# Chapter 4: Properties of Single Loci

## Introduction

too easy

## Allele and genotype frequencies

too easy

## The transmission of genetic information

### The Hardy-Weinberg principle

$$p^2 + 2pq + q^2 = 1$$

where p = allele frequency of first allele at a locus and q = allele frequency of the second allele at that same locus.

Assumptions of H-W:

- No selection
- No mutation
- Random mating
- No differential migration
- No random drift

Even though these assumptions will never be met completely in the real world, for the majority of the time the H-W prinicple holds regardless.

Assuming assumptions are met, 2 important points from H-W:

1. It takes no more than a single generation equilibriate and stabilize gene frequencies in the two sexes.
2. Only one additional generation is required for the stabilisation of the genotype frequencies into the predictible Hardy-Weinberg proportions.

#### Sex-linked loci

Alleles on sex chromosomes in diploid organisms are obviously different. Sons can only receive and X chromosome from their mother so the frequency of X linked loci in the sons is equal to that of their mothers. Daughters receive both an X chromosome from Mum + from Dad.

Overall this means allele frequencies oscilate around an equilibrium state, but continually get closer to that state over the generations (see Figure 4.2 and page 56 for equation).

#### Polyploidy

Skipped over this section because it's not relevant to human quant gen. Buuut, essentially it just details how to derive allele frequencies under a certain case of polyploidy. Also, it should be noted that of course H-W does not hold under polyploidy!

### Age structure

Age structure also complicates our idealised model of H-W. In populations composed of several age classes, the generations overlap, and this causes the approach of genotype frequencies towards the H-W expectations to be gradual (rather than just by 1/2 generations), even in the case of an autosomal locus. Doesn't explain this very much, but it's covered elsewhere. Importantly, when newly founded populations have significant age structure, fluctuations in both gene and genotoype frequencies may occur for a substantial period of time even in the abscence of selection!

### Testing for Hardy-Weinberg proportions

Says in the book that LRT can be used to test for departures from HWE, buuuut I'm pretty sure that even now the most common method is the chi-squared test (or Fisher's exact test if the sample size is tiny and the allele is rare.). Essentially, in a population, at a specific locus, you can calculate the allele frequencies (and from that expected genotype frequencies) from the observed genotype frequencies then test if there is a difference between the observed and expected values. LRT equation for it given on page 60. **RECREATE THE CHI-SQUARED TEST IN CODE!!!!**

Should remember (as pointed out above), that just because some assumptions are violated, doesn't mean you'd get a departure from HWE!

## Characterising the influence of a locus on the phenotype

If a trait is entirely influenced by a single locus then the genetic effect on that trait can be characterised pretty easily and the dominance and additive effects of the alleles can be calculated. So if a locus has genotypes $B_1B_1$, $B_1B_2$, $B_2B_2$, then the values given to these genotypes can be said to be: $-a$, $(1+k)a$ and $+a$. Now if you have genotype data at that locus and data on the trait you can work out the effect of the $B_2$ allele by taking the mean phenotypic value of individuals with $B_2B_2$ and subtracting the mean phenotypic value of individuals with $B_1B_1$ and dividing by 2 i.e.

$$B_{2eff} = \frac{p_{B2} - p_{B1}}{2}$$

where $B_{2eff}$ is the effect of allele $B_2$, $p_{B2}$ is the mean phenotypic value of individuals with $B_2B_2$ and $p_{B1}$ is the mean phneotypic value of individuals with $B_1B_1$.

As $B_{2eff} = a$ you can then substitute this in to $(1+k)a$ to get the dominance coefficient $k$. Of course if $k = 0$ then there is no dominance (in reality you would calculate probability of dominance).

## The basis of dominance

Confusing part... Don't really get the enzyme activity bit...

Main point (I think) is that new deleterious mutations are very likely to be recessive and new mutations with a slight deleterious effect interact in an almost entirely additive fashion (no dominance!).

## Fisher's decomposition of the genotypic value

Recalling that the phenotypic value can be partitioned like so:

$$z = G + E$$

where z is the phenotype, G is the genotypic value and E is the environmental value.

The genotypic value of a specific locus can be partitioned into it's "expected" values based on there being only additive effects ($\hat{G}$) and the deviations from the expected values or dominance effects ($\delta$). So for genotype $B_iB_J$:

$$G_ij = \hat{G_i}j + \delta_ij$$

This can be formalised (whatever the fuck that means) by regressing the genotypic values on the number of $B_1$ and $B_2$ alleles in the genotype ($N_1$ and $N_2$):

$$G_ij = \hat{G_i}j + \delta_ij = \mu_G + \alpha_1 N_1 + \alpha_2 N_2$$

$\mu_G$ = the mean genotypic value in the population, $\alpha_1$ and $\alpha_2$ are the slopes of the regression, $N_1$ and $N_2$ are the number of $B_1$ and $B_2$ alleles. So the regression is:

$$ G_ij ~ N_2 + N_1

By noting that for any individual, $N_1 = 2 - N_2$ you can reduce the multiple regression model into an easier to work with univariate model. Give it a go:

$$G_ij = ...$$

If you plotted the genotypic value ($G$) against gene content ($N_2$ or number of $B_2$ alleles) and calculated residuals these residuals would be $\delta$ (see Figure 4.6).

The rest of the chapter uses this regression and what we know about genotype frequencies to derive a formula for the average effect of allelic substitution:

$$\alpha = a[1 + k(p_1 - p_2)]$$

where $a$ = genotypic value of $B_2$ (see above), $k$ is the dominance coefficient and $p_1$ and $p_2$ are the frequencies of $B_1$ and $B_2$. This value $\alpha$ represents the average change in genotypic value that results when a $B_2$ allele is randomly substituted for a $B_1$ allele. If no dominance ($k = 0$) then $\alpha = a$. Except in the case of additivity, the average effect of allelic substitution is not simply a function of the inherent physiological properties of the allele. It can only be defined in the context of the population!

## Partioning the genetic variance.