Population Genetics: The Hardy-Weinberg Model for Two Loci

Sorin Istrail

Department of Computer Science
Brown University, Providence
sorin@cs.brown.edu

September 21, 2010

Outline

Linkage Equilibrium (LE)

- Consider locus **A**. Random mating implies alleles at this locus are combined at random in genotypes according to the Hardy-Weinberg proportions. If the two alleles are A_1, A_2 and their frequencies are p_1, p_2 with $p_1 + p_2 = 1$. Then the HW Law tells us that genotypes A_1A_1, A_1A_2, and A_2A_2 are expected in proportions $p_1^2, 2p_1p_2, p_2^2$ respectively.
- Consider locus **B**. Random mating implies alleles at this locus are combined at random in genotypes according to the Hardy-Weinberg proportions. If the two alleles are B_1, B_2 and their frequencies are q_1, q_2 with $q_1 + q_2 = 1$. Then the HW Law tells us that genotypes B_1B_1, B_1B_2, and B_2B_2 are expected in proportions $q_1^2, 2q_1q_2, q_2^2$ respectively.

Linkage Equilibrium

- The term random association is used when the frequency of any gamete that is made out of a particular combination of alleles equals the product of the frequencies of those alleles. Therefore, we say that alleleA_1 is in random association with allele A_2, and we say that alleleB_1 is in random association with allele B_2. Genes or sites in random association are said to be in Linkage Equilibrium (LE).

Gamete	Frequency
$A_1 B_1$	$p_1 q_1$
$A_1 B_2$	$p_1 q_2$
$A_2 B_1$	$p_2 q_1$
$A_2 B_2$	$p_2 q_2$

Recombination Rate r

- The rate of approaching Linkage Equilibrium between two genes or two sites depends on the rate of recombination in genotypes heterozygous for both genes or sites.
- There are two genotypes that are called double heterozygotes: A_1B_1/A_2B_2 and A_1B_2/A_2B_1. Both the A site and B site in each genotype has both alleles and so the name pointing out this property. The symbol"/" denotes the union of the two gametes.
- Such double heterozygotes genotypes have the property that their gametes in the next generation are of 4 types: (1) A_1B_1, (2) A_2B_2, (1) A_1B_2, (1) A_2B_1.

Recombination Rate r

- The games (1) and (2) are called non-recombinant : their alleles are associated in the same manner as in the previous generation, A_1 with B_1 and A_2 with B_2. By Mendelian segregation the frequency of the gamete (1) is the same as the frequency of the gamete (2).
- The games (3) and (4) are called recombinant: their alleles are associated in differently from the previous generation, A_1 with B_2 and A_2 with B_1.By Mendelian segregation the frequency of the gamete (3) is the same as the frequency of the gamete (4). By Mendelian segregation the frequency of the gamete (3) is the same as the frequency of the gamete (4).
- The combined frequency of the gametes of types (1) and (2) is different in general from than the combined frequency of the gametes of types (3) and (4). It is the recombination rate that determines the proportion of the two classes.

Definition of the recombination rate r

Definition

The recombination rate r is the proportion of recombinant gametes produced by a double heterozygote.

- If the two loci are on the same chromosome, the recombination rate depends on physical distance on the chromosome of the two loci. The closer the two sites the less likely that a breakage takes place in the region between the sites.
- For loci on different chromosomes, $r = 0.5$. In this case all four gametic types are produced in equal frequency.

Definition of the recombination rate r

The values of r have the following interpretations:

-
  - $0 \leq r \leq 0.5$;
  - $r = 0$ means no recombination;
  - $r = 0.5$ means sites are on different chromosomes or far apart on the same chromosome;
  - When $r < 0.5$ the two sites are on the same chromosome and are said to be *linked*.

There are now 10 types of genotypes

- A_1B_1/A_1B_1 double homozygous
- A_1B_1/A_1B_2 single heterozygous (at B)
- A_1B_1/A_2B_1 single heterozygous (at A)
- A_1B_1/A_2B_2 double heterozygous (cis)
- A_1B_2/A_1B_2 double homozygous
- A_1B_2/A_2B_1 double heterozygous (trans)
- A_1B_2/A_2B_2 single heterozygous (at A)
- A_2B_1/A_2B_1 double homozygous
- $A_2B_1A_2B_2$ single heterozygous (at B)
- A_2B_2/A_2B_2 double homozygous

The double heterozygote A_1B_1/A_2B_2

This genotype produces the following gametes:

Gamete	Frequency
A_1B_1	$\frac{1-r}{2}$
A_2B_2	$\frac{1-r}{2}$
A_1B_2	$\frac{r}{2}$
A_2B_1	$\frac{r}{2}$

The double heterozygote A_1B_2/A_2B_1

This genotype produces the following gametes:

Gamete	Frequency
A_1B_1	$\frac{r}{2}$
A_2B_2	$\frac{r}{2}$
A_1B_2	$\frac{1-r}{2}$
A_2B_1	$\frac{1-r}{2}$

Linkage Disequilibrium parameter D

Let us denote $g_{A_1B_1}$ and $g'_{A_1B_1}$ the frequency of gamete A_1B_1 in the current generation, and next genera

$$g'_{A_1B_1} = 1 * g^2_{A_1B_1} + \frac{1}{2}(2g_{A_1B_1}g_{A_1B_2}) + \frac{1}{2}(2g_{A_1B_1}g_{A_2B_1}) + \frac{1}{2}(1-r)(2g_{A_1B_1}g_{A_2B_2}) + \frac{1}{2}r(2g_{A_1B_2}g_{A_1B_1}) =$$

$$g_{A_1B_1}(g_{A_1B_1} + g_{A_1B_2} + g_{A_2B_1} + (1-r)g_{A_2B_2}) + rg_{A_1B_2}g_{A_2B_1} =$$

$$g_{A_1B_1}(g_{A_1B_1} + g_{A_1B_2} + g_{A_2B_1} + g_{A_2B_2}) + rg_{A_1B_2}g_{A_2B_1} - rg_{A_1B_1}g_{A_2B_2} =$$

$$g_{A_1B_1} + r(g_{A_1B_2}g_{A_2B_1} - g_{A_1B_1}g_{A_2B_2}) = g_{A_1B_1} - rD$$

D

$$\begin{array}{|c|}\hline g'_{A_1B_1} = g_{A_1B_1} - rD \\ \hline g'_{A_1B_2} = g_{A_1B_2} + rD \\ \hline g'_{A_2B_1} = g_{A_2B_1} - +D \\ \hline g'_{A_2B_2} = g_{A_2B_2} - rD \\ \hline \end{array}$$

Single locus HW: constancy of frequencies = no evolution; Two loci HW: potential for evolution via D and r

- The qualitative difference between single locus and two loci HW is the potential for evolution. In the single locus case, equilibrium is reached in a single generation of random mating and preserved thereafter. For the two loci case, the gamete frequencies in the next generation is potentially different than their frequencies in the previous generation; the parameters quantifying the difference are r and D.
- When is Evolution occuring? – ie., change in the frequencies of genes in the gene pool (gametes). Clearly if $r > 0$, i.e., some recombination is happening, and $D \neq 0$, i.e., there is some linkage disequilibrium, then $g'_{A_1B_1} \neq g_{A_1B_1}$ and Evolution takes place!
- When $r = 0$ no evolution is occuring. In this case, the two loci model is equivalent with the single locus model with four

$D = 0$

- $D = 0$ happens when the two loci gamete frequencies are the product of their respective single locus alleles frequencies.
- We have $p_{A_1} = g_{A_1B_1} + g_{A_1B_2}$ and $q_{B_1} = g_{A_1B_1} + g_{A_2B_1}$. Then $p_{A_1}q_{B_1} = (g_{A_1B_1} + g_{A_1B_2})(g_{A_1B_1} + g_{A_2B_1}) =$ $g^2_{A_1B_1} + g_{A_1B_1}g_{A_2B_1} + g_{A_1B_1}g_{A_1B_2} + g_{A_1B_2}g_{A_2B_1} =$ $g_{A_1B_1}(g_{A_1B_1} + g_{A_1B_2} + g_{A_2B_1}) + g_{A_1B_2}g_{A_2B_1} =$ $g_{A_1B_1} - g_{A_1B_1}g_{A_2B_2} + g_{A_1B_2}g_{A_2B_1} = g_{A_1B_1} - D$
- therefore
- $D = g_{A_1B_2} - p_{A_1}q_{B_1}$

D is the deviation from random

- $D = g_{A_1B_2} - p_{A_1}q_{B_1}$
- $D =$ the deviation of the two loci gamete frequencies from the product of the respective single locus allele frequencies.