Using DNA to advance family research

By David Bradford

New developments in molecular biology and genetics have combined to provide genealogists new insights into their deepest ancestral roots. When traditional paper-based research proves inadequate, DNA testing offers another reliable "source" for investigation of some kinds of brick-wall situations.

Background

In both plants and animals, traits pass to offspring by means of long molecular strands called Deoxyribonucleic Acid (DNA). Virtually all human cells contain nine meters of nuclear DNA in forty-six tightly wound chromosomes. DNA is a bit like coiled telephone wire—portions of each DNA molecule, called genes, carry information, while other portions, called junk DNA, surround the genes. These instruction-laden genes determine our gender, eye, hair and skin color, disease resistance, and even future personality traits.

As parents contribute genes to children, almost all of their DNA is scrambled, producing offspring whose appearance and DNA are somewhat different from either parent or other siblings. This DNA shuffling is crucial to human evolution, because it allows the expressions of diverse traits that help descendants survive in the face of changing environments.

Recent scientific advances have led to an explosion of DNA testing for paternity determinations, disease vulnerability, criminal investigations, and genealogical research. Most of these tests look within genes.

But genealogical DNA testing examines junk DNA that escapes scrambling and continues virtually unchanged from one generation to the next. Testing these stable DNA markers can establish evidence of kinship when test results from two individuals are identical or nearly so.

DNA can answer many kinds of genealogical questions, but there are limitations. Nevertheless, if a hypothesis is testable by DNA analysis, new understandings can bridge gaps in traditional paper records.

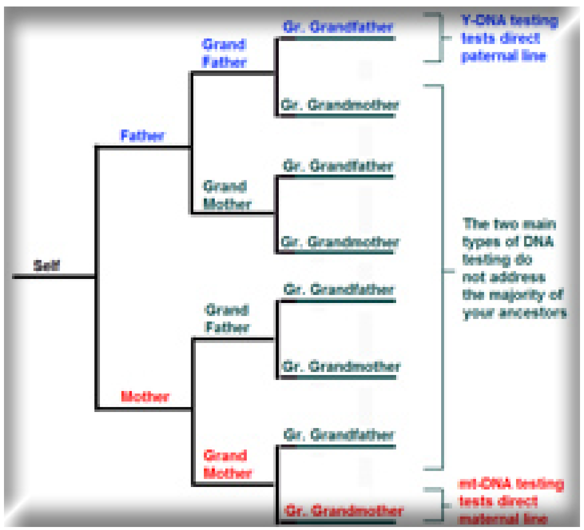
DNA testing—the basics

This is one test that doesn't require studying. Here is the lowdown:

- DNA is painlessly collected by swabbing the inside of the cheek using a kit from one of several DNA testing labs. The swab is mailed to the lab along with a signed consent form and payment.
- Pricing depends on the lab, the test type and the number of markers measured. DNA tests analyze between ten and sixty-seven markers and cost between $100 and $350.
- DNA labs sponsor DNA Surname Projects that focus on individual surnames, and the project administrators may offer discounts.
- DNA labs provide online resources. These hyperlinks are a mix of several labs' offerings: FAQs, Tutorials, Glossaries, Newsletters, Case Studies, Readings, Privacy Policies, and Links.

- In a couple weeks, the lab returns a DNA "Report Card" (genotype) consisting of a numerical value (allele) for each DNA marker measured.
- DNA results may also be posted anonymously to the lab's online database under the kit number.
- Long-dead ancestors need not be disinterred for testing; rather, their inherited DNA is available by testing their living descendants (or those of related branches).

The power of most DNA analysis comes from comparing test results of two suspected branches to establish (or refute) a family link. Logically, the more alike are two people's DNA test results, the more closely related are the individuals.



What DNA testing won't do

DNA testing is a powerful tool, but there are also some limits inherent in DNA testing:

- Genealogical DNA test results are not a unique personal DNA fingerprint; rather, test results reflect a family genotype; close kin have identical results.
- Genealogical DNA test results will not reveal genetic disorders because the lab looks at the junk DNA, not genes.
- Genealogical DNA testing examines only two branches of a pedigree: the direct maternal and paternal branches. Most ancestors fall in between and are invisible to the more common DNA tests.
- Genealogical DNA tests are well suited to disproving a genetic relationship; two people with very different test results can be certain that they are unrelated.
- Genealogical DNA testing will not define the exact degree of kinship.

DNA is well suited to revealing whether a relationship exists, but because it is based on probability calculations, DNA provides an inexact estimate of the who and when of newly discovered relationships.

The big three DNA tests

Y-Chromosome DNA (Y-DNA) reveals a man's father's father's line (a.k.a., the surname line). Y-DNA is conveyed virtually unchanged from father to son in the Y (male) chromosome without any maternal contribution. Only males may contribute cells for Y-DNA testing. Female genealogists often persuade male relatives to contribute DNA on their behalf. Y-DNA results must be compared to another suspected relative's results to establish kinship. The closer are two people's Y-DNA results, the closer is their relationship.

Despite avoiding the genetic shuffle of each generation, Y-DNA still undergoes a slow and predictable rate of spontaneous DNA change (mutation). The average mutation rate of individual Y-DNA markers is once every twenty generations—or about every five hundred years. These slow, predictable changes make Y-DNA useful for many European and American genealogists, since the earliest use of surnames began five to eight hundred years ago. Thus, Y-DNA is the most common and genealogically useful DNA test. When two people's Y-DNA test results are identical or differ by one or two marker values, it is considered evidence of kinship.

Y-DNA online databases already exist: Y-Base, SMGF, and Y-Search. Y-DNA databases are free and open to the public. Oxford Ancestors also hosts a members-only Y-DNA database.

Mitochondrial DNA (mt-DNA) reveals a person's mother's mother's line. Mitochondrial DNA is conveyed virtually unchanged from mother to children in the mother's egg without any paternal contribution. Men and women both receive their mother's mt-DNA, and either may donate cells for mt-DNA testing. But mt-DNA is considered a "paternal dead end" because only women pass their mt-DNA to later generations. mt-DNA results must be compared to another suspected relative's results to help establish kinship.

Because mt-DNA mutates only once every five hundred generations (ten thousand years), mt-DNA is more archeological than genealogical in character. But, mt-DNA can help with certain narrow questions (e.g., to which of great-granddad's three wives does my maternal grandmother belong?). Unlike Y-DNA, only exact matches of two people's mt-DNA results are considered genealogically relevant.

A few searchable online mt-DNA databases exist: MitoSearch, SMGF, and the FBI's mt-DNA Population Database. Oxford Ancestors also hosts a members-only mt-DNA database.

Ancient Ancestral DNA (Haplogroup Analysis) reveals broad geographic origins of pre-historical ancestors. Depending on the lab and on the research question, Ancient Ancestral DNA can be obtained from mitochondrial or chromosomal sources. Choose a lab whose Ancient Ancestral DNA test best answers the hypothesis being examined:

- Y-DNA reveals Ancient Ancestors in the direct paternal line;
- mt-DNA reveals Ancient Ancestors in the direct maternal line;
- Autosomal DNA (chromosomes other than the X or Y) reveals all Ancient Ancestors.

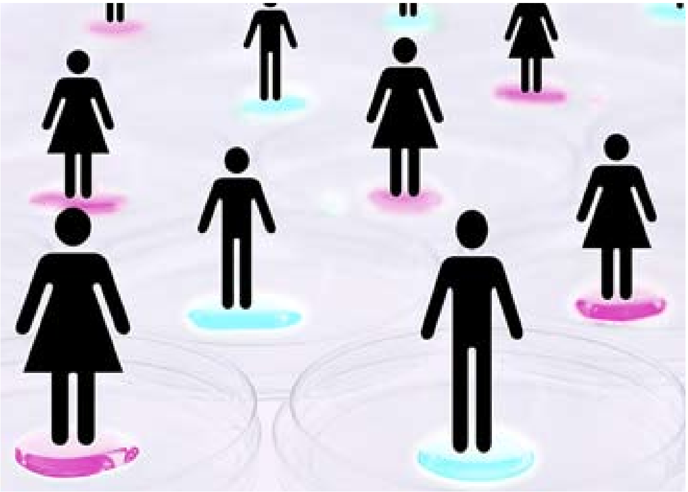
Ancient ancestral results are stand-alone findings and are meaningful without comparison. Some labs offer qualitative results (Nordic ancestry: Yes or No), while other labs provide quantitative results (Nordic: 14 percent, Native American: 5 percent, etc.).

YHRD is one Ancient Ancestral DNA database.

Most recent common ancestor (MRCA)

Most DNA testing seeks to provide proof of relatedness between two individuals. The universal measure of DNA relatedness is called "Time to Most Recent Common Ancestor"—most often shown as MRCA. MRCA comes in two flavors:

- **Expected MRCA** is based on traditional records research; it is the expected (hypothesized) generational difference between the two suspected living family members. For example, hold up four fingers to represent a boy, his father, his paternal grandfather, and his paternal great-grandfather. Between the four fingers there are three generational gaps (mutational opportunities). This boy and his paternal great-grandfather would have an Expected MRCA value of three.
- **Calculated MRCA** is based on a comparison of two DNA test results—this is the probable generational difference estimated in the lab's look-up chart. The number of DNA differences (actual mutations) indicates the amount of time (generational gaps) separating the individuals.



Comparing an Expected MRCA to the Calculated MRCA is the essence of genealogical DNA testing; identical or very similar values for Expected and Calculated MRCA is strong evidence of kinship.

Develop a DNA testing strategy

DNA cannot solve every research question. Before purchasing a DNA test, first see whether a genealogy question meets these requirements:

- Does the question involve a branch of the pedigree that can be revealed by DNA testing?

Example: Does the question deal with a father's father's line (Y-DNA), a mother's mother's line (mt-DNA), or relate to a broader question, like, "do I have Native Americans among my forebears?" (Ancient Ancestral DNA).

- Is there sufficient conventional research to propose a hypothesis to answer the question?

Example: Based on Kay Smith's vital records research, she suspects, but cannot yet prove, that she and James Smith share a direct, unbroken, paternal great x third great-grandfather, Ulysses Smith.

- Do living relatives exist to test the proposed answer to this research question?

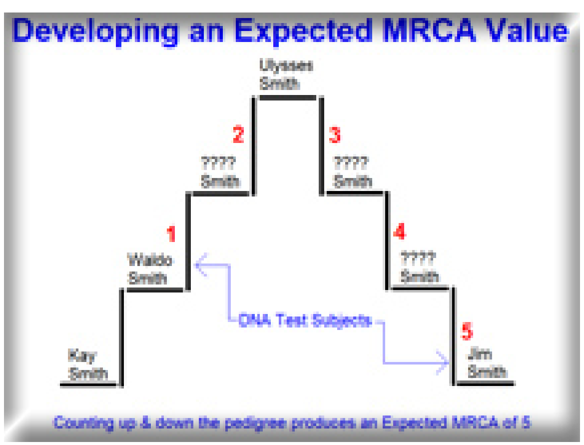
Example: Kay Smith is asking a paternal line question, and because she is female, Kay must persuade a brother, father, paternal uncle, or other qualifying male relative to be tested in her place. Jim Smith is living and is male, so he may be directly tested.

- Will these people agree to the inconvenience, cost, and revelation of DNA results?

Example: Both Jim Smith and Kay Smith's qualifying male relative must allow their Y-DNA to be tested and then agree to share the results for comparison. Kay must also pay for these two tests or convince family members to contribute.

Identify the DNA test type, number of markers, and test subjects

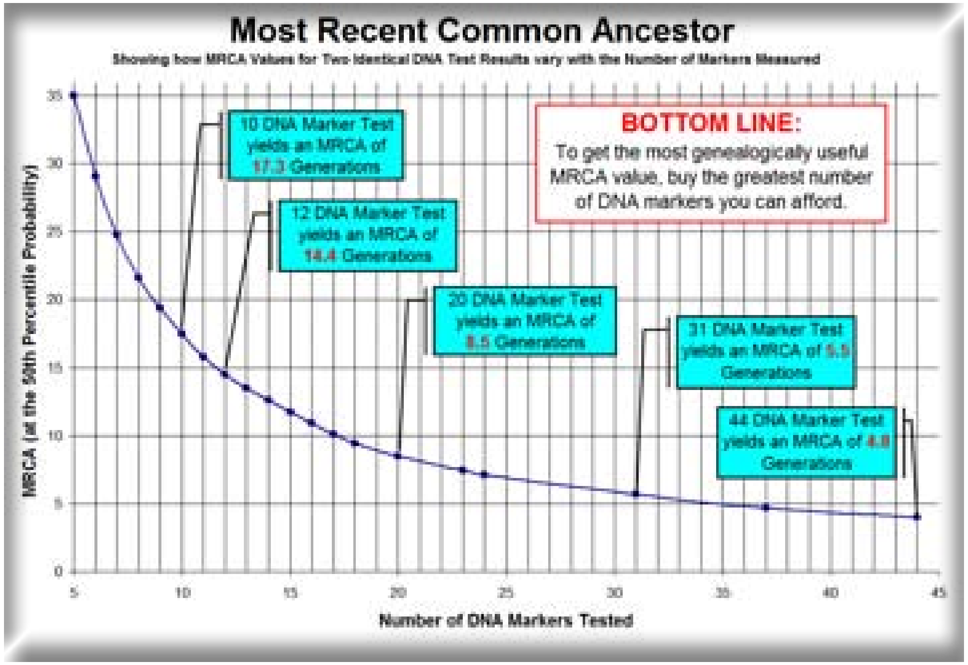
In this example, we suspect that Kay Smith and Jim Smith share a common paternal ancestor, so Y-DNA will be tested. This means Kay (a female) must convince both Jim Smith and one of her male relatives with a direct unbroken father–son line to Ulysses Smith to be tested. Let's assume Kay convinces Jim and her father Waldo to request thirty-one marker tests from a DNA lab.



Determine the expected MRCA

Based on her vital records research, Kay counts the number of generations that she believes separate Waldo from James. This requires that she tally the expected (hypothesized) generational gaps between the two test subjects going both up and down the pedigree:

Kay counts five gaps between Jim Smith and her father. This gives her an Expected MRCA value of 5.



Compute the calculated MRCA

A calculated MRCA is arrived at by comparing two DNA test report cards. The number of DNA markers with values that differ (regardless of by how much) is tallied as evidence of mutation indicating the passage of generational time. Then, based on the total number of markers measured, we can estimate how many generations most probably separate the two test subjects. There's no need for a calculator, because DNA labs provide easy-to-use online calculated MRCA look-up charts.

The more markers that are measured, the more narrowly predictive and genealogically useful the "Report Cards" will be. For this reason, your DNA test should include as many markers as you can afford.

The calculated MRCA is the most probable, but not perfectly predictive, number of generations separating two individuals whose DNA tests are compared.

There is uncertainty reflected in the calculated MRCA; in fact, for the same set of facts, multiple MRCA values are provided for 50, 90, and 95 percent probability. The value selected depends on the level of certainty desired.

It is important to note that calculated MRCA values assume that our two DNA test subjects are randomly selected (unrelated). However, genealogists select individuals for DNA testing precisely because of strongly suspected kinship by dint of shared surname, common geography or other traditional research evidence. This pre-existing likelihood of relatedness is nowhere reflected in the calculated MRCA. With good traditional research strongly suggesting kinship, genealogists often cautiously adopt the 50 percent probability level. In other words, meaningful DNA testing requires a strong suspicion of relatedness based on prior paper-based research; DNA is a supplement, not a substitute for traditional research methods.

In our example, let's assume that Waldo and James Smith have agreed to be tested and that all thirty-one of their DNA marker values are identical. We then use the lab's online look-up tool to find the calculated MRCA is 5.5 generations at a 50 percent probability.

Test the hypothesis

To test her hypothesis, Kay compares her two MRCA values:

- Expected MRCA—Based on her traditional records research, she expects Waldo and Jim Smith to occupy positions on their family pedigree that are five generations apart, and
- Calculated MRCA—Based on DNA test results from Waldo and Jim Smith, she found that their identical DNA marker values predict positions on their family pedigree consistent with about 5.5 generational opportunities to mutate with a 50 percent probability.

Draw conclusions

Comparing Expected MRCA (suspected from her traditional "paper" research) to Calculated MRCA (from two people's DNA results), Kay can draw one of three possible conclusions:

- Identical or very close values provide strong contributory (but not absolute) evidence of the suspected relationship. In the presence of other traditional research evidence, this DNA finding can be the confirmatory bit of proof a researcher has been seeking. This is the conclusion Kay draws from her family's DNA testing.
- Somewhat similar results mean that some relationship exists, but that the two test subjects are either more closely or more distantly related than originally suspected. For instance, if Kay had a calculated MRCA of one generation (or, for that matter ten generations) it would have suggested that Kay was on the right track, but that she needs to revise her theory about kinship to better reflect the reality of the DNA records.
- Significantly different results allow the certain conclusion that no relationship exists. This is also true anytime one or more mt-DNA or three or more Y-DNA marker values differ between test subjects. For instance, if Kay's two DNA report cards were complete mismatches, it would mean Kay's suspicion was entirely wrong, regardless of what her paper research indicated.

Wholly unexpected findings of non-relatedness may reflect an undocumented adoption, an altered or assumed surname, a previously unknown marriage, or maternal infidelity somewhere in the family line.

The incidence of such events in any one generation is statistically small but it is cumulative across the pedigree. The further back one's DNA investigation probes, the greater the possibility of an unexplained interruption in the familial DNA chain.

Testing other family branches can often confirm or clarify unexpected findings. Remember that such surprises can prove embarrassing to family members (e.g., Uncle Melvin who is the president of the heritage society to which you've just proved "non-relatedness").

Conclusion

Genetic genealogy, or the use of DNA testing to advance one's family history research, is becoming increasingly popular. DNA is best suited for investigating the direct, unbroken lines of one's father's father's father (Y-DNA) or one's mother's mother's mother (mt-DNA). DNA of ancient ancestral geography is also available, but not as useful to genealogists.

Though DNA will not pinpoint a specific ancestor, it can produce strong confirmatory evidence for many research roadblocks when traditional paper-based sources are incomplete. DNA tests can also provide absolute evidence of non-relatedness and save a researcher years of futile inquiry.

Careful evaluation of your genealogical questions and your relatives' availability and willingness to submit to testing are required before selecting from among the varied and expensive DNA tests available. DG

David Bradford is a Wisconsin native with a B.S. in medicine from UW Madison and an MBA from UW Milwaukee. He is a longstanding member of the Rock County Genealogical Society and currently serves as its webmaster and board chairman. He researches pioneers from southern Wisconsin for others, and has indexed several historical Rock County, Wisconsin, books and atlases. He lectures regionally on topics related to genealogy, medicine, and technology.