Background information on testing of the Y chromosome

The science of genetics and DNA research has come a long way since Watson and Crick first discovered DNA (deoxyribonucleic acid) over 50 years ago. DNA is stored in the nucleus of each cell and controls the cell and its interactions with other cells in the body. DNA is composed of long strands of 4 different subunits called bases that are then paired to each other. The four different bases are cytosine (C), adenine (A), thymine (T), and guanine (G). The DNA in humans is organized in 46 chromosomes, of which 44 are autosomal chromosomes and 2 are sex chromosomes, the X and Y chromosomes. Males have one Y chromosome and one X chromosome. Females have two X chromosomes. The Y chromosome is passed down from each father to his sons. If a father has a daughter he passes on his X chromosome to his daughters. Mothers pass one of their X chromosomes on to both their daughters and their sons.

Since about 95% of our DNA doesn't code for any known function it has been subject to random mutations over the millennia. These mutations are potentially helpful to genealogists if they are carefully studied. The mutations generally occur in two forms: single nuclear polymorphisms (SNPs) and short tandem repeats (STRs). SNPs are focal single base pair mutations where one base is substituted for another at a particular location on a chromosome. STRs are longer DNA segments that consist of repeated 3 to 6-base pair sequences in which an additional sequence is occasionally either added or deleted.

An example of an STR composed of a 4-base pair sequence that is repeated five times is as follows: CATACATACATACATACATA. In this example each letter is an abbreviation for the base that is in that segment of DNA. The number of times that any one sequence is repeated (in this case CATA) is the value for that particular STR marker. In this example, the value for the STR marker would be 5. In general, the values for any particular marker can vary, but are usually between 5 and 35. There are usually between 5 and 10 different values that are possible for any particular marker.

There are hundreds, if not thousands, of STR markers on each chromosome. As these markers are identified they are given a name by geneticists, such as DYS390. If one tests multiple different markers on a chromosome or chromosomes (usually the Y chromosome), the resulting set of values for each group of markers is called a haplotype. Geneticists have discovered that the Y chromosomes of all men in world can arranged in 18 different groupings known as haplogroups. Each haplogroup is defined by whether or not the Y chromosomes of the men in that haplogroup have a particular set of SNPs and has been named by a specific letter of the alphabet (A to R).

As a general rule SNPs occur relatively infrequently, usually only about one time per every 50,000,000 bases that are replicated. However, since there are about 3.2 billion base pairs in the human genome and about 60 million base pairs in the Y chromosome, on the average about 60 SNPs should occur in each new child that is born that don't appear in their parents and about one SNP should be found in the Y chromosome of each boy that is born that his father's Y chromosome doesn't have.

Mutations occur much more frequently in STR markers than SNPs occur. On the average, there is about a 0.4% chance that any given STR marker will mutate in any given generation. When mutations occur the value of the marker usually increases by one or decreases by one depending on whether an extra copy of the repeated sequence was inserted into the DNA or whether one of the repeated sequences was deleted. On rare occasions a mutation can result in

an STR marker's value increasing or decreasing by two or more. As a general rule, the higher any particular marker's value, the more it is prone to mutate.

In the past 6 years DNA analysis has been increasingly used by genealogists as a tool to answer genealogical questions or to confirm assumed relationships. Up to this point most research has focused on the Y chromosome, which is only found in males and is passed down from father to son. If two males share a recent common ancestor then their Y chromosomes should be identical, or at least nearly identical. The more dissimilar any two males' Y chromosomes, the more distantly related they are. Genealogists have been analyzing the DNA samples from many men with the same surname to see if they have share a common recent ancestor.

Available genetic evidence indicates that all men descend from a common ancestor who likely lived in Africa thousands of years ago. There have been multiple mutations in the Y chromosomes of that man's male descendents as they dispersed throughout the world over the years. Those mutations have been in the form of both STR mutations and SNPs. As geneticists have studied these variations in the Y chromosome over the past 20 years or so they have categorized the Y chromosome results into haplogroups. Each haplogroup is defined by a specific SNP. Within each haplogroup various subclades have been discovered. Each subclade is also defined by a specific SNP. The Y chromosome haplogroup tree is frequently being updated as more and more SNPs are discovered. The most up-to-date Y chromosome haplogroup tree may be found at http://www.isogg.org/tree/index.html. The Genographic Project web site at https://www3.nationalgeographic.com/genographic/index.html has an excellent discussion of the various Y chromosome haplogroups and is well worth exploring. The most common haplogroup among men of European ancestry is haplogroup R1b. Other haplogroups that are common in Europe include R1a, I, E3b, and J.

Once one has obtained their Y chromosome STR marker results, one generally will want to know which haplogroup they are in. Many of the commercial DNA testing companies offer SNP testing that will determine one's Y chromosome haplogroup with certainty. STR marker results can be used to predict one's Y chromosome haplogroup with relatively high accuracy. The best program available to predict one's Y chromosome haplogroup using STR marker results is Whit Athey's program, which may be found at https://home.comcast.net/~hapest5/index.html. This program is good at predicting the haplogroup, but can't accurately predict which subclade within a haplogroup to which one may belong. SNP testing is required to determine the subclades.

A number of programs are available for analyzing STR data. The best available free program is Dean McGee's Y-DNA comparison utility program, which may be found at http://www.mymcgee.com/tools/yutility.html. This program can be used to determine the genetic distance between groups of Y chromosome haplotypes and can also be used to predict the time to the most recent common ancestor (TMRCA) for any two haplotypes. It can also be used to generate Fluxus .ych files, which is helpfully for creating diagrams of haplogroup data.

The following are some additional web sites besides those mentioned above and those mentioned on the home page that discuss Y chromosome testing and genetic genealogy in general in greater detail:

1. International Society of Genetic Genealogists: http://www.isogg.org
2. World Families.net web site: http://worldfamilies.net
3. Charles Kerchner's web site: http://www.kerchner.com/dna-info.htm
4. Ron Scott's web site: http://freepages.genealogy.rootsweb.com/~ncscotts

5. John McEwan's web site: http://www.geocities.com/mcewanjc
6. Ken Nordtvedt's web site: http://knordtvedt.home.bresnan.net
7. Short tandem repeats database: http://www.cstl.nist.gov/biotech/strbase/index.htm

The following are DNA testing companies and foundations:

1. Sorenson Molecular Genealogy Foundation: http://smgf.org (offers free testing)
2. Family Tree DNA: www.familytreedna.com
3. National Geographic Society Genographic Project: https://www3.nationalgeographic.com/genographic
4. Ancestry.com: http://www.dna.ancestry.com
5. RelativeGenetics: www.relativegenetics.com (now merged with Ancestry.com)
6. EthnoAncestry: http://www.ethnoancestry.com
7. Oxford Ancestors: www.oxfordancestors.com
8. DNA Heritage: http://www.dnaheritage.com
9. Gene Tree: http://www.genetree.com

The following are various DNA e-mail lists that one can join if they are so inclined:

1. RootsWeb Genealogy-DNA list http://archiver.rootsweb.com/th/index/GENEALOGY-DNA; most active list with an average of about 2000 messages per month
2. FTDNA http://www.familytreedna.com/forum; very active
3. ISOGG Newbies; about 350 messages per month http://groups.yahoo.com/group/DNA-NEWBIE
4. Genealogy.com DNA forum; http://genforum.genealogy.com/dna; about 50 messages per month
5. World Family Network www.wfnforum.net; minimally active

The following are informative books about DNA testing

1. Trace Your Roots with DNA by Megan Smolenyak and Ann Turner, 2004.
2. DNA and Family History: How Genetic Testing Can Advance Your Genealogical Research by Chris Pomery, 2004.
3. Unlocking Your Genetic History by Thomas Shawker, 2004.
4. The History and Geography of Human Genes by L. Luca Cavalli-Sforza, Paola Menozzi, and Alberto Piazza, 1994.
5. The Journey of Man by Spencer Wells, 2002.
6. Deep Ancestry by Spencer Wells, 2006.
7. The Seven Daughters of Eve by Bryan Sykes, 2001.

Tim Janzen
E-mail: tjanzen@comcast.net
Mennonite genealogy resources web site: www.timjanzen.com