

14

**Molecular Evolution and
Population Genetics**

Population Genetics

- Gene pool = the complete set of genetic information in all individuals within a population
- Genotype frequency = proportion of individuals in a population with a specific genotype
- Genotype frequencies may differ from one population to another
- Allele frequency = proportion of any specific allele in a population
- Allele frequencies are estimated from genotype frequencies

- Allele frequencies when mating is random

Maternal gamete	Frequency	Paternal gamete	Frequency	Zygote genotype	Frequency
A	p	A	p	AA	p^2
		a	q	Aa	pq
a	q	A	p	Aa	pq
		a	q	aa	q^2

Identity of maternal allele may be A or a .

Whatever the identity of the maternal allele, that of the paternal allele may be A or a .

Random combinations of maternal and paternal alleles form zygote genotypes.

Hardy–Weinberg Principle

- When gametes containing either of two alleles, A or a , unite at random to form the next generation, the genotype frequencies among the zygotes are given by the ratio

$$p^2 : 2pq : q^2$$

this constitutes the Hardy–Weinberg (HW) Principle

p = frequency of a dominant allele A

q = frequency of a recessive allele a

$$p + q = 1$$

Hardy–Weinberg Principle

- One important implication of the HW Principle is that allelic frequencies will remain constant over time if the following conditions are met:
 - The population is sufficiently large
 - Mating is random
 - Allelic frequencies are the same in males and females
 - Selection does not occur = all genotypes have equal in viability and fertility
 - Mutation and migration are absent

Hardy–Weinberg Principle

- Another important implication is that for a rare allele, there are many more heterozygotes than there are homozygotes for the rare allele

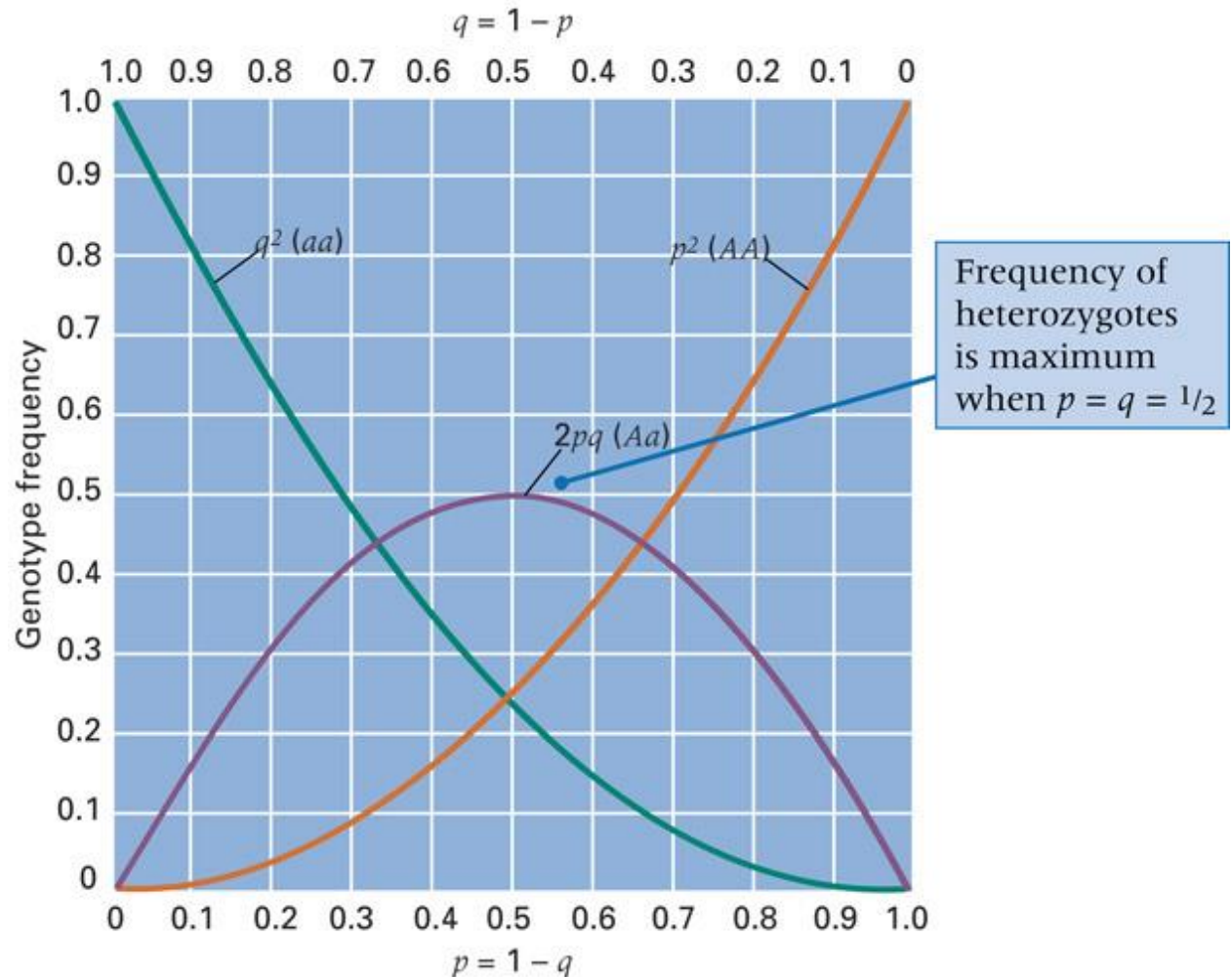


Fig. 14.12

Hardy–Weinberg Principle

- HW frequencies can be extended to multiple alleles:

Frequency of any homozygous genotype = square of allele frequency = p_i^2

Frequency of any heterozygous genotype = 2 x product of allele frequencies = $2p_i p_j$

~~Evolution~~

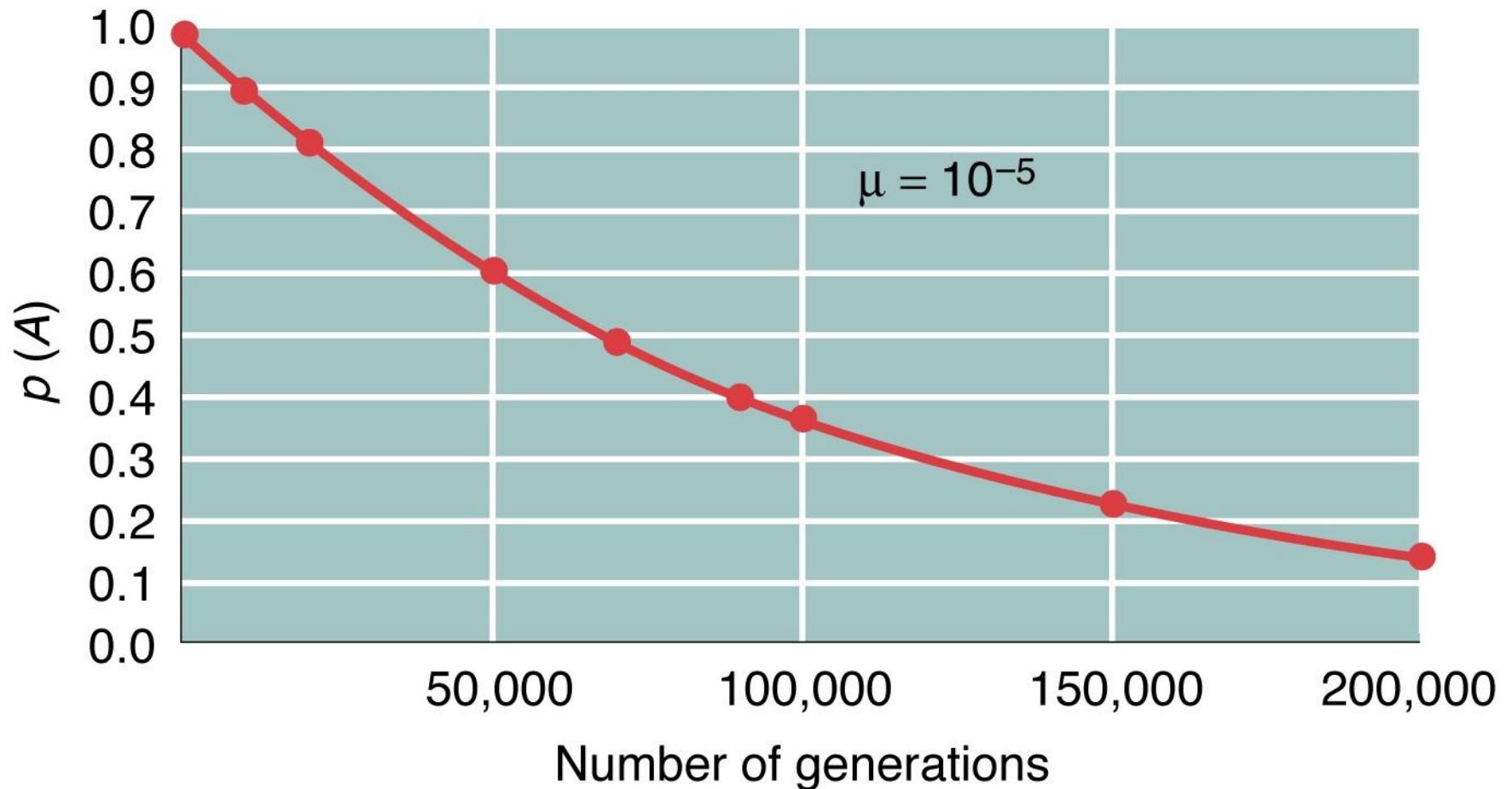
Changing Allele Frequencies

- 1. Mutation = the origin of new genetic capabilities in populations = the ultimate source of genetic variation**
- 2. Natural selection = the process of evolutionary adaptation = genotypes best suited to survive and reproduce in a particular environment give rise to a disproportionate share of the offspring**
- 3. Migration = the movement of organisms among subpopulations**
- 4. Random genetic drift = the random, undirected changes in allele frequencies, especially in small populations**

mutations

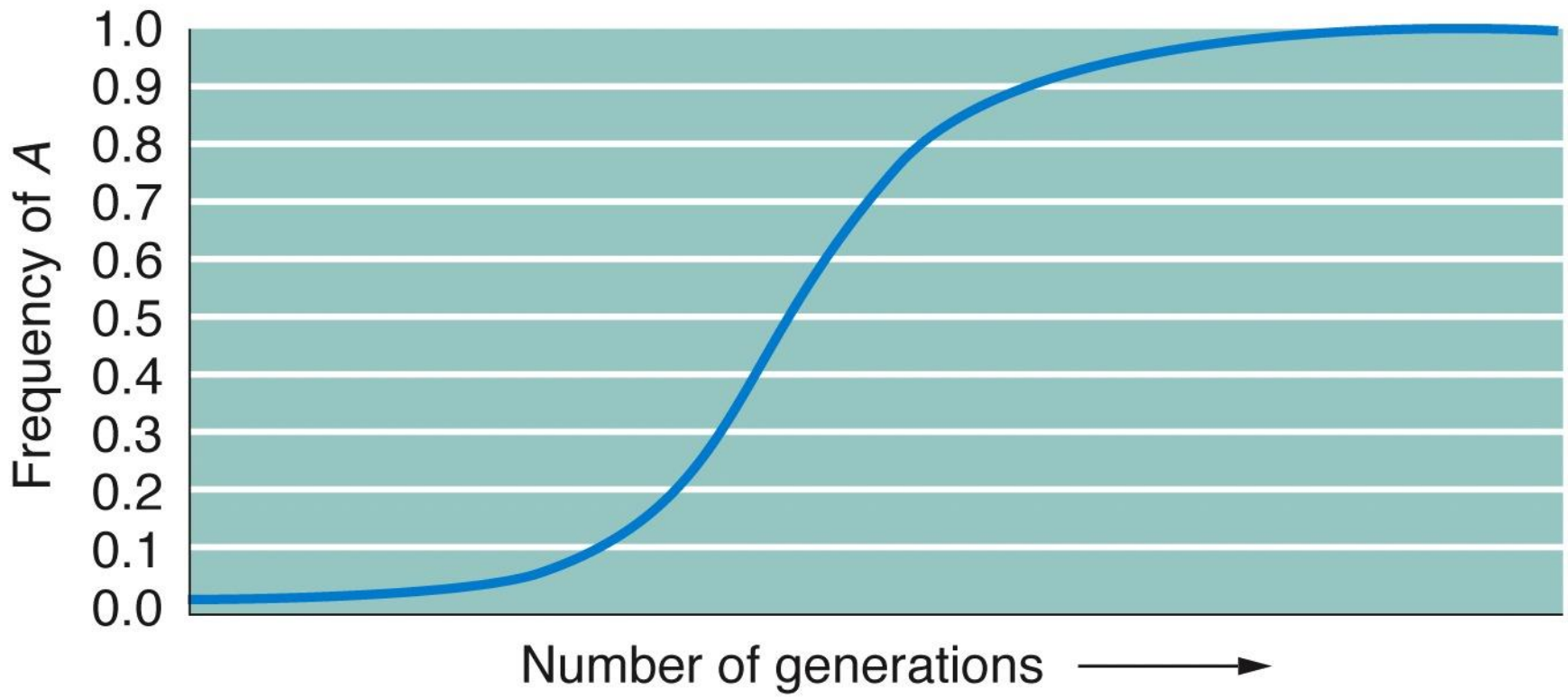
1 generation
 $\Delta p = -\mu p$

n generations
 $p_n = p_{(0)} e^{-\mu n}$



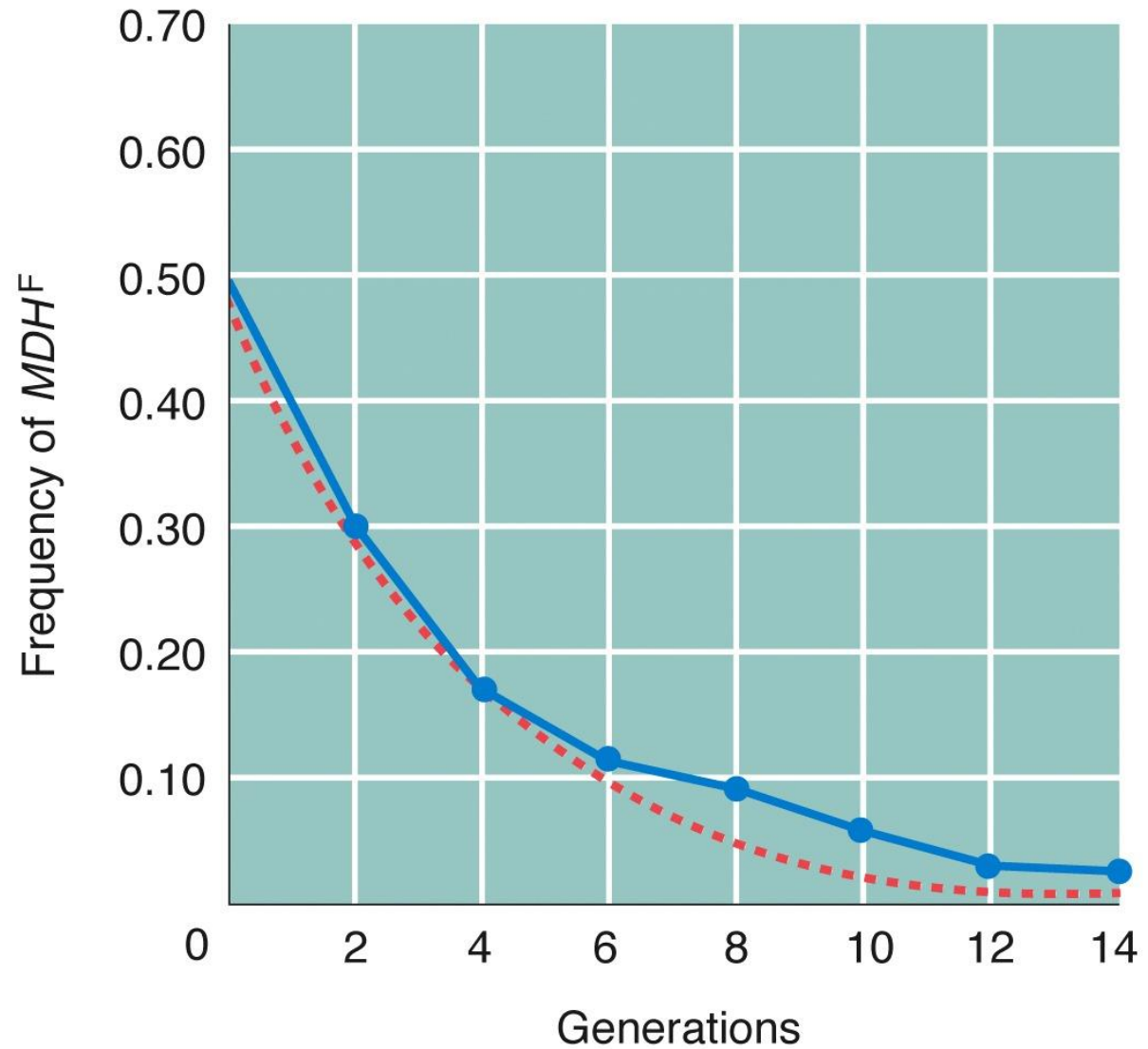
Selection and Fitness

- Organisms differ in their ability to survive and reproduce, and some of these differences are due to genotype
- Fitness (W) is the relative ability of genotypes to survive and reproduce
- Relative fitness measures the comparative contribution of each parental genotype to the pool of offspring genotypes in each generation
- Selection coefficient refers to selective disadvantage of a disfavored genotype



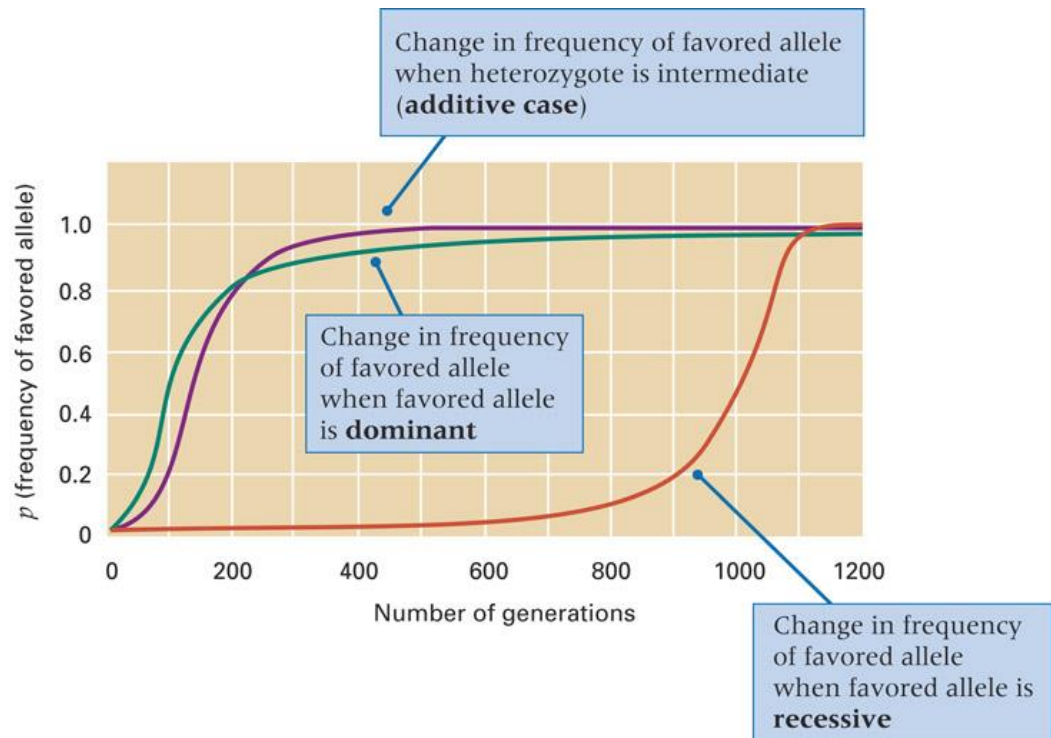
Selection can affect frequencies quite rapidly

$$W_{AA} = 1.00$$
$$W_{Aa} = 0.75$$
$$W_{aa} = 0.40$$



Selection in Diploids

- Frequency of favored dominant allele changes slowly if allele is common
- Frequency of favored recessive allele changes slowly if the allele is rare
- Rare alleles are found most frequently in heterozygotes
- When favored allele is dominant, recessive allele in heterozygotes is not exposed to natural selection



Heterozygote Superiority

- Heterozygote superiority = fitness (measurement of viability and fertility) of heterozygote is greater than that of both homozygotes
- When there is heterozygote superiority, neither allele can be eliminated by selection
- In sickle cell anemia, allele for mutant hemoglobin is maintained in high frequencies in regions of endemic malaria because heterozygotes are more resistant to this disease

Selection in Diploids

- Selection can be balanced by new mutations
- New mutations often generate harmful alleles and prevent their elimination from the population by natural selection
- Eventually the population will attain a state of equilibrium in which the new mutations exactly balance the selective elimination

Migration

Population on island (i) has migrants (m) from the mainland

Allele frequency in the next generation is the weighted average of the two populations

$$p'_{(i)} = (1-m)p_{(i)} + (m)p_{(m)}$$

Random Genetic Drift

- Some changes in allele frequency are random due to genetic drift
- Random genetic drift comes about because populations are not infinitely large
- Only relatively few of the gametes participate in fertilization = sampling
- With random genetic drift, the probability of fixation of an allele is equal to its frequency in the original population

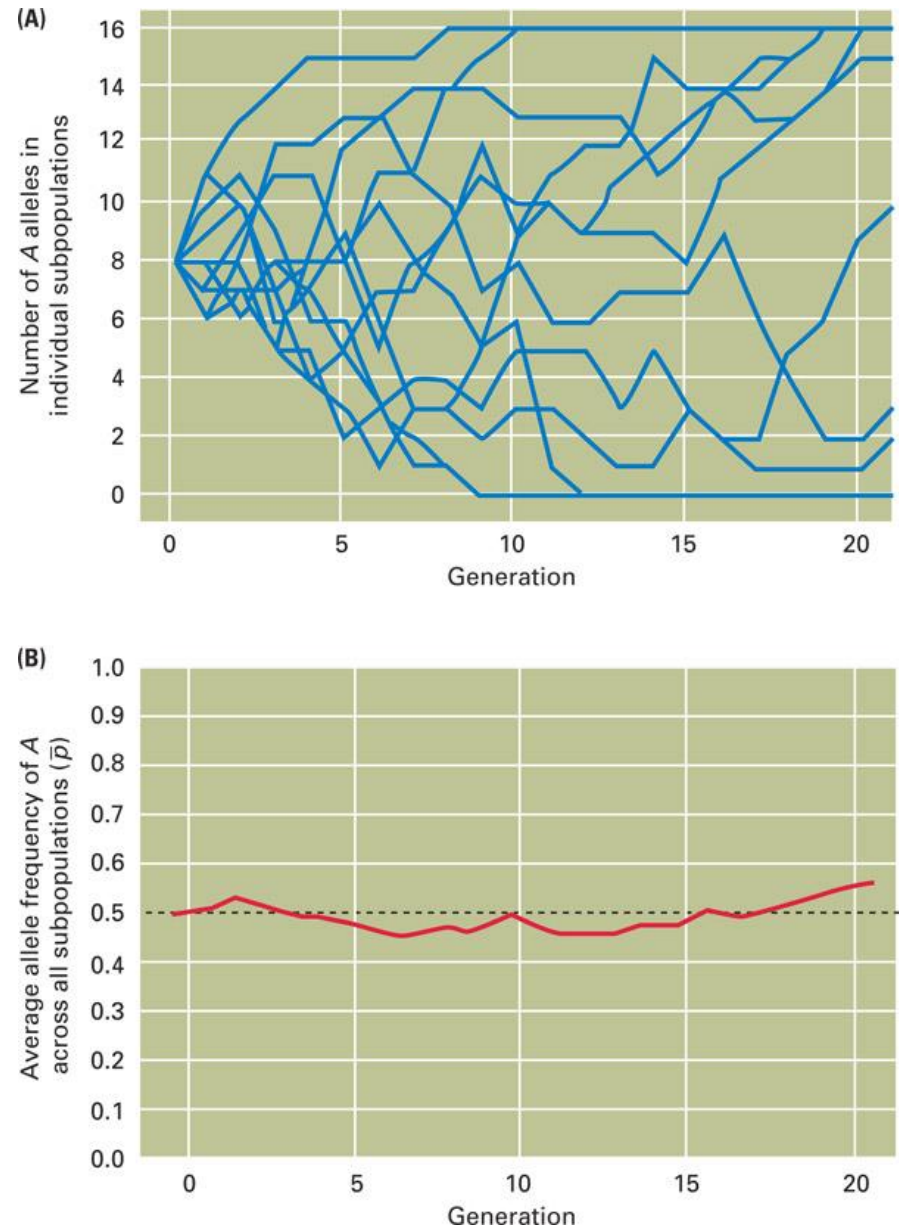


Fig. 14.27

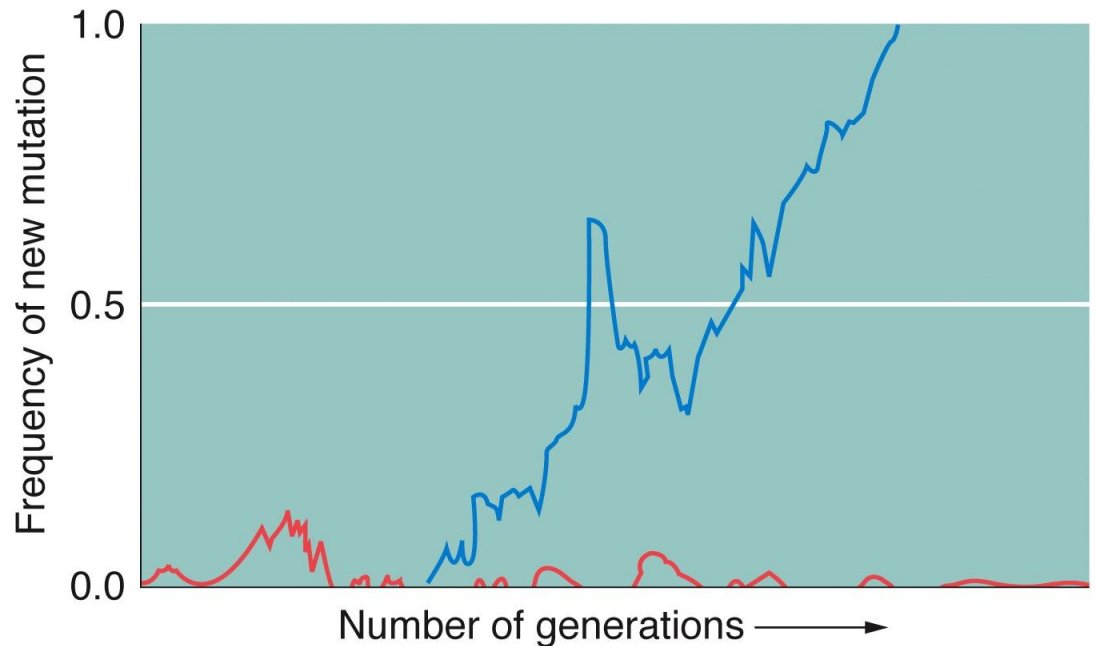
Random Genetic Drift

- Chance that a new mutation will be lost:

$$(2N-1)/(2N)$$

- Chance that a new mutation will become fixed in the population:

$$1/(2N)$$



Inbreeding

- Inbreeding means mating between relatives
- Inbreeding results in an excess of homozygotes compared with random mating
- In most species, inbreeding is harmful due to rare recessive alleles that wouldn't otherwise become homozygous

<i>AA</i>	<i>Aa</i>	<i>aa</i>	Genotype
0	1	0	Initial frequency
$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	After one generation of self-fertilization
$\frac{3}{8}$	$\frac{1}{4}$	$\frac{3}{8}$	After two generations of self-fertilization
$\frac{7}{16}$	$\frac{1}{8}$	$\frac{7}{16}$	After three generations of self-fertilization
$\frac{15}{32}$	$\frac{1}{16}$	$\frac{15}{32}$	After four generations of self-fertilization

With repeated self-fertilization, the frequency of heterozygous genotypes is reduced by half in each successive generation.

Inbreeding

- A convenient measure of effects of inbreeding is based on the reduction of heterozygosity H_i and is called the inbreeding coefficient F

$$F = (2pq - H_i)/2pq$$

- The overall genotype frequencies in the inbred population are

$$f(AA) = p^2(1 - F) + pF$$

$$f(Aa) = 2pq(1 - F)$$

$$f(aa) = q^2(1 - F) + qF$$

Molecular evolution

- The discovery that DNA is the genetic material made it possible to compare corresponding genes even in distantly related species
- DNA and protein sequences contain information about evolutionary relationships among species
- Comparative studies of macromolecules, the study of how and why their sequences change through time constitutes molecular evolution

Molecular evolution

- Accumulation of sequence differences through time is the basis of molecular systematics, which analyses them in order to infer evolutionary relationships
- A gene tree is a diagram of the inferred ancestral history of a group of sequences
- A gene tree is only an estimate of the true pattern of evolutionary relations
- Neighbor joining = one of the way to estimate a gene tree
- Bootstrapping = a common technique for assessing the reliability of a node in a gene tree
- Taxon = the source of each sequence

(A)

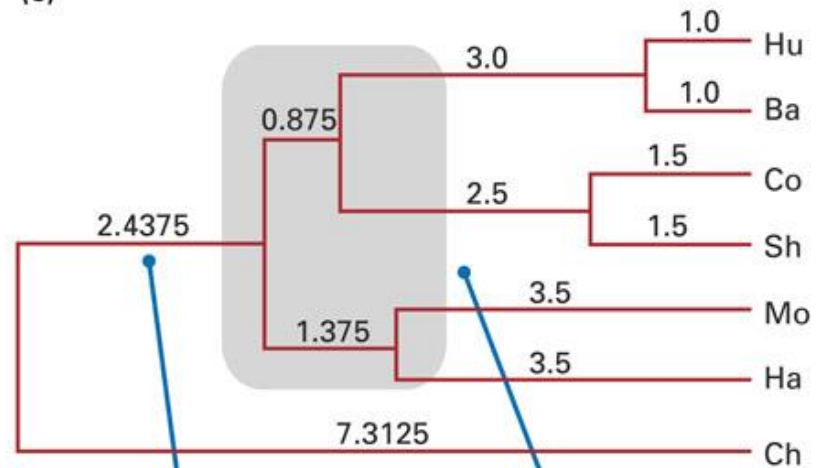
```
Human (Hu)  MVHLTPEEKSAVTALWGVNVDVGGGEALGRLLVVYPWTQRFFESFGDLS
Baboon (Ba)  . . . . . N . . . . . D . . . . .
Cow (Co)     . . . . . A . . . . . F . . . . . K . . . . .
Sheep (Sh)   . . . . . A . . . . . GF . . . . . K . . . . . A . . . . . H . . . . .
Mouse (Mo)   . . . . . DA . . . . . SG . . . . . A . . . . . Y . D . . . . .
Hamster (Ha) . . . . . DA . . . . . AL . . . . . G . . . . . A . A . . . . . H . . . . .
Chicken (Ch) . . . . . W . A . . . . . QLI . G . . . . . A . C . A . . . . . A . . . . . I . . . . . A . . . . . N . . . . .
```

(B)

	Ba	Co	Sh	Mo	Ha	Ch
Hu	2	6	9	8	9	13
Ba		7	10	7	10	13
Co			3	11	12	16
Sh				12	9	15
Mo					7	16
Ha						14

Each entry in the distance matrix equals the number of amino acid differences between the two sequences.

(C)



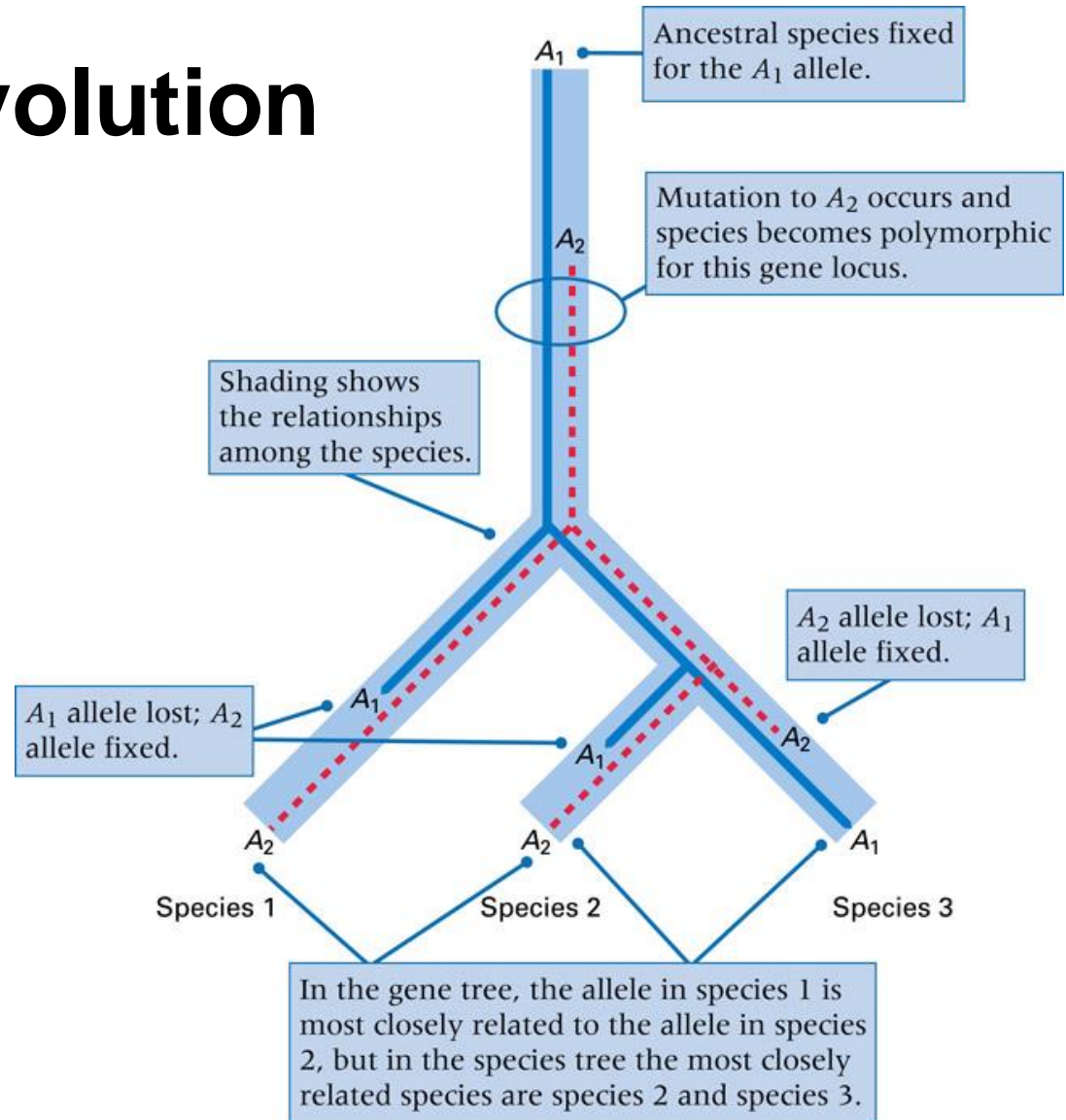
The UPGMA gene tree has the horizontal branch lengths scaled to the distances between the nodes.

Bootstrapping indicates that the branching order in the shaded area cannot be resolved.

Fig. 14.1

Molecular evolution

- A gene tree does not necessarily coincide with a species tree:
 - The sorting of polymorphic alleles in the different lineages
 - Recombination within gene make it possible for different parts of the same gene to have different evolutionary histories



Molecular evolution

- Rate of sequence evolution = the fraction of sites that undergo a change in some designated time interval = number of replacements per site per billion years
- Rates of evolution can differ dramatically from one protein to another

Molecular evolution

- There are different kind of nucleotide sites and nucleotide substitutions depending on their position and function in the genome
- Synonymous substitution = no change in amino acid sequence = primarily at the third codon position
- Nonsynonymous substitution = amino acid replacement
- Rates of evolution of nucleotide sites differ according to their function

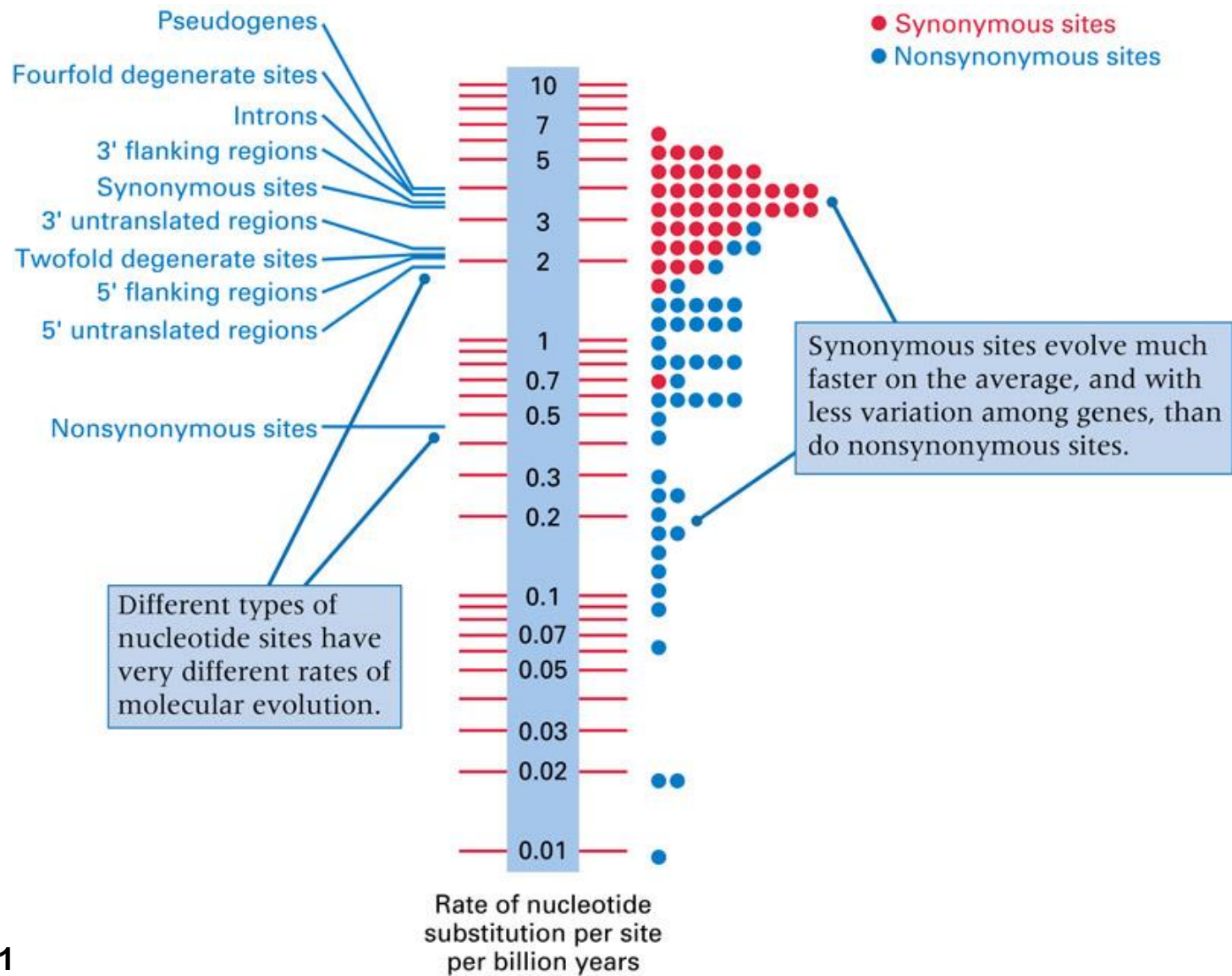


Fig. 1

Molecular evolution

- New genes usually evolve through duplication and divergence
- Ortologous genes = duplicated as an accompaniment to speciation, retain the same function
- Paralogous genes = duplicated in the genome of the same species, acquire new or more specialized function
- Pseudogenes = duplicated genes that have lost their function

Population Genetics

- Population genetics = application of genetic principles to entire populations of organisms
- Population = group of organisms of the same species living in the same geographical area
- Subpopulation = any of the breeding groups within a population among which migration is restricted
- Local population = subpopulation within which most individuals find their mates