

Genes and Mutations

As explained earlier, genes are segments of DNA and DNA is the hereditary material of cells. Although this is clearly understood now, it was not obvious when biologists first used the terms **gene** and **genetics**; for this reason, the definitions of many of the terms associated with this section have undergone considerable change over time. Because not all references change at the same rate, significant variation exists with respect to terminology and definitions used by various sources. Providing current definitions for some commonly used terms (as appropriate for this course) is therefore useful.

Definitions for some important terms:

1. **Gene** – A gene is a unit of heredity. The term gene is usually applied to segments of DNA; however, since the genetic material within some viruses is RNA, viral genes may be segments of RNA. Most genes encode proteins, ie, code for m-RNA that is translated into amino acid sequences. For our purposes, these will be designated as **structural genes**. In eukaryotic cells, structural genes carry regions not **expressed** as amino acid sequence; these regions include promoter sequences and the introns removed during post-transcriptional modification. Prokaryotic genes are often associated with operons. When they are, a single promoter is involved, and multiple structural genes are transcribed as one, long m-RNA molecule. The process and the genes are **polycistronic**.

Not all genes are structural genes. Segments of DNA encoding t-RNA, r-RNA, s-RNA, etc. are transcribed to produce RNA molecules, but these are not translated. Some authors consider promoter sites, operator sites, attenuator sites, etc. to be regions within genes, but according to the Sequence Ontology Project definition, these regions (with their own distinct function) are separate genes.

Alternate forms of a gene are called **alleles**. These are nucleotide sequences coding for polypeptides with the same function in general, but with slightly different amino acid sequences. For example, normal hemoglobin and sickle-cell hemoglobin, or the various different flagellin proteins made by bacteria.

2. **Genetics** – Genetics is the science or study of genes, heredity and factors that influence the variations observed within organisms. Heredity has been studied at multiple levels since humans first began to domesticate animals and breed plants for food production. Classical genetics (Mendelian inheritance) is based primarily on the work of Gregor Mendel and his studies involving pea plants (1860s). Modern geneticists focus more on the structure and function of genes at the molecular level, i.e., DNA and RNA and their interactions with various proteins. Needless to say, there is considerable variation in the field of genetics.
3. **Genome** – The term genome can be defined as the total DNA content of the chromosome (or multiple chromosomes within eukaryotic organisms), or as the total DNA content of the organism. Since the term genome was generated from the terms

gene and **chromosome**, the first definition is more accurate. The genome includes all portions of the chromosome including non-coding regions such as promoters, operators and introns. It does not include plasmids, mitochondrial DNA, chloroplast DNA or viral DNA.

4. **Genotype** – The genotype is the genetically determined characteristics of an organism or the genetic potential of that organism. As we have already seen, genes may be present within an organism, but not be **expressed**, i.e., transcribed and translated to produce proteins, because transcription is being repressed. Some bacteria carry genes that are not expressed for long periods of time (e.g., *E. coli* cells living inside adult cows), and some carry genes that are never expressed due to slight changes in their DNA, i.e., due to mutation.
5. **Phenotype** – The phenotype is the observed characteristics of an organism, or the genotype expressed. The **environment** (both inside and outside cells) has a significant influence on phenotype, as we have seen in the laboratory. Recall the variation in morphology of *E. coli* colonies grown on MAC, EMB, T-7 and nutrient agar. Bacteria often express certain genes only under certain circumstances, and gene expression may or may not result in an observable characteristic even when it is occurring.
6. **Locus** – The term locus (plural = loci) usually refers to the location of a gene on a chromosome, but can also be applied to other structures. The locus is a fixed location on a chromosome, and is usually occupied by a specific gene or codon, etc.

The genome of an organism and therefore the genotype is not necessarily static over time. Within populations, organisms can appear with novel characteristics, encoded by genes not previously present. There are two basic mechanisms involved in bringing about such changes, these are **mutations** and **genetic exchange mechanisms** (sexual reproduction). Under some circumstances it is difficult to distinguish between the two.

Mutations:

A **mutation** is any change in the nucleotide sequence of DNA within a cell, or RNA within certain types of viruses. Though some references define mutations as heritable changes in DNA, most mutations are not heritable because they cause cells to die. The majority of mutations are lethal to the cells they occur in. In multicellular organisms such as ourselves, these changes frequently go unnoticed, but in the case of single-celled organisms, they end existence.

Mutations can be caused by a variety of factors including various types of chemicals and radiation. When the specific cause of a mutation cannot be identified, the mutation is said to occur spontaneously. Though exact numbers undoubtedly change over time, the **spontaneous mutation** rate within living organisms is around 1 per 100 million copies of DNA. This means polymerase enzymes can copy a single DNA molecule about 100 million times, incorporating only one error. Though this seems initially like a very low rate for mutations, remember that bacteria growing under optimal conditions often reach an m-concentration of 10^9 cells/ml of culture medium in less than 24 hours. This means a typical population of bacteria will experience one mutation every day.

Mutation rate can be significantly increased by specific factors (chemicals and physical factors) called **mutagenic agents** or **mutagens**. Since mutations can sometimes lead to tumor development, mutagens are also often **carcinogens**, i.e., cancer-causing agents.

Types of Mutations:

In multicellular eukaryotic organisms, mutations are often divided into two categories based on the types of cells or tissues involved. **Germ line** mutations occur within cells giving rise to **gametes** (sex cells) and can be passed to future generations, while **somatic mutations** occur within cells or tissues making up the body of the organism and cannot be passed on. When dealing with single-celled organisms, this distinction becomes meaningless, because any mutation that does not kill the cell is likely to be passed to the next generation (although there are exceptions). For this section, we will concentrate on the types of mutations occurring within single-celled organisms, specifically bacteria.

1. **Point mutations** – Point mutations are those involving single-base changes, i.e., changes in just **one nucleotide** within a gene or chromosome. These single-base changes can involve:
 - a. **Additions** – The addition of one extra base within the nucleotide sequence.
 - b. **Deletions** – The removal of one base from the nucleotide sequence.
 - c. **Substitutions** – The replacement of one base in the sequence with a different one.

Of these mutation types, **substitutions** are least likely to be lethal. Substitution type point mutations usually involve the replacement of a purine with a different purine or a pyrimidine with a different pyrimidine, but can sometimes involve purine for pyrimidine substitutions or vice versa. Substitution mutations may or may not result in changes to polypeptides; this is because the genetic code contains considerable redundancy. Consider the following nucleotide sequences.

DNA nucleotide sequence = TAC CCG GTT AAA CTG CGG TTT ACT

m-RNA nucleotide sequence = AUG GGC CAA UUU GAC GCC AAA UGA

A polypeptide encoded by this m-RNA sequence would contain methionine, glycine, glutamine, phenylalanine, aspartic acid, alanine, and lysine. The last codon "UGA" is "Umber", a terminator codon, so would not code for any amino acid. If we allow a substitution type point mutation to change the third base in the second codon (any letter will do), the mutation will have no effect on the polypeptide being formed. This is because all codons containing GG as the first two bases code for glycine. A mutation like this, having no effect on the amino acid sequence being formed is called a **silent mutation**. Changing the last base in fifth codon would also result in a silent mutation.

Substitution type point mutations are not always silent. Sometimes a single base change will cause the amino acid encoded to change; for example, if we substitute a "C" for the "A" in the third position of the third codon, the amino acid at that position will be changed from glutamine to histadine. Changing a single amino acid may or may not be significant, but the difference between normal hemoglobin and sickle-cell hemoglobin involves a single amino acid change.

If a substitution type point mutation occurred within the initiator codon (AUG), translation could not occur, because the fmet-t-RNA could not bind to m-RNA at the ribosome. In this case, if the protein being encoded was essential to cell survival, the cell would die. So although substitution type point mutations can be silent, they can also be lethal to the cell.

Addition and deletion type point mutations always have a significant impact. If one base is added, or one base is removed, the result will be a shift in the codon "reading frame", either to the left or right. This is called a **frame shift** mutation. In either case, the amino acid sequence is very likely to be changed. A substitution type point mutation can change a single amino acid, but an addition or a deletion will likely change **all** the amino acids being encoded beyond the mutation point. Under such circumstances, it is highly unlikely that a functional protein will be formed, but it is **not impossible**. Novel proteins with unique functions do sometimes arise within populations, and this is one mechanism allowing such changes.

If a mutation allows an individual within a population to gain an advantage over others within the population, it can be maintained and passed to future generations. For example, consider a population of bacteria living within a host organism. A point mutation occurring within one of the cells allows it to produce a novel protein, an enzyme called β -lactamase. This will have little effect under most circumstances, but if β -lactam antibiotics are present within the environment (as they would be within a host taking oral penicillin), the organism producing β -lactamase suddenly has a significant advantage over other cells because it is resistant to the drug. Cells susceptible to the drug will die, and the resistant organism will reproduce freely, having lost competitors using common resources. The result will be a new population of organisms, all resistant to the β -lactam drug.

Organisms do not develop resistance to drugs because they are exposed to the drugs. They develop resistance because they mutate, and the presence of the drug in the environment exerts selective pressure on the population. The process is called **natural selection**, and it is a major factor influencing **evolution**. Unfortunately the current, widespread use of antimicrobial drugs is exerting selective pressure on existing bacterial populations, and many important pathogens are becoming resistant. Bacteria mutate (recall mutation is a characteristic of life), and humans are exerting selective pressure on their populations by trying to control them. Control measures are necessary under some circumstances, but exceedingly foolish when applied without considering the consequences.

2. **Non-point Mutations** – Non-point mutations involve regions of DNA containing multiple nucleotides, sometimes one or more genes in length. A variety of such changes can occur, but some of the most common types include:
 - a. **Translocation or transposition** – the movement of a segment of DNA from one location to another within or between chromosomes or plasmids.
 - b. **Deletions** – The removal of segments of DNA one or more genes in length.
 - c. **Inversions** – The reversal of a segment of DNA within a chromosome.

Translocations or transpositions sometimes involve segments of DNA called **transposable elements** or **transposons**. These are segments of DNA that can initiate their own translocation, i.e., movement from one location to another within a chromosomes, between chromosomes or between a chromosome and a plasmid.

Transposons, also called "jumping genes" or "mobile genetic elements", were first observed by Barbara McClintock during her studies involving corn plants. Though their mechanism for movement is variable, some transposons initiate movement by producing an enzyme called **transposase** that recognizes and cuts DNA within regions containing **inverted repeat** sequences. The enzyme breaks phosphodiester bonds within these regions and allows the DNA segment to exit the chromosome and then insert itself in a new location. Some types of transposons reproduce themselves, and only the new copies move. This allows them to increase in number within the genome. Bacterial transposons often carry genes encoding antibiotic resistance, so are involved in the development of new drug resistant strains. A DNA sequence called Alu, associated with the human genome is a transposon about 300bp long. This sequence occurs with variable frequency within humans but is typically repeated between 300,000 and one million times.

Translocation also occurs during meiosis when eukaryotic chromosomes line up in homologous pairs. Sections of chromosome cross one another, break, and then reattach in new locations. This process adds considerable genetic variation to eukaryotic genomes.

Deletions can be initiated by a variety of means, but sometimes result from transposon activity. They can also be caused by errors in chromosome separation during meiosis and by ionizing radiation.

Inversions can occur when segments of DNA form loops, break, and then attach themselves in an inverted position. In this case no DNA is lost, but genes may be significantly altered.

Mutations involving the rearrangement of large sections of DNA are often lethal to cells, but not always. Different types of proteins often contain regions with common amino acid sequences, just as different words often contain common letter sequences (consider the number of words containing the prefix "trans"). If the beginning sequence of one gene is moved and attached to the ending sequence of a different gene, the new combination is more likely to result in a functional protein than is a completely novel amino acid sequence (as would be produced as the result of a frame shift mutation). Non-point mutations are probably responsible for much of the variation currently present within fully functional genomes.

Some Examples of Mutagenic Agents:

A thorough coverage of mutagenic agents is beyond the scope of this class, because they are too numerous and their mechanisms of action are too varied. A brief presentation including a few specific examples will be sufficient. Mutagenic agents (mutagens) can be categorized as either **chemical agents**, or **physical factors** as indicated below:

1. Chemical mutagens

Chemical mutagens are chemical agents known to increase mutation rates or to cause changes in nucleotide sequences within DNA molecules. Two categories of chemical mutagens are:

- a. **Base analogs** – Chemicals resembling naturally occurring nitrogenous bases.

b. **Alkylating agents** – Chemicals that add methyl or ethyl groups to molecules (among other things).

Base analogs are chemicals that resemble naturally occurring nitrogenous bases (A, T, C and G) so can be incorporated into nucleic acids in place of these. When the DNA containing a base analog replicates, the analog codes for the wrong complimentary base and causes a substitution type point mutation. Examples of base analogs include 5'-bromouracil, 5'-fluorouracil, Zidovudine (also called Azidothymidine (AZT) or Retrovir), and many other chemicals.

Analog of uracil, a pyrimidine base with a single ring in its structure, typically resemble thymine (5'-methyl uracil), and are incorporated into DNA in the place of thymine. When the DNA replicates, these chemicals code for thymine instead of adenine, so cause the substitution of a pyrimidine for what would normally have been a purine. Some base analogs cause chain termination during replication, recall the **dideoxynucleotides** described in lab.

Alkylating agents can cause the addition of methyl and ethyl groups to various types of organic compounds. When this involves nucleotides, the result can be substitution type point mutations or deletions. Alkylating agents can also cause purine bases such as guanine to form cross-links between complimentary strands in DNA and these prevent the strands from separating. If strands cannot separate, replication and transcription cannot occur. Sometimes alkylating agents cause DNA repair enzymes to fragment DNA, basically destroying it.

2. **Physical mutagens**

Physical factors recognized as mutagenic agents include various forms of electromagnetic radiation, e.g., ultraviolet light, x-rays and gamma rays.

Ultraviolet light causes damage to DNA by inducing the formation of dimers between adjacent thymine bases, i.e., **thymine-thymine dimers**. When a DNA strand containing a T-T dimer replicates, the dimer will code for only one adenine rather than two, resulting in a deletion type point mutation. Since deletions cause frame shifts, ultra violet light is often lethal to cells.

Many types of bacteria are exposed to sunlight for prolonged periods of time, and have developed resistance to ultra violet rays. The DNA within these cells can still be damaged by ultra violet radiation, i.e., thymine-thymine dimers are still formed; but some cells also have mechanisms for repairing this type of damage.

Some bacteria use "light repair" mechanisms to repair damage caused by ultra violet light. In these organisms, enzymes activated by visible light locate the thymine-thymine dimers and break them apart. Other organisms use "dark repair" mechanisms that do not require light. The enzymes involved in this type of repair can locate thymine-thymine dimers, remove regions of DNA containing these, and replace the damaged regions with new, correct bases.

Ionizing radiation including x-rays and gamma rays are recognized as powerful mutagens. These can cause ionization of various molecules within cells, including

nucleic acids, proteins, lipids, etc. The effects of ionizing radiation on DNA vary, but may include loss of or damage to bases, breaks in one or both strands resulting in deletions and/or chromosome rearrangements, or cross linking within DNA strands inhibiting replication and transcription. This type of radiation is usually lethal to cells, but some bacteria e.g., *Deinococcus radiodurans*, and *D. radiopugnans* can repair damage caused by ionizing radiation. Researchers are very interested in the enzymes formed by these organisms as they may have potential use in the clinical setting.