

= Genetics =

Genetics is the study of genes , genetic variation , and heredity in living organisms . It is generally considered a field of biology , but it intersects frequently with many of the life sciences and is strongly linked with the study of information systems .

The father of genetics is Gregor Mendel , a late 19th @-@ century scientist and Augustinian friar . Mendel studied ' trait inheritance ' , patterns in the way traits were handed down from parents to offspring . He observed that organisms ( pea plants ) inherit traits by way of discrete " units of inheritance " . This term , still used today , is a somewhat ambiguous definition of what is referred to as a gene .

Trait inheritance and molecular inheritance mechanisms of genes are still primary principles of genetics in the 21st century , but modern genetics has expanded beyond inheritance to studying the function and behavior of genes . Gene structure and function , variation , and distribution are studied within the context of the cell , the organism ( e.g. dominance ) and within the context of a population . Genetics has given rise to a number of sub @-@ fields including epigenetics and population genetics . Organisms studied within the broad field span the domain of life , including bacteria , plants , animals , and humans .

Genetic processes work in combination with an organism 's environment and experiences to influence development and behavior , often referred to as nature versus nurture . The intra- or extra @-@ cellular environment of a cell or organism may switch gene transcription on or off . A classic example is two seeds of genetically identical corn , one placed in a temperate climate and one in an arid climate . While the average height of the two corn stalks may be genetically determined to be equal , the one in the arid climate only grows to half the height of the one in the temperate climate due to lack of water and nutrients in its environment .

= = Etymology = =

The word genetics stems from the Ancient Greek ????????? genetikos meaning " genitive " / " generative " , which in turn derives from ???????? genesis meaning " origin " .

= = The gene = =

The modern working definition of a gene is a portion ( or sequence ) of DNA that codes for a known cellular function or process ( e.g. the function " make melanin molecules " ) . A single ' gene ' is most similar to a single ' word ' in the English language . The nucleotides ( molecules ) that make up genes can be seen as ' letters ' in the English language . Nucleotides are named according to which of the four nitrogenous bases they contain . The four bases are cytosine , guanine , adenine , and thymine . A single gene may have a small number of nucleotides or a large number of nucleotides , in the same way that a word may be small or large ( e.g. ' cell ' vs. ' electrophysiology ' ) . A single gene often interacts with neighboring genes to produce a cellular function and can even be ineffectual without those neighboring genes . This can be seen in the same way that a ' word ' may have meaning only in the context of a ' sentence ' . A series of nucleotides can be put together without forming a gene ( non coding regions of DNA ) , like a string of letters can be put together without forming a word ( e.g. udkslk ) . Nonetheless , all words have letters , like all genes must have nucleotides .

A quick heuristic that is often used ( but not always true ) is " one gene , one protein " meaning a singular gene codes for a singular protein type in a cell ( enzyme , transcription factor , etc . ) .

The sequence of nucleotides in a gene is read and translated by a cell to produce a chain of amino acids which in turn folds into a protein . The order of amino acids in a protein corresponds to the order of nucleotides in the gene . This relationship between nucleotide sequence and amino acid sequence is known as the genetic code . The amino acids in a protein determine how it folds into its unique three @-@ dimensional shape , a structure that is ultimately responsible for the protein 's function . Proteins carry out many of the functions needed for cells to live . A change to the DNA in a

gene can alter a protein 's amino acid sequence , thereby changing its shape and function and rendering the protein ineffective or even malignant ( e.g. sickle cell anemia ) . Changes to genes are called mutations .

= = History = =

The observation that living things inherit traits from their parents has been used since prehistoric times to improve crop plants and animals through selective breeding . The modern science of genetics , seeking to understand this process , began with the work of Gregor Mendel in the mid @-@ 19th century .

Prior to Mendel , Imre Festetics , a Hungarian noble , who lived in K?szeg before Gregor Mendel , was the first who used the word " genetics " . He described several rules of genetic inheritance in his work The genetic law of the Nature ( Die genetische Gesätze der Natur , 1819 ) . His second law is the same as what Mendel published . In his third law , he developed the basic principles of mutation ( he can be considered a forerunner of Hugo de Vries . )

Other theories of inheritance preceded his work . A popular theory during Mendel 's time was the concept of blending inheritance : the idea that individuals inherit a smooth blend of traits from their parents . Mendel 's work provided examples where traits were definitely not blended after hybridization , showing that traits are produced by combinations of distinct genes rather than a continuous blend . Blending of traits in the progeny is now explained by the action of multiple genes with quantitative effects . Another theory that had some support at that time was the inheritance of acquired characteristics : the belief that individuals inherit traits strengthened by their parents . This theory ( commonly associated with Jean @-@ Baptiste Lamarck ) is now known to be wrong ? the experiences of individuals do not affect the genes they pass to their children , although evidence in the field of epigenetics has revived some aspects of Lamarck 's theory . Other theories included the pangenesis of Charles Darwin ( which had both acquired and inherited aspects ) and Francis Galton 's reformulation of pangenesis as both particulate and inherited .

= = = Mendelian and classical genetics = = =

Modern genetics started with Gregor Johann Mendel , a scientist and Augustinian friar who studied the nature of inheritance in plants . In his paper " Versuche über Pflanzenhybriden " ( " Experiments on Plant Hybridization " ) , presented in 1865 to the Naturforschender Verein ( Society for Research in Nature ) in Brünn , Mendel traced the inheritance patterns of certain traits in pea plants and described them mathematically . Although this pattern of inheritance could only be observed for a few traits , Mendel 's work suggested that heredity was particulate , not acquired , and that the inheritance patterns of many traits could be explained through simple rules and ratios .

The importance of Mendel 's work did not gain wide understanding until the 1890s , after his death , when other scientists working on similar problems re @-@ discovered his research . William Bateson , a proponent of Mendel 's work , coined the word genetics in 1905 . ( The adjective genetic , derived from the Greek word genesis ? ??????? , " origin " , predates the noun and was first used in a biological sense in 1860 . ) Bateson both acted as a mentor and was aided significantly by the work of women scientists from Newnham College at Cambridge , specifically the work of Becky Saunders , Nora Darwin Barlow , and Muriel Wheldale Onslow . Bateson popularized the usage of the word genetics to describe the study of inheritance in his inaugural address to the Third International Conference on Plant Hybridization in London , England , in 1906 .

After the rediscovery of Mendel 's work , scientists tried to determine which molecules in the cell were responsible for inheritance . In 1911 , Thomas Hunt Morgan argued that genes are on chromosomes , based on observations of a sex @-@ linked white eye mutation in fruit flies . In 1913 , his student Alfred Sturtevant used the phenomenon of genetic linkage to show that genes are arranged linearly on the chromosome .

= = = Molecular genetics = = =

Although genes were known to exist on chromosomes , chromosomes are composed of both protein and DNA , and scientists did not know which of the two is responsible for inheritance . In 1928 , Frederick Griffith discovered the phenomenon of transformation ( see Griffith 's experiment ) : dead bacteria could transfer genetic material to " transform " other still @-@ living bacteria . Sixteen years later , in 1944 , the Avery ? MacLeod ? McCarty experiment identified DNA as the molecule responsible for transformation . The role of the nucleus as the repository of genetic information in eukaryotes had been established by Hämmerling in 1943 in his work on the single celled alga *Acetabularia* . The Hershey ? Chase experiment in 1952 confirmed that DNA ( rather than protein ) is the genetic material of the viruses that infect bacteria , providing further evidence that DNA is the molecule responsible for inheritance .

James Watson and Francis Crick determined the structure of DNA in 1953 , using the X @-@ ray crystallography work of Rosalind Franklin and Maurice Wilkins that indicated DNA had a helical structure ( i.e. , shaped like a corkscrew ) . Their double @-@ helix model had two strands of DNA with the nucleotides pointing inward , each matching a complementary nucleotide on the other strand to form what looks like rungs on a twisted ladder . This structure showed that genetic information exists in the sequence of nucleotides on each strand of DNA . The structure also suggested a simple method for replication : if the strands are separated , new partner strands can be reconstructed for each based on the sequence of the old strand . This property is what gives DNA its semi @-@ conservative nature where one strand of new DNA is from an original parent strand .

Although the structure of DNA showed how inheritance works , it was still not known how DNA influences the behavior of cells . In the following years , scientists tried to understand how DNA controls the process of protein production . It was discovered that the cell uses DNA as a template to create matching messenger RNA , molecules with nucleotides very similar to DNA . The nucleotide sequence of a messenger RNA is used to create an amino acid sequence in protein ; this translation between nucleotide sequences and amino acid sequences is known as the genetic code .

With the newfound molecular understanding of inheritance came an explosion of research . A notable theory arose from Tomoko Ohta in 1973 with her amendment to the neutral theory of molecular evolution through publishing the nearly neutral theory of molecular evolution . In this theory , Ohta stressed the importance of natural selection and the environment to the rate at which genetic evolution occurs . One important development was chain @-@ termination DNA sequencing in 1977 by Frederick Sanger . This technology allows scientists to read the nucleotide sequence of a DNA molecule . In 1983 , Kary Banks Mullis developed the polymerase chain reaction , providing a quick way to isolate and amplify a specific section of DNA from a mixture . The efforts of the Human Genome Project , Department of Energy , NIH , and parallel private efforts by Celera Genomics led to the sequencing of the human genome in 2003 .

= = Features of inheritance = =

= = = Discrete inheritance and Mendel 's laws = = =

At its most fundamental level , inheritance in organisms occurs by passing discrete heritable units , called genes , from parents to progeny . This property was first observed by Gregor Mendel , who studied the segregation of heritable traits in pea plants . In his experiments studying the trait for flower color , Mendel observed that the flowers of each pea plant were either purple or white ? but never an intermediate between the two colors . These different , discrete versions of the same gene are called alleles .

In the case of the pea , which is a diploid species , each individual plant has two copies of each gene , one copy inherited from each parent . Many species , including humans , have this pattern of inheritance . Diploid organisms with two copies of the same allele of a given gene are called

homozygous at that gene locus , while organisms with two different alleles of a given gene are called heterozygous .

The set of alleles for a given organism is called its genotype , while the observable traits of the organism are called its phenotype . When organisms are heterozygous at a gene , often one allele is called dominant as its qualities dominate the phenotype of the organism , while the other allele is called recessive as its qualities recede and are not observed . Some alleles do not have complete dominance and instead have incomplete dominance by expressing an intermediate phenotype , or codominance by expressing both alleles at once .

When a pair of organisms reproduce sexually , their offspring randomly inherit one of the two alleles from each parent . These observations of discrete inheritance and the segregation of alleles are collectively known as Mendel 's first law or the Law of Segregation .

== Notation and diagrams ==

Geneticists use diagrams and symbols to describe inheritance . A gene is represented by one or a few letters . Often a " + " symbol is used to mark the usual , non @-@ mutant allele for a gene .

In fertilization and breeding experiments ( and especially when discussing Mendel 's laws ) the parents are referred to as the " P " generation and the offspring as the " F1 " ( first filial ) generation . When the F1 offspring mate with each other , the offspring are called the " F2 " ( second filial ) generation . One of the common diagrams used to predict the result of cross @-@ breeding is the Punnett square .

When studying human genetic diseases , geneticists often use pedigree charts to represent the inheritance of traits . These charts map the inheritance of a trait in a family tree .

== Multiple gene interactions ==

Organisms have thousands of genes , and in sexually reproducing organisms these genes generally assort independently of each other . This means that the inheritance of an allele for yellow or green pea color is unrelated to the inheritance of alleles for white or purple flowers . This phenomenon , known as " Mendel 's second law " or the " Law of independent assortment " , means that the alleles of different genes get shuffled between parents to form offspring with many different combinations . ( Some genes do not assort independently , demonstrating genetic linkage , a topic discussed later in this article . )

Often different genes can interact in a way that influences the same trait . In the Blue @-@ eyed Mary ( *Omphalodes verna* ) , for example , there exists a gene with alleles that determine the color of flowers : blue or magenta . Another gene , however , controls whether the flowers have color at all or are white . When a plant has two copies of this white allele , its flowers are white ? regardless of whether the first gene has blue or magenta alleles . This interaction between genes is called epistasis , with the second gene epistatic to the first .

Many traits are not discrete features ( e.g. purple or white flowers ) but are instead continuous features ( e.g. human height and skin color ) . These complex traits are products of many genes . The influence of these genes is mediated , to varying degrees , by the environment an organism has experienced . The degree to which an organism 's genes contribute to a complex trait is called heritability . Measurement of the heritability of a trait is relative ? in a more variable environment , the environment has a bigger influence on the total variation of the trait . For example , human height is a trait with complex causes . It has a heritability of 89 % in the United States . In Nigeria , however , where people experience a more variable access to good nutrition and health care , height has a heritability of only 62 % .

== Molecular basis for inheritance ==

== DNA and chromosomes ==

The molecular basis for genes is deoxyribonucleic acid ( DNA ). DNA is composed of a chain of nucleotides , of which there are four types : adenine ( A ) , cytosine ( C ) , guanine ( G ) , and thymine ( T ) . Genetic information exists in the sequence of these nucleotides , and genes exist as stretches of sequence along the DNA chain . Viruses are the only exception to this rule ? sometimes viruses use the very similar molecule , RNA , instead of DNA as their genetic material . Viruses cannot reproduce without a host and are unaffected by many genetic processes , so tend not to be considered living organisms .

DNA normally exists as a double @-@ stranded molecule , coiled into the shape of a double helix . Each nucleotide in DNA preferentially pairs with its partner nucleotide on the opposite strand : A pairs with T , and C pairs with G. Thus , in its two @-@ stranded form , each strand effectively contains all necessary information , redundant with its partner strand . This structure of DNA is the physical basis for inheritance : DNA replication duplicates the genetic information by splitting the strands and using each strand as a template for synthesis of a new partner strand .

Genes are arranged linearly along long chains of DNA base @-@ pair sequences . In bacteria , each cell usually contains a single circular genophore , while eukaryotic organisms ( such as plants and animals ) have their DNA arranged in multiple linear chromosomes . These DNA strands are often extremely long ; the largest human chromosome , for example , is about 247 million base pairs in length . The DNA of a chromosome is associated with structural proteins that organize , compact and control access to the DNA , forming a material called chromatin ; in eukaryotes , chromatin is usually composed of nucleosomes , segments of DNA wound around cores of histone proteins . The full set of hereditary material in an organism ( usually the combined DNA sequences of all chromosomes ) is called the genome .

While haploid organisms have only one copy of each chromosome , most animals and many plants are diploid , containing two of each chromosome and thus two copies of every gene . The two alleles for a gene are located on identical loci of the two homologous chromosomes , each allele inherited from a different parent .

Many species have so @-@ called sex chromosomes that determine the gender of each organism . In humans and many other animals , the Y chromosome contains the gene that triggers the development of the specifically male characteristics . In evolution , this chromosome has lost most of its content and also most of its genes , while the X chromosome is similar to the other chromosomes and contains many genes . The X and Y chromosomes form a strongly heterogeneous pair .

### == = Reproduction == =

When cells divide , their full genome is copied and each daughter cell inherits one copy . This process , called mitosis , is the simplest form of reproduction and is the basis for asexual reproduction . Asexual reproduction can also occur in multicellular organisms , producing offspring that inherit their genome from a single parent . Offspring that are genetically identical to their parents are called clones .

Eukaryotic organisms often use sexual reproduction to generate offspring that contain a mixture of genetic material inherited from two different parents . The process of sexual reproduction alternates between forms that contain single copies of the genome ( haploid ) and double copies ( diploid ) . Haploid cells fuse and combine genetic material to create a diploid cell with paired chromosomes . Diploid organisms form haploids by dividing , without replicating their DNA , to create daughter cells that randomly inherit one of each pair of chromosomes . Most animals and many plants are diploid for most of their lifespan , with the haploid form reduced to single cell gametes such as sperm or eggs .

Although they do not use the haploid / diploid method of sexual reproduction , bacteria have many methods of acquiring new genetic information . Some bacteria can undergo conjugation , transferring a small circular piece of DNA to another bacterium . Bacteria can also take up raw DNA fragments found in the environment and integrate them into their genomes , a phenomenon known as transformation . These processes result in horizontal gene transfer , transmitting fragments of

genetic information between organisms that would be otherwise unrelated .

= = Recombination and genetic linkage = = =

The diploid nature of chromosomes allows for genes on different chromosomes to assort independently or be separated from their homologous pair during sexual reproduction wherein haploid gametes are formed . In this way new combinations of genes can occur in the offspring of a mating pair . Genes on the same chromosome would theoretically never recombine . However , they do via the cellular process of chromosomal crossover . During crossover , chromosomes exchange stretches of DNA , effectively shuffling the gene alleles between the chromosomes . This process of chromosomal crossover generally occurs during meiosis , a series of cell divisions that creates haploid cells .

The first cytological demonstration of crossing over was performed by Harriet Creighton and Barbara McClintock in 1931 . Their research and experiments on corn provided cytological evidence for the genetic theory that linked genes on paired chromosomes do in fact exchange places from one homolog to the other .

The probability of chromosomal crossover occurring between two given points on the chromosome is related to the distance between the points . For an arbitrarily long distance , the probability of crossover is high enough that the inheritance of the genes is effectively uncorrelated . For genes that are closer together , however , the lower probability of crossover means that the genes demonstrate genetic linkage ; alleles for the two genes tend to be inherited together . The amounts of linkage between a series of genes can be combined to form a linear linkage map that roughly describes the arrangement of the genes along the chromosome .

= = Gene expression = =

= = = Genetic code = = =

Genes generally express their functional effect through the production of proteins , which are complex molecules responsible for most functions in the cell . Proteins are made up of one or more polypeptide chains , each of which is composed of a sequence of amino acids , and the DNA sequence of a gene ( through an RNA intermediate ) is used to produce a specific amino acid sequence . This process begins with the production of an RNA molecule with a sequence matching the gene 's DNA sequence , a process called transcription .

This messenger RNA molecule is then used to produce a corresponding amino acid sequence through a process called translation . Each group of three nucleotides in the sequence , called a codon , corresponds either to one of the twenty possible amino acids in a protein or an instruction to end the amino acid sequence ; this correspondence is called the genetic code . The flow of information is unidirectional : information is transferred from nucleotide sequences into the amino acid sequence of proteins , but it never transfers from protein back into the sequence of DNA ? a phenomenon Francis Crick called the central dogma of molecular biology .

The specific sequence of amino acids results in a unique three @-@ dimensional structure for that protein , and the three @-@ dimensional structures of proteins are related to their functions . Some are simple structural molecules , like the fibers formed by the protein collagen . Proteins can bind to other proteins and simple molecules , sometimes acting as enzymes by facilitating chemical reactions within the bound molecules ( without changing the structure of the protein itself ) . Protein structure is dynamic ; the protein hemoglobin bends into slightly different forms as it facilitates the capture , transport , and release of oxygen molecules within mammalian blood .

A single nucleotide difference within DNA can cause a change in the amino acid sequence of a protein . Because protein structures are the result of their amino acid sequences , some changes can dramatically change the properties of a protein by destabilizing the structure or changing the surface of the protein in a way that changes its interaction with other proteins and molecules . For

example , sickle @-@ cell anemia is a human genetic disease that results from a single base difference within the coding region for the ? @-@ globin section of hemoglobin , causing a single amino acid change that changes hemoglobin 's physical properties . Sickle @-@ cell versions of hemoglobin stick to themselves , stacking to form fibers that distort the shape of red blood cells carrying the protein . These sickle @-@ shaped cells no longer flow smoothly through blood vessels , having a tendency to clog or degrade , causing the medical problems associated with this disease .

Some DNA sequences are transcribed into RNA but are not translated into protein products ? such RNA molecules are called non @-@ coding RNA . In some cases , these products fold into structures which are involved in critical cell functions ( e.g. ribosomal RNA and transfer RNA ) . RNA can also have regulatory effects through hybridization interactions with other RNA molecules ( e.g. microRNA ) .

= = = Nature and nurture = = =

Although genes contain all the information an organism uses to function , the environment plays an important role in determining the ultimate phenotypes an organism displays . This is the complementary relationship often referred to as " nature and nurture " . The phenotype of an organism depends on the interaction of genes and the environment . An interesting example is the coat coloration of the Siamese cat . In this case , the body temperature of the cat plays the role of the environment . The cat 's genes code for dark hair , thus the hair @-@ producing cells in the cat make cellular proteins resulting in dark hair . But these dark hair @-@ producing proteins are sensitive to temperature ( i.e. have a mutation causing temperature @-@ sensitivity ) and denature in higher @-@ temperature environments , failing to produce dark @-@ hair pigment in areas where the cat has a higher body temperature . In a low @-@ temperature environment , however , the protein 's structure is stable and produces dark @-@ hair pigment normally . The protein remains functional in areas of skin that are colder ? such as its legs , ears , tail and face ? so the cat has dark @-@ hair at its extremities .

Environment plays a major role in effects of the human genetic disease phenylketonuria . The mutation that causes phenylketonuria disrupts the ability of the body to break down the amino acid phenylalanine , causing a toxic build @-@ up of an intermediate molecule that , in turn , causes severe symptoms of progressive mental retardation and seizures . However , if someone with the phenylketonuria mutation follows a strict diet that avoids this amino acid , they remain normal and healthy .

A popular method in determining how genes and environment ( " nature and nurture " ) contribute to a phenotype is by studying identical and fraternal twins or siblings of multiple births . Because identical siblings come from the same zygote , they are genetically the same . Fraternal twins are as genetically different from one another as normal siblings . By analyzing statistics on how often a twin of a set has a certain disorder compared to other sets of twins , scientists can determine whether that disorder is caused by genetic or environmental factors ( i.e. whether it has ' nature ' or ' nurture ' causes ) . One famous example is the multiple birth study of the Genain quadruplets , who were identical quadruplets all diagnosed with schizophrenia .

= = = Gene regulation = = =

The genome of a given organism contains thousands of genes , but not all these genes need to be active at any given moment . A gene is expressed when it is being transcribed into mRNA and there exist many cellular methods of controlling the expression of genes such that proteins are produced only when needed by the cell . Transcription factors are regulatory proteins that bind to DNA , either promoting or inhibiting the transcription of a gene . Within the genome of Escherichia coli bacteria , for example , there exists a series of genes necessary for the synthesis of the amino acid tryptophan . However , when tryptophan is already available to the cell , these genes for tryptophan synthesis are no longer needed . The presence of tryptophan directly affects the activity of the genes ?

tryptophan molecules bind to the tryptophan repressor ( a transcription factor ) , changing the repressor 's structure such that the repressor binds to the genes . The tryptophan repressor blocks the transcription and expression of the genes , thereby creating negative feedback regulation of the tryptophan synthesis process .

Differences in gene expression are especially clear within multicellular organisms , where cells all contain the same genome but have very different structures and behaviors due to the expression of different sets of genes . All the cells in a multicellular organism derive from a single cell , differentiating into variant cell types in response to external and intercellular signals and gradually establishing different patterns of gene expression to create different behaviors . As no single gene is responsible for the development of structures within multicellular organisms , these patterns arise from the complex interactions between many cells .

Within eukaryotes , there exist structural features of chromatin that influence the transcription of genes , often in the form of modifications to DNA and chromatin that are stably inherited by daughter cells . These features are called " epigenetic " because they exist " on top " of the DNA sequence and retain inheritance from one cell generation to the next . Because of epigenetic features , different cell types grown within the same medium can retain very different properties . Although epigenetic features are generally dynamic over the course of development , some , like the phenomenon of paramutation , have multigenerational inheritance and exist as rare exceptions to the general rule of DNA as the basis for inheritance .

= = Genetic change = =

= = Mutations = =

During the process of DNA replication , errors occasionally occur in the polymerization of the second strand . These errors , called mutations , can affect the phenotype of an organism , especially if they occur within the protein coding sequence of a gene . Error rates are usually very low ? 1 error in every 10 ? 100 million bases ? due to the " proofreading " ability of DNA polymerases . Processes that increase the rate of changes in DNA are called mutagenic : mutagenic chemicals promote errors in DNA replication , often by interfering with the structure of base @-@ pairing , while UV radiation induces mutations by causing damage to the DNA structure . Chemical damage to DNA occurs naturally as well and cells use DNA repair mechanisms to repair mismatches and breaks . The repair does not , however , always restore the original sequence .

In organisms that use chromosomal crossover to exchange DNA and recombine genes , errors in alignment during meiosis can also cause mutations . Errors in crossover are especially likely when similar sequences cause partner chromosomes to adopt a mistaken alignment ; this makes some regions in genomes more prone to mutating in this way . These errors create large structural changes in DNA sequence ? duplications , inversions , deletions of entire regions ? or the accidental exchange of whole parts of sequences between different chromosomes ( chromosomal translocation ) .

= = Natural selection and evolution = =

Mutations alter an organism 's genotype and occasionally this causes different phenotypes to appear . Most mutations have little effect on an organism 's phenotype , health , or reproductive fitness . Mutations that do have an effect are usually deleterious , but occasionally some can be beneficial . Studies in the fly *Drosophila melanogaster* suggest that if a mutation changes a protein produced by a gene , about 70 percent of these mutations will be harmful with the remainder being either neutral or weakly beneficial .

Population genetics studies the distribution of genetic differences within populations and how these distributions change over time . Changes in the frequency of an allele in a population are mainly influenced by natural selection , where a given allele provides a selective or reproductive advantage



to the organism , as well as other factors such as mutation , genetic drift , genetic draft , artificial selection and migration .

Over many generations , the genomes of organisms can change significantly , resulting in evolution . In the process called adaptation , selection for beneficial mutations can cause a species to evolve into forms better able to survive in their environment . New species are formed through the process of speciation , often caused by geographical separations that prevent populations from exchanging genes with each other . The application of genetic principles to the study of population biology and evolution is known as the " modern synthesis " .

By comparing the homology between different species ' genomes , it is possible to calculate the evolutionary distance between them and when they may have diverged . Genetic comparisons are generally considered a more accurate method of characterizing the relatedness between species than the comparison of phenotypic characteristics . The evolutionary distances between species can be used to form evolutionary trees ; these trees represent the common descent and divergence of species over time , although they do not show the transfer of genetic material between unrelated species ( known as horizontal gene transfer and most common in bacteria ) .

= = = Model organisms = = =

Although geneticists originally studied inheritance in a wide range of organisms , researchers began to specialize in studying the genetics of a particular subset of organisms . The fact that significant research already existed for a given organism would encourage new researchers to choose it for further study , and so eventually a few model organisms became the basis for most genetics research . Common research topics in model organism genetics include the study of gene regulation and the involvement of genes in development and cancer .

Organisms were chosen , in part , for convenience ? short generation times and easy genetic manipulation made some organisms popular genetics research tools . Widely used model organisms include the gut bacterium *Escherichia coli* , the plant *Arabidopsis thaliana* , baker 's yeast ( *Saccharomyces cerevisiae* ) , the nematode *Caenorhabditis elegans* , the common fruit fly ( *Drosophila melanogaster* ) , and the common house mouse ( *Mus musculus* ) .

= = = Medicine = = =

Medical genetics seeks to understand how genetic variation relates to human health and disease . When searching for an unknown gene that may be involved in a disease , researchers commonly use genetic linkage and genetic pedigree charts to find the location on the genome associated with the disease . At the population level , researchers take advantage of Mendelian randomization to look for locations in the genome that are associated with diseases , a method especially useful for multigenic traits not clearly defined by a single gene . Once a candidate gene is found , further research is often done on the corresponding gene ? the orthologous gene ? in model organisms . In addition to studying genetic diseases , the increased availability of genotyping methods has led to the field of pharmacogenetics : the study of how genotype can affect drug responses .

Individuals differ in their inherited tendency to develop cancer , and cancer is a genetic disease . The process of cancer development in the body is a combination of events . Mutations occasionally occur within cells in the body as they divide . Although these mutations will not be inherited by any offspring , they can affect the behavior of cells , sometimes causing them to grow and divide more frequently . There are biological mechanisms that attempt to stop this process ; signals are given to inappropriately dividing cells that should trigger cell death , but sometimes additional mutations occur that cause cells to ignore these messages . An internal process of natural selection occurs within the body and eventually mutations accumulate within cells to promote their own growth , creating a cancerous tumor that grows and invades various tissues of the body .

Normally , a cell divides only in response to signals called growth factors and stops growing once in contact with surrounding cells and in response to growth @-@ inhibitory signals . It usually then divides a limited number of times and dies , staying within the epithelium where it is unable to

migrate to other organs . To become a cancer cell , a cell has to accumulate mutations in a number of genes ( 3 ? 7 ) that allow it to bypass this regulation : it no longer needs growth factors to divide , it continues growing when making contact to neighbor cells , and ignores inhibitory signals , it will keep growing indefinitely and is immortal , it will escape from the epithelium and ultimately may be able to escape from the primary tumor , cross the endothelium of a blood vessel , be transported by the bloodstream and will colonize a new organ , forming deadly metastasis . Although there are some genetic predispositions in a small fraction of cancers , the major fraction is due to a set of new genetic mutations that originally appear and accumulate in one or a small number of cells that will divide to form the tumor and are not transmitted to the progeny ( somatic mutations ) . The most frequent mutations are a loss of function of p53 protein , a tumor suppressor , or in the p53 pathway , and gain of function mutations in the ras proteins , or in other oncogenes .

= = = Research methods = = =

DNA can be manipulated in the laboratory . Restriction enzymes are commonly used enzymes that cut DNA at specific sequences , producing predictable fragments of DNA . DNA fragments can be visualized through use of gel electrophoresis , which separates fragments according to their length .

The use of ligation enzymes allows DNA fragments to be connected . By binding ( " ligating " ) fragments of DNA together from different sources , researchers can create recombinant DNA , the DNA often associated with genetically modified organisms . Recombinant DNA is commonly used in the context of plasmids : short circular DNA molecules with a few genes on them . In the process known as molecular cloning , researchers can amplify the DNA fragments by inserting plasmids into bacteria and then culturing them on plates of agar ( to isolate clones of bacteria cells ) . ( " Cloning " can also refer to the various means of creating cloned ( " clonal " ) organisms . )

DNA can also be amplified using a procedure called the polymerase chain reaction ( PCR ) . By using specific short sequences of DNA , PCR can isolate and exponentially amplify a targeted region of DNA . Because it can amplify from extremely small amounts of DNA , PCR is also often used to detect the presence of specific DNA sequences .

= = = DNA sequencing and genomics = = =

DNA sequencing , one of the most fundamental technologies developed to study genetics , allows researchers to determine the sequence of nucleotides in DNA fragments . The technique of chain @-@ termination sequencing , developed in 1977 by a team led by Frederick Sanger , is still routinely used to sequence DNA fragments . Using this technology , researchers have been able to study the molecular sequences associated with many human diseases .

As sequencing has become less expensive , researchers have sequenced the genomes of many organisms , using a process called genome assembly , which utilizes computational tools to stitch together sequences from many different fragments . These technologies were used to sequence the human genome in the Human Genome Project completed in 2003 . New high @-@ throughput sequencing technologies are dramatically lowering the cost of DNA sequencing , with many researchers hoping to bring the cost of resequencing a human genome down to a thousand dollars .

Next generation sequencing ( or high @-@ throughput sequencing ) came about due to the ever @-@ increasing demand for low @-@ cost sequencing . These sequencing technologies allow the production of potentially millions of sequences concurrently . The large amount of sequence data available has created the field of genomics , research that uses computational tools to search for and analyze patterns in the full genomes of organisms . Genomics can also be considered a subfield of bioinformatics , which uses computational approaches to analyze large sets of biological data . A common problem to these fields of research is how to manage and share data that deals with human subject and personally identifiable information . See also genomics data sharing .

= = Society and culture = =

On 19 March 2015 , a leading group of biologists urged a worldwide ban on clinical use of methods , particularly the use of CRISPR and zinc finger , to edit the human genome in a way that can be inherited . In April 2015 , Chinese researchers reported results of basic research to edit the DNA of non @-@ viable human embryos using CRISPR .