Subject: Bioinformatics

B.Sc 1st Year (Semester-I)

Course Title: Foundations of Bioinformatics-I

Unit-I

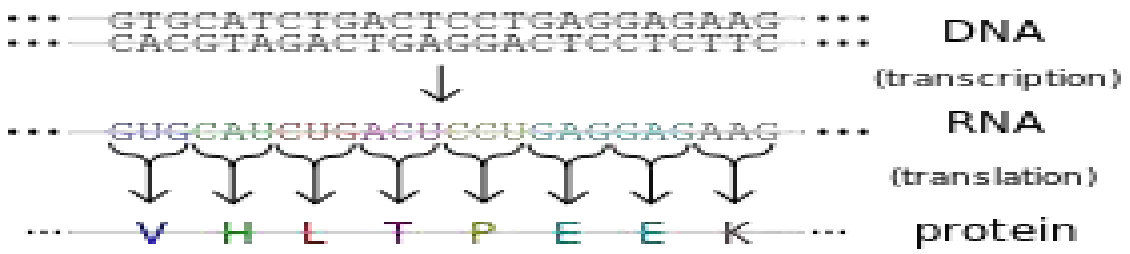
Genetics is the study of genes-what they are, what they do, and how they work. Genes are made up of molecules inside the nucleus of a cell that are strung together in such a way that the sequence carries information: that information determines how living organisms inherit phenotypic traits, (features) determined by the genes they received from their parents and thereby going back through the generations. For example, offspring produced by sexual reproduction usually look similar to each of their parents because they have inherited some of each of their parents' genes. Genetics identifies which features are inherited, and explains how these features pass from generation to generation. Genetics helps us to explain why offsprings resemble both their parents.

Genes are made of DNA, which is divided into separate pieces called chromosomes. Chromosomes link one generation to the next and are responsible for the similarities between the offspring and their parents. Chromosomes are made up of tiny bodies known as genes. Genes reside in the long molecule of DNA that exists within all cells.DNA in conjugation with a protein matrix, form the nucleoproteins and become organized into the chromosome structure. These bodies are the ultimate units of heredity.

Genes are pieces of DNA that contain information for synthesis of ribonucleic acids (RNAs) or polypeptides. Genes are inherited as units, with two parents dividing out copies of their genes to their offspring. This process can be compared with mixing two hands of cards, shuffling them, and then dealing them out again. Humans have two copies of each of their genes, and make copies that are found in eggs or sperm-but they only include one copy of each type of gene. An egg and sperm join to form a complete set of genes. The eventually resulting offspring has the same number of genes as their parents, but for any gene one of their two copies comes from their father, and one from their mother. The effects of this mixing depend on the types (the alleles) of the gene. If the father has two copies of an allele for red hair, and the mother has two copies for brown hair, all their children get the two alleles that give different instructions, one for red hair and one for brown. The hair color of these children depends on how these alleles work together. If one allele dominates the instructions from another, it is called the dominant allele, and the allele that is overridden is called the recessive allele.

The function of genes is to provide the information needed to make molecules called proteins in cells. Cells are the smallest independent parts of organisms: the human body contains about 100 trillion cells, while very small organisms like bacteria are just one single cell. A cell is like a miniature and very complex factory that can make all the parts needed to produce a copy of itself, which happens when cells divide. There is a simple division of labor in cells-genes give instructions and proteins carry out these instructions, tasks like building a new copy of a cell, or repairing damage. Each type of protein is a specialist that only does one job, so if a cell needs to do something new, it must make a new protein to do this job. Similarly, if a cell needs to do something faster or slower than before, it makes more or less of the protein responsible. Genes tell cells what to do by telling them which proteins to make and in what amounts.

Genes are expressed by being transcribed into RNA, and this RNA then translated into protein.



Mendelian Genetics

Mendelian inheritance is inheritance of biological features that follows the laws proposed by Gregor Johann Mendel in 1865 and 1866 and re-discovered in 1900. It was initially very controversial. When Mendel's theories were integrated with the Boveri–Sutton chromosome theoryof inheritance by Thomas Hunt Morgan in 1915, they became the core of classical genetics while Ronald Fisher combined them with the theory of natural selection in his 1930 book *The Genetical Theory of Natural Selection*, putting evolution onto a mathematical footing and forming the basis for Population genetics and the modern evolutionary synthesis.

Mendels Experiments:

Mendel choose garden pea as plant material for his experiments, because of the following advantages: a) well defined characters b) bisexual flowers c) self-fertilization d) easy hybridization. Besides these features, garden pea being self- fertilized, had pure lines due to natural self-fertilization for a number of years.

Mendel discovered that, when he crossed purebred white flower and purple flower pea plants (the parental or P generation), the result was not a blend. Rather than being a mix of the two, the offspring (known as the F_1 generation) was purple-flowered. When Mendel self-fertilized the F_1generation pea plants, he obtained a purple flower to white flower ratio in the F_2 generation of 3 to 1. Mendel also discovered that the F_1 individuals showed only one of the traits in the F_1generation which he expressed as dominance of one trait over the other .the trait which appeared in the F_1generation was called dominant and the other which did not appear in F_1generation was called recessive.

He then conceived the idea of heredity units, which he called "factors". Mendel found that there are alternative forms of factors-now called genes-that account for variations in inherited characteristics. For example, the gene for flower color in pea plants exists in two forms, one for purple and the other for white. The alternative"forms" are now called alleles. For each biological trait, an organism inherits two alleles, one from each parent. These alleles may be the same or different. An organism that has two identical alleles for a gene is said to be homozygous for that gene (and is called a homozygote). An organism that has two different alleles for a gene is said be heterozygous for that gene (and is called a heterozygote).

Mendel also hypothesized that allele pairs separate randomly, or segregate, from each other during the production of gametes: egg and sperm. Because allele pairs separate during gamete production, a sperm or egg carries only one allele for each inherited trait. When sperm and egg unite at fertilization, each contributes its allele, restoring the paired condition in the offspring. This is called the **Law of Segregation**. Mendel also found that each pair of alleles segregates independently of the other pairs of alleles during gamete formation. This is known as the **Law of Independent Assortment**.

Law of Segregation of genes (the "First Law")

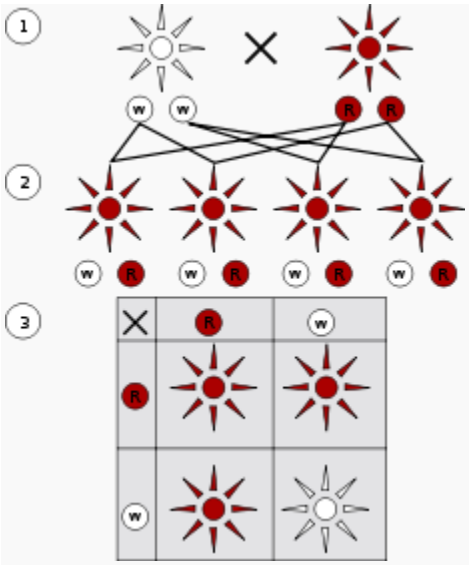


Figure 1 Dominant and recessive phenotypes.
(1) Parental generation.
(2) F_1 generation.
(3) F_2 generation. Dominant (red) and recessive (white) phenotype look alike in the F_1 (first) generation and show a 3:1 ratio in the F_2 (second) generation.

The Law of Segregation states that every individual organism contains two alleles for each trait, and that these alleles segregate (separate) during meiosis such that each gamete contains only one of the alleles. An offspring thus receives a pair of alleles for a trait by inheriting homologous chromosomes from the parent organisms: one allele for each trait from each parent.

Molecular proof of this principle was subsequently found through observation of meiosis by two scientists independently, the German botanist Oscar Hertwig in 1876, and the Belgian zoologist Edouard Van Beneden in 1883. Paternal and maternal chromosomes get separated in meiosis and the alleles with the traits of a character are segregated into two different gametes. Each parent contributes a single gamete, and thus a single, randomly successful allele copy to their offspring and fertilization.

Law of Independent Assortment (the "Second Law")

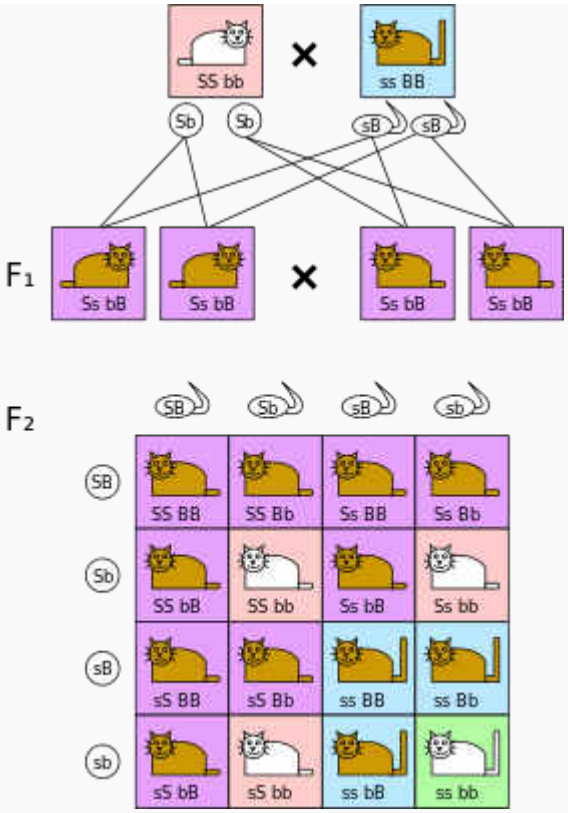


Figure 2 Dihybrid cross. The phenotypes of two independent traits show a 9:3:3:1 ratio in the F_2 generation. In this example, coat color is indicated by **B** (brown, dominant) or **b** (white), while tail length is indicated by **S** (short, dominant) or **s** (long). When parents are homozygous for each trait (**SSbb** and **ssBB**), their children in the F_1 generation are heterozygous at both loci and only show the dominant phenotypes (**SsbB**). If the children mate with each other, in the F_2 generation all combinations of coat color and tail length occur: 9 are brown/short (purple boxes), 3 are white/short (pink boxes), 3 are brown/long (blue boxes) and 1 is white/long (green box).

The Law of Independent Assortment states that alleles for separate traits are passed independently of one another from parents to offspring. That is, the biological selection of an allele for one trait has nothing to do with the selection of an allele for any other trait. Mendel found support for this law in his dihybrid cross experiments (Fig. 1). In his monohybrid crosses, an idealized 3:1 ratio between dominant and recessive phenotypes resulted. In dihybrid crosses, however, he found a 9:3:3:1 ratios (Fig. 2). This shows that each of the two alleles is inherited independently from the other, with a 3:1 phenotypic ratio for each.

Law of Dominance (the "Third Law") Mendel's Law of Dominance states that recessive alleles will always be masked by dominant alleles. Therefore, a cross between a homozygous dominant and a homozygous recessive will always express the dominant phenotype, while still having a heterozygous genotype. Law of Dominance can be explained easily with the help of a mono hybrid cross experiment:- In a cross between two organisms pure for any pair (or pairs) of contrasting traits (characters), the character that appears in the F1 generation is called "dominant" and the one which is suppressed (not expressed) is called "recessive." Each character is controlled by a pair of dissimilar factors. Only one of the characters expresses. The one which expresses in the F1 generation is called Dominant. It is important to note however, that the law of dominance is significant and true but is not universally applicable.

According to the latest revisions, only two of these rules are considered to be laws. The third one is considered as a basic principle but not a genetic law of Mendel.

Population genetics

The study of genetics at the population level is called population genetics. Population genetics is a subfield of genetics that deals with genetic differences within and between populations, and is a part of evolutionary biology. Studies in this branch of biology examine such phenomena as adaptation, speciation, and population structure.

Population genetics was a vital ingredient in the emergence of the modern evolutionary synthesis. Population genetic models are used both for statistical inference from DNA sequence data and for proof/disproof of concept.

Population genetics began as a reconciliation of Mendelian inheritance and biostatistics models. Natural selection will only cause evolution if there is enough genetic variation in a population. The Hardy-Weinberg principle provides the solution to how variation is maintained in a population with Mendelian inheritance. According to this principle, the frequencies of alleles (variations in a gene) will remain constant in the absence of selection, mutation, migration and genetic drift.

The main processes influencing allele frequencies are natural selection, genetic drift, gene flow and recurrent mutation

Natural selection, which includes sexual selection, is the fact that some traits make it more likely for an organism to survive and reproduce. Population genetics describes natural selection by defining fitness as a propensity or probability of survival and reproduction in a particular environment. The fitness is normally given by the symbol **w**=1-**s** where **s** is the selection coefficient. Natural selection acts on phenotypes, so population genetic models assume relatively simple relationships to predict the phenotype and hence fitness from the allele at one or a small number of loci. In this way, natural selection converts differences in the fitness of individuals with different phenotypes into changes in allele frequency in a population over successive generations.

.

Mutation

Mutation is the ultimate source of genetic variation in the form of new alleles. In addition, mutation may influence the direction of evolution when there is mutation bias, i.e. different probabilities for different mutations to occur. For example, recurrent mutation that tends to be in the opposite direction to selection can lead to mutation-selection balance.

Mutations can involve large sections of DNA becoming duplicated, usually through genetic recombination. This leads to copy-number variation within a population. Duplications are a major source of raw material for evolving new genes. Other types of mutation occasionally create new genes from previously noncoding DNA.

Genetic drift

Genetic drift is a change in allele frequencies caused by random sampling. That is, the alleles in the offspring are a random sample of those in the parents. Genetic drift may cause gene variants to disappear completely, and thereby reduce genetic variability. In contrast to natural selection, which makes gene variants more common or less common depending on their reproductive success, the changes due to genetic drift are not driven by environmental or adaptive pressures, and are equally likely to make an allele more common as less common.

Gene flow

Gene flow is the transfer of alleles from one population to another population through immigration of individuals. In this example, one of the birds from population A immigrates to population B, which has fewer of the dominant alleles, and through mating incorporates its alleles into the other population.

Applications

Population genetics models are used to infer which genes are undergoing selection. One common approach is to look for regions of high linkage disequilibrium and low genetic variance along the chromosome, to detect recent selective sweeps.

Another approach is the McDonald–Kreitman test. The McDonald–Kreitman test compares the amount of variation within a species (polymorphism) to the divergence between species (substitutions) at two types of sites, one assumed to be neutral.

Another approach to demographic inference relies on the allele frequency spectrum.

Population genetic models are created to describe the evolution of dominance and other forms of robustness, theevolution of sexual reproduction and recombination rates, the evolution of mutation rates, the evolution of evolutionary capacitors, the evolution of costly signalling traits, the evolution of ageing, and the evolution of co-operation. For example, most mutations are deleterious, so the optimal mutation rate for a species may be a trade-off between the damage from a high deleterious mutation rate and the metabolic costs of maintaining systems to reduce the mutation rate, such as DNA repair enzymes.

Gene concept

A **gene** is a locus (or region) of DNA which is made up of nucleotides and is the molecular unit of heredity. The transmission of genes to an organism's offspring is the basis of the inheritance of phenotypic traits. Most biological traits are under the influence of polygenes (many different genes) as well as the gene–environment interactions. Some genetic traits are instantly visible, such as eye colour or number of limbs, and some are not, such as blood type, risk for specific diseases, or the thousands of basic biochemical processes that comprise life.

The concept of a gene continues to be refined as new phenomena are discovered. For example, regulatory regions of a gene can be far removed from its coding regions, and coding regions can be split into several exons. Some viruses store theirgenome in RNA instead of DNA and some gene products are functional non-coding RNAs. Therefore, a broad, modern working definition of a gene is any discrete locus of heritable, genomic sequence which affect an organism's traits by being expressed as a functional product or by regulation of gene expression.

A gene can be described as a polynucleotide chain, which is a segment of DNA. It is a functional unit controlling a particular trait such as eye colour.

Beadle and Tatum concluded by various experiments that gene is a segment of DNA that codes for one enzyme. They proposed one gene-one enzyme hypothesis. But as some genes code for proteins that are not enzymes, the definition of gene was changed to one gene-one protein hypothesis.

The concept of gene has undergone further changes as the new facts came to light. Since proteins are polypeptide chains of amino acids translated by mRNA, gene was defined as one gene-one polypeptide relationship.

Even the one gene-one polypeptide definition is not complete as it does not include gene which codes for rRNA and tRNA. Only mRNA is translated into proteins. Therefore genes which code for polypeptides and RNAs are called structural genes.

In addition to structural genes, DNA also contains some sequences that have only regulatory function. These regulatory genes constitute signals, which “turn on” and “turn off” the transcription of structural genes and perform various other regulatory functions. In this way the definition of gene includes structural genes as well as regulatory genes.

Benzer coined terms for the gene, they are Cistron which is the unit of function, Recon which is the unit of recombination and Muton which is the unit of mutation.

Molecular Definition of a Gene

According to Lodish and others, gene is defined as the entire nucleic acid sequence that is necessary for the synthesis of a functional gene product, which may be polypeptide or any type of RNA. In addition to structural genes (coding genes) it also includes all the control sequences and non-coding introns.

Most prokaryotic genes transcribe polycistronic mRNA and most eukaryotic genes transcribe monocistronic mRNA.

Fine Structure of a Gene:

A gene is present only in one strand of DNA, which is a double stranded helix. A gene consists of several different regions. The main region is the coding sequence which carries information regarding amino acid sequence of polypeptides. The region on the left side of coding sequence (upstream or minus region) and on the right side (downstream or plus region) consists of fairly fixed regulatory sequences.

Alleles:

The word "allele" is a short form of **allelomorph** ("other form", a word coined by William Bateson), which was used in the early days of genetics to describe variant forms of a genedetected as different phenotypes. It derives from the Greek word meaning "reciprocal" or "each other".

Most multicellular organisms have two sets of chromosomes; that is, they are diploid. These chromosomes are referred to as homologous chromosomes. If both alleles at a gene (or locus) on the homologous chromosomes are the same, they and the organism are homozygous with respect to that gene (or locus). If the alleles are different, they and the organism areheterozygous with respect to that gene.

Dominant and Recessive alleles

In many cases, genotypic interactions between the two alleles at a locus can be described as dominant or recessive, according to which of the two homozygous phenotypes the heterozygotemost resembles. Where the heterozygote is indistinguishable from one of the homozygotes, the allele involved is said to be dominant to the other, which is said to be recessive to the former. The degree and pattern of dominance varies among loci. This type of interaction was first formally described by Gregor Mendel. However, many traits defy this simple categorization and the phenotypes are modeled by co-dominance and polygenic inheritance.

The term "wild type" allele is sometimes used to describe an allele that is thought to contribute to the typical phenotypic character as seen in "wild" populations of organisms, such as fruit flies (*Drosophila melanogaster*). Such a "wild type" allele was historically regarded as dominant (overpowering - always expressed), common, and normal, in contrast to "mutant" alleles regarded as recessive, rare, and frequently deleterious. It was formerly thought that most individuals were homozygous for the "wild type" allele at most gene loci, and that any alternative "mutant" allele was found in homozygous form in a small minority of "affected" individuals, often as genetic diseases, and more frequently in

heterozygous form in "carriers" for the mutant allele. It is now appreciated that most or all gene loci are highly polymorphic, with multiple alleles, whose frequencies vary from population to population, and that a great deal of genetic variation is hidden in the form of alleles that do not produce obvious phenotypic differences.

Multiple alleles

The genetic systems proposed thus far have been limited to a single pair of alleles. The maximum number of alleles at a gene locus that any individual possesses is two, with one on each of the homologus chromosomes. But since a gene can be changed to alternative forms by the process of mutation, a large number of alleles is theoretically possible in a population of individuals. Whenever more than two alleles are identified at a gene locus in a population, then it is called as multiple allelic series

Transposable element (TE or transposon)

A transposable element is a DNA sequence that can change its position within a genome, sometimes creating or reversing mutations and altering the cell's genome size. Transposition often results in duplication of the TE.

Barbara McClintock discovered the first TEs in maize (*Zea mays*) at the Cold Spring Harbor Laboratory in New York. McClintock was experimenting with maize plants that had broken chromosomes.

In the winter of 1944-1945, McClintock planted corn kernels that were self-pollinated, meaning that the silk (style) of the flower received pollen from its own anther. These kernels came from a long line of plants that had been self-pollinated, causing broken arms on the end of their ninth chromosomes. As the maize plants began to grow, McClintock noted unusual color patterns on the leaves. For example, one leaf had two albino patches of almost identical size, located side by side on the leaf. McClintock hypothesized that during cell division certain cells lost genetic material, while others gained what they had lost. However, when comparing the chromosomes of the current generation of plants with the parent generation, she found certain parts of the chromosome had switched position. This refuted the popular genetic theory of the time that genes were fixed in their position on a chromosome. McClintock found that genes could not only move, but they could also be turned on or off due to certain environmental conditions or during different stages of cell development.

McClintock also showed that gene mutations could be reversed. She presented her report on her findings in 1951, and published an article on her discoveries in *Genetics* in November 1953 entitled "Induction of Instability at Selected Loci in Maize."

Her work would be largely dismissed and ignored until the late 1960s-1970s when it would be rediscovered after TEs were found in bacteria. She was awarded a Nobel Prize in Physiology or Medicine in 1983 for her discovery of TEs, more than thirty years after her initial research.

Transposable elements make up a large fraction of the genome and are responsible for much of the C-value of eukaryotic cells. There are at least two classes of TEs: Class I TEs generally function via reverse transcription, while Class II TEs encode the protein transposase, which they require for insertion and excision, and some of these TEs also encode other proteins. It has been shown that TEs are important in genome function and evolution. Transposons are also very useful to researchers as a means to alter DNA inside a living organism.

Transposable elements represent one of several types of mobile genetic elements. TEs are assigned to one of two classes according to their mechanism of transposition, which can be described as either *copy and paste* (Class I TEs) or *cut and paste* (Class II TEs).

Class I (retrotransposons)

Class I TEs are copied in two stages: first, they are transcribed from DNA to RNA, and the RNA produced is then reverse transcribed to DNA. This copied DNA is then inserted back into the genome at a new position. The reverse transcription step is catalyzed by a reverse transcriptase, which is often encoded by the TE itself. The characteristics of retrotransposons are similar to retroviruses, such as HIV.

II (DNA transposons)

The cut-and-paste transposition mechanism of class II TEs does not involve an RNA intermediate. The transpositions are catalyzed by several transposase enzymes. Some transposases non-specifically bind to any target site in DNA, whereas others bind to specific target sequences. The transposase makes a staggered cut at the target site resulting in single-strand 5' or 3' DNA overhangs, so-called "sticky ends". This step cuts out the DNA transposon, which is then ligated into a new target site; the process involves activity of a DNA polymerase that fills in gaps and of a DNA ligase that closes the sugar-phosphate backbone. This results in duplication of the target site. The insertion sites of DNA transposons may be identified by short direct repeats (created by the staggered cut in the target DNA and filling in by DNA polymerase) followed by a series of inverted repeats important for the TE excision by transposase. Cut-and-paste TEs may be duplicated if their transposition takes place during S phase of the cell cycle, when a donor site has already been replicated but a target site has not yet been replicated. Such duplications at the target site can result in gene duplication, which plays an important role in genomic evolution. Not all DNA transposons transpose through the cut-and-paste mechanism. In some cases, a replicative transposition is observed in which a transposon replicates itself to a new target site.

Genetic linkage and Mapping

Genetic linkage is the tendency of alleles that are close together on a chromosome to be inherited together during the meiosis phase of sexual reproduction. Genes whose loci are nearer to each other are less likely to be separated onto different chromatids during chromosomal crossover, and are therefore said to be genetically *linked*. In other words, the nearer two genes are on a chromosome, the lower is the chance of a swap occurring between them, and the more likely they are to be inherited together.

Genetic linkage was first discovered by the British geneticists William Bateson, Edith Rebecca Saunders and Reginald Punnett shortly after Mendel's laws were rediscovered. The understanding of linkage was expanded by the work of Thomas Hunt Morgan.Alfred Sturtevant, a student of Morgan's, first developed genetic maps, also known as linkage maps. Sturtevant proposed that the greater the distance between linked genes, the greater the chance that non-sister chromatids would cross over in the region between the genes. By working out the number of recombinants it is possible to obtain a measure for the distance between the genes. This distance is expressed in terms of a **genetic map unit (m.u.)**, or a **centimorgan** and is defined as the distance between genes for which one product of meiosis in 100 is recombinant.

Linkage map

A linkage map is a genetic map of a species or experimental population that shows the position of its known genes or genetic markers relative to each other in terms of recombination frequency, rather than a specific physical distance along each chromosome. Linkage mapping is critical for identifying the location of genes that cause genetic diseases.

A genetic map is a map based on the frequencies of recombination between markers during crossover of homologous chromosomes. The greater the frequency of recombination (segregation) between two genetic markers, the further apart they are assumed to be. Conversely, the lower the frequency of recombination between the markers, the smaller the physical distance between them. Historically, the markers originally used were detectable phenotypes (enzyme production, eye

color) derived from coding DNA sequences; eventually, confirmed or assumed noncoding DNA sequences such as microsatellites or those generating restriction fragment length polymorphisms (RFLPs) have been used.

Genetic maps help researchers to locate other markers, such as other genes by testing for genetic linkage of the already known markers.

A genetic map is **not** a physical map (such as a radiation reduced hybrid map) or gene map.

Recombination frequency is a measure of genetic linkage and is used in the creation of a genetic linkage map. Recombination frequency (θ) is the frequency with which a single chromosomal crossover will take place between two genes during meiosis. A centimorgan (cM) is a unit that describes a recombination frequency of 1%. In this way we can measure the genetic distance between two loci, based upon their recombination frequency. This is a good estimate of the real distance. Double crossovers would turn into no recombination. In this case we cannot tell if crossovers took place. If the loci we're analysing are very close (less than 7 cM) a double crossover is very unlikely. When distances become higher, the likelihood of a double crossover increases. As the likelihood of a double crossover increases we systematically underestimate the genetic distance between two loci.

During meiosis, chromosomes assort randomly into gametes, such that the segregation of alleles of one gene is independent of alleles of another gene. This is stated in Mendel's Second Law and is known as **the law of independent assortment**. The law of independent assortment always holds true for genes that are located on different chromosomes, but for genes that are on the same chromosome, it does not always hold true.

As an example of linkage, consider the classic experiment by William Bateson and Reginald Punnett. They were interested in trait inheritance in the sweet pea and were studying two genes-the gene for flower colour (*P*, purple, and *p*, red) and the gene affecting the shape of pollen grains (*L*, long, and *l*, round). They crossed the pure lines *PPLL* and *ppll* and then self-crossed the resulting *PpLl* lines. According to Mendelian genetics, the expected phenotypes would occur in a 9:3:3:1 ratio of PL:Pl:pL:pl. To their surprise, they observed an increased frequency of PL and pl and a decreased frequency of Pl and pL (see table below).

Bateson and Punnett experiment

Phenotype and genotype	Observed	Expected from 9:3:3:1 ratio
Purple, long (*P_L_*)	284	216
Purple, round (*P_ll*)	21	72
Red, long (*ppL_*)	21	72
Red, round (*ppll*)	55	24

Their experiment revealed **linkage** between the *P* and *L* alleles and the *p* and *l* alleles. The frequency of *P* occurring together with *L* and with *p* occurring together with *l* is greater than that of the

recombinant *Pl* and *pL*. The recombination frequency is more difficult to compute in an F2 cross than a backcross but the lack of fit between observed and expected numbers of progeny in the above table indicate it is less than 50%.

The progeny in this case received two dominant alleles linked on one chromosome (referred to as **coupling** or **cis arrangement**). However, after crossover, some progeny could have received one parental chromosome with a dominant allele for one trait (e.g. Purple) linked to a recessive allele for a second trait (e.g. round) with the opposite being true for the other parental chromosome (e.g. red and Long). This is referred to as **repulsion** or a **trans arrangement**. The phenotype here would still be purple and long but a test cross of this individual with the recessive parent would produce progeny with much greater proportion of the two crossover phenotypes. While such a problem may not seem likely from this example, unfavorable repulsion linkages do appear when breeding for disease resistance in some crops.

The two possible arrangements, cis and trans, of alleles in a double heterozygote are referred to as gametic phases, and *phasing* is the process of determining which of the two is present in a given individual.

When two genes are located on the same chromosome, the chance of a crossover producing recombination between the genes is related to the distance between the two genes. Thus, the use of recombination frequencies has been used to develop **linkage maps** or **genetic maps**.

However, it is important to note that recombination frequency tends to underestimate the distance between two linked genes. This is because as the two genes are located farther apart, the chance of double or even number of crossovers between them also increases. Double or even number of crossovers between the two genes results in them being cosegregated to the same gamete, yielding a parental progeny instead of the expected recombinant progeny. As mentioned above, the Kosambi and Haldane transformations attempt to correct for multiple crossovers

Genetic disorders

A **genetic disorder** is a genetic problem caused by one or more abnormalities in the genome, especially a condition that is present from birth (congenital). Most genetic disorders are quite rare and affect one person in every several thousands or millions.

Genetic disorders may be hereditary, passed down from the parents' genes. In other genetic disorders, defects may be caused by newmutations or changes to the DNA. In such cases, the defect will only be passed down if it occurs in the germ line. The same disease, such as some forms of cancer, may be caused by an inherited genetic condition in some people, by new mutations in other people, and mainly by environmental causes in other people. Whether, when and to what extent a person with the genetic defect or abnormality will actually suffer from the disease is almost always affected by the environmental factors and events in the person's development.

Some types of recessive gene disorders confer an advantage in certain environments when only one copy of the gene is present.

A **single-gene disorder** is the result of a single mutated gene. Over 4000 human diseases are caused by single-gene defects.Single-gene disorders can be passed on to subsequent generations in several ways. Genomic imprinting and uniparentaldisomy, however, may affect inheritance patterns. The divisions between recessive and dominant types are not "hard and fast", although the divisions betweenautosomal and X-linked types are (since the latter types are distinguished purely based on the chromosomal location of the gene). For example, achondroplasia is typically considered a dominant disorder, but children with two genes for achondroplasia have a severe skeletal disorder of which achondroplasics could be viewed as carriers. Sickle-cell anemia is also considered a recessive condition, butheterozygous carriers have increased resistance to malaria in early childhood, which could be described as a related dominant conditionWhen a couple where one partner or both are sufferers or

carriers of a single-gene disorder wish to have a child, they can do so through *in vitro* fertilization, which means they can then have a preimplantation genetic diagnosis to check whether the embryo has the genetic disorder

Autosomal dominant

Only one mutated copy of the gene will be necessary for a person to be affected by an autosomal dominant disorder. Each affected person usually has one affected parent The chance a child will inherit the mutated gene is 50%. Autosomal dominant conditions sometimes have reduced penetrance, which means although only one mutated copy is needed, not all individuals who inherit that mutation go on to develop the disease. Examples of this type of disorder are Huntington's disease neurofibromatosis type 1, neurofibromatosis type 2, Marfan syndrome, hereditary nonpolyposis colorectal cancer, hereditary multiple exostoses (a highly penetrant autosomal dominant disorder),Tuberous sclerosis, Von Willebrand disease, and acute intermittent porphyria. Birth defects are also called congenital anomalies.

Autosomal recessive

Two copies of the gene must be mutated for a person to be affected by an autosomal recessive disorder. An affected person usually has unaffected parents who each carry a single copy of the mutated gene (and are referred to as carriers). Two unaffected people who each carry one copy of the mutated gene have a 25% risk with each pregnancy of having a child affected by the disorder. Examples of this type of disorder are AcrodermatitisEnteropathica Albinism, Medium-chain acyl-CoA dehydrogenase deficiency, cystic fibrosis, sickle-cell disease, Tay-Sachs disease, Niemann-Pick disease, spinal muscular atrophy, and Roberts syndrome. Certain other phenotypes, such as wet versus dry earwax, are also determined in an autosomal recessive fashion

X-linked dominant

X-linked dominant disorders are caused by mutations in genes on the X chromosome. Only a few disorders have this inheritance pattern, with a prime example being X-linked hypophosphatemic rickets. Males and females are both affected in these disorders, with males typically being more severely affected than females. Some X-linked dominant conditions, such as Rett syndrome, incontinentiapigmenti type 2, and Aicardi syndrome, are usually fatal in males either *in utero* or shortly after birth, and are therefore predominantly seen in females. Exceptions to this finding are extremely rare cases in which boys with Klinefelter syndrome (47,XXY) also inherit an X-linked dominant condition and exhibit symptoms more similar to those of a female in terms of disease severity. The chance of passing on an X-linked dominant disorder differs between men and women. The sons of a man with an X-linked dominant disorder will all be unaffected (since they receive their father's Y chromosome), and his daughters will all inherit the condition. A woman with an X-linked dominant disorder has a 50% chance of having an affected fetus with each pregnancy, although it should be noted that in cases such as incontinentiapigmenti, only female offspring are generally viable. In addition, although these conditions do not alter fertility *per se*, individuals with Rett syndrome or Aicardi syndrome rarely reproduce

X-linked recessive

X-linked recessive conditions are also caused by mutations in genes on the X chromosome. Males are more frequently affected than females, and the chance of passing on the disorder differs between men and women. The sons of a man with an X-linked recessive disorder will not be affected, and his daughters will carry one copy of the mutated gene. A woman who is a carrier of an X-linked recessive disorder (XX) has a 50% chance of having sons who are affected and a 50% chance of having daughters who carry one copy of the mutated gene and are therefore carriers. X-linked recessive conditions include the serious diseases hemophilia A, Duchenne muscular dystrophy, and Lesch-Nyhan syndrome, as well

as common and less serious conditions such as male pattern baldness and red-green color blindness. X-linked recessive conditions can sometimes manifest in females due to skewed X-inactivation or monosomy X (Turner syndrome).

Y-linked

Y-linked disorders, also called holandric disorders, are caused by mutations on the Y chromosome. These conditions display may only be transmitted from the heterogametic sex (e.g. male humans) to offspring of the same sex. More simply, this means that Y-linked disorders in humans can only be passed from men to their sons; females can never be affected because they do not possess Y-allosomes.

Y-linked disorders are exceedingly rare but the most well-known examples typically cause infertility. Reproduction in such conditions is only possible through the circumvention of infertility by medical intervention.

Many genes

Genetic disorders may also be complex, multifactorial, or polygenic, meaning they are likely associated with the effects of multiple genes in combination with lifestyles and environmental factors. Multifactorial disorders include heart disease and diabetes. Although complex disorders often cluster in families, they do not have a clear-cut pattern of inheritance. This makes it difficult to determine a person's risk of inheriting or passing on these disorders. Complex disorders are also difficult to study and treat, because the specific factors that cause most of these disorders have not yet been identified. Studies which aim to identify the cause of complex disorders can use several methodological approaches to determine genotype-phenotype associations. One method, the genotype-first approach, starts by identifying genetic variants within patients and then determining the associated clinical manifestations. This is opposed to the more traditional phenotype-first approach, and may identify causal factors that have previously been obscured by clinical heterogeneity, penetrance, and expressivity.