From Mendel to Human Genome: Solving the Heredity Puzzle

Organisms produce offspring. The offspring usually look pretty much like their parents. Catfish offspring are catfish. Tomato-plant offspring are tomato plants. Monarch-butterfly offspring are monarch butterflies. Clearly some kind of information was passed from the parents to the next generation. Something inherited by the offspring from its parents carried the message for how to develop just like Mom and Dad.

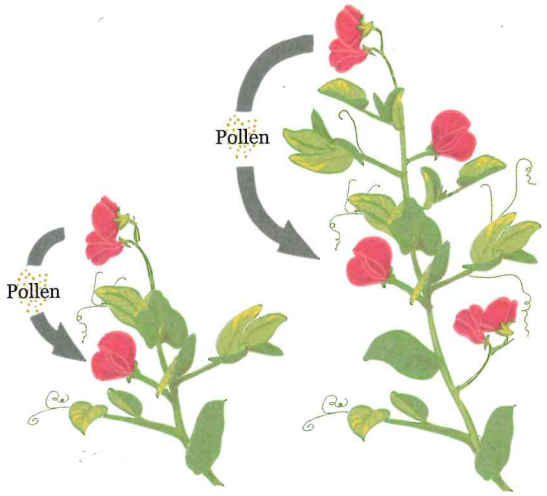
This fact was known for thousands of years, based on observation. Distinct facial features were passed from generation to generation. The mating of large, strong horses usually produced large, strong offspring. Goats with long, silky hair usually produced more of their kind. People used common sense to select and breed plants and animals with desirable characteristics. The offspring inherited something that produced the desired characteristics. But what was inherited? It was a mystery.

Early Research into Heredity

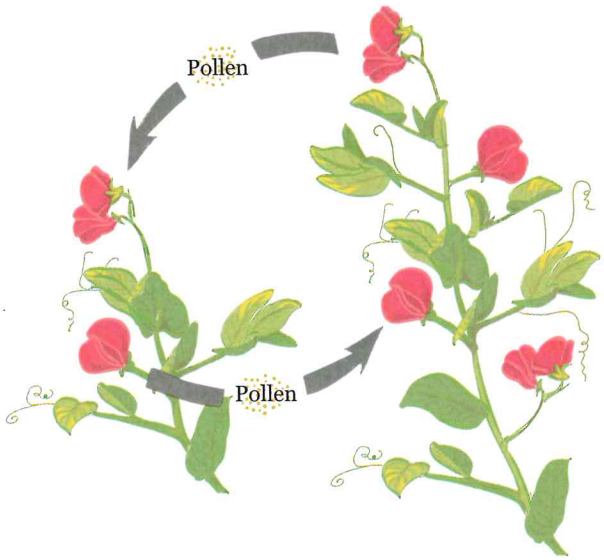
Gregor Mendel was born in 1822 in a poor farming community in what is now the Czech Republic in central Europe. He was a bright student, but his family didn't have enough money to send Gregor to the university. In order to continue his studies he joined a monastery. There he made a life of teaching and researching the question of heredity. By conducting careful experiments over many years, Mendel made landmark discoveries in heredity and established a new science: genetics.

Mendel was a keen observer of nature. He noticed that the common garden pea, *Pisum sativum*, had a significant amount of variation from plant to plant. Some plants produced green seeds, others yellow seeds; some plants produced red flowers, others white flowers; some plants were tall and others short; and so on. Mendel decided to see what he could find out about the distribution of the characteristics.

The experiments Mendel designed proceeded in three phases. First, he raised several generations of pea plants, making sure that the plants self-pollinated. That means that the flowers on every plant were fertilized only with pollen from the same plant. No flower was fertilized with pollen from another plant. In this way Mendel obtained pure breeding strains of plants for a trait, such as plant height. Tall plants produced only offspring that became tall plants when mature. Another strain produced only offspring that were short.



Next Mendel allowed his pure breeding strains to crossbreed. Pollen from tall plants was placed on the flowers of short plants, and pollen from short plants was placed on the flowers of tall plants.



The plants that he allowed to cross-pollinate Mendel called the parents. He identified them as the **P generation.** The first generation of offspring from the parents Mendel called the first filial generation (filial means sons and daughters). He identified them as the **F_1 generation.**

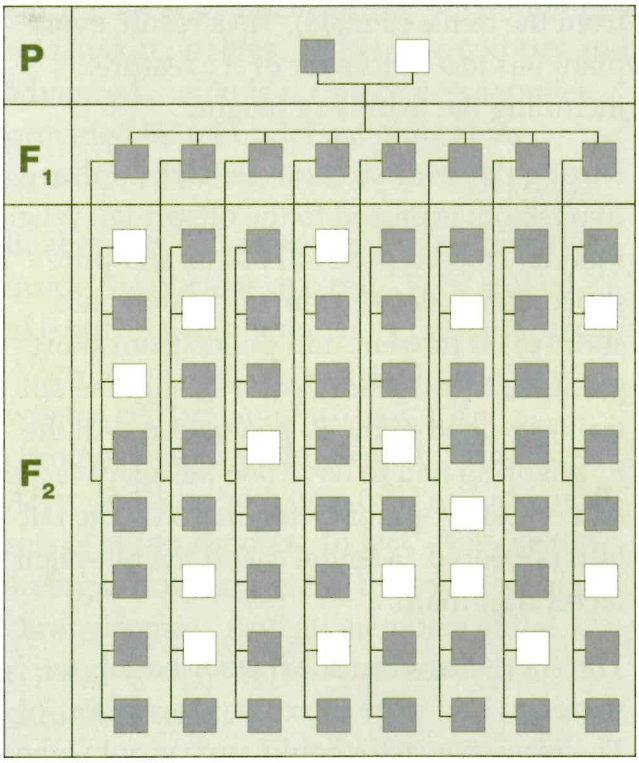
The third phase of Mendel's experiment involved letting the F_1 generation self-pollinate. The offspring of the self-pollination was the F_2 generation.

Mendel's Results

Mendel crossed his tall-plant and short-plant parents to produce an F_1 generation. When the F_1 generation matured, all the F_1 plants were tall. There were no short plants in the F_1 generation.

The tall F_1 plants were isolated from one another. Pollen from the flowers on each plant was used to pollinate the other flowers on the same plant. Some of the offspring, the F_2 generation, were tall plants and some were short plants!

Mendel had to make sense out of the results. The original parents were pure breeders for height (either tall or short). The F_1 generation produced by crossing a short plant and a tall plant were all tall. The tall F_1 generation produced both tall and short mature plants when allowed to self-pollinate. Below is a diagram of Mendel's results. (Each square represents a plant and indicates its height, with gray being tall.)



When Mendel counted the number of tall and short plants in the F_2 generation, he found one short plant for every three tall plants. The ratio of tall to short plants was 3:1. What could cause one out of four plants to be short, when short plants were completely absent from the F_1 generation?

Mendel's hypothesis revolutionized our understanding of heredity. This is what he reasoned.

Height is a **feature** of pea plants. The feature has two **traits,** tall and short. Offspring inherit something (Mendel didn't know what) that determines the trait from two sources, one from the male and one from the female. Mendel called the inherited something a factor.

The pea plants, which go through sexual reproduction, must get one height factor from the pollen (male) and one height factor from the ovule (female). As a result, every plant has *two* factors for every feature, including the feature of height.

During pollination, *one or the other* of the two factors is passed to the offspring. When Mendel crossed pure tall plants and pure short plants (pollen from tall plants onto short-plant flowers, and pollen from short plants onto tall-plant flowers), only the tall trait appeared in the F_1 generation. All the F_1 offspring had to have one tall-plant factor and one short-plant factor, but only the tall trait appeared. Mendel called the tall-plant factor **dominant.**

The short-plant factor, which Mendel reasoned had to be there, was **recessive.** The recessive factor could appear only when the offspring inherited the recessive factor from both parents. In that case the short-plant trait could appear because there would be no dominant factor.

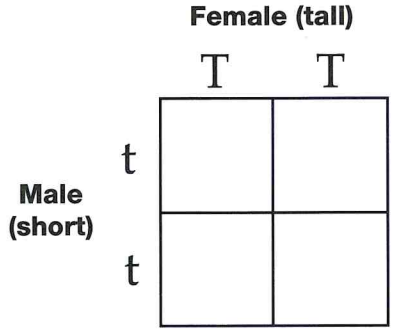
Punnett Squares

Mendel's hypothesis explained how traits could disappear in one generation and reappear in the next. His ideas explained observation in the real world. Furthermore, he could predict the number of offspring that would have a dominant trait and a recessive trait.

The method we use today to predict the traits of offspring is the Punnett square. Reginald Punnett, an early 20th-century heredity scientist, introduced it. His method uses a simple two-coordinate system to show the probability of traits in offspring. This is how it works.

The feature of plant height is represented by the letter *t* for tallness. An uppercase letter *(T)* refers to the dominant factor and a lowercase letter *(t)* refers to the recessive factor.

Mendel's true-breeding tall pea plants had to have two dominant factors (TT). The true-breeding short pea plants had to have two recessive factors (tt). A Punnett square set up to represent Mendel's cross of the two true-breeding pea plants looks like this.



Filling in the squares with the factors produces four possible offspring, and they all have the same pair of factors (Tt). Because all the offspring have a dominant factor, they all have the tall-plant trait.

Male (short) \ Female (tall)	T	T
t	Tt	Tt
t	Tt	Tt

Let's see what happened when Mendel self-pollinated the F_1 generation of tall plants. The male (pollen) and female (ovule) factors would have to both be Tt.

Male (tall) \ Female (tall)	T	t
T	TT	Tt
t	Tt	tt

Of the four possible offspring, three have the dominant factor, one TT and two Tt. The fourth has two recessive factors (tt), however, producing a short-plant offspring.

Mendel's experiments, completed in 1865, uncovered two principles that proved to be huge breakthroughs in the science of heredity.

1) Two factors determine traits. One factor comes from each parent. 2) Factors can be dominant or recessive. Recessive factors can be present in an organism even when no trait confirms that they are there.

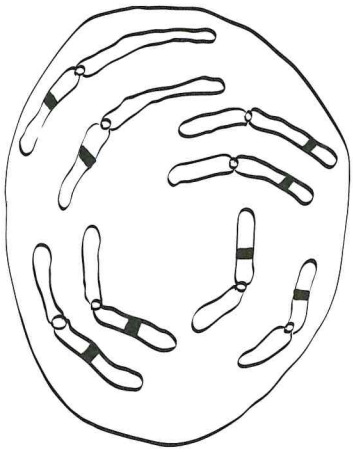
Understanding the Heredity Factor

Mendel's discoveries were so advanced that other scientists could not understand or accept their meaning. His work drifted out of the mainstream of science for 35 years. During those years, however, microbiology was making strides. Better microscopes and advanced methods for preparing samples provided better understanding of the structure of the nucleus of the cell.

In about 1875 chromosomes were observed inside the nucleus of a cell. They looked like little Xs and hot dogs of different sizes. A few years later scientists observed that, when a cell prepared to divide, the chromosomes first duplicated themselves. After division, both new cells had identical sets of chromosomes. In fact, every cell in an organism had exactly the same set of chromosomes. And all members of the same species have the same number of chromosomes.

In 1902 Walter S. Sutton observed that the chromosomes in a nucleus could be sorted into almost identical pairs. The grasshopper that Sutton was studying had 18 chromosomes. When he looked closely, he could see nine almost identical pairs. He named the two members of a pair **homologues** from a Greek word homologos.

The nucleus of a make-believe animal, the larkey, has eight chromosomes, organized in four homologous pairs. A larkey nucleus looks like this.

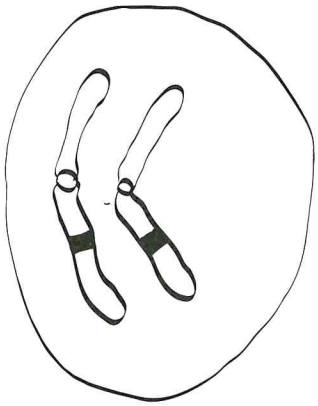


Further investigation revealed that, when cells divided to produce sex cells (sperm and eggs), the homologous chromosomes separated and went into separate cells. The sperm and egg cells had just one set of chromosomes, not two.

This discovery was remarkable. Sutton realized that, if the factors of heredity (now known as genes) were located on chromosomes, the two genes that determined a trait could be on homologous chromosomes. One form of the gene (perhaps a dominant gene) could be on one of the homologues and another form of the gene (maybe a recessive gene) could be on the other homologue. When a cell divided to form two sperm cells in a male (or egg cells in a female), the two homologous chromosomes separated. One chromosome in the pair went into one sperm cell, and the other chromosome went into the other sperm cell. Sperm cells and egg cells have just one set of chromosomes in their nuclei.

During sexual reproduction a sperm cell and an egg cell fuse. This is called fertilization. The set of chromosomes from the sperm and the set of chromosomes from the egg are united to form a *new* set of homologous chromosomes. The chromosomes are again paired, but the homologous pairs are different than those of either parent. The whole mechanism of heredity fell into place.

During the next decade it was confirmed that genes were indeed carried on chromosomes, and the two forms of the gene on homologous (paired) chromosomes could be identical (both dominant or both recessive), or they could be different (one of each).



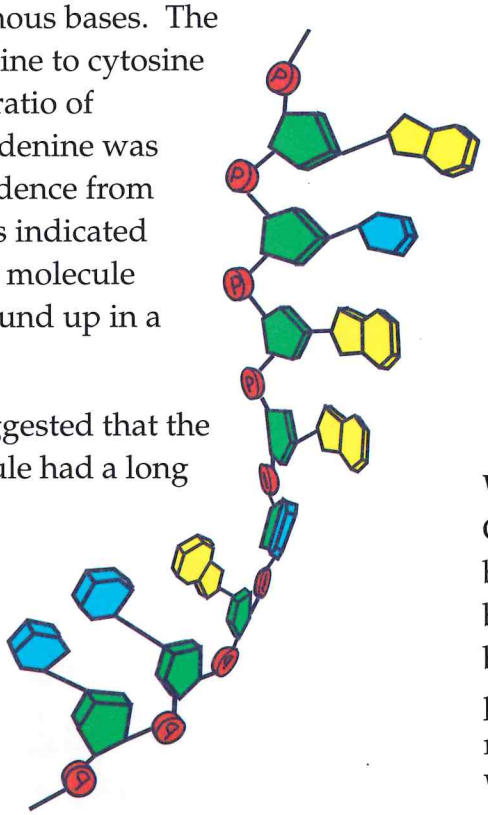
The vocabulary was refined to make communication more precise. The form of a gene on a single chromosome was an **allele,** and the two interacting alleles on homologous chromosomes constituted a **gene.**

What Is a Gene?

Understanding the gene came from understanding the structure and function of one remarkable molecule, deoxyribonucleic acid, better known simply as **DNA.** It was first extracted from the nucleus of cells in 1869. Because it is slightly acidic, it was named nucleic acid.

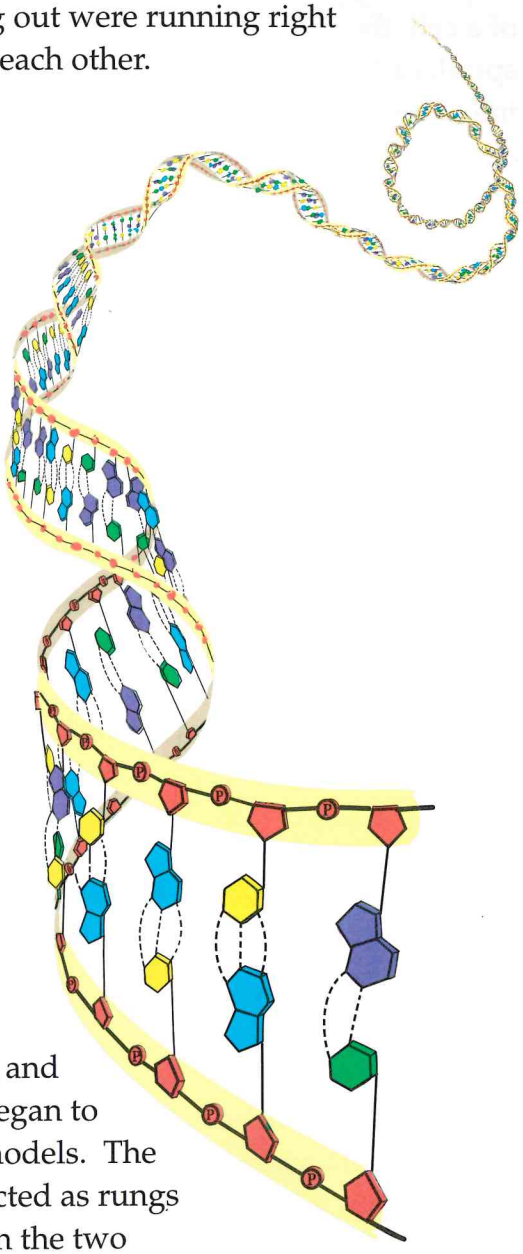
During the 1930s the great biochemist Phoebus A. Levene analyzed the DNA molecule and found that it was composed of several subunits—a phosphate unit, a sugar, and four nitrogenous bases. Levene got the analysis of the components of DNA exactly right. What he failed to understand was how they all fit together in a long repeating chain. This error in determining how the phosphates, sugars, and bases went together stalled further understanding of the role of DNA.

In the early 1950s two young scientists met at the Cavendish Laboratory in Cambridge, England. James D. Watson was a molecular biologist, and Francis H. Crick was a physicist. They proceeded to study what was known about the structure of the DNA molecule. The chemistry of the parts was known, as was the ratio of the four nitrogenous bases. The ratio of guanine to cytosine was 1:1; the ratio of thymine to adenine was also 1:1. Evidence from other sources indicated that the long molecule might be wound up in a spiral.



The data suggested that the DNA molecule had a long backbone of phosphates and sugars with the nitrogenous bases sticking out to the sides.

The single-stranded molecule didn't, however, form a coil. After exhaustive further study of their own data, and using crystallography structured data from another scientist, Rosalind Franklin, Watson and Crick deduced that the DNA molecule was not a single long strand wound into a spiral, but a double strand. Two backbones of phosphates and sugars with bases sticking out were running right next to each other.



Watson and Crick began to build models. The bases acted as rungs between the two phosphate-sugar rails. The model began to look like a twisted ladder. When they realized that the base cytosine could form a rung only with guanine, and adenine could form a rung only with thymine, they had it! The structure of the

magnificent molecule of heredity had been discovered, and it fit all the criteria for the central player in the story of heredity.

The DNA molecule is extremely long. The DNA in one of your cells might be 150 centimeters (59 inches) long if it were stretched out straight. To fit in the nucleus of a cell, the molecule is twisted into a spiral, called a helix. That helix is twisted into a second helix, and that helix is wound again, and so forth, until it forms the chromosome. That's what a chromosome is—a DNA molecule, twisted into a compact structure.

So where is the gene? A section of DNA—a sequence of bases—codes for a feature. A sequence may be short or long, and any sequence of bases along the length of the millions of bases might be a gene. Corresponding sequences on two homologous chromosomes—the two alleles—constitute a gene. Genes direct the manufacture of proteins, which go out into the body to make things happen.

Genotype and Phenotype

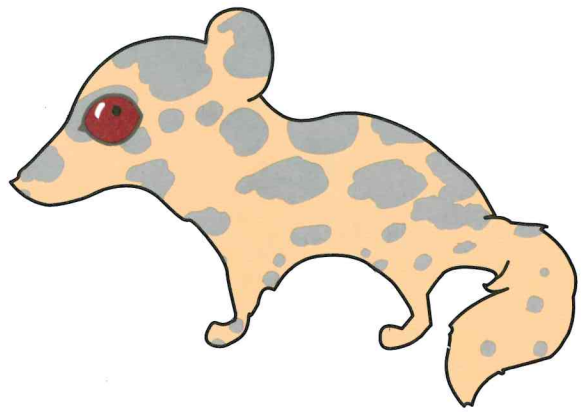
Genes determine features. The two alleles a larkey has for eye color interact to produce the trait of gray eyes or red eyes. If the larkey has two dominant alleles (EE) for eye color, the eyes will be red. If the larkey has one dominant and one recessive allele (Ee) for eye color, the eyes will still be red. Only if the larkey has two recessive alleles (ee) will the eyes be gray. The two alleles for a feature on a larkey's paired homologous chromosomes is its **genotype** for that feature.

If a gene is represented by two identical alleles, the larkey is **homozygous** for that feature. A homozygous genotype can be either homozygous dominant (EE) or homozygous recessive (ee). If a gene is represented by one dominant and one recessive allele, the larkey is **heterozygous** for eye color. The heterozygous genotype for larkey eye color is Ee.

Of course, larkeys have lots of other features. Larkeys have a pair of alleles that constitutes the gene for every feature. The larkey's complete set of paired alleles is its overall genotype. Here are three of the many possible genotypes for the four larkey genes being investigated in this course.

A a	A a	A A
E E	E E	E e
f f	f f	f f
T T	T t	T T

Genes code for features. Particular pairs of alleles—genes—will determine what traits an organism has. If a larkey has even one dominant gene for leg length in its genotype (that is, it is either homozygous dominant or heterozygous), the larkey will have short legs. How an organism looks as a result of its genotype is its **phenotype.** Because homozygous dominant (EE) and heterozygous (Ee) genotypes both produce

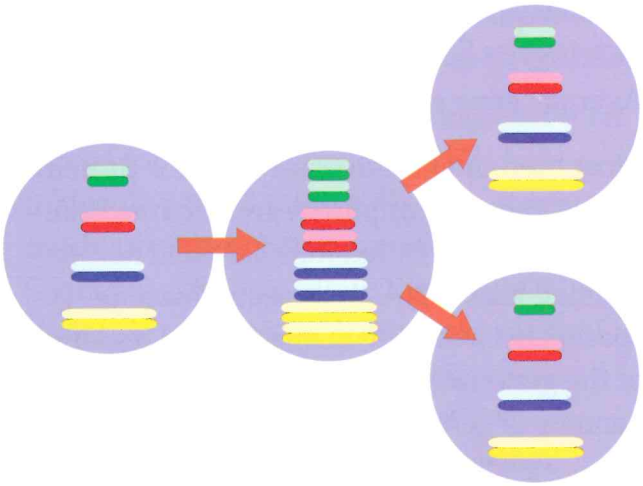


phenotypes with the dominant trait, genotype is not always apparent from the phenotype. The three larkey genotypes used as examples earlier actually produce identical phenotypes.

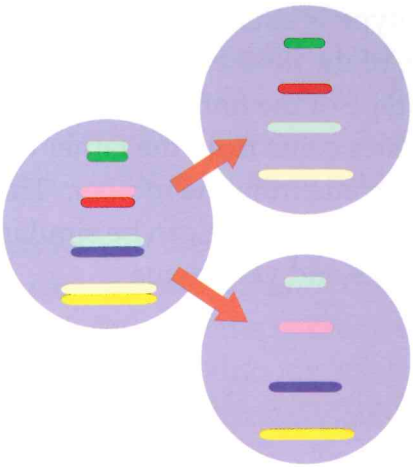
How Do Genes Get Mixed Up from Generation to Generation?

Organisms grow and maintain themselves by producing new cells. New cells result from cell division—a cell simply divides into two cells exactly like the old cell. Just like that, there are twice as many cells.

In preparation for cell division, a cell first produces a complete set of new chromosomes. When cell division happens, one set of chromosomes goes into each new cell. In the larkey case, the cell briefly has 16 chromosomes. Then it divides, and each cell gets one set of eight chromosomes (four homologous pairs). Both cells are exactly the same. This process is known as mitosis.

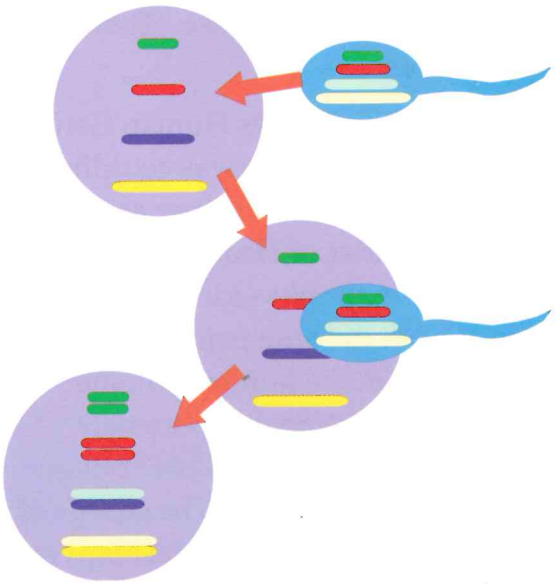


The story is different when cells divide to form sperm and egg cells as they do in organisms that reproduce sexually. After dividing once to form two identical cells, each of those divides again. But the chromosomes do not duplicate before



division. Instead, one member of each chromosome pair goes into one of the new cells, and the other chromosome goes into the other cell. These are the sperm cells or the egg cells. As a result each larkey sperm or egg cell has only four chromosomes instead of the usual eight. This process is known as meiosis.

During fertilization, the sperm cell nucleus fuses with the nucleus of the egg cell. At that time the four sperm chromosomes match up with the four egg chromosomes. The set of chromosomes is again complete—four pairs of homologous chromosomes. That creates the genotype for a new larkey.



The genotype is usually different from the genotype of the mother and different from the genotype of the father. Half the alleles for the genes come from the father and the other half come from the mother. This explains how offspring can be similar to their parents and yet unique.

Mapping the Human Genome

You have a genome. A genome is an organism's complete set of DNA, the incredibly long molecule that chromosomes are made of. If you could take apart all your chromosomes and look at the long string of base pairs along the double helix, you would be looking at your genome.

Every organism has a unique genome. At the same time, however, every organism has almost exactly the same genome as all other members of its species. More than 99.9% of your nucleotide bases (adenine, thymine, cytosine, and guanine) are exactly the same as your mother's, your brother's, your best friend's, your favorite rock star's, Nelson Mandela's, and everyone else's on Earth in Europe, Asia, Africa, the South Pacific islands, and everywhere else. There is a human genome.

In 1990 the United States Human Genome Project was launched. It was coordinated by the U.S. Department of Energy and the National Institutes of Health. The goal of the project was simple—identify the sequence of all the base pairs in all the chromosomes in the human genome. Accomplishing the goal is extremely difficult. There are about 3.1647 billion (3,164,700,000) base pairs. The strings of DNA in the individual chromosomes range from 50 million to more than 250 million bases. The job was huge.

The project was originally expected to take 15 years, but rapidly advancing laboratory technologies shortened the time by a couple of years. The first working draft of the sequence of bases in the human genome was released in June 2000. Genetics researchers had some interesting results.

Humans have a surprisingly small number of genes. Early in the project, scientists estimated that the human genome had between 80,000 and 120,000 genes. As the project draws close to completion, the number has been revised to 30,000 to 35,000 genes. Genes, sometimes called locations on chromosomes, are sequences of bases along the genome that code for the synthesis of a protein. If a particular sequence of bases were to move from location A to location B on the DNA molecule, the gene would be unchanged, and it would still code for a protein.

Gene sequences can be as small as a few hundred or as big as 2.4 million base pairs. Average gene size is about 3000 base pairs.

Most human DNA doesn't code for proteins; that is, it has no apparent genetic function. It is sometimes called junk DNA. The inactive junk DNA areas seem to act as spacers between the active genetic regions of the genome. Only about 2% of the genome is genetically active.

But what a marvelous job that 2% does! The human genome, which is now written down in its simple four-letter alphabet, G, C, A, and T, is available for all to read. The amazing story it tells is how to make a human being. The directions for building and maintaining the most complex system

known is written in living code and tucked away in the nucleus of every cell in your body.

Conclusion

In 1865 Mendel announced to the scientific community that organisms pass units of information to their offspring during reproduction. This inheritance allows the offspring to develop just like their parents. He didn't know what the units were, but he understood how they acted. We now know that Mendel discovered the existence of genes and described how they work.

For a number of reasons Mendel's discoveries were forgotten for 35 years. During those years the field of biochemistry was making advances. The DNA molecule was extracted from the chromosomes in the nucleus, and it was analyzed. The phosphate group, sugar molecules, and four nucleotide bases were identified. But how, or even if, this huge mass of chemical units contributed to heredity was still a mystery.

In 1953 Watson and Crick announced their model for the structure of the DNA molecule—the double helix. Overnight the mechanism of inheritance was clarified. Sequences of bases along the huge DNA molecule acted as the instruction manual for synthesizing proteins that make life possible. And the mechanisms for replicating the DNA molecule further clarified the story of inheritance. The mystery was solved in the mid-20th century.

The quest for understanding continues, however. At the close of the 20th century, not even 150 years after Mendel puzzled over his pea plants, the scientific community proposed a bold project to read the entire human genome—to identify every single base pair of the DNA of a human being. The 15-year project is drawing to a close, and the genome is available for all to read.

What stories will the genome tell? Is the cure for cancer written there? Can the information coded in the DNA be used to make proteins to repair body structures or manufacture specific drugs? Can the story written in the genome describe how we are related to all other life-forms? As usual, the closing of one chapter in science marks the opening of another. The completion of the sequencing of the human genome opens the doors to new science, with applications that we can only dream about today.