

## DNA and Genealogy: Putting the Genes Back into Genealogy

Stephen McDonald, MD 10 June 2006

Genetics is the scientific study of inheritance particularly of specific visible traits. Much has been learned about inheritance since Gregor Mendel did his early work in the 1800's. With the discovery of chromosomes in cells and DNA in the 1950's genetics took an important leap forward. The code for all the genes in man was published in 2001 (the human genome). The potential for using DNA to look at ones genealogy is becoming more important and available. A brief introduction to the terminology of genetics will provide the genealogist a background to make decisions about whether it would be helpful for ones own genealogy dilemmas.

Genealogical DNA testing has great potential to help solve many dilemmas. It is important to realize what DNA testing can and can't do. The two types of testing most used are YDNA ( the Y chromosome checking direct male ancestors) and the MtDNA (mitochondrial DNA which checks direct maternal lines.) DNA testing:

### •Can't

- replace conventional (paper) research
- provide 100% confidence
- establish unknown family connections
- remove all brick walls

### •Can

- test family traditions
- test suspected family connections
- test connections between different family lines
- map established family trees

## Glossary of Terms

*Extracted from "Trace Your Roots with DNA by Megan Smolenyak*

**Allele** – one of the alternative versions of a gene or genetic marker that can exist in a particular location on a chromosome; in genealogy, most often used to refer to the number of repeats in a given STR (e.g., 14-15-16 etc.); the variation in the number of these repeats is used to differentiate people

**Amino acids** – building blocks of protein

**Autosomal** – pertaining to a gene or genetic marker in any chromosome other than the sex chromosomes; in genealogy, frequently used to refer to tests other than Y-DNA and mtDNA (such as the BioGeographical Ancestry test)

**Bases** – building blocks of DNA named adenine (A), cytosine (C), guanine (G), and thymine (T); see nucleotide

**Cambridge Reference Sequence (CRS)** - the complete sequence (16,569 bases long) of mitochondrial DNA; first analyzed in a woman from Cambridge, England

**CODIS markers** – the 13 autosomal STR markers used by the FBI and included in the Combined DNA Index System, a database of records with DNA profiles from criminal offenders, crime scene evidence, and missing persons and their relatives

**Collateral lines** – branches that are descended from a common ancestor but not through the same direct line; in genealogy, researching these lines often produces DNA-testing candidates for other lines whose DNA has daughtered or petered out

**DNA** – DeoxyriboNucleic Acid (dee-ox-ee-rye-boh-new-lee-ic acid); double-stranded molecule (that forms the familiar double-helix) that encodes genetic information; composed of phosphate, deoxyribose (a sugar), and the four bases (A, C, G, and T)

**DNA polymerase** – enzyme associated with DNA replication (It copies separated strands of DNA and uses each copy as a template for the next.) and thus critical to the duplication of genetic information; see *Polymerase Chain Reaction* (PCR)

**Dominant** – pertains to alleles (and corresponding traits) that always manifest whenever present; see *recessive* for contrast

**DYS-(DNA Y Segment)** -- a number designating a given marker on the Y chromosome (e.g, DYS19, DYS390 etc.); system established by international convention.

**Gene** – fundamental unit of heredity; a sequence of nucleotides on a chromosome that codes for proteins (and ultimately, some aspect of an organism's development)

**Genealogy** – the merger of genetics and genealogy; the use of DNA to learn about one's roots

**Genome** – all the genetic material in an organism or a cell (sometimes excluding mtDNA); in humans, this is the whole set of 23 chromosomes with their more than 3 billion base pairs

**Genotype** – the actual DNA sequence (i.e., combination of alleles) for some region of interest (i.e., selected loci located on paired or unpaired chromosomes)

**Haplogroup** – a large cluster of people who share the same UEP and whose ancestry converges on the person who was the founding father or founding mother; used to define genetic populations; in Y-DNA testing, mostly defined by SNPs; more loosely, a cluster of similar haplotypes

**Haplotype** – the complete set of results from multiple sites tested on a chromosome inherited from one parent (e.g., the Y or mtDNA); in Y-DNA testing, expressed as series of numbers (each one representing the allele at a specific STR marker), which are compared to others' haplotypes for indications of relationship; in mtDNA testing, expressed as differences from the Cambridge Reference Sequence

**Homozygous** – when chromosomes have two copies of the same allele at a given gene locus; occurs most often when an allele is common in the general population; contrasting word is heterozygous

**Junk DNA** – non-coding (i.e., not used for making proteins) stretches of DNA with no known function, which represent an estimated 95 percent of our DNA; because it does not affect traits or medical conditions, it acts as a silent recorder, accumulating mutations and preserving one's ancestral history; used by population geneticists to study the migrations of ancient peoples and by genealogists to learn about their origins

**Karyotype** – a picture of all the chromosomes of an organism, often arranged in homologous pairs according to decreasing size; frequently used in genetic testing

**Locus** – specific location on a chromosome; position where a particular marker is located

**Marker** – a distinctive landmark that occurs in an otherwise feature-less stretch of DNA; a DNA sequence with known genetic characteristics that can be tested for purposes of comparison (e.g., SNP, STR)

**Mitochondria** – my-toe-CON-dree-uh, the plural form of mitochondrion; plentiful organelles in the cytoplasm of cells that provide energy for the cells; see *mitochondrial DNA*

**Mitochondrial DNA (mtDNA)** – genetic material found in mitochondria; passed from mothers to their children, but only daughters are able to pass it on; useful to genealogists for learning about their maternal roots; also valuable for the identification of degraded remains

**Most Recent Common Ancestor (MRCA)** – the shared ancestor of two or more people who represents their closest (and therefore, most recent) link; for instance, the MRCA of a pair of second cousins as their mutual great-grandfather or great-grandmother

**Non-paternity event** – catch-all term for situations where the Y chromosome is unlinked from the surname; includes informal and casual adoption, infidelity, illegitimacy, etc.

**Non-Recombining Y (NRY)** – that portion of the Y chromosome that is passed essentially unchanged (except for occasional mutations) from father to son down through the generations; all Y-DNA tests for genealogy use markers on the NRY

**Nucleotide** – a base plus a phosphate group; the building blocks of DNA; see *bases*

**Nucleus** – the central region of the cell that houses the chromosomes and is separated by a membrane from the cytoplasm

**Petering out** – when an mtDNA line has died out because only sons (who are now deceased) were born; see *daughtering out* for reverse

**Phenotype** – observable traits of an organism (e.g., hair color); may or may not be genetically related

**Polymerase Chain Reaction (PCR)** – technique developed by Kary B. Mullis to mimic the replication process of the cell, allowing scientists to efficiently amplify (i.e., make millions of copies of) small, selected segments of DNA; sometimes referred to as molecular photocopying or Xeroxing; used in genealogy to amplify samples submitted for analysis; see *DNA Polymerase*

**Polymorphisms** – (from the work roots poly = many and morph = forms) inherited differences in genetic markers among individuals and populations that play a key role in genealogy; see *mutation and polymorphic*

**Recessive** – pertains to alleles (and corresponding traits) that will only manifest when the dominant allele/trait is absent; recipient must have two copies of the recessive allele (one from each parent) for the trait to be evident; see *dominant* for contrast

**Restriction Fragment Length Polymorphisms (RFLP)** – (pronounced riff-lips), collection of DNA fragments produced when DNA is cut (at designated sequences) with restriction enzymes; mutation in the slowly changing coding region of mtDNA result in different fragment lengths in people, and as detected by RFLP testing, constitute the formal definitions of mtDNA haplogroups

**Sequence** – the order of nucleotide bases in a DNA molecule (e.g., AGCTTTACGGA) that encodes for proteins; the sequence of the human genome is 3 billion DNA bases

**Sex chromosomes** – the X and Y chromosomes; the Y is involved in sex determination and is the basis for most present genealogical testing

**Sex-linked** – refers to traits and diseases that derive from genes on the X chromosome, such as color blindness

**Short Tandem Repeat (STR)** – a *short* pattern (often two to five bases in length) *repeated* a number of times in a row (in *tandem*); for instance, GATAGATAGATA, three repeats of the GATA sequence; the differences in the STRs at selected markers on the Y chromosome provide a bases for comparison among individuals and populations and are used extensively for most Y-DNA genealogical testing; also called a microsatellite

**Silent mutation** – a mutation that is not expressed and has no effect on the phenotype of an organism; such mutations accumulate over time because they do not affect survival but can provide useful means of discriminating between people and populations

**Single Nucleotide Polymorphism (SNP)** – (pronounced “snip”), a small genetic change or variation that occurs within a DNA sequence when a single nucleotide, such as an A, replaces one of the other three nucleotide letters: C,G, or T; occur so infrequently that they are used to define haplogroups

**Unique Event Polymorphism (UEP)** – a class of mutations in which the mutation rate is so low that it can be considered a one-time event or essentially unique-for instance, SNPs and indels; useful for exploring deep ancestry

**Y-DNA** – a genetic material found in the Y chromosome; passed from fathers to their sons essentially unaltered down through the generations except for occasional mutations; used for tests designed to explore one’s paternal ancestry

### **Some useful Websites:**

Family Tree DNA [www.familytreedna.com](http://www.familytreedna.com)

Relative Genetics [www.relativegenetics.com](http://www.relativegenetics.com)

National Geographic/IBM Genographic project

<https://www3.nationalgeographic.com/genographic/index.html>