

Genetic Genealogy Journey

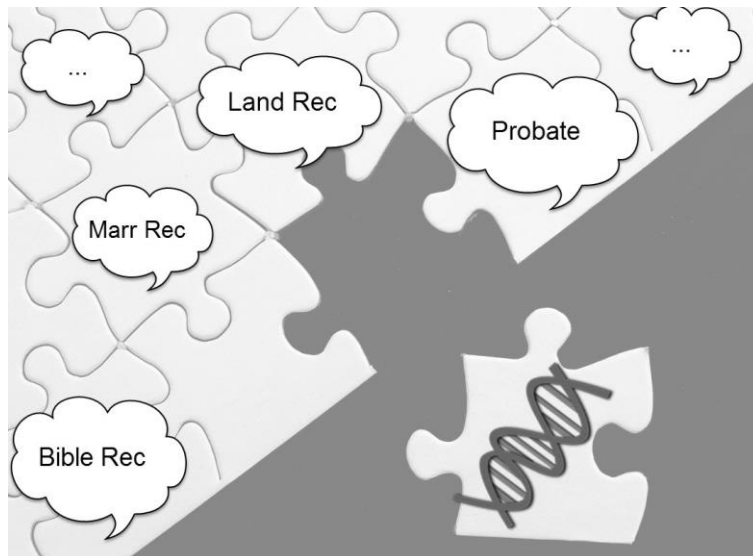
What We Don't Know

by Debbie Parker Wayne, CGSM, CGLSM

Bertrand Russell said, “[W]e can prove generally that we know what we know. But it remains conceivable that we don't know what we don't know.”¹

As more of us study genetics and publish our findings, our knowledge of DNA and genealogy grows. Once we thought Neanderthals never mated with modern humans; now we believe some modern humans carry some Neanderthal DNA.² Once we thought much of our DNA was junk with no purpose; now we believe that DNA contains switches and controls for the genes.³ We are beginning to create faces from nothing but DNA.⁴ Before 2009, many scientists thought only Y-DNA and mtDNA could be used effectively for genealogy. Advances in laboratory processes and DNA analysis techniques opened new worlds for genealogical uses of atDNA. We cannot predict what amazing new discoveries we will make tomorrow. Those discoveries could affect how we interpret DNA results for medical and legal uses and for genetic genealogy.

Our conclusions should consider the accuracy and completeness of our genealogy, our genetic knowledge level, and the current understanding of how to apply genetics to genealogy.



All URLs accessed 2 November 2015.

¹ Bertrand Russell, *The Principles of Mathematics* (Cambridge, England: Cambridge University Press, 1903):355; electronic copy, *Internet Archive* (<https://archive.org/details/cu31924001610421>).

² “Why Am I Neanderthal?,” The Genographic Project, *National Geographic* (<https://genographic.nationalgeographic.com/neanderthal/>).

³ Stephen S. Hall, “Hidden treasures in Junk DNA,” 18 September 2012, *Scientific American* (<http://www.scientificamerican.com/article/hidden-treasures-in-junk-dna/>).

⁴ Peter Aldhous, “Genetic mugshot recreates faces from nothing but DNA,” 20 March 2014, *New Scientist* (<https://www.newscientist.com/article/mg22129613-600-genetic-mugshot-recreates-faces-from-nothing-but-dna/>).

Our Genealogy

We want our family trees to extend as far back as possible, to depict our ancestral lines, and not to include someone else's ancestors. Accurately identifying all of our ancestors back six, eight, or more generations, and as many of their descendants as possible, will help us in our analysis of DNA matches. Deeper trees help us determine whether we may have more than one ancestor in common with a particular person who has atDNA that matches ours. Sharing multiple ancestors will have a big impact on our interpretation of atDNA evidence.

Our Genetic Knowledge

Genetic genealogy is still new to many of us. A depth of knowledge allows us to apply the DNA evidence accurately. Even those of us who have been involved with genetic genealogy for ten or fifteen years must study constantly to keep up with new developments.

A researcher who is new to using DNA may not realize that only men have Y-DNA. If my brother takes a Y-DNA test and I, as a woman with no Y-DNA, take an atDNA test, my brother will not show up on my list of atDNA matches. It does not mean he is not my brother. It means that I do not share Y-DNA with him. If we both take an mtDNA and/or atDNA test, we can expect to be on each other's list of matches.

We will have DNA matches on Y-DNA and mitochondrial DNA (mtDNA) that go much further back than six or eight generations to a common ancestor. Because Y-DNA and mtDNA do not recombine in the way atDNA does, we can sometimes be more sure of our interpretation of the Y-DNA and mtDNA evidence.

To begin using DNA evidence correctly, we must understand basic inheritance patterns. We must keep abreast of changes in minimum recommendations for number of markers to test and how to interpret the DNA results as applied to our family tree.

We will usually share more autosomal DNA (atDNA) with cousins with whom we share a more recent common ancestor. Recombination of atDNA as it passes from parent to child usually reduces the amount of DNA we may share with a cousin. In some situations, we may share detectable amounts of atDNA with cousins where the common ancestor is further back in time. Moreover, in some cases, we may have a second unknown common ancestor making it look as if we share more DNA than predicted with a cousin.

Applying Genetics to Genealogy

Proper application of genetics to genealogy requires an understanding of the limits of DNA. We are still in the exploration phase, pushing to discover the boundaries for the application of genetic genealogy. It is normal for there to be differences of opinion during the discovery phase of any new technology. We should keep an open mind. What we don't know we don't know might change our genealogical conclusion once we learn more.

A lot of statistical data is available about DNA matches. How much DNA might be shared by third cousins, **as predicted by statistical averages?** What percentage of third cousins will actually share a detectable amount of DNA, **as predicted by statistical averages?**

What we cannot forget as we answer these questions is that random recombination and random mutations in DNA make the data for many specific families fall outside of those statistical averages. Until we have more published data from in-depth family studies, our conclusions are not as credible as they will be when we have more data. **More data** is almost always better in genetic genealogy.

Blaine Bettinger's "Shared cM Project" is one of the few published studies indicating how much DNA two testers with a known relationship share. The data was self-reported so there might be errors in the relationships or amount of shared DNA indicated. Real numbers from real people demonstrate an eye-opening variation in the amount of shared DNA in many relationships.⁵ Some of the amounts are far outside of the statistical average predicted. There are statistical averages and there is real life. More real life data will help us define real boundaries during this discovery phase of genetic genealogy.

⁵ Blaine Bettinger, "The Shared cM Project – An Update," 25 May 2015, *The Genetic Genealogist* blog (<http://thegeneticgenealogist.com/2015/05/25/the-shared-cm-project-an-update/>).

Relationship	#	Min	Median	Average	Max
1C	662	82.65	884.81	880.81	1559.00
1C1R	606	54.18	434	439.74	902.59
1C2R	134	26.90	244.76	239.72	412.74
2C	598	47	235.78	245.98	760.10
2C1R	590	0	131.22	139.66	465.99
2C2R	101	0	84.96	94.04	574.60
3C	517	0	83.41	89.40	334.06
3C1R	264	0	67.92	76.03	332.26
3C2R	36	16.4	53.41	51.90	121.66
Aunt/Uncle	722	121.34	1730.26	1703.44	2226.6
Aunt/Nephew	106	847	1724.14	1708.33	2175
Aunt/Niece	137	684.62	1722.3	1709.52	2226.6
Uncle/Nephew	67	739.71	1743.88	1723.71	2038.52
Uncle/Niece	101	892.5	1742	1704.56	2004
Aunt/Uncle	311	121.34	1730.13	1694.38	2192
Grandparent/Grandchild	195	875	1757.84	1759.54	2365.1
Great Aunt/Uncle	123	235.71	858.12	843.84	1301.1
Great Grandparent/Grandchild	25	547	969.64	855.66	1110.02

Blaine Bettinger, "The Shared cM Project - An Update," 25 May 2015, *The Genetic Genealogist* blog (<http://thegeneticgenealogist.com/2015/05/25/the-shared-cm-project-an-update/>). See post for complete chart. Shared under Attribution-NonCommercial-ShareAlike CC license.

Random recombination results in each sibling inheriting some of the same DNA and some different segments of DNA from a parent. Because of this, we must be careful how we interpret DNA matches. We must be even more careful when we have fewer tested relatives we can compare. More data is better.

How do we get more data? We can wait for other descendants of our ancestors to take a DNA test then work with them to identify our common ancestor. We can ask targeted relatives to take a DNA test to help us determine which ancestral line the unknown cousins link to. Targeted testing can give us answers more quickly than just waiting for the right cousin to show up on our match list.

When analyzing our DNA match list, we must remember a threshold level determines who is on the list. If we are looking at a match list at a DNA testing company the company selects the threshold used to determine if two testers are related. Each testing company uses a different threshold. If we are looking at a list provided by a third-party tool, we may have the option to change the default threshold values.

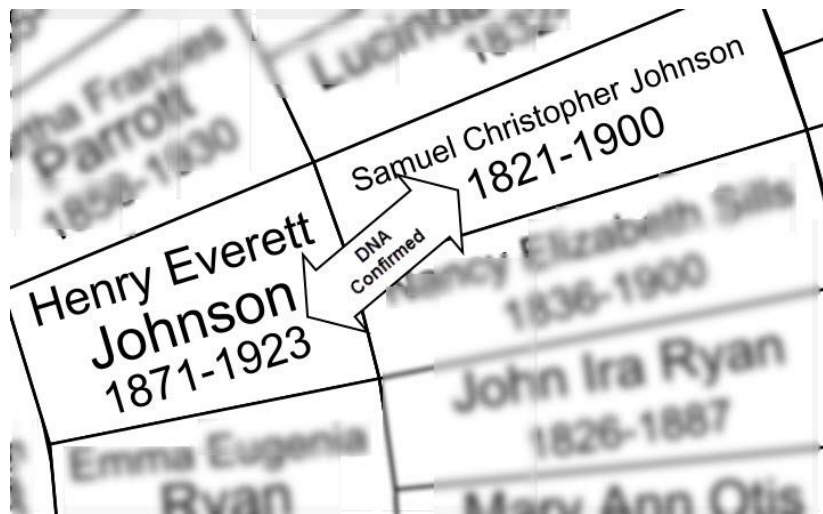
If the thresholds are set too low, there will be false positives in our list of matches—people in our match list with which we do not share a recent common ancestor that we can document. If the thresholds are set too high, there will be false negatives—people who are not in our match

list with which we share a documentable recent common ancestor. Random recombination results in each person inheriting differing amounts of DNA from any specific common ancestor. There will be real cousins who may not show up on our list of matches no matter how low we set the threshold, but we must use a reasonable threshold for the total amount of DNA we share with a person and for the size of any individual DNA segment. We must consider all possibilities as we correlate DNA and documentary evidence to reach a conclusion.

With all of this randomness, is genetic genealogy useful? As with many things in genealogy, the answer is, “Yes, but.” DNA is one more tool in our toolbox; it does not stand alone. We correlate DNA evidence with our documentary evidence and use all of the evidence to try to answer a genealogical question. There are times when we need more DNA evidence and/or more documentary evidence to reach a credible conclusion. Let’s look at some examples.

I first ordered atDNA testing only for one full brother, one half-sister, and myself. Several acquaintances showed only as matches to my half-sister. My natural assumption was that they were related to my sister through her paternal line that is different from the paternal line I share with my full siblings. However, after I tested additional full siblings, some of my full siblings also match the people that were only matches to my half-sister. In this case, my full brother and I did not inherit some segments from our common mother that our other full siblings and our half-sister did inherit. Without testing the additional siblings, I did not have enough data to reach the right conclusion.

We might have three cousins on our DNA match list who all trace their lineage to one common ancestor named Samuel. If those three cousins all copied the family tree we posted online a decade ago and they did no verification of that tree, the DNA match does not prove a link to Samuel. Shared DNA proves we are all descended from a common ancestor, but if there is an error in the tree, we have the wrong ancestor. If all three cousins have well-researched trees, created independently, and all include Samuel, the link to Samuel is more credible.



Sharing another unidentified common ancestor could change our conclusion. Confirming that the DNA match list of each of those cousins includes all of the other cousins adds more credibility.

If each of the cousins shares DNA “in common with” all of the others that provides stronger evidence of one common ancestor, but not indisputable proof. If any of the cousins have trees where some ancestral lines do not go back many generations, the common ancestor might not be Samuel but some ancestor not yet identified by the researcher.

If some, but not all, of the cousins share DNA “in common with” others we may still have a credible genetic network of cousins. Random recombination means we may not all have inherited the same DNA segment from any specific common ancestor. Two sets of “in common with” matches could mean each set of cousins inherited a different DNA segment from the common ancestor. It could also mean we are related to one set of cousins through our paternal line and the other set through our maternal line.

If a significant number of the tested cousins have well-researched, independent trees going back many generations, include Samuel as an ancestor and have no other common ancestors, and share an overlapping segment of DNA of significant size on the same chromosome,⁶ we have a triangulated group. This adds more credibility to our DNA evidence, but it is important to know if there are any other common ancestors besides Samuel from whom we may have inherited this DNA segment.

Conclusion

More data is better. Complete family trees and many tested cousins with whom we can compare DNA leads to more credible conclusions about kinship links.

Accurately-researched family trees with deep lines and DNA results from multiple testers are both essential to a credible genealogical conclusion as we search further back in time.

Debbie Parker Wayne, CG, CGL, is experienced using DNA analysis, as well as more traditional techniques, for genealogical research in Texas, the South and West. She coordinates the genetic genealogy courses at several genealogy institutes and is the Texas State Genealogical Society’s DNA Project Director. See <http://debbiewayne.com/> for more information.

⁶ Debbie Parker Wayne, "Logging and Analyzing atDNA Matches," National Genealogical Society *NGS Magazine* 41 (January-March 2015): 58-61; linked online at (http://debbiewayne.com/presentations/gatagacc_biblio.php). See discussion of segment sizes.