SQUIGGLY LINES AND SPIT

HOW DNA IS CHANGING FAMILY HISTORY RESEARCH

Dave Obee
Dave@CanGenealogy.com

DNA has become one of the most important tools in genealogical research, thanks to the availability of affordable genetic genealogy tests based on our saliva. These tests will usually reveal hundreds of matches – people who have a few centiMorgans of DNA that are the same as yours. Depending on the size of the match, these people could be siblings, cousins, children or parents; a relative, of some sort.

Genetic genealogy is still a young science. The first mass-market DNA test, only for male-line research, was offered by FamilyTreeDNA in 2000, but its high cost deterred many genealogists. In 2007, 23andMe was the first company to offer autosomal DNA tests. The largest DNA database today is that of Ancestry. In 2012, Ancestry started selling a mass-market DNA test in the United States, and expanded it to Canada, the United Kingdom, Australia and New Zealand in 2015.

But what is DNA? Deoxyribonucleic acid – better known as DNA – is a fancy name for a molecule composed of two chains that coil around each other to form a double helix carrying the genetic instructions used in the growth, development, functioning, and reproduction of all known living organisms and many viruses. There you go.

There are three kinds of tests:

• **Autosomal**, the most common test, includes all your ancestral lines, with a limit of perhaps three to six generations back. Males and females can do this test.

• **Y-DNA** tests the male-male-male line, and can go much further back, but it is difficult to know where the connection might be. Only males can do this test.

• **mtDNA** tests the female-female-female line, and goes many generations back. It is the least useful of the tests. It can be done by males and females.

You received your DNA from your parents, half from your mother and half from your father. They received their DNA from their parents, and so on. The halves we received from each parent were random, and they were random for our siblings as well. You will be a close match to your siblings, but not an exact match. Your parents did not match their siblings either – and that will have an impact on the matches of their descendants. You are a closer match to your first cousins than to your second cousins, and they in turn are closer than your third cousins.

But wait! Thanks to the random nature of DNA, it is quite possible that bits have been passed down more or less intact for two or three generations. You got 50 per cent from your father, but not necessarily 25 per cent from your father's father. That is because of the random nature of DNA; that 50 per cent from your father could have included 90 per cent from his mother and only 10 per cent from his father.

DNA test results on their own will have little value; these results must be used in combination with traditional genealogical research. Our DNA by itself will not reveal much, other than a rough estimate of our ethnic origins, a guesstimate based on what other testers have reported. Ethnicity results are basically for entertainment, not research.

Matches are far more valuable. You will make the most genealogical progress determining how your matches fit into your family.

Companies offering DNA tests

AncestryDNA – Autosomal; 27 million people in the database. Easy to use and offers links to trees. Does not allow uploads from other sites.

23andMe – Autosomal; 15 million people in the database. Shows whether matches relate to each other. No links to trees and no uploads. Also offers health tests.

FamilyTreeDNA – Autosomal, Y-DNA and mtDNA matching; six million people in the database. Has a chromosome browser and can upload from other sites.

MyHeritage – Autosomal; 6.8 million people in the database. Most popular in Europe. Has nice matching tools and allows uploads.

LivingDNA – Autosomal, Y-DNA and mtDNA; 300,000 people in the database. British focus. Allows uploads. No tree provision. Partners with FindMyPast.

Results from different companies can be uploaded and compared on other sites. The most popular is GEDmatch, which has 1.4 million members.

Making the most of your DNA matches

Paired with regular genealogical research, your DNA matches can solve mysteries, blast through brick walls, confirm research, and help us find more relatives than we ever thought possible. To get the best results, build out your tree, including collateral lines.

Link your tree to your DNA test so others can get a sense of how you might tie in. You do not need to use your entire tree to do this; a small one, just for DNA match purposes would be enough, and would help your matches save time.

Test with as many companies as possible. The larger the pool of matches, the more results you are likely to have. Two companies – Ancestry and 23andme – do not allow uploads from other companies, so start by testing at one of them. Then, download your raw data from that test, and upload it to MyHeritage, FTDNA, and Living DNA, as well as GEDmatch, which brings together results from all companies.

Match lists might show close relatives and more distant ones. It is good to sort out how people are connected to you, and it is easier with close matches. Consider getting other people to test. Your siblings will get a different random selection from your parents, so test them as well. Full siblings will have matches that you do not have, and you will have some they do not have. First cousins will help you narrow down the ancestral line for a match. If you have a relative who is of your parents' generation, get them to test; their DNA will be valuable to you.

When it comes to pushing back your lineage, more distant cousins will be needed – matches with whom you share 100 centiMorgans or less. Odds are, you will find plenty on the major DNA sites, and with many it will be impossible to determine the exact connection. Check as many of these matches as you can. It might mean researching dozens of families, but the reward can be huge.

It is possible to find common matches in the 1700s, further back than we believed a few years ago. The thinking had been that the basic autosomal tests could not be trusted when dealing with ancestors more than four or five generations ago; that a match of only a small number of centimorgans – 10 cM or less – might be a false match, or too distant to even be proven. But as more people have tested, the limits have been pushed. It is only a matter of time before someone finds common matches in the 1600s.

The most reliable matches are the ones with more centiMorgans in common, and larger segments in common. Small numbers might indicate false matches. To get a better sense of potential relationships in centimorgan ranges, use the Shared cM Project Relationship Chart: https://thegeneticgenealogist.com/wp-content/uploads/2020/03/Shared-cM-Project-Relationship-Chart.png

Most companies allow you to make notes on your matches. These are highly recommended so you do not research the same matches over and over. It is also good to save these notes offline as well, because sometimes people delete their tests.

Use searches and filters to zero in on matches of interest. Search to see which names each match might have. Searches on Ancestry will get results from linked trees, even if they are private. If a match does not have a tree, check for matches in common to see if you can at least determine which ancestral line the person is on. Or, if the name is uncommon, you could do basic research into the person's ancestry.

Building trees for matches is often necessary. Online trees are notoriously inaccurate, so do not rely on what they have posted. Also, be careful with hints – just because Ancestry keeps suggesting something does not mean it is correct.

Organize your matches into clusters, using tools available on each site. Ancestry and MyHeritage, for example, have colour coding options. Starting with your closest matches, mark the different ancestral lines. Be warned that a match might appear on several different lines, and might show up as a common match with a cousin on your maternal side and with a cousin on your paternal side.

Create clusters of your unknowns, the people who all seem to be connected but you cannot place. Start with your top unknown, the one with the highest match numbers, and cluster them with common matches. Then take your second highest unknown, other than the first cluster of unknowns, and create another group. Then a third, then a fourth. This method is sometimes called the ABCD method, sometimes the Leeds method, and sometimes it is just called common sense. Looking at all the matches in one of these clusters might provide clues that could help you identify the connection to you.

Try to contact your matches if they are of interest – especially your strong matches that you cannot place in your tree. That would allow you to collaborate with them, and perhaps add valuable information to your research. A few people will respond right away, but others will be slower, and (odds are) more than half will never respond. Do not despair, just do all that you can to research the family yourself. Armed with more information, you might have better luck in making contact. Finally, if you make contact, switch to email as soon as possible. It is more lasting than messages sent on the DNA test company sites.

There are many ways to sort through your match information. Chromosome browsers, available on some test sites, are popular because they can often pinpoint connections more easily. Other sites, such as DNApainter, might also help. Clustering tools provided by the test companies can help, and some researchers say the best idea is to put match information onto spreadsheets. Different tools and strategies work for different people, so do not feel obligated to use a system that others recommend.

DNA matches can be useful for adoptees to find birth parents, and for birth parents to find their children. But there is a flip side to this: Sometimes there are surprises. As an example, people who do a test might find that their father is not a match. Some people assume the DNA test is flawed and want to do another test, but what they need to do is talk with their mother.

Some sites, including 23andMe, offer health insights based on DNA results, but approach those with caution.

Genealogists are not the only people using DNA testing to solve mysteries. Genetic matches are also helping police services solve cold cases, including murders, assaults and rapes. This is not without controversy, and there are questions about personal privacy being violated. The results obtained by police should inspire genealogists – because in the end, they are looking for cousins, the same thing we are doing.

Types of DNA tests

Autosomal tests look at Chromosomes 1-22 and X. Chromosomes 1-22, the autosomes, are inherited equally from both parents and roughly equally from grandparents to about great-great-great-grandparents. Inheritance is more random and unequal from more distant ancestors. The major benefit of an autosomal test is matching other individuals. The companies offer estimates to show the relationships to these matches, but because of the random nature of DNA inheritance, estimates are only approximate, and have a wider margin of error with every generation back.

The **X-chromosome** follows a special inheritance pattern; it cannot be passed from one male to another. The X-chromosome can be used to narrow down possible ancestor lines – for example, an X-chromosome match with a male can only have come from his maternal side. FTDNA and 23andme show X matches. More information is on Blaine Bettinger's TheGeneticGenealogist website.

Y-DNA tests look at the father's father's father's line, or the direct patrilineal lineage. The Y-Chromosome is only present in males, and is passed almost unchanged from one male to another. Women looking for their direct paternal DNA ancestry need to ask a relative on a male line to test. A Y-DNA test can be like looking for the proverbial needle in a haystack; most of your matches will not be related within a genealogically useful time frame. Identify likely matches and ask them to test, then compare your results with them. This test can be used to find a person's haplogroup, which is often associated with an ethnic group or region.

mtDNA tests, or mitochondrial DNA, look at the mother's mother's mother's line. A mother passes mtDNA to a child of either gender, so males and females can take this test. Identifying a relative can be difficult because of name changes in every generation and because there can be little change over many generations. A perfect match to another person's mtDNA test results indicates shared ancestry that could be one to fifty generations ago, which is of little value. Use this test to confirm a match, rather than to find one. This test can help determine a person's haplogroup.

Definitions (the boring science stuff)

Centimorgan (cM) is the unit for segments of DNA. A full human genome is about 6500 cM. Most companies will show customers how many cMs they share, and across how many segments. A person will share about 3200 to 3700 cM with a parent; 1300 to 2400 cM with a grandparent; 500 to 1300 with a first cousin; 50 to 500 cM with a second cousin; and no more than 80 cM with a third cousin. A match of 7 cM could lead a researcher to a common ancestral match in the 1700s, so do not ignore the small ones!

SNP (for single-nucleotide polymorphism) is a change at a single point in genetic code. A DNA test might test about 700,000 SNPs. The more SNPs in common between two individuals, the higher the likelihood they share a segment of DNA. That makes it more likely that they match. Typical Y-DNA SNP tests look at about 20,000 to 35,000

SNPs. A SNP test allows a much higher resolution than an STR test. A match should have 700 SNPs or more.

STR (for short tandem repeat) markers represent the number of times a certain section of DNA repeats. The results of two individuals are then compared to see if there is a match. STRs mutate fairly frequently, and typical tests examine between 30 and 120 STR markers. DNA companies will usually provide estimates of how closely related two matches are, based on the difference between their results. The test thresholds vary by company.

MRCA, or most recent common ancestor. The great (times x) grandparent you share with another person tested.

A **haplogroup** indicates a single line of descent, usually dating back thousands of years. Haplogroups are normally identified by an initial letter of the alphabet, and refinements consist of additional number and letter combinations. A haplogroup is made up of a group of haplotypes; these are groups of genes that are inherited as a whole from a single parent. Haplogroups are usually determined through Y-DNA and mtDNA tests.

All human men descend in the paternal line from a single man, Y-chromosomal Adam, who lived (best guess) 200,000 to 400,000 years ago. Different branches of this tree are different haplogroups. Most haplogroups can be further subdivided multiple times into groups called sub-clades.

For more information and more inspiration

The best book: Blaine T. Bettinger's The Family Tree Guide to DNA Testing and Genetic Genealogy (second edition)

International Society of Genetic Genealogy (ISOGG): https://isogg.org

The Power of Genetic Genealogy with Diahann Southard (from the International Symposium on Human Identification): https://www.youtube.com/watch?v=qshi9qORPF0

The Gene Hunters: How genetic genealogy is helping police solve 37-year-old cold cases (from The Fifth Estate): https://www.youtube.com/watch?v=ggYz5VlCld0

Resolving a Case of Unknown Identity Through Genetic Genealogy (from the International Symposium on Human Identification): https://www.youtube.com/watch?v=32SEh8VO--k