0   Inbreeding and Population Structure.

[This chapter is just barely started and does not yet have a particular "case study" associated with it. But I'm going to post this anyway in case you find this useful for reviewing your lecture notes.]

The Hardy Weinberg Equilibrium was defined for the case of random mating. However it is rare that individuals in a population truly mate at random[1]. For populations with spatial structure, individuals are more likely to mate with others that are nearby than with individuals from more distant sites. Although it would be nice to define a population as the collection of individuals that are mating together randomly, the boundaries of a population are often very indistinct. A plant that is growing on one side of a meadow is much more likely to pollinate others growing nearby, so in that sense two sides of a meadow may be genetically separated. But when the plants are scattered evenly across that meadow, where would you draw the boundary?

In addition, there may be behavioral patterns that cause individuals to mate non-randomly. If individuals live in family structured groups they may be likely to mate with relatives. An extreme form of non-random mating occurs in plants that routinely self-fertilize.

The result is that populations are often genetically structured, at a variety of scales. Behavioral traits may result in individuals that are inbred with respect to the local subpopulation. And spatial separation may result in sub-populations that are "inbred" with respect to the population as a whole. In the latter case, inbreeding arises because individuals are on average more related to others within their own subpopulation than they are to others in more distant sites. If mating takes place within a subpopulation then they are mating with closer relatives than they would with a truly random sample of mates from the population as a whole.

Sewall Wright developed one index of population structure which he called the "fixation index" or F. F_{IS} is the inbreeding coefficient of an individual with respect to the local subpopulation and F_{ST} is the average inbreeding coefficient of subpopulations relative to the total population.

0.1   Effect of inbreeding on allele and genotype frequencies.

Self fertilization: Lets start with the extreme case of self fertilization. How do allele and genotype frequencies change after on generation of self fertilization? As before we'll consider a single locus with two alleles, A and a.   We'll use upper case letters for the

---

[1] It is important to remember that we are talking about random mating *with respect to a particular locus*. Birds may choose mates non-randomly with respect to plumage, but since unlinked loci are inherited independently, they may very well be mating at random with respect to a certain enzyme locus.

genotype frequencies in that population. Let P be the frequency of AA homozygotes, H be the frequency of heterozygotes and Q be the frequency of aa homozygotes.

$P+H+Q=1$.

After selfing, all of the AA homozygotes produce only AA offspring and all of the aa homozygotes produce only aa offspring. However, selfing within the heterozygotes produces all three genotypes of offspring: 1/4 will be AA, 1/2 will be Aa and 1/4 will be aa. The result is that the frequency of the homozygous genotypes increases with each generation of selfing, and the frequency of heterozygotes decreases.

Genotype	Initial Frequency	Frequency after selfing	
AA	P	$P + 1/4\ H$	
Aa	H	$1/2\ H$	
aa	Q	$Q + 1/4\ H$	

In particular, the change in heterozygosity with selfing is:
$$H' = \frac{1}{2}H$$
The frequency of heterozygotes decreases by half each generation. Eventually, after many generations of selfing the frequency of heterozygotes will decline to zero.

Effect of selfing on allele frequency: The initial allele frequency is $p = P + \frac{1}{2}H$

and after selfing the allele frequency will be $p' = P' + \frac{1}{2}H'$. Substituting the new values for P and H we get

$$p' = P + \frac{1}{4}H + \frac{1}{2}\left(\frac{1}{2}H\right)$$

$$p' = P + \frac{1}{2}H$$

$$p' = p$$

which is simply the original allele frequency.

Non random mating, even in the most extreme form of self fertilization, has no effect on allele frequency. Selfing causes genotype frequencies to change as the frequency of homozygotes increases and the frequency of heterozygotes decreases, but the allele frequency remains constant.

Because non-random mating only reshuffles genotype frequencies with respect to their HW expectations, we can use the deviation of genotype frequencies from their expected values as a measure of inbreeding. If Ho is the observed frequency of heterozygotes and

He is the expected frequency under random mating, then the inbreeding coefficient or fixation index is

$$F = 1 - \frac{H_o}{H_e} .$$

If there are no heterozygotes in the population then the inbreeding coefficient is 1.0. When the frequency of heterozygotes equals the HW expectation then the inbreeding coefficient is 0. In cases where there is an excess of observed heterozygotes, then the inbreeding coefficient can be negative.

Example: Let's assume the population starts out in HWE. That means the genotype frequencies will be p2, 2pq and q2.

Genotype	Initial Frequency	Frequency after selfing
AA	p^2	
Aa	2p(1-p)	
aa	$(1-p)^2$	

- What will be the genotype frequencies after selfing?
- Show that the allele frequency after selfing is still p.

EXTRA: What happens if there is only partial inbreeding? For example, many other plant species reproduce with a mixture of self-fertilization and random outcrossing. Other species may have brother-sister mating or some other intermediate level of inbreeding. In that case the inbreeding rate can vary anywhere between 0 and 1.

In that general case, we can assume that there is a fraction F of seeds that are completely inbred, while the remainder, (1-F) are produced through random mating. The random mating fraction will produce offspring in HW proportions, whereas the inbred portion will produce offspring that have both alleles identical by descent.[2]

What will be the proportion of AA genotypes?
For the random mating fraction, which occurs with probability (1-F), then the AA homozygotes will be produced with probability p^2. For the completely inbred fraction, which occurs with probability F, if one allele is A, then the other one must also be A (because we assume they are completely inbred). Therefore the inbred offspring will

_____

[2] The real seeds do not have to fall into two different classes. They may all have the same intermediate level of inbreeding. But mathematically we can treat the case where every seed is 50% inbreed as exactly the same as when 50% percent of the seeds are the product of random mating and 50% are completely inbred.

either be AA (with probability p) or aa (with probability q). There is no way to produce a heterozygote if there is complete inbreeding.

Genotype	Probability of random mating	Genotype probability	Probability of inbreeding	Genotype Probability
AA	(1-F)	p^2	F	p
Aa	(1-F)	2p(1-p)	F	0
aa	(1-F)	$(1-p)^2$	F	q

$P' = (1-F)p^2 + Fp$
$H' = (1-F)\ 2pq$
$Q' = (1-F)\ q^2 + Fq$

What is the new allele frequency after partial inbreeding?

$p' = P' + \dfrac{1}{2}H'$

$p' = (1-F)p^2 + Fp + \dfrac{1}{2}(1-F)2pq$

$p' = p^2 - Fp^2 + Fp + pq - Fpq$

Now substitute q=(1-p)

$p' = p^2 - Fp^2 + Fp + p - p^2 - Fp + Fp^2$

Notice that almost all of the terms are both added and subtracted. All that is left is

$p' = p$

Again, inbreeding does not alter allele frequencies, it just alters the genotype proportions.

0.2   Estimating population level inbreeding

When there is inbreeding, then the fraction of heterozygotes will be less than the fraction expected under random mating. In particular, the fraction of heterozygotes will be reduced by (1-F).

That suggests that we can use the deviation of observed and expected proportion of heterozygotes as a way of estimating the degree of inbreeding.

$H_{obs} = (1-F)H_{exp}$

$F = 1 - \dfrac{H_{obs}}{H_{exp}}$

This is Wright's measure F_{IS}, if the expected heterozygosity is based on the allele frequency in the local subpopulation.

0.3 Spatial structure

As we stated at the beginning, spatial structuring of the population creates a kind of inbreeding. If individuals mate with others in their own subpopulation they will be mating with others with whom they are more related than if they mated randomly over the entire subpopulation.

We can estimate the inbreeding imposed by population structure with an equation that is analogous to the individual inbreeding coefficient. Sewall Wright labeled that inbreeding associated with spatial structure as F_{ST}: the fixation index (or inbreeding) of a subpopulation relative to the total population. Again we can measure that level of inbreeding by the decrease in heterozygosity within subpopulations relative to what we would expect if there were random mating over the entire population.

$$F_{ST} = 1 - \frac{\text{Average Expected Heterozygosity within Subpopulations}}{\text{Expected Heterozygosity of the Total Population}}$$

$$F_{ST} = 1 - \frac{\overline{H}_s}{H_T}$$

If individuals mate at random over the entire population, then the expected total heterozygosity $H_T = 2p(1-p)$.

On the other hand, if there is spatial structure and individuals mate within subpopulations then the frequency of heterozygotes will depend on the local allele frequency in that subpopulation: $H_i = 2p_i(1-p_i)$ for subpopulation i. If there are a total of k subpopulations, then the average expected heterozygosity within subpopulations is:

$$\overline{H}_S = \frac{1}{k}\sum_{i=1}^{k} 2p_i(1-p_i).$$

Example: Imagine that a population is divided into two spatially separated subpopulations. Subpopulation 1 has p=0.8 and subpopulation 2 has p=0.5. Both of those subpopulations have genotype frequencies that precisely match the HW expectations. On the other hand, if we look at the entire population as a whole the genotypes deviate from HW proportions. There are fewer observed heterozygotes than we would expect, reflecting the "inbreeding" imposed by population structure.

The allele frequency in the combined population is 0.65, so the expected heterozygosity is Ht=0.455. The average expected heterozygosity in the two subpopulations is Hs=0.41.

		Genotype			Allele Frequency
		AA	Aa	aa	p_A
Subpopulation 1		128	64	8	0.8
Expected		128	64	8	
Subpopulation 2		50	100	50	0.5
Expected		50	100	50	
Total Population		178	164	58	0.65
Expected		169	182	49	

$$F_{ST} = 1 - \frac{\overline{H}_s}{H_T} = 1 - \frac{0.410}{0.455} = 0.099$$

To understand the properties of F_{ST}, we can look at some extreme conditions. If individuals are mating completely at random over the entire population, then there will be no local variation in allele frequency and each of the subpopulations will have the same expected heterozygosity as the total population. In that case Fst will be 0: there is no differentiation among subpopulations. At the other extreme, each subpopulation may be completely isolated from all of the other subpopulations and each subpopulation may have become fixed for a different allele. That is the maximum level of differentiation. If each subpopulation is fixed for one allele or the other, then there is no heterozygosity within subpopulations (H_S =0). In that case Fst will be 1.0.

For most species of animals, F_{ST} ranges from 0 to 0.2. Plant populations usually show somewhat higher degrees of spatial structure, since plants are rooted in place. Yet even for plants F_{ST} is usually less than 0.4.