

# Kerchner's

## Genetic Genealogy Glossary

Written: June 2003

Last Revision: 30 November 2004

Copyright © 2003-2004 Charles F. Kerchner, Jr.  
All Rights Reserved

Notice: Establishing or posting a link to this page is encouraged and permitted. But, reuse or reprinting it in its entirety or in part in other websites or mail lists, or in any other media or publication, for profit or not for profit, without my expressed permission, is not permitted. Printing a hard copy of this report for your own personal, non-commercial use is permitted.

Document URL link address: <http://www.kerchner.com/books/glossary.pdf>

Visit my Genetic Genealogy help page: <http://www.kerchner.com/dna-info.htm>

Order a Genetic Genealogy DNA Testing Dictionary (lots more terms and more) :

<http://www.kerchner.com/books/ggdictionary.htm>

Additions, Corrections, Comments, and Suggestions invited. Contact me at:

<http://www.kerchner.com/contact.htm>

**Adenine** - One of the four bases which are the nucleotides in DNA that are represented by the letters A, T, C, and G. Adenine is the "A". The others are Thymine, Guanine, and Cytosine. When these bases form the base pairs which are the rungs of the DNA ladder, Adenine always pairs with Thymine and Guanine always pairs with Cytosine. See Base Pair.

**Allele** - One of the variant, alternative forms of a gene, nucleotide, or non-gene DNA sequence pattern, at a particular locus (location), on a chromosome. Different alleles, if located in a gene produce variation in inherited characteristics such as hair color or blood type or even diseases. In an individual, one allele (the dominant form) may be expressed more than another form (the recessive one). Different alleles of DNA sequences when not located in genes do not produce variations in inherited characteristics or diseases. Mutations when they occur outside the gene areas do not affect the survival, function, or characteristics of the organism and thus are easily passed along in the DNA replication process for many generations with no effect on the organism. But these different alleles (different nucleotides or different DNA sequence patterns) found at certain known locations in the genome DNA sequence can be used to easily tell one DNA genome from another. The term allele is commonly used in Genealogy by DNA in referring to test results in non-gene areas of the genome to mean the sequenced base (A, T, C, or G) letter character symbol of a DNA nucleotide at a specific nucleotide known location in the genome DNA sequence, the specific DNA sequence pattern at a specific known location in the genome DNA sequence, or the numeric value of the number of repeats of a particular repeating DNA sequence pattern at a certain known location in the genome DNA sequence. See DYS, SNP, STS, Mini-Satellite, Micro-Satellite, and STR.

**Allele Frequency** - The proportion of a particular allele found among the chromosomes carried by individuals in a population.

**AMH (Atlantic Modal Haplotype)** - See Atlantic Modal Haplotype.

**Amino Acids** - A group of 20 different kinds of small molecules made of three bases formed from triplet base combinations of the bases of A, T, C, and G, such as CAG which codes for the amino acid named Glycine. The various triplet base combinations are then linked together in long chains to form proteins. Amino acids are often referred to as the “building blocks” of proteins. See Base and Protein.

**Amplification** - An increase in the number of copies of a specific DNA fragment or the entire DNA molecule. See Polymerase Chain Reaction (PCR).

**Atlantic Modal Haplotype (AMH)** - A descriptive term used to characterize the most common haplotype in parts of Europe. The DYS markers and STR allele repeat values for the AMH are: DYS388=12, DYS390=24, DYS391=11, DYS392=13, DYS393=13, and DYS394 (aka DYS19)=14.

**Autosomal Chromosomes** - The 44 non-sex chromosomes which are arranged in 22 pairs or sets. See Autosome.

**Autosome** - Any chromosome other than a sex determining chromosome. Humans have 22 pairs of autosomes. Autosomes are chromosomes not involved in sex determination. The human genome contains 46 chromosomes. These 46 chromosomes are also called the diploid. The 46 chromosomes are divided into two groups, i.e., the 22 homologous pairs of autosomes numbered by size from 1-22 and 1 pair of sex determining chromosomes (the X and Y chromosomes ... XX = female and XY = male), which is numbered the 23<sup>rd</sup> chromosome pair. Note: The Y sex chromosome is much smaller and very different than the X sex chromosome and thus in males which are XY in the 23<sup>rd</sup> sex determining chromosome pair, this XY pair of chromosomes is non-homologous. Compare to Sex Chromosome. See X Chromosome, Y Chromosome, and Diploid.

**Base** - A small chemical molecule which is the information portion of the nucleotides in DNA. The bases are: Adenine, Thymine, Cytosine, and Guanine. These bases are represented by the letters A, T, C, and G, respectively. When these bases form the base pairs which are the rungs of the DNA ladder, Adenine always pairs with Thymine (AT or TA) and Guanine always pairs with Cytosine (GC or CG). See base pairs.

**Base Pair (bp)** - Two of the bases A, T, C, and G, which weakly link to form a “rung of the DNA ladder.” A DNA nucleotide is made of a molecule of sugar, a molecule of phosphoric acid, and a molecule called a base. The bases are the “letters” that spell out the genetic code and are the data bits of the encoded information in the DNA. The code letters are A, T, C, and G, which stand for the chemicals Adenine, Thymine, Cytosine, and Guanine, respectively. When these bases form the base pairs, which are the rungs of the DNA ladder, Adenine always pairs with Thymine (AT or TA), and Cytosine always pairs with Guanine (CG or GC). The base pair is the portion of the DNA molecule in which the basic nucleotide instruction bit or data point of information of the genetic code is embedded. Base pair combinations provide a chemically stable and reliable method of storing the DNA information at the data bit level. These base pairs are the rungs of the ladder in the DNA molecule. The human genome contains about 3.2 billion base pairs (bp) or rungs of the ladder, or in other words 3.2 billion pieces of information. If each nucleotide piece of data in the genome DNA sequence was stored in a text file on a computer’s hard disk using its representative base letter, i.e., A, T, C, or G, it would take at least a 3.2 giga-byte hard drive to hold it all.

**BP** - A frequently used abbreviation for Base Pair. See Base Pair.

**Base Sequence** - The order of nucleotide bases in a DNA molecule. See DNA Sequence.

**Base Sequence Analysis** - A method, usually automated, for determining the base sequence in a DNA molecule. See DNA Sequencing.

**Buccal Cells** - Moist inner cheek lining cells easily scraped or swabbed off and which are used to easily, painlessly, and inexpensively obtain DNA samples. Use of buccal cells eliminates the need to have blood drawn for DNA tests.

**Cambridge Reference Sequence (CRS)** – The first sequence completed for the human mtDNA molecule. This sequence is now the reference to which all other human maternal line mtDNA sequences are compared. The human mtDNA molecule is circular shaped and is approximately 16, 540 base pairs long. The original sequence has been corrected/revised and that revised sequence is known as the Revised Cambridge Reference Sequence. See DNA Sequence. See Mitochondrial DNA.

**Candidate Gene** - A gene, located in a chromosome region suspected of being involved in a disease, whose protein product suggests that it could be the disease gene in question. DNA tests used in Genealogy by DNA do not test for any genes, or any other patterns or markers or indicators involved with diseases or inherited physical traits or characteristics.

**Carrier** - An individual who possesses one copy of a mutant allele that causes disease only when two copies are present. Although the carrier (when the gene is recessive) is not affected by the disease, two carriers can produce a child who has the disease. DNA tests used in Genealogy by DNA do not test for any genes, or any other patterns or markers or indicators involved with diseases or inherited physical traits or characteristics.

**Cells** - The basic unit of any living organism. It is a small, watery, compartment filled with chemicals. Except for mature red blood cells and the specialized sex cells, each cell also contains a complete copy of the organism's DNA genome. Note: white blood cells do contain a complete copy of the organism's DNA genome and that is where test labs look for the DNA when a blood test is used as the source of DNA.

**Chromosome** - The self-replicating genetic structures of cells containing the cellular, nuclear DNA that contains in its nucleotide sequence the linear array of genes. One of the threadlike "packages" of genes and other DNA in the nucleus of a cell. Different kinds of organisms have different numbers of chromosomes. Humans have 23 pairs of chromosomes, 46 in all. 44 autosomes and two sex chromosomes. Each parent contributes one chromosome to each pair, so children get half of their chromosomes from their mothers and half from their fathers. The mother always contributes an X chromosome in her egg cell. The father can contribute either an X or Y chromosome via his sperm cell. When these two sex determining chromosomes pair up by the union of the sperm cell and the egg cell, the fertilized egg with its new chromosome pair combination (one derived from the mother and one from the father) determines the sex of the child which will develop from the fertilized egg. XX = female and XY = male. The 23 chromosome pairs are often displayed in a pictorial form called a karyotype. The first 22 autosomal pairs are numbered thereon 1-22 based on chromosome size, from the largest to the smallest pair. The sex chromosomes are numbered the 23<sup>rd</sup> chromosome pair in a karyotype. Thus the 23<sup>rd</sup> chromosome pair determines one's sex. See Sex Chromosome, X Chromosome, Y Chromosome, and Autosome.

**CentiMorgan (cM)** - The average length of DNA estimated from exchange of homologous genetic material, between chromosomes during meiosis, averaging 1 cross-over per 100 gametes. A unit of measure of recombination frequency. One centimorgan is equal to a 1% chance that a marker at one genetic locus will be separated from a marker at a second locus due to crossing-over in a single generation. In human beings, 1 centimorgan is equivalent, on average, to 1 million base pairs. See Gamete and Cross-Over.

**Code** - See Genetic Code.

**Coded Regions** - Regions in the DNA molecule where the useful information to the organism or genes are found. Coded regions make up less than 3% of the human genome.

**Coding Strand** - The strand of DNA which contains the encoded information. The unzipped double helix, spiral ladder shaped, DNA molecule contains two strands. One strand, the Coding Strand, contains the coding information and the other non-coding strand contains the mirror, complementary image of the coding information. The coding strand of the DNA molecule is the strand that carries the information. This information is encoded in the gene sequences located at various positions along the strand. The genes contain the necessary information to make proteins. Compare to Non-coding DNA and Non-coding DNA Strand.

**Codon** - Three bases in a DNA or RNA sequence which specify a single amino acid. See Genetic Code.

**CODIS** - Combined DNA Index System (CODIS). A set of genetic markers located in the autosomal chromosome sets which are used to uniquely identify a human being. CODIS numbers appear as two numbers usually separated by a comma which represent the allele value of the pair of STR/VNTR repeats of the genetic marker at certain known locations on the homologous autosomal chromosome pairs. One allele value is from the chromosome received from the father and the other allele value is from the corresponding chromosome in that set received from the mother, e.g., (15,18). The CODIS system typically uses 14 CODIS marker locations in the autosomal chromosomes. A database of CODIS test results is maintained by the FBI and some states and is used to identify people and solve crimes. CODIS type markers are also used in Paternity Tests to see if the candidate parents could have contributed the allele values for the CODIS markers in the combined diploid of the child, to produce the child whose parentage is in question. See Genetic Marker.

**Complementary Sequence** - Nucleic acid base sequences that can form a double-stranded structure by matching base pairs. The complementary sequence to G-T-A-C is C-A-T-G.

**Conserved Sequence** - A base sequence in a DNA molecule (or an amino acid sequence in a protein) that has remained essentially unchanged throughout evolution.

**Cross-Over** - A process occurring during meiosis, the creation of sex cells, whereby genes or segments of DNA from one chromosome in a chromosome pair cross over and swap with similar segments of DNA in the other chromosome in the pair. If this occurs with a segment containing genes then traits will be swapped from one chromosome to the other. If the cross-over occurs at a locus of a genetic marker then it will result in exchange of the genetic marker alleles (marker values) between chromosomes. See Gamete, Sex Chromosome, Allele, and Genetic Markers.

**CRS** - A frequently used abbreviation for Cambridge Reference Sequence. See Cambridge Reference Sequence.

**Cytosine** - One of the four bases which are the nucleotides in DNA that are represented by the letters A, T, C, and G. Cytosine is the "C". The others are Adenine, Thymine, and Guanine. When these bases form the base pairs which are the rungs of the DNA ladder, Adenine always pairs with Thymine and Guanine always pairs with Cytosine. See Base Pair.

**Deletion** - A particular kind of DNA mutation, i.e., loss of a piece of DNA from a chromosome. Deletion of a gene, or part of a gene (even a single nucleotide in a gene), or even an entire chromosome, can lead to a disease or abnormality and the resultant mutated organism may or may not survive. A deletion of a nucleotide or a segment of DNA at some point in time in a non-gene, non-coding portion of the chromosome has no effect on the organism and can provide a genetic marker location to differentiate one DNA genome sequence from another, since the mutation location does not affect the survival, function, or characteristics of the organism. Such markers can be used to trace the evolution of the mutated DNA molecule, as this new mutated DNA molecule is replicated over time, and as more and more mutations occur and accumulate in non-gene areas. See Genetic Markers and Monosome.

**Diploid** - A full set of nuclear genetic material, consisting of paired chromosomes – one chromosome from each parental set. A diploid is expressed as a number, i.e., the number of chromosomes in most cells except as found in the gametes which are the sex cells which only have a haploid, half a set. In humans the diploid number is 46. Compare to Haploid.

**DNA Amplification** - The process of increasing the number of copies of a specific DNA fragment or the entire DNA molecule. See Polymerase Chain Reaction (PCR).

**DNA (DeoxyriboNucleicAcid)** - The large chemical molecule inside the nucleus of a cell that carries the genetic instructions for making living organisms.

**DNA Fingerprinting** - A term for testing DNA markers to a sufficient degree of resolution to specifically identify one person from another with a very high degree of certainty. See CODIS.

**DNA Marker** - A gene or some non-coding segment of DNA whose location in the genome is known and whose different allele values at the location can be used to sort versions of the DNA genome into groups and/or to distinguish one DNA genome from another. See Genetic Marker.

**DNA Replication** - The process by which the DNA double helix unzips and makes an exact copy of itself.

**DNA Segment** – A section or piece of a DNA molecule. Typically, once the segment or piece of DNA is obtained it is then sequenced. Compare DNA Sequence. See DNA Sequencing.

**DNA Sequence** - The exact order of the A, T, C, and G nucleotides in a segment of DNA or the entire DNA molecule. Some sequences, the genes, convey information useful to the organism. But most sequences do not convey information useful to the organism and is considered non-coding, junk DNA. See DNA Sequencing.

**DNA Sequencing** - The process of determining the exact order of the nucleotide bases in a segment of DNA. This is one of the main and important steps in the Genealogy by DNA testing process which is done by the testing lab. See DNA Sequence.

**Dominant** - A gene that almost always results in a specific physical trait or characteristic such as brown eyes or type A blood, even though the person's genome possesses only one copy of this dominant gene in its genotype. The other gene in this genotype case is called recessive. With a dominant gene, the chance of passing on the gene and therefore the physical characteristic to children is 50-50 in each pregnancy. Compare to Recessive.

**Double Helix** - The shape that the two linear strands of DNA assume when bonded together. The structure arrangement of DNA, which looks something like a very long ladder twisted into a helix, coil, or spiral staircase. The sides of the "ladder" are formed by a backbone of sugar and phosphate molecules, and the "rungs" or base pairs consist of nucleotide bases joined weakly in the middle by hydrogen bonds. See Base Pair.

**Duplication** - A particular kind of DNA mutation, i.e., production of one or more extra copies of some segment of the DNA within the genome. Duplication of a gene, or part of a gene (even a single nucleotide in a gene), or even an entire chromosome, can lead to a disease or abnormality and the resultant mutated organism may or may not survive. A duplication of a nucleotide or a segment of DNA at some point in time in a non-gene, non-coding portion of the chromosome has no effect on the organism and can provide a genetic marker location to differentiate one DNA genome sequence from another, since the mutation location does not affect the survival, function, or characteristics of the organism. Such markers can be used to trace the evolution of the mutated DNA molecule, as this new mutated DNA molecule is replicated over time, and as more and more mutations in non-gene areas occur. See Genetic Markers.

**DYS (DNA Y-Chromosome Segment)** - A locus (location) on the Y-Chromosome DNA sequence in a non-gene, non-coding area used as a genetic marker. There are many known DYS locations on the Y-Chromosome. The nomenclature for determining and counting the repeats, the allele value of the DYS, and assignment of a new DYS number is controlled by the HUGO Gene Nomenclature Committee in the effort to maintain industry standardization, e.g., DYS390 is a typical DYS location number used in Y-Chromosome haplotype tests. The typical Y-DNA haplotype test currently looks at 12 DYS marker locations (the low resolution test) or 25 DYS marker locations (the high resolution test). The 12 DYS marker test is used for low cost initial screening of males in a surname project, and for use in estimating the male's Y-Chromosome Haplogroup. The 25 DYS marker test is used to more provide a high resolution haplotype which provides higher resolution when comparing two males Y-Chromosome Haplotype to determine if they shared a common male ancestor in recent genealogical history. Testing only 12 DYS markers or less can lead to false positive match and false negative non-match conclusions. Therefore in most Genealogy by DNA Y-DNA surname projects, the higher resolution 25 DYS marker test is recommended. In general, the more markers looked at by a test the better, i.e., the more marker information obtained per dollar of cost the better. See Genetic Marker, Haplotype, and Haplogroup.

**Electrophoresis** - A method of separating large molecules (such as DNA fragments or proteins) from a mixture of similar molecules. An electric current is passed through a medium (usually a gel) containing the mixture, and each kind of molecule travels through the medium at a different rate, depending on its electrical charge and size. Separation is based on these differences. Agarose and acrylamide gels are the media commonly used for electrophoresis of proteins and nucleic acids.

**Enzyme** - A protein that encourages a biochemical reaction, usually speeding it up. Organisms could not function if they had no enzymes.

**Evolutionarily Conserved** - See Conserved Sequence.

**Expressed Gene** - See Gene Expression.

**Expressed Sequence Tag (EST)** - See Sequence Tagged Site.

**Gamete** - A mature germ sex cell (sperm or ovum/egg cell) possessing a haploid chromosome set (23 for humans) and capable of initiating formation of a new individual by fusion with another gamete, i.e., the union of the sperm and egg cell. See Sex Chromosome, Haploid, and Zygote.

**Gene** - The fundamental functional and physical unit of heredity passed from parent to offspring. Genes are the life function information bearing segments of DNA. Most genes contain the information for making a specific protein. Genes are sequences of DNA information containing several hundred to a few thousand DNA nucleotide instruction data points. Each gene performs a fundamental and specific function such as providing the information to make a protein. Genes control heredity and control day-to-day functions of cells and control the manufacture of proteins. The human genome has about 25,000 separate genes. See Genetic Code.

**Gene Expression** - The process by which a gene's coded information is converted into the structures present and operating in the cell.

**Gene Families** - Groups of closely related genes that make similar products. As an example there are a family of genes which control eye color.

**Gene Mapping** - Determination of the relative position of genes on a DNA molecule and of the distance between them.

**Gene Product** - The biochemical material, either RNA or protein, resulting from expression of a gene. The amount of gene product is used to measure how active a gene is.



**Generation** - The average interval of time between the birth of parents and the birth of their offspring. For purposes of DNA projects, some people use 18 or 20 years, and others may use 25, 30, or 35 years. Also the number of years on average per generation may vary over the centuries within a specific culture as acceptable marital age varies over time, and also can vary from culture to culture. When you are trying to estimate the time in years to the most recent common ancestor (TMRCA), the result will differ depending on the number of years used for a generation.

**Genetic Code** - The instructions in a gene that tell the cell how to make a specific gene product such as a protein from the corresponding amino acids. The letters A, T, C, G are the symbols in the DNA code. The letters symbolize the chemicals Adenine, Thymine, Cytosine, and Guanine respectively, that are the physical nucleotide bases of DNA. Certain combinations of 3 DNA bases correspond to the chemical makeup for the various amino acids. Each gene's genetic code combines the four chemicals in various triplet codes of 3 bases each to spell out 3-letter "words" called Codons, e.g., "CAG". These triplet codes are arranged in the proper sequence in the gene to make a protein, i.e., "code for" a protein. The sequence specifies which amino acid is needed at every step in the process in making a protein by a cell. As an example, the triplet code CAG is a Codon for the amino acid Glycine. Also some triplet codes are stop codes such as "TAG" which stops the protein making process when the protein is complete and properly assembled by the cell, as specified by the genetic code sequence in the gene. See Gene.

**Genetic Map** - See Linkage Map and Physical Map.

**Genetic Marker** - A segment of DNA with an identifiable physical location on a chromosome and whose inheritance can be followed from generation to generation of offspring. A marker can be a gene, but usually it is some segment of DNA with no known function, i.e., in the non-coding "junk DNA" which makes up most (97%+) of the human DNA genome. Genetic markers located in various chromosomes are used to differentiate one human genome from another. Polymorphisms (changes) occurring at known genetic marker locations are used to trace and track the evolution of the human genome over time. See SNP, UEP, STR, STS, Micro-satellite, Mini-satellite, and VNTR.

**Genetic Material** - See Genome.

**Genetic Profile** - A collection of information about a person's genes.

**Genetics** - The science which studies the patterns of inheritance of specific traits.

**Genome** - All the DNA contained in a non-reproductive cell of the organism necessary to replicate the organism. An organism's complete set of DNA instructions necessary for making the complete organism. This includes both the nuclear DNA in the chromosomes within the nucleus and the DNA in a mitochondria organelle which are located within the cell, but outside the nucleus. In humans there are two types of DNA, nuclear and mitochondrial. The mitochondria DNA is a relatively small DNA molecule shaped as a circle located within all the mitochondria organelles within the cells, but which are outside the nucleus. The nuclear DNA is thread-like in shape and is divided up into 46 chromosomes or packets of DNA arranged in 23 sets within the cell nucleus. Some chromosomes contain more DNA than others and thus some chromosomes are large and some are small. The male sex determining Y chromosome happens to be a small chromosome. Its small size plays a significant part in why it is passed down virtually unchanged from father to son over the generations. See Sex Chromosome, Y Chromosome, X Chromosome, Autosome, and Homologous Chromosomes.

**Genome Project** - Research and technology efforts aimed at mapping and sequencing some or all of the genome of human beings and other organisms. See Human Genome Project.

**Genotype** - The actual allele value present and measured in an individual at a particular locus on the DNA molecule. The genetic identity of an individual at the DNA level which may or may not show as an outward, observable characteristic. Compare to Phenotype.

**Germ Line** - The inherited material that comes from the egg or sperm and is passed on to offspring.

**Guanine** - One of the four bases which are the nucleotides in DNA that are represented by the letters A, T, C, and G. Guanine is the "G". The others are Adenine, Thymine, and Cytosine. When these bases form the base pairs which are the rungs of the DNA ladder, Adenine always pairs with Thymine and Guanine always pairs with Cytosine. See Base Pair.

**Haploid** - The number of chromosomes in a sperm or egg cell, which is half the full set, i.e., half the diploid number. In humans the diploid number is 46. Thus in humans the haploid number is 23. One haploid (half set of the diploid) of 23 chromosomes is provided via the mother's egg cell and another haploid (half set of the diploid) of 23 chromosomes is provided via the father's sperm cell. Compare to Diploid.

**Haplogroup** - A group of similar patterned and related descendant haplotypes which share a common ancestor defined by a unique event polymorphism (a one-time SNP mutation) at a specific locus in their DNA sequence, i.e., a UEP. Haplogroups are assigned alpha numeric designators. Different sets of alpha-numeric designators are given to male Y-DNA haplogroups and female mtDNA haplogroups. A commonly occurring paternal line Y chromosome haplogroup designator is the alpha-numeric R1b. A commonly occurring maternal line mtDNA haplogroup designator is the letter H. Maternal line mtDNA haplogroups have been even further personalized by Dr. Brian Sykes in his book, "The Seven Daughters of Eve." Since the haplogroups represent common maternal lines he gave the haplogroups female names which

correspond with the letters. For example my maternal line mtDNA haplogroup is the letter H. And the female name Dr. Brian Sykes gave to that haplogroup is Helena. See UEP and SNP. Compare to Haplotype.

**Haplotype** - A set of allele values for genetic markers (a set of gene or genetic marker DNA sequences) inherited as a unit. A contraction of the phrase “haploid genotype”. Different combinations of polymorphisms at a set of polymorphic sites are known as haplotypes. Commonly used term in Genealogy by DNA for the series of DYS STR numbers which are the allele values of the test results of a set of genetic markers of a Y-Chromosome paternal line DNA test. The results of the maternal line mitochondria DNA test is also known as a haplotype. See Atlantic Modal Haplotype. Compare to Haplogroup and Genetic Marker.

**Heredity** - The handing down of certain traits from parent to their offspring. The process of heredity occurs through passing of genes from parent to their offspring. See Genes.

**Heterozygous** - The presence of different allele values on each chromosome at one or more loci of a homologous chromosome pair. Possessing two different forms of a particular gene, DNA sequence pattern, or nucleotide allele at a genetic marker location, one inherited from each parent. Compare to Homozygous. See CODIS.

**Highly Conserved Sequence** - A DNA sequence that is very similar in several different kinds of organisms. Scientists regard these cross species similarities as evidence that a specific gene performs some basic function essential to many forms of life and that evolution has therefore conserved its structure by permitting few mutations to accumulate in it. See Conserved Sequence.

**Homeobox** - A short stretch of nucleotides whose base sequence is virtually identical in all the genes that contain it. It has been found to occur in the genes of many organisms from fruit flies to humans.

**Homologous Chromosomes** - A pair of same sized and same function chromosomes containing the same linear gene sequences, each derived from one parent. However, the allele value of the respective gene from each parent may be the same or different. The autosomal chromosomes which associate themselves in pairs or sets according to their size and function,  $\frac{1}{2}$  the set from the father and  $\frac{1}{2}$  the set from the mother are homologous within each pair. Each chromosome in the homologous set is identical in size and function to the other but with each chromosome in the set usually having differences in genes and marker allele values at various loci. As an example a homologous gene in one chromosome of the chromosome pair may code for brown eyes and the other homologous gene in the other chromosome of the chromosome pair may code for blue eyes. In this example the Brown gene is dominant and the Blue gene is recessive so the person would have brown eyes. See Autosome, Autosomal Chromosomes, Heterozygous, and Homozygous. Compare to Sex Chromosome.

**Homozygous** - The presence of the same allele values on each chromosome at one or more loci of a homologous chromosome pair. Possessing two identical forms of a particular gene of DNA sequence pattern, or nucleotide allele at a genetic marker location, one inherited from each parent. Compare to Heterozygous. See CODIS.

**HUGO (Human Genome Organization)** - See Human Genome Organization.

**Human Genome Organization (HUGO)** - The Human Genome Organization (HUGO) is the international organization of scientists involved in the Human Genome Project, the global initiative to map and sequence the human genome. The sequencing was completed in 2003.

**Human Genome Project** - An international research project to map each human gene and to completely sequence human DNA, i.e., to sequence the entire human genome. The human genome sequencing was completed in 2003. See Genome.

**HVR** – Hyper Variable Region. The region in the circular maternal line mtDNA molecule which is not used for any known cell function and thus mutations in this region accumulate over time. Also known as HVS or Hyper Variable Segment. See Mitochondrial DNA. See DNA Sequence.

**HVR1** – The region in the circular shaped mtDNA molecule from base pair 16,001 to base pair 16,540. The total molecule length is about 16,540 base pairs. This was the first region used as a reference for maternal line mtDNA Genetic Genealogy testing. It was labeled HVR1 because it was the first region used even though it is in the second place position in the mtDNA circular molecule as one proceeds around the circular molecule from base pair 1 to base pair 16,540. Also known as HVS1 for Hyper Variable Segment. Compare HVR2. See Mitochondrial DNA. See DNA Sequence.

**HVR2** – The region from base pair 61 to base pair 570 in the circular mtDNA molecule. Also known as HVS2 for Hyper Variable Segment. Compare to HVR1. See Mitochondrial DNA. See DNA Sequence.

**HVS** – Hyper Variable Segment. See HVR – Hyper Variable Region.

**IBD** – Identical by Descent. When a person's haplotype is identical to another person's haplotype because they share a common ancestor and the haplotype has been preserved and passed down through the generations in their respective lines of descent from the common ancestor.

**IBS** – Identical by State. When a person's haplotype is identical to another person's haplotype because the mutations accumulating over time in the two originally different lines just happen to be in a direction that makes their haplotypes more similar.

**Independent Assortment, Law of** - See Law of Independent Assortment.

**Inherited** - Transmitted through genes and chromosomes from parents to offspring.

**Insertion** - An extra nucleotide or DNA segment sequence is inserted into the DNA within the parent DNA sequence. See STR and Polymorphism.

**Interphase** - The period in the cell cycle when DNA is replicated in the nucleus; followed by mitosis. See Mitosis.

**Junk DNA** - Non-coding DNA located in DNA areas between the genes. Junk DNA refers to the DNA sequences located between the gene (protein making) sequences in the coding strand of the human genome. These non-coding areas in the coding strand have no known function and the DNA sequences in these areas are often referred to by many as “Junk DNA.” Junk DNA makes up over 97% of the human genome. But while the Junk DNA is of no known use to the organism it is very useful to geneticists and genealogists for it contains many loci (locations) in which many easily measurable and identifiable Genetic Markers are found. Compare to Coded Regions. See Genetic Marker.

**Karyotype** - The chromosomal complement of an individual including the numeric identification of the chromosomes. The autosomal chromosome sets are arranged in the karyotype or picture of the set of chromosomes from the largest in size to the smallest and are numbered from 1-22. The 23<sup>rd</sup> set contains the sex determining chromosomes, i.e., XX=female or XY=male. The term karyotype is also used to refer to a photograph or sketch of an individual's set of chromosomes.

**Kilobase (kb)** - Unit of length of DNA fragments equal to 1000 nucleotides. Compare to Megabase.

**Knockout** - Inactivation of specific genes. Knockouts are often created in laboratory organisms such as yeast or mice so that scientists can study the knockout organism as a model for a particular inherited characteristic or disease.

**Law of Independent Assortment** - A law proposed by Mendel, the father of genetics, which is his second law. Mendel's Second Law - the law of independent assortment states that during gamete (sex cell) formation the segregation of the alleles of one allelic pair is independent of the segregation of the alleles of another allelic pair. See Gamete, Sex Chromosome, Allele, and Genetic Marker.

**Linkage** - The proximity of two or more markers (e.g. genes or genetic markers) on a chromosome. The closer together the genes or markers are the lower the probability that they will be separated during DNA replication and hence the greater the probability that they will be inherited together.

**Linkage Map** - A map of the relative positions of genetic loci on a chromosome determined on the basis of how often the loci are inherited together. Distance is measure in centimorgans (cM), which is measuring unit about 1 million base pair long. Compare to Physical Map.

**Loci** - The plural of Locus.

**Locus** - The place or location on a chromosome or DNA sequence where a specific gene or genetic marker is located. A kind of address for the gene or maker. The plural is "loci", not "locuses". See DYS and Genetic Markers.

**LOD Score** - A statistical estimate of whether two loci are likely to lie near each other on a chromosome and are therefore likely to be inherited together. A LOD score of three or more is generally taken to indicate that the two loci are close.

**Map** - See Linkage Map and Physical Map.

**Marker** - An identifiable physical locus (location) on a chromosome whose inheritance can be monitored. Markers can be genes or some segment of DNA with no known coding function but whose pattern of inheritance can be determined. See Genetic Marker.

**Megabase (mb)** - Unit of length of DNA fragments equal to 1 million nucleotides. Compare to Kilobase. See CentiMorgan.

**Meiosis** - The process of cell division which creates sex cells. The process of two consecutive cell divisions in the diploid (46 chromosome) progenitors of sex cells. Meiosis results in four rather than two daughter cells, each with a haploid (23 chromosomes) set of chromosomes which have been selected randomly from the parent diploid and which set of autosome chromosomes in the haploid are not exactly the same as any in the parent diploid set of chromosomes due to cross-over effects during synapsis in the early part of the meiosis process. A notable exception to this cross-over effect during meiosis is the Y chromosome. Because of its very different size and different genetic content and being non-homologous with its partner X chromosome in the gender set of chromosomes, cross-over effects in the coding regions of the Y chromosome do not occur and it is passed virtually unchanged from a male parent cell to its male producing sex cells, i.e., the sperm cells bearing a Y chromosome. Thus the Y chromosome is passed from father to son virtually unchanged. Compare to Mitosis. See Law of Independent Assortment and Y Chromosome.

**Mendel, Johann Gregor** - Austrian biologist, born in 1822 and died in 1884, who laid the foundations for the science of genetics. Mendel was a monk whose controlled experiments with breeding peas in the monastery garden led him to conclude that the heritable units we now call genes were not blends of parental traits but separate physical entities passed individually in specific promotions from one generation to the next. Mendel's discoveries were ignored for several decades, but other biologists finally recognized their significance early in the 20<sup>th</sup> century.

**Mendelian Inheritance** - The manner in which genes and traits are passed from parents to children. Examples of Mendelian inheritance include autosomal dominant, autosomal recessive, and sex-linked genes.

**mRNA** - See Messenger RNA.

**Messenger RNA (mRNA)** - RNA that serves as a template for protein synthesis. Compare to Transfer RNA. See Genetic Code.

**Metaphase** - A stage in mitosis or meiosis during which the chromosomes are aligned along the equatorial plane of the cell. See Meiosis and Mitosis.

**Micro-Satellite** - Repetitive stretches of very short sequences of DNA, usually 3, 4, or 5 nucleotides, such as GATA, repeated many times, e.g., GATAGATAGATA within the stretch of DNA. They are used as genetic markers to track inheritance. Also known as an STR. The number of repeats, the allele value of the marker, can be many times such as 6-32 times, etc. Compare to Mini-Satellite and Sequence Tagged Site. See STR and Genetic Marker.

**Mini-Satellite** - Repetitive stretches of short sequences of DNA, usually 15-100 nucleotides, e.g. MSY1, CACAATATACATGATGTATATTATA, repeated many times within the stretch of DNA in the Y-Chromosome. They are used as genetic markers to track inheritance. The number of repeats, the allele value of the marker, can be 20-50 times, etc. Compare to Micro-Satellite and Sequence Tagged Site. See VNTR and Genetic Marker.

**Mitochondria** - The organelles located within the cell that generates energy for the cell. Mitochondria are passed down from mother's to all their offspring via her egg cell and then via her daughters' egg cells to their offspring.

**Mitochondrial DNA (mtDNA)** - The genetic material of the mitochondria which contains the code to produce more Mitochondria. The mtDNA molecule is a very small, simple, and efficient molecule compared to the nuclear DNA. Human mtDNA is only 16,569 base pairs in length compared to about 3.2 billion base pairs for the nuclear DNA. The mtDNA molecule is also shaped differently. It is shaped as a circle with no beginning or end, unlike nuclear DNA which is in long thread-like molecules with a beginning and an end. MtDNA is passed down from females to all their offspring via her egg cell virtually unchanged over many generations, except for an occasional natural mutation. In turn her daughter's egg cells pass her mtDNA to her offspring. Thus mtDNA is passed down virtually unchanged except for occasional natural mutations over many generations via the direct female line of descent. While males inherit the same mtDNA in their cells as their mother, they of course cannot pass it along since males do not produce egg cells. But since all offspring of the mother inherit her mtDNA, it can be used to determine if two people are from the same biological mother. It can also be used, using the slight mutational differences in non-coding regions in the molecule, where genetic markers are found, to sort humans into groups called Haplogroups which share similar mtDNA genetic marker mutation patterns, which have been passed down from our ancient maternal ancestor. It can also be used to see if two people descend via separate recent maternal lines from a common recent maternal ancestor, in a genealogical time frame. Thus mtDNA is a useful tool for Genealogy by DNA purposes. See Genetic Markers.

**Mitosis** - The process of cell division which exactly replicate themselves. The process of nuclear cell division that produces two daughter cells that are genetically identical to each other and to the parent cell. Compare to Meiosis.

**Molecule** - The smallest particle of a substance that retains the properties of the substance and is composed of one or more atoms.

**Monosomy** - Possessing only one copy of a particular chromosome instead of the normal two copies. See Autosomal.

**Multiplexing** - A sequencing approach that uses several pooled samples simultaneously greatly increasing sequencing speed. See DNA Sequencing.

**MRCA** - Most Recent Common Ancestor. The most recent common male or female ancestor from which two individuals both biological are descended from. Compare to TMRCA.

**mtDNA** - See Mitochondrial DNA.

**Mutation** - Any heritable change in a DNA sequence. A permanent structural alteration or change in the DNA sequence. In most cases DNA changes have no effect on the organism. But sometimes the changes do cause harm and the organism will not thrive or survive. And occasionally a mutation can improve an organism's chance of surviving and it will pass the beneficial change on to its descendants. This is the principal of natural selection. A mutation in a non-gene, non-coding area of the DNA sequence has no effect on the organism and thus will be easily and readily passed along to its descendants. Many of these mutations in non-gene, non-coding areas of the DNA sequence are used as Genetic Markers and are used to trace the evolution of the DNA sequence over time. See Genetic Marker and Polymorphism.

**Nitrogenous base** - A nitrogen-containing molecule having the chemical properties of a base. See Base.

**Non-coding Regions** - Regions of the DNA molecule which do not contain any genes with information useful to the organism. Also called Junk DNA. Non-coding regions make up over 97% of the human genome. See Junk DNA.

**Non-coding Strand and Non-coding DNA** - The non-coding strand of DNA is the strand that does not carry the information necessary to make a proteins. The unzipped DNA spiral ladder shaped molecule contains two strands. One strand contains the coding information and the other strand contains the mirror, complementary image of the coding information. The non-coding strand is also known as the anti-sense strand. The term non-coding DNA is also used to refer to the DNA sequences located between the gene (protein making) sequences in the coding strand. These non-coding areas in the coding strand have no known function and the DNA sequences in these areas are often referred to as "junk DNA." Junk DNA makes up over 97% of the human genome. But the junk DNA while of no known use to the organism is useful to geneticists and genealogists for it contains many locations in which many easily measurable and identifiable genetic markers are found. See Genetic Marker.



**Non-Reproductive cells** - Cells which are not Sex Cells. Compare to Sex Cells.

**Nonsense Mutation** - A single DNA base substitution resulting in a stop codon. See Base and Codon.

**Nucleic Acid** - A large molecule composed of nucleotide sub-units.

**Nucleotide** - A sub-unit of DNA or RNA. The basic building blocks of the information portion of the DNA molecule. There are four separate nucleotide bases: [A] Adenine, [T] Thymine, [C] Cytosine, and [G] Guanine. These four bases are chemically only able to combine into 2 unique base pair combinations, AT and CG. The two base pair types form the base pairs in the rungs of the ladder of the DNA molecule. Compare to Base Pair.

**Nucleus** - The cellular organelle, i.e., the central cell structure, that houses the chromosomes.

**Organelle** - A small specialized part of a cell that performs some function for the cell that is analogous to an organ in our body. See Mitochondria and Nucleus.

**Organism** - A living thing with complex structure and interdependent and subordinate parts whose relationship and properties are largely determined by their function in the whole structure and which whole structure is constituted to carry on the activities of life.

**PCR** - Polymerase Chain Reaction. See Polymerase Chain Reaction.

**Pedigree** - A simplified diagram of a family's genealogy that shows family members' relationships to each other. In genealogy it is commonly called a Pedigree Chart. When used for genetic purposes, it shows how a genetic marker, particular trait, characteristic, or even a disease has been inherited.

**Peptide** - Two or more amino acids joined by a peptide bond. Compare to Protein. See Amino Acid.

**Phenotype** - The observable traits or characteristics of an organism. For example hair color, eye color, weight, or the presence or absence of a disease. Phenotypic traits are not necessarily genetic. Your picture is one format of recording and representing your Phenotype. Compare to Genotype.

**Physical Map** - A map of the locations of identifiable landmarks on DNA, i.e., genes and genetic markers. Compare to Linkage Map.

**Polymerase Chain Reaction (PCR)** - A fast, inexpensive chemical process and technique for making an unlimited number of copies of any segment of DNA or the entire DNA molecule. Sometimes called "molecular photocopying," PCR has had an immense impact on biology and medicine, especially genetic research. A method of DNA analysis that amplifies a specific DNA segment or gene region of the DNA molecule, or the entire DNA molecule, and which allows less costly and quicker DNA analysis than prior methods. Dr. Kary Mullis of the USA won ½ the Nobel Prize in Chemistry in 1993 for his invention of the PCR method of copying and amplifying DNA molecules.

**Point Mutation** - A change in a single nucleotide. See SNP and UEP.

**Polymorphic** - Something such as a genotype which exists in many forms.

**Polymorphism** - A naturally occurring or induced variation in the sequence of genetic information on a segment of DNA. See Genetic Marker. A commonly observed variation in the sequence of DNA among individuals. See STR and Genetic Marker.

**Polymorphic Site** - A site in the genome DNA sequence within a chromosome such as a gene or genetic marker which is found in two or more forms, i.e., which has at least two, but usually many more, identifiable DNA sequences (allele values) at that site, found within the human population. Also called a polymorphic locus. See Genetic Marker.

**Primer** - A short oligonucleotide sequence (a chain of 2-10 nucleotides) used in a polymerase chain reaction.

**Probe** - A piece of labeled DNA or RNA or an antibody used to detect the function of a gene.

**Pro-nucleus** - The nucleus of a sperm or an egg prior to fertilization. Sperm and egg cells carry half the number of chromosomes of other non-reproductive cells. When the pro-nucleus of a sperm fuses with the pro-nucleus of an egg, their chromosomes combine and become part of a single nucleus in the resulting zygote, containing a full set of chromosomes. See Haploid. Compare to Nucleus.

**Protease** - A protein that digests other proteins.

**Protein** - A large complex molecule made up of one or more chains of amino acids. Proteins perform a wide variety of activities in the cell. Compare to Peptide. See Amino Acid.

**Pseudo-gene** - A sequence of DNA that is very similar to a normal gene but that has been altered slightly so it is not expressed. Such genes were probably once functional but over time acquired one or more mutations that rendered them incapable of producing a protein product.

**Purine** - A nitrogen-containing, single-ring, basic compound that occurs in nucleic acids. The purines in DNA and RNA are adenine and guanine. Compare to Pyrimidine. See Base.

**Pyrimidine** - A nitrogen-containing, double-ring, basic compound that occurs in nucleic acids. The pyrimidines in DNA are cytosine and thymine and in RNA they are cytosine and uracil. Compare to Purine. See Base.

**Recessive** - A genetic trait or disorder that appears only in individuals who have received two copies of the recessive gene, one from each parent. Compare to Dominant.

**Recombination** - The natural process by which progeny derive a random combination of genes and genetic markers from both parents. In humans and many higher organisms this occurs during the creation of sex cells (egg cell and sperm cell) by the process of independent assortment and cross-over. Compare to Recombinant DNA.

**Recombinant DNA** - Recombinant DNA refers to DNA which has been altered by non-natural, external intervention by joining genetic material from two different sources using recombinant DNA technology. A variety of techniques that molecular biologists use to manipulate DNA molecules to study the expression of a gene.

**Regulatory Regions or Sequences** - A DNA base sequence that controls gene expression. See Gene Expression.

**Reproductive Cells** - See Sex Cells.

**Restriction Enzyme, Endonuclease** - A protein that recognizes specific, short nucleotide sequences and cuts DNA at those sites. Bacteria contain over 400 such enzymes that recognize and cut over 100 different DNA sequences. See Restriction Enzyme Cutting Site.

**Restriction Enzyme Cutting Site** - A specific nucleotide sequence of DNA at which a particular restriction enzyme cuts the DNA. Some sites occur frequently in DNA (e.g., every several hundred base pairs), others much less frequently (e.g., every 10,000 base pairs).

**Restriction Fragment Length Polymorphisms (RFLP)** - Genetic variations at the site where a restriction enzyme cuts a piece of DNA. Such variations affect the size of the resulting fragments. These sequences can be used as genetic markers on physical maps and linkage maps. RFLP is also pronounced "rif" lip. See Genetic Markers and Polymorphism.

**RFLP** – See Restriction Fragment Length Polymorphisms.

**RNA (RiboNucleicAcid)** - A chemical similar to a single strand of DNA. In RNA, the letter U, which stands for Uracil, is substituted for T in the genetic code. RNA delivers DNA's genetic message to the cytoplasm of a cell where proteins are made.

**Ribonucleotides** - See Nucleotide.

**Ribosome** - Organelles within a cell that are the site of protein synthesis.

**Sequence** - The base sequence. See DNA Sequence.

**Sequence Tagged Site (STS)** - A short DNA segment of about 200-500 base pairs that occurs only once in the human genome and whose exact location and exact sequence order of bases are known. And these sequences do not repeat. Because each is unique, an STS is helpful for chromosome placement of the mapping and sequencing data from many different laboratories. An STS serves as a landmark on the physical map of the human genome. Compare to Micro-Satellite, STR, Mini-Satellite, and VNTR. See Genetic Marker and DYS.

**Sequencing** - See DNA Sequencing.

**Sex Cells** - The sperm cells in males and the egg cells in females. Compare to Somatic Cells. See Gamete.

**Sex Chromosome** - One of the two chromosomes that determine an organism's genetic sex. Humans have two kinds of sex chromosomes, one called X and the other called Y. Normal females possess two X chromosomes and normal males one X and one Y. The Y chromosome which makes a child male is thus passed from fathers to sons. The human genome diploid consists of 46 chromosomes. 22 pairs of autosomes and 1 pair of sex chromosomes (the X and Y chromosomes ... XX diploid cell = female and XY diploid cell = male). Thus it can be seen that the human genome has 2 sex chromosomes. The sex chromosomes are numbered as the 23<sup>rd</sup> chromosome pair in a karyotype, which is a photograph of all the chromosomes arranged in a specified order. Compare to Autosome. See Y Chromosome, X Chromosome, and Karyotype.

**SNP (Single Nucleotide Polymorphisms)** - Common variations in the nucleotide allele value at a certain nucleotide position in the human genome that occurred over time. These nucleotide allele value variations in the human genome DNA sequence (a base A becomes the base T, a base A becomes a G, or other similar variations) occur at a frequency of about one in every 1,000 bases in the genome. When the change occurs it is called a polymorphism. These variations can be used to track inheritance in families and the evolution of the human genome over time. SNP is pronounced "snip". Compare to UEP and STR. See Genetic Marker.

**Somatic Cells** - All body cells, except the reproductive gamete cells and their precursors. Compare to Sex Cells.

**Species** - A single, distinct class of living creature with features that distinguish it from others.

**STR (Short Tandem Repeats)** - Patterns in the DNA sequence which repeat over and over again in tandem, i.e., right after each other. As an example: GATAGATAGATAGATAGATA. In this example the tetra-nucleotide (four nucleotide) pattern of GATA is repeated 5 times. Thus this STR would have an allele or marker value of 5. An STR is also known as a micro-satellite. Compare to Mini-Satellite and VNTR. See Micro-Satellite and Genetic Marker.

**STS** - See Sequenced Tagged Site.

**Substitution** - Replacement of one nucleotide in a DNA sequence by another nucleotide, i.e., A for T, C for G, A for G, etc., or replacement of one amino acid in a protein by another amino acid. See Genetic Marker and Polymorphism.

**Synapsis** - The very close alignment and association of the homologous chromosomes during the first prophase of Meiosis held to be the mechanism for cross-over of genetic material between the two homologous chromosomes in each matching pair of the autosomal chromosomes.

**Tandem Repeat Sequence** - Multiple copies of the same base sequence on a chromosome used as a marker in physical mapping. See STR, Physical Mapping, and Genetic Marker.

**Telomere** - The ends of chromosomes. These specialized structures are involved in the replication and stability of the linear DNA molecules. See DNA Replication.

**Thymine** - One of the four bases which are the nucleotides in DNA that are represented by the letters A, T, C, and G. Thymine is the "T". The others are Adenine, Cytosine, and Guanine. When these bases form the base pairs which are the rungs of the DNA ladder, Adenine always pairs with Thymine and Guanine always pairs with Cytosine. See Base Pairs.

**Time to Most Recent Common Ancestor (TMRCA)** - A length of time calculated based on the average mutation rate per marker per transmission event of the genetic markers being used and the number of mutational differences observed in their respective haplotypes which is then used to determine how far back in time two individuals shared a common male or common female ancestor. Compare to MRCA.

**TMRCA** - See Time to Most Recent Common Ancestor.

**Traits** - Ways of looking, thinking, or being. Traits that are genetic are passed down via the genes in our DNA from parents to offspring.

**Transcription** - The synthesis of an RNA copy from a sequence of DNA (a gene). The first step in gene expression. Compare to Translation. See Gene Expression.

**tRNA** - See Transfer RNA.

**Transfer RNA (tRNA)** - A class of RNA having structures with triplet nucleotide sequences that are complementary to the triplet nucleotides coding of mRNA (messenger RNA). The role of tRNAs in protein synthesis is to bond with amino acids and transfer them to the ribosomes, where proteins are assembled according to the genetic code carried by the mRNA. Compare to Messenger RNA. See Genetic Code.

**Transformation** - A process by which the genetic material carried by an individual cell is altered by incorporation of exogenous DNA into its genome. Compare Mutation.

**Transition** - A type of base pair, nucleotide mutation involving the replacement of a purine with another purine, or of a pyrimidine with another pyrimidine (e.g., replacing base pair GC with base pair AT). In this example the pyrimidine base C is replaced with the pyrimidine base T in the coding strand. This type of mutation is much more common than a transversion. Compare to Transversion. See SNP, UEP, Base, and Mutation.

**Translation** - The process in which the genetic code carried by mRNA (messenger RNA) directs the synthesis of proteins from amino acids. Compare to Transcription.

**Transversion** - A type of base pair, nucleotide mutation involving replacement of a purine with a pyrimidine, or a pyrimidine with a purine (e.g., replacing base pair GC with base pair TA). In this example the pyrimidine base C is replaced with the purine base A in the coding strand. This type of mutation is much less common than a transition. Compare to Transition. See SNP, UEP, Base, and Mutation.

**Trisomy** - Possessing three copies of a particular chromosome instead of the normal two copies.

**UEP** - See Unique Event Polymorphism.

**UEP Site** - A polymorphic site at which only two allele values or forms of the genotype are observed in the entire human population at that site and thus it is assumed that a single historical mutation event was responsible for the observed polymorphism. The one allele being the original, ancestral value and the other being the mutated value.

**Unique Event Polymorphism (UEP)** - A change at a locus which is assumed to have had only one single historical mutation event at the locus.

**Uracil** - One of the four bases (A, U, C, and G) in RNA (RiboNucleicAcid). The others are Adenine, Guanine, and Cytosine. Uracil replaces Thymine, which is the fourth base in DNA (DeoxyriboNucleicAcid). Like Thymine, Uracil always pairs with Adenine.

**Variable Number Tandem Repeats (VNTR)** - A defined region of DNA containing multiple copies of short sequence patterns of bases, typically 9 to 80 bases, which are repeated a number of times in tandem, i.e., immediately adjacent to each other. The number of repeats varying among individuals in the population. See Mini-satellite and Genetic Marker.

**Vector** - An agent, such as a virus or a small piece of DNA called a plasmid, that carries a modified or foreign gene. When used in gene therapy, a vector delivers the desired gene to a target cell.

**Virus** - A molecular biological entity that can reproduce only within a host cell. Viruses consist of nucleic acid covered by protein. Some animal viruses are also surrounded by a membrane. Inside the infected cell, the virus uses the synthesizing capability of the host cell to produce progeny virus.

**VNTR** - See Variable Number Tandem Repeats.

**X Chromosome** - One of the two types of sex determining chromosomes. The unfertilized egg cell always is created with one X chromosome in the haploid of the human diploid in the unfertilized egg cell. The other sex chromosome is provided by the male's sperm cell. An X chromosome with the haploid in the sperm cell from the male and the fertilized egg produces a female offspring. A Y chromosome with the haploid in the sperm cell from the male and the fertilized egg produces a male offspring. Thus the Y chromosome is passed from fathers to sons. It should be noted that the X chromosome is much larger than its Y chromosome counterpart. Also, when the two X chromosomes, one from each parent, are paired with each other in a fertilized egg cell, and thus ultimately producing a female offspring, these two X chromosomes are of equal size and are thus homologous in the female offspring. Thus the X chromosomes become subject to cross-over effects during subsequent egg cell creation in the offspring, and thus the homologous gene alleles and genetic marker alleles in both those X chromosomes can randomly swap positions in the next generations making it very difficult to track a particular X chromosome over more than just a couple of generations. Determining a common ancestor for an X chromosome is very difficult beyond a couple generations. Thus the X chromosome is not a very useful tool for Genealogy by DNA purposes. However the Y chromosome because of its small and very different size is not homologous with its X chromosome counterpart in a male offspring and is thus not subjected to cross-over effects and thus the Y chromosome is passed down from fathers to sons unchanged over many generations, except for occasionally occurring natural mutations. Compare to Y Chromosome. See Sex Chromosome and Autosome.

**YCAII** - A highly polymorphic di-nucleotide (two nucleotide) repeating pattern locus in the Y-Chromosome DNA sequence. It is now also known as DYS413. Sometimes this marker was given in two parts YCAIIa and YCAIIb. See Genetic Marker and DYS.

**YGATA** - A highly polymorphic tetra-nucleotide (four nucleotide) polymorphic repeating pattern locus in the Y-Chromosome DNA Sequence. YGATA A7.1 is now known as DYS460. Y GATA A7.2 is now known as DYS461. YGATA A4 is now known as DYS439. See Genetic Marker and DYS.

**Y Chromosome** - One of the two types of sex determining chromosomes, which are the X and Y chromosomes. Men have an X and a Y chromosome. The Y chromosome is the chromosome that determines a person is male. Since a female has two X chromosomes, the unfertilized egg cell of the mother is always created with one X chromosome in its haploid, i.e., the  $\frac{1}{2}$  set of the human diploid contributed by the mother in the unfertilized egg cell. The other sex chromosome required to fertilize the egg cell is provided by the male's sperm cell's  $\frac{1}{2}$  set of the human diploid. If another X chromosome is provided to the egg cell via the male sperm cell, then the fertilized egg cell will develop into a female child. If a Y chromosome is provided to the egg cell via the sperm cell from the male, then the fertilized egg cell will develop into a male child. Thus the Y chromosome is passed down from fathers to sons. The Y chromosome is very unique. The Y Chromosome is much smaller and very different from its X chromosome partner in the 23<sup>rd</sup> chromosome sex determining pair. These two chromosomes are not homologous like the other 22 autosome chromosome pairs are. Thus during the metaphase of meiosis (sex cell creation process) because the Y chromosome is so relatively small and very different from the X chromosome, they do not line up very well during the metaphase of meiosis. Synapsis does not occur. Thus, except for the very ends of the chromosomes, they do not exchange genetic material readily via cross-over like the homologous autosome chromosomes do. Thus during creation of male, Y chromosome bearing, sperm cells the Y chromosome is passed to the Y sperm cell virtually unchanged. Thus the genes and genetic markers in the Y chromosome are passed from father to son over many generations virtually unchanged, except for an occasional natural mutation. The Y chromosome is not subject to genetic material mixing by cross-over effects over the generations like all the other chromosomes are. Thus the Y chromosome is unique and is a valuable and useful tool for determining common ancestry in the male population since if two men share a recent common ancestor, in a genealogical time frame, their Y chromosomes will be virtually identical. The Y chromosome is thus a very useful tool for Genealogy by DNA purposes. Compare to X Chromosome. See Sex Chromosome, Autosome, and Genetic Marker.

**Zygote** - A fertilized cell. A cell containing a complete diploid set of chromosomes formed by the union of two gametes, i.e. the union of the male sperm cell with the female egg cell. See Gamete and Sex Chromosome.