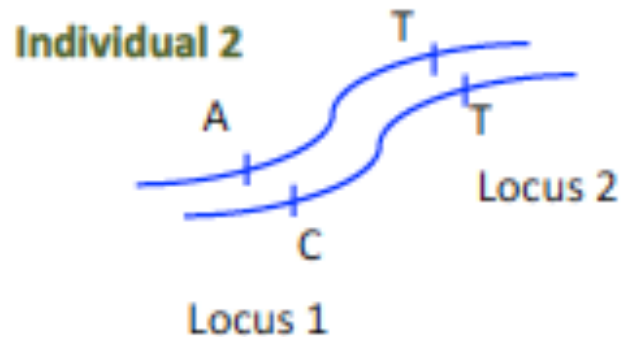
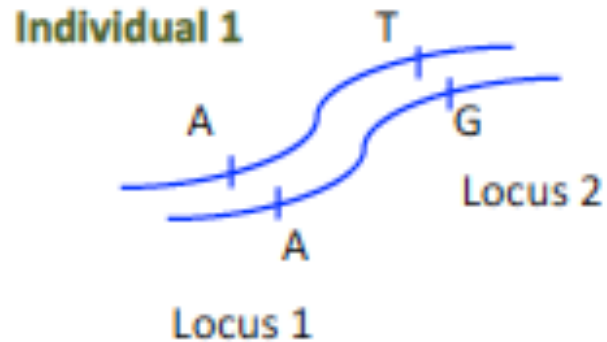


Quantitative Population Genomics

Important concepts for population genetics

- Heritability, Mutations, Genetic Drift, Linkage, and Recombination
- List of words to define:
 - Get in groups of 2-3 and spend 5 minutes seeing if you can define these words.



Allele: variant form of a gene

Can be dominant or recessive

Variance is at a specific genetic locus

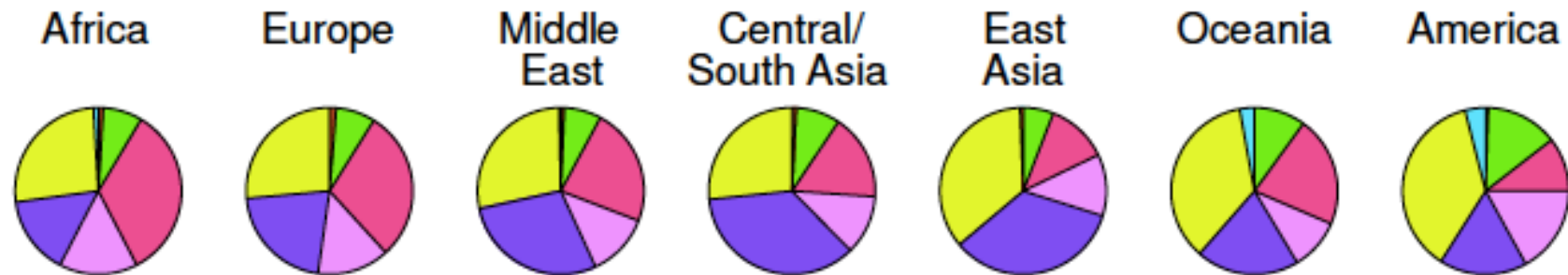
Frequency of observing this allele in a population can be determined

Penetrance is the proportion of individuals carrying a particular allele that also expresses an associated phenotype

GATCTTCGTACTGAGT
GATCTTCGTACTGAGT
GATTTCGTACGGAAT
GATTTCGTACTGAGT
GATCTTCGTACTGAAT
GATTTCGTACGGAAT
GATTTCGTACTGAAT
GATCTTCGTACGGAAT

Binary (biallelic) single nucleotide polymorphisms

D6S474



D12S2070

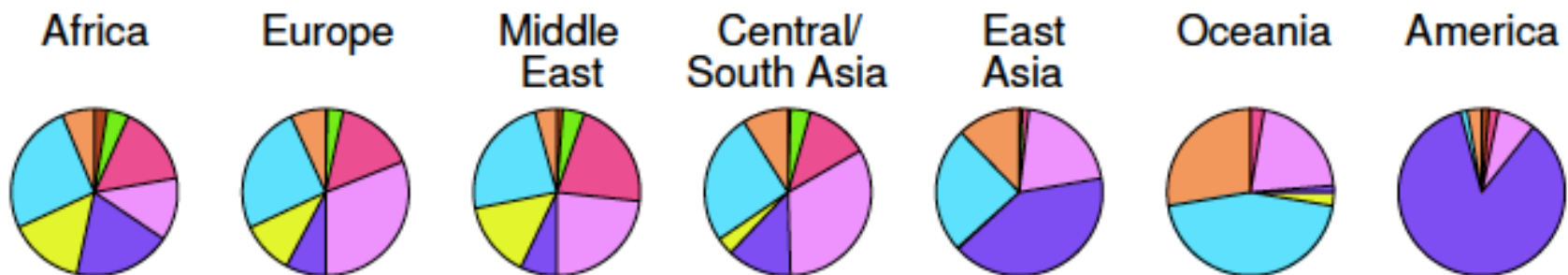


Fig. 13.1. [This figure also appears in the color insert.] Allele frequencies associated with two microsatellite markers in human populations taken from seven different regions. Each allele has a different color code, and the size of the sector in the pie chart indicates allele frequency. Reproduced with permission. Copyright 2002 NA Rosenberg.

Gene	Allele Frequencies			
	Population I		Population II	
	1	2	1	2
<i>A</i>	0.8	0.2	0.2	0.8
<i>B</i>	0.8	0.2	0.2	0.8
<i>C</i>	0.8	0.2	0.2	0.8

Given this population data on these three alleles, how would you go about identifying which population a person most likely belonged to? What do you predict would happen in a person from population I and a person from population II had a child?

GATCTTCGTACTGAGT	CTG	CTG/CTG
GATCTTCGTACTGAGT	CTG	
GATTTCGTACGGAAT	TGA	TGA/TTG
GATTTCGTACTGAGT	TTG	

An individual has 2 haplotypes per set of loci

What are potential haplotypes for a offspring
between these two individuals?

GATCTTCGTACTGAGT	CTG	CTG/CTG
GATCTTCGTACTGAGT	CTG	
GATTTTCGTACGGAAT	TGA	TGA/TTG
GATTTTCGTACTGAGT	TTG	

GATCTTCGTACTGAGT	CTG	CTG/TGA
GATTTTCGTACGGAAT	TGA	

GATCTTCGTACTGAGT	CTG	CTG/TTG
GATTTTCGTACTGAGT	TTG	

Does this always hold?

Linkage

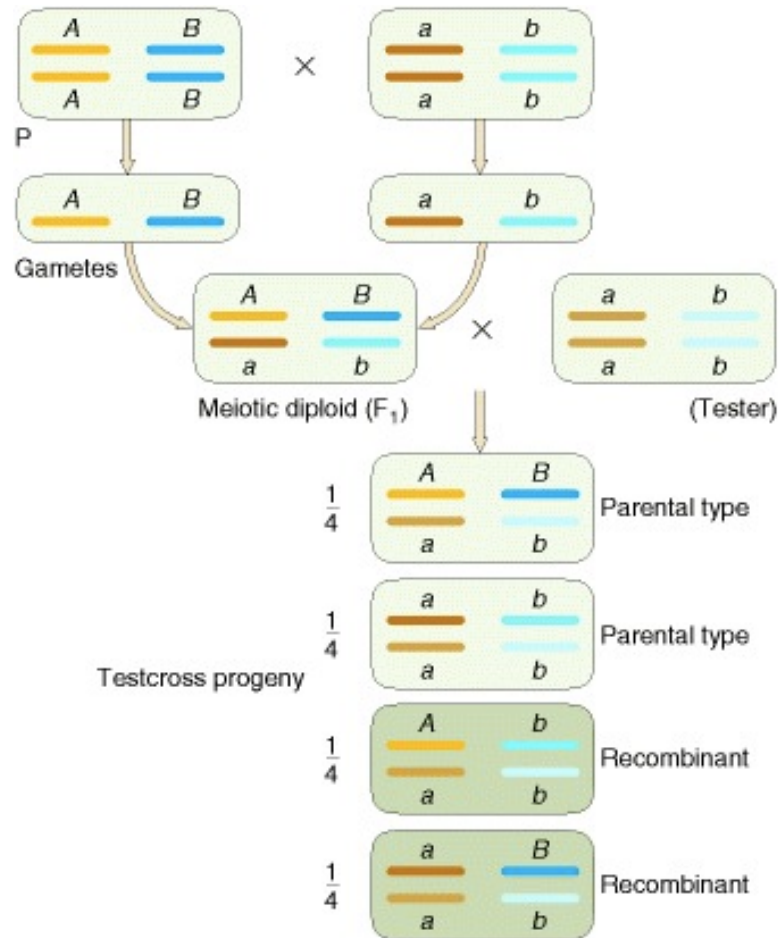
Phenotype	Mendelian expected numbers	Actual results
Purple, long	3911	4831
Purple, round	1303	390
Red, long	1303	393
Red, round	435	1338

Why?

What data are unexplained by this explanation?

How might this occur?

Recombination



Recombination
Frequency:

$RF = \text{recombinants} / \text{total}$

If not linked,
recombination
frequency is 50%

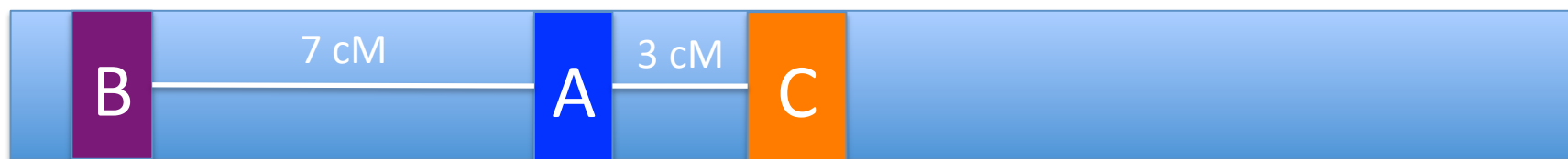
Recombination via Crossovers

	Meiotic chromosomes	Meiotic products	
Meiocytes with no crossover between the genes			Parental
			Parental
			Parental
			Parental
Meiocytes with a crossover between the genes			Parental
			Recombinant
			Recombinant
			Parental

Griffiths AJF, Miller JH, Suzuki DT, et al. An Introduction to Genetic Analysis. 7th edition. New York: W. H. Freeman; 2000.
Recombination. Figure 5-7.

There are 3 genes A B and C. A and B have a recombination frequency of 7%. B and C have a recombination frequency of 10%.

Are genes A and C linked? What are the potential cMs between them?



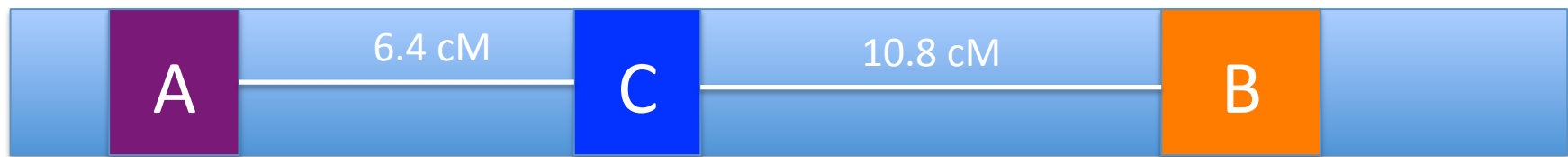
These are the results of a testcross.

In cross 1: you cross female heterozygotes of gene *A*, *B*, and *C* with homozygous recessive males.

You get the following flies:

Genotype	Number of flies
A•B•C	625
a•b•c	620
A•B•c	5
A•b•C	76
A•b•c	45
a•B•C	40
a•B•c	73
a•b•C	7

Draw a map with the appropriate distances between *A*, *B*, and *C*.

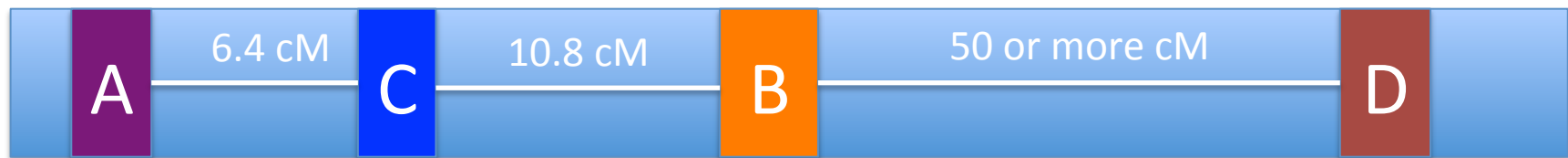


In cross 2: you cross female heterozygotes of gene *A*, *B*, and *D* with homozygous recessive males.

You get the following flies:

Genotype	Number of flies
A•B•D	527
a•b•d	528
A•B•d	530
a•b•D	525
A•b•D	97
A•b•d	99
a•B•D	98
a•B•d	99

Add gene *D* to your map from the 1st cross.



OR



- The closer two alleles are on chromosome, the less likely they are to assort independently
- When they don't assort independently they are out of equilibrium
- This lack of equilibrium is known as **Linkage Disequilibrium (LD)**
- This LD is decreased by recombination event

Probability of being linked

- 2 Haplotypes: A_1B_1 and A_2B_2
- Probability of each haplotype is $p_{A_1B_1}$ and $p_{A_2B_2}$
- Allele frequencies are:

$$p_{A_1}, p_{B_1}, p_{A_2}, p_{B_2}$$

- These probabilities need to be calculated from population genetic data
- How would we determine linkage?
 - Difference in probability of haplotype from the probability of getting those two alleles separately

Linkage Disequilibrium Value

D is the amount of disequilibrium

$$D = p_{A_1B_1} - p_{A_1} * p_{B_1} = p_{A_2B_2} - p_{A_2} * p_{B_2}$$

$$-D = p_{A_2B_1} - p_{A_2} * p_{B_1} = p_{A_1B_2} - p_{A_1} * p_{B_2}$$

$$D = (p_{A_1B_1} * p_{A_2B_2}) - (p_{A_1B_2} * p_{A_2B_1})$$

Ideally compare linkages – normalize
to expected values

$$D' = D / D_{\max}$$

$$D_{\max} = \min\{p_{A_2}p_{B_1}, p_{A_1}p_{B_2}\}, \text{ if } D > 0,$$

$$D_{\max} = \min\{p_{A_1}p_{B_1}, p_{A_2}p_{B_2}\}, \text{ if } D < 0.$$

If complete linkage $|D'| = 1$

SNP number				
1	2	3	4	5
--A-----	T--G--T-----			T-----
--T-----	T--G--T-----			T-----
--A-----	A--G--T-----			A-----
--T-----	T--G--T-----			A-----
--A-----	T--G--T-----			A-----
--T-----	T--C--A-----			T-----
--A-----	A--C--A-----			A-----
--T-----	A--C--A-----			A-----
--A-----	T--C--A-----			T-----
--T-----	A--C--A-----			A-----

For 1,2 and 2,3:
Determine $|D'|$

SNP number				
1	2	3	4	5
--A-----	T--G--T-----			T-----
--T-----	T--G--T-----			T-----
--A-----	A--G--T-----			A-----
--T-----	T--G--T-----			A-----
--A-----	T--G--T-----			A-----
--T-----	T--C--A-----			T-----
--A-----	A--C--A-----			A-----
--T-----	A--C--A-----			A-----
--A-----	T--C--A-----			T-----
--T-----	A--C--A-----			A-----

[1,2] $p_{AT}=0.3$, $p_{TA}=0.2$, $p_{AA}=0.2$, $p_{TT}=0.3$

$p_{1(A)}=0.5$, $p_{1(T)}=0.5$, $p_{2(T)}=0.6$, $p_{2(A)}=0.4$

$D=0.00$, $D_{\max}=0.20$ $|D'|=0.00$

SNP number				
1	2	3	4	5
--A-----	T--G--T-----			T-----
--T-----	T--G--T-----			T-----
--A-----	A--G--T-----			A-----
--T-----	T--G--T-----			A-----
--A-----	T--G--T-----			A-----
--T-----	T--C--A-----			T-----
--A-----	A--C--A-----			A-----
--T-----	A--C--A-----			A-----
--A-----	T--C--A-----			T-----
--T-----	A--C--A-----			A-----

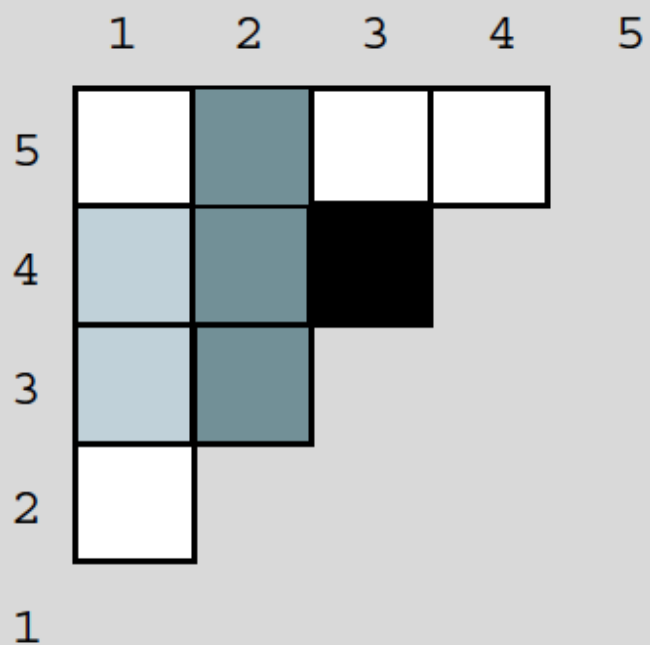
$[2,3] \ p_{TG}=0.4, p_{AC}=0.3, p_{TC}=0.2, p_{AG}=0.1$

$p_{2(T)}=0.6, p_{2(A)}=0.4, p_{3(G)}=0.5, p_{3(C)}=0.5,$

$D=0.1, D_{\max}=0.20 \ |D'|=0.5$

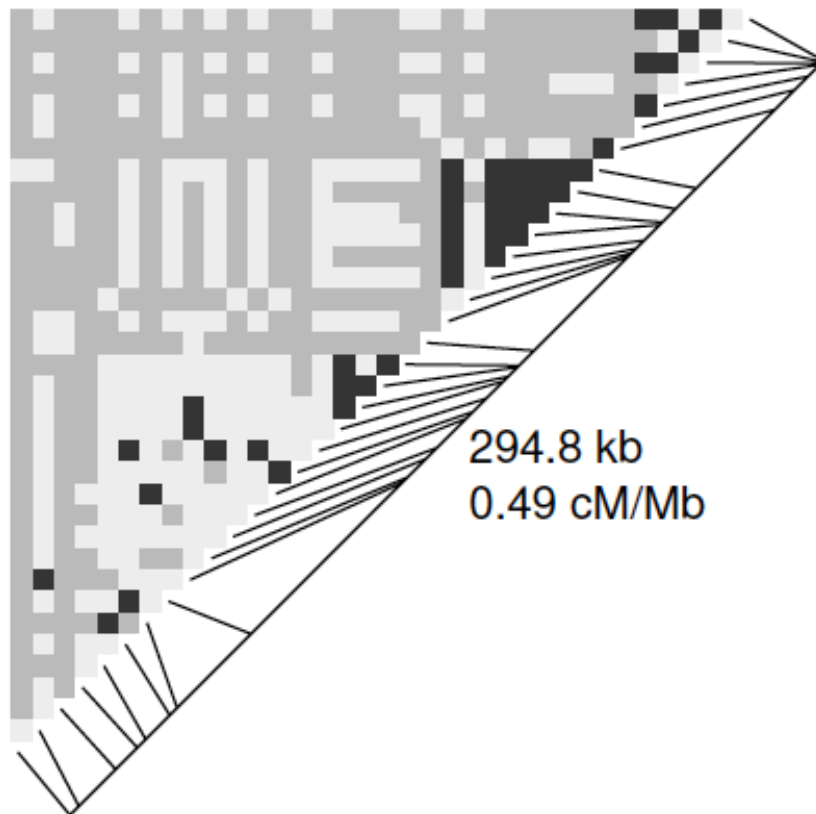
SNP number				
1	2	3	4	5
--A-----	T--	G--	T-----	T-----
--T-----	T--	G--	T-----	T-----
--A-----	A--	G--	T-----	A-----
--T-----	T--	G--	T-----	A-----
--A-----	T--	G--	T-----	A-----
--T-----	T--	C--	A-----	T-----
--A-----	A--	C--	A-----	A-----
--T-----	A--	C--	A-----	A-----
--A-----	T--	C--	A-----	T-----
--T-----	A--	C--	A-----	A-----

$[2,5] \ p_{TA}=0.2, \ p_{AT}=0.0, \ p_{TT}=0.4, \ p_{AA}=0.4,$
 $p_{2(T)}=0.4, \ p_{2(A)}=0.6, \ p_{5(A)}=0.6, \ p_{5(T)}=0.4,$
 $D=-0.16, \ D_{\max}=0.24 \ |D'|=.67$

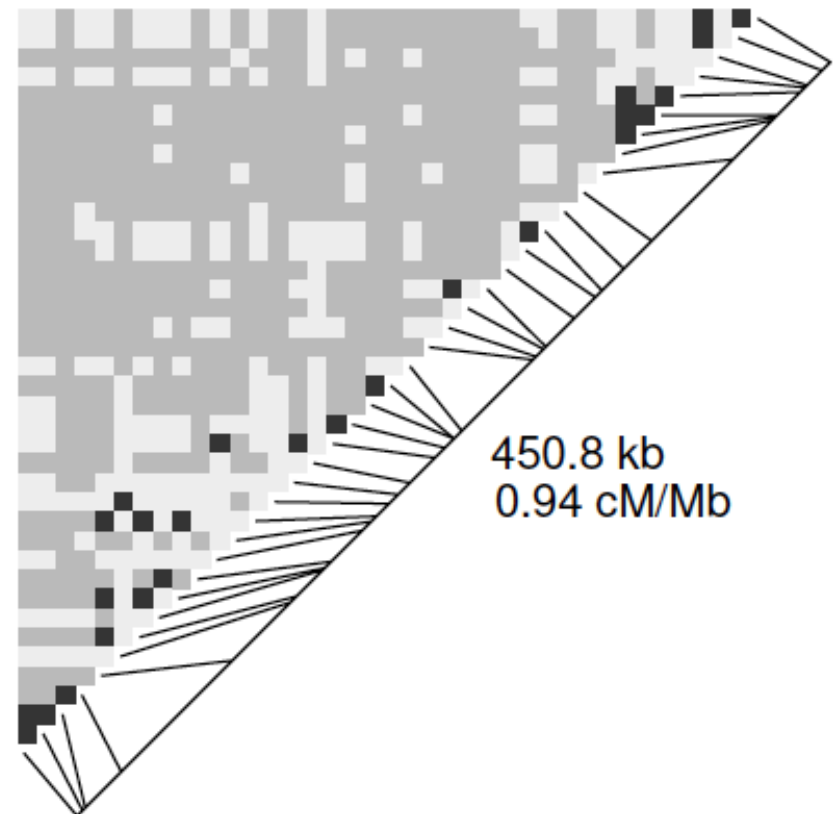


White: $|D'| \leq 0.2$
 Light Grey: $0.2 < |D'| \leq 0.5$
 Dark Grey: $0.5 < |D'| \leq 0.8$
 Black: $0.8 < |D'| \leq 1.0$

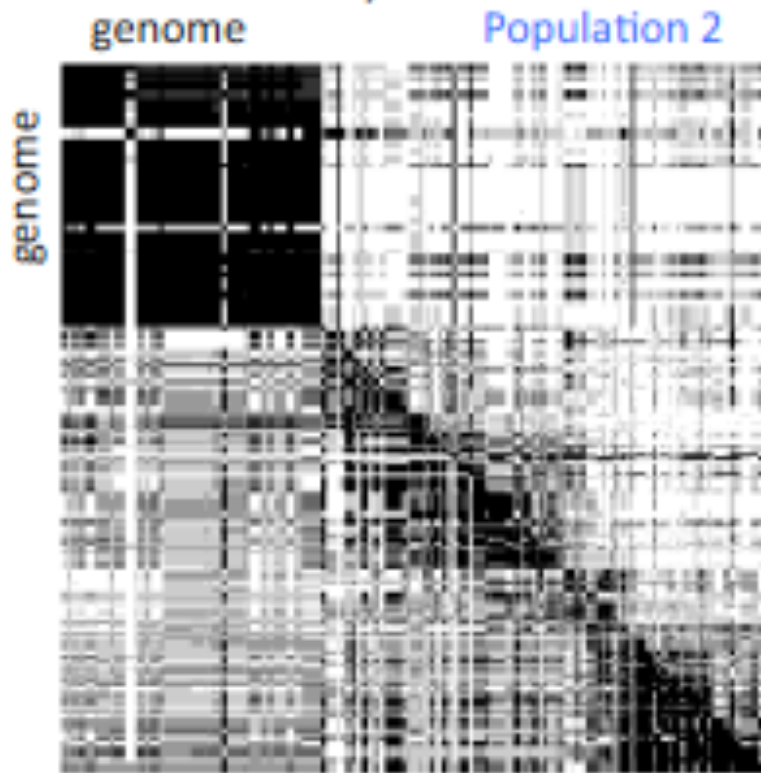
A. Sub-Saharan African sample,
region 19a



B. Sub-Saharan African sample,
region 32a

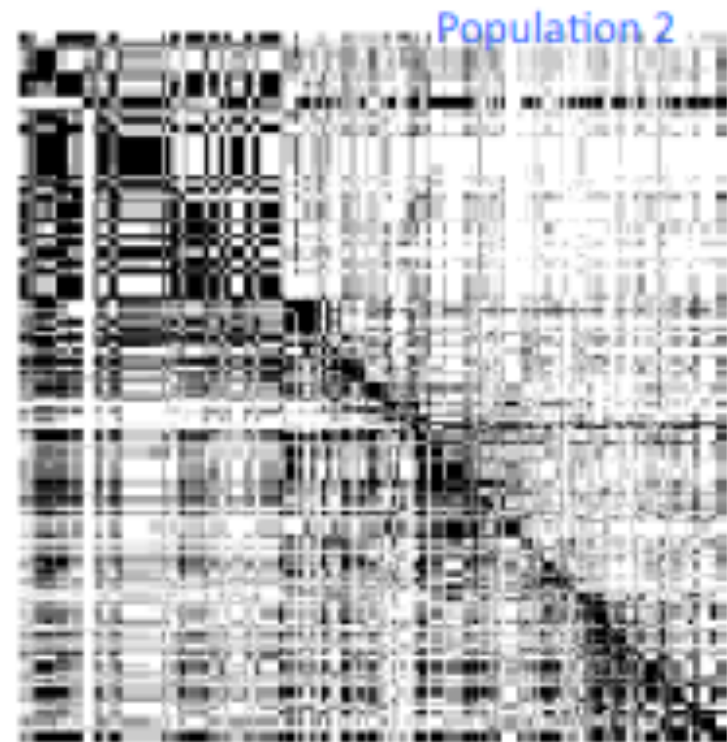


- r^2 in SNP data from a population of individuals (Black: $r^2=1$, white: $r^2=0$)



Population 1

Reproduced from Seyoung Kim



Population 1

$$r^2 = \frac{D^2}{p_{A1}p_{A2}p_{B1}p_{B2}}$$

Linkage isn't equal across the genome

Determining recombination rates

$$p'_{A_1B_1} = (1-r)p_{A_1B_1} + rp_{A_1} * p_{B_1}$$

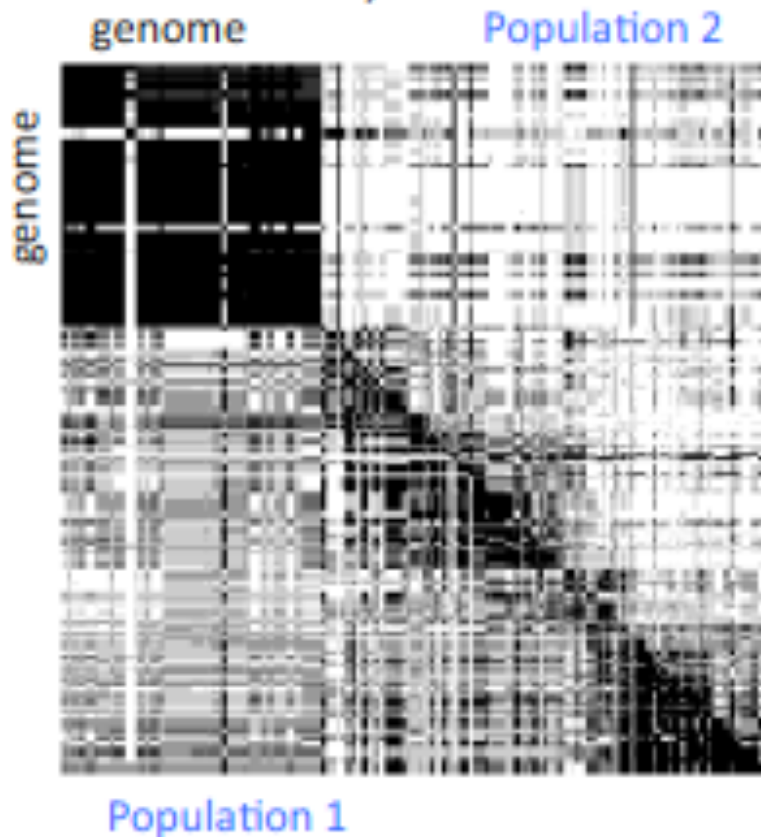
$$D'' = (1-r)D$$

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

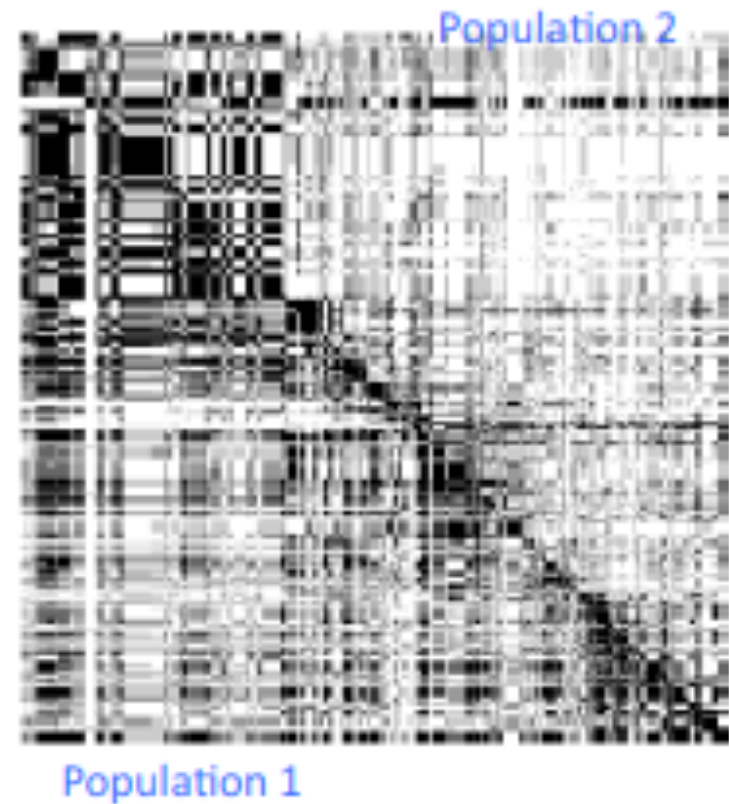
$$0 \leq r \leq 0.5$$

For genes that are close (1% recombination), ~69 generations or 1400 years to come to equilibrium

- r^2 in SNP data from a population of individuals (Black: $r^2=1$, white: $r^2=0$)



Reproduced from Seyoung Kim



Recombination isn't equally likely across the genome

How does evolution with a population
affect genotype and haplotype?

D6S474

Africa



Europe



Middle
East



Central/
South Asia



East Asia



Oceania



America



D12S2070

Africa



Europe



Middle
East



Central/
South Asia



East Asia



Oceania

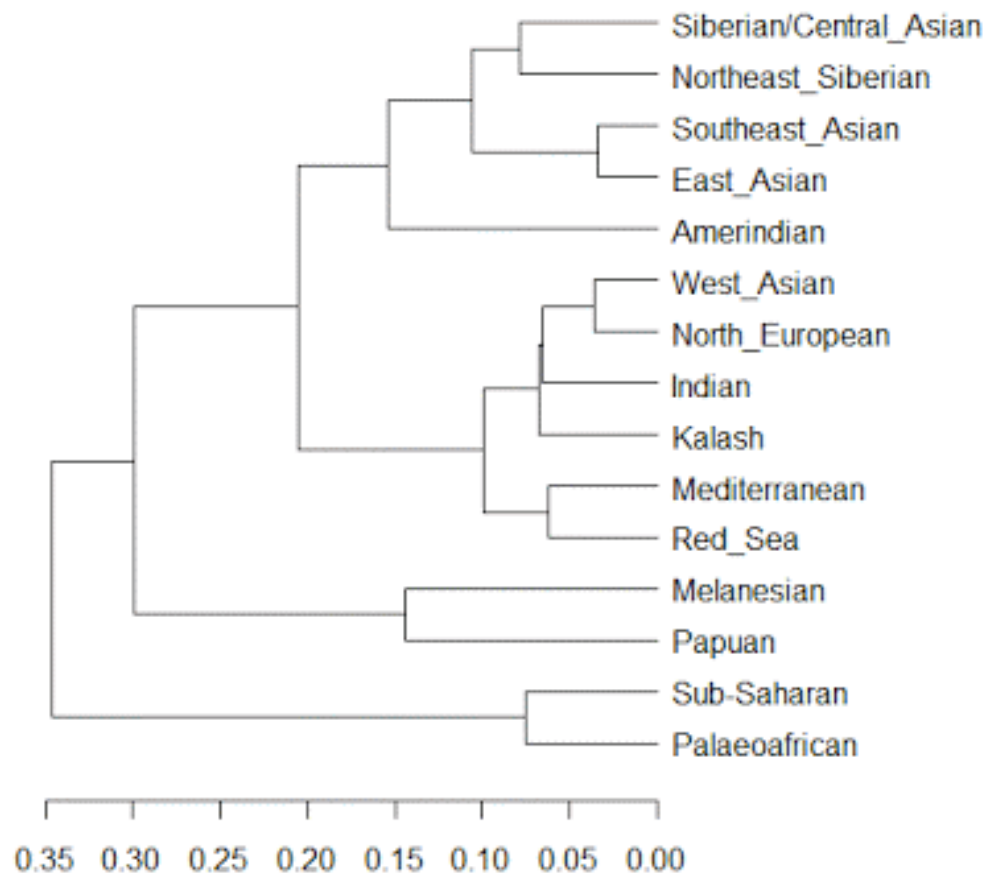


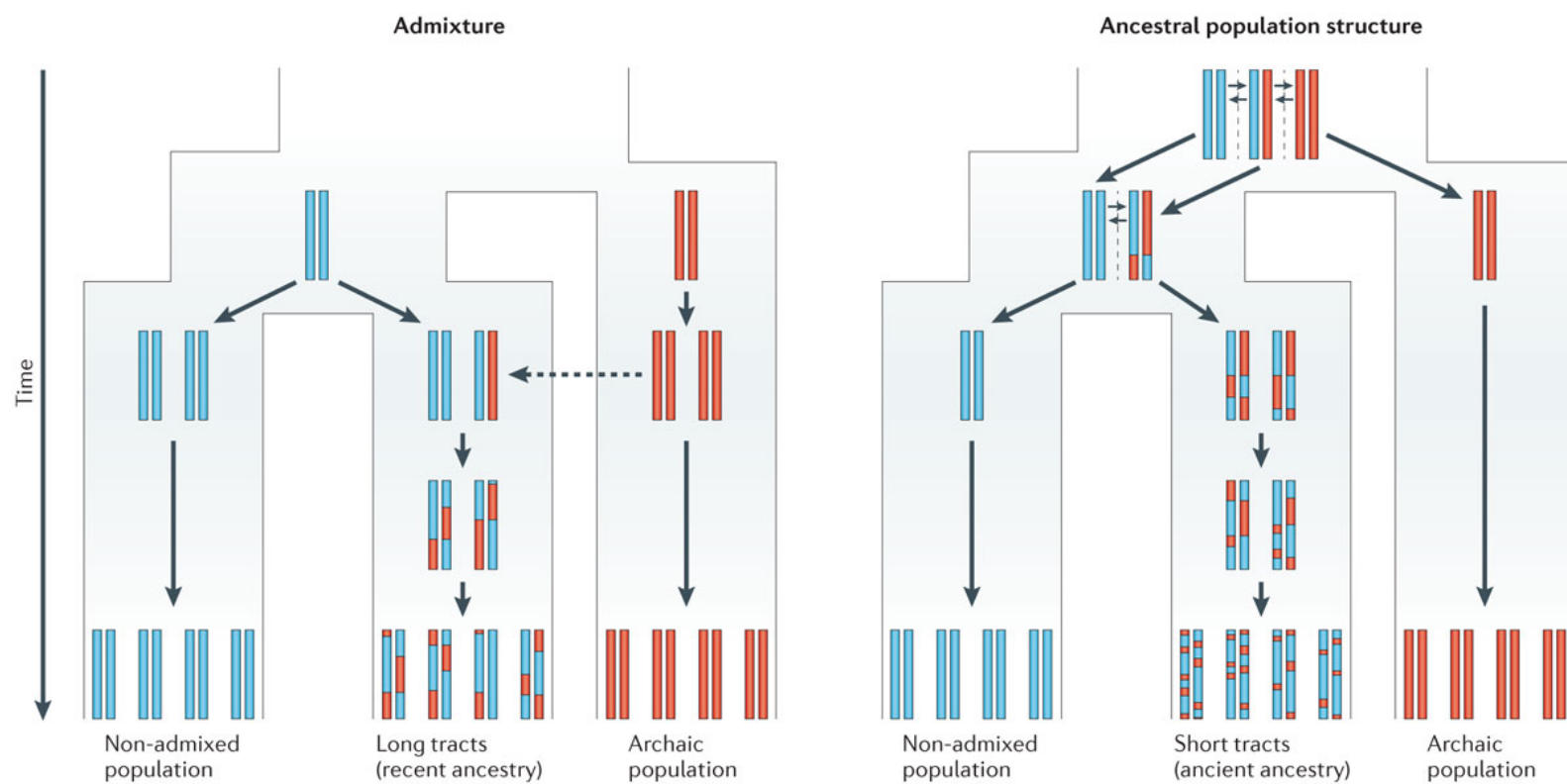
America



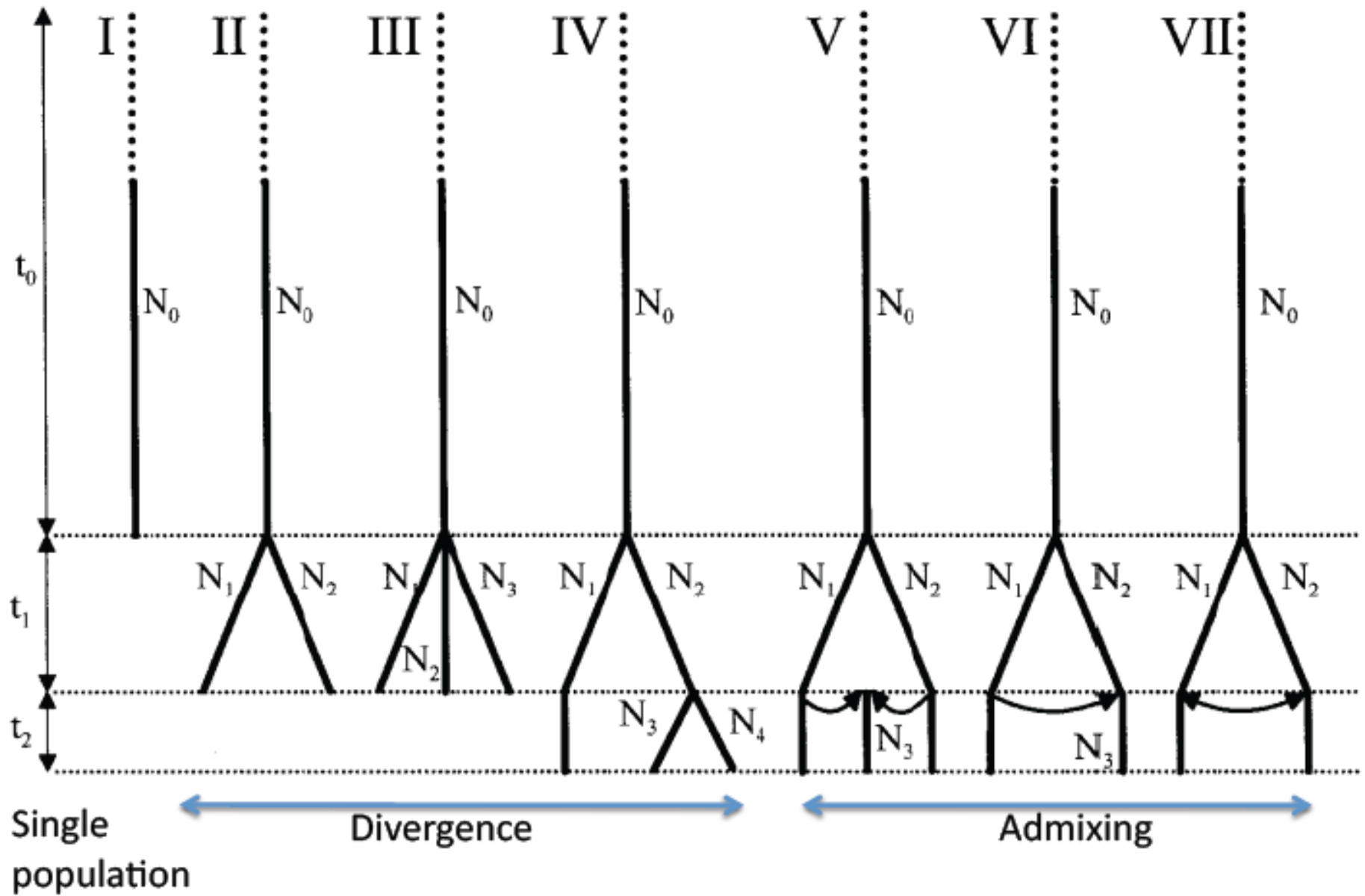
Genographic Project / The Human Journey: Migration Routes







Nature Reviews | **Genetics**



All Genomes are a mixture of ancestral genomes



Summarization of Class