Genetic Drift

- **Genetic drift is the random, stochastic change in allele frequency from one generation to the next due to sampling errors in a finite population.**

I) Genetic Drift Introduction

A) Probability of allele frequency change.

- **For a small, finite population the probability of changing significantly in allele frequency is larger than in a large population.**

Remember binomial expansion:

$$\mathrm{Pr}(i) = \frac{2N!}{i!(2N-i)!} p^{i} q^{2N-i}$$

The probability of fixation of a allele (i.e., i = 2N):

$$\mathrm{Pr}(2N) = \frac{\cancel{2N}!}{\cancel{2N}!(2N-2N)!} p^{2N} q^{2N-2N}$$

$$= p^{2N}$$

Population Size (N)

	1	2	4	100
Probability of Fixation	25%	6.25%	0.39%	6.223×10^{-60}%

Selection Simulations: http://gsoft.smu.edu/GSoft.html

PopBio Simulations: http://academic.reed.edu/biology/Software.html

These are also available on the web page as links to the programs (sorry, I think they are Macintosh only).

B) Drift as Inbreeding – the overall effect of drift can be viewed as inbreeding or the probability of autozygosity.

- **N individuals contribute 2 gametes to the next generation**
- **Drawing 2 alleles from the gamete pool with replacement, there are two ways to produce an autozygous zygote**
  - **the probability of drawing the same allele twice is $\frac{1}{2N}$.**
  - **The probability of drawing two different alleles is $1 - \frac{1}{2N}$. The probability that the different allele is identical by descent (IBD) to the first is F (i.e., the inbreeding coefficient we discussed earlier).**
- **The probability then that two alleles drawn from a pool are IBD is:**

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

The probability of not being autozygous is then:

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

Or for any generation t:

$$1 - F_t = \left(1 - \frac{1}{2N}\right)^t \left(1 - F_0\right)$$

In an Infinite Allele Model the probability of not being IBD is also the probability of being heterozygous, so:

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

With drift, heterozygosity is reduced by $1 - \frac{1}{2N}$ every generation.

OH 12.1

An effective population size of at least 50 is necessary to maintain genetic variation with drift.

C) Drift as a Diversifying Force – the same action that results in drift acting to decrease heterozygosity within single populations, results in multiple, isolated (completely or partially) to diverge.

1) **If we start with many populations all at $p = 0.5$ and allow only drift, ultimately (regardless of population size – speed), half of the populations will fix for one allele and half will be fixed for the other. In other words, drift increases the variance between the populations over time.**

OH 12.2, 12.2A, 12.2B

D) Drift – Mutation Equilibrium – just as with selection, mutation and migration, there is an equilibrium between mutation and drift and the standard assumption becomes a Selection – Mutation – Migration – Drift Equilibrium.

OH 12.3

- **As we've seen before, if heterozygosity is decreased each generation regardless of the size of the population, all populations should be homozygous given enough time.**

Remember the probability of Autozygosity is:

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

Adding mutation:

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

Where $(1 - \mu)^2$ is the probability that neither, autozygous allele has mutated

An equilibrium exists between mutation and drift in that mutation acts to increase heterozygosity and drift acts to decrease it.

$$\hat{F} = \frac{(1-\mu)^2}{4N\mu + 1} \approx \frac{1}{4N\mu + 1} \quad \text{and} \quad \hat{H} = \frac{4N\mu}{4N\mu + 1}$$

II) Genetically Effective Population Size (N_e)

- **Census Population Size – number of individuals counted or estimated based on observed count number**
- **Evolutionary (or Genetic) Effective Population Size – the number of individuals in an theoretically ideal population that would have the same magnitude of genetic drift (or increase in inbreeding) as the observed population.**
- **Evolutionary Effective Size ≠ Census Size (usually <<).**

A) Three types of Evolutionary Effective Population Sizes

1) **Inbreeding Effective Size – this estimate uses the change in the inbreeding coefficient (F) to estimate the size.**
2) **Variance Effective Size – this estimate uses the change in the variance in allele frequency to estimate the size.**
3) **Eigenvalue Effective Size – this estimate uses the rate of loss of heterozygosity to estimate the size.**

B) Inbreeding Evolutionary Effective Population Size – Several different population parameters can make the EEPS < Census size.

1) **Different Sex Ratios – when the sex ratio is not equal, offspring can be half – sibs of each other.**
   - **Here N_e is the probability that two gamete alleles are from the same grandparent.**

Consider a population with equal sex ratio:

P(male) = 0.5 and P(female) = 0.5

In any generation, the probability that a zygote carries homologous chromosomes in the GRANDPARENT generation is:

P(male) and P(male) = P(0.5) X P(0.5) = 0.25

P(male) and P(female) = 2 X P(0.5) X P(0.5) = 0.5

P(female) and P(female) = P(0.5) X P(0.5) = 0.25

Given one, male gamete, the probability a copy of that gamete is selected again is: $\frac{1}{N_M}$. Given one female gamete, the probability a copy of that gamete is selected again is: $\frac{1}{N_F}$.

Combining the probabilities:

The probability that the homologous chromosomes forming a gamete are both from males and are both from the same male is: $\frac{1}{4N_M}$. The probability that the homologous chromosomes forming a gamete are both from females and are both from the same female is: $\frac{1}{4N_F}$.

$$\frac{1}{N_e} = \frac{1}{4N_F} + \frac{1}{4N_M}$$

$$N_e = \frac{4N_F N_M}{N_F + N_M}$$

OH 12.4

2) **Variation in Fecundity –The assumption in the Wright–Fisher model of genetic drift is that offspring are produced according to a Poisson distribution. When this is not the case, some individuals can be over represented in the next generation. In the special case of a diploid organism with constant population size with a mean contribution per individual is 2 gametes:**

$$N_e = \frac{4N - 2}{V_k + 2}$$ **where V_k is the variance in number of gametes.**

If the distribution is Poisson, then the probability that gametes come from a particular parent follows the binominal distribution and the variance in the number of gametes is:

$$\mathbf{V_k = \frac{2(N-1)}{N}}$$

Substituting we can see:

$$\mathbf{N_e = \frac{4N-2}{V_k + 2}}$$

$$\mathbf{= \frac{4N-2}{\frac{2(N-1)}{N} + 2} = \frac{4N-2}{\frac{4N-2}{N}} = (4N-2) \times \frac{N}{(4N-2)}}$$

$$\mathbf{N_e = N}$$

Alternatively if $V_k = 0.0$ (i.e., all individuals have exactly the same number of offspring as can be the case in some captive breeding programs):

$$\mathbf{N_e = \frac{4N-2}{2} = 2N-1}$$

$$\mathbf{N_e \approx 2N}$$

a) **Selection – Effective Population Size Paradox – when selection is operating, this is (by definition) a variance in the number of offspring contributed to the next generation so the effective populations size is reduced. The greater the variance in fecundity (or fitness) the smaller the effective population size. Therefore, as the strength of selection increases, so does the magnitude of the effect of drift.**

3) **Variation in the Population Size – with most of our models we dealt with a stable population size. When population size varies from one generation to the next, this will reduce the effective population size relative to the census size.**

$$N_e = \frac{t}{\sum_{i=1}^{t} \frac{1}{N_i}} \quad \text{(i.e., the harmonic mean)}$$

t	N
1	10,000
2	9,800
3	8,987
4	8,894
5	8,879
6	12,004
7	10
8	8,700
9	9,751
10	8,429
11	9,185
Sum	94,639

With t_7: AM = 8,603.55, HM = 108.84

Without t_7: AM = 9,462.90, HM = 9,374.82

Population size can change in two ways:

a) Bottleneck – populations experience a sever reduction in population size in the current range

OH 12.5

  i) Bottlenecks have an effect on heterozygosity only if very small population sizes are reached (i.e., ≤ 10) and long time to recover (i.e., low intrinsic rate of population growth)

  ii) greatest impacts on bottlenecks are on allelic diversity. Rare alleles are lost from the population. Rare alleles, however, contribution little to heterozygosity.

b) Founder Event – propagules leave the parental range and found a new, small population in a new geographic area.

  i) Essentially the same as above, only the new habitat can have unique selection profile. This can result in rapid adaptive changes ala Shifting Balance Theory.

III) Coalescent Theory

- **Genetic Drift and inbreeding can also be (and some times better so) dealt with using the coalescence of lineages in pedigrees.**

OH 12.6A

As before, the probability that two alleles forming a zygote are from the same individual in the previous generation is:

$$\text{P(2 alleles are IBD from t – 1)} = \frac{1}{2N}$$

$$\text{P(2 alleles not IBD from t – 1)} = 1 - \frac{1}{2N}$$

$$\text{P(3}^{\text{rd}}\text{ allele is not 1}^{\text{st}}\text{ or 2}^{\text{nd}}\text{)} = 1 - \left(\frac{1}{2N} + \frac{1}{2N}\right)$$

$$= \frac{2N-2}{2N} = 1 - \frac{2}{2N}$$

$$\text{P(3 alleles are not IBD from t – 1)} = \left(1 - \frac{1}{2N}\right)\left(1 - \frac{2}{2N}\right)$$

P(k alleles are not IBD from any generation)

$$= \prod_{i=1}^{k-1}\left(1 - \frac{i}{2N}\right) \approx 1 - \left(\frac{k!}{2!(k-2)!}\right)\frac{1}{2N}$$

P(k alleles coalesced t generations ago)

$$= \frac{1}{2N}\left[1 - \frac{1}{2N}\right]^{t}$$

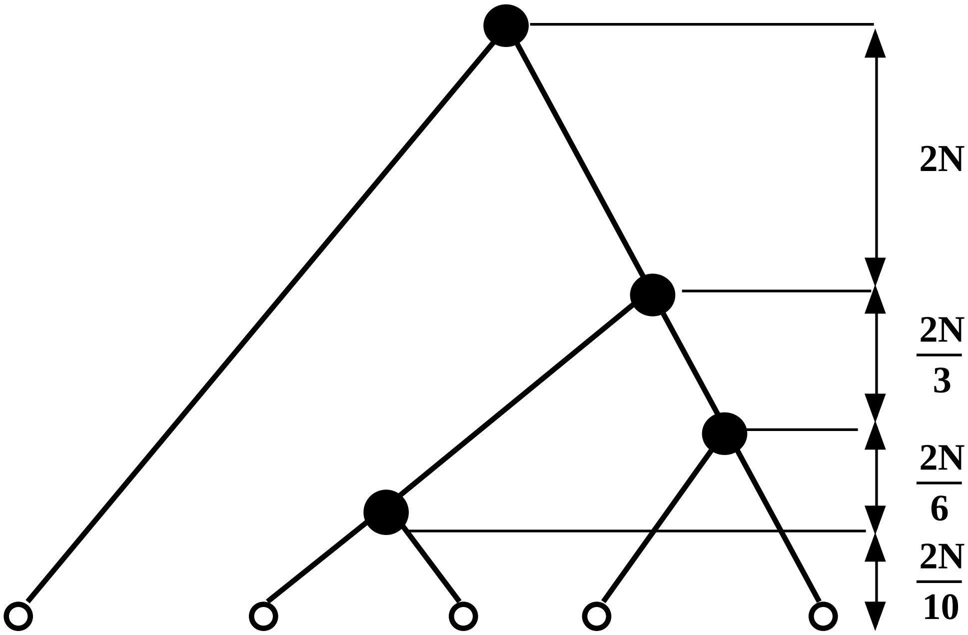
OR

the probability that two alleles were the same in any previous generation (i.e., coalesced) ($\frac{1}{2N}$) and were different for t generations $\left[1 - \frac{1}{2N}\right]^{t}$

The expected time for which there are k numbers of lineages is:

$$E(T_k) = \frac{4N}{k(k-1)}.$$

e.g., for 5 lineages:



The expected time to coalescence of all n lineages is:

$$E(t) = \sum_{i=2}^{k} E(T_i) = 2N\left(1 - \frac{1}{k}\right)$$

e.g., for 5 lineages = 3.2N generations

This process works as well for alleles as it does for lineages

OH 12.6A, 12.6B