing selection
- Mutation-selection balance
- Mutation-genetic drift balance (Neutral theory).

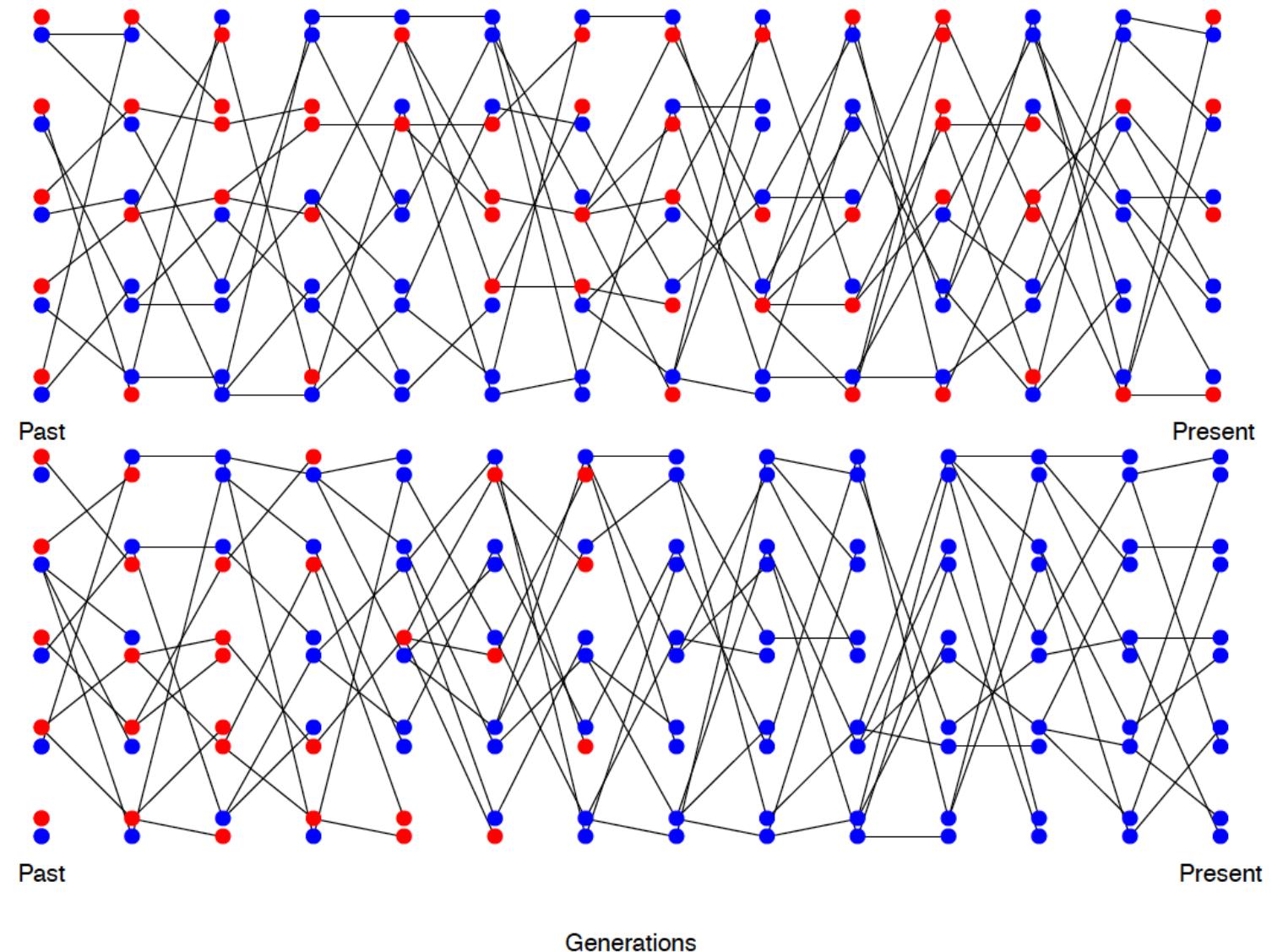
Neutral theory of molecular evolution

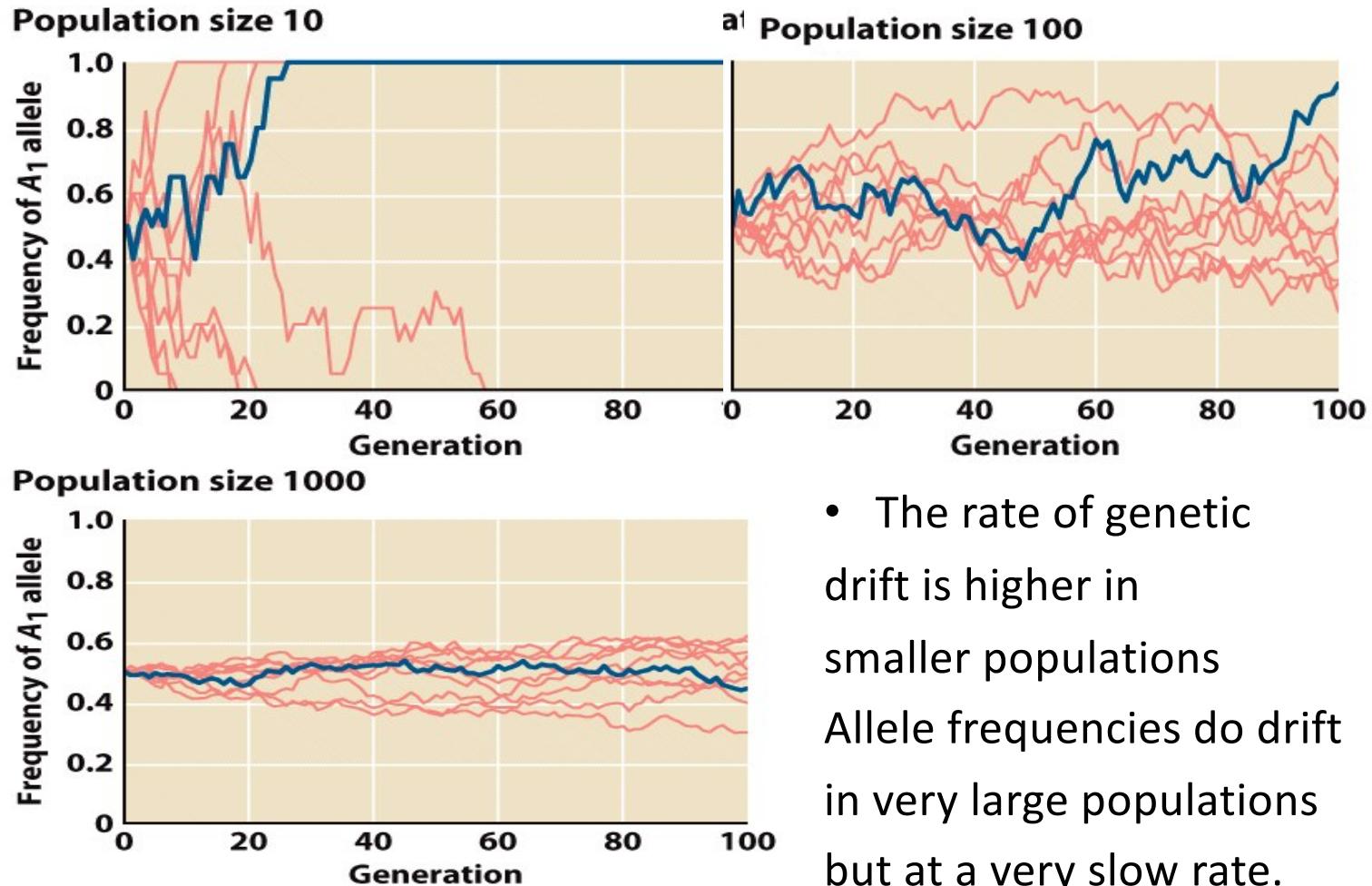
- Kimura 1968; King and Jukes 1969
- Claimed:
 - Most new mutations are deleterious and are lost immediately
 - Most of the observed molecular polymorphism and substitutions are neutral



Claimed that this is consistent with:

- High levels of genetic polymorphism
- The molecular clock





Evolution, 1/e Figure 8.3

© 2012 W. W. Norton & Company, Inc.