

Mutation

COMP 571 - Fall 2010
Luay Nakhleh, Rice University

Outline

- (1) The source of all genetic variation
- (2) The fate of a new mutation
- (3) Mutation models
- (4) The influence of mutation on allele frequency and autozygosity
- (5) The coalescent with mutation

(1) The Source of All Genetic Variation

- * So far, we have discussed processes, such as non-random mating, gene flow, etc., that shape or change existing genetic variation
- * But, where does genetic variation come from in the first place?

(1) The Source of All Genetic Variation

- * The process of **mutation**, the permanent incorporation of random errors in DNA that results in differences between ancestral and descendant copies of DNA sequences, is the ultimate source of all genetic variation

(1) The Source of All Genetic Variation

- * **Somatic mutations** occur in non-reproductive cells and are not passed onto offspring
- * **Germ line mutations**, those that occur in reproductive cells like eggs and sperm, can be passed to offspring and are the only mutations that matter to large-scale evolution

(1) The Source of All Genetic Variation

- * Mutation is a broad term that encompasses a wide variety of events that lead to alterations in DNA sequences
- * **Point mutation:** replacement of a single base pair by another nucleotide
- * **Transitions:** **purine** to purine (A \leftrightarrow G) and **pyrimidine** to pyrimidine (C \leftrightarrow T)
- * **Transversions:** purine to pyrimidine or pyrimidine to purine

(1) The Source of All Genetic Variation

- * **Synonymous mutations** result in the same translation of a DNA sequence into a protein due to the redundant nature of the genetic code
- * **Nonsynonymous mutations** result in a codon that does change the resulting amino acid sequence

		Second letter				
		U	C	A	G	
First letter	U	UUU } Phe UUC } UUA } Leu UUG }	UCU } UCC } Ser UCA } UCG }	UAU } Tyr UAC } UAA Stop UAG Stop	UGU } Cys UGC } UGA Stop UGG Trp	U C A G
	C	CUU } CUC } Leu CUA } CUG }	CCU } CCC } Pro CCA } CCG }	CAU } His CAC } CAA } Gln CAG }	CGU } CGC } Arg CGA } CGG }	U C A G
	A	AUU } AUC } Ile AUA } AUG Met	ACU } ACC } Thr ACA } ACG }	AAU } Asn AAC } AAA } Lys AAG }	AGU } Ser AGC } AGA } Arg AGG }	U C A G
	G	GUU } GUC } Val GUA } GUG }	GCU } GCC } Ala GCA } GCG }	GAU } Asp GAC } GAA } Glu GAG }	GGU } GGC } Gly GGA } GGG }	U C A G

(1) The Source of All Genetic Variation

- * **Insertion or deletion (indels)** of DNA sequences is another type of mutations
- * Indels within coding regions result in **frameshift mutations** if the change in sequence length is not an even multiple of three

(1) The Source of All Genetic Variation

- * **Gene duplication** gives rise to multiple copies of homologous genes, called **multigene families**
- * Some of the duplicate genes may lose function due to the accumulation of mutations, becoming **pseudogenes**
- * **Gene conversion** may result in the homogenization of the sequences of multiple loci within multigene families

(1) The Source of All Genetic Variation

- * At the genome/chromosome level, a mutation can take the form of:
 - * **inversion**: a chromosomal region forms a loop structure that results in a segment breaking and being repaired in reversed orientation
 - * **translocation**: segments of chromosome break free from one chromosome and are incorporated by repair mechanisms into a non-homologous chromosome
 - * **Transposable elements** are frequent causes of translocations

(1) The Source of All Genetic Variation

- * **Horizontal, or lateral, gene transfer (HGT)** is another form of mutation and refers to the movement and incorporation of DNA segments between different individuals within a population and even between different species

(1) The Source of All Genetic Variation

- * The probability that a locus or base pair will experience a mutation is a critical parameter in population genetics since the **rate of mutation** describes how rapidly novel genetic variation is added to populations
- * Mutation rates are quite difficult to estimate with precision in many types of organisms
- * The most general rule of mutation is that it is a rare event with a low probability of occurrence

(1) The Source of All Genetic Variation

Table 5.1 Per-locus mutation rates measured for five loci that influence coat-color phenotypes in inbred lines of mice (Schlager & Dickie 1971). Dominant mutations were counted by examining the coat color of F1 progeny from brother–sister matings. Recessive mutations required examining the coat color of F1 progeny from crosses between an inbred line homozygous for a recessive allele and a homozygous wild-type dominant allele. The effort to obtain these estimates was truly incredible, involving around 7 million mice observed over the course of 6 years.

Locus	Gametes tested	Mutations observed	Mutation rate per locus $\times 10^{-6}$ (95% CI)
Mutations from dominant to recessive alleles			
Albino	150,391	5	33.2 (10.8–77.6)
Brown	919,699	3	3.3 (0.7–9.5)
Dilute	839,447	10	11.9 (5.2–21.9)
Leaden	243,444	4	16.4 (4.5–42.1)
Non-agouti	67,395	3	44.5 (9.2–130.1)
All loci	2,220,376	25	11.2 (7.3–16.6)
Mutations from recessive to dominant alleles			
Albino	3,423,724	0	0 (0.0–1.1)
Brown	3,092,806	0	0 (0.0–1.2)
Dilute	2,307,692	9	3.9 (1.8–11.1)
Leaden	266,122	0	0 (0.0–13.9)
Non-agouti	8,167,854	34	4.2 (2.9–5.8)
All loci	17,236,978	43	2.5 (1.8–3.4)

95% CI, 95% confidence interval.

(1) The Source of All Genetic Variation

Table 5.2 Rates of spontaneous mutation expressed per genome and per base pair for a range of organisms. The most reliable estimates come from microbes with DNA genomes whereas estimates from RNA viruses and eukaryotes have greater uncertainty. Full explanation of the assumptions and uncertainties behind these estimates can be found in Drake et al. (1998).

Organism	Mutation rate per replication	
	Per genome	Per base pair
Lytic RNA viruses		
Bacteriophage Q β	6.5	
Poliovirus	0.8	
Vesicular stomatitis virus	3.5	
Influenza A	≥ 1.0	
Retroviruses		
Spleen necrosis virus	0.04	
Rous sarcoma virus	0.43	
Bovine leukemia virus	0.027	
Human immunodeficiency virus	0.16–0.22	
DNA-based microbes		
Bacteriophage M13	0.0046	7.2×10^{-7}
Bacteriophage λ	0.0038	7.7×10^{-8}
Bacteriophages T2 and T4	0.0040	2.4×10^{-8}
<i>Escherichia coli</i>	0.0025	5.4×10^{-10}
<i>Neurospora crassa</i>	0.0030	7.2×10^{-11}
<i>Saccharomyces cerevisiae</i>	0.0027	2.2×10^{-10}
Eukaryotes		
<i>Caenorhabditis elegans</i>	0.018	2.3×10^{-10}
<i>Drosophila</i>	0.058	3.4×10^{-10}
Human	0.49	1.8×10^{-10}
		2.5×10^{-8a}
Mouse	0.16	5.0×10^{-11}

^aEstimate from Nachman and Crowell (2000) based on pseudogene divergence between humans and chimpanzees.

(1) The Source of All Genetic Variation

- * How can such a low-probability event like mutation add more than a trivial amount of genetic variation to populations?
- * To answer this question, let's take humans as an example
- * Mutation rate is about 1×10^{-9} mutations per base pair per generation
- * Each of the two copies of the genome has about 3.2×10^9 base pairs

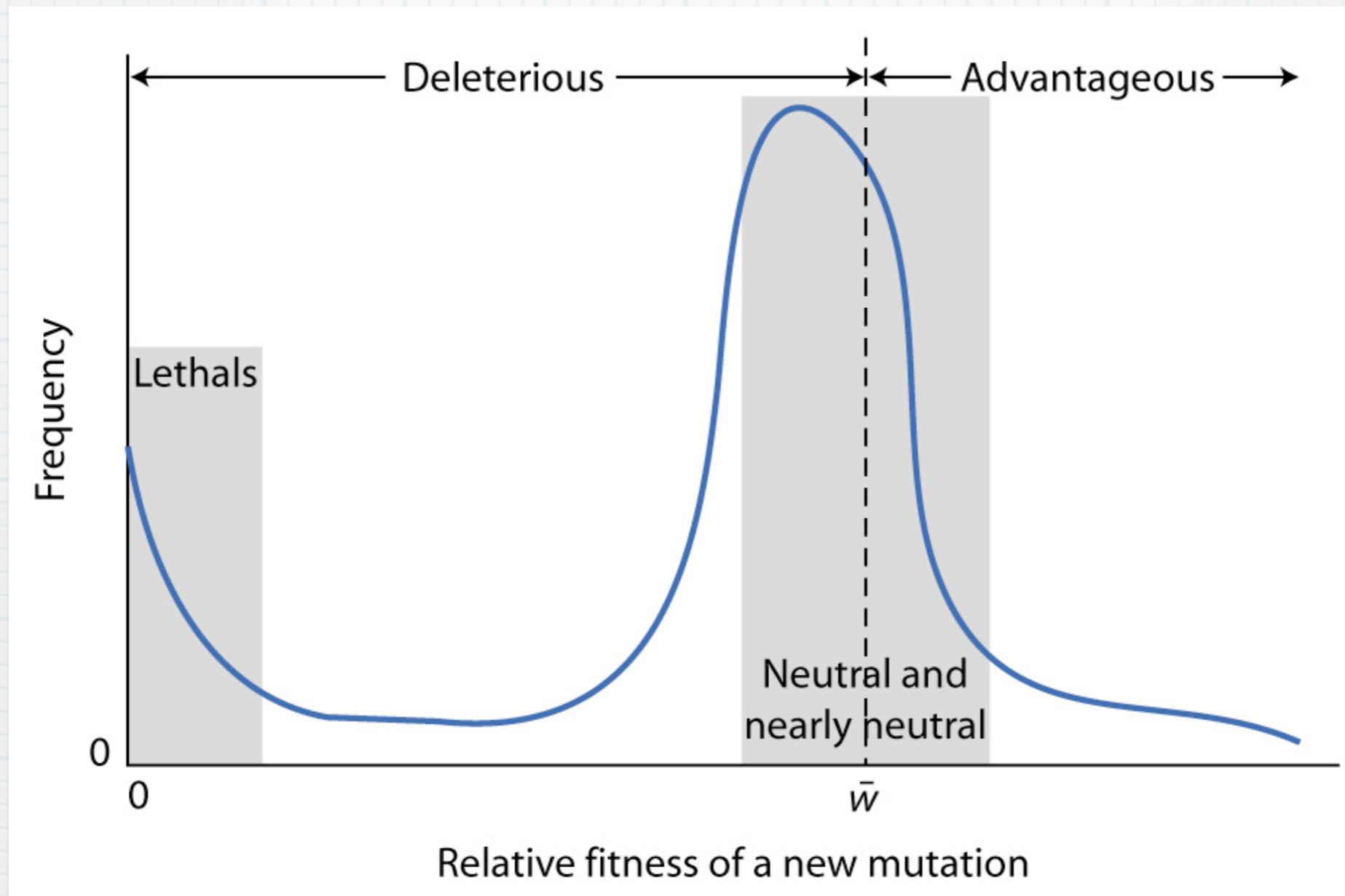
(1) The Source of All Genetic Variation

- * Each genome of each diploid individual will have $1 \times 10^{-9} \times 2 \times 3.2 \times 10^9 = 6.4$ mutations
- * Assuming the human population size is about 6.486 billion, there are a total of $6.4 \times 6.486 \times 10^9 = 41.5 \times 10^9$ mutations
- * This means that over 41 billion mutations are expected every generation
- * The number of mutations per generation depends on the rate of mutation, the genome size, and the population size

(1) The Source of All Genetic Variation

- * The impact that a mutant allele has on the phenotype of an individual can vary greatly
- * The phenotype is often considered in the context of its survivorship and reproduction, or **fitness**
- * The range of the possible fitnesses for an individual mutant allele can be thought of as a **mutation fitness spectrum**

(1) The Source of All Genetic Variation



(1) The Source of All Genetic Variation

- * **Deleterious mutations** reduce survival and reproduction, while mutations that improve survival and reproduction are **advantageous**
- * Mutations that have small positive or negative effects on fitness are called **neutral** or **nearly neutral**
- * Mutations that increase fitness above the average fitness of the population are called **beneficial**

(1) The Source of All Genetic Variation

- * Strongly deleterious or strongly beneficial mutations will be steadily and predictably driven to loss or to fixation, respectively, by **natural selection**
- * Fixation and loss of neutral or nearly neutral mutations is due in whole or in part to **random genetic drift**
- * A consequence is that mildly deleterious mutations may reach fixation by chance and accumulate in a population over time, and some mildly beneficial mutations may be lost from the population by chance

(1) The Source of All Genetic Variation

- * In asexual organisms, beneficial mutations may arise in two different individuals, and “compete” against each other, with the result being the loss of one of them; this is known as **clonal interference**
- * In diploid organisms, this phenomenon may be overcome by **genetic recombination**
- * In asexual organisms, **homologous recombination** (whose effect is similar to gene conversion) may also ameliorate clonal interference (more later)

(1) The Source of All Genetic Variation

- * Next, we will address questions related to:
 - * the fate of a new mutation,
 - * the impact of mutation on allele frequencies,
 - * the balance between genetic drift and mutation, and
 - * how to incorporate mutations into the coalescent model

(2) The Fate of a New Mutation

- * The mutation rate dictates how often a new mutation will appear in a population
- * But, once a mutation has occurred, population genetic processes acting on it will determine whether it increases or decreases in frequency

(2) The Fate of a New Mutation

* Here, we will consider four different perspectives on the fate of a new mutation, each of which makes different assumptions about the population context:

(A) Chance of mutation loss due to Mendelian segregation

(B) Fate of a new mutation in a finite population

(C) Geometric model of mutations fixed by natural selection

(D) Muller's Ratchet and the fixation of deleterious mutations

(2) The Fate of a New Mutation

(A) Chance of mutation loss due to Mendelian segregation

- * Assume existing alleles A_1, A_2, \dots at a locus
- * A mutation gives rise to new allele A_m , which must be found in a heterozygous genotype $A_x A_m$, for some existing allele A_x
- * For every progeny produced by $A_x A_m$, there is a $1/2$ chance the mutant allele is inherited, and a $1/2$ chance it is not
- * The total chance that $A_x A_m$ passes the mutant allele on to the next generation depends on the number of progeny produced

(2) The Fate of a New Mutation

(A) Chance of mutation loss due to Mendelian segregation

- * If k is the number of progeny produced by $A_x A_m$ and there is independent assortment of alleles, then the probability that the mutant allele is not transmitted to the next generation in any of the progeny is

$$P(\text{mutation lost}) = \left(\frac{1}{2}\right)^k$$

(2) The Fate of a New Mutation

(A) Chance of mutation loss due to Mendelian segregation

- * For a given family of size k (assuming every pair of parents produce two progeny on average), the probability that a mutant allele is not transmitted to the next generation is

$$P(\text{mutation lost}) = \left(\frac{2^k}{k!} \right) e^{-2} \left(\frac{1}{2} \right)^k$$

(2) The Fate of a New Mutation

(A) Chance of mutation loss due to Mendelian segregation

- * For a given family of size k (assuming every pair of parents produce two progeny on average), the probability that a mutant allele is not transmitted to the next generation is

$$P(\text{mutation lost}) = \left(\frac{2^k}{k!} \right) e^{-2} \left(\frac{1}{2} \right)^k$$

the expected
frequency of
parental pairs

(2) The Fate of a New Mutation

(A) Chance of mutation loss due to Mendelian segregation

- * For a given family of size k (assuming every pair of parents produce two progeny on average), the probability that a mutant allele is not transmitted to the next generation is

$$P(\text{mutation lost}) = \left(\frac{2^k}{k!} \right) e^{-2} \left(\frac{1}{2} \right)^k$$

the expected
frequency of
parental pairs

the chance of
not transmitting
the mutant allele

(2) The Fate of a New Mutation

(A) Chance of mutation loss due to Mendelian segregation

Table 5.3 The expected frequency of each family size per pair of parents (k) under the Poisson distribution with a mean family size of 2 ($\bar{k} = 2$). Also given is the expected probability that a mutant allele A_m would not be transmitted to any progeny for a given family size. Note that $0!$ equals one.

Family size per pair of parents (k) ...	0	1	2	3	4 ... k
Expected frequency	e^{-2}	$2e^{-2}$	$2e^{-2}$	$\frac{4}{3}e^{-2}$	$\frac{2}{3}e^{-2} \dots \frac{2^k}{k!}e^{-2}$
Chance that A_m is not transmitted	1	$\frac{1}{2}$	$\left(\frac{1}{2}\right)^2$	$\left(\frac{1}{2}\right)^3$	$\left(\frac{1}{2}\right)^4 \dots \left(\frac{1}{2}\right)^k$

(2) The Fate of a New Mutation

(A) Chance of mutation loss due to Mendelian segregation

- * The total probability that the mutant allele is not transmitted to the next generation is the sum over all possible family sizes from zero to infinity:

$$P(\text{mutant lost}) = \sum_{k=0}^{\infty} \binom{2^k}{k!} e^{-2} \left(\frac{1}{2}\right)^k = e^{-2} \sum_{k=0}^{\infty} \binom{2^k}{k!} \left(\frac{1}{2}\right)^k = e^{-2} \sum_{k=0}^{\infty} \frac{1}{k!}$$

- * Using the fact that as k goes to infinity, we have

$$\sum_{k=0}^{\infty} \frac{1}{k!} \rightarrow e$$

we get

$$P(\text{mutant lost}) = e^{-1}$$

(2) The Fate of a New Mutation

(A) Chance of mutation loss due to Mendelian segregation

- * A general expression for the cumulative probability of a mutation being lost from the population over time is

$$P(\text{mutant lost in generation } t) = e^{x-1}$$

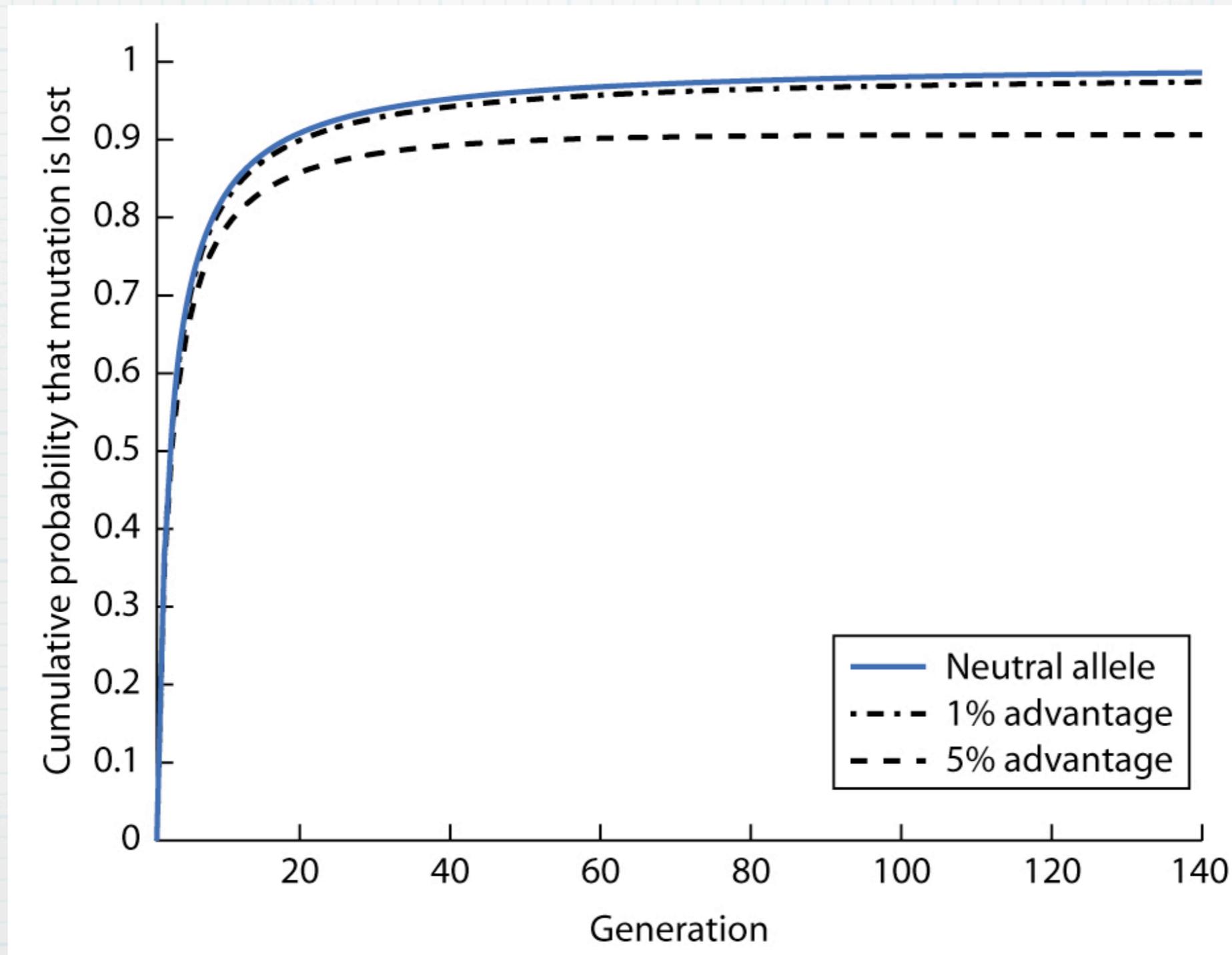
where x is the probability of loss in generation $t-1$

- * If a new mutation has selective advantage c ($c=1$ indicates neutrality), then the cumulative probability that an allele is lost at generation t is

$$P(\text{mutant lost in generation } t) = e^{c(x-1)}$$

(2) The Fate of a New Mutation

(A) Chance of mutation loss due to Mendelian segregation



(2) The Fate of a New Mutation

(B) Fate of a new mutation in a finite population

- * So far, we have assumed the population size is very large, which allowed us to use the expected values for the proportion of parental pairs for each family size under the Poisson distribution and the chance of an allele being lost for each family size, probabilities that should only be met in the limit of many parental pairs that span a wide range of family sizes
- * We now consider the fate of a new mutation in a finite population

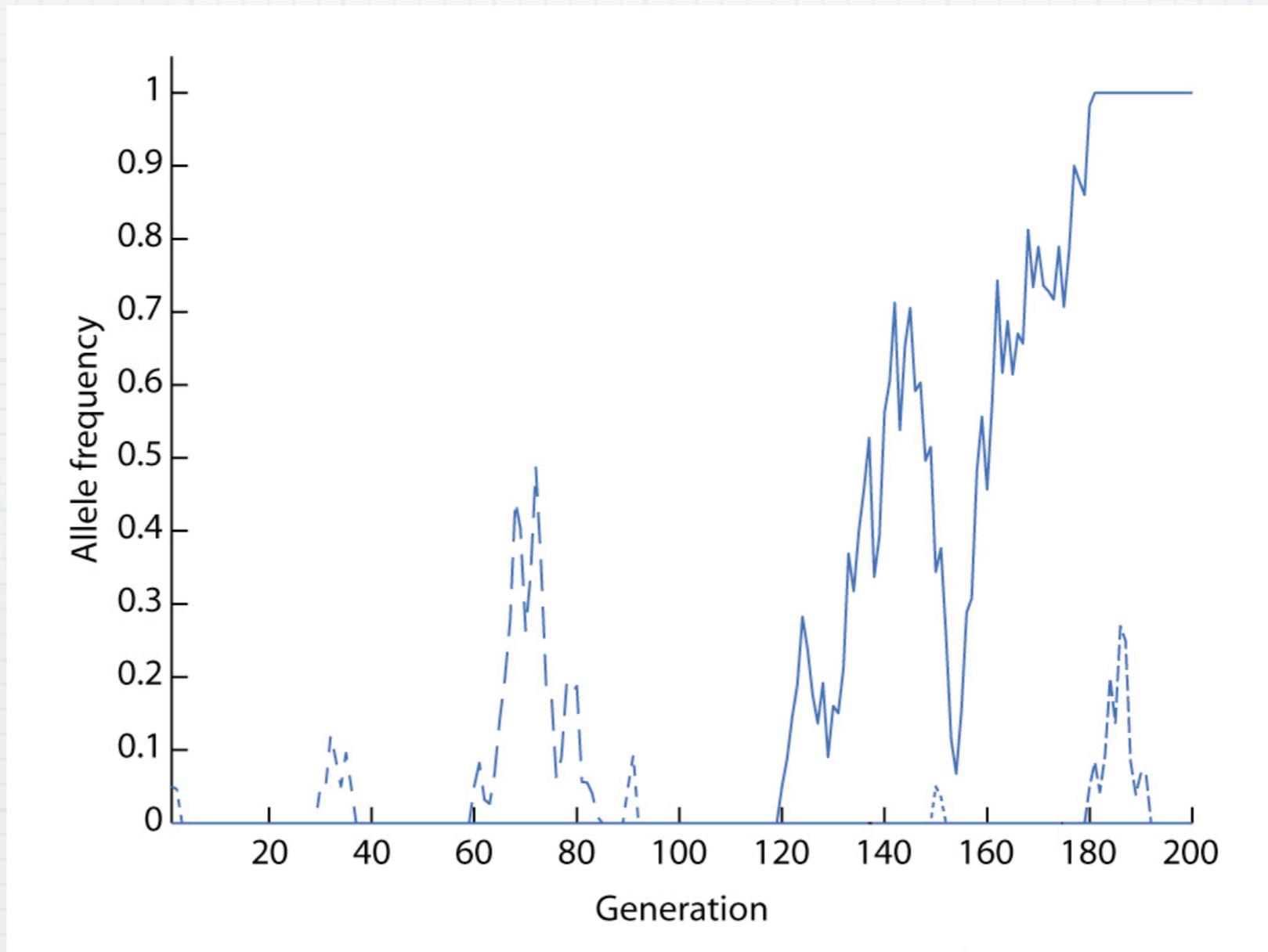
(2) The Fate of a New Mutation

(B) Fate of a new mutation in a finite population

- * The initial frequency of a new mutation is $p_0 = 1/(2N_e)$
- * If the frequency is determined strictly by genetic drift, then the new mutation has a probability that equals p_0 of going to fixation, and a probability that equals $1-p_0$ of going to loss

(2) The Fate of a New Mutation

(B) Fate of a new mutation in a finite population



One new mutation is introduced into the population every 30 generations, and $N_e = 10$

(2) The Fate of a New Mutation

(B) Fate of a new mutation in a finite population

- * If the population were observed at a single point in time, it is possible that it would be polymorphic
- * Observing many such loci at one point in time, it would be very likely that at least some of them would be polymorphic
- * This observation forms the basis of the **neutral theory** of molecular evolution, the hypothesis that genetic variation in populations is caused by genetic drift (more later)

(2) The Fate of a New Mutation

(C) Geometric model of mutations fixed by natural selection

- * We will now consider the fate of new beneficial mutations, first looking at natural selection alone and then at the combined effect of natural selection and genetic drift
- * Mutations may have a range of effects on fitness as well as on any phenotype with variation that has a genetic basis
- * Fisher sought to determine the distribution of the effect sizes of the beneficial mutations that are fixed by natural selection over time

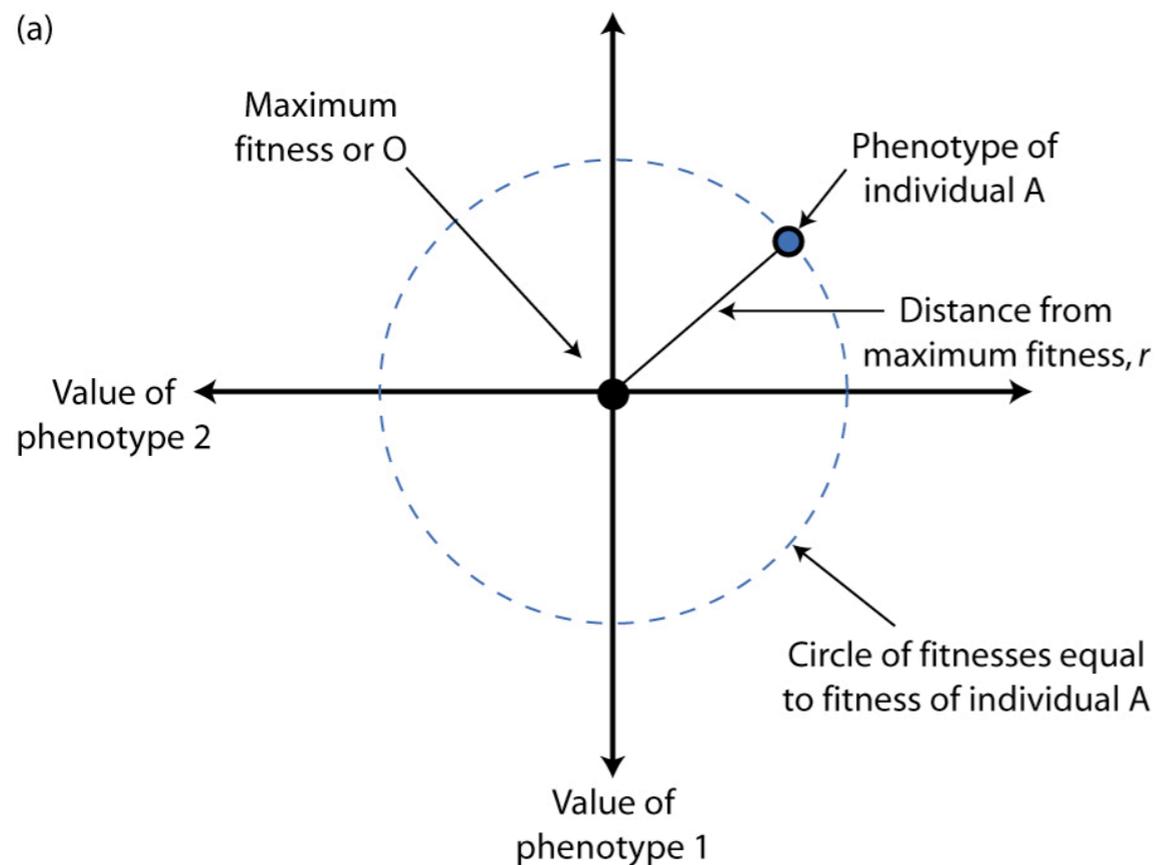
(2) The Fate of a New Mutation

(C) Geometric model of mutations fixed by natural selection

- * **Micromutationalism**: the view that beneficial mutations fixed by natural selection have small effects and therefore that the process of adaptation is marked by gradual genetic change
- * The model that leads to this view is called the **geometric model of mutation**

(2) The Fate of a New Mutation

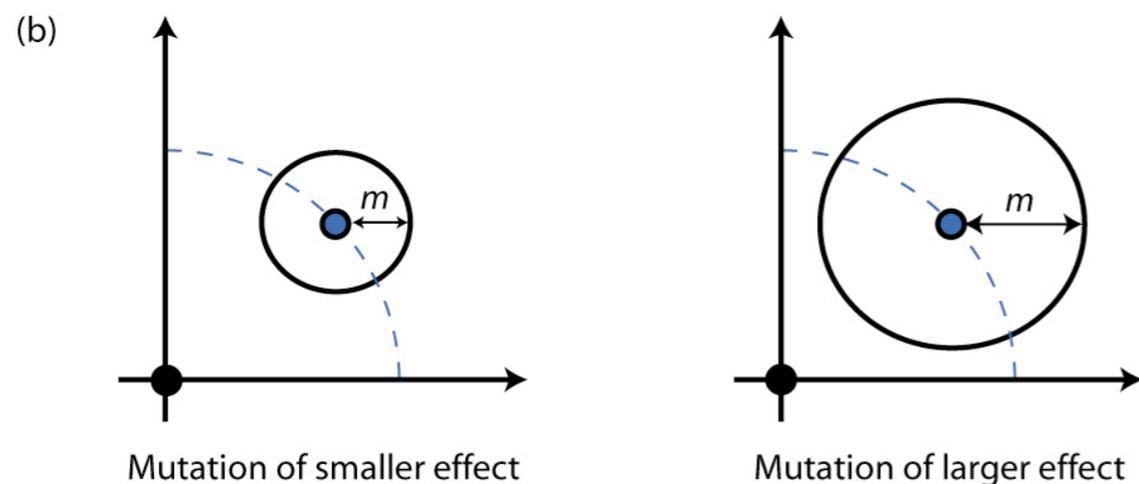
(C) Geometric model of mutations fixed by natural selection



Mutations with a very large effect (greater than $2r$) cannot get the phenotype any closer to the optimum. Such mutations will never be fixed by natural selection.

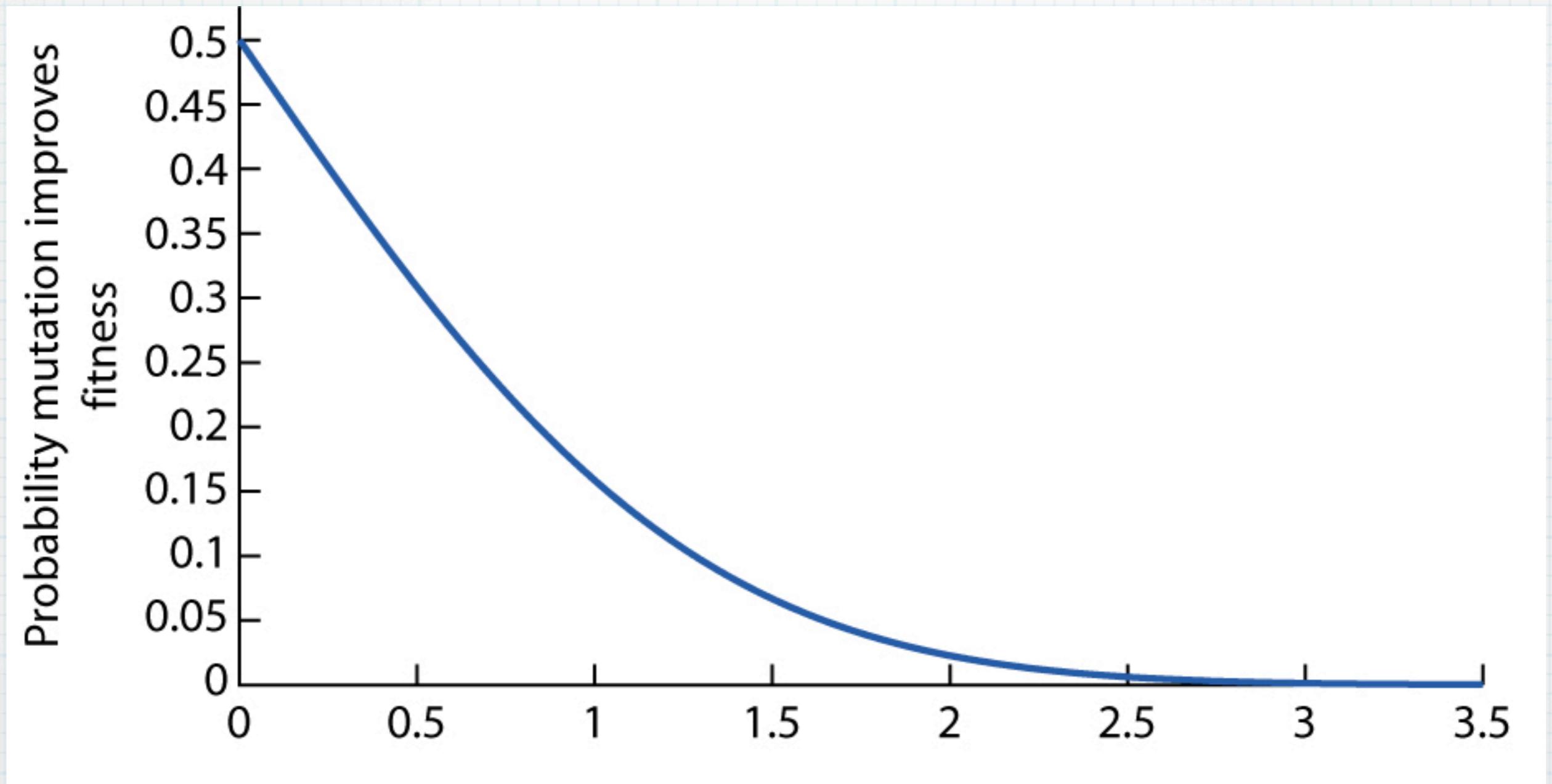
For mutations with a smaller effect (radius m , smaller than $2r$) we have

$$P(\text{mutation improves fitness}) = \frac{1}{2} \left(1 - \frac{m}{2r} \right)$$



(2) The Fate of a New Mutation

(C) Geometric model of mutations fixed by natural selection



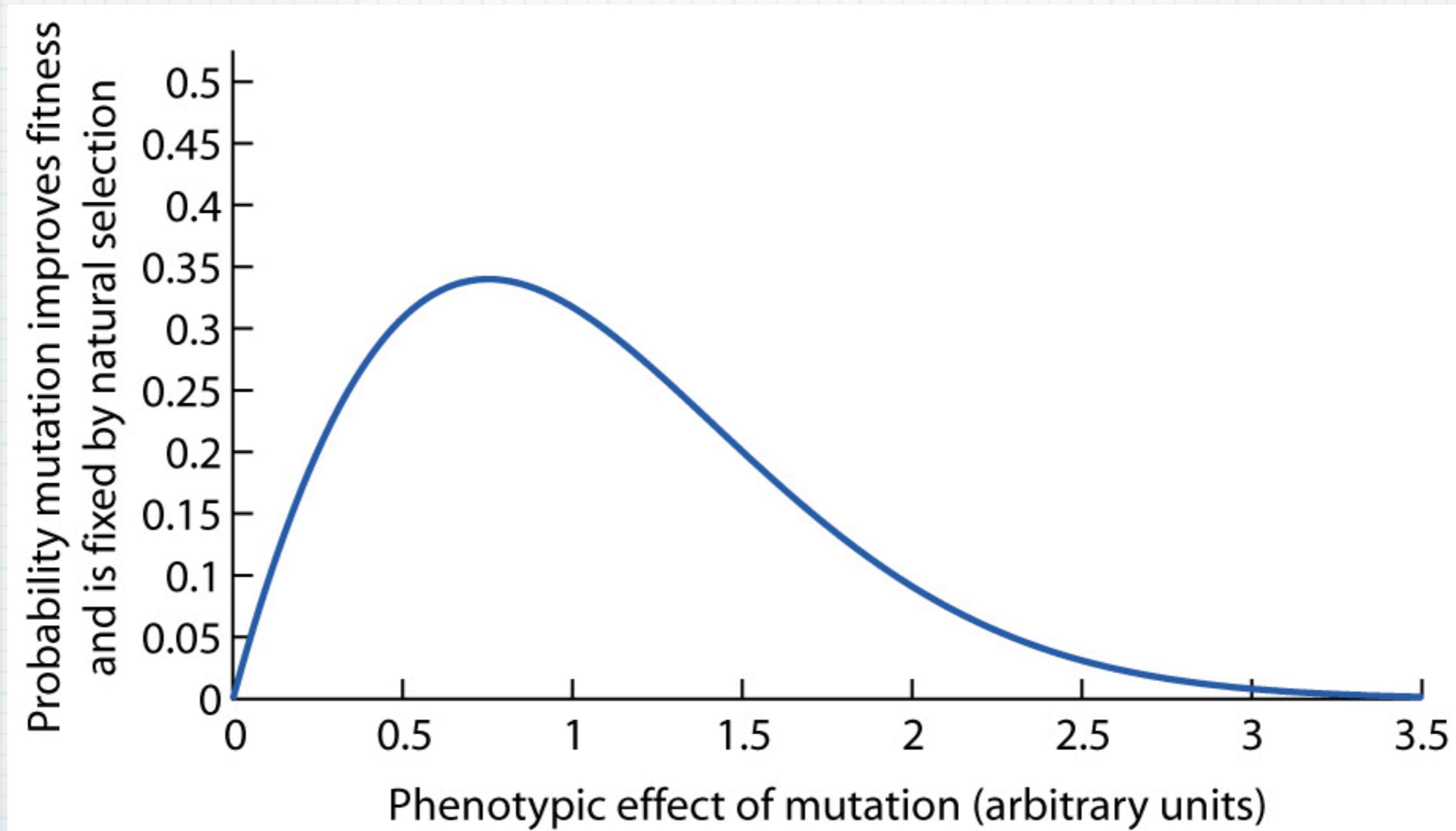
(2) The Fate of a New Mutation

(C) Geometric model of mutations fixed by natural selection

- * Kimura later reevaluated the predictions of the geometric model by relaxing Fisher's implicit assumption of an infinite effective population size
- * This change allows genetic drift to operate on the frequency of mutations along with natural selection
- * Natural selection will only determine the fate of an allele if it is stronger than the randomizing effect of genetic drift

(2) The Fate of a New Mutation

(C) Geometric model of mutations fixed by natural selection



New mutations with an intermediate effect on fitness are the most likely to fix under natural selection in a finite population

(2) The Fate of a New Mutation

(D) Mullers' Ratchet and fixation of deleterious mutations

- * The combination of mutation, genetic drift, and natural selection results in the progressive loss of the class of individuals in a population with the fewest mutations, in a phenomenon called **Muller's Ratchet**
- * Muller's Ratchet results in the accumulation of more and more mutations in a population, which leads to ever-declining average fitness in populations if most mutations are deleterious
- * Thus, Muller's Ratchet demonstrates a selective advantage of recombination under some conditions

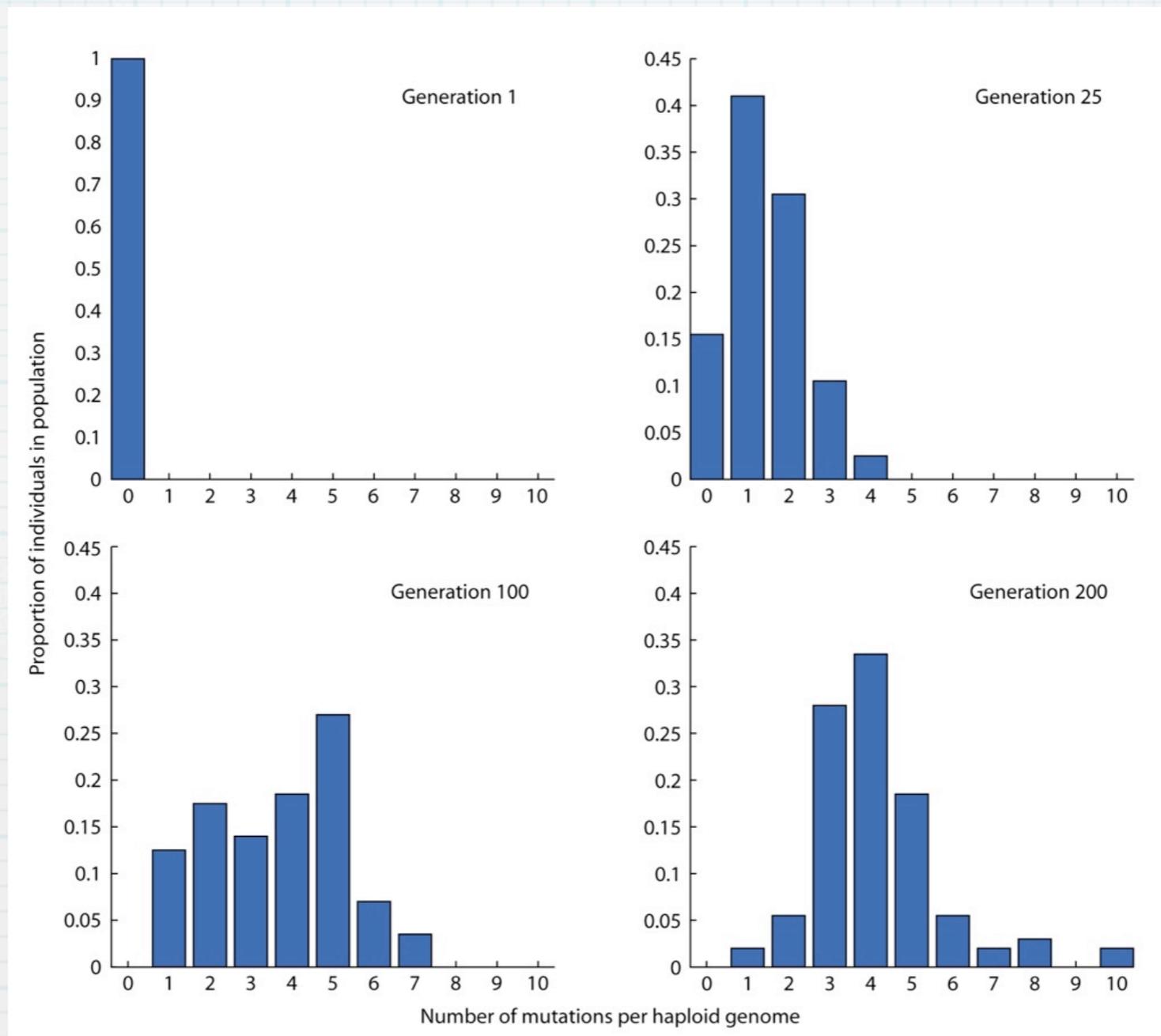
(2) The Fate of a New Mutation

(D) Mullers' Ratchet and fixation of deleterious mutations

- * Consider a finite haploid population
- * Assume all mutations at all loci are equally deleterious and acted against by natural selection to the same degree
- * The selective disadvantage is s at each locus with a mutation and the total selection coefficient against an individual with n mutated loci is $(1-s)^n$
- * Further, assume mutations are deleterious and irreversible

(2) The Fate of a New Mutation

(D) Mullers' Ratchet and fixation of deleterious mutations



$N_e=200$, mutation rate = 0.06, each mutation reduces the chance of reproduction by 1%, and each individual has 100 loci

(3) Mutation Models

- * **Mutation models** attempt to capture the essence of the genetic changes caused by mutation while at the same time simplifying the process of mutation into a form that permits generalizations about allele frequency changes
- * We'll describe some mutation models for discrete alleles and others for DNA sequences

(3) Mutation Models

Mutation models for discrete alleles

- * The **infinite alleles model** assumes that each mutational event creates a new allele unlike any other allele currently in the population
- * It is an assumption used to guarantee that identity in state is equivalent to identity by descent

(3) Mutation Models

Mutation models for discrete alleles

- * Under the infinite alleles model:
 - * mutation produces the original copy of each allele, but is not an ongoing process influencing the frequency of any allele already in the population
 - * the evolutionary distance between all alleles is the same, since all alleles are produced by a single mutational event and alleles can never accumulate multiple mutations

(3) Mutation Models

Mutation models for discrete alleles

- * The **finite alleles model** assumes that there is a finite number, k , of possible alleles
- * In this model, each allele can mutate with equal probability to each of the other $k-1$ possible allelic states
- * In this model, the same allele can be created by mutation repeatedly, blurring the equivalence of identity by state and identity by descent

(3) Mutation Models

Mutation models for discrete alleles

- * As the number of possible alleles decreases and the mutation rate increases, allelic state becomes a poorer and poorer measure of identity by descent since an increasing proportion of alleles with identical states have completely independent histories
- * The term **homoplasy** refers to allelic states that are identical in state without being identical by descent

(3) Mutation Models

Mutation models for discrete alleles

- * The **stepwise mutation model** assumes that allelic states are somehow ordered and the allelic states produced by mutation depend on the initial state of an allele

(3) Mutation Models

Mutation models for discrete alleles

- * The mutation model allows defining measures that express the genetic similarity or dissimilarity of individuals or populations, called **genetic distances**
- * The **standard genetic distance**, or **D measure**, has been widely employed
- * Assume frequencies of alleles A_1, \dots, A_k in two populations, so that we have p_1, \dots, p_k and q_1, \dots, q_k

(3) Mutation Models

Mutation models for discrete alleles

$$J_{11} = \sum_{i=1}^k p_i^2 \quad J_{22} = \sum_{i=1}^k q_i^2$$

$$J_{12} = \sum_{i=1}^k p_i q_i$$

$$I = \frac{J_{12}}{\sqrt{J_{11} J_{22}}}$$

$$D = -\ln(I)$$

(3) Mutation Models

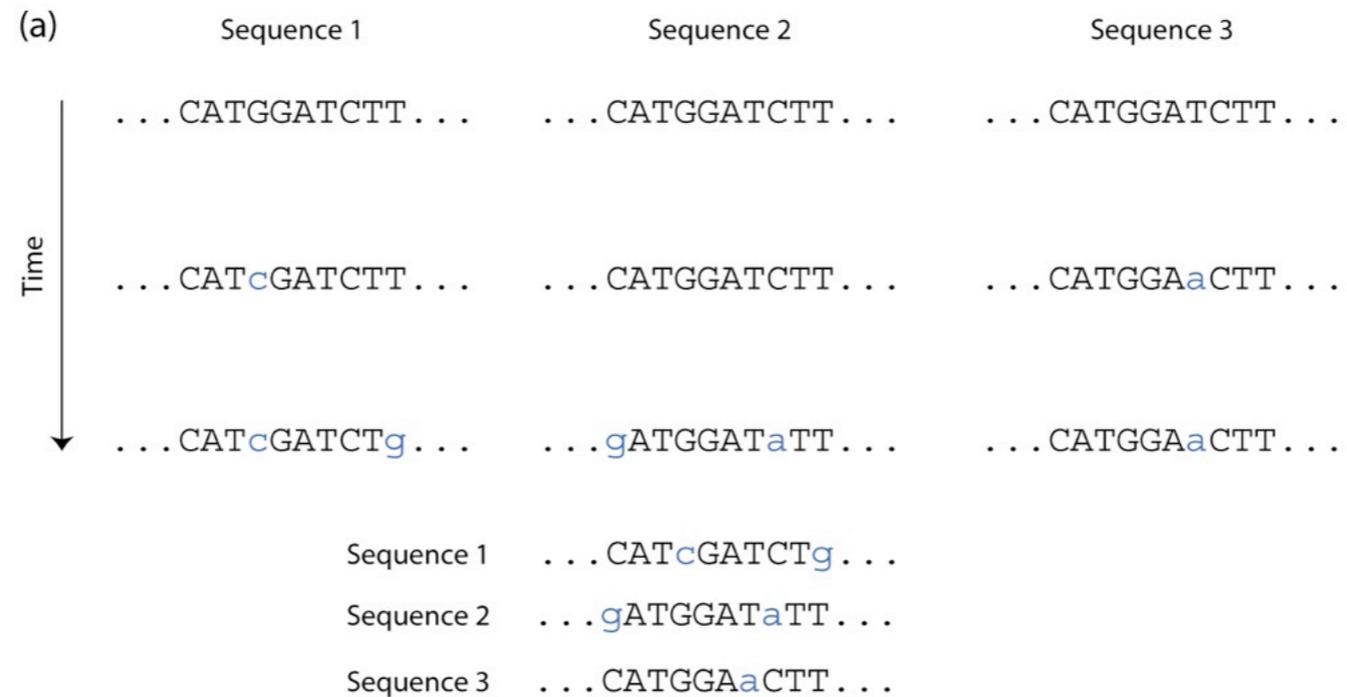
Mutation models for DNA sequences

- * The **infinite sites model**: mutation acts on infinitely long DNA sequences where each mutation occurs at a different position along the DNA sequence and the same position cannot experience a mutation more than once
- * The **finite sites model**: mutation acts on DNA sequences of finite length so that the same site may experience a mutation more than once

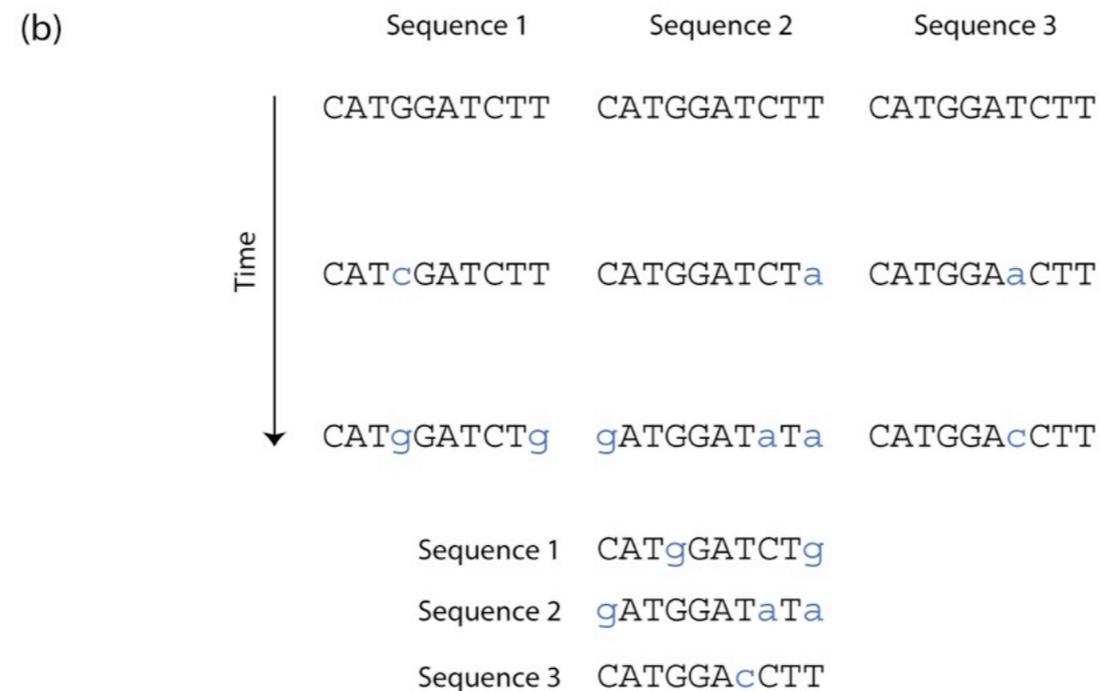
(3) Mutation Models

Mutation models for DNA sequences

Infinite sites model



finite sites model



(4) The Influence of Mutation on Allele Frequency and Autozygosity

- * We will focus on the impact of constant mutation on allele frequencies in a single panmictic population that is very large
- * Consider one locus with two alleles A (frequency p) and a (frequency q)
- * Assume an **irreversible mutation model**: all mutations change A alleles into a alleles, but does not change a alleles into A alleles

(4) The Influence of Mutation on Allele Frequency and Autozygosity

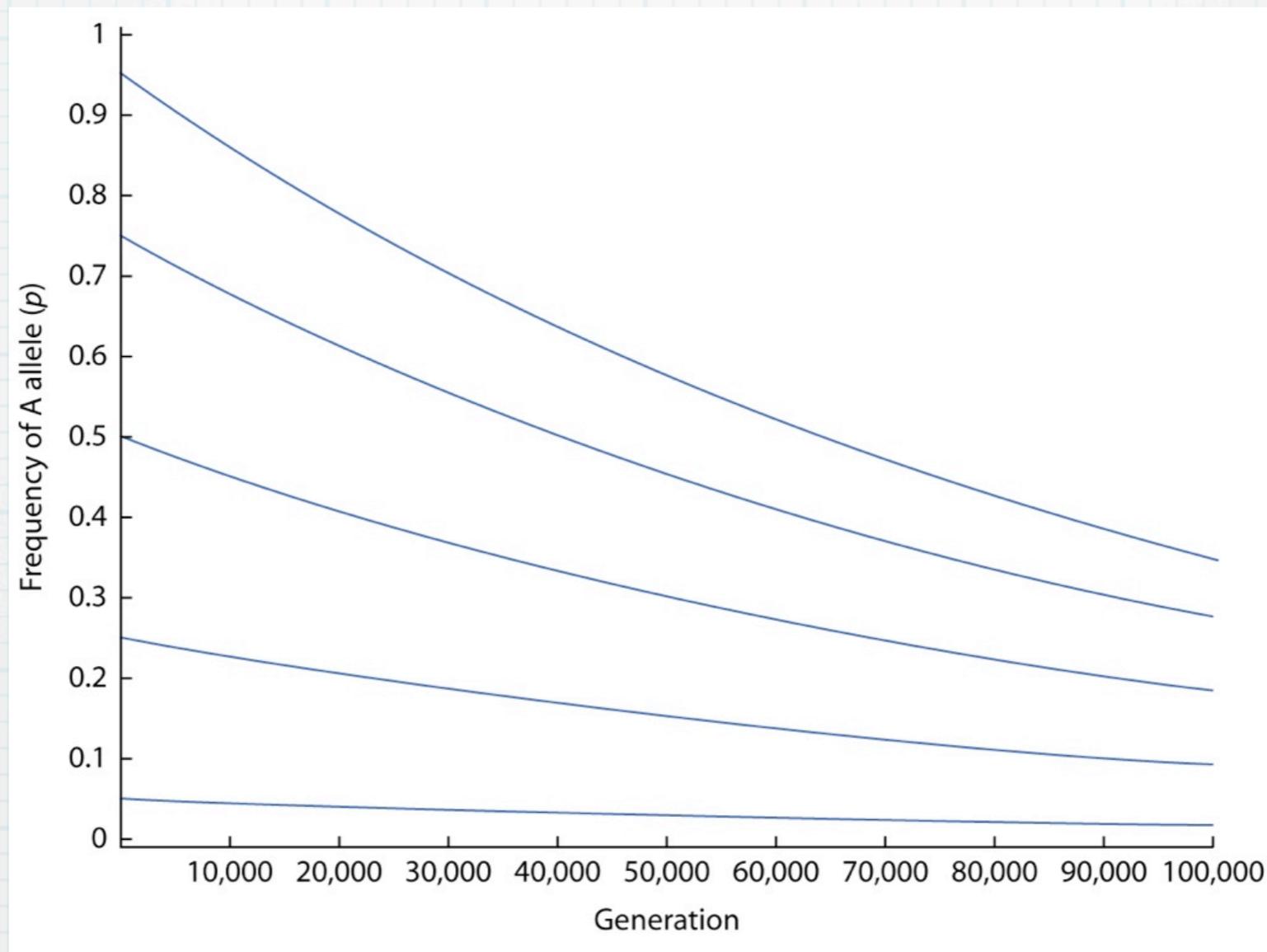
- * Assuming that the chance that mutation changes A into a in a generation is μ , the frequency of the A allele after one generation of mutation is

$$p_{t+1} = p_t(1 - \mu)$$

- * After an arbitrary number of generations:

$$p_t = p_0(1 - \mu)^t$$

(4) The Influence of Mutation on Allele Frequency and Autozygosity



$$\mu = 1 \times 10^{-5}$$

(4) The Influence of Mutation on Allele Frequency and Autozygosity

* Let's now consider the **reversible mutation model**: rate of forward mutation (μ) and reverse mutation (ν)

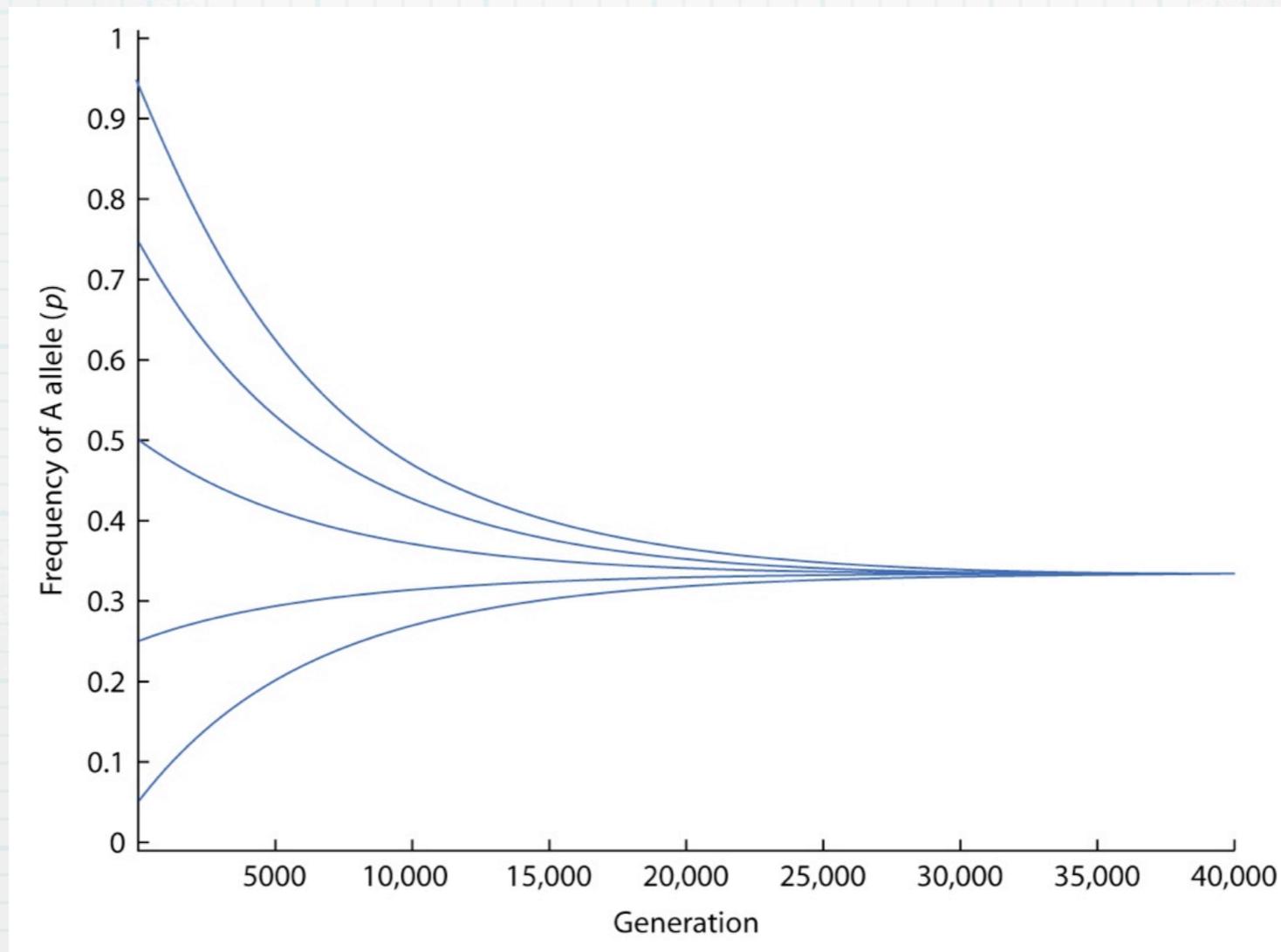
* The allele frequency of A after one generation is

$$p_{t+1} = p_t(1 - \mu) + (1 - p_t)\nu$$

* The equilibrium value of the frequency of A is

$$p_{\text{equilibrium}} = \frac{\nu}{\mu + \nu}$$

(4) The Influence of Mutation on Allele Frequency and Autozygosity



$$\mu = 1 \times 10^{-4} \qquad \nu = 5 \times 10^{-5}$$
$$\frac{\nu}{\mu + \nu} = 0.3333$$

(4) The Influence of Mutation on Allele Frequency and Autozygosity

- * Notice that allele frequency change due to mutation under the irreversible mutation model is identical to allele frequency change due to gene flow under the continent-island model
- * Further, allele frequency change due to mutation under the reversible mutation model is identical to allele frequency change due to gene flow under the two-island model

(4) The Influence of Mutation on Allele Frequency and Autozygosity

- * Using these observations, we can derive a formula for the autozygosity

$$F_t = \frac{1}{2N_e}(1 - \mu)^2 + \left(1 - \frac{1}{2N_e}\right)(1 - \mu)^2 F_{t-1}$$

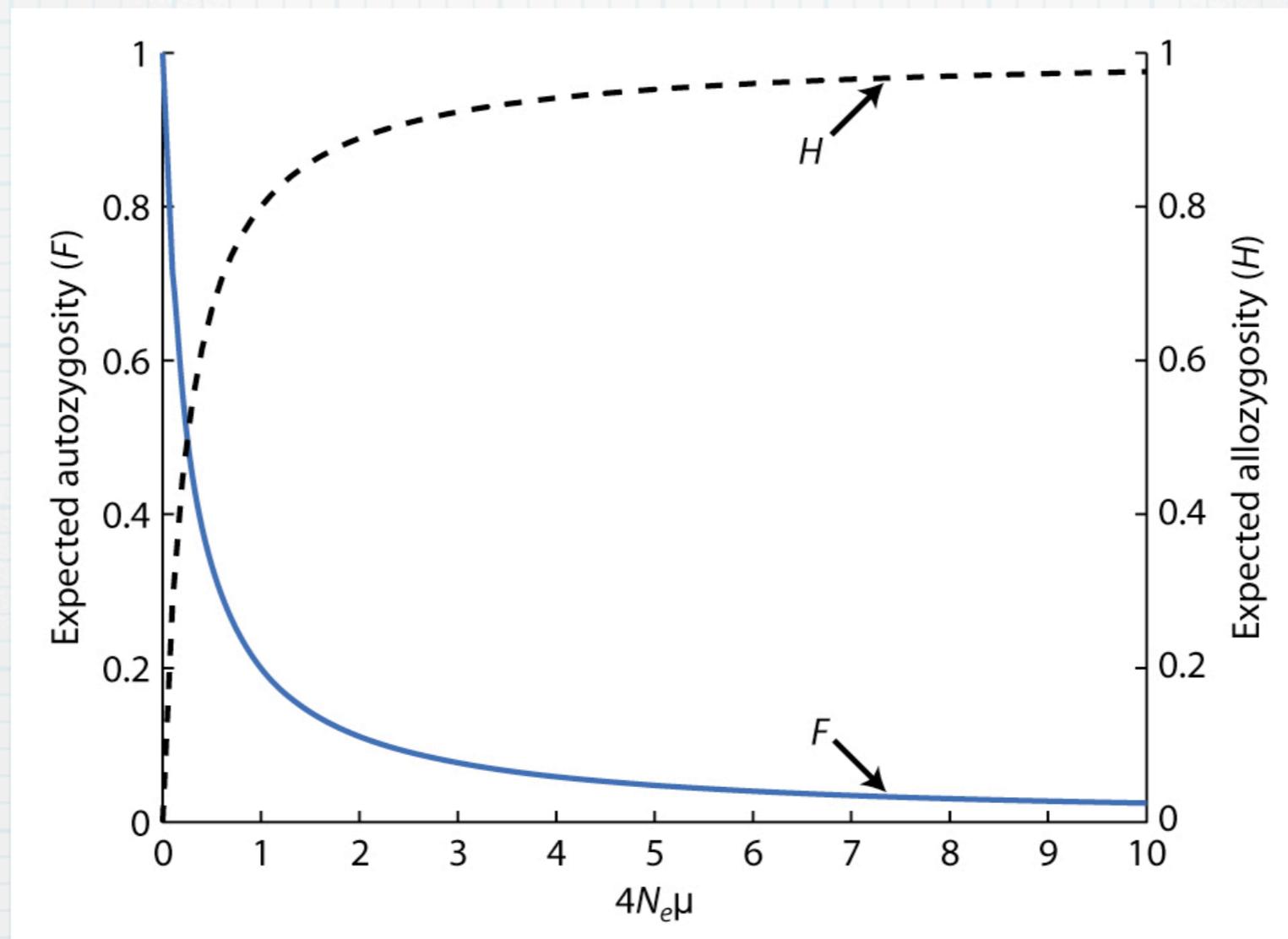
- * Assuming that the mutation rate is small and much, much less than N_e , an approximation for the expected amount of autozygosity at equilibrium in a finite population experiencing mutation is

$$F_{equilibrium} = \frac{1}{4N_e\mu + 1}$$

- * The expected heterozygosity is

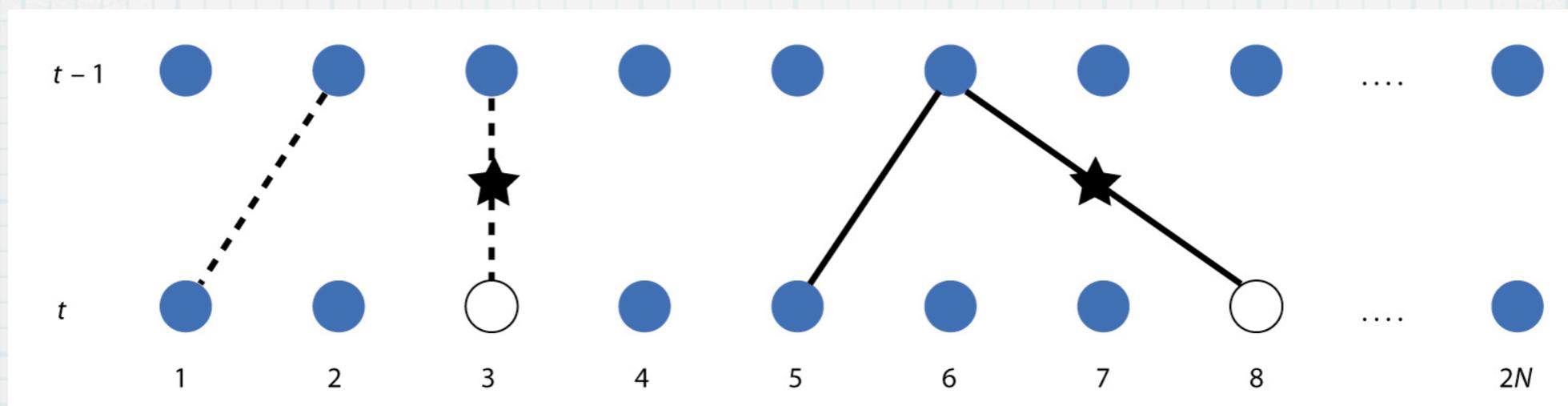
$$H_{equilibrium} = \frac{4N_e\mu}{4N_e\mu + 1}$$

(4) The Influence of Mutation on Allele Frequency and Autozygosity



(5) The Coalescent With Mutation

- * We will focus on coupling the process of coalescence and the process of mutation while moving back in time toward the MRCA



(5) The Coalescent With Mutation

- * We will assume that both coalescent and mutation events are rare, so that the two events are mutually exclusive
- * The chance that t generations pass before a mutation event occurs is

$$P(T_{mutation} = t) = (1 - \mu)^{t-1} \mu$$

which can be approximated as

$$P(T_{mutation} = t) = e^{-t \frac{\theta}{2}}$$

where

$$\theta = 4N_e \mu$$

(5) The Coalescent With Mutation

- * The total chance of mutation for k lineages is approximated as

$$P(T_{mutation} = t) = e^{-t \frac{\theta}{2} k}$$

- * The chance that a mutation occurs in one of k lineages at or before a certain time is approximated as

$$P(T_{mutation} \leq t) = 1 - e^{-t \frac{\theta}{2} k}$$

(5) The Coalescent With Mutation

- * Accounting for both types of events, we have

$$P(T_{event} \leq t) = 1 - e^{-t \left[\frac{\theta}{2}k + \frac{k(k-1)}{2} \right]}$$

- * The total chance that the event is a coalescence or a mutation is

$$\frac{\theta}{2}k + \frac{k(k-1)}{2}$$

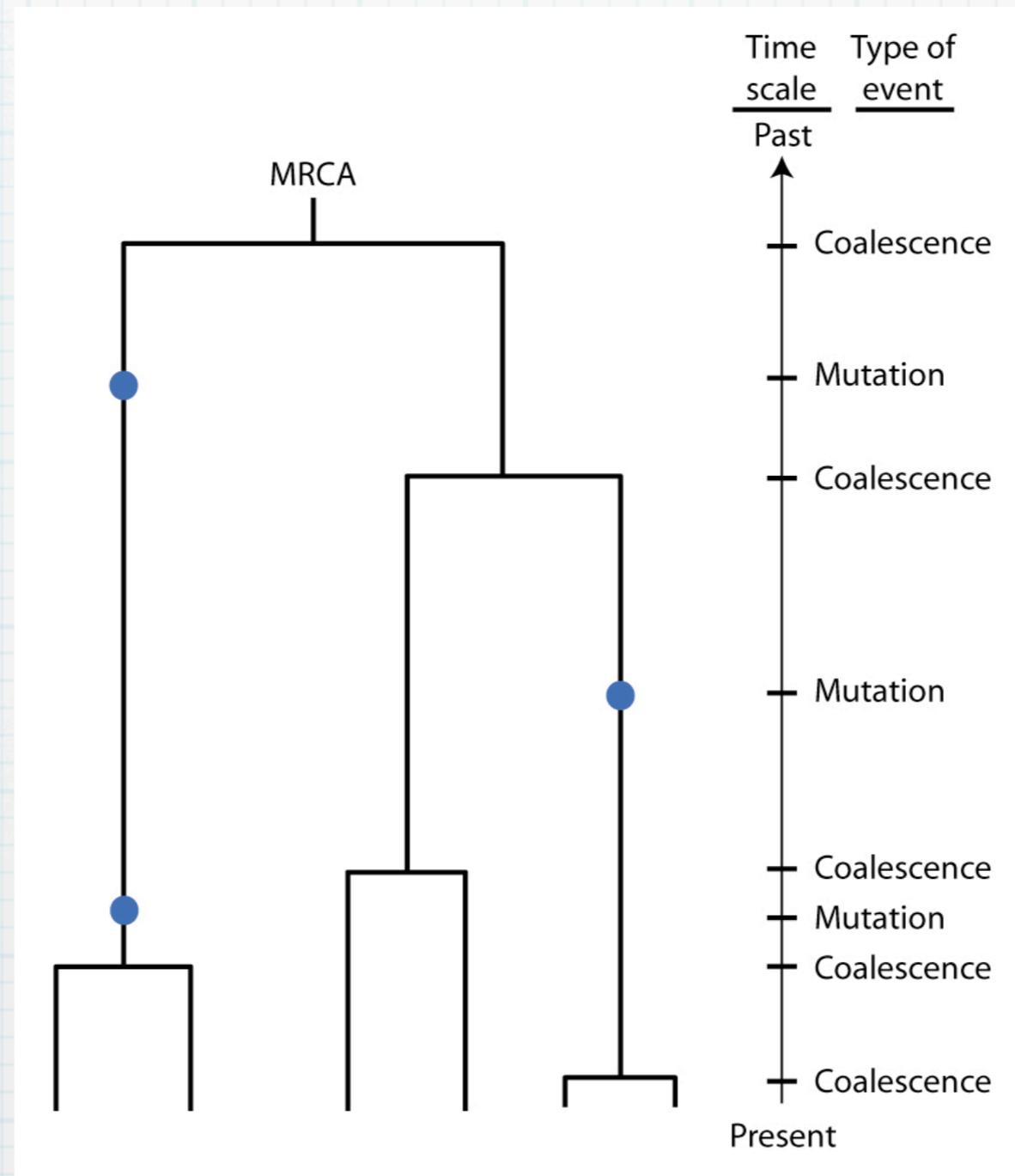
- * The chance that the event is a mutation is

$$\frac{\frac{\theta}{2}k}{\frac{\theta}{2}k + \frac{k(k-1)}{2}} = \frac{\theta}{k-1 + \theta}$$

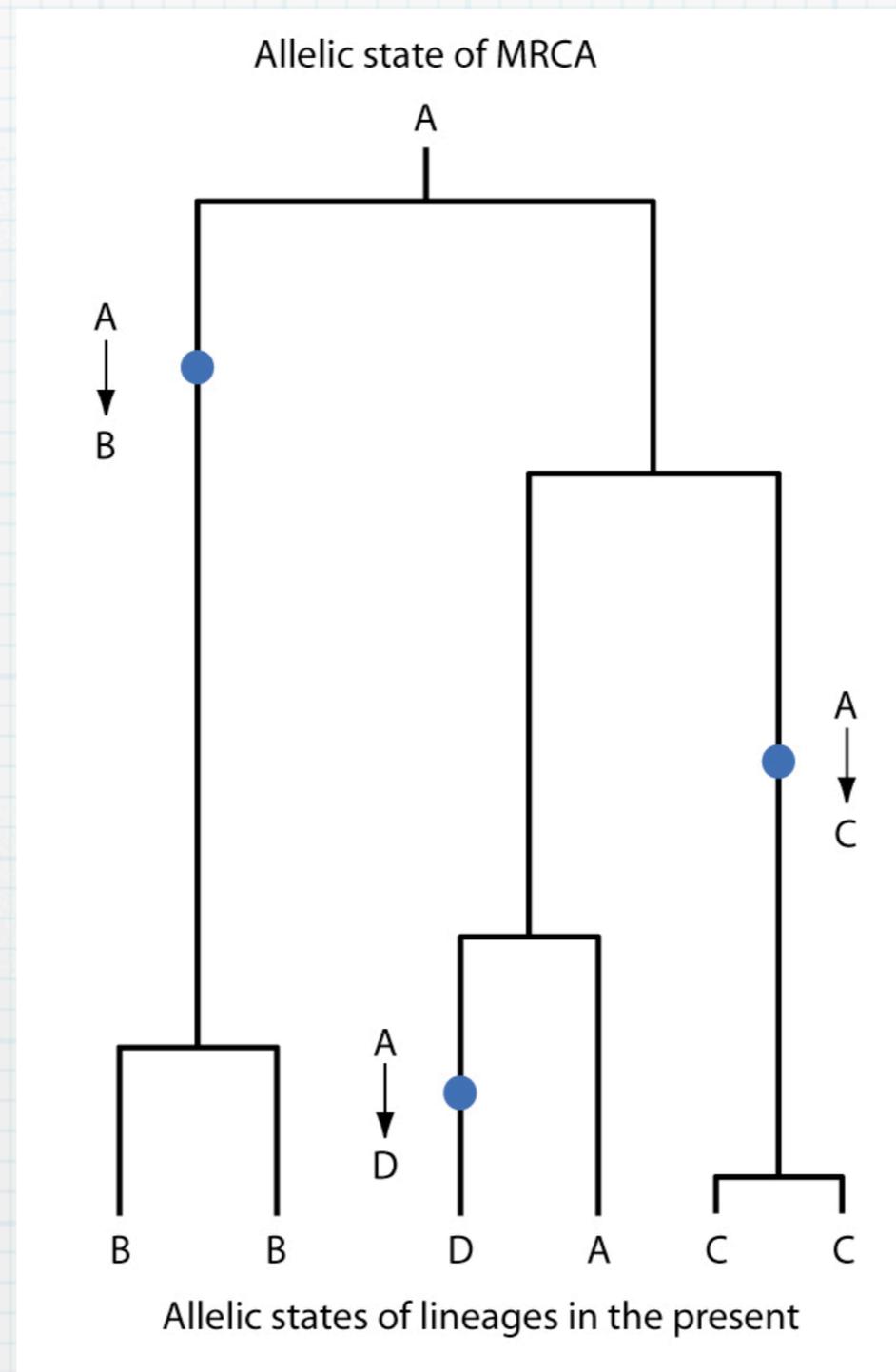
- * The chance that the event is a coalescence is

$$\frac{\frac{k(k-1)}{2}}{\frac{\theta}{2}k + \frac{k(k-1)}{2}} = \frac{k-1}{k-1 + \theta}$$

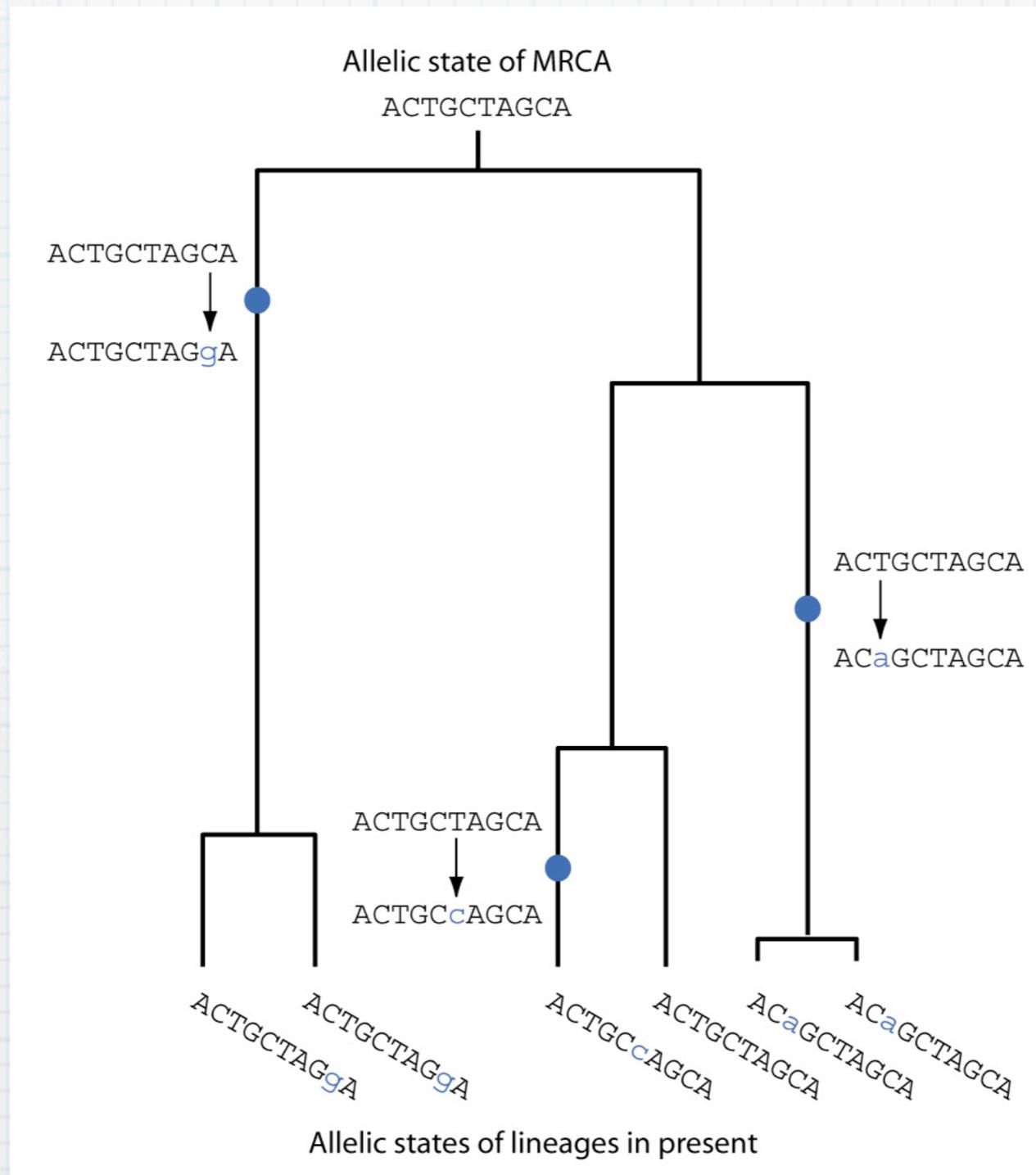
(5) The Coalescent With Mutation



(5) The Coalescent With Mutation



(5) The Coalescent With Mutation



Summary

- * The spectrum of relative fitness for genotypes containing mutations expresses the frequencies of mutations with a range of fitness effects
- * New mutations may be lost simply by Mendelian segregation
- * New selectively neutral mutations have a chance of fixation equal to their initial frequency
- * Fisher's geometric model of mutation shows that mutations with small effects on phenotype are more likely to be fixed by natural selection
- * The combination of mutation, genetic drift, and natural selection in genomes where recombination is absent or restricted leads to accumulation of deleterious mutations (Muller's Ratchet)

Summary

- * The infinite alleles model assumes mutations always create novel alleles
- * The infinite sites model assumes that each mutation changes a DNA site that didn't experience mutation before
- * Irreversible mutations will eventually lead to the loss of the original allele, whereas that's not the case under the reversible mutation model
- * The coalescent model can be augmented to with mutations to yield predictions for the number and frequency of alleles expected under the process or processes that produced the genealogy