Some Basic Genetics:
Most living things pass on their characteristics to the next generation as information encoded by a chemical called DNA. There are many ways for species to reproduce, but here we'll only be interested in organisms that reproduce sexually. These organisms have two copies of each DNA strand and the copies are not necessarily the same. Each chunk of DNA that contains the information for one particular feature (like eye color, hair color, blood type and diseases such as sickle-cell anemia and cystic fibrosis) is called a *gene*.

The different possibilities for each gene are called *alleles*. For example the gene for eye color could have the allele that codes for blue eyes, or the allele that codes for brown eyes. Let's let A represent the allele coding for brown eyes and a represent the allele coding for blue eyes. Then, since each individual has two copies of each gene, any one individual could have one of three *genotypes*: AA, Aa, aa. The *genotype* of an individual refers to the specific types of each gene that they have. If an individual has two of the same alleles (AA or aa) we call them a *homozygote*. If an individual has two different alleles (Aa), we call them a *heterozygote*.

When individuals mate, they each give one of their genes to their offspring, so the babies have one copy of a gene from their mother and one from their father. For example, if a mother and father both have AA genotypes, all of the children will have AA genotypes. However, if one parent has the AA genotype and one has the aa genotype, all of the children will have the Aa genotype. Likewise, if one parent has the AA genotype and the other has the Aa genotype, half of the children will have AA and half will have Aa.

In the field of mathematical biology, it is common to make simplifying assumptions before proceeding with the mathematical analysis of a problem. For our purposes we'll assume the following:

(1) Non-overlapping Generations: This means that individuals in one generation die before the members of the next generation are born. This probably doesn't seem like a reasonable assumption for humans, and is actually more reasonable for things like annual plants and some invertebrates. However, this assumption will make our work a lot easier and will still give us some useful information.

(2) Random Mating: This means that mating takes place completely at random. This may also seem a bit unreasonable since movie stars tend to marry other movie stars rather than your average professor at the University. However, it is pretty unlikely that a person chooses their boyfriend, girlfriend, husband or wife based on their blood type. Likewise, a person

with brown eyes may be more attracted to another person with brown eyes, but they won't care if that person has the genotype AA or Aa. So, for some genetic traits, this may be a pretty good assumption.

(3) Sexes are Equivalent: For our models, we'll assume that males and females have the exact same genotypes in the same proportions. As you'll see in our classroom simulations, we'll all be able to mate with everyone else.

(4) No Migration: For our models, we'll assume that no one is coming into or leaving our population. We can imagine a population of individuals living on an island that is very difficult to get to and is rarely visited by outsiders.

Basic Notation
We will use the following notation.

D = fraction of the population that is type AA (Dominant Homozygous)
H = fraction of the population that is type Aa (Heterozygous)
R = fraction of the population that is type aa (Recessive Homozygous)

We assume that the three genotypes given above (AA, Aa, and aa) are the only possible genotypes for the trait we are studying. Therefore, $D + H + R = 1$. We are also interested in the fractions of the individual alleles (A and a). Each individual carries two alleles, so the total number of alleles in the population is two times the population size. If we are calculating the fraction of alleles that are of type A, we have two contributions from AA individuals and one contribution from Aa individuals. Likewise, for alleles of type a, there is one contribution from type Aa individuals and two from type aa. We use p to represent the fraction of alleles in the population that are of type A and q to represent the fraction that are a. We must have $q = 1 - p$ since there are only two types of alleles.

$$p = \frac{2D + H}{2(D + H + R)} = D + \frac{H}{2} \qquad\qquad q = \frac{H + 2R}{2(D + H + R)} = \frac{H}{2} + R$$

Hardy-Weinberg Equilibrium
For our first example we'll assume that the three genotypes have no effect on whether you live or die. For example, they might determine the color of your eyes or hair. Let's assume that people with the AA or Aa genotype have brown eyes and the people with the aa genotype have blue eyes. There's no reason to think that people with brown eyes will be more likely to survive and reproduce than people with blue eyes (and vice-versa), so we say that there is no *selection* for one genotype over another. As in our class simulation, we'll assume that the initial population starts out with $1/2$ AA and $1/2$ aa. What will happen over time?

2

Using our notation defined above, we have genotype frequencies of $D_0 = 1/2$, $H_0 = 0$, $R_0 = 1/2$ and allele frequencies of $p_0 = 1/2$ and $q_0 = 1/2$. We will make a mating square to determine all of the possible offspring types and their proportions. The first column and row represent the genotypes of the parents and the interior boxes represent the offspring. The first interior box (top left) gives the probability that an AA mates with an AA. This is $\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$. Each child gets one allele from each of its parents. In this first pairing, all of the children are AA. The second interior box in the top row represents the pairing of an AA with an aa, which also has probability $\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$. All of the children from this pairing are type Aa since they receive an A allele from one parent and an a allele from the other parent. We fill in the other boxes using similar reasoning.

	AA (1/2)	aa (1/2)
AA (1/2)	AA (1/4)	Aa (1/4)
aa (1/2)	Aa (1/4)	aa (1/4)

We can use the table we've just created to determine both the fractions of the three types of individuals in the next generation and the new allele frequencies.

$$D_1 = \frac{1}{4} \qquad H_1 = \frac{1}{4} + \frac{1}{4} = \frac{1}{2} \qquad R_1 = \frac{1}{4}$$

$$p_1 = D_1 + \frac{H_1}{2} = \frac{1}{4} + \frac{1/2}{2} = \frac{1}{2} \qquad q_1 = \frac{H_1}{2} + R_1 = \frac{1/2}{2} + \frac{1}{4} = \frac{1}{2}$$

You may have noticed something interesting about the last two calculations. If not, notice that our new allele frequencies p_1 and q_1 are exactly the same as they were in the parent generation! Does this make sense? At the start, we assumed that there is no advantage to having one genotype over another (brown eyes versus blue eyes), so its not that surprising that the relative number of A and a alleles in the population stays the same. However, they seem to have reorganized themselves into different pairings. We started out with parents that had either two A's or two a's. Now, we see that half of the population is a new type Aa. Let's see what happens when we look at the grandchildren of our original population. To do this, we'll make another mating square, adding a third row and column for the Aa

type.

	AA (1/4)	Aa (1/2)	aa (1/4)
AA (1/4)	AA (1/16)	AA $\left(\frac{1}{2}\cdot\frac{1}{8}=1/16\right)$ Aa $\left(\frac{1}{2}\cdot\frac{1}{8}=1/16\right)$	Aa (1/16)
Aa (1/2)	AA $\left(\frac{1}{2}\cdot\frac{1}{8}=1/16\right)$ aA $\left(\frac{1}{2}\cdot\frac{1}{8}=1/16\right)$	AA $\left(\frac{1}{4}\cdot\frac{1}{4}=1/16\right)$ Aa $\left(2\cdot\frac{1}{4}\cdot\frac{1}{4}=2/16\right)$ aa $\left(\frac{1}{4}\cdot\frac{1}{4}=1/16\right)$	Aa $\left(\frac{1}{2}\cdot\frac{1}{8}=1/16\right)$ aa $\left(\frac{1}{2}\cdot\frac{1}{8}=1/16\right)$
aa (1/4)	Aa (1/16)	aA $\left(\frac{1}{2}\cdot\frac{1}{8}=1/16\right)$ aa $\left(\frac{1}{2}\cdot\frac{1}{8}=1/16\right)$	aa (1/16)

Again, we'll use the table we've just created to determine the fractions of the three types of individuals and the allele frequencies in the second generation.

$$D_2 = 4\cdot\frac{1}{16} = \frac{1}{4} \qquad H_2 = 8\cdot\frac{1}{16} = \frac{1}{2} \qquad R_2 = 4\cdot\frac{1}{16} = \frac{1}{4}$$

$$p_2 = D_2 + \frac{H_2}{2} = \frac{1}{4} + \frac{1/2}{2} = \frac{1}{2} \qquad q_2 = \frac{H_2}{2} + R_2 = \frac{1/2}{2} + \frac{1}{4} = \frac{1}{2}$$

Hopefully, you've noticed something interesting about all of these calculations. If not, notice that we got the exact same frequencies in our first generation! It looks like our population is at some sort of equilibrium. In fact, after only one generation our population has reached something called the Hardy-Weinberg equilibrium.

We started with a population that was half AA (brown eyes) and half aa (blue eyes) so the allele frequencies were both 1/2. What would have happened if we had started with

unequal proportions? Assume that D_0 is the fraction of AA in our parent generation, H_0 is the fraction of Aa and R_0 is the fraction of aa. Let's make a mating square and see what happens.

	AA (D_0)	Aa (H_0)	aa (R_0)
AA (D_0)	AA (D_0^2)	AA $\left(\frac{1}{2} \cdot D_0 H_0\right)$ Aa $\left(\frac{1}{2} \cdot D_0 H_0\right)$	Aa $(D_0 R_0)$
Aa (H_0)	AA $\left(\frac{1}{2} \cdot H_0 D_0\right)$ aA $\left(\frac{1}{2} \cdot H_0 D_0\right)$	AA $\left(\frac{1}{4} \cdot H_0^2\right)$ Aa $\left(2 \cdot \frac{1}{4} \cdot H_0^2\right)$ aa $\left(\frac{1}{4} \cdot H_0^2\right)$	Aa $\left(\frac{1}{2} \cdot H_0 R_0\right)$ aa $\left(\frac{1}{2} \cdot H_0 R_0\right)$
aa (R_0)	Aa $(R_0 D_0)$	aA $\left(\frac{1}{2} \cdot R_0 H_0\right)$ aa $\left(\frac{1}{2} \cdot R_0 H_0\right)$	aa (R_0^2)

First, we'll look at the fractions of the three types of individuals in the next generation.

$$D_1 = D_0^2 + \frac{1}{2}D_0 H_0 + \frac{1}{2}D_0 H_0 + \frac{1}{4}H_0^2 = \left(D_0 + \frac{H_0}{2}\right)^2$$

$$H_1 = \frac{1}{2}D_0 H_0 + D_0 R_0 + \frac{1}{2}D_0 H_0 + \frac{1}{2}H_0^2 + \frac{1}{2}H_0 R_0 + R_0 D_0 + \frac{1}{2}R_0 H_0 = 2\left(D_0 + \frac{H_0}{2}\right)\left(\frac{H_0}{2} + R_0\right)$$

$$R_1 = \frac{1}{4}H_0^2 + \frac{1}{2}H_0 R_0 + \frac{1}{2}R_0 H_0 + R_0^2 = \left(\frac{H_0}{2} + R_0\right)^2$$

Once we've factored the terms, we notice that $D_1 = p_0^2$ where p_0 is the initial fraction of A alleles in the population. Likewise, $H_1 = 2p_0 q_0$ and $R_1 = q_0^2$, where q_0 is the initial fraction of a alleles in the population. Now, let's look at the new allele frequencies.

$$p_1 = D_1 + \frac{H_1}{2} = \left(D_0 + \frac{H_0}{2}\right)^2 + \left(D_0 + \frac{H_0}{2}\right)\left(\frac{H_0}{2} + R_0\right)$$

5

$$= \left(D_0 + \frac{H_0}{2}\right) \left(D_0 + \frac{H_0}{2} + \frac{H_0}{2} + R_0\right) = \left(D_0 + \frac{H_0}{2}\right) = p_0$$

$$q_1 = \frac{H_1}{2} + R_1 = \left(D_0 + \frac{H_0}{2}\right) \left(\frac{H_0}{2} + R_0\right) + \left(\frac{H_0}{2} + R_0\right)^2$$

$$= \left(\frac{H_0}{2} + R_0\right) \left(D_0 + \frac{H_0}{2} + \frac{H_0}{2} + R_0\right) = \left(\frac{H_0}{2} + R_0\right) = q_0$$

These are exactly the same as before! Once again we see that the fraction of each allele stays constant over time. In our calculation of D_1, H_1 and R_1, we found that instead of using the mating square, we could use the formulas $D_1 = p_0^2$, $H_1 = 2p_0q_0$ and $R_1 = q_0^2$. If p_1 is exactly the same as p_0 and q_1 is exactly the same as q_0, the genotype frequencies D_2, H_2 and R_2 will be exactly the same as D_1, H_1 and R_1. So, we seem to have reached an equilibrium. It turns out that after only one generation, any population will reach Hardy-Weinberg equilibrium.

What we've done so far suggests that after only one generation, a population will contain the exact same fractions of all genotypes in all future generations. This means that if half of the population has brown eyes now, half of the population will have brown eyes 100 years from now. Do you think this is what happens in real life? If we're talking about eye color, this doesn't seem like a bad prediction. However, remember that at the start we assumed that eye color will have no effect on whether you live or die. What if we look at a gene for something more important than eye color? Will our analysis still hold?

Before we can answer this question, we introduce the biological concept of *fitness*. If I ask you how fit you are, you may answer by telling me how strong you are, or how far you can run. In biological terms, physical strength is not necessarily important to an individuals fitness. In genetics, all that really matters is passing on your genes to your children, so that your genes carry on even after you die. So, when we are thinking of genetics, the *fitness* of an individual is the number of children he or she will leave in the next generation. If you live to be 100 years old, but never have children, your biological fitness is zero. On the other hand, if you have quintuplets at the age of 25 and die the next day, your biological fitness is still pretty high. With this idea in mind, let's look at some examples.

Lethal Recessive
Cystic Fibrosis is a genetic disease that you may have heard of. A person with cystic fibrosis often has large amounts of mucus gathering in their lungs, which can lead to respiratory infections and eventually death. Before modern medicine, people with cystic fibrosis did not live long enough to reproduce and have children. So, in genetic terms, their fitness was zero.

(With current medical techniques a person with this disease can live quite a long time, but for our example, we'll assume that advanced medical techniques are not yet available.)

It turns out that cystic fibrosis occurs when a person has two copies of a recessive allele (aa). On the other hand, if a person has only one copy of the allele (Aa), they are just as healthy as a person without the allele at all (AA). If we are thinking only of genetics, a person born with cystic fibrosis effectively dies at birth, since they make no contribution to future generations. So, we might expect that after some time, the a allele will disappear from the population. Let's see if this is the case.

For this example, we'll start with an entire population of heterozygotes (Aa). Using our notation from before $D_0 = R_0 = 0$, $H_0 = 1$, $p_0 = 1/2$ and $q_0 = 1/2$. We can use the results of the Hardy-Weinberg equilibrium to make our calculations a little easier. According to Hardy-Weinberg, the fractions of the three genotypes in the next generation will be:

$$D_1 = p_0^2 = \left(\frac{1}{2}\right)^2 = \frac{1}{4} \qquad H_1 = 2p_0q_0 = 2\left(\frac{1}{2}\cdot\frac{1}{2}\right) = \frac{1}{2} \qquad R_1 = q_0^2 = \left(\frac{1}{2}\right)^2 = \frac{1}{4}$$

In this example the aa individuals will all die before they reproduce, so we really have $R_1 = 0$. Now, we'll have to rethink the proportions we have above. One way to do this, is to think of a smaller total population size equal to $D_1 + H_1 = p_0^2 + 2p_0q_0$. To figure out the actual fractions of AA and Aa individuals in the population, we can divide the results above by our total size $p_0^2 + 2p_0q_0$. Therefore, the first generation of children is

$$D_1 = \frac{p_0^2}{p_0^2 + 2p_0q_0} = \frac{1/4}{1/4 + 1/2} = \frac{1}{3} \qquad H_1 = \frac{2p_0q_0}{p_0^2 + 2p_0q_0} = \frac{1/2}{1/4 + 1/2} = \frac{2}{3}$$

and the new allele frequencies are

$$p_1 = D_1 + \frac{H_1}{2} = \frac{1}{3} + \frac{1}{3} = \frac{2}{3} \qquad q_1 = \frac{H_1}{2} + R_1 = \frac{1}{3}$$

The first thing you may notice is that the allele frequencies are not staying the same from one generation to the next. This makes sense as individuals with two a alleles die and do not pass on any alleles to future generations. As expected, the fraction of a alleles in the population is decreasing (from $1/2$ to $1/3$). Let's see what happens after one more generation.

$$D_2 = \frac{p_1^2}{p_1^2 + 2p_1q_1} = \frac{(2/3)^2}{(2/3)^2 + 2\,(2/3)\,(1/3)} = \frac{1}{2}$$

$$H_2 = \frac{2p_1 q_1}{p_1^2 + 2p_1 q_1} = \frac{2\,(2/3)\,(1/3)}{(2/3)^2 + 2\,(2/3)\,(1/3)} = \frac{1}{2}$$

$$p_2 = D_2 + \frac{H_2}{2} = \frac{1}{2} + \frac{1}{4} = \frac{3}{4} \qquad\qquad q_2 = \frac{H_2}{2} + R_2 = \frac{1}{4}$$

Again, the fraction of a alleles in the population has decreased from $1/3$ to $1/4$ and the fraction of A alleles seems to be increasing towards 1. Will the fraction of A alleles actually go to 1 or to some other large fraction?

Before we get into a lot of algebra, let's go ahead and consider the more general case where p_0 is not specified from the start. Once we have a general result, we can compare it to our class simulation by setting $p_0 = 1/2$. First, we write down a general equation for p_1 remembering that $q_0 = 1 - p_0$.

$$p_1 = D_1 + \frac{H_1}{2} = \frac{p_0^2}{p_0^2 + 2p_0 q_0} + \frac{p_0 q_0}{p_0^2 + 2p_0 q_0} = \frac{p_0\,(p_0 + 1 - p_0)}{p_0\,(p_0 + 2 - 2p_0)} = \frac{1}{2 - p_0}$$

This formula will hold in general for any step so we can write

$$p_{n+1} = \frac{1}{2 - p_n}$$

Let's solve for the equilibrium points, or the points where $p_{n+1} = p_n$. We will call the equilibrium point p^*. We have

$$p^* = \frac{1}{2 - p*} \;\Rightarrow\; 2p^* - (p^*)^2 = 1 \;\Rightarrow\; (p^*)^2 - 2p^* + 1 = 0 \;\Rightarrow\; (p^* - 1)^2 = 0$$

The only equilibrium point is $p^* = 1$. This corresponds to the case when A is the only allele in the population. Will the population actually go to this equilibrium?

To answer this question, let's first see if we can come up with a way to calculate p after several generations without going through each of the iterations (i.e. p_0 to p_1 to p_2 ...). Do you see a pattern in what we've calculated so far for the special case $p_0 = 1/2$? We have $p_0 = 1/2$, $p_1 = 2/3$ and $p_2 = 3/4$. It looks like a formula for this case might be $p_n = (n+1)/(n+2)$. Now, let's see if we can find a pattern for the general formula. We'll start by looking at the

8

first few iterations.

$$p_1 = \frac{1}{2 - p_0}$$

$$p_2 = \frac{1}{2 - p_1} = \frac{1}{2 - 1/(2 - p_0)} = \frac{2 - p_0}{3 - 2p_0}$$

$$p_3 = \frac{1}{2 - p_2} = \frac{1}{2 - (2 - p_0)/(3 - 2p_0)} = \frac{3 - 2p_0}{4 - 3p_0}$$

For general p_0 it looks like

$$p_n = \frac{n - (n - 1)p_0}{(n + 1) - np_0}$$

might be what we are looking for. Let's use a method called proof by induction to prove that it is in fact the correct formula. First, we notice that the formula gives us the correct answer when $n = 1$.

$$p_1 = \frac{1 - (0)p_0}{2 - (1)p_0} = \frac{1}{2 - p_0}$$

Let's assume the formula is correct for $n = k$ and show that it works for $n = k + 1$. This will show that the formula must work for all n.

$$p_{k+1} = \frac{1}{2 - p_k} = \frac{1}{2 - \frac{k - (k-1)p_0}{(k+1) - kp_0}}$$

$$= \frac{(k + 1) - kp_0}{2(k + 1) - 2kp_0 - k + (k - 1)p_0} = \frac{(k + 1) - kp_0}{(k + 2) - (k + 1)p_0}$$

This is exactly the form our formula predicts. Now, we can use our formula for p_n to predict what will happen after a long time has passed. We can rewrite the formula by grouping the n's together.

$$p_n = \frac{n - (n - 1)p_0}{(n + 1) - np_0} = \frac{n(1 - p_0) + p_0}{n(1 - p_0) + 1}$$

From here it is easy to see that as $n \to \infty$, $p_n \to 1$.

9

It looks like the "bad" allele will eventually disappear, but how long will it take for this to happen? First, we'll look at our specific example where $p_0 = 1/2$ and make a table of values.

p_0	p_1	p_2	p_3	p_4	p_5	$\cdots$	p_{10}	p_{20}
0.5	0.67	0.75	0.8	0.83	0.86	$\cdots$	0.92	0.95

We're definitely getting closer to 1, but not very quickly. To get an idea of how fast p will approach one, we will look instead at $q = 1 - p$ and see how fast it approaches zero. That is, we'll be looking at how fast the bad a allele disappears.

$$q_n = 1 - p_n = 1 - \frac{n(1 - p_0) + p_0}{n(1 - p_0) + 1} = \frac{n(1 - p_0) + 1 - n(1 - p_0) - p_0}{n(1 - p_0) + 1}$$
$$= \frac{1 - p_0}{n(1 - p_0) + 1} = \frac{1}{n + \frac{1}{1 - p_0}}$$

From this equation, we see that q_n (the fraction of allele a in the population) approaches zero at a rate related to n, or the generation time. This rate also tells us how fast the A allele takes over. We'll come back to this result later when we compare this example to others. In summary, we've found that if a recessive allele is lethal when it occurs as a homozygote (aa), and has no effect when it occurs as a heterozygote (Aa), it will eventually disappear from the population. However, this disappearance will take a long time.

10