DNA Analysis Methodology:
Defeat the Genealogy Gremlin
With Pedigree Evaluation, Mitigation, and Reasoning

20 October 2023

Reisinger Memorial Lecture

Sponsored by the Board for Certification of Genealogists

Karen Stanbary, LCSW, AM, CG® ◊ https://karenstanbarygenealogy.com ◊ info@karenstanbarygenealogy.com

Today's digital click-through age conditions us to expect immediate answers. Too often we buzz through match lists, clicking on trees and forming hasty conclusions. In this intoxicating excitement, we might forget a tried- and-true genealogy tenet: Try as hard to disprove your hypothesis as you try to prove it. Most genealogists begrudgingly welcome this tenet after failed research attempts. The phrase echoes in our minds, but do we apply it when working with DNA test results?

Two standards express the principle:

- Standard 50, "Assembling conclusions from evidence," says "Once a genealogist resolves conflicting evidence, all remaining relevant evidence items are compatible with *a single answer to the research question"*[1]
- Standard 17, "Extent," says "thorough research gathers sufficient data to test— *and to support or reject— hypotheses*."[2]

A common mistake in DNA analysis occurs when we *rule in* conclusions instead of *ruling out* competing hypotheses. Many naively scan match lists, seeking common surnames, and then hastily conclude their shared autosomal DNA came through the ancestral line with the common surname. We see what we want to see. Confirmation bias pollutes our conclusions.

Think about it: Anyone alive on planet earth at the time of a child's conception could be the source of the child's shared DNA. As ridiculous as this sounds, the underlying logic will help us achieve accurate results. Start big, and then eliminate candidates until one answer remains. Consider all ancestral lines as competing hypotheses until your evidence analysis, evidence correlation, and reasoning rule them out.

Standard 52, "Analyzing DNA test results," provides best practice guidance for pedigree evaluation. The gaps in incomplete or shallow pedigrees represent other possibilities for the common ancestor. We consider the possibility of more than one common ancestor in any comparison of two DNA test takers.

We always begin by carefully crafting a focused research question seeking unknown information about an ancestor (the research subject). We then recognize and control assumptions about our research's starting point. We select base DNA test takers who provide good coverage of the research subject's genome. When selecting the base test takers, we consider the unique DNA each individual brings to the table, e.g., the X-DNA, YDNA, and mtDNA of the research subject. We confirm that the biological relationships among the base test takers are consistent with the documented relationship. We assess the accuracy of the DNA test takers' pedigree. We note any instances of pedigree collapse. We document any undocumented parent-child relationships. We build each DNA test taker's pedigree to at least the depth of the research subject's parents.

1. Board for Certification of Genealogists (BCG), *Genealogy Standards*, 2nd ed. rev. (Nashville: Ancestry, 2021), 28–9.
2. Ibid., 14–15.

Then we seek a group of matches who may be related on the studied ancestral line. We conduct research and reconstruct those matches into one family. This provides the basis to hypothesize how our base test takers may fit into that reconstructed family. Part of the analysis process includes the evaluation of each match's pedigree.

We conduct our own research to validate the match's tree. We document the undocumented parent-child relationships we find. We research to discover unknown maiden names of mothers. We examine the compared pedigrees for completeness. Most have gaps that may hide the source of the shared DNA, despite known common ancestors in the compared pedigrees.

But evaluation alone is insufficient. It is not enough to simply evaluate compared pedigrees and report on the percentage of identified and proven ancestors. Evaluation without mitigation of the problems is meaningless. We must mitigate any problems before making a conclusion about a biological relationship. Unmitigated problems represent possibilities for another shared ancestor. The chain is only as strong as the weakest link.

The September 2023 issue of *OnBoard* provides details about fifteen strategies to mitigate problems encountered in pedigree evaluation:

1. **Substitute a test taker with a more complete pedigree.**
2. **Conduct additional documentary research** to fill pedigree gaps.
3. **Employ advanced atDNA analysis techniques** like clustering, genetic networks, chromosome painting, visual phasing, and segment triangulation.
4. **Use advanced analysis techniques to add a layer to the analysis** that eliminates competing hypotheses.
5. **Articulate reasoning arguing that gaps in the pedigree are irrelevant** to the research question.
6. **Harness the power of X-DNA's unique inheritance path** to rule out ancestral lines as impossible.
7. **Correlate matches from both sides of the ancestral couple** to justify the elimination of irrelevant ancestral lines even when those lines include unknown ancestors.
8. **Select and analyze DNA test takers who are related to multiple base test takers**.
9. **Maximize the research subject's genome coverage** to capture DNA test takers descending from more distant ancestors.
10. **Analyze shared segments for a unique admixture estimate.**
11. **Consider members of genetic clusters as the FAN club for DNA analysis[3].**
12. **Avoid using DNA test takers who share only small segments.**
13. **Correlate chromosome painting of multiple common segments** may help to eliminate the competition.
14. **Integrated genetic and document evidence** may support only one answer to the research question.
15. **Utilize the power of Big-Y testing** to identify patrilineal kin who may not appear on Y-STR tests.

3. A documentary FAN club, coined by Elizabeth Shown Mills, identifies associated people (Friends-Associates-Neighbors) of a research subject. A biological FAN club is a group of genetically related test takers who likely descend from the same distant ancestor of the research subject.

CASE STUDY

Who are the biological parents of John Weyer who married Elizabeth Weidmann on 11 January 1876 in Burlington, Iowa?

Base test taker: John is my mother—Nancy Stumpf's, paternal great-grandfather. Nancy is an only child and provided an autosomal DNA sample. She does not have any first cousins on this line available for DNA testing. Alone, her DNA only covers about 6.25% of the genomes of each of John's biological parents.

Documentary obstacles:

- John was born about 1852, probably in Indiana, but the paper trail provides conflicting information about his birthplace. There is no civil birth registration for this time and place.
- According to family lore, he was orphaned at age seven. There is no death information for his father. The Catholic cemetery records are missing for that time.
- John's death certificate names his father "Joseph Weyer" and his mother "No Record."
- John's census records report differing birthplaces for his father.
- There are twelve men with the name of Joseph Weyer in Indiana in 1850.
- John's obituary names a sister. Her death certificate names different parents with totally different birthplaces (Mississippi).

Research strategy:

1. Identify serendipitous matches (second to third cousins) who provide additional coverage of the genomes of John's biological parents. Strive to identify matches who descend from different children of John Weyer and Elizabeth Weidmann. Seek collaboration and access to their match lists.
2. Group third cousins into likely Weyer and likely Weidmann buckets.
3. Seek unknown matches who share atDNA with the likely Weyer bucket.
4. Reconstruct matches into one family.
5. Conduct pedigree evaluation and mitigate any problems.
6. Correlate the amount of shared atDNA with the traced relationships and compare to statistical studies.
7. Hypothesize parents for John Weyer, one parent at a time.
8. Try as hard to disprove the hypothesis as you try to prove it.

Examples of selected pedigree mitigation strategies:

- **Maximize the research subject's genome coverage** to capture DNA test takers descending from more distant ancestors.

  I begin with limited test takers for the Weyer problem. My mother is an only child with no paternal first cousins available for testing. Since John Weyer's biological parents are her second great-grandparents, it is best to seek additional collaborative test takers. Searching the Ancestry, FTDNA, My Heritage *and* GEDmatch *databases identifies a group of second cousins. I seek collaboration and access to their match lists.*

- **Substitute a test taker with a more complete pedigree.**

  Even though TG shares 195 cM with Nancy, his tree provides little useful information. His Galvin surname is a known surname in John's eldest daughter Elizabeth Rogosia's descendants. TG is unresponsive via Ancestry messaging. I find another match who shares 53 cM but who has a complete pedigree linking back to John Weyer through his daughter Elizabeth Rogosia. I use this tree to identify the Galvin connection. Newspaper and living people databases identify TG. I add both Galvin descendants to the hypothetical descendant tree, increasing coverage of the genome.

- **Conduct additional documentary research** to fill pedigree gaps.

  SM (Ancestry) shares 58 cM with Nancy on two segments. The match is ICW proven Weyer-Weidmann descendants. The attached pedigree names a father, Joseph A. Moriarty, born in Knoxville, Tennessee and died in Kingsport, Tennessee. With this information in hand, I can build a tree that leads to Seraphine Weyer, a son of Joseph Weyer and Ann Spengler.

- **Articulate reasoning arguing that gaps in the pedigree are irrelevant** to the research question.
  Presley's Ancestry profile is not linked to a tree. But his FTDNA *profile provides a partial pedigree. He also says he descends from Joseph Weyer and Anne Spengler. The pedigree does not name one paternal great-grandmother and four paternal second great-grandparents. Three maternal second great-grandparents are unidentified. If Presley is related to Nancy through the Joseph Weyer and Anne Spengler couple, then Presley would be a third cousin/half third cousin to Nancy, predicted to share, on average 49 cM (half 3C) or 75 cM (3C). Presley shares 197 cM* (FTDNA) *or 185 cM (Ancestry), way more than one standard deviation from the mean. Pedigree evaluation is essential to evaluate the possibility of multiple common ancestors. His paternal gaps can be discounted using reasoning. But that reasoning does not apply to the gaps on his maternal side. Additional research does not reveal common surnames nor German locations.*

- **Employ advanced atDNA analysis techniques** like clustering, genetic networks, chromosome painting, visual phasing, and segment triangulation.

  Presley and Nancy have Ancestry *shared matches on a Kurtz line and a Moriarty line forming a genetic cluster. They share just one distinct cluster with other descendants from Joseph Weyer's first marriage. The cluster is only connected to the Weyer-Weidmann cluster. There is not enough data to analyze chromosome painting.*

- **Select and analyze DNA test takers who are related to multiple base test takers**. *An extensive group of test takers helps to reach a defensible conclusion that John Weyer is the biological son of Joseph Weyer and his second wife Margaretha Birge. Seven Weyer-Weidmann descendants share DNA with four Weyer-Spengler descendants from three different children. Furthermore, due to the half-relationships resulting from two marriages, the most recent common ancestral couple for all the test takers are the parents of Frantz Joseph Weyer—Joseph Weyer and Elisabetha Gasche of Engenthal, Bas-Rhin. John Presley may be related to my mom on a second line but that does not impact the conclusion that he is related to her on her Weyer-Gasche line. The fact that multiple independent descendants from Frantz Joseph Weyer's first marriage match multiple descendants from his second marriage reduces the burden of full pedigree evaluation among the matches.*