

Essential Considerations for DNA Evidence

Blaine T. Bettinger, Ph.D., J.D.

blainebettinger@gmail.com / www.DNA-Central.com

Does the DNA Evidence Actually Support Your Conclusion?

Whether DNA evidence *supports* a hypothesis or conclusion is a difficult question, requiring an in-depth analysis of the many specific issues that arise when utilizing DNA. Indeed, EVERY type of evidence involves its own issues that must be considered. However, genealogists are still elucidating the metes and bounds of the application of DNA evidence to traditional genealogical research. Together we'll look at just a few of the issues that can arise and must be considered before you declare a genealogical conclusion.

1. Confirmation Bias

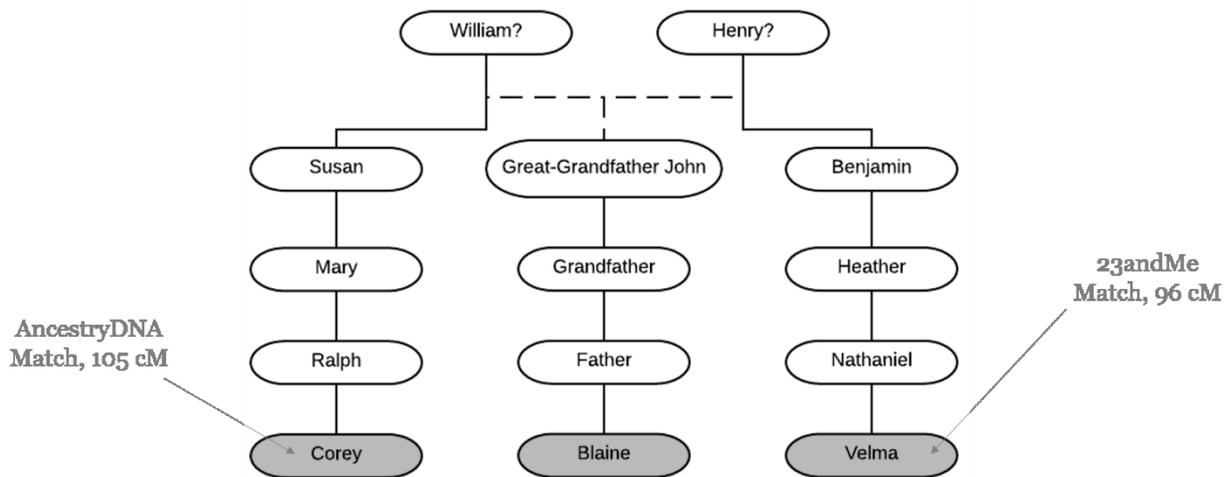
There is an extreme danger of confirmation bias in genealogy. However, this is an area where genealogy can learn from and borrow from science. The goal of science and the scientific method is NOT to confirm a theory or hypothesis; it is to **reject** a theory or hypothesis! Are you gathering evidence to support or reject your hypotheses?

Psychologists studying human behavior have demonstrated that humans have *an innate tendency to confirm existing beliefs*, rather than questioning them or seeking new ones. This makes sense from an evolutionary standpoint. Although the human brain is the most powerful and dynamic evidence analysis machine in existence, analyzing evidence requires a lot of energy. Additionally, a human would traditionally encounter very little new information during their lifetime, so any new information likely fit very well within pre-existing information and concepts.

The modern human encounters massive amounts of new information daily, much of which is completely foreign to a person's large mental framework of the world. Unfortunately, the modern human must actively work against that innate tendency to fit new information into old molds. On the surface, it is just so much easier and satisfying to find evidence that fits our preconceived ideas!

How do we fight confirmation bias? By one or more of: (1) recognizing that confirmation bias exists and is something your brain favors; (2) formulating multiple hypotheses for every genealogical mystery; and (3) trying to find evidence that *disproves* a hypothesis, rather than only looking for evidence that supports a hypothesis.

In the following image, most genealogists will find a match to *either* Corey *or* Velma and will stop there. However, by testing the second descendant we discover that our conclusion is not as solid as we thought it was. This is an example of attempting to avoid confirmation bias, and trying to DISPROVE our hypothesis or conclusion rather than simply gathering and relying on biased evidence.



- Mohajer, Sia. “Confirmation Bias and the Power of Disconfirming Evidence,” *Farnham Street* (<https://fs.blog/2017/05/confirmation-bias>).
- Starmans, Barbara J. “The Dangers of Confirmation Bias,” *genealogy + news*. (<https://www.outofmytreegenealogy.com/dangers-of-confirmation-bias/>).

2. Tree Completeness

How complete are the trees you are comparing? In other words, how confident are you that the DNA couldn't have come from other shared ancestors? This is one of the biggest problems currently faced by those utilizing DNA evidence. Without sufficient data, it can appear that a segment comes from one ancestor, only to learn with additional testing or other data that the segment came from another *previously unknown or unconsidered* ancestor or ancestral couple.

For example, you have (on average) 32 great-great-great-grandparents. How many of those do you know? If you're comparing with 4th cousins how sure are you that you're comparing all your ancestors and all your match's ancestors back to that generation?

To triangulate a segment of DNA or an ancestor, it is necessary to compare the family trees of the matches. In the following chart, we see the percentage of identified ancestors in the trees of the three people used to triangulate to a possible ancestor at the **4Great-Grandparent** level:

	Total Percentage of Identified Ancestors [(Known Ancestors / Possible Ancestors) x 100]		
	Me	Match #1	Match #2
Grandparent	100%	100%	100%
2Great-Grandparent	100%	100%	100%
3Great-Grandparent	88%	88%	50%
4Great-Grandparent	78%	75%	12%
5Great-Grandparent	60%	55%	5%

Is this triangulation reliable? Shouldn't reviewers have this information available when considering the evidence?

How much of your tree do you “know”?

Ancestor	Cousin Relationship	Total # Known	Percentage
Grandparents	1C	out of 8	
Great-Grandparents	2C	out of 16	
GG-Grandparents	3C	out of 32	
GGG-Grandparents	4C	out of 64	
GGGG-Grandparents	5C	out of 128	
GGGGG-Grandparents	6C	out of 256	

- Bettinger, Blaine. How Much of Your Family Tree Do You Know? And Why Does That Matter?, *The Genetic Genealogist*, 11 August 2015 (<https://thegeneticgenealogist.com/2015/08/11/how-much-of-your-family-tree-do-you-know-and-why-does-that-matter/>).
- Bettinger, Blaine. The DNA Era of Genealogy, *The Genetic Genealogist*, 17 December 2016 (<https://thegeneticgenealogist.com/2016/12/17/the-dna-era-of-genealogy/>).

3. Segment Size and False Matching

Are you avoiding small segments and ensuring that the shared segment size(s) are reasonable in view of the hypothesized relationship? Many to most small segments are problematic, so we should either avoid them entirely or provide considerable reasoning as to why it can be utilized.

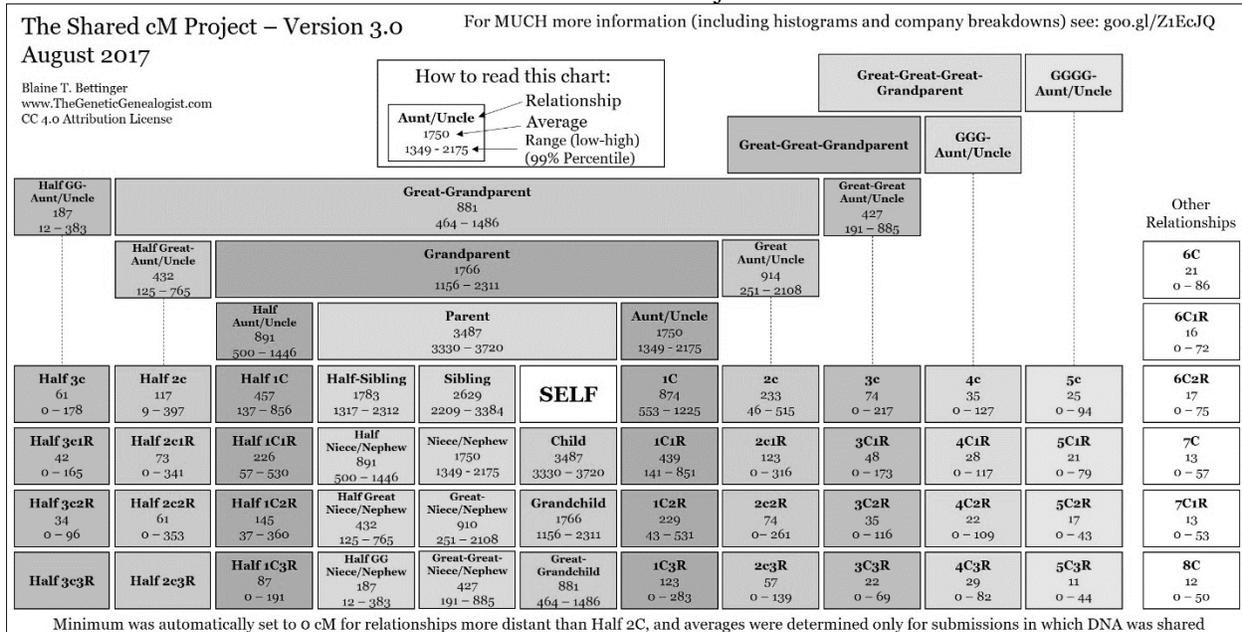
- Bettinger, Blaine. A Small Segment Round-Up, *The Genetic Genealogist*, 29 December 2017 (<https://thegeneticgenealogist.com/2017/12/29/a-small-segment-round-up/>).
- ISOGG Wiki, Identical By Descent (https://isogg.org/wiki/Identical_by_descent).
- Speed & Balding. Relatedness in the post-genomic era: is it still useful? *Nature Reviews Genetics*. 2015: 16: 33-34 (<https://www.nature.com/articles/nrg3821>). Paywall, see ISOGG Wiki, Identical By Descent (https://isogg.org/wiki/Identical_by_descent) for information about this paper.

4. Total Shared DNA

Are you ensuring that the total shared cM is reasonable in view of the hypothesized relationship? The Shared cM Project is a collaborative data collection and analysis project created to understand the ranges of shared centiMorgans associated with various known relationships. As of August 2017, total shared cM data for more than 25,000 known relationships has been provided.

- Bettinger, Blaine. August 2017 Update to the Shared cM Project, *The Genetic Genealogist*, 26 August 2017 (<https://thegeneticgenealogist.com/2017/08/26/august-2017-update-to-the-shared-cm-project/>).
- Bettinger, Blaine. The Shared cM Project – Version 3.0 (August 2017), *The Genetic Genealogist*, 26 August 2017 (https://thegeneticgenealogist.com/wp-content/uploads/2017/08/Shared_cM_Project_2017.pdf).
- Perl, Jonny. The Shared cM Project 3.0 tool v4, *DNA Painter* (<https://dnapainter.com/tools/sharedcmv4>).

The Shared cM Project:



See Bettinger, Blaine. August 2017 Update to the Shared cM Project, *The Genetic Genealogist*, 26 August 2017 (<https://thegeneticgenealogist.com/2017/08/26/august-2017-update-to-the-shared-cm-project/>) for the full chart.

5. Segment Frequency

We care about segment frequency because segments that are frequently found within a population can cause issues when we are trying to assign them to a particular ancestor, or use them to support a genealogical conclusion. The more common a segment, the greater the probability that you could have inherited it from an ancestor OTHER than the ancestor you think it is! Additionally, it has been suggested that the more common a segment, the older the segment may be, which makes sense (although these studies tend to focus on the smaller segments that we should largely be staying away from – see HERE for more). The more time a segment has been around, the greater the chance that it could be widespread.

- Bettinger, Blaine. Analyzing Segment Frequency at GEDmatch, *The Genetic Genealogist*, 13 August 2017 (<https://thegeneticgenealogist.com/2017/08/13/analyzing-segment-frequency-at-gedmatch/>).

Other Resources:

- Mills, Elizabeth Shown. QuickLesson 24: Evaluating DNA as Evidence, *Evidence Explained*, (<https://www.evidenceexplained.com/content/quicklesson-24-evaluating-dna-evidence>).