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

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

  - e.g.,

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

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?

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

  - where k is family size for a pair of diploid individuals
  - what is k for a stable population?
    - poisson distribution: mean = variance

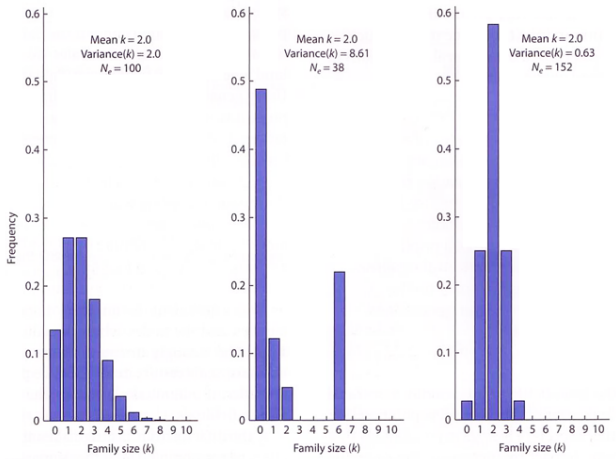


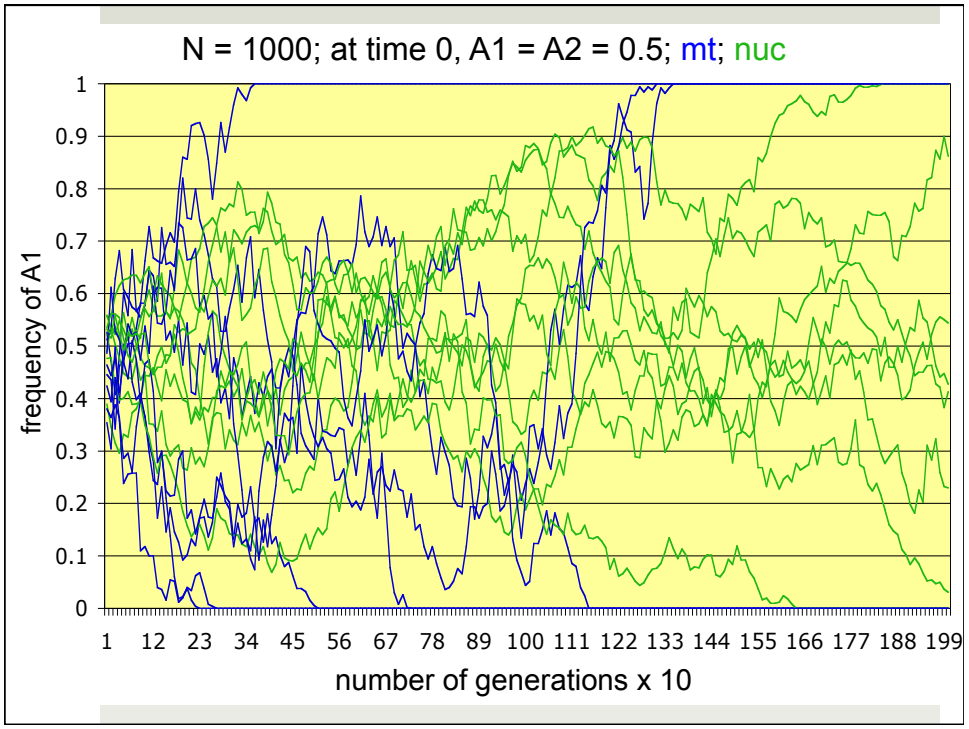
Figure 3.17 Distributions of family size. The variance equals the mean as expected for a Poisson distribution on the left. However, the center distribution has a few families that are very prolific while 75% of the families produce two or fewer progeny with most individuals failing to reproduce. The 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.

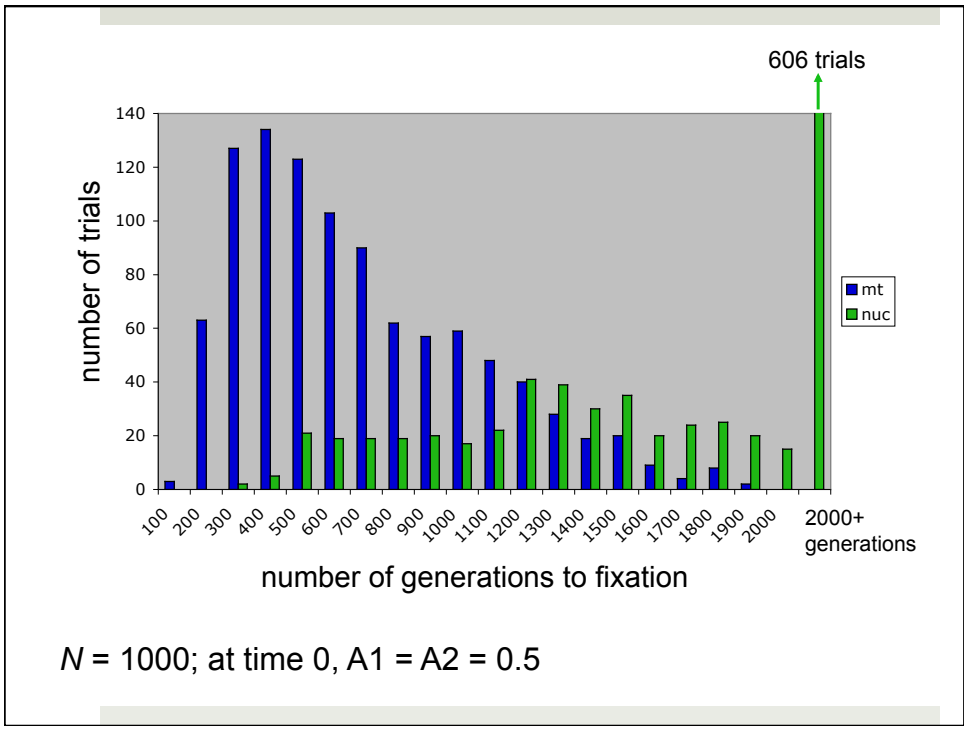
Factors affecting N_e

- sex chromosomes
  - the "population" size of sex chromosomes is smaller than for autosomes
  - N_e for X-linked genes:

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

  - if $N_m = N_f$, then $N_e = 3N/4$
  - mtDNA and Y-chromosome $N_e = N/4$
    - assuming equal variance in RS between males and females





N = 1000; at time 0, A1 = A2 = 0.5

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)

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

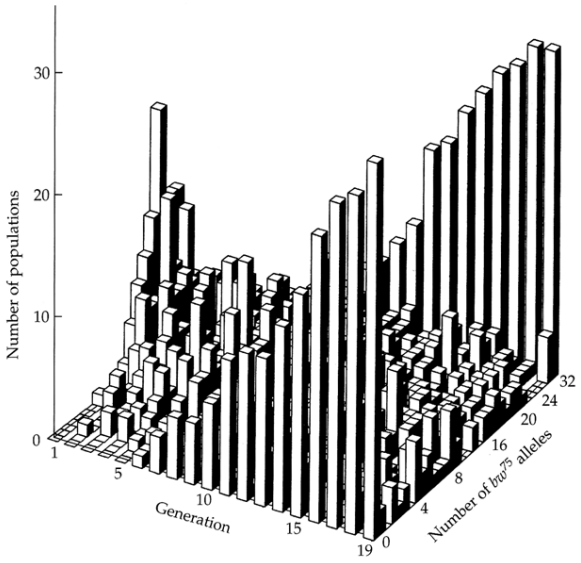
  - …where N is the population size in each of D demes and m is the migration rate between demes

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

Fig. 3.4 Random genetic drift in 107 *Drosophila* populations; 8 males, 8 females each generation

Why different from ideal expectations ?



"Identical by Descent"

- concept used to derive theoretical expectations for N_e
- Hartl & Clark develop the concept using a structured population model (Fig. 3.9)

- Fig. 3.9

Overall average allele frequency stays the same but heterozygosity declines

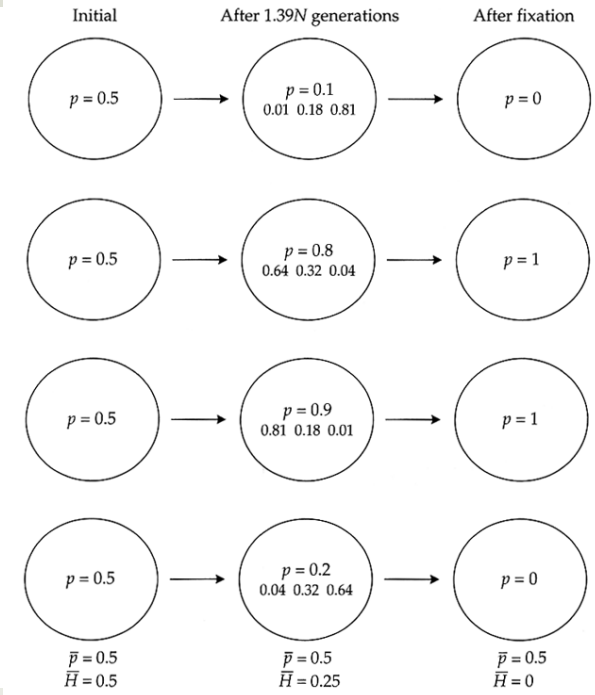
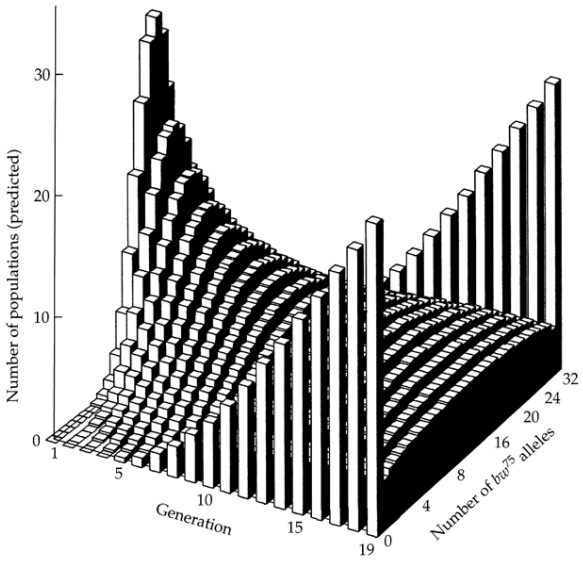


Fig. 3.5 Predicted distributions of allele frequencies in replicate populations of *N* = 16

Fig. 3.9 (last slide) is showing the exact same process as this figure



"Identical by Descent"

- what is the probability that two randomly sampled alleles are identical by descent (i.e., "replicas of a gene present in a previous generation")?
  - Wright's "fixation index" F
- at the start of the process (time 0), "declare" all alleles in the population to be unique or unrelated, $F_t = 0$ at $t = 0$
- in the next generation, the probability of two randomly sampled alleles being copies of the same allele from a single parent = $1/(2N)$, so…

"Identical by Descent"

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

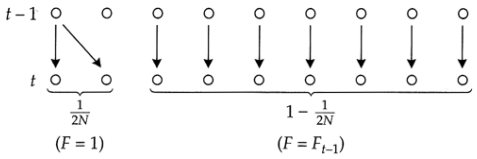


FIGURE 3.10 Diagram illustrating the reasoning behind the recursion for F in a finite population. When the gametes are drawn to make up the population at generation t, there is a chance $1/(2N)$ that any pair of alleles will have been identical in generation $t - 1$. If this happens, the probability of identity is 1. For the allele pairs drawn in generation t from two distinct alleles at generation $t - 1$ [the probability of this happening is $1 - 1/(2N)$], the probability of identity is F_{t-1}. Adding the probabilities of these two events, we get $F_t = 1/(2N) + [1 - 1/(2N)]F_{t-1}$.

"Identical by Descent"

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

= probability that alleles are copies of the same gene from the immediately preceding generation *plus* the probability that the alleles are copies of the same gene from an earlier generation

or

$$F_t = 1 - \left(1 - \frac{1}{2N}\right)^t$$

assuming $F_0 = 0$

compare to: mean time to fixation for new mutant = $\sim 4N$

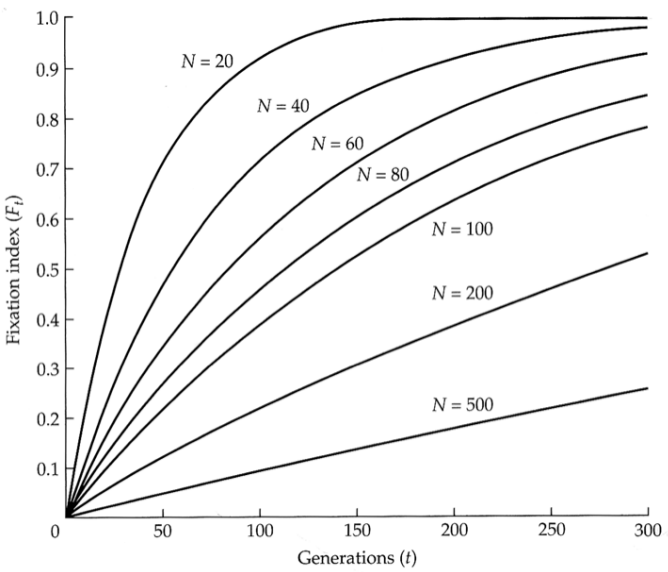
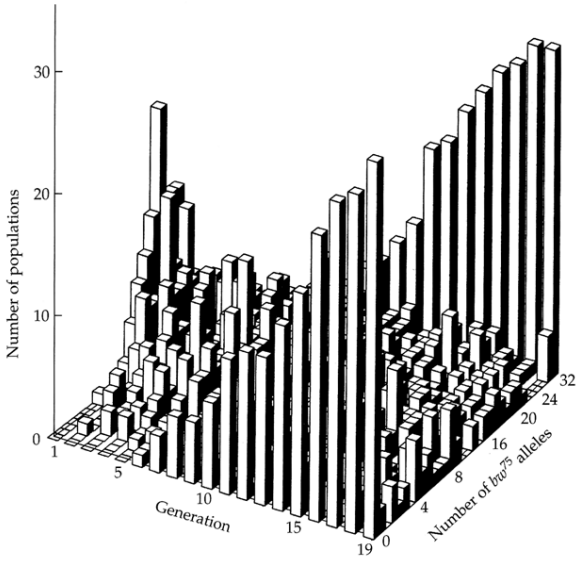
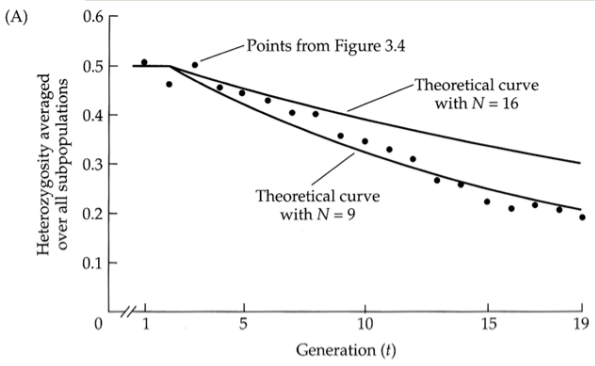


FIGURE 3.11 Increase of F_t in ideal populations as a function of time and effective population size N.

Random genetic drift in 107 *Drosophila* populations;
8 males, 8 females each generation





$$H_t = \left(1 - \frac{1}{2N_e}\right)^t H_0$$

$$\approx H_0 e^{-t/2N_e}$$

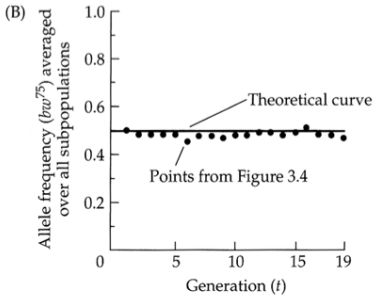


FIGURE 3.12 Theoretical curves for average heterozygosity among subpopulations (A) with $N = 9$ or $N = 16$, along with actual values (plotted as points) from the experiment in Figure 3.4. Part (B) shows the theoretically expected average allele frequency among the 107 subpopulations and the observed average. (Data from Buri 1956.)

Deriving the population size effect

- Start here: $$F_t = \frac{1}{2N} + \left(1 - \frac{1}{2N}\right) F_{t-1}$$
- Finish here: $$N_e = t \Big/ \left(\frac{1}{N_0} + \frac{1}{N_1} + \ldots + \frac{1}{N_{t-1}} \right)$$
- other effects on N_e are similarly based on the logic of identity by descent

(Hartl & Clark: pp. 116-122)