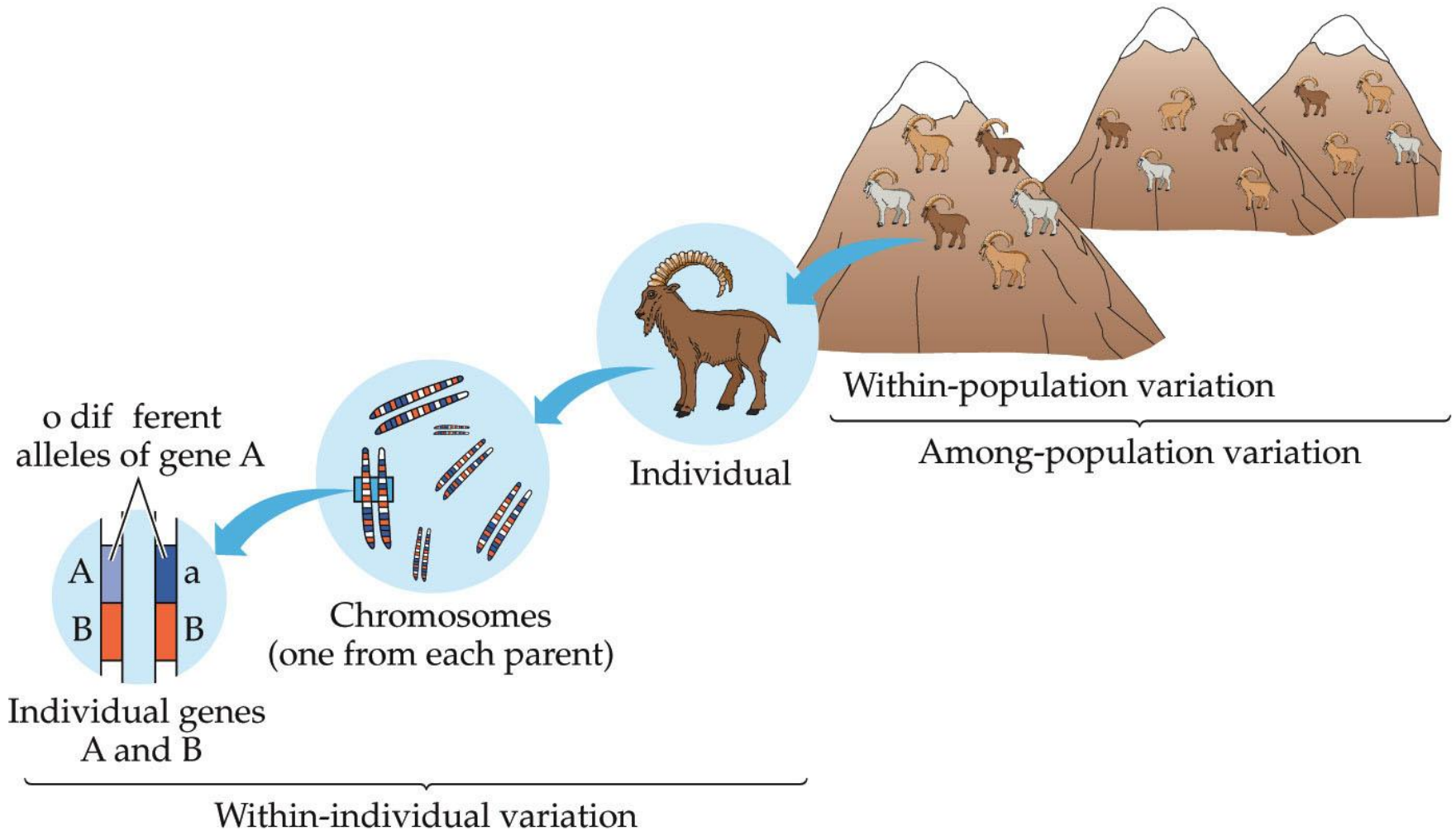


# Primer in Population Genetics

# Hierarchical Organization of Genetics Diversity



# Defining Genetic Diversity within Populations

- Polymorphism – number of loci with > 1 allele
- Number of alleles at a given locus
- Heterozygosity at a given locus
- Theta or  $\theta = 4N_e\mu$  (for diploid genes)

where

$N_e$  = effective population size

$\mu$  = per generation mutation rate

# Defining Genetic Diversity

## Among Populations

- Genetic diversity among populations occurs if there are differences in allele and genotype frequencies between those populations.
- Can be measured using several different metrics, that are all based on allele frequencies in populations.
  - $F_{st}$  and analogues
  - Genetic distance, e.g., Nei's  $D$
  - Sequence divergence

# Estimating Observed Genotype and Allele Frequencies

- Suppose we genotyped 100 diploid individuals ( $n = 200$  gene copies)....

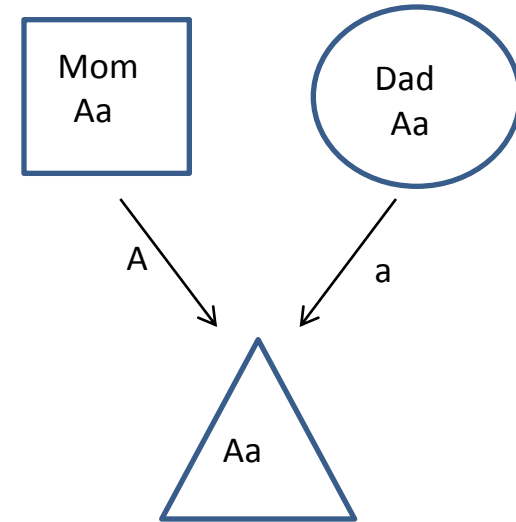
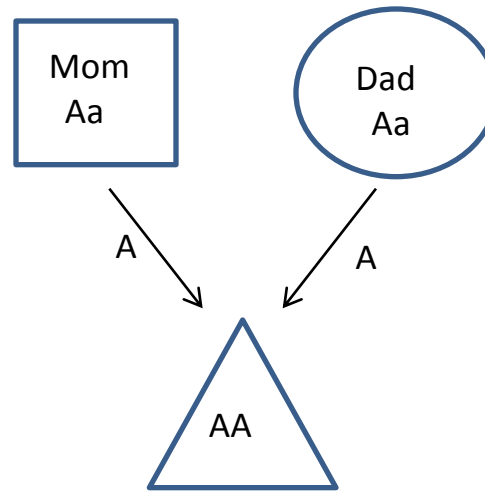
	<u>Genotypes</u>		
	<u>AA</u>	<u>Aa</u>	<u>aa</u>
Number	58	40	2
Genotype Frequency	0.58	0.40	0.02

	# obs. for genotype			
<u>Allele</u>	<u>AA</u>	<u>Aa</u>	<u>aa</u>	<u>Observed Allele Frequencies</u>
A	116	40	0	$156/200 = \mathbf{0.78 (p)}$
a	0	40	4	$44/200 = \mathbf{0.22 (q)}$

# Estimating Expected Genotype Frequencies

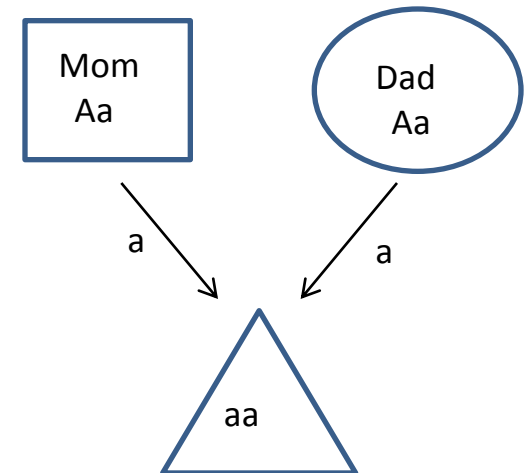
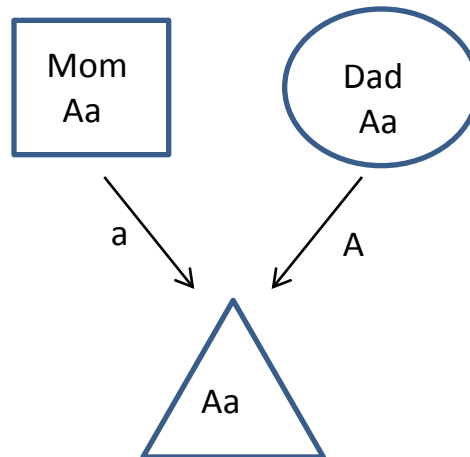
## Mendelian Inheritance

- Offspring inherit one chromosome and thus one allele **independently** and **randomly** from each parent



- Mom and dad both have genotype Aa, their offspring have three possible genotypes:

AA    Aa    aa



# Estimating Expected Genotype Frequencies

- Much of population genetics involves manipulations of equations that have a base in either probability theory or combination theory.

- Rule 1:** If you account for all possible events, the probabilities sum to 1. [e.g.,  $p + q = 1$  for a two-allele system].

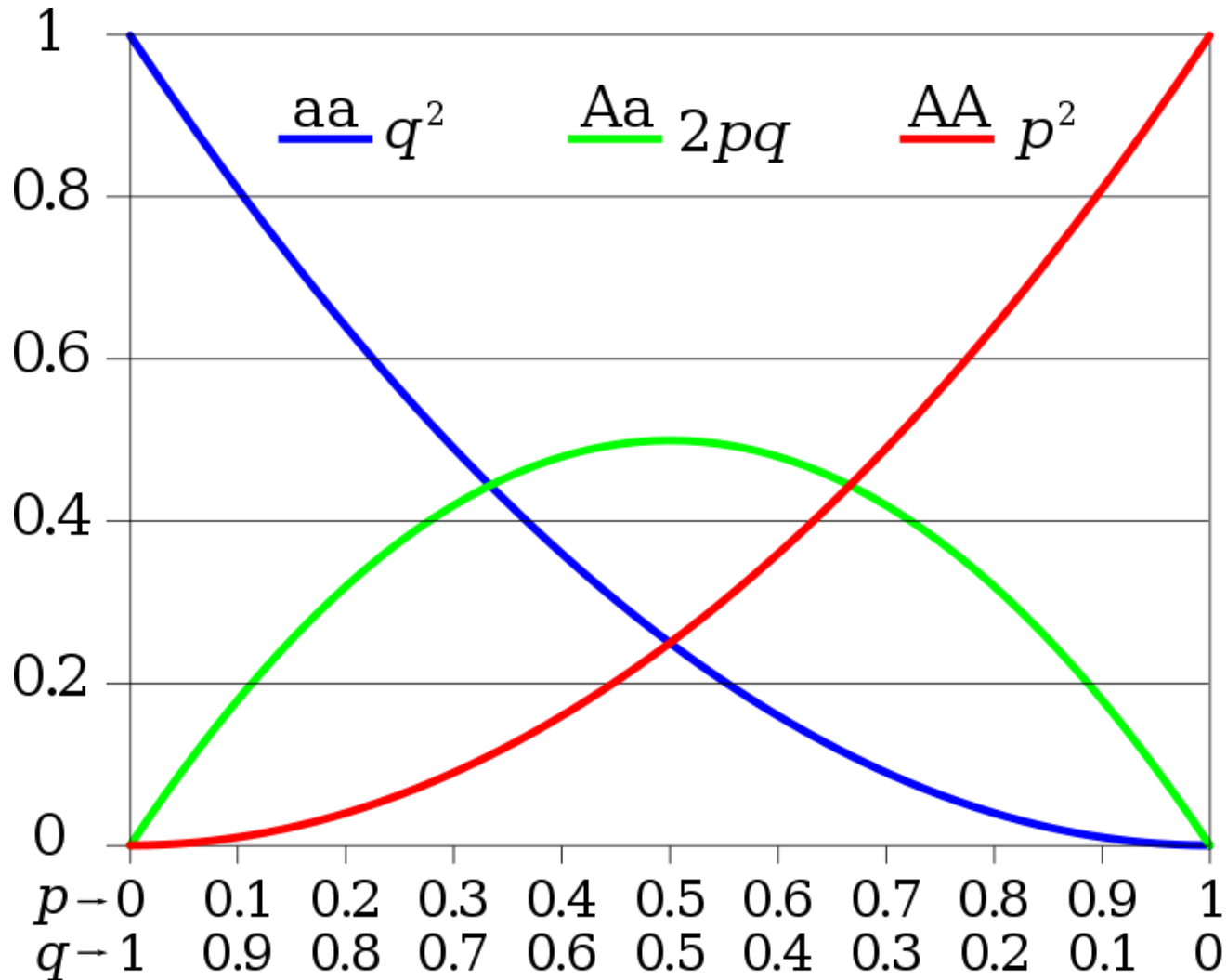
- Rule 2:** The probability that two independent events occur is the **product** of their individual probabilities.

[e.g., probability of a homozygote with **aa** genotype is  $q * q = q^2$ ].

- Thus, under “ideal” conditions, expected genotype frequencies are  $p^2$  for AA,  $2pq$  for Aa, and  $q^2$  for aa; and the genotype frequencies sum to 1 such that:

$$p^2 + 2pq + q^2 = 1 \text{ (Hardy Weinberg Equilibrium)}$$

# Expected Genotype Frequencies under HWE





# Testing for Deviations from HWE

	Genotypes			Total	
	AA	Aa	aa		
Observed numbers (O)	16	20	4	40	$p = 52/80 = 0.65$ $q = 28/80 = 0.35$
Expected frequencies	$p^2$	$2pq$	$p^2$	1.0	
	$0.65^2$	$2*0.65*0.35$	$0.35^2$	1.0	
	0.42	0.46	0.12	1.0	
Expected numbers (E) (expected frequency * 40)	17	18	5	40	

# Testing for Deviation from HWE with a Chi-square test

$$\chi^2 = \sum (O - E)^2/E$$

$$\chi^2 = (16-17)^2/17 + (20-18)^2/18 + (4-5)^2/5$$

$$\chi^2 = 0.4$$

- Probability of obtaining a  $\chi^2$  of 7.2 or greater (1 df) = 0.53
- Thus, observed genotypes do not deviate from HWE

# Estimating Expected Heterozygosity when >2 Alleles

- When there are more than 2 alleles, a simple way to calculate HW expected heterozygosity is:

$$H_e = 1 - \sum p_i^2$$

- For example, if allele frequencies for 3 alleles are 0.5, 0.3, and 0.2, HW expected heterozygosity is:

$$\begin{aligned} H_e &= 1 - (0.5^2 + 0.3^2 + 0.2^2) = 1 - (0.25 + 0.09 + 0.04) \\ &= 0.62 \end{aligned}$$

# Estimating Genomic Diversity

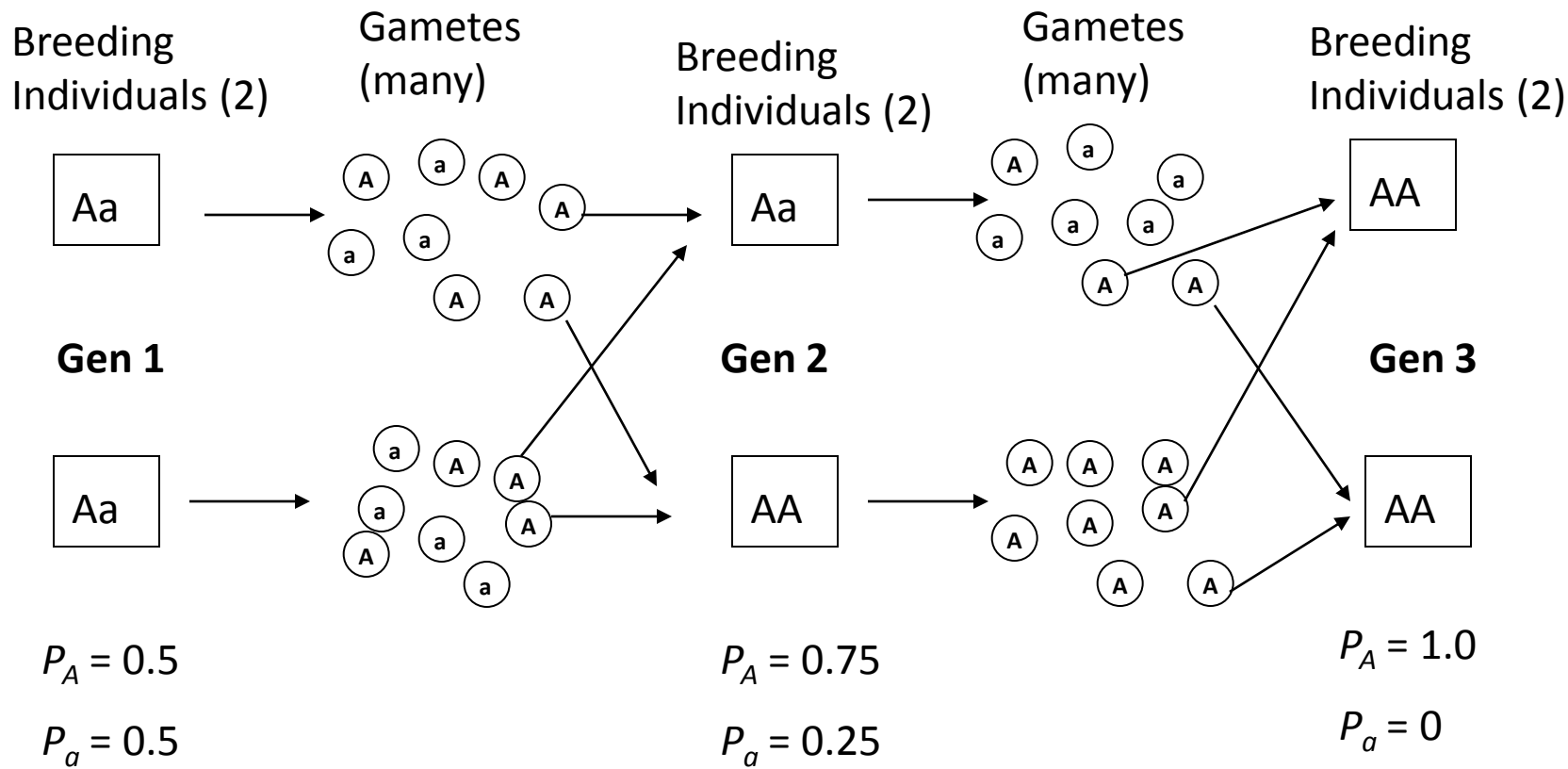
- To fully assess the demographic history and evolutionary potential of species, genome-wide assessment of genetic diversity is needed (mammals have ~35,000 loci).
- Genetic diversity measures are estimated over several loci that are presumed to be a random sample of the genome
- Heterozygosity is often averaged over multiple loci to obtain an estimate of genome-wide genetic diversity

# Evolutionary Processes that Influence Genetic Diversity

- **Genetic Drift** – random change in allele frequencies in a population from generation to generation due to finite population size.
- **Mutation** – an error in the replication of DNA that causes a structural change in a gene. **Only source of new genetic variation in populations (sex cells only).**
- **Gene Flow** – exchange of genetic information among population via migration of individuals.
- **Natural Selection** – differential contribution of genotypes to the next generation due to differences in survival and reproduction.

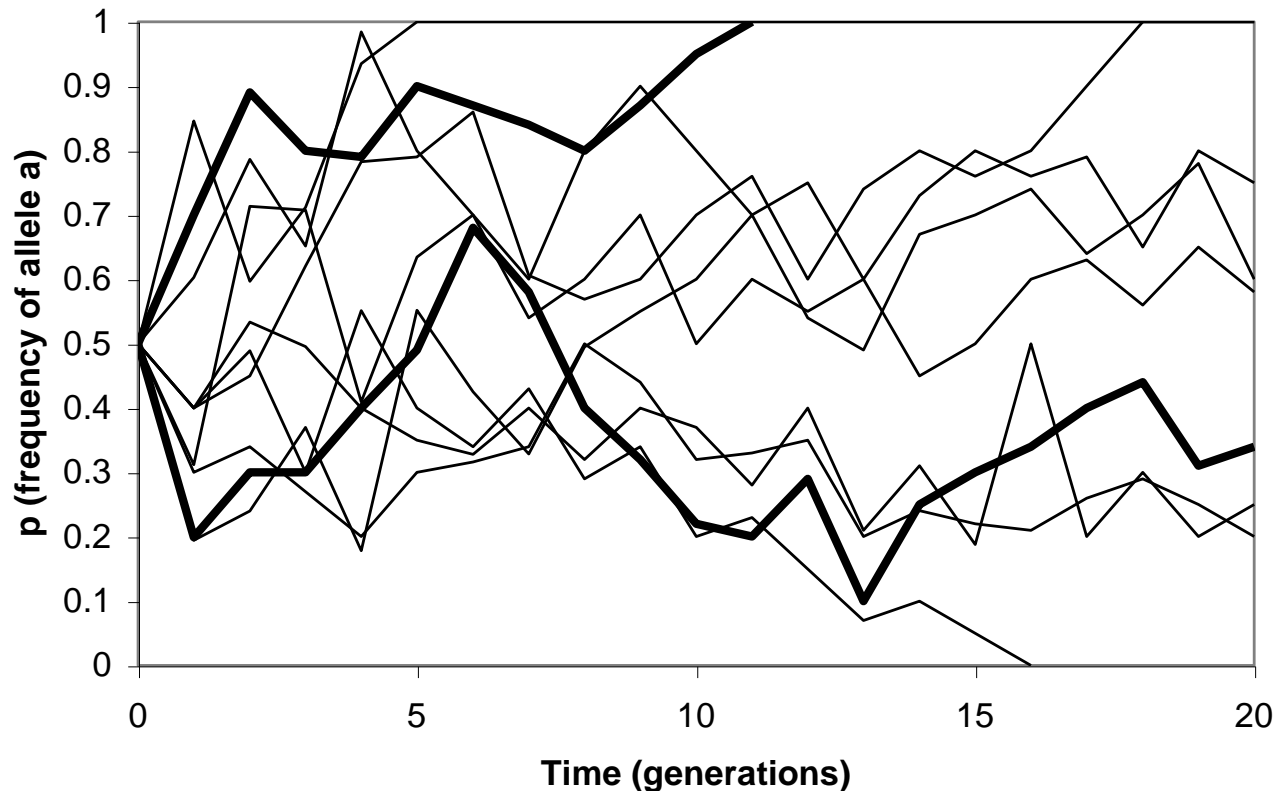
# Genetic Drift

Random changes in allele frequencies across generations due to finite population size



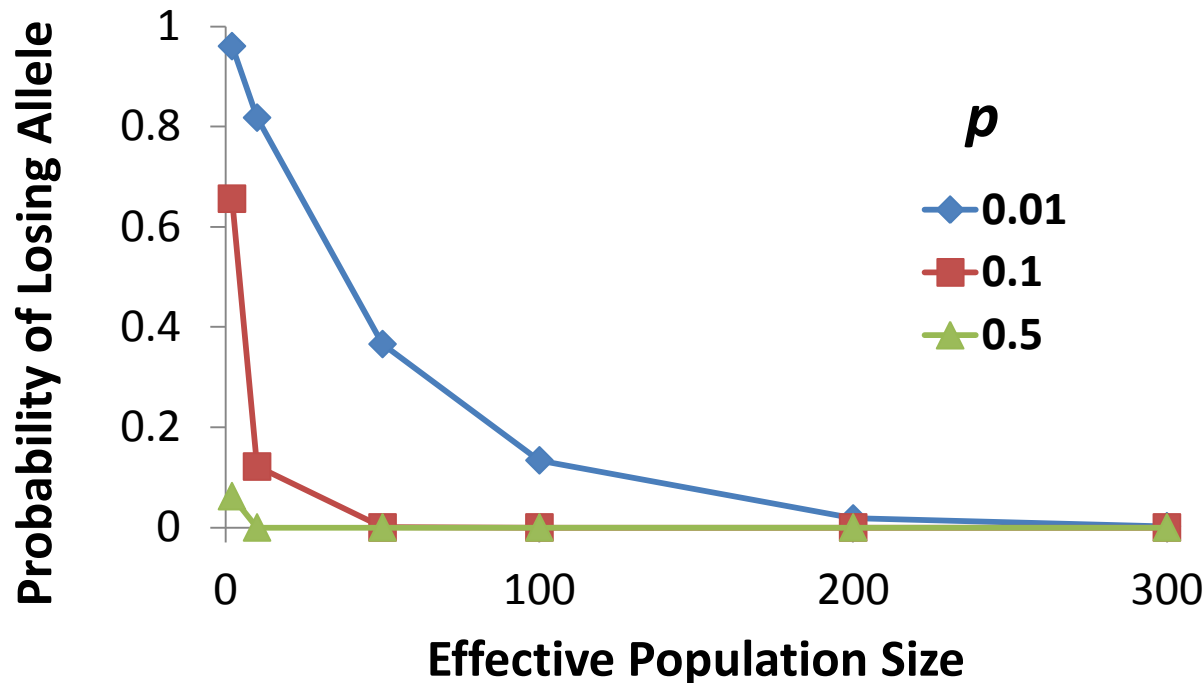
# Genetic Drift

- Allele frequencies change over time randomly and some alleles can go extinct or become “fixed”.
- Which alleles frequencies can “drift” in different directions for different populations, resulting in greater differences among populations



# Loss of Allelic Diversity

- The probability of an allele being lost during a bottleneck of size  $N_e$ :  $(1 - p)^{2N_e}$ , where  $p$  is the frequency of the allele, in the generation following the bottleneck



**Rare alleles are lost first during bottlenecks**

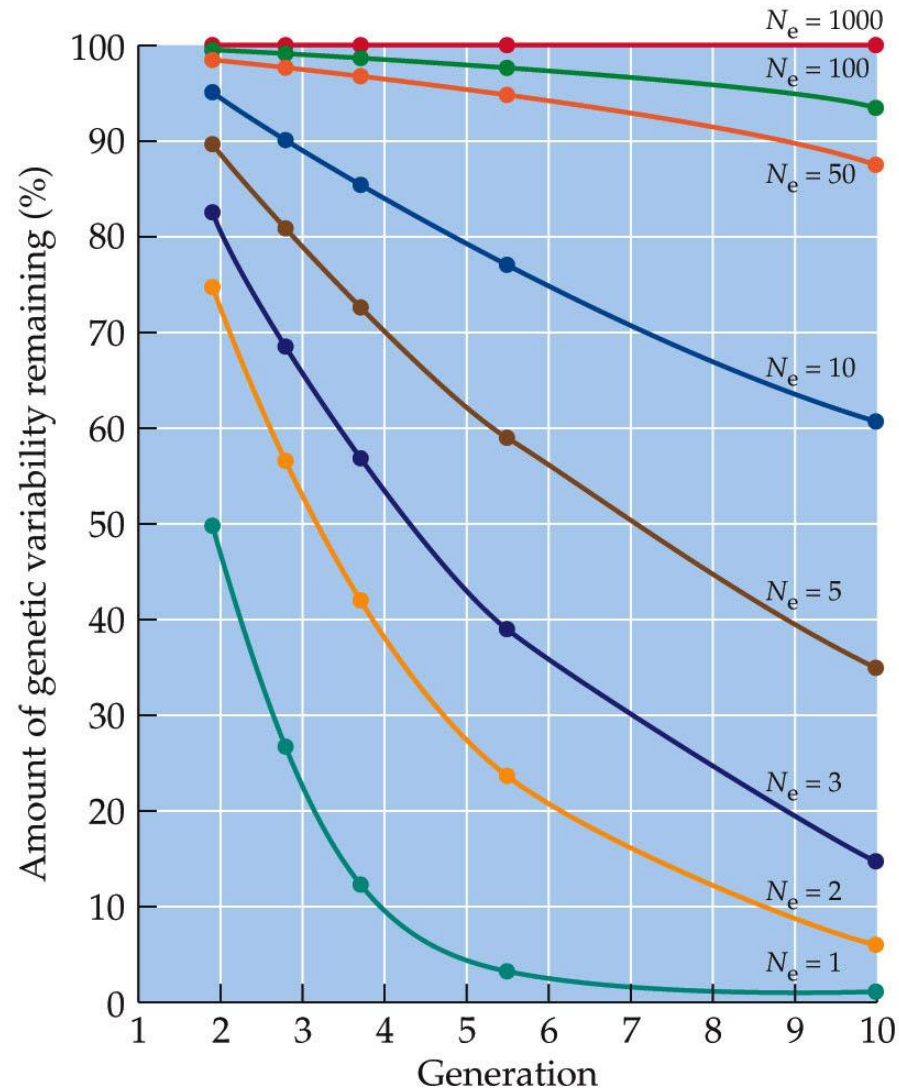


# Loss of Heterozygosity

- Loss of heterozygosity occurs due to the loss of alleles, but occurs more slowly, particularly compared to rare alleles.
- Per generation loss of heterozygosity (increase in homozygosity) =  $1/2N_e$
- Over  $t$  generations, the loss of heterozygosity =  $1 - (1 - 1/2N_e)^t$

# Loss of Genetic Diversity in Small Populations

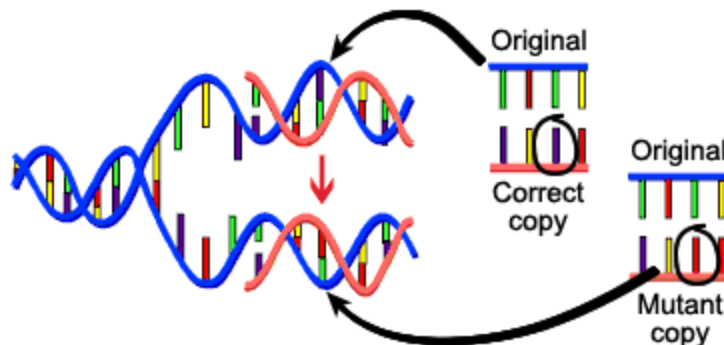
## Theoretical Expectations



# Mutations

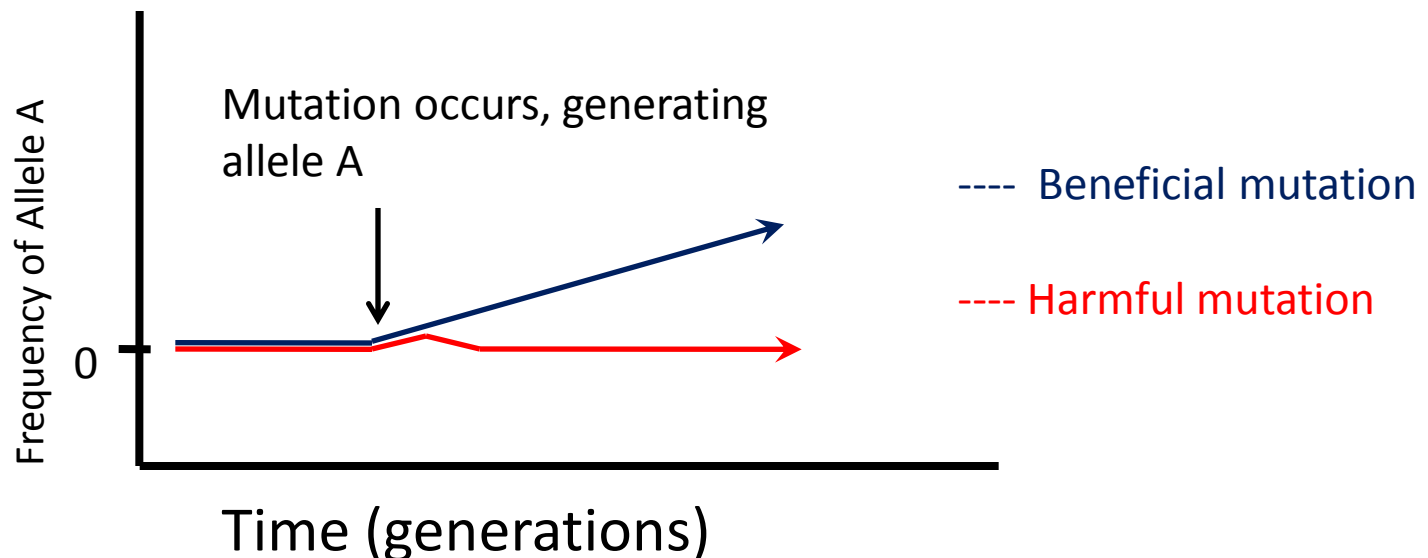
- **Mutation** – an error in the replication of DNA that causes a structural change in a gene.
  - **Entire chromosomal complements**
  - **Translocations:** the movement of nucleotides from one part of the genome to another.
  - **Duplication:** small number of nucleotides or large pieces of chromosomes
  - **Single nucleotides:** removals, substitutions, or insertions
- E.g., a substitution...

CTGGAG  
CTGGGG



# Mutations

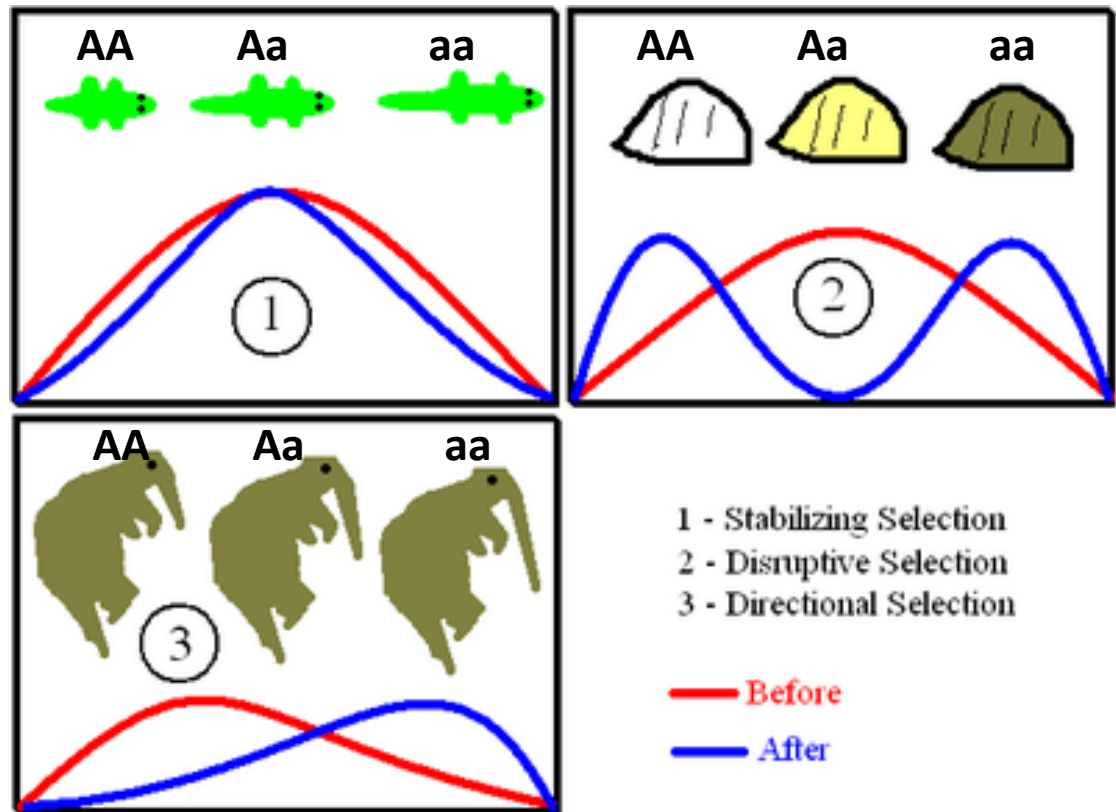
- Mutations can offset loss of genetic diversity due to genetic drift, but mutation rates in nature are low,  $\sim 10^{-9}$  mutations per locus per generation in protein coding nuclear genes.
- Also, most mutations are harmful and get weeded out of the population, relatively few mutations are beneficial



# Selection

- Whether selection increases within-population diversity depends on if selection is stabilizing, disruptive, or directional

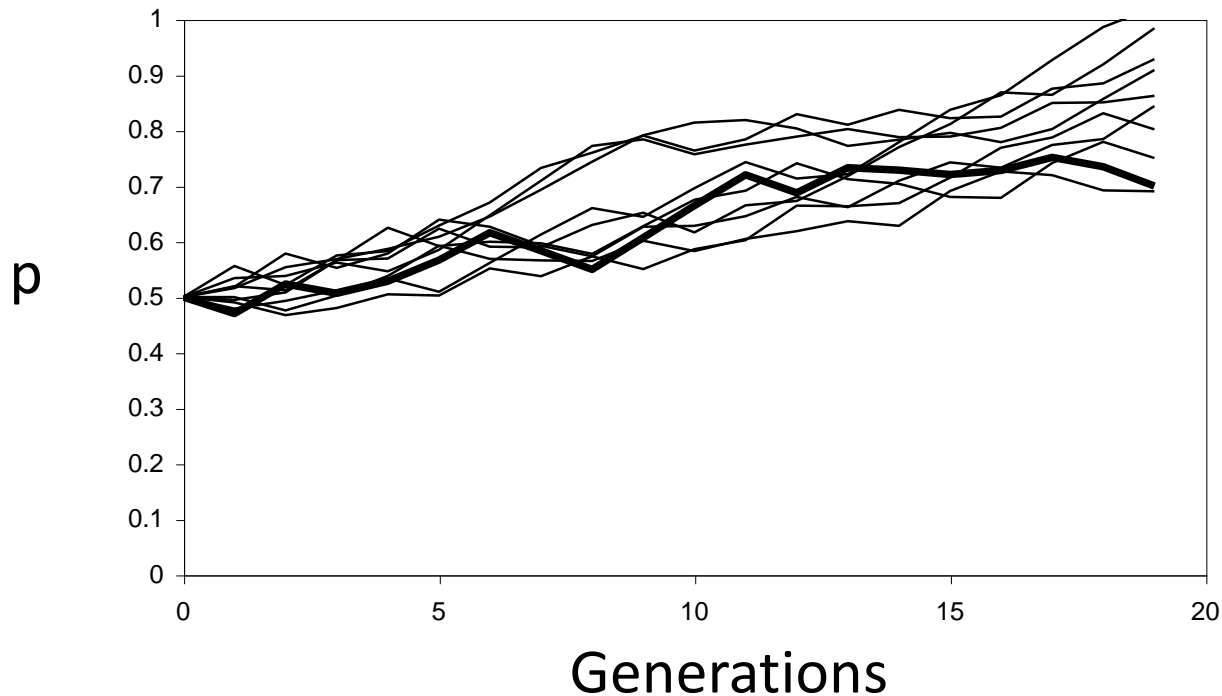
- Whether selection increases within-population diversity also depends on if selection is stabilizing, disruptive, or directional



Frequency of Phenotype

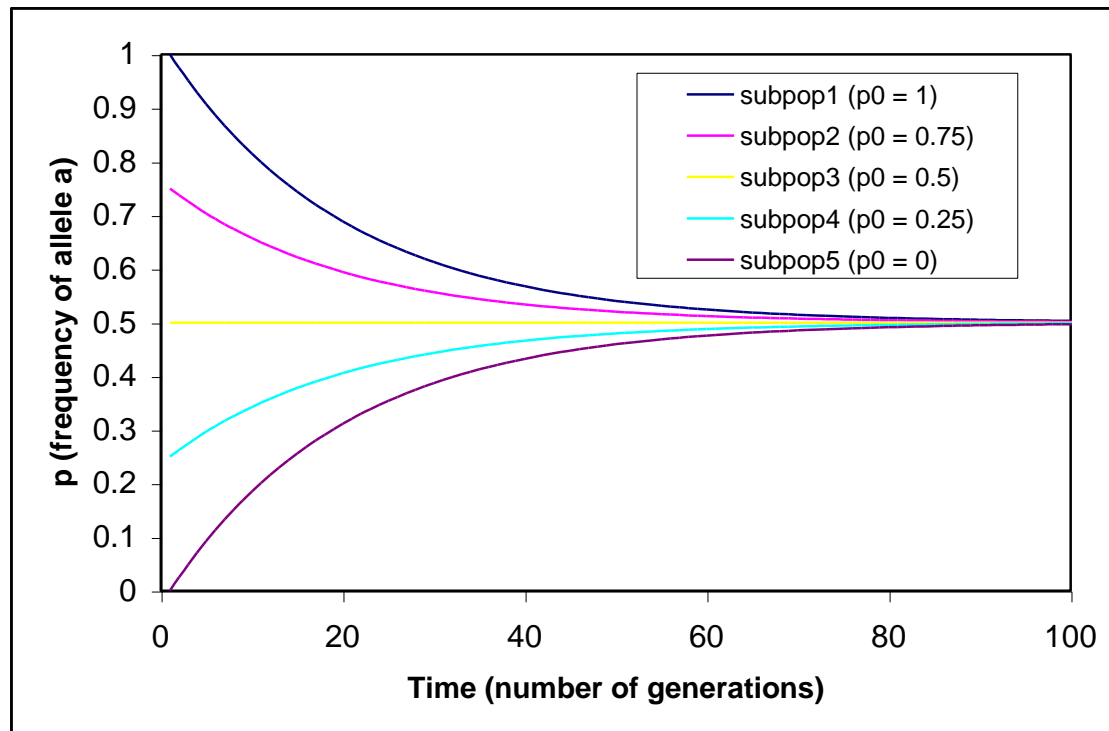
# Selection

- In populations of finite size, selection is not the only factor responsible for changes in allele and genotype frequencies.
- For example, stochastic fluctuations occur in the frequency of allele **A** due to genetic drift, despite a general increase due to directional selection.



# Gene Flow

- Increases genetic variation within populations because it brings in new alleles.
- Reduces genetic differences among populations, because alleles are being exchanged
- E.g., Five populations with different initial frequencies ( $p$ ) of allele  $a$  connected by a migration rate ( $m$ ) of 0.05.



# Evolutionary Processes that Influence Genetic Diversity

## Level of Genetic Variation

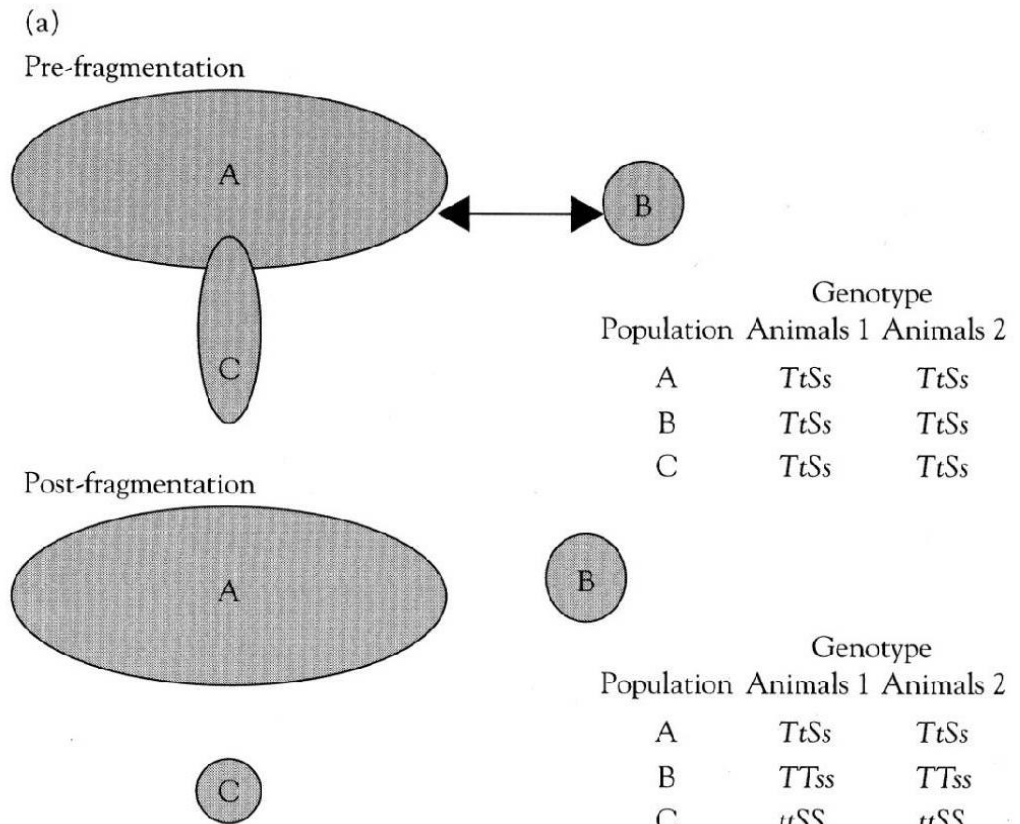
	<u>Within Pops</u>	<u>Among Pops</u>
Genetic Drift	↓	↑
Mutation	↑	↑
Gene Flow	↑	↑
Selection	↑↓	↑↓



# Potential Effects of Habitat Fragmentation on Genetic Diversity

- Disrupting gene flow between population A and B and fragmenting habitat between A and C reduces genetic variation in A and C.

- Ultimately allele *t* or *T* and allele *s* or *S* gets “fixed” in populations B and C.



# Increase in Genetic Variation among Populations

(b)

