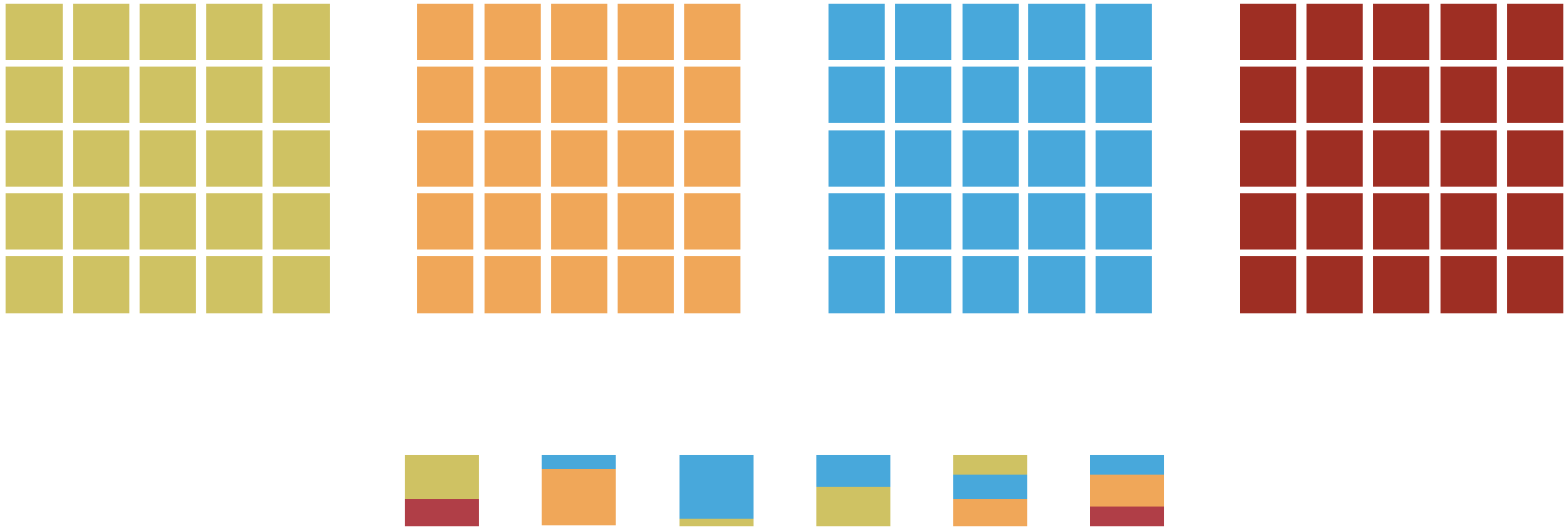
Section 1.7.1 of notes

Reference Populations (known or unknown)



Identify **ancestry proportions** for individuals with **admixed** ancestry

Approaches: Structure (MCMC, Bayesian)
Or ADMIXTURE (quadratic programming)

Adapted from John Novembre slide

Genetic Structure of Human Populations

Noah A. Rosenberg,[1]* Jonathan K. Pritchard,[2] James L. Weber,[3]
Howard M. Cann,[4] Kenneth K. Kidd,[5] Lev A. Zhivotovsky,[6]
Marcus W. Feldman[7]

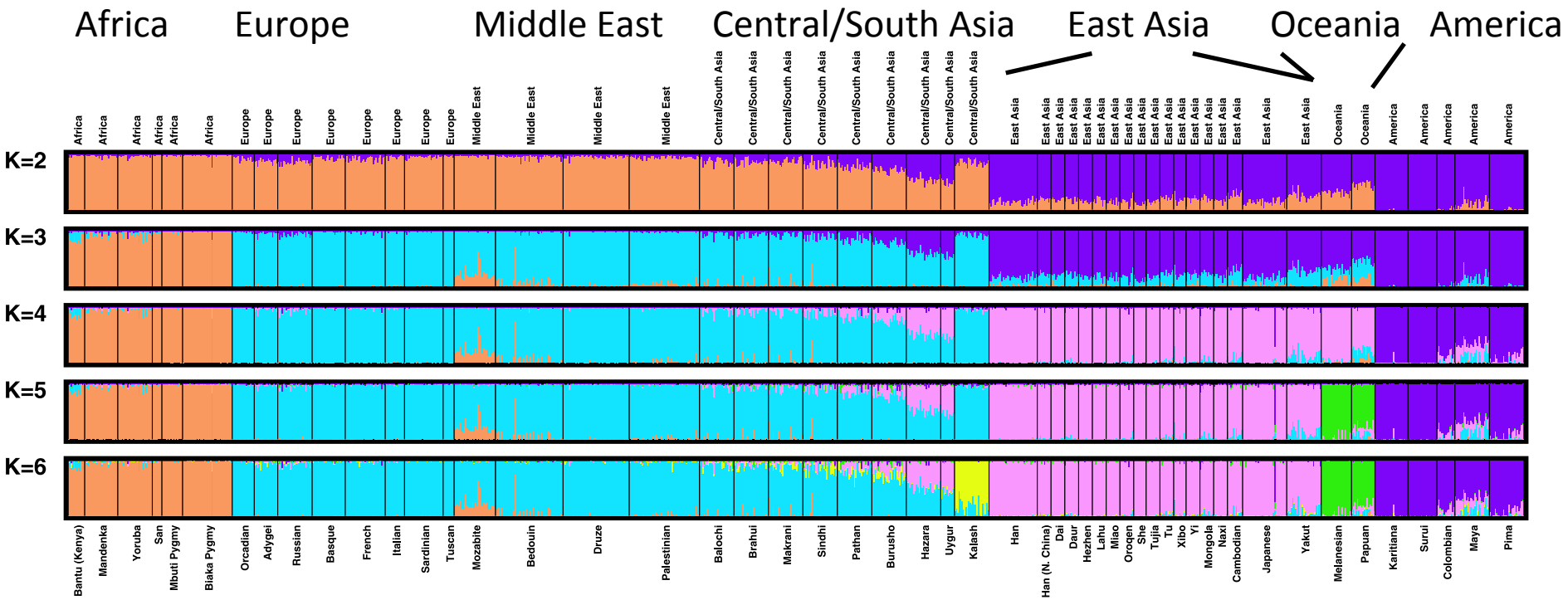
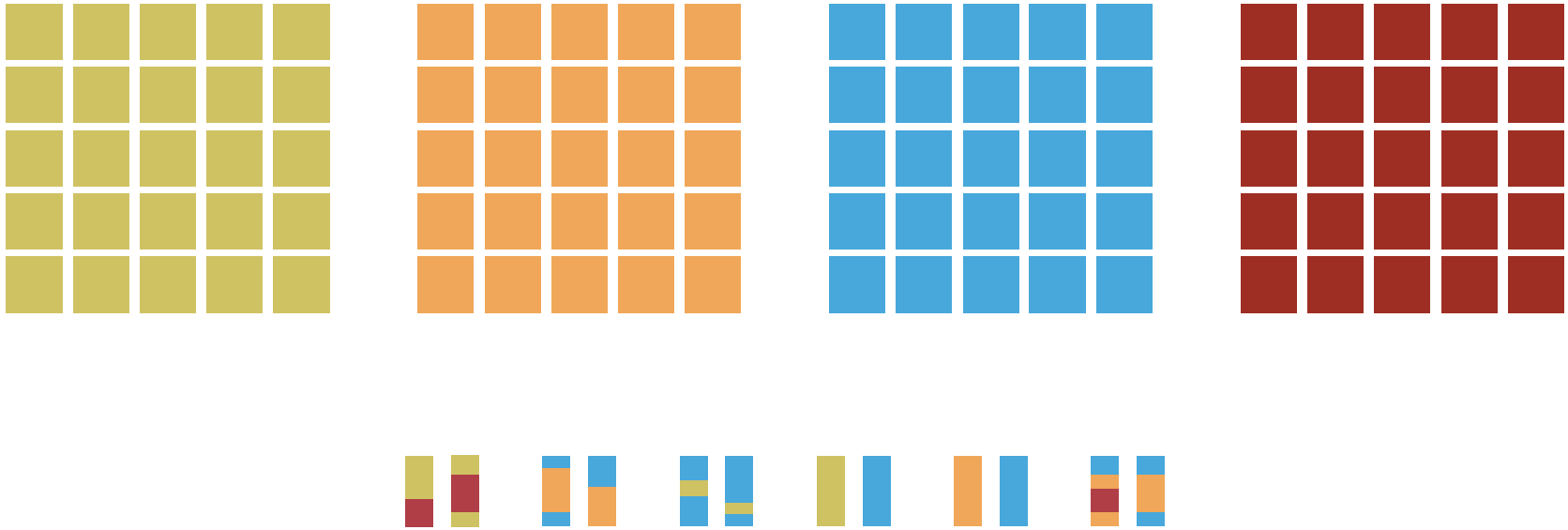


Fig. 1. Estimated population structure. Each individual is represented by a thin vertical line, which is partitioned into *K* colored segments that represent the individual's estimated membership fractions in *K* clusters. Black lines separate individuals of different populations. Populations are labeled below the figure, with their regional affiliations above it. Ten *structure* runs at each *K* produced nearly identical individual membership coefficients, having pairwise similarity coefficients above 0.97, with the exceptions of comparisons involving four runs at $K = 3$ that separated East Asia instead of Eurasia, and one run at $K = 6$ that separated Karitiana instead of Kalash. The figure shown for a given *K* is based on the highest probability run at that *K*.

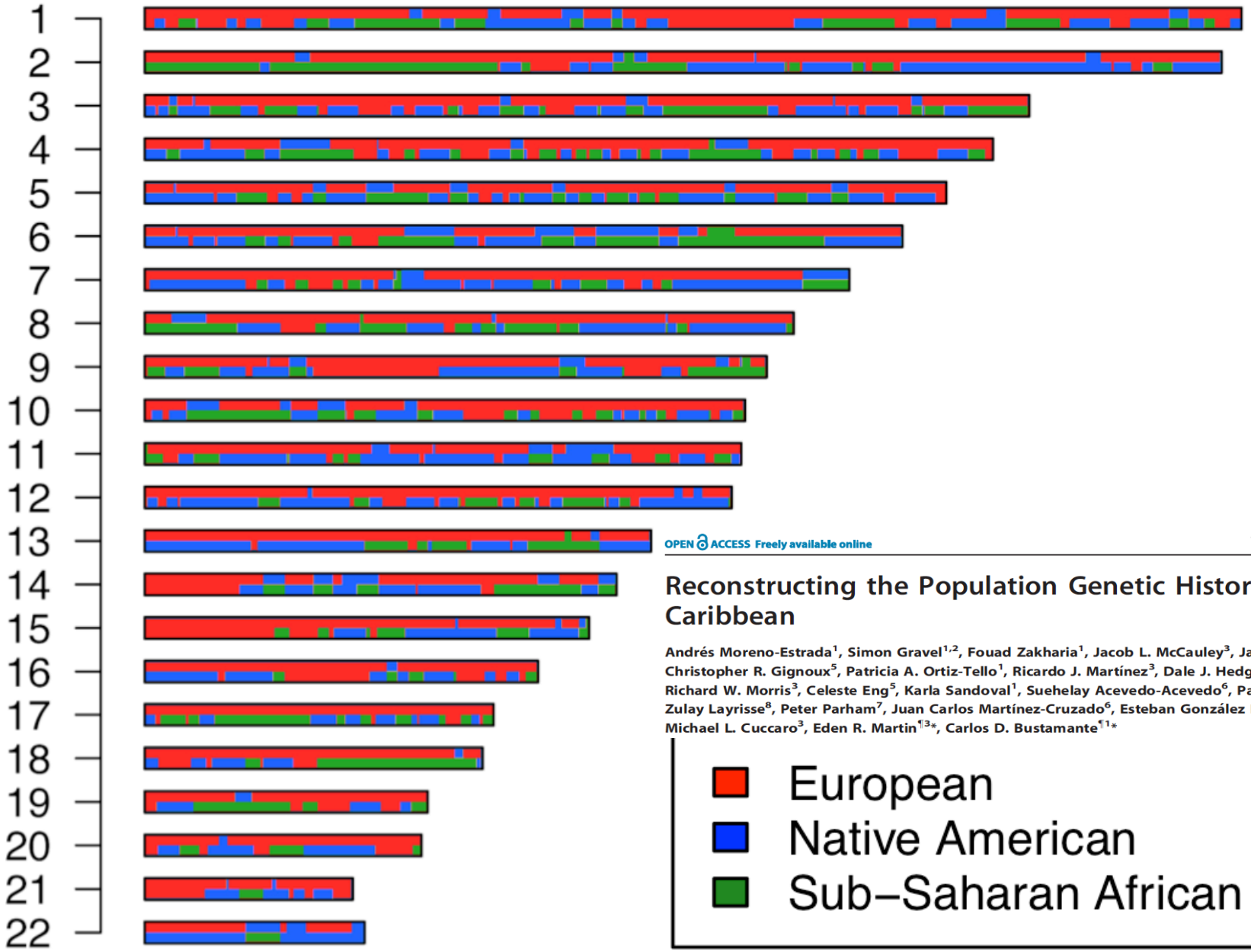
Reference Populations



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

Approaches: Based on Hidden Markov Models

Local ancestry



OPEN ACCESS Freely available online

PLOS | GENETICS

Reconstructing the Population Genetic History of the Caribbean

Andrés Moreno-Estrada[1], Simon Gravel[1,2], Fouad Zakharia[1], Jacob L. McCauley[3], Jake K. Byrnes[1,4], Christopher R. Gignoux[5], Patricia A. Ortiz-Tello[1], Ricardo J. Martínez[3], Dale J. Hedges[3], Richard W. Morris[3], Celeste Eng[5], Karla Sandoval[1], Suehelay Acevedo-Acevedo[6], Paul J. Norman[7], Zulay Layrisse[8], Peter Parham[7], Juan Carlos Martínez-Cruzado[6], Esteban González Burchard[5], Michael L. Cuccaro[3], Eden R. Martin[¶3]*, Carlos D. Bustamante[¶1]*

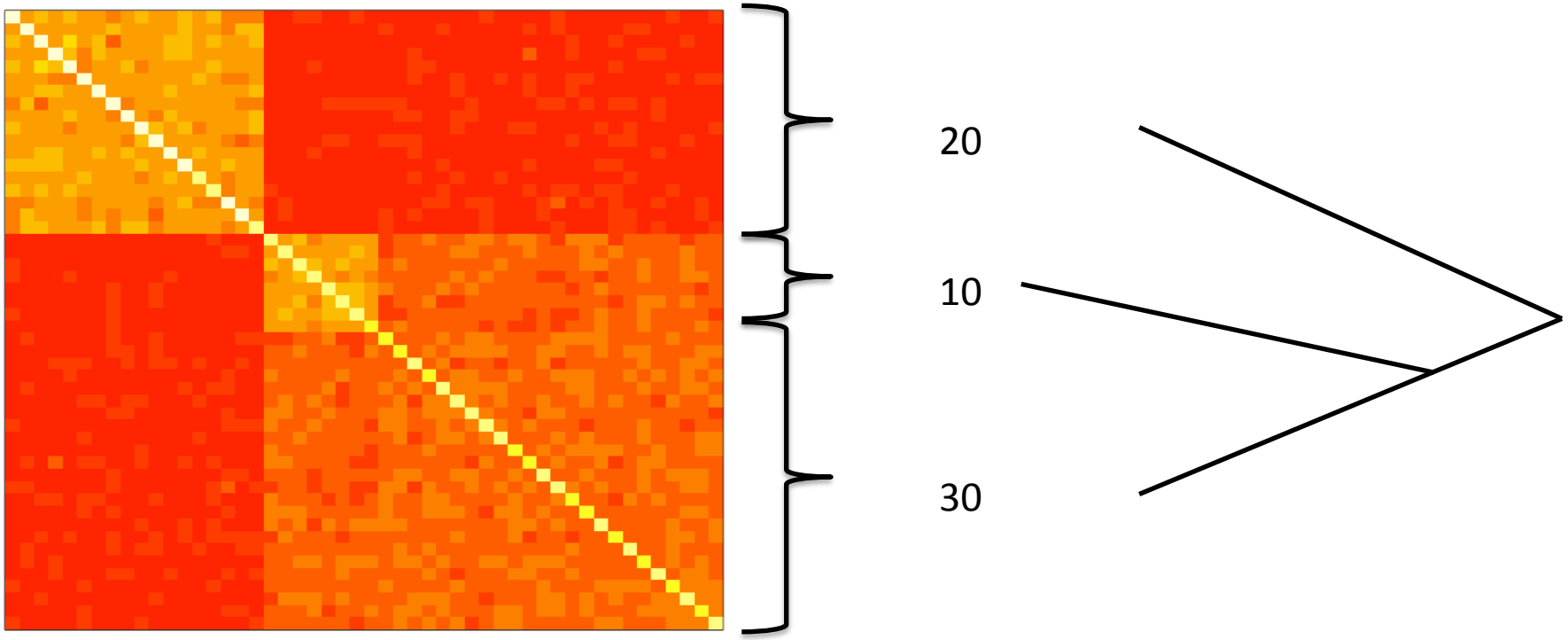
Principal components analysis

Example

Section 1.7.2 of notes

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

Relatedness matrix R



i^{th} and j^{th} entry =average over loci (l) of $(X_{li}-\overline{X_l})(X_{lj}-\overline{X_l})$

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

Modified from slide by Gavin Band

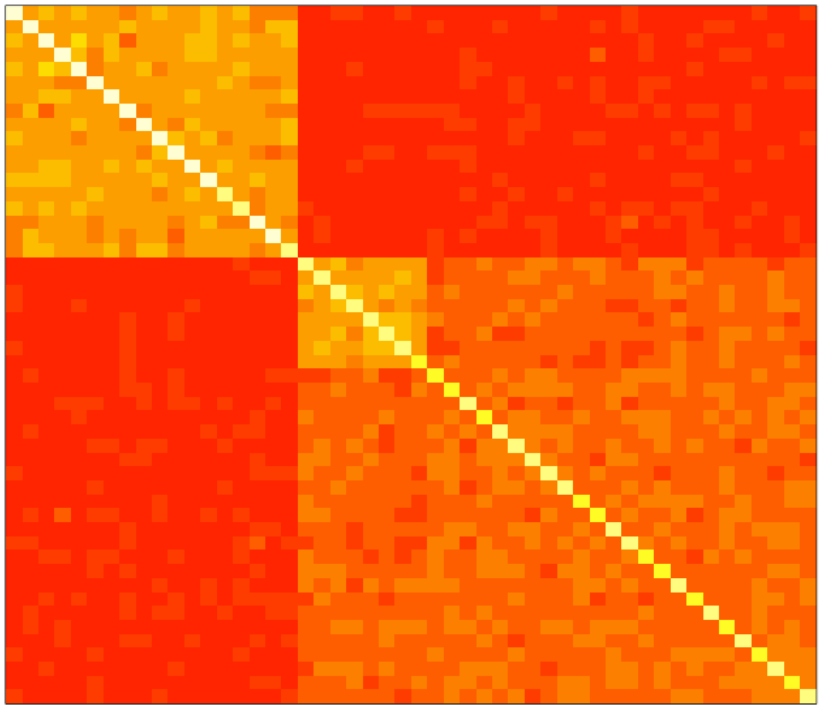
Principal components analysis Example

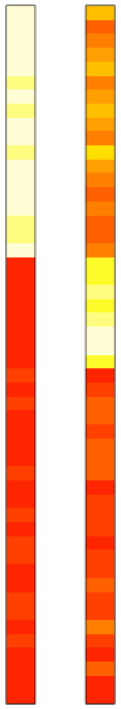
(Simulated data, 50 individuals, 1000 SNPs)

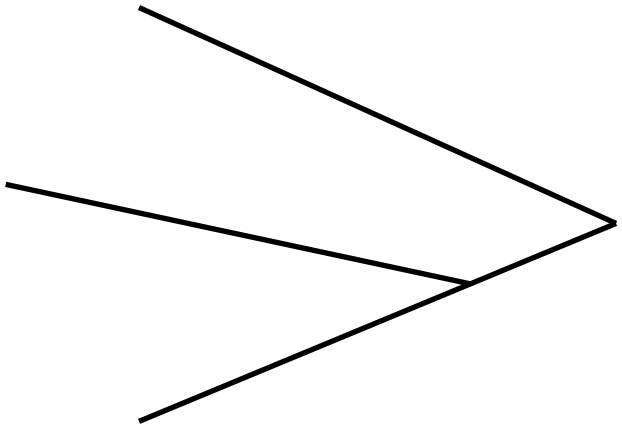
Relatedness matrix R

Principal components
Eigen-vectors)

v_1 v_2







Modified from slide by Gavin Band

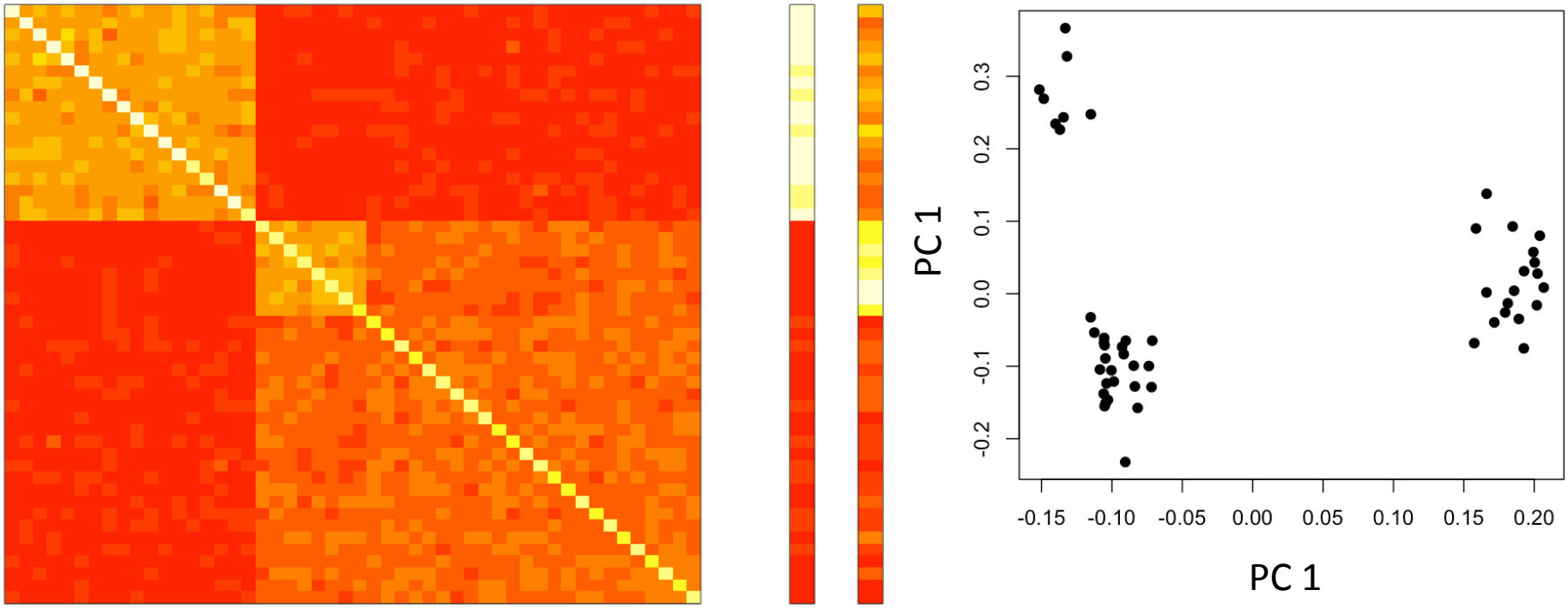
Principal components analysis Example

(Simulated data, 50 individuals, 1000 SNPs)

Relatedness matrix R

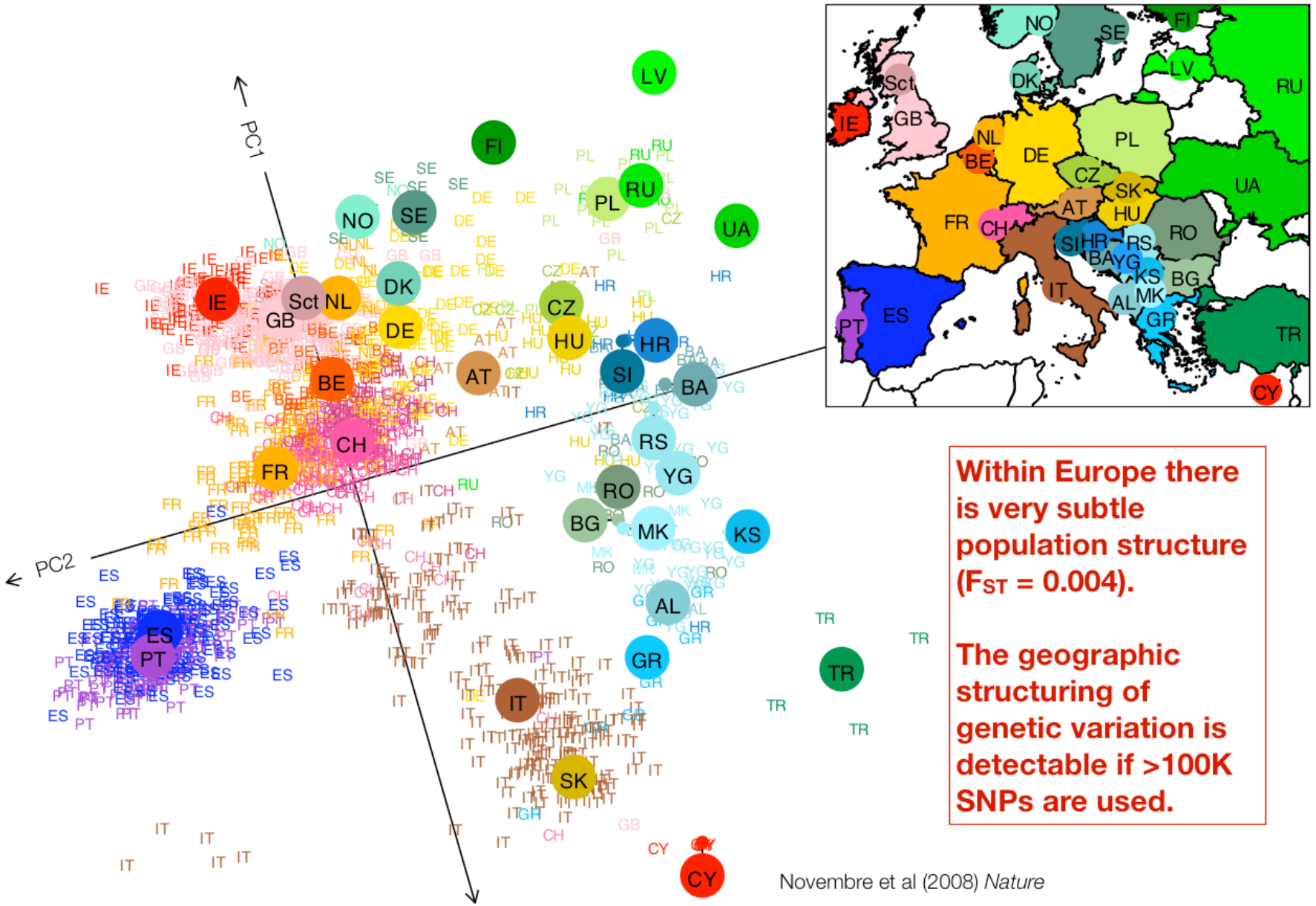
Eigenvectors

v_1 v_2

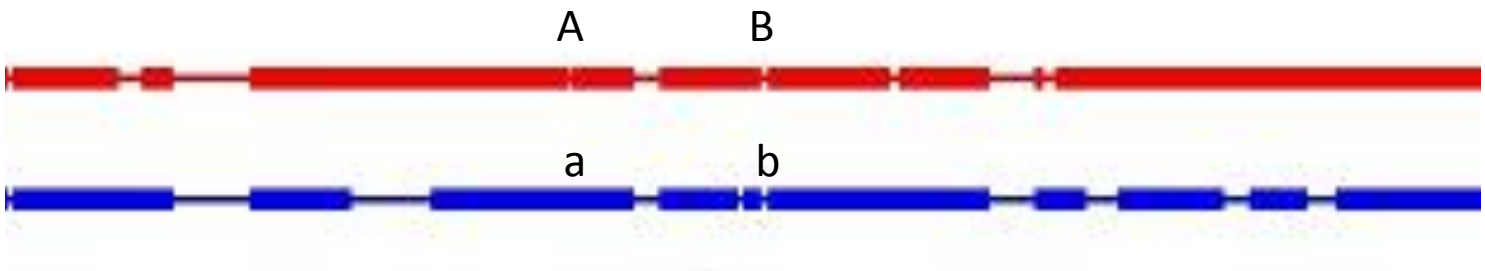


Modified from slide by Gavin Band

Principal Component Analysis of Europeans

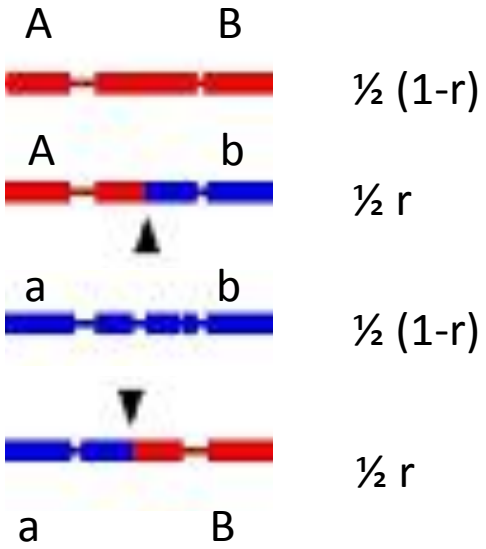


Recombination and Linkage Disequilibrium (LD)

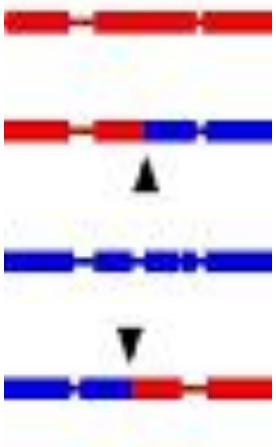


r = recombination fraction
probability of an odd
Number of crossovers occur
Between our markers

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



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



p_{AB} = frequency of AB

p_{ab} = frequency of ab

p_{Ab} = frequency of Ab

p_{aB} = frequency of aB

If independent the expected frequency of gametes (haplotypes)

$p_A X p_B$

$p_a X p_b$

$p_A X p_b$

$p_a X p_B$

Define "D"

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

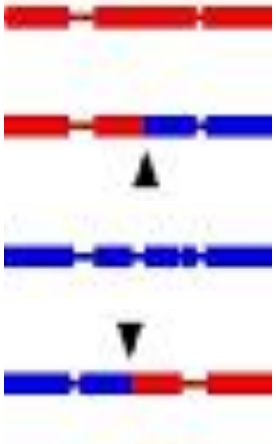
$D_{ab} = p_{ab} - p_a p_b$

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

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

The covariance of A and B.

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



Define "D"

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

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

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

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

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

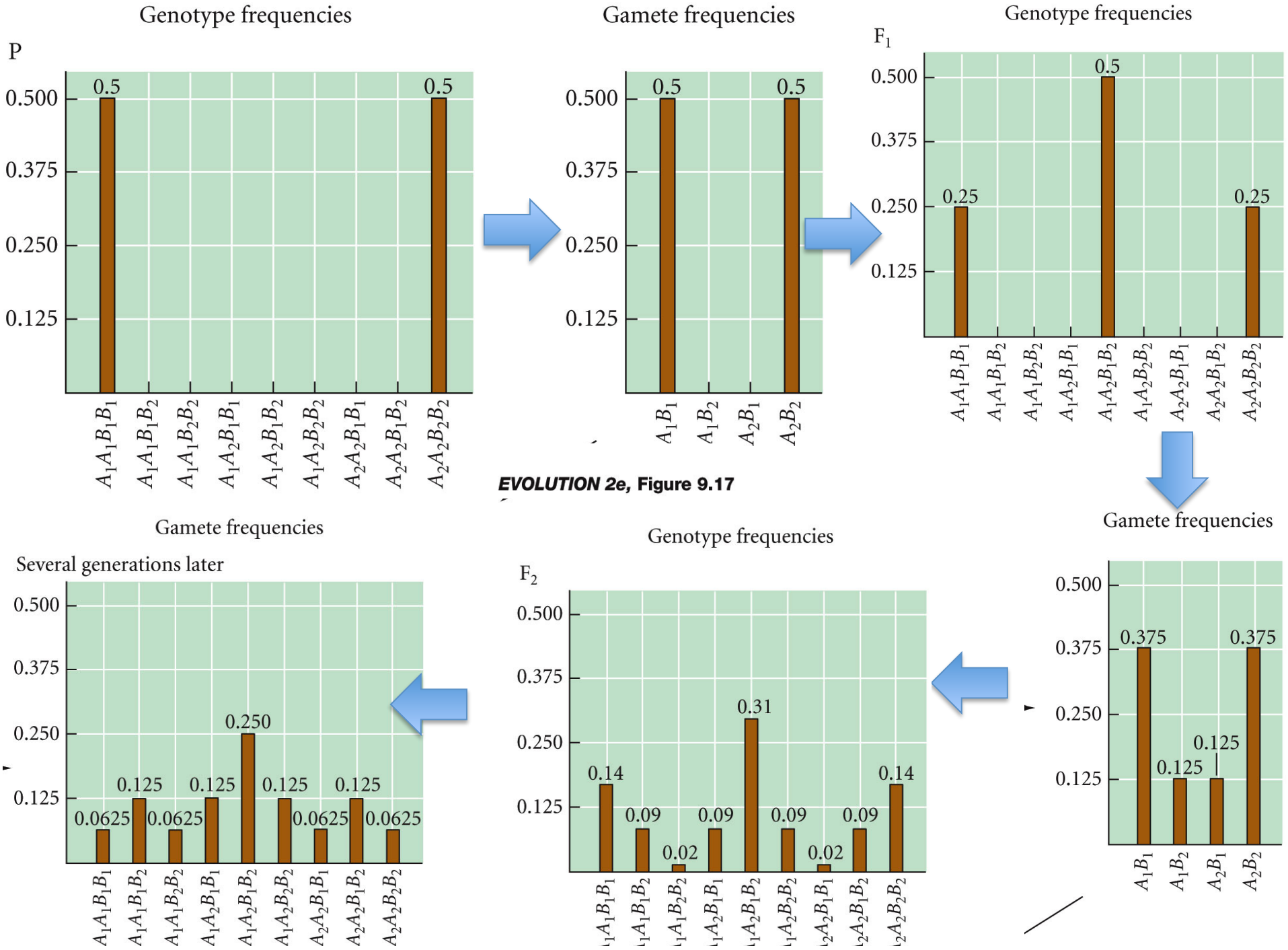
$D_{AB} = D_{ab}$ and $D_{Ab} = D_{aB}$

(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 $p_{AB} = p_A p_B + D$



***EVOLUTION 2e*, Figure 9.17**

Decay of LD in a very large boring randomly mating population

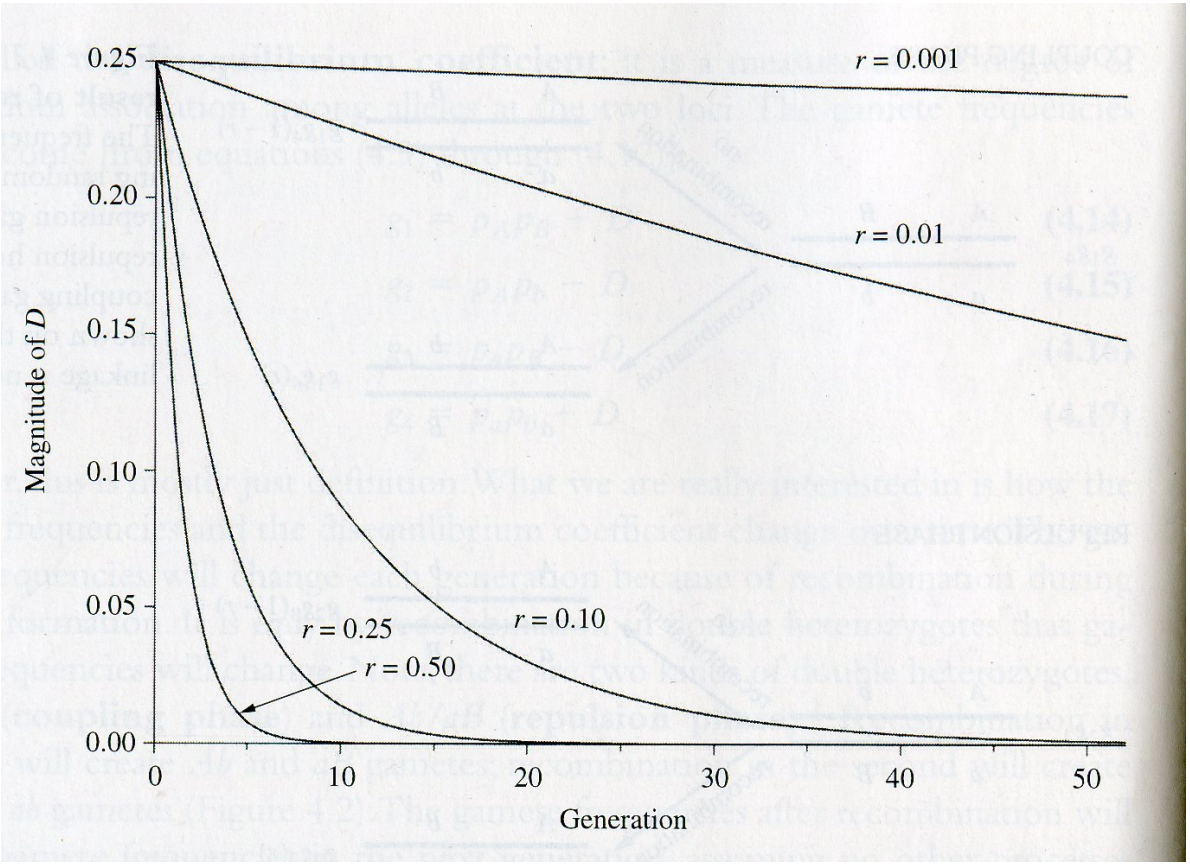
$$D_t = (1- r)^t D_o$$

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

linkage disequilibrium

How does LD change over time due to recombination?

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



Note: more distant markers recombine more!

So eventually recombination leads to *D*=0.

Even with free recombination (r=0.5), it isnt instantaneous