

Models of Genetic Drift

- ❖ Binomial Distribution
- ❖ Markov Chains
- ❖ Diffusion Approximation (pp. 67-73)
 - ❖ based on continuous rather than discrete variables (i.e., time and allele frequencies)

Wright-Fisher transition probabilities

- ❖ binomial results in a matrix of transition probabilities that can be multiplied to predict the distribution of allele frequencies at any number of generations in the future

		to j								
		1	0	0	0	0	0	0	0	0
matrix for $2N = 8$	from i	0.344	0.393	0.196	0.056	0.010	0.001	0	0	0
		0.100	0.267	0.311	0.208	0.087	0.023	0.004	0	0
		0.023	0.112	0.235	0.282	0.211	0.101	0.030	0.005	0
		0.004	0.031	0.109	0.219	0.273	0.219	0.109	0.031	0.004
		0	0.005	0.030	0.101	0.211	0.282	0.235	0.112	0.023
		0	0	0.004	0.023	0.087	0.208	0.311	0.267	0.100
		0	0	0	0.001	0.010	0.056	0.196	0.393	0.344
		0	0	0	0	0	0	0	0	1
		0	0	0	0	0	0	0	0	0

Predicted distributions of allele frequencies in replicate populations of $N = 16$

$$P_{i \rightarrow j} = \frac{(2N)!}{j!(2N-j)!} p^j q^{2N-j}$$

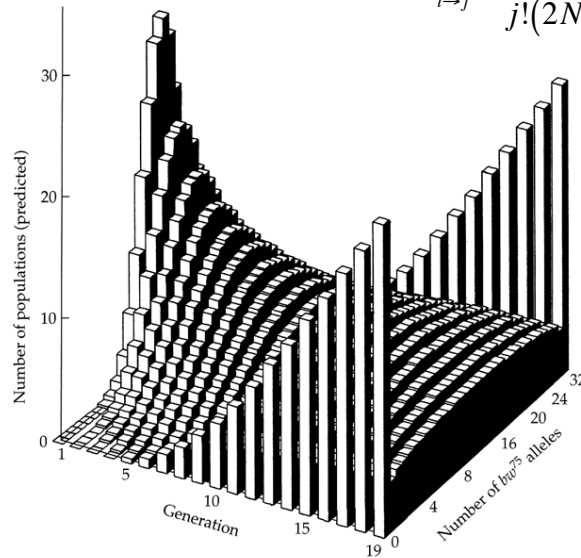
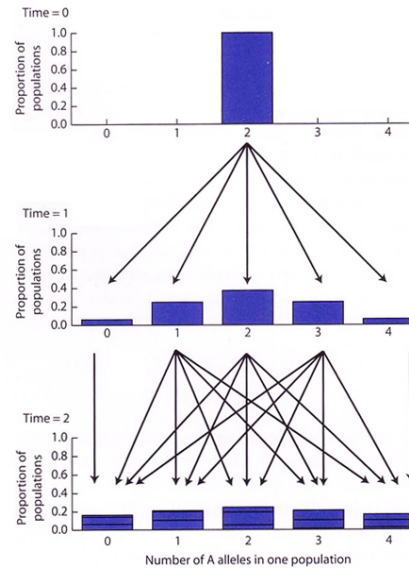


Figure 3.10 Genetic drift modeled by a Markov chain. In this case, the sample size is two diploid genotypes ($2N = 2$) or four gametes per generation. Initial allele frequencies in all populations are $p = q = 0.5$. In one generation, sampling error shifts some proportion of the initial populations that contain two copies of each allele to states of zero, one, two, three, or four copies of one allele. Between generations one and two, sampling error again shifts some proportion of the initial populations to states of zero, one, two, three, or four copies of one allele. However, in generation one there are populations present with all allelic states. The arrows represent the possible allelic states produced by sampling error in the third generation for each of the states in the second generation. The bars in the histogram for the third generation are divided by horizontal lines to show the contributions of each second generation allelic state to the total frequency of populations with a given allelic state (some contributions are very small and are difficult to see). As the Markov process continues, the frequency distribution accumulates more and more of the populations at states of zero and four alleles, eventually reaching fixation or loss for all populations.



Loss of Genetic Diversity due to Drift

- ❖ Pg. 66
- ❖ "...rate at which genetic variation is lost from the collection of many populations is $1-1/2N$ "
- ❖ put another way, in the next generation, genetic variation will be only $(1-1/2N)$ times as great as in the previous generation
- ❖ same applies for many loci within one population

The Diffusion Approximation

$$\begin{aligned} \phi(p,x;t + \Delta t) - \phi(p,x;t) = & \\ -[M(x)\phi(p,x;t) - M(x - \Delta x)\phi(p,x - \Delta x;t)] & \\ + \frac{1}{2} \{ [V(x + \Delta x)\phi(p,x + \Delta x;t) - V(x)\phi(p,x;t)] & \\ - [V(x)\phi(p,x;t) - V(x - \Delta x)\phi(p,x - \Delta x;t)] \} & \end{aligned}$$

- ❖ "the discussion may seem quite mysterious, but please do not be discouraged because a detailed understanding is not necessary to understand the rest of the chapter or anything later in the book"

Expectations based on the diffusion approximation

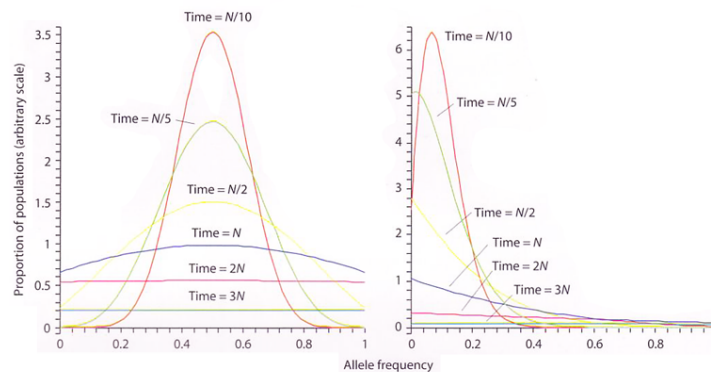


Plate 3.13 Probability densities of allele frequency for many replicate populations predicted using the diffusion equation. The initial allele frequency is 0.5 on the left and 0.1 on the right. Each curve represents the probability that a single population would have a given allele frequency after some interval of time has passed. The area under each curve is the proportion of alleles that are not fixed. Time is scaled in multiples of the effective population size, N . Both small and large populations have identically shaped distributions, although small populations reach fixation and loss in less time than large populations. The populations that have reached fixation or loss are not shown for each curve.

Insights...

- ❖ time to fixation and time to loss

$$\bar{T}_{fix} = -4N \frac{(1-p) \ln(1-p)}{p}$$

$$\bar{T}_{loss} = -4N \frac{p \ln(p)}{1-p}$$

- ❖ persistence time (for polymorphism)

$$\bar{T}_{segregate} = -4N [p \ln(p) + (1-p) \ln(1-p)]$$

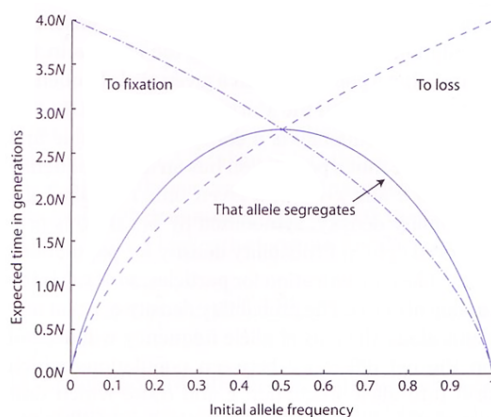


Figure 3.14 Average time that an allele segregates, takes to reach fixation, or takes to reach loss depending on its initial frequency when under the influence of genetic drift alone. Alleles remain segregating (persist) for an average of $2.8N$ generations when their initial frequency is $1/2$. Fixation or loss takes up to an average of $4N$ generations when alleles are initially very rare or nearly fixed, respectively. Since these are average times, alleles in individual populations experience longer and shorter fixation, loss, and segregation times. Time is scaled in multiples of the population size.

Insights...

- ❖ probability of fixation for a new, neutral mutation = $1/(2N)$
 - ✧ time to fixation = $\sim 4N$
- ❖ probability of loss = $1 - (1/(2N))$
 - ✧ time to loss = $\sim 2 \ln(2N)$
- * these are *average* times!

Effective population size N_e

- ❖ the number of individuals in a theoretically ideal population having the same magnitude of random genetic drift as the actual population
- ❖ for most natural populations, N_e is **smaller** than the **census** size
 - ❖ often substantially smaller!

Factors affecting N_e

- ❖ changes in population size across generations
 - ❖ N_e = harmonic mean population size

$$N_e = t / \left(\frac{1}{N_0} + \frac{1}{N_1} + \dots + \frac{1}{N_{t-1}} \right)$$

- ❖ e.g.,

$$N_e = 4 / \left(\frac{1}{1000} + \frac{1}{2000} + \frac{1}{20} + \frac{1}{1000} \right) = 76.2$$

Factors affecting N_e

- ❖ sex ratio (of breeding individuals)

- ❖ “bottleneck” through the less numerous sex

$$N_e = 4 \frac{N_m N_f}{N_m + N_f}$$

- ❖ e.g.,

$$N_e = 4 \times \frac{10 \times 100}{10 + 100} = 36.4$$



Factors affecting N_e

- ❖ variation in “family size” (i.e., fitness)

- ❖ What is N_e in a haploid population in which every individual leaves one offspring?

- ❖ Diploid?

$$N_e = \frac{4N_{t-1}}{\text{var}(k) + \bar{k}^2 - \bar{k}}$$

- ❖ where k is family size for a pair of diploid individuals

- ❖ what is k for a stable population?

- ❖ poisson distribution: mean = variance

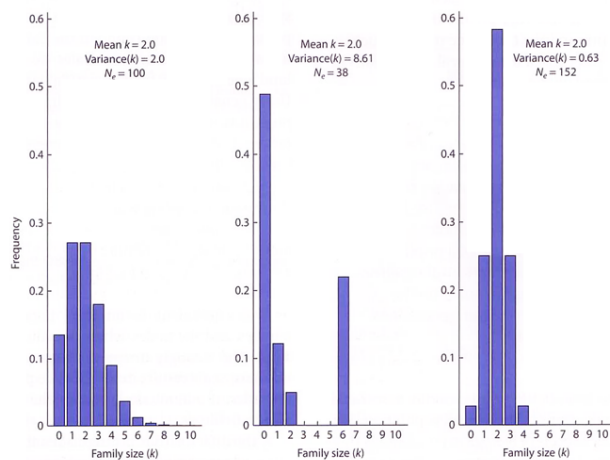


Figure 3.17 Distributions of family size. The variance equals the mean as expected for a Poisson distribution on the left. However, the center distribution has a few families that are very prolific while 75% of the families produce two or fewer progeny with most individuals failing to reproduce. The distribution on the right has less variance in family size than expected for a Poisson distribution with most families of size two. The Poisson distribution is taken as the standard with an effective size of 100. By comparison, the center distribution has a smaller effective population size and the distribution on the right a larger effective population size.

Factors affecting N_e

❖ sex chromosomes

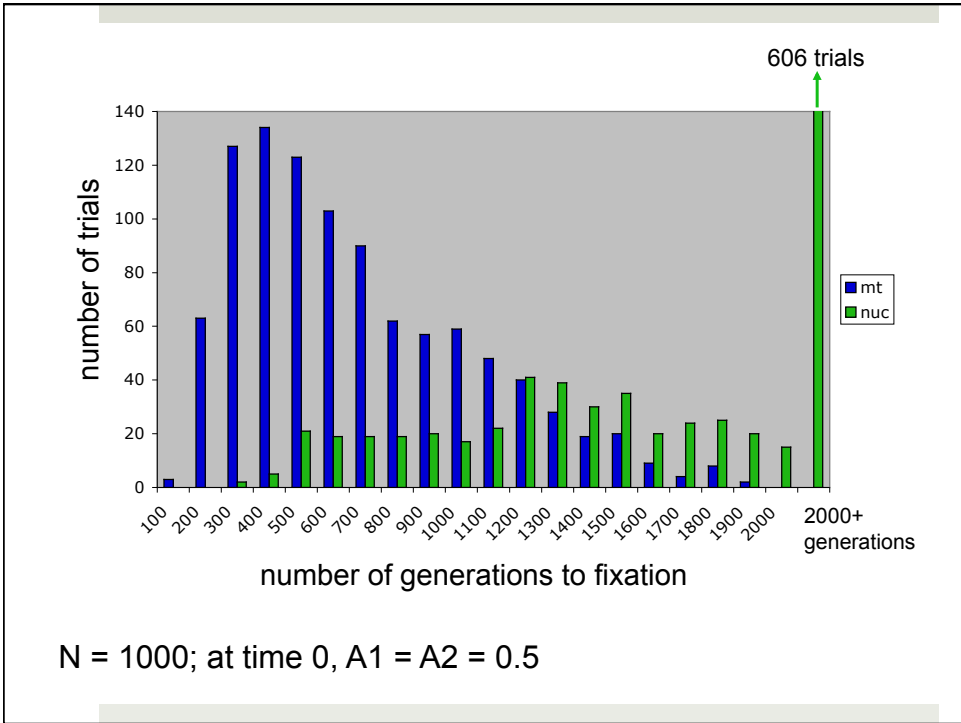
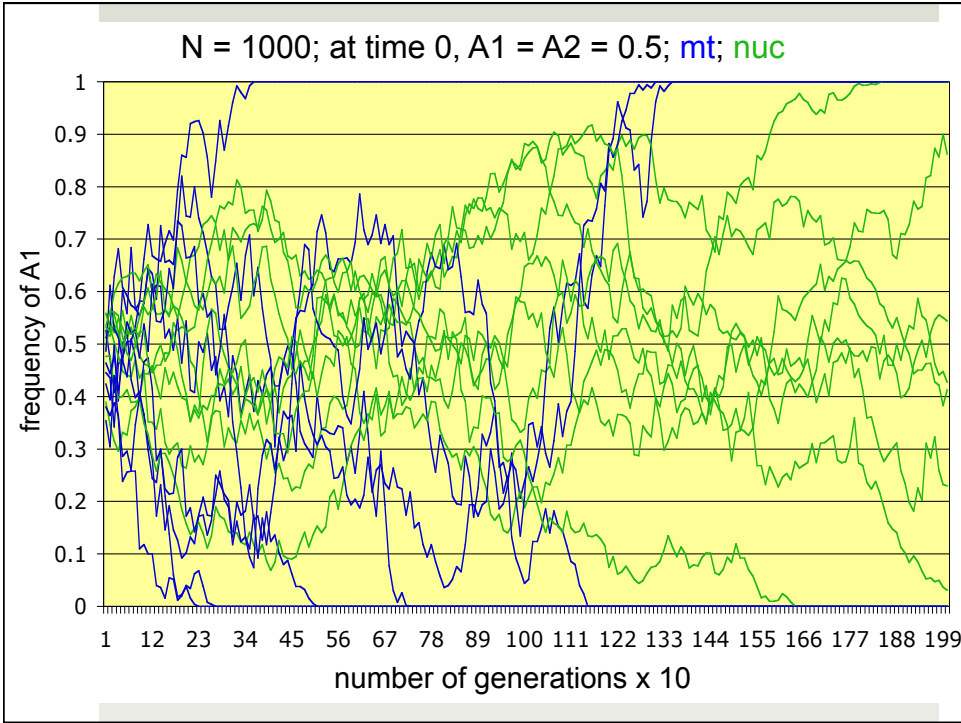
❖ the “population” size of sex chromosomes is smaller than for autosomes

❖ N_e for X-linked genes:

$$N_e = \frac{9N_m N_f}{4N_m + 2N_f}$$

❖ if $N_m = N_f$, then $N_e = 3N/4$

❖ mtDNA and Y-chromosome $N_e = N/4$



Factors affecting N_e

- ❖ population structure with gene flow
 - ❖ population subdivision maintains relatively greater genetic diversity (slows the process of drift to fixation in the overall population)

$$N_e = ND \left(1 + \frac{1}{4Nm} \right)$$

- ❖ ...where N is the population size in each of D demes and m is the migration rate between demes

Factors affecting N_e

- ❖ breeding sex ratio
- ❖ fluctuation in population size
- ❖ variance in family size (or # offspring)
- ❖ number of gene copies
 - ❖ sex chromosomes, mtDNA
- ❖ population structure