

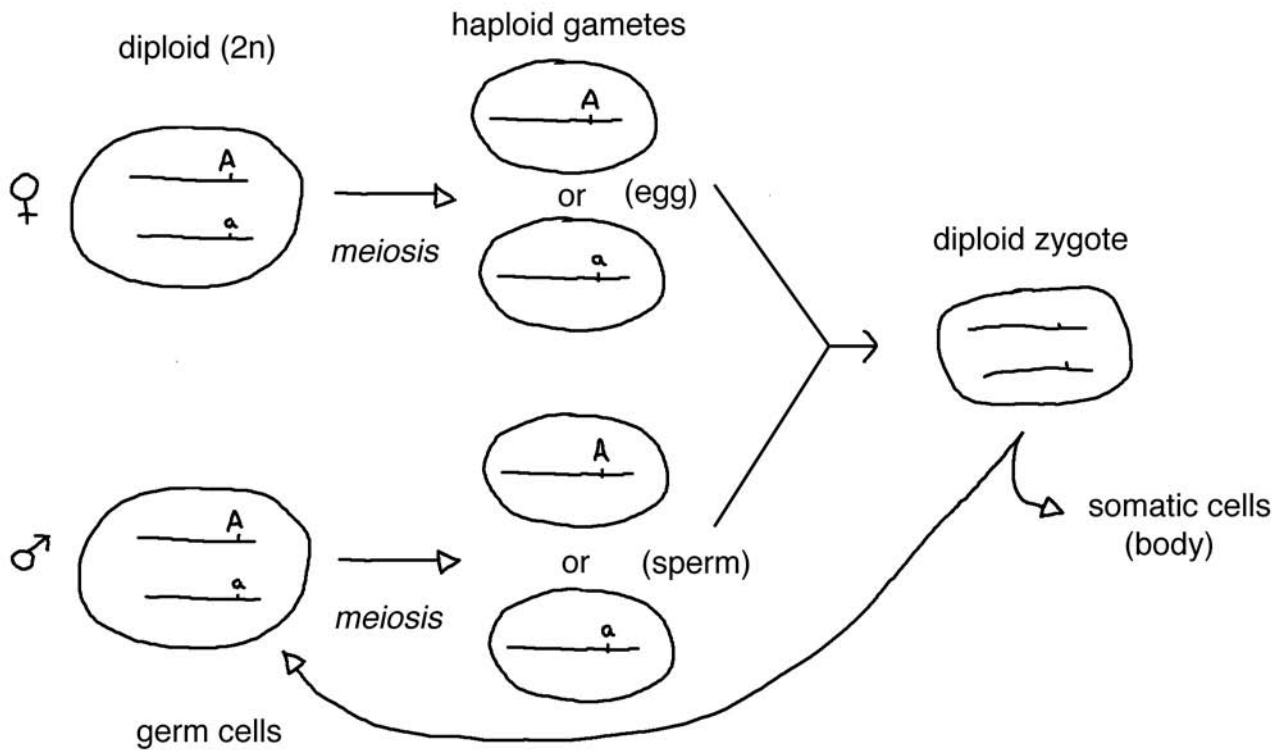
# Genetics Lecture Notes

7.03 2005

Lectures 3 - 5

### Lecture 3

Now let's consider diploid organisms:



The genotype of the zygote will depend on which alleles are carried in the gametes.

Allele in gamete		sperm	
		A	a
egg	A	A/A	A/a
	a	a/A	a/a

When heterozygotes mate their offspring will have different phenotypes: If **A** is dominant to **a**, the two possible phenotypes will be the phenotype of **a/a** or the phenotype of **A/A** and **A/a**.

When we do breeding experiments it is important to know the genotypes of the parents. But as you can see from the example above individuals with the dominant trait could be either **A/A** or **A/a**. A method to control this type of variation is to start with populations that we know to be homozygous. One way to do this is to keep inbreeding individuals until all crosses among related individuals always produce identical offspring. This is known as a true-breeding population and all individuals can be assumed to be homozygous.

**True Breeding:** homozygous for all genes

Say we have a true breeding line of shibire flies; these flies are paralyzed and have genotype **shi<sup>-</sup>/shi<sup>-</sup>**.

First, we can test to see whether the shibire allele is dominant or recessive.

$$\begin{array}{ccc} \mathbf{shi^{-}/shi^{-}} & \times & \mathbf{shi^{+}/shi^{+}} \text{ (wild-type)} \\ \downarrow & & \\ \text{all are } \mathbf{shi^{-}/shi^{+}} & & \end{array}$$

(The offspring from a cross of two true breeding lines is known as the F<sub>1</sub> or first filial generation). The F<sub>1</sub> flies appear like wild type therefore **shi<sup>-</sup>** is recessive (not expressed in heterozygote).

Say we have isolated a new paralyzed mutant that we call **par**.

We start with a true breeding **par<sup>-</sup>** strain that we mate to wild type. We find that the mutation is not expressed in the F<sub>1</sub> heterozygotes and therefore is recessive.

To find out whether **par<sup>-</sup>** is the same as shi- we can do a complementation test since both mutations are recessive. For this test, we cross a true breeding **par<sup>-</sup>** strain to a true breeding **shi<sup>-</sup>** strain.

$$\begin{array}{ccc} \mathbf{par^{-}/par^{-}} & \times & \mathbf{shi^{-}/shi^{-}} \\ \downarrow & & \\ \text{F}_1 \text{ (these flies must inherit both } \mathbf{shi^{-}} \text{ and } \mathbf{par^{-}}) & & \end{array}$$

Possible outcome	Complementation?	Explanation	Inferred genotype
F <sub>1</sub> not paralyzed	<b>shi<sup>-</sup></b> and <b>par<sup>-</sup></b> complement	<b>par<sup>-</sup></b> genotype can supply function missing in <b>shi<sup>-</sup></b> and vice versa	<b>par<sup>-</sup>/par<sup>-</sup>, shi<sup>-</sup>/shi<sup>-</sup></b>
F <sub>1</sub> paralyzed	<b>shi<sup>-</sup></b> and <b>par<sup>-</sup></b> do not complement	<b>par<sup>-</sup></b> has lost function needed to restore <b>shi<sup>-</sup></b>	<b>shi<sup>-</sup>/shi<sup>-</sup></b>

Let's look more carefully at gene segregation in a cross between F<sub>1</sub> flies.

$$\mathbf{shi^{-}/shi^{+}} \quad \times \quad \mathbf{shi^{-}/shi^{+}}$$

What is the probability of a paralyzed fly in the next (F<sub>2</sub>) generation?

Definition:  $p(\mathbf{a}) = \frac{n_{\mathbf{a}}}{N}$        $n_{\mathbf{a}}$  = number of outcomes that satisfy condition **a**

$N$  = total number of outcomes (of equal probability)

Probability problems can be solved by accounting for every outcome, but usually it is easier to combine probabilities.

$p(\text{paralyzed } F_2 \text{ fly}) = p(\text{inherit } \mathbf{shi}^- \text{ from mother and inherit } \mathbf{shi}^- \text{ from father})$

Product rule:  $p(\mathbf{a} \text{ and } \mathbf{b}) = p(\mathbf{a}) \times p(\mathbf{b})$

(note the product rule only applies if **a** and **b** are independent which is the case here since the allele from mother does not affect the allele from the father)

$p(\mathbf{shi}^- \text{ from mother}) = 1/2$ ,  $p(\mathbf{shi}^- \text{ from father}) = 1/2$

$p(\text{paralyzed}) = 1/2 \times 1/2 = 1/4$

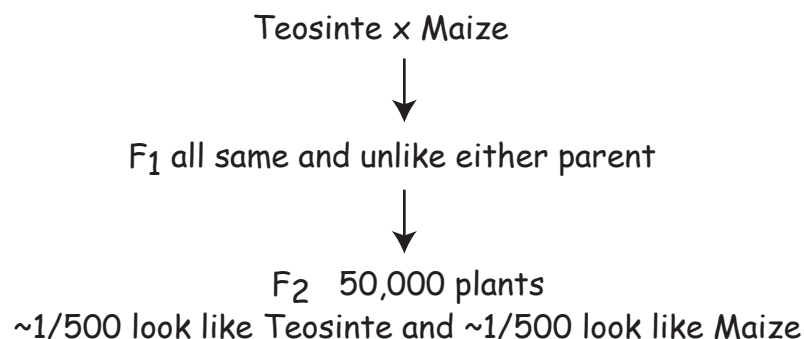
$p(\text{not paralyzed}) = 1 - 1/4 = 3/4$

Thus in the  $F_2$  generation the phenotypic ratio will be, 1 paralyzed : 3 not paralyzed

A 1 : 3 phenotypic ratio among the  $F_2$  in a breeding experiment shows that alleles of a single gene are segregating.

This actually constitutes a third definition of a gene. Historically, this was the first definition of the gene developed by Gregor Mendel in the 1860s. Mendel was able to detect single genes segregating in pea plants because he looked at simple traits and started with true-breeding strains.

Let's see how these ideas can be applied to a very interesting problem in the evolution of corn. Domestic corn is derived from wild progenitor Teosinte. There is no historical record of how the breeding was done to produce Maize but there is a genetic record of the differences between Teosinte and Maize recorded the genomic differences between these two species. Maize and Teosinte can be crossed to give viable progeny.



How many genes contribute to the differences between the two kinds of plants?

Let's designate the genes that differ as **A, B, C, D ...**

For each gene there are two alleles: the allele present in Teosinte and the allele present in Maize.

For the **A** gene we will designate these alleles **A<sub>T</sub>** and **A<sub>M</sub>** respectively. For the **B** gene there will be alleles **B<sub>T</sub>** and **B<sub>M</sub>** and so on for all the genes that differ.

Let's follow the **A** gene through the cross between Maize and Teosinte

$$\begin{array}{c} A_T/A_T \times A_M/A_M \\ \downarrow \\ F_1: A_T/A_M \end{array}$$

Because the **F<sub>1</sub>** don't look like either parent, let's assume that the alleles are incompletely dominant.

**Incomplete dominance:** heterozygote expresses a trait intermediate between the traits of either homozygous parents.

(Alternatively, the genes that differ could have a mixture of dominant and recessive alleles)

$$\begin{array}{ccc} F_2: & A_T/A_T & A_T/A_M & A_M/A_M \\ & 1 & 2 & 1 \\ & 1/4 & 1/2 & 1/4 \end{array}$$

1/4 will look like Teosinte

For two genes that differ: **A<sub>T</sub>/A<sub>T</sub> B<sub>T</sub>/B<sub>T</sub>**

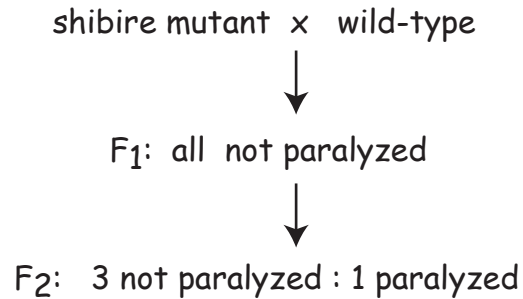
$$1/4 \times 1/4 = 1/16 \text{ will look like Teosinte}$$

Similarly, for three genes the probability will be 1/64. For four genes it will be 1/256, and for five genes it will be 1/1024.

Since ~1/500 look like Teosinte the conclusion is that 4-5 genes differ between wild corn (Teosinte) and domestic corn (Maize). Using modern methods, it has been confirmed that there are about five significantly different alleles and several of these have been located using mapping methods.

## Lecture 4

From the last lecture, we followed gene segregation in a cross between a true breeding strain with a shibire mutation and flies from a wild-type strain.



This is the segregation pattern expected for a single gene. But in an actual experiment how do we know that the phenotypic ratio is really 3 : 1 ?

There is no logical way to prove that we have exactly a 3 : 1 ratio. Nevertheless, we can think of an alternative hypothesis then show that the alternative hypothesis does not fit the data. Usually, we then adopt the simplest hypothesis that still fits the data.

A possible alternative hypothesis is that recessive mutations in two different genes are needed to get a paralyzed fly.

In this case a true breeding paralyzed fly would have genotype: **a/a , b/b**

Whereas wild-type would have genotype: **A/A , B/B**

F<sub>1</sub>: **A/a B/b** not paralyzed

F<sub>2</sub>:  $p(\mathbf{a/a \text{ and } b/b}) = (1/4)^2 = 1/16$

$p(\mathbf{a/a \text{ and } B/-}) = 1/4 \times 3/4 = 3/16$

$p(\mathbf{A/- \text{ and } b/b}) = 3/16$

$p(\mathbf{A/- \text{ and } B/-}) = \text{the rest} = 9/16$

This is the classic ratio for two gene segregation 9 : 3 : 3 : 1  
paralyzed

For our hypothesis we should see a phenotypic ratio of 15 not paralyzed : 1 paralyzed.

Therefore, to distinguish one-gene segregation from two-gene segregation we need a statistical test to distinguish 3 : 1 from 15 : 1. Intuitively, we know that in order to get statistical significance, we need to look at a sufficient number of individuals.

For a **chi-square test** you start with a specific hypothesis that gives a precise expectation. The test is then applied to the actual experimental results and will give the probability of obtaining the results under the hypothesis. The test is useful for ruling out hypotheses that would be very unlikely to give the actual results.

Say we look at 16 flies in the F<sub>2</sub> and observe 14 not paralyzed and 2 paralyzed flies.

Under the hypothesis of two genes we expect 15 not paralyzed flies and 1 paralyzed fly.

We calculate the value  $\chi^2$  using the formula below. Where O is the number of individuals observed in each class and E is the number of individuals expected for each class.

$$\chi^2 = \sum_{\text{(all classes)}} \frac{(O - E)^2}{E} = \frac{1}{15} + \frac{1}{1} = 0.067 + 1 = 1.067$$

degrees of freedom (df) = number of classes - 1

From the table using 1 df,  $0.05 < p < 0.5$

The convention we use is that  $p \leq 0.05$  constitutes a deviation from expectation that is significant enough to reject the hypothesis. Therefore, on the basis of this sample of 16 flies we can't rule out the hypothesis that two genes are required.

Say we look at 64 F<sub>2</sub> flies and find that 12 are paralyzed. For the hypothesis of two genes the expectation is that 4 would be paralyzed. The  $\chi^2$  for this data:

$$\chi^2 = \frac{8^2}{60} + \frac{8^2}{4} = 1.07 + 16 = 17.1$$

From the table  $p < 0.005$  so we reject the two-gene hypothesis.

Let's use this data to test the hypothesis of one gene segregation which would be expected to give 16 paralyzed flies from 64 F<sub>2</sub> flies.

$$\chi^2 = \frac{4^2}{48} + \frac{4^2}{16} = 0.33 + 1 = 1.33$$

From the table using 1 df,  $0.5 < p < 0.5$ . Thus the data still fits the hypothesis of one-gene segregation.

So far, the hypothesis that one gene is responsible for the paralyzed trait is the simplest explanation that fits the data.

The way to distinguish most easily between a heterozygote and a homozygote expressing a dominant trait is to cross to a homozygous recessive test strain.

**Test cross:** cross to homozygote recessive

$A/A \times a/a$  gives all  $A/a$  i.e. all offspring will express the dominant trait.

$A/a \times a/a$  gives  $1/2 A/a$  and  $1/2 a/a$  i.e. one half of the offspring will express the dominant trait.

### **Mendelian inheritance in humans**

For humans we can't do test crosses, of course, but by following inheritance of a trait for several generations the modes of inheritance can usually be identified by applying basic principles of Mendel. The following are guidelines for identifying different modes of inheritance in pedigrees.

#### **Autosomal dominant**

i) Affected individuals must have at least one affected parent

Exceptions to this rule will occur if a new mutation arises in one of the parents (in real life a more likely explanation is extramarital paternity). Another possibility is incomplete penetrance, where other genetic or environmental factors prevent the trait from being expressed in one of the parents.

ii) For rare dominant traits, if one parent is affected most likely half of the children will be affected.

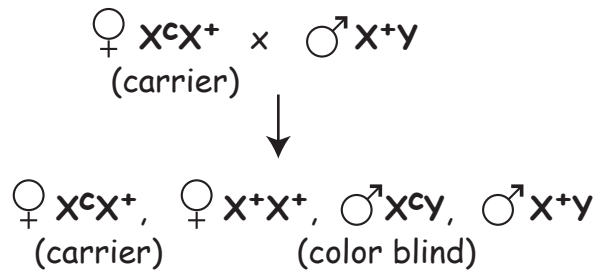
#### **Autosomal recessive**

i) When both parents are carriers, on average  $1/4$  of the children will be affected.

ii) When both parents are affected, then all of the children will be affected.

iii) If the trait is very rare then consanguinity is likely. That is, it is likely that parents of affected children are themselves related (e.g. cousins).

## X-linked inheritance



i) When parents are a carrier ♀ and an unaffected ♂, then on average, 1/2 of the daughters will be carriers and 1/2 of the sons will be affected.

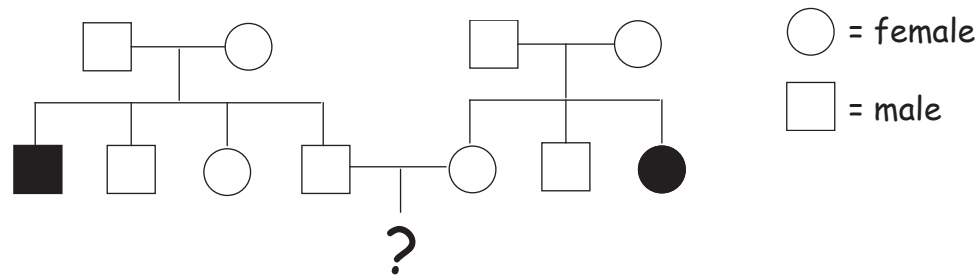
If the trait is rare then the vast majority of affected individuals will be male which is the hallmark of X-linked traits.

ii) Affected sons inherit the allele from mother

- Maternal uncles often affected
- Since inherited only from mother, inbreeding doesn't increase the probability of an affected ♂.

## Lecture 5

Consider the following pedigree of an autosomal recessive trait.



$p(\text{affected child}) = p(\text{mother carrier and father carrier and affected child})$

$$= 2/3 \times 2/3 \times 1/4 = 1/9$$

However, if they have a child that is affected we must reassess the probability that their next child will be affected.

$p(\text{both parents carriers}) = 1$ . So,  $p(\text{next child affected}) = 1/4$

This example shows how probability calculations are based on information. The probability changes not because the parents have changed but because our information about them has.

Now consider the case that the two parents have an unaffected child, with this new information we can recalculate the probability that the next child will be affected. An unaffected child does not establish definitively that both parents are **not** carriers, but it should be apparent that our estimate of the probability should be somewhat less than the original probability of  $4/9$ .

In order to calculate such probabilities we need to introduce a new concept known as **conditional probability**.

$p(X|Y)$  = probability that event **X** will occur given that **Y** has occurred

$$p(X|Y) = \frac{p(X \text{ and } Y)}{p(Y)}$$

Therefore:

$$p(X \text{ and } Y) = p(X|Y) \cdot p(Y)$$

## Bayes Theorem

We can use this identity to perform a very useful type of probability calculation. In many simple probability calculations we can readily calculate the probability of a particular outcome (effect) given a particular circumstance (cause). Bayes' theorem allows us to calculate the inverse probability of a particular cause based on an observed effect. This method depends on knowing the probabilities of the measured event occurring given each of the possible causes and on knowing the a priori probabilities of obtaining the conditions for each of the possible causes.

Stated in terms of conditional probabilities, Bayes' theorem allows us to express  $p(X|Y)$  in terms of  $p(Y|X)$ , where  $X$  represents a particular circumstance (cause) and  $Y$  represents an observed effect.

$$p(X|Y) = \frac{p(X \text{ and } Y)}{p(Y)} = \frac{p(Y | X) \cdot p(X)}{p(Y)}$$

Consider a simple case in which there are only two possible circumstances/hypotheses (ie both parents are carriers or both parents are not carriers). We will express the complement of  $X$  (not  $X$ ) as  $\bar{X}$ . Thus:

$$p(Y) = [p(Y|X) \cdot p(X)] + [p(Y|\bar{X}) \cdot p(\bar{X})]$$

$$p(X|Y) = \frac{p(Y | X) \cdot p(X)}{[p(Y|X) \cdot p(X)] + [p(Y|\bar{X}) \cdot p(\bar{X})]}$$

To apply this formula to the pedigree problem we will define  $X$  = both parents carriers; and  $\bar{X}$  = not both parents carriers;  $Y$  = first child unaffected.

Accordingly,  $p(X) = 4/9$ ,  $p(\bar{X}) = 5/9$ ,  $p(Y|X) = 3/4$ ,  $p(Y|\bar{X}) = 1$

$$p(X|Y) = \frac{3/4 \cdot 4/9}{3/4 \cdot 4/9 + 1 \cdot 5/9} = \frac{3/9}{3/9 + 5/9} = 3/8$$

The probability that the next child will be affected is  $3/8 \times 1/4 = 3/32 = 0.094$ , which is slightly less than the probability we calculated before the couple had an unaffected child  $1/9 = 0.111$

Here is an alternative way to express Bayes theorem, which shows how this expression fits well with common sense.

$$p(X|Y) = \frac{p(Y | X) \cdot p(X)}{p(Y)}$$

$$p(Y|\bar{X}) = \frac{p(Y | \bar{X}) \cdot p(\bar{X})}{p(Y)}$$

Thus:

$$\frac{p(X|Y)}{p(\bar{X}|Y)} = \frac{p(Y|X)}{p(Y|\bar{X})} \cdot \frac{p(X)}{p(\bar{X})}$$

Posterior odds = Likelihood ratio · Prior odds

Let's now consider two examples to show the power of Bayes' theorem to analyze a wide variety of interesting problems.

What is the probability of a misdiagnosis of AIDS infection using a diagnostic test for HIV that has both a false positive and false negative rate = 0.005.

The probability that an individual will test positive given they are infected  
 $p(\text{Pos} | \text{Inf}) = 0.995$

The probability that an individual will test positive given they are not infected  
 $p(\text{Pos} | \text{NInf}) = 0.005$

The a priori probability that an individual in the US is infected  $p(\text{Inf}) = 0.001$

Using Bayes' theorem:

$$p(\text{Inf} | \text{Pos}) = \frac{p(\text{Pos} | \text{Inf}) \cdot p(\text{Inf})}{p(\text{Pos} | \text{Inf}) \cdot p(\text{Inf}) + p(\text{Pos} | \text{NInf}) \cdot p(\text{NInf})}$$

$$= \frac{0.995 \cdot 0.001}{(0.995 \cdot 0.001) + (0.005 \cdot 0.999)}$$

$$= 0.16$$

The remarkable conclusion is that although the AIDS test has a very low error rate (for individuals who test positive there is a 99.5% chance that they have the disease) when the test is used broadly, only a minority of the positive tests would actually be AIDS cases.

The Monty Hall problem.

Three boxes - one contains keys to a new car

You initially get to pick a box; the probability you pick the box with the keys =  $1/3$

After you pick a box (call this box 1), without letting you see whether you have picked the box with the key, Monty Hall opens one of the remaining boxes (call this box 2), then gives you the option of choosing the remaining unopened box (call this box 3). Are you better off keeping box 1 or switching to box 3?

To analyze this problem we need precise rules for Monty Hall's behavior. Monty Hall knows where the key is and he will always show you an empty box. If you have picked the box with the key he will select another box at random.

Using Bayes' theorem to express the probability the key is in box 3:

$$p(\text{key in box 3} \mid \text{box 2 shown empty}) = \frac{p(\text{box 2 shown empty} \mid \text{key in box 3}) \cdot p(\text{key in box 3})}{p(\text{box 2 shown empty})}$$

$$\begin{aligned} p(\text{box 2 shown empty}) &= p(\text{box 2 shown empty} \mid \text{key in box 1}) \cdot p(\text{key in box 1}) + \\ &\quad p(\text{box 2 shown empty} \mid \text{key in box 2}) \cdot p(\text{key in box 2}) + \\ &\quad p(\text{box 2 shown empty} \mid \text{key in box 3}) \cdot p(\text{key in box 3}) \\ &= (1/2 \cdot 1/3) + (0 \cdot 1/3) + (1 \cdot 1/3) = 1/6 + 1/3 = 1/2 \end{aligned}$$

$$p(\text{key in box 3} \mid \text{box 2 shown empty}) = \frac{1/3}{1/2} = 2/3$$

Therefore the probability the key is in box 3 is twice the probability it is in box 1 and you would be better off taking Monty Hall's offer to switch boxes.