BIS 2B Math Equations

Dylan Ang

May 2021

Contents

1 Linkage Mapping

Chromosomal crossover, or crossing over, is the exchange of genetic material during sexual reproduction between two homologous chromosomes' non-sister chromatids that results in recombinant chromosomes [1].

$$\% \text{ Recombination} = \frac{\text{\# of Recombinant Offspring}}{\text{Total \# of Offspring}} \tag{1}$$

Note: 1% Recombination = 1 map unit

Consider the following sample data, which shows the number of individuals in a sample population with a certain genotype. Note that there are 1000 total individuals.

Abc 18	aBc 112	abc 308	ABc 66
abC 59	ABC 320	AbC 102	aBC 15

You may be asked to calculate the distance between two genes a and b. To find this distance, note which genotypes feature a crossover event between a and b. i.e when a and b do not match case. These are

- Abc 18
- aBc 112
- AbC 102
- aBC 15

Plug into equation (1)

$$\frac{18 + 112 + 102 + 15}{1000} = 24.7\% = 24.7 \text{ map units}$$

1.1 Identifying Gene Order

In addition to calculating distance between genes, you may be asked to identify the order of the genes.

Consider two sister chromatids,

Figure 1: Sister Chromatids

Figure 2: Single Recombination Event

Figure 3: Double Recombination Event

As you can see in Figure 3, double recombination events require two "cuts" compared to just one. This makes successful double recombination much more rare than single or no recombination.

If I told you that Abc and aBC were single recombination events, you'd be able to tell that the A gene is the one crossing over. You'd be able to conclude that A is on the left side or right side of the genome. Unfortunately this doesn't help us very much because this means the gene order could be ABC or ACB, as well as the inverses of these, CBA and BCA. So while it narrows down the order, it doesn't give us the answer.

Now consider if I told you that Abc and aBC represented the double recombination event. You'd know that there must be two genes crossing over (Hence the "double" in double recombination), and these genes must be B and C. If B and C are the two genes crossing over, you can see from Figure 3 that the odd gene out is in the center position. Therefore A must be in the center of the genome. In summary, If you know which genotype corresponds to a Double Recombination Event, then you can find the order of the alleles on the genome.

In order identify which genotype corresponds to a Double Recombination Event, find the two least frequent genotypes. Recall that Double Recombination events are more rare than single or no mutation events, so we know that these two are the Double Recombination events. Now consider the following sample data, I've circled the two least common genotypes.

Abc 18	aBc 112	abc 308	ABc 66
abC 59	ABC 320	AbC 102	aBC 15

We can conclude that Abc and aBC are the Double Recombination events, and thus the A allele is in between B and C on the genome. Note that we technically don't know if the order is BAC or CAB, but these are actually considered the same ordering.

2 Hardy Weinberg

In population genetics, the Hardy–Weinberg principle, also known as the Hardy–Weinberg equilibrium, model, theorem, or law, states that allele and genotype frequencies in a population will remain constant from generation to generation in the absence of other evolutionary influences. These influences include genetic drift, mate choice, assortative mating, natural selection,

sexual selection, mutation, gene flow, meiotic drive, genetic hitchhiking, population bottleneck, founder effect and inbreeding [2].

2.1 Equations and Notation

Frequencies are written as f(The thing we are testing for). For example, the frequency of the AA genotype is $f(AA)$, and the frequency of the A allele is written $f(A)$.

$$f(AA) = \frac{\text{\# of AA individuals}}{\text{\# of individuals (N)}} \tag{2}$$

$$f(A) = \frac{\text{\# of A alleles}}{\text{total \# of alleles (2N)}} \tag{3}$$

There are two ways to calculate allele frequencies from given frequencies. The equation on the right hand side requires that you know the total number of individuals, N. Whereas the left hand side only requires that you have the frequencies of the genotypes.

$$f(A) = f(AA) + \frac{f(Aa)}{2} \tag{4}$$

$$= \frac{2 * (\text{\# of AA}) + (\text{\# of Aa})}{2N} \tag{5}$$

Variable	Allele
p	Dominant Allele
q	Recessive Allele

Variable	Genotype
p^2	Homozygous Dominant
$2pq$	Heterozygous
q^2	Homozygous Recessive

2.2 Example Problem: Counting Fruit Fly Phenotypes

A and a code for the eye color of the fruit flies. The A allele codes for red eyes and the a allele codes for white eyes. There is incomplete dominance, and heterozygotes (Aa) have pink eyes.

This question asks you to perform analysis of data on fruit fly genotypes. Given the number of each genotype in each generation, can you determine if there is evolution happening? The basic process for solving these problems is

1. Calculate Observed Genotype Frequencies using the data table
2. Use these values to calculate Allele Frequencies
3. Use the calculated allele frequencies to predict the next generation's genotype frequencies.
4. Calculate the observed genotype frequencies of the next generation using the same method as used for the first generation.

	# Red eyed flies	# pink eyed flies	# white eyed flies
S_0	83	32	106
S_1	85	29	100
F_1	8	22	10

2.2.1 Step 1

First, we need to calculate the observed frequencies of the genotypes. To do this, refer to equation (2). The frequency of the different genotypes are

$$f(\text{Red Eyes}) = f(AA) = \frac{83}{83+32+106} = 0.376$$
$$f(\text{Pink Eyes}) = f(Aa) = \frac{32}{83+32+106} = 0.145$$
$$f(\text{White Eyes}) = f(aa) = \frac{106}{83+32+106} = 0.480$$

We have $83+32+106$ in the denominator, as this is N, the total number of individuals in the S_0 population. From here we can move on to step 2 by using one of the two equations (4) and (5) to calculate the allele frequencies of A and a. I will show both.

2.2.2 Step 2

The first method is with the equation (4). We just found $f(AA) = 0.376$ and $f(Aa) = 0.145$, so

$$f(A) = 0.376 + \frac{0.145}{2} = 0.448$$

Finding $f(a)$ is similar, except we replace $f(AA)$ with $f(aa)$.

$$f(a) = 0.480 + \frac{0.145}{2} = 0.552$$

The second method uses equation (5) where N is the total number of individuals. From the table, there are 83 AA individuals, 32 pink eyed individuals, and 106 white eyed individuals. In total, there are $N = 221$ individuals. Therefore

$$f(A) = \frac{2 * 83 + 32}{2 * 221} = 0.448$$
$$f(a) = \frac{2 * 106 + 32}{2 * 221} = 0.552$$

Both methods result in the same answer, the only reason to use one or the other is if you can not use the original counts and only have the frequencies to go off of. Then, it would make sense to use the equation that is based on frequency.

2.2.3 Step 3

If the populations are in Hardy Weinberg Equilibrium, then subsequent generations will have expected genotype frequencies matching $f(AA) = p^2$, $f(Aa) = 2pq$, $f(aa) = q^2$ where $p = f(A)$ and $q = f(a)$. The expected genotype frequencies for subsequent generations of S_0 are

$$f(AA) = p^2 = (0.448)^2$$
$$f(Aa) = 2pq = 2 * (0.448) * (0.552)$$
$$f(aa) = q^2 = (0.552)^2$$

2.2.4 Step 4

Now that we know what we are expecting to see in S_1, calculate the observed genotype frequencies out of the data table in order to compare them to your expected values. If there is a significant difference, then this population is not in Hardy Weinberg Equilibrium.

References

[1] Wikipedia contributors. *Chromosomal crossover — Wikipedia, The Free Encyclopedia*. [Online; accessed 16-May-2021]. 2021. URL: https://en.wikipedia.org/w/index.php?title=Chromosomal_crossover&oldid=1022943486.

[2] Wikipedia contributors. *Hardy–Weinberg principle — Wikipedia, The Free Encyclopedia*. [Online; accessed 16-May-2021]. 2021. URL: https://en.wikipedia.org/w/index.php?title=Hardy%E2%80%93Weinberg_principle&oldid=1021474919.