Chapter 7
Mendel & Morgan

Gregor Mendel

Introduction

In the middle of the nineteenth century, an Austrian monk, Gregor Mendel, toiled for almost 10 years systematically breeding pea plants and recording his results. Like many of his contemporaries, Mendel was intrigued with heredity and wanted to uncover the laws behind it. In 1865, just five years after the publication of Charles Darwin's *Origin of Species*, Mendel presented his results to the local natural history society in Brunn[1] which published his paper in their proceedings one year later.

To be honest, many historians surmise that Mendel's presentation and his paper were quite boring. They were crammed with numbers and percentages about green versus yellow peas, round versus wrinkled peas, axial versus terminal infloresences, yellow versus green pods, red-brown versus white seed coats, etc. To make matters worse, many of these traits were cross tabulated. Hence, his audience had to listen to numbers about round and yellow versus wrinkled and yellow versus round and green versus wrinkled and green pea plants. Perhaps as a consequence of this, no one paid attention to Mendel, and the basic principles of genetics that he elucidated in his paper and his presentation went unrecognized until shortly after the turn of the 20th century.

Mendel postulated three laws: (1) segregation, (2) dominance, and (3) independent assortment. Table 7.1 presents these laws and their definitions. In the following sections, we will examine some of Mendel's actual data and try to deduce how Mendel may have arrived at them.

Table 7.1. Mendel's three laws of heredity.

Law	Name	Explanation
1	Dominance	When two different hereditary factors are present, one will be dominant and the other will be recessive.
2	Segregation	Each organism has two hereditary factors and passes one of these, at random, to an offspring.
3	Independent Assortment	The hereditary factors for one trait (e.g., color of pea) are transmitted independently of the hereditary factors for another trait (e.g., shape of pea)/

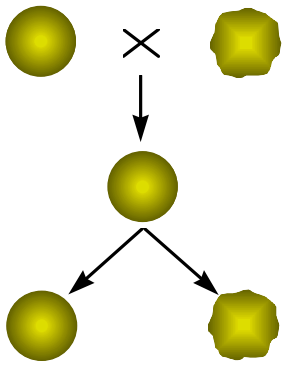
Mendel's Laws: Dominance

Figure 7.1 presents the results of one of Mendel's breeding experiments.[2] Mendel began with two lines of yellow peas that always bred true.[3] One line consistently gave

[1] Brunn is currently located in the Czech Republic.

[2] The actual breeding design was Mendel's dihybrid cross of round versus wrinkled and yellow versus green peas.

Figure 7.1. An example of Mendel's results.



round peas while the second always gave wrinkled peas. Mendel cross bred these two strains by fertilizing the round strain with pollen from the wrinkled strain and fertilizing the wrinkled strain with pollen from the round plants. The seeds from the next generation were all round. At this point, Mendel probably asked himself, "Whatever happened to the hereditary information about making a wrinkled pea?"

Mendel did not stop at this point. He cross-fertilized all the pea plants in this generation with pollen from other plants in the same generation. When there progeny matured, he noticed a very curious phenomenon—wrinkled peas reappeared! How could this happen when all the parents of these plants had round seeds? Obviously, the middle generation in Figure 7.1, despite being all round, still possessed the hereditary information for making a wrinkled pea. But somehow that information was not being expressed. Hence, Mendel surmised, some hereditary factors are dominant to other hereditary factors. In this case, a round shape is dominant to a wrinkled shape.

Mendel's Laws: Segregation

Figure 7.1 gives the number and percentage of round and wrinkled peas in the third generation. There were just about three round seeds for every wrinkled seed. This ratio of 3 plants with the dominant hereditary factor to every one plant with the recessive factor kept coming up time and time again in Mendel's breeding program. There were 3.01 yellow seeds for every green seed; 3.15 colored flowers to every white flower; 2.95 inflated pods to every constricted pod. Mendel must have spent considerable time cogitating over a sound, logical reason why the 3 to 1 ratio should always appear.

To answer this question, let us return to Figure 7.1 and consider the middle generation in light of the law of dominance. Obviously, because they are themselves round, these plants must have the hereditary information to make a round pea. They must also have the hereditary information to make a wrinkled pea because one quarter of their progeny are wrinkled. Mendel's stroke of genius lay in applying elementary probability theory to this generation—what would one expect if these two types of hereditary information were discrete and combined at random in the next generation?

The situation for this is depicted in Table 7.2 where *R* denotes the hereditary information for making a round pea and *W*, the information for a wrinkled pea. The probability that a plant in the middle generation transmits the *R* information simply the probability of a heads on the flip of a coin or 1/2. Thus, the probability of transmitting the

---

[3] Indeed, Mendel's discoveries, like many important scientific breakthroughs, were partly the result of dumb luck. Without knowing the actual genetics behind what they were doing, people had been inbreeding peas for many years, creating what are now known as *inbred strains*. An inbred strain is a strain that is homozygous at all loci. These inbred strains formed the starting point of Mendel's breeding designs.

W information is also 1/2. Hence, the probability that the male parent transmits the *R* information and the female parent also transmits the *R* information is 1/2 x 1/2 = 1/4. All of the progeny that receive two *R*s will be round. Similarly, the probability that the male transmits *R* while the female transmits *W* is also 1/4 as is the probability that the male transmits *W* and the female, *R*. Thus, 1/4 + 1/4 = 1/2 of the offspring will have both the *R* and the *W* information. They, however, will all be round because *R* dominates *W*. Thus far, 3/4 of all the offspring will be round.

Table 7.2. Illustration of Mendel's law of segregation for round (*R*) and wrinkled (*W*) hereditary factors in peas.

		Female Parent:	
Male Parent:		*R* information 1/2	*W* information 1/2
R information	1/2	*RR* 1/2	*RW* 1/4
W information	1/2	*WR* 1/4	*WW* 1/4

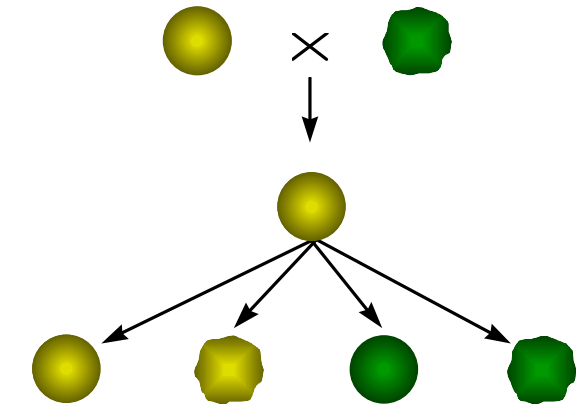
The only other possibility is that of both the male and female plant transmitting *W*. The probability of this is also 1/2 x 1/2 = 1/4. So the remaining 1/4 of the progeny will receive only the hereditary information on making a wrinkled pea and consequently will be wrinkled themselves. "Voila!" Mendel must have thought, "The hereditary factors are discrete. Every plant has two hereditary factors and passes only one, at random, to an offspring."

Mendel's Laws: Independent Assortment

A scientist who achieves success using one particular technique always uses that technique in the initial phase of solving the next problem. Mendel was probably no exception. His success in using the mathematics of probability to develop the law of segregation undoubtedly influenced his approach to his next problem, that of dealing with two different traits at once.

Figure 7.2 gives the results of his breeding round, yellow plants with wrinkled green plants and keeping track of both color and shape in the subsequent generations. The middle generation is all yellow and round. Hence, yellow is the dominant hereditary factor for color; and round, as we have seen, is the dominant for shape. The next generation is both confirmatory and troubling. First, there are a total of 423 round and 133 wrinkled peas, giving a ratio of 3.18 to 1, very close to the 3:1 ratio expected from the laws

Figure 7.2. Mendel's data for a dihybrid cross between round, yellow peas and wrinkled, green peas.



of dominance and segregation. Also confirming these predictions is the 2.97 to 1 ratio of yellow to green peas.

This generation, however, has combinations of traits not seen in either of the previous generations—wrinkled, yellow peas and round, green peas. What can explain this? Once again, an hypothesis and the mathematics of probability gave a solution. Mendel's hypothesis was that the hereditary factors for pea color are independent of the hereditary factors for pea shape. The mathematical calculations for deriving the expected number of plants of each type in the third generation are more complicated than those for deriving the law of segregation, but they follow the same basic principles of probability theory.

The middle generation will have four discrete hereditary factors, the *R* and *W* factors that we have already discussed and the *Y* (for yellow) and *G* (for green) factors that determine color. If the hereditary factors for shape are independent of those for color, then the plants in the second generation can give any of four different combinations to their offspring. The combinations are *RY*, *RG*, *WY*, and *WG* and the probability of passing on any single combination is 1/4 (see Table 7.3).

Table 7.3. The four types of gametes from the middle generation of Figure 7.2 if the hereditary factors for shape assort independently of those for color.

	Color Information:			
Shape Information:	*Y*	1/2	*G*	1/2
R 1/2	*RY*	1/2	*RG*	1/4
W 1/2	*WY*	1/4	*WG*	1/4

Table 7.4 gives the results of these four types of gametes, as we now call them, coming together in the offspring in the third generation. Nine-sixteenths of the offspring are expected to be round and yellow, three-sixteenths will be round and green, another three-sixteenths will be wrinked and yellow, and the remaining one-sixteenths should be wrinkled and green. Indeed, this expected 9:3:3:1 ratio is very close that observed by Mendel. Were this to happen today, Mendel would have high-fived all the monklettes who helped him to plant and to count the peas (and probably to harvest and dine on them as well) and dashed off a paper to *Science* or *Nature*. Instead, he patiently tabulated his results and embarked on the lengthy journey to Brunn.

Table 7.4. Expected frequencies of peas in the third generation if pea shape and pea color assort independently.

	Female gametes:			
Male:	RY 1/4	WY 1/4	RG 1/4	WG 1/4
RY 1/4	RR,YY	RW,YY	RR,YG	RW,YG
WY 1/4	WR,YY	WW,YY	WR,YG	WW,YG
RG 1/4	RR,GY	RW,GY	RR,GG	RW,GG
WG 1/4	WR,GY	WW,GY	WR,GG	WW,GG

Mendel's Laws: Exceptions

The majority of seminal scientific discoveries never get things completely right. Instead, they turn science in a different direction and make us think about problems in a different way. It often takes years of effort to fill in the fine points and find the exceptions to the rule. Mendel's laws follow this pattern. None of the three laws is completely correct. We know now that some hereditary factors are codominant, not completely dominant, to others; one can cross red with white petunias and get pink offspring, not red or white offspring as Mendel would have predicted. We also know that the law of segregation is not always true in its literal sense. In humans, the X and the Y chromosome are not passed along entirely at random from a father—slightly more boys than girls are conceived. And we also know that not all hereditary factors assort independently. Those that are located close together on the same chromosome tend to be inherited as a unit, not as independent entities.

These exceptions are individual trees within the forest. Mendel's great accomplishment was to orient science toward the correct forest. Hereditary factors do not "blend" as Darwin thought; they are discrete and particulate, as Mendel postulated. As Mendel conjectured, we have two hereditary factors, one of which we received from our father and the other from our mother. We do not have 23 hereditary factors, one on each chromosome, as the early cell biologist Weissman theorized. And two different hereditary factors, provided that they are far enough away on the same chromosome or located on entirely different chromosomes, are transmitted independently of each other. Mendel's basic concepts provided a paradigm shift and sparked the nascent science of genetics at the turn of the century, an achievement that the humble monk was never recognized for during his life.

Application of Mendel's Laws: The Punnett Rectangle

In high school biology, many of you were exposed to a Punnett square. Indeed, we have seen examples of the Punnett square in Tables 7.2, 7.3, and 7.4. High school has also taught us that a square is a specific case of a more general geometric form, the rectangle. We can now generalize and develop the concept of a Punnett rectangle to apply Mendel's concepts of heredity to calculating the genotypes and phenotypes of offspring from a given mating type. The formal rules for constructing a Punnett rectangle are given in Table 7.5. They should be carefully reviewed.

Table 7.5. Steps in constructing a Punnett rectangle.

Step	Operation
1	Write down the genotypes of the mother's gametes along with their associated probabilities. These will label the rows of the rectangle.
2	Write down the genotypes of the father's gametes along with their associated probabilities. These will label the columns of the rectangle. (Naturally, it is possible to switch the steps—mother's gametes forming the columns and father's gametes, the rows—without any loss of generality.)
3	The genotype for each cell within the rectangle is the genotype of the row gamete united with the genotype of the column gamete. Enter these for all cells in the rectangle
4	The probability for each cell within the rectangle equals the probability of the row gamete multiplied by the probability of the column gamete. Enter these for all cells in the rectangle.
5	Make a table of the unique genotypes of the offspring and their probabilities. If a genotype occurs more than once within the cells of the Punnett rectangle, then add the cell probabilities together to get the probability of that genotype.
6	If needed, make a table of the unique phenotypes and their probabilities from the table of genotypes. If two or more genotypes in the genotypic table give the same phenotype, then add the probabilities of those genotypes together to get the probability of the phenotype.

The Punnett Rectangle: An example

As an example of this technique, consider the mating of woman who has genotype *AO* at the ABO blood group locus and a man who also has genotype *AO*. From Step 1, the mother's egg can contain either allele *A* (with a probability of 1/2) or allele *O* (also with a probability of 1/2). From Step 2, the father's sperm may contain either *A* or *O* , each with a probability of 1/2. The row and column labels for this Punnett rectangle given in Table 7.6.

Table 7.6. Setting up the row and column labels for a Punnett rectangle for a mating between an *AO* mother and an *AO* father.

		Father's gametes and probability:	
		A 1/2	*O* 1/2
Mother's gametes and probability:	*A* 1/2		
	O 1/2		

Step 3 requires entry of the offspring genotypes into the cells of the rectangle. The upper left cell represents the fertilization of mother's *A* egg by father's *A* sperm, so the genotypic entry is *AA*. The upper right cell represents the fertilization of mother *A* egg by father's *O* sperm, giving genotype *AO*. Following these rules, the genotypes for the lower left and lower right cells are, respectively, *OA* and *OO*.

The probabilities for Step 4 require us to multiply the row probability by the column probability. For the upper left cell, the row probability is 1/2 and the column

probability is also 1/2, so the cell probability is 1/2 x 1/2 = 1/4. It is obvious that each cell for this example will have a probability of 1/4. The completed Punnett rectangle is given in Table 7.7.

Table 7.7. Punnett rectangle for a mating between an *AO* mother and an *AO* father.

			Father's gametes and probability:			
			A	1/2	*O*	1/2
Mother's gametes	*A*	1/2	*AA*	1/4	*AO*	1/4
and probability:	*O*	1/2	*OA*	1/4	*OO*	1/4

In Steps 5 and 6, we must complete a table of genotypes and, if needed, phenotypes. To complete Step 5, we note that there are three unique genotypes in the cells of the Punnett rectangle—*AA*, *AO*, and *OO*. Genotypes *AA* and *OO* occur only once, so the probability of each of these genotypes is 1/4. Genotype AO, on the other hand, occurs twice, once in the upper right cell and again in the lower left. Consequently, the probability of the heterozygote *AO* is the sum of these two cell probabilities or 1/4 + 1/4 = 1/2. Table 7.8 gives the genotypes and their frequencies.

Table 7.8. Expected offspring genotypes and their probabilities from an *AO* by *AO* mating.

Genotype	**Probability**
AA	1/4
AO	1/2
OO	1/4

Step 6 requires calculation of the phenotypes from the genotypes. Because allele *A* is dominant in the ABO blood system, both genotypes *AA* and *AO* will have phenotype *A*. The probability of phenotype *A* will be the sum of the probabilities for these two genotypes or 1/4 + 1/2 = 3/4. Genotype *OO* will have phenotype *O*. Because genotype *OO* occurs only once, the probability of phenotype *O* is simply 1/4. The complete table of genotypic and phenotypic frequencies is given below in Table 7.9.

Table 7.9. Expected offspring genotypes and their probabilities from an *AO* by *AO* mating.

Genotype	**Probability**		**Phenotype**	**Probability**
AA	1/4		*A*	3/4
AO	1/2		*O*	1/4
OO	1/4			
	Sum = 1.0			Sum = 1.0

The column probabilities should always sum to 1.0. Although this is visually obvious for the present example, it is recommended that the sums be calculated to avoid errors, especially for more complicated problems.

The Punnett rectangle: A two locus example.

The logic of the Punnett rectangle may be applied to genotypes at more than one locus. The only requirement is that all the loci are unlinked, i.e., no two loci are located close together on the same chromosome. (Later on, we deal with the case of Punnett rectangles with linked loci.) The Punnett rectangle for two loci will be illustrated by calculating traditional blood types.

Traditional blood typing for transfusions use phenotypes at two genetic loci. The first is the ABO locus and the second is the rhesus locus. Although the genetics of the rhesus locus are actually quite complicated, we will assume that there are only two alleles, a "plus" or + allele and a "minus" or - allele[4]. The + allele is dominant to the - allele, so the two rhesus phenotypes are + and -. The blood types used for transfusions and blood donations concatenate the ABO phenotype with the rhesus phenotype, giving phenotypes such as *A*+, *B*-, *AB*+, etc. What are the expected frequencies for the offspring of a father with genotype *AO*/+- (read "genotype *AO* at the ABO locus and genotype +- at the rhesus locus") and a mother who is genotype *OO*/+-?

The trick to the problem of two unlinked loci is to go through the Punnett rectangle steps three times. In the first pass, the Punnett rectangle is used to obtain the genotypes for the maternal gametes. The second pass calculates the paternal gametes, and the third and final pass uses the results from the first two passes to get the offspring genotypes.

The mother in this problem has genotype *OO*/+-. Because the ABO and rhesus loci are unlinked, the probabilities for the ABO locus are independent of those for the rhesus locus. This permits us to use a Punnett rectangle to derive the maternal gametes. The rows of the rectangle are labeled by the contribution to the maternal egg from mother's ABO alleles and their probabilities. Here, mother can give only an *O*. Consequently there will be only one row to the rectangle, and it will be labeled *O* and have a probability of 1.0. The columns are labeled by the contribution of mother's rhesus alleles and their probabilities. In the present case, the columns will be labeled + and -, each with probability 1/2. The completed Punnett rectangle needed to get the maternal gametes and their probabilities is given below in Table 7.10.

Table 10.10. Maternal gametes from a mother who is *OO*/+-

			Rhesus allele and probability:			
ABO allele			+	1/2	-	1/2
and probability	*O*	1.0	*O*+	1/2	*O*-	1/2

From this table, we see that there are two possible maternal gametes, each with a probability of 1/2. The first has one of mother's *O* alleles at the ABO locus and the + allele at the rhesus locus, giving gamete *O*+. The second gamete, *O*-, contains one of mothers *O* alleles but her - allele at rhesus.

[4] In reality, there is more than two alleles for the rhesus locus, just as there are more than three alleles for the ABO locus. We used the outdated rhesus and ABO systems because treating all the alleles for both blood groups introduces unneeded complexity for the introductory student.

The second pass through the Punnett rectangle is used to calculate father's gametes. His contribution to the gamete from his alleles at the ABO locus may be either allele *A*, with probability 1/2, or *O*, also with probability 1/2. Thus, there will be two rows to father's Punnett rectangle, one labeled *A* and the other labeled *O*, each with probability 1/2. Like mother, father is a heterozygote at rhesus. Hence the columns for his table will equal those for the maternal table. The Punnett rectangle for the paternal gametes is given below in Table 7.11.

Table 7.11. Father's gametes when father is genotype *AO*/+-.

			Rhesus allele and probability:			
			+	1/2	-	1/2
ABO allele and probability:	*A*	1/2	*A*+	1/4	*A*-	1/4
	O	1/2	*O*+	1/4	*O*-	1/4

Father has four different paternal gametes, each with a probability of 1/4. The first possible gamete carries father's *A* allele at the ABO locus and father's + allele at the rhesus locus (*A*+). Analogous interpretations apply to the remaining three gametes—*A*-, *O*+, and *O*-.

We may now construct the last Punnett rectangle to find the offspring genotypes and their frequencies. In this rectangle, the rows are labeled by the maternal gametes and their probabilities and the columns by the paternal gametes and their probabilities. This will give a rectangle with two rows and four columns. The completed rectangle is given below in Table 7.12.

Table 7.12. Punnett rectangle for offspring genotypes.

			Father's gametes and probability:							
			A+	1/4	*A*-	1/4	*O*+	1/4	*O*-	1/4
Mother's gametes and probability:	*O*+	1/2	*AO*/++	1/8	*AO*/+-	1/8	*OO*/++	1/8	*OO*/+-	1/8
	O-	1/2	*OA*/-+	1/8	*OA*/--	1/8	*OO*/-+	1/8	*OO*/--	1/8

From this Punnett rectangle, we can the table of expected genotypes, expected phenotypes, and their frequencies. This is given below. In Table 7.13.

Table 7.13. Expected offspring genotypes, phenotypes and probabilities from a mating of genotypes *OO*/+- and *AO*/+-.

Genotypes:	**Probability:**		**Phenotypes:**	**Probability:**
AO/++	1/8		*A*+	3/8
AO/+-	2/8		*A*-	1/8
AO/--	1/8		*O*+	3/8
OO/++	1/8		*O*-	1/8
OO/+-	2/8			
OO/--	1/8			

Hence, 3/8 or 37.5% of the offspring are expected to have blood type *A*+; 1/8 or 12.5% will have blood type *A*-; 3/8 or 37.5 will have blood type *O*+; and 1/8 or 12.5% will have blood type *O*-.

The Punnett Rectangle: X-linked loci..

Genetic females, of course, have two *X* chromosomes while genetic males have one *X* chromosome and one *Y* chromosome. The *Y* chromosome is much smaller than the *X* chromosome and contains many fewer loci than the *X*. Thus, many genes on the *X* chromosome—in fact, the overwhelming number of genes—do not have counterparts on the *Y*[5]. Such loci are called *X*-linked genes. For such genes, females have two alleles, one on each *X*, whereas males have only one allele, the one on their single *X* chromosome.

The trick to dealing with this situation in a Punnett squares is to include the chromosome when writing a gamete. Hence, if a mother is genotype *Aa* for a locus on the *X* chromosome, we will write her gametes as *X*-*A* and *X*-*a* instead of just *A* and *a*. A father who is genotype *A*, for example, will have his games written as *X*-*A* and *Y*. Note that father's gamete containing the *Y* chromosome does not have an allele associated with it because the locus does not exist on the *Y*. To illustrate this technique, consider the offspring from a mating between an *Aa* female and an *a* male where allele *a* causes hemophilia, a locus located on the *X* chromosome but not on the *Y* chromosome. Hemophilia is a recessive disorder, so in phenotypic terms, this mating is between a hemophiliac male and a normal, but carrier, female. The Punnett rectangle appears below in Table 7.14.

Table 7.14. Punnett rectangle for a mating between an *Aa* woman and an *a* man for a sex-linked locus.

		Father's gametes and probability:	
		X-*a* 1/2	*Y* 1/2
Mother's gametes	*X*-*A* 1/2	*XX*-*Aa* 1/4	*XY*-*A* 1/4
and probability:	*X*-*a* 1/2	*XX*-*aa* 1/4	*XY*-*a* 1/4

The table of expected genotypes, phenotypes, and probabilities is:

Table 7.15. Expected genotypes, phenotypes, and probabilities for the offspring of a male hemophiliac and a female carrier for hemophilia.

Genotype:	**Probability:**		**Phenotype:**	**Probability:**
XX-*Aa*	1/4		female, carrier	1/4
XX-*aa*	1/4		female, hemophilia	1/4
XY-*A*	1/4		male, normal	1/4
XY-*a*	1/4		male, hemophilia	1/4

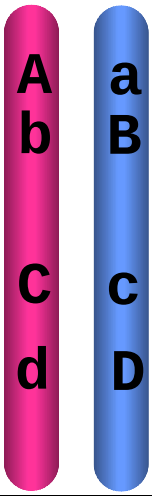
[5] There are a few regions of the *Y* that do contain loci found on the *X*. These genes are called pseudoautosomal loci.

Thomas Hunt Morgan[6]

Introduction

Thomas Hunt Morgan was a famous geneticist who, in the initial years of the 20th century, studied *Drosophila*, the fruit fly, in his lab at Columbia University in New York City. Morgan's choice of *Drosophila* was both fortuitous and prescient not only for his own historic findings but also for introducing a model organism that has evolved to become a major workhorse in the science of genetics. These tiny flies reproduce quickly and leave a large number of progeny. While Mendel had to wait months to plant and harvest two generations of peas, Morgan could study have a dozen generations in the same time. Moreover, Drosophila have only three chromosomes.

Figure 7.3. The linear arrangement of genes along a chromosome.



Two of Morgan's many findings stand out. Despite all the complicated looping of the DNA around chromosomal proteins, Morgan found that the genes on a chromosome have a remarkable statistical property –namely, statistically genes appear as if are linearly arranged along the chromosome. Thus, one can draw a schematic of a chromosome as a single strait line, with the genes in linear order along that straight line, even though the actual physical construction of the chromosome is a series of looped and folded DNA. Figure 7.3 provides an example of the linear arrangements of genes on chromosomes. Morgan's second finding was no less important. He discovered that chromosomes can recombine and exchange genetic material.

Recombination

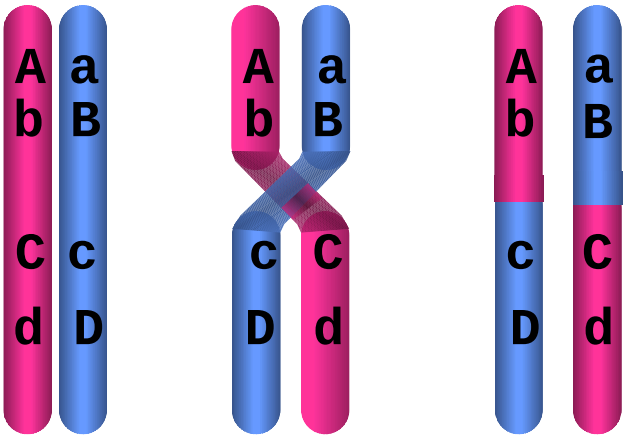
Recombination or *crossing over*, as it also called, refers to the fact that in the genesis of a sperm or egg, the maternal chromosome pairs with its counterpart paternal chromosome and exchanges genetic material. We have already discussed recombination in Chapter 2 under the topic of meiosis. Here, we will deal with the statistical implications of crossing over. The process is diagrammed in Figure 7.4.

The probability that a recombination event occurs between two loci is a function of the distance between the two loci. The alleles at two loci that are far apart on a chromosome are more likely to recombine than the alleles for two that are close together on the chromosome. For example, in Figure 7.3, it is more likely that loci A and D will have a recombination event between them that it is that loci A and B will involve a recombination. To same the same thing in a different way, the alleles for two genes that are physically close together on a chromosome are more likely to be inherited as a unit than the alleles for two genes that are far apart on the same chromosome.

---

[6] This part of the chapter contains some advanced material. For the undergraduate student it is important to know the following: (1) who Morgan was; (2) what recombination is; (3) that recombination is a function of physical distance; (4) the definition of a linkage study; (5) the meaning of the recombination fraction, θ; and (6) the meaning of $\theta = 0$ and $\theta = .5$. It is also important to inspect the pedigrees and read the text in order to arrive at some appreciation for linkage analysis. You will not be given data to compute θ, a lod score, or the probability of offspring from a mating using linkage.

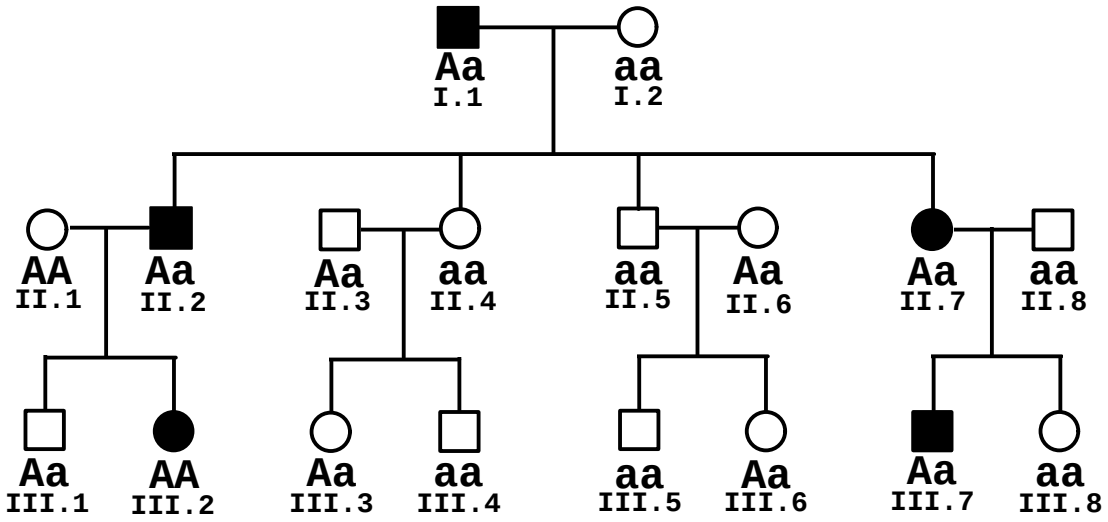
Figure 7.4. Recombination of homologous chromosomes. The maternal (pink) and paternal (blue) chromosomes pair up (left panel) and exchange segments of DNA (middle panel). The resulting chromosomes that can be transmitted via sperm or egg is a combination of the parent's maternal and paternal chromosomes (right panel).



The facts that (1) genes are arranged linearly on chromosomes and (2) recombination is a function of the physical distance between two genes permit human geneticists to locate genes for disorders and traits among the 23 human chromosomes. The procedure for doing this is called *linkage analysis*, a statistical technique for *tracing the within-pedigree cosegregation of one or more genetic markers with a trait.* Careful wording has been used to construct this definition, so let us take it apart, phrase by phrase, starting from the end and working backwards. The *trait* in human linkage analysis is usually a disorder. (The special case in which the trait is a score on a continuous variable will be discussed later in the text). The *markers* are genes whose location on a particular chromosome are already known. The term *cosegregation* refers to the fact that the allele for the trait and the allele(s) for the marker(s) tend to be transmitted together because they are physically close together on the same chromosome. Finally, the *within-pedigree* phrase denotes that the cosegregation of the marker allele(s) and the trait allele takes place *within* a pedigree.

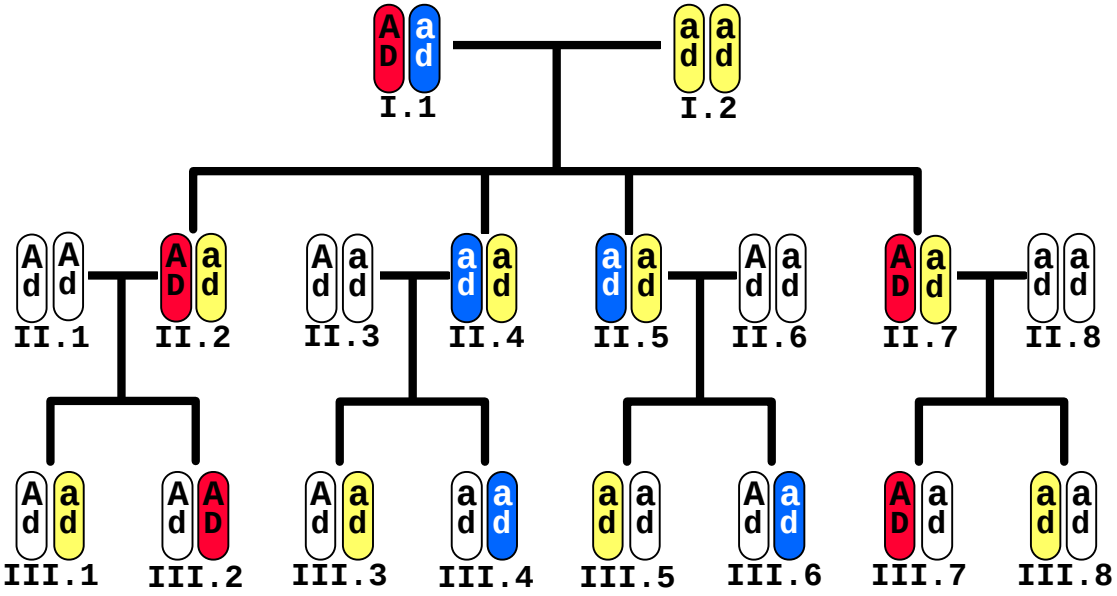
The pedigree depicted in Figure 7.5 illustrates the principles of linkage between a marker locus and a gene for a rare dominant disorder. For the disorder, alleles D causes the disorder while allele d is the normal allele. The two marker alleles are denoted as A and a. The grandfather in this pedigree has *haplotype* AD/ad. This means that

Figure 7.5. An example of linkage between a marker locus with alleles A and a and a dominant disorder with alleles D (which causes the disorder) and d (the normal allele). The grandfather is haplotype AD/ad and the grandmother is haplotype ad/ad.



grandfather has allele A on the same chromosome that contains allele D and, conversely, allele a on the chromosome with d. Grandmother is haplotype ad/ad. In generation II, the two offspring that received grandfather's allele A (i.e., persons II.2 and II.7) also have the disorder because they inherited the AD chromosome from their father. The two aa offspring in generation II (i.e., II.4 and II.5) are unaffected because they inherited the ad chromosome from their father.

Figure 7.6. The pedigree in Figure 7.5 expressed in terms of chromosomes.



The grandchildren are consistent with linkage. Granddaughter III.2 received the AD chromosome from her father which was transmitted to him by his father. Her marker genotype is AA because she inherited the other A from her mother. Her brother, III.1, deserves some mention. He has marker genotype Aa, the same marker genotype as his affected father, his affected grandfather, and his affected aunt (II.7). Why is he not affected? The reason is that he received his A allele from his unaffected mother, not from his father. His father gave the ad chromosome that he inherited from his mother (I.2). Note that two other grandchildren also have marker genotype Aa but are unaffected (III.3 and III.6). Once again, this is due to their inheritance of allele A from unaffected parents, II.3 and II.6, respectively. Grandson III.7 is affected because he inherited the AD chromosome from his mother.

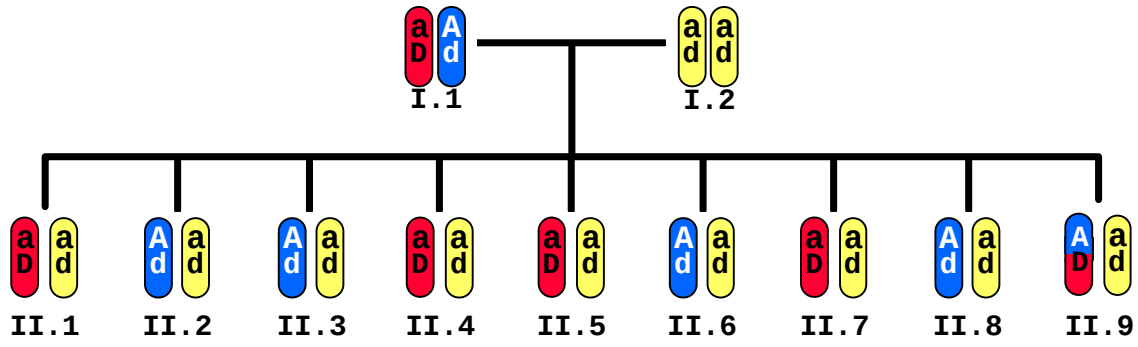
It takes time and considerable mental concentration to trace alleles through pedigrees, so do not despair if you find the description of this pedigree somewhat confusing. To assist you in decoding it, examine Figure 7.7 which gives a color-coded schematic for the transmission of the chromosomes. The red chromosome is the one that causes the disorder because it has allele D, so everyone in the pedigree with a red chromosome will have the disorder. Grandfather's other chromosome is blue, both of grandmother's chromosomes are yellow, and the chromosomes of people who marry into the pedigree are white. Notice how all the unaffected grandchildren with genotype Aa inherited the A allele from someone who married into the pedigree.

Advanced Topics

Estimating recombination

This pedigree in Figure 7.6 is idealized in the sense that no recombination has

Figure 7.8. Linkage between a marker locus and a dominant disorder illustrating a recombinant (person II.9).



taken place between the A and the D loci. The situation becomes more complicated when the trait locus and the marker are far enough apart that recombination can occur. This situation is illustrated by the pedigree in Figure 7.8. Note how in this pedigree the allele for the disorder, D, is on the same chromosome as allele a whereas in the previous pedigree, D was located on the same chromosome as marker allele A. Consequently affected offspring, with the expectation of II.9, have haplotype[7] aD/ad. This is the key point about linkage—in some pedigrees the disease allele, D, will be associated with marker allele A while in other pedigrees it will be associated with marker allele a. The central feature is that *within any single pedigree, the disease allele will consistently cosegregate with the same marker allele*. This is what is meant by the phrase *within-pedigree cosegregation* in the definition of linkage.

Offspring II.9 in Figure 7.8 deserves comment. This person is a termed a *recombinant* because the chromosome that he/she inherited from parent I.1 has undergone a recombination between the marker and the disease locus. Instead of receiving either a whole red chromosome or a whole blue chromosome, this person has been given the top portion of the blue chromosome recombined with the lower portion of the red chromosome.

Recombination is measured by the recombination fraction customarily denoted by θ (lowercase Greek theta). Many people working in the field of genetics define θ as the probability of a recombination between two loci, and even though that definition is not quite correct, we shall use it here[8]. The lowest possible value for θ is 0 and its upper limit is .5. When θ = 0, there is no recombination between the marker and the trait locus. In practice, this would mean one of two things: (1) either the maker is actually the trait

---

[7] A *haplotype* is defined as the alleles on a chromosome. If a person is haplotype Ab/aB, then that person has one chromosome with alleles A and b on it and a second chromosome with alleles a and B on it.

[8] Strictly, θ is a conditional probability and answers the following question: given that one inherits the allele at locus 1 from the maternal chromosome, what is the probability that one will inherit the allele at locus 2 from the paternal chromosome?

locus, or (2) the two loci are so close together that they rarely recombine and the estimate of θ rounds off to 0. When θ = .5, then the loci are not linked. The marker and trait loci are either on entirely different chromosomes or they are so far apart on the same chromosome that they act as independently assorting loci.

Conceptually, it is quite easy to estimate θ--simply count the number of recombinants among offspring and divide by the total number of offspring. For example, the pedigree in Figure 7.8 has one recombinant offspring and nine total offspring, so the estimate of θ is 1/9 or .11. In practice, however, most geneticists use sophisticated computer algorithms that simultaneous estimate θ and test for linkage.

Our old friend the Punnett rectangle can be used to find the expected frequency of offspring genotypes in the presence of linkage. The logic remains the same—one parent's gametes and their probabilities form the rows of the rectangle while the other parent's games and probabilities form the columns. The only trick is that the probabilities of the gametes are a function of θ.

As an example, consider a father who has haplotype AD/ad and a mother with haplotype ad/ad. All of mother's gametes will be genotype ad. Father, however, may have one of four different gametes, given by the rows in Table 7.16. The probability that father gives gamete AD equals the probability of two independent events—that probability father gives allele A (versus allele a) at the marker (which equals 1/2) and the probability that a recombination does not occur between the marker and trait locus which equals the quantity (1 - θ). Mathematically, the probability father transmits AD is the product of these two probabilities or 1/2(1 - θ).

Similarly, the probability that father's gamete has genotype Ad is the product of the probabilities of two independent events—the probability that father transmits A at the marker (again, 1/2) and the probability that a recombination occurs between the marker and the trait locus (or θ in this case). Consequently, the probability of transmitting gamete Ad equals 1/2θ. Similar logic will give the probability for father's gametes aD and ad. It is not necessary to understand the probability theory that goes into these calculations. The important point is that the logic of the Punnett rectangle can be applied to the case of linkage just as it can be applied to the case of two independently assorting loci.

Table 7.16. Expected frequency of children from a mating where father is haplotype AD/ad and mother is haplotype ad/ad.

			Mother's gametes:	
Father's gametes:	Recombinant?	Probability	ad Probability = 1.0	
AD	No	1/2(1 - θ)	AD/ad	1/2(1 - θ)
Ad	Yes	1/2θ	Ad/ad	1/2θ
aD	Yes	1/2θ	aD/ad	1/2θ
ad	No	1/2(1 - θ)	ad/ad	1/2(1 - θ)

Table 7.17 gives numerical estimates of the expected frequency among offspring as an function of the recombination fraction, θ. When θ = 0, father can give only two types of gametes, AD or ad. As θ increases, the proportion of recombinants (i.e., Ad and aD)

to nonrecombinants (i.e., AD and ad) increased. When θ reaches its upper limit of .5, then all four types of gametes are equally probable. Compare this column in Table 7.17 with the Punnett rectangle in Table 7.4 used to illustrate independent assortment in Mendel's peas. It is clear that when θ = .50, the two loci assort independently.

Table 7.17. Expected frequency of children from a mating where father is haplotype AD/ad and mother is haplotype ad/ad as a function of the recombination fraction, θ.

			θ =					
Offspring Haplotype:	Recombinant?	Probability	0	.10	.20	.30	.40	.50
AD/ad	No	1/2(1 - θ)	.50	.45	.40	.35	.30	.25
Ad/ad	Yes	1/2 θ	.00	.05	.10	.15	.20	.25
aD/ad	Yes	1/2 θ	.00	.05	.10	.15	.20	.25
ad/ad	No	1/2 (1 - θ)	.50	.45	.40	.35	.30	.25

Detecting Linkage

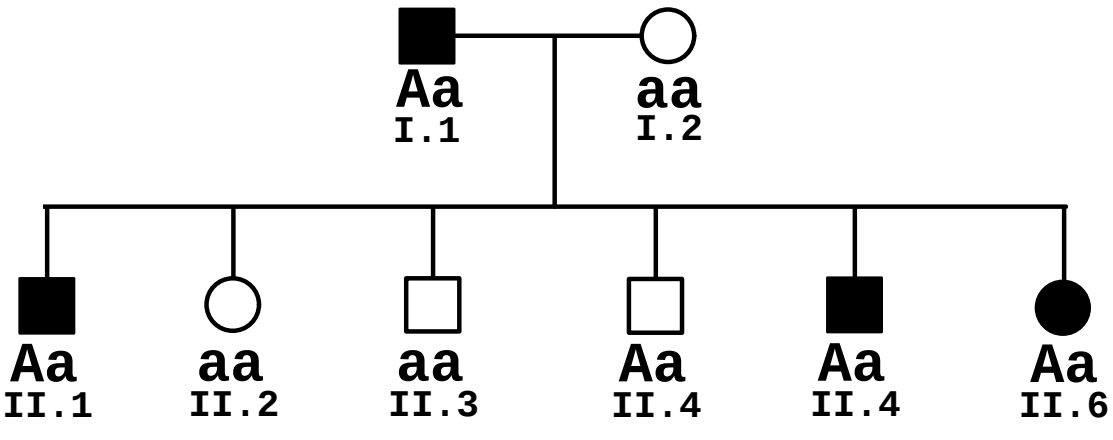
Dramatic advances in genetics have given us over a thousand different marker loci scattered over the 23 human chromosomes. When geneticists have a genetic disorder but do not know where the gene is, they gather a large number of families in which the disorder runs and obtain some biological specimen (usually blood, but sometimes cheek scrapings) from each member. DNA is extracted from the biological specimen and then genotyped on a large number of the marker loci. Ideally, marker loci are selected so that they are evenly spaced but cover each and every chromosome. Sophisticated mathematical procedures, implemented in equally sophisticated computer algorithms, are then used to test for linkage through the genome. This type of procedure is called a *whole genome scan*.

Although we treated linkage as the study of a single marker with a trait locus, in practice geneticists prefer to examine several linked markers simultaneously, a procedure known as *multipoint linkage*. Multipoint linkage has the major statistical advantage of being able to detect linkage more powerfully than single point linkage (the analysis of only a single marker). The actual mathematics of multipoint linkage are too complicated for us to explore here, but once again, the Punnett rectangle could still be used to calculate the expected frequency of offspring.

There are two generic types of statistical techniques used in linkage analysis, *parametric* linkage analysis and *nonparametric* linkage analysis. Parametric linkage analysis uses statistical procedures to estimate θ and sometimes other quantities. An important term in parametric linkage analysis is the *lod score*. The term lod is specific to genetics and refers to the *common logarithm* (i.e., the base 10 logarithm*) of the odds for linkage*. The odds for linkage equal the ratio of two probabilities. The numerator is the probability of observing the data given that θ is some value less than .50 (i.e., the marker and the trait loci are linked) and the denominator is probability of observing the data given that θ = .50 (i.e., the marker and the trait loci are not linked). It is not necessary for us to learn how to compute lod scores, but it is helpful to go through an example.

Figure 7.9 depicts a pedigree for a dominant disorder. We illustrate the calculation

Figure 7.9. An example of a pedigree used for the calculation of a lod score.



of the lod score by assuming that the affected father has haplotype *AD/ad* and that θ equals .10. We begin the calculating by finding the probability of each offspring under the hypothesis that θ = .10. We can use Table 7.16 for this purpose. The probability of II.1 is .45, the probability of II.2 is .45, the probability of II.3 is .45, the probability of II.4 is .05, and the probabilities of II.5 and II.6 are both .45. We next multiply all these probabilities together to arrive at the probability of observing this family under the hypothesis of linkage with θ = .10. This quantity is $.05(.45)^5 = .0009226$.

We next calculate the probability of the family under the hypothesis that θ = .50. Again, start by finding the probability for each offspring under the hypothesis that θ = .50. From Table 7.16, this probability will be equal for all offspring and is .25. By multiplying .25 by itself six times, we have the probability of this family when θ = .50--$.25^6$ = .000241.

The lod score is then the base 10 log of the ratio of these two probabilties, or

$$\text{lod} = \log_{10}\left(\frac{.0009226}{.0002414}\right) = 0.577 .$$

In traditional linkage analysis of Mendelian traits, a lod score of 3 or more is required in order to be confident of linkage. The present lod score is much less than 3, so we would not take this pedigree as sufficient evidence to claim linkage between the *A* locus and the disorder. In reality, linkage studies involve studying many pedigrees and lod scores are added over pedigrees.

Parametric linkage and lod scores are suitable for single gene disorders. For complex disorders, like all of psychopathology, many geneticists prefer to use the nonparametric approach. The advantage of the nonparametric techniques is that it is not necessary to make assumptions about the mode of inheritance for the disorder, something that we quite frankly do not know about in the case of psychopathology. The disadvantage of nonparametric approaches is that they are less powerful than the parametric techniques.

It is not possible to survey all the nonparametric techniques here. Instead, we will illustrate one of them, the *affected sib-pair method*. Here, the geneticist gathers data on a large number of sibships to locate those that have at least two members of a sibship who

are affected with the disorder. These affected sib pairs are then genotyped at the marker locus[9], and the sib pairs are placed into one of two mutually exclusive categories based on their genotypes at the marker. The first category includes all sib pairs who have the same genotype at the marker; we can term these the *marker-concordant* pairs. The second category is for the *marker-discordant* pairs, i.e., those sib pairs who have different genotypes at the marker.

If the marker is not linked to the gene for the disorder, then we should expect an equal number in both categories[10]. However, if the marker is linked to the disease locus, then there should be more marker-concordant pairs than marker-discordant pairs. Hence, one simply performs a statistical test to see whether the number of marker-concordant pairs is significantly larger than the number of marker-discordant pairs.

[9] In practice, the parents of the sibs are also genotyped. I omit this complication to make the logic of the design easier to understand.

[10] Strictly speaking, when there is no linkage, the ratio of marker concordant to marker discordant pairs is a complicated function of the frequencies of the marker alleles. The example in the text assumes that there are a very large number of alleles so that the frequency of any single allele is always quite small.