Revealing American Indian and Minority Heritage Using Y-line, Mitochondrial, Autosomal and X Chromosomal Testing Data Combined with Pedigree Analysis

Roberta Estes

Abstract

As a project administrator of several historically based genetic genealogy projects, such as the Lost Colony, Cumberland Gap, Melungeon, Carolina Native Heritage and Hatteras Island projects which involve thousands of participants, I routinely receive questions from individuals who have an oral history of Native American heritage and would like to use genetic genealogical tools to prove, or disprove, their oral history. This paper documents the various discovery steps and processes using different types of DNA testing for a typical individual participant and appropriate family members who carry an oral Native history combined with genealogical evidence that has been forthcoming during the elapsed years since genetic testing for genealogy first because available. Each test along with associated benefits and detriments are discussed in relation to the analysis of minority ancestry. The conclusion combines the information from all the various tests, pedigree analysis and genealogical evidence, discussing which tests are beneficial and most accurate, and which ones are not useful, and why.

Oral History

Oral history is an important component of genealogical research. For most people, it's where they begin their search. I've been working with oral history for many years in relation to the various projects I administer, genealogy and my clients. I've discovered that oral history tends to be relatively accurate for 2-3 generations. After that, parts of the story are preserved, but the generations, individuals and timeline are often askew. For example, most people know their grandparents, but generally don't know their great-grandparents, so the stories of their great-grandparents generation are conveyed second hand through either their parents or their grandparents.

Memories fail people, and the essence of the story about Indian ancestry may be accurate, but who and in which generation may have been forgotten or only partially remembered. With each ensuing generation, more detail is lost until only the primary topic itself is remembered accurately, that the family has Indian ancestry. In the hundreds of people I have worked with, I have yet to find one person whose ancestor that they thought was "full blooded Indian" actually was Native as proven by either DNA or traditional genealogy. Needless to say, people are often very disappointed.

DNA is often our only path to unravel the truth, as Indian ancestry was hidden in the 1800s and early 1900s due to laws that labeled anyone with Indian heritage as a "person of color", meaning black or more accurately stated, "not white". Along with that label came the discriminatory practices of that time[1].

Native American History

Many American families carry oral histories of Native American heritage. Most often, we think of either the Western tribes who still reside in or near their indigenous homes, or the Cherokee who were displaced in the 1830s, forced to march from Appalachia to Oklahoma in the dead of winter, an event subsequently known as the Trail of Tears[2].

In truth, the history of Native American heritage in North America is much, much more complex. It is probable that many of the people who carry oral history of

Address for correspondence: Roberta Estes, Roberta@dnaexplain.com or robertajestes@att.net

Received: Dec. 24, 2009; Accepted: July 17, 2010; Published Dec. 19, 2010

Open Access article distributed under Creative Commons License Attribution License 3.0 (http://creativecommons.org/licenses/by-nc-sa/3.0/) which permits noncommercial sharing, reproduction, distribution, and adaptation provided there is proper attribution and that all derivative works are subject to the same license.

[1] http://en.wikipedia.org/wiki/Race_in_the_United_States
[2] http://en.wikipedia.org/wiki/Trail_of_Tears

"Cherokee heritage" are actually descended from a tribe other than the Cherokee, at least initially. The Cherokee were well known for accepting remnants of other tribes whose members and numbers had been decimated by disease or war. Sometimes alliances were created for mutual protection or benefit.

The Cherokees were not the only tribe in the Eastern United States. The Eastern seaboard was widely populated by varying tribes, some related and affiliated, and some not. There were in fact three major language groupings, Algonquin, Souian and Iroquoian scattered throughout the Eastern seaboard northward into Canada, westward to Appalachia and south to the Gulf of Mexico[3].

People from Africa were also imported very early, often, but not always, as slaves. Jamestown shows evidence of individuals of African heritage as early as 1619. Those who were later brought specifically as slaves sometimes ran away, escaping into the Native population. Conversely, eastern seaboard Indians were often taken or sold by defeating tribes as well as colonists and traders into slavery[4].

In the early years of settlement, European women were scarce. Some men immigrated with wives and families, but most did not, and few women came alone. Therefore, with nature taking its course, it is not unreasonable to surmise that many of the early settlers traded with, worked alongside and married into indigenous families, especially immigrants who were not wealthy. Wealthy individuals traveled back and forth across the Atlantic and could bring a bride on a subsequent journey.

What does this mean to the family historian who is trying to prove their genealogy and understand better just who they are and where they come from?

If a family has a long-standing oral history of Native American heritage, it is probably true at some level. Historically, Native people were classified as "non-white" which severely limited (and sometimes prevented) their ability to function as free people with equal rights. This means that "free people of color" often could not vote, own land, bear arms nor attend schools along with white people, if at all.

Furthermore, laws varied and how much non-white heritage constituted a "person of color" ranged from the infamous "one drop" rule to lesser admixture, sometimes much more liberal, to only the third generation. In essence, as soon as individuals could become or pass for "white" they did. It was socially and financially advantageous. It is not unusual to find a family who moved from one location to another, generally westward, and while they were classified as mulatto in their old home, they were white in their new location.

Generally there were only three or sometimes four classifications available, white, negro or black, mulatto and Indian. Sometimes Indian was a good thing to be, because in colonial states, reservation or tribally affiliated Indians weren't taxed. However, this also means their existence in a particular area often went unrecorded. Often, poor "free people of color" weren't taxed either because they lived and worked on someone else's land and they had nothing of value. Because of their poverty and resulting lack of records, they became invisible to the genealogist.

Any classification other than white meant in terms of social and legal status that these people were lesser citizens. Therefore, Native American or African heritage that was not visually obvious was hidden and sometimes renamed to much less emotionally and socially charged monikers, such as Black Dutch, Black Irish and possibly also Portuguese. In Hawkins and Hancock County, Tn, the Melungeons were also prevalent and have proven to be a genetically tri-racially admixed group[5], although the term Melungeon tended to be a social epithet, a label one may have used to refer to darker neighbors, but never to describe one's own family.

For genealogists who are fortunate, there are records confirming their Native heritage, such as the Dawes Rolls[6] and other legal documents. More often, there are only hints, if even that, such as a census where an ancestor is listed as mulatto, or some other document that suggests their mixed heritage. Most often though, the stories are very vague and were suppressed for generations. References may be oral or found in old letters or documents. Supporting documentation is often missing. Physical traits associated with Native heritage, the most common and readily apparent of which are shovel shaped incisors, may be present[7].

Many times, it was the woman of the couple who was admixed initially, of course leading to admixed children,

[3] Where Have All the Indians Gone? Native American Eastern Seaboard Dispersal, Genealogy and DNA in Relation to Sir Walter Raleigh's Lost Colony of Roanoke (JOGG 2009 Vol 5,#2) Estes

[4] Indian Slavery in Colonial America, (2009) by Alan Gallay, University of Nebraska Press

[5] Melungeons and DNA, Fall 2009, Melungeon Historical Society Newsletter, Roberta Estes

[6] Final Rolls of the Citizens and Freedmen of the Five Civilized Indian Tribes, National Archives, http://www.archives.gov/genealogy/tutorial/dawes/

[7] University of Illinois at Chicago, School of Oral Science, 2009, *Hominid Evolution, Dental Anthropology, and Human Variation, Non-Metric Variation in Tooth Form*, http://www.uic.edu/classes/osci/osci590/10_1Non-Metric.htm

but with 50% less admixture than their mixed parent. It was much more common for a male of European stock to intermarry with Native or admixed women, rather than the other way around.

This means to genetic genealogists today, that they are likely to meet with frustration when attempting to document Native heritage in both male and female lines. Because this process is both time consuming and often frustrating due to their inability to locate a suitable Y-line or mitochondrial DNA candidate for critical lines, many genetic genealogists turn to autosomal DNA testing with the hope of confirming their Native heritage.

The process of proving Native Heritage can use many tools in the genetic genealogists toolbox, beginning with the standard Y chromosomal and mitochondrial DNA tests.

Creating a DNA Pedigree Chart

The only tools available to the genetic genealogist that answer the question of Native ancestry definitively, and provide the exact line, are the Y chromosomal and mitochondrial DNA tests. In both cases, the haplogroup assignment will determine whether the paternal or maternal line being tested is Native, African, Indo-European or Asian. In some cases, these haplogroups overlap to some extent, but the Native American haplogroups are definitive, with one exception. Doubt about the ethnic association of haplogroup Q1a3 remains.

The first step in evaluating Native heritage is to create a DNA pedigree chart. Using a color coded template that will assist the individual in determining both Y chromosomal and mitochondrial lineage, I suggest that they overlay their genealogical information onto the template for planning purposes. This allows the participant to focus on the lines most likely to have Native ancestry and to eliminate others from consideration.

To assist my clients in this endeavor, I constructed a color coded pedigree chart (Fig. 1) that I encourage them to use in order to identify appropriate individuals to test or who have perhaps tested already in surname projects. On the chart, squares are male and ovals are female lines.

Identifying appropriate individuals to test for various Y

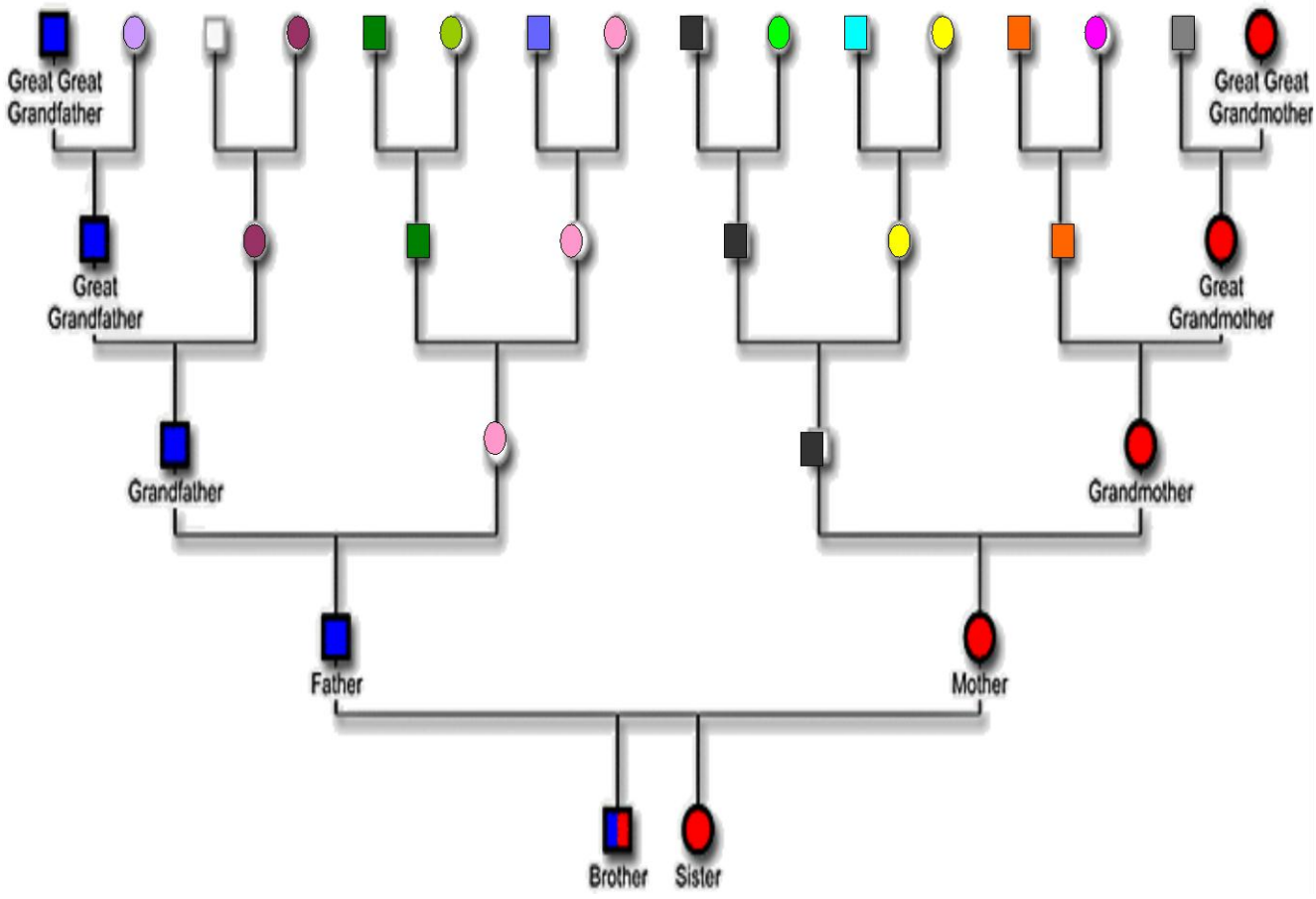

Fig. 1: Color Coded Pedigree Chart

chromosomal (Y-line) and mitochondrial (mtDNA) maternal lines is relatively easy using the DNA pedigree chart, allowing the participant to move backward in time, up the chart, until a line can be found with appropriate living descendants. In the case of Y-line testing, those descendants would be males carrying the family surname, and in the case of mitochondrial lines, it would be a living descendant, male or female, who has descended from the woman in question through only females.

Males

- Carry Y-line DNA inherited from their father
- Represented by blue boxes on the pedigree chart
- Pass the Y chromosome from father to son
- The Y chromosome is what makes a male, male
- Carry mitochondrial DNA inherited from their mother
- Don't pass their mitochondrial DNA on to their children (their wives contribute mtDNA to their children.)
- Y-line DNA is not admixed with any DNA from the mother

Females

- Carry only mitochondrial DNA, not Y-line
- Represented by the ovals on the pedigree chart
- Pass their mtDNA to children of both sexes
- Only female children pass it on
- MtDNA is not admixed with any DNA from the father

Given the above, the blue father carries the blue Y-line DNA of his father, but he also carries the pink mtDNA of his mother. If he is living, he can test for both. If he has died, then another individual descended only through females from any of the individuals designated by pink ovals can be tested for the same mtDNA that the father carried. These are known as proxy or surrogate tests, where another individual tests "in place of" the person whose DNA one would like to test. When searching for Native ancestors, it is necessary to find various family members to proxy test for each genetic line.

Y-Line Native Haplogroups

Using the DNA Pedigree chart as a reference, if the belief is that the father's paternal line is the line carrying the Native heritage, the brother or father would be appropriate candidates to test. Results of Y chromosomal testing are straightforward and definitive in terms of haplogroup results. Haplogroups Q1a3a (M3) and subgroups[8], and C3b(P39)[9] are the only Native American haplogroups, with the possible exception of haplogroup Q1a3 which is as yet indeterminate. SNP[10] testing is required in order to determine the full haplogroup designation.

Mitochondrial DNA Native Haplogroups

Using the DNA pedigree chart as a reference, if the maternal line is believed to carry Native heritage, either the brother, sister or the mother would be appropriate to test. Native American mitochondrial DNA haplogroups are limited to 5 main haplogroups, A, B, C, D and X and within those, 15 subgroups: A2, A2a, A2b, B2, C1b, C1c, C1d, C1d1, C4c, D1, D2a, D3, D4h3a, X2a, and X2g.[11] There is no known haplogroup overlap outside of the Americas with the possible exception of Northeastern Asia. For those with traditional American genealogy, without any East Asian ancestors, the haplogroup will be a definitive identifier of Native ancestral origins. Full sequence testing is required to determine the full haplogroup designation.

Further Mitochondrial and Y-Line Testing

If the expected ancestor does not produce a Native haplogroup, moving on up that particular line to test contributing lines is the next step. For example, if the maternal line, the red ovals, did not produce a Native haplogroup, the next step would be to test the individual's contributing lines, such as the mother's father, Mr. Black or the grandmother's father, Mr. Orange. The goal with this testing is to test until the pedigree chart of the various ancestors has been completed. One by one, each ancestor gains a genetic identity and the Native ancestor will either be identified, eliminated as a possibility, or the participant will eventually run out of ancestors with descendants that can be located to proxy test.

This is the point at which many seasoned genealogists become frustrated and turn to autosomal testing in an

[8] Subgroups Q1a3a1, Q1a3a2 and Q1a3a3 have only been found in South America.

[9] Stephen L. Zegura, Tatiana M. Karafet, Lev A. Zhivotovsky, and Michael F. Hammer, "High-Resolution SNPs and Microsatellite Haplotypes Point to a Single, Recent Entry of Native American Y Chromosomes into the Americas," *Molecular Biology and Evolution* 21(1):164–175. 2004

[10] SNP=single nucleotide polymorphism. This type of testing is available by individual SNP or grouped by haplogroup from Family Tree DNA through their deep clade tests for existing customers.

[11] The Initial Peopling of the Americas: A Growing Number of Founding Mitochondrial Genomes from Beringia, 2010, Genome Research, Perego et al.

attempt to obtain an answer to the question that has remained elusive.

Autosomal DNA Testing

Unlike Y-line and mtDNA testing where the DNA of the father or mother is passed to the offspring unmixed with that of the other parent, autosomal testing tests portions of the DNA of an individual that they receive from both parents. As the field of genetic genealogy has moved forward, research indicates that certain markers are found in higher or lower frequencies in different ethnic or geographic populations.

For example, if someone has the Duffy Null allele, or genetic marker, we know they positively have African admixture. We don't know how much African admixture, or from which line, or when that individual with African roots entered their family tree, but we know for sure they existed.

Attempting to determine the population frequency of varying markers and what that means relative to other populations is the key to this analysis. Few markers are simply present or absent in populations, but are found in varying frequencies. Some populations are widely studied in the research literature, and others are virtually untouched. The process of compiling this information in a meaningful manner so that it can be analyzed is a formidable task, as the information is often found in nearly inaccessible academic and forensic research publications. It's difficult to determine sometimes if the DNA analysis of 29 individuals in a small village in northern Italy is, for example, representative of that village as a whole, of northern Italy, or more broadly for all of Italy as a whole. Is it representative of Italy today or Italy historically? These and other similar questions have to be answered fully before the data from autosomal testing can be useful and reliable.

If the DNA tests being performed aren't mtDNA or Y-line, then they are autosomal tests, meaning they are performed on the balance of the DNA contributed by both parents to their offspring.

Autosomal DNA testing has some unique challenges to overcome. Scientists are still learning about how DNA is passed and recombined. Therefore, we are trying to measure something that we aren't sure how is selected to be passed from parent to child. We know children receive 50% of their DNA from each parent, but we also know that the grandparents DNA is not passed in 25% increments to each child. In many cases, DNA is passed in bundles. We are trying to measure values that appear with more or less frequency in various populations and we are trying to ascertain if the values reported in scientific literature are relevant for entire populations or geographies.

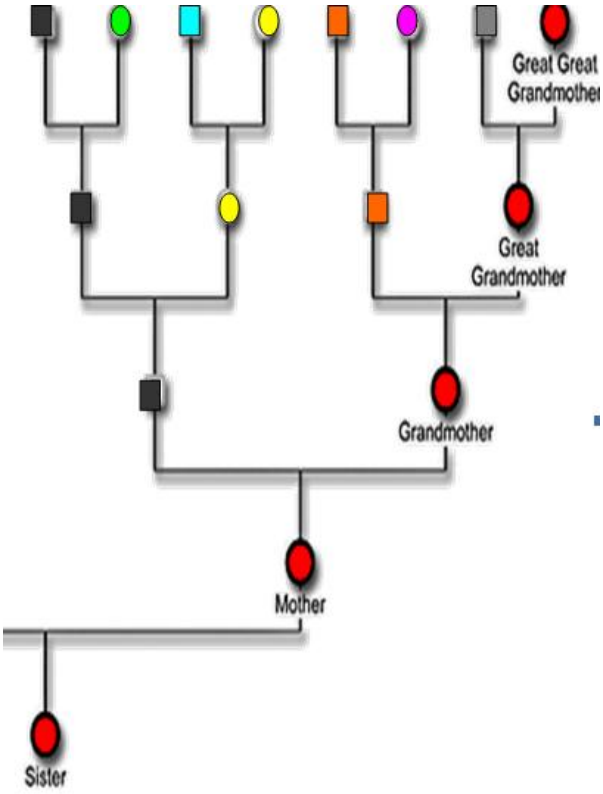

Fig. 2: Maternal Inheritance

Before discussing the various kinds of autosomal tests and what they mean to genealogists, let's review the autosomal inheritance process and how it really works.

Inheritance - Passing of Autosomal DNA from Parents to Children

Autosomal DNA is the DNA contributed by both parents to each child which excludes the Y chromosome and mitochondrial DNA. A female child receives an X chromosome from each parent and in this circumstance, the X chromosome is recombined the same fashion as autosomal DNA. The X chromosome will be discussed in a later section.

Everyone knows that you inherit half of your DNA from your mother and half from your father. While this is true, it does not mean that you receive 25% of your DNA from each grandparent.

While each child does on the average receive 25% from each grandparent, the actual inheritance pattern varies much more than that and each sibling may receive far more, or less, than 25% of their markers from any grandparent.

We don't understand today how inheritance traits are selected to be passed to children. Each parent receives

50% from each of his or her parents, but how this is combined and reduced to the "half" that is passed to each child is unknown. Every individual has 2 chromosomes in each pair, one from each parent, but their DNA recombines to create one "new" chromosome to be passed to each of their children.

Some "groups" of genetic material are inherited together, and you may wind up with more or less genetic material from one of your grandparents. In time, certain genetic "traits" will be lost in some descendants, while not in others. Therefore, you can't figure actual inheritance percentages by using the 50% rule. This means that if your father was 50% Native American, you are not necessarily 25%, genetically speaking. You may receive 10% of his Native genes and your sibling may receive 40%.

The following diagram, The Rainbow Shuffle, (Fig. 3) shows that child 1 receives only one A from Grandpa A but three Bs from Grandma B. In the next generation, child 1 is less likely to pass on the A to their children and more likely to pass on a copy of B. Child 2, by comparison, is more likely to pass on an A than their single copy of B.

Let's use the Duffy Null allele as an example. The Duffy Null allele is found only in African populations, and is therefore an important informative marker to determine African heritage. Currently this marker is found in about 68% of American blacks and in 88-100% of African blacks[12]. This marker could have entered the DNA pedigree chart with a grandmother who carried the allele but had no obvious visible African ancestral traits, or from a parent who might have been born in Africa and is visibly African. The Duffy Null allele, which is just one marker, could have been passed in the inheritance of DNA for many generations, far after any visible physical African traits had disappeared, or it could be one of many African traits passed from parent to child. It is also possible that an individual who is admixed, whether they know it or not, and physically appears to be of African descent, has lost the Duffy Null allele someplace along the line in recombination and transmission.

The relevance of the Duffy Null allele is determined by the number of other "African" markers that appear in high quantity. If there are few other African markers, then African ancestry was likely further back in time. If there are many, then African ancestry was likely more recent. These statistical calculations are how the importance of autosomal markers is determined and how percentages or estimates of ethnicity are calculated.

Any one allele or marker can be lost permanently in any generation. Each child receives one gene from each parent. In the example below, let's say that the mother carried genetic markers A and B, and the father C and D, and D is the Duffy Null allele.

[12] The Duffy Blood Group, NCBI, 2009 http://www.ncbi.nlm.nih.gov/bookshelf/br.fcgi?book=rbcantigen&part=ch09Duffy

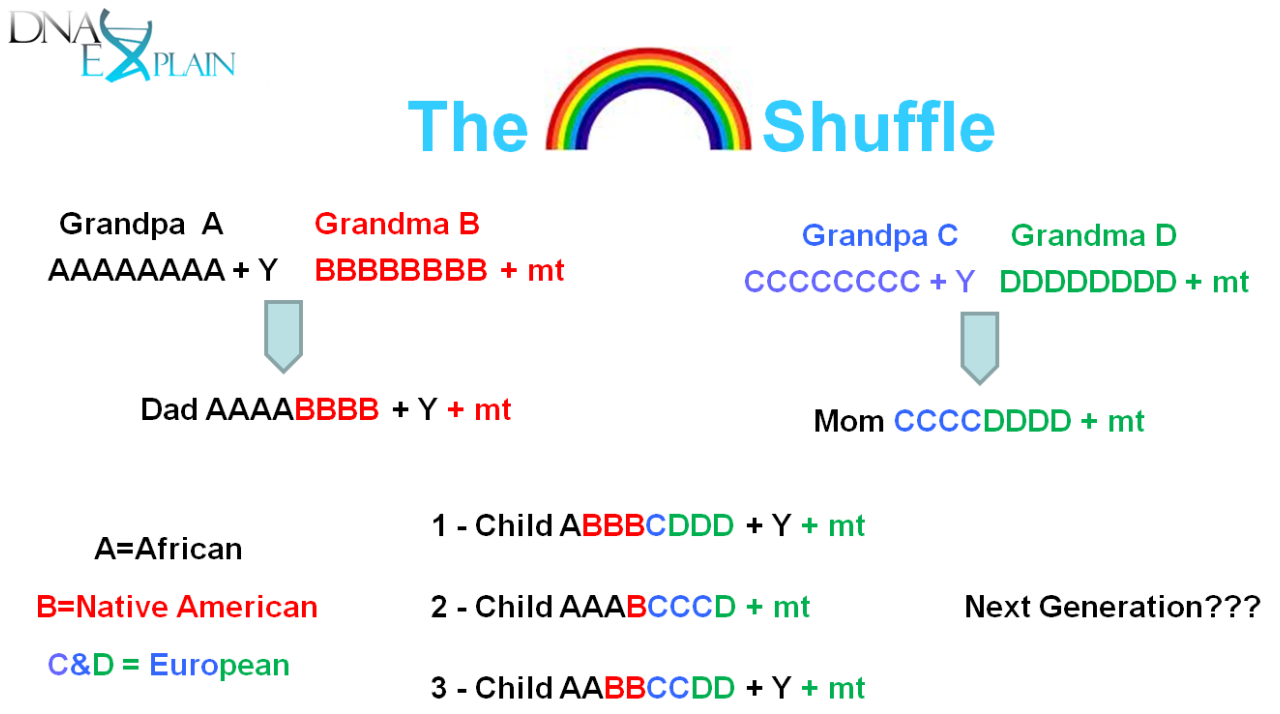

Fig. 3: The Rainbow Shuffle

	Mother		Father	
	|	|	|	|
Markers:	A	B	C	D

Child 1 – A and C
Child 2 – A and D
Child 3 – B and C
Child 4 – B and D

You can see that half the children received the D marker, but each inheritance event was a random recombination of the markers. It is also possible that none of the children would receive the D marker, or all of them would receive it. Statistically speaking, half will receive the marker, but statistics and individual inheritance are two different things. Random recombination is the reason why siblings who take autosomal tests sometimes show significantly different results.

You can also see how a marker that is very old ancestrally, meaning introduced many many generations ago, could be absent in one entire descendant line and present in another line.

From the above examples, we see that we have two variables that we need to deal with when attempting to use autosomal DNA for genealogy.

First, we need to take into consideration inheritance patterns which we can't determine retrospectively without testing several descendant lines. So, in essence, we can only deal with, and test, what we personally carry today as our genetic inheritance.

The second variable is determining population frequency for a particular marker and understanding its significance to us through comparative population genetics.

This is why autosomal testing can give us important hints, but in some cases is considered "unreliable". The results are highly subjective today, but increase in accuracy as more markers are used and research is completed, compiled, published and analyzed.

Pedigree Analysis - What Do I Really Know About My Genealogy?

This question seems odd, but the process of analyzing genetic genealogy, especially autosomal DNA has caused me to look at genealogy from a different perspective and to evaluate what is definitively known using traditional genealogical research methods.

Using a pedigree chart generated by genealogy software[13], a summary spreadsheet was created recording the ancestry contribution percent by location from both parents by determining the contributed percent of the oldest ancestor in each line. Table 1, as an example, shows the "end line" ancestor, the percentage of DNA contributed by that ancestor, and their heritage, if known.

The goal of this methodology is to determine, by geographic location, how much of the participants DNA is positively accounted for, and how much could be of Native ancestry, and in what lines it could occur.

Using the pedigree analysis method described above, our participant's geographic origins total are shown in Table 2. In broader categories, and combining those that are similar, we find the results in Table 3.

Now that we know what the majority ancestry looks like and any minority ancestry that we are aware of, let's take a look at the various types of autosomal testing available.

Low Marker Resolution Tests

The first genealogical autosomal test to enter the marketplace was a test using 71 markers to determine ethnicity percentages. Today, with tests available that scan half a million markers, these older test are grouped together as Low Marker Resolution Tests because they were first generation tests that used comparatively few markers. They are the DNAPrint test, Omnipop which uses the CODIS markers and DNATribes.

DNA Print

In 2003, DNA Print Genomics introduced the DNA Print test, the results of which would tell the participant what percentages of 4 ethnic groups their DNA carried.

The participant's results were reported as:

- Indo-European: 75%
- East-Asian: 15%
- Native American: 10%
- African: 0%

One confusing aspect of this report was the East-Asian component which was higher than the Native American component. The participant has no East Asian heritage. After discussing this situation with the scientists at DNAPrint Genomics, it was determined that the East-

[13] I use Personal Ancestral File, available free from the Church of Jesus Christ of Later Day Saints website, but all genealogy software has the ability to generate a pedigree chart. www.familysearch.org

End Line Ancestor	German	Unknown	English	European	Native	Scotland	Wales	Holland
Jacob Lentz	3.125							
Fredericka Moselman	3.125							
Nabby (poss Curtis)		3.125						
Abraham Estes			0.1954					

Table 1

Geographic Origins	Percent of DNA Contribution	Comments
German	23.0228	
Probably German	0.7813	
Unknown	17.1889	These individuals have surnames but their heritage is unknown.[14]
English	6.2024	
Probably English	4.1022	
European by DNA	6.8362	Surname and DNA are both European, but the European location is unknown.
British Isles	2.344	
Possibly Native American	0.049	
Native American	0.2443	
Scotland	4.1018	
Wales	0.3907	
Holland	14.5511	
Ireland	2.9299	
Probably British	2.5394	
Turkish	.0031	
Probably French	1.7764	
French	4.8349	
Switzerland	.7813	
Total	92.6797	
Total known	75.4908	Total less the unknown category above
Total Unknown	24.5092	Includes unknown category with surnames and all others not accounted for.

Table 2

Geography	Percent
Germany	23.8041
British Isles	22.6104
Holland	14.5511
European by DNA	6.8362
France	6.6113
Switzerland	.7813
Native American	.2933
Turkish	.0031

Table 3

[14] The ethnic heritage of several line could probably be inferred by surname or ethnicity of marriage partner. However, I have avoided the temptation to make inferences within the United States, as the Native or African ancestry may well lie with one of these ancestors. These are in fact the perfect candidates and to eliminate them from consideration by inferring origins would be a disservice.

Asian and the Native American were actually reflecting the same heritage, Native American, and the values should be combined. This made sense given that the American Indians migrated from Asia between 12,000 and 15,000 years ago[15].

DNAPrint version 2.0 utilized 71 Ancestral Informative Markers (AIMs). Version 2.5 of the test introduced some years later referenced 175 markers.

Version 2.0 results were also delivered with a triangle chart that showed the various confidence levels shown as bands. These confidence levels sparked a great deal of heated debate on the Genealogy-DNA Rootsweb list[16]. Some individuals felt that these results were right in line with their known genealogy and others received surprising results in both directions. For our participant, 25% American Indian seemed quite high given what was known about their genealogy.

The black, blue and yellow circles shown in Figure 4 are confidence bands. This means that while the most likely result is the red dot which represents the values provided in the report, the results can indeed fall anyplace within the confidence bands. These values are calculated to be between 2 and 10 times less likely than the red dot, but are possible nonetheless[17]. This variance is known as statistical noise.

Customer satisfaction with the product seemed to often be driven by whether or not the expected or desired result was obtained.

Later versions of this test included bar graphs with the percent graphed and the confidence range displayed in percent as well. Confidence range values were often as high as 15%. Values reported at zero (such as the African value in the example shown) are often given a +5% range. Obviously, one cannot go below zero, so zero values are often shown as "possibly 5%", although they are 2, 5 or 10 times less likely to be 5% than they are to be zero.[18]

Most results typically have a 10% variance, especially the later results. Applying this to the participant results, they could have the ranges of the various ethnic heritages using the 10% variance, as shown in Table 4.

Group	Reported %	As Low As (10%)	As High As (10%)
Indo European	75	65	85
East-Asian	15	5	25
Native American	10	0	20
African	0	0	10

Table 4

Earlier results could potentially have the results listed in Table 5 using 15% variance.

Group	Reported %	As Low As (15%)	As High As (15%)
Indo European	75	60	90
East-Asian	15	0	30
Native American	10	0	25
African	0	0	15

Table 5

According to the documentation provided by DNAPrint, it is not possible that the participant's Native American or East Asian percent is zero, because the confidence band does not reach below the zero point which is the bottom left corner of the triangle.

For consumers, the important information delivered by this product was the actual ethnic percentages. The question became whether this technology was mature enough to deliver a reliable and reasonable product to the consumer? Let's take a look at this question outside of the science of the situation and from the genealogical perspective.

What is Statistical Noise?

In layman's terms, statistical noise is the term we use and the percentage variance given for "slop", or what we don't know or can't account for mathematically. In this case, the variance could be based on variables like the lack of population normalization, a small sample size or other unknown quantities such as variability in the measurements of the occurrence rates of specific allelic values within populations.

To demonstrate these concepts, let's use an example sampling of a population that is supposed to represent people in Michigan. The sampler tests 50 people in Detroit. Questions that give rise to statistical noise are questions like whether the 50 people who were tested in Detroit are actually representative of the entire state of Michigan? Are they representative of even the City of Detroit itself? The answer could be no in both cases

[15] Perego et al, *The Phylogeny of the Four Pan-American MtDNA Haplogroups: Implications for Evolutionary and Disease Studies* (2008 PLoS ONE) and Fagundez et al, *Mitochondrial Population Genomics Supports a Single Pre-Clovis Origin with a Coastal Route for the Peopling of the Americas* (2008, American Journal of Human Genetics)
[16] http://archiver.rootsweb.ancestry.com/cgi-bin/search
[17] This information is documented in the AncestrybyDNA User Manual (no version, no date) delivered with the product.
[18] This information is documented in the AncestrybyDNA User Manual (no version, no date).

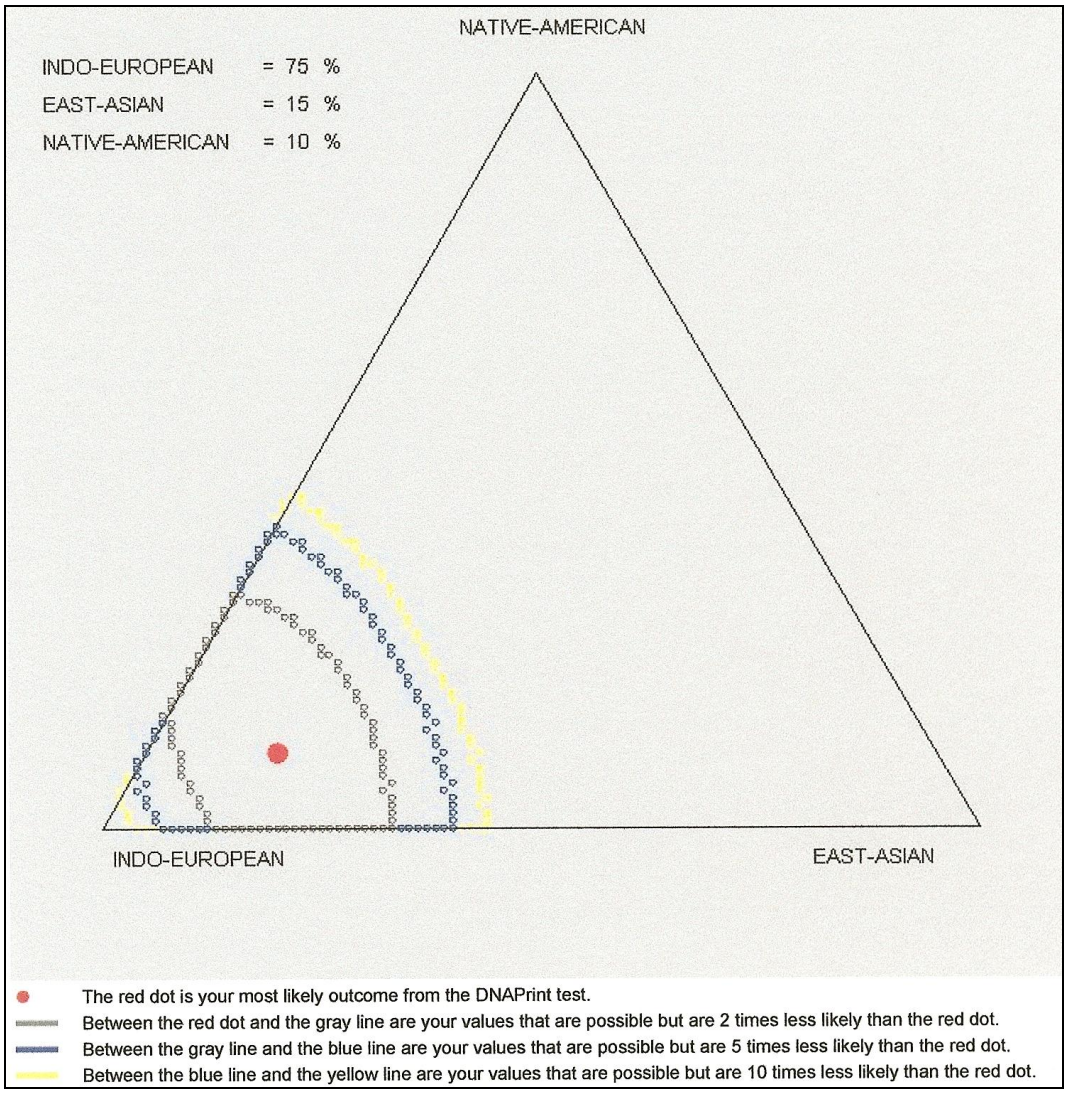

Figure 4: DNAPrint Test Results

depending on which neighborhood the sampler visited. If the sample size is small and localized, the greater the doubt about the relevance of that sample to the larger area being measured, in this case, Michigan, and that doubt has to be translated mathematically into a plus or minus number.

Why Does Sample Size Matter?

Using our illustration where we are sampling in Michigan, a larger sample size gives any DNA (or statistical) anomaly the opportunity to become normalized by virtue of a larger sample size. For example, if 50 people are sampled in Detroit in the Polish ethnic neighborhood, and they happen to all be related within a few generations (a situation unknown to the sampler and possibly also to the participants themselves), one might well obtain a falsely high reading on a few particular DNA locations that are prevalent in the Polish or Polish/Detroit population. Extrapolating this information to apply to all of Michigan would cause faulty conclusions that Michigan people have a particular value at a particular location, when that is not the case.

On the other hand, if 5000 people were sampled from all over Michigan (preferably in proportion to their population distribution), the elevated values for the 50 people from the Detroit Polish neighborhood would become just 1% of the group and their elevated numbers would not skew the data for the entire population of Michigan[19].

[19] This is known as the Bernouli principle which is often referred to as the "law of averages". The Bernouli principle guarantees stable long-

Obviously researchers attempt to deal with this issue, but many times the data they have to work with is data collected by others and published in medical, forensic, research or academic publications and the circumstances surrounding the data collection may not be entirely known. How does this affect genealogy?

Genealogy and Statistical Noise

As a genealogist, the question unable to be answered from the DNA Print test is whether or not the results are meaningful to the participant's research. In the pedigree chart in Figure 5, the percentage of DNA that is contributed by each generational ancestor, on average, is shown in Figure 5.

Most people know their grandparents and some know their great grandparents. If any of those individuals were descended from a specific ethnic group, the oral history would be readily available. As we can see in the chart above, our great grandparents contribute 12.5% each (on average) to our genetic makeup. We can also see that 10% or 15% statistical noise falls on either side of our great-grandparents generation, making minority ancestral percentages greater than 10-15% questionable.

Applying this to our participant's example situation, using the pedigree chart below, 25% Native American would equate to one grandparent or two great grandparents or 4 great-great grandparents. Given that the mother's ancestry was primarily German with only one unknown line (the Lore male line shown in orange in Figure 6), this scenario seemed highly unlikely as the entire Native contribution had to have been contributed by the father's line. Had any two great-grandparents been Native, that situation would certainly have been a known or discussed fact within the family

Unfortunately, this calls into question the credibility of this particular test. Adding or subtracting the statistical noise component only makes things worse, given that the amount of Native ancestry could increase to a total of 40% or decrease to as low as 10%. This range means that the participant could have one great grandparent who was almost entirely Native (10%) or that both of the Estes side grandparents (who contributed an approximate total of 50% of the participants DNA) were nearly entirely Native (40%) which is known to be untrue.

term results for random events. Results are very reliable with many events, and not reliable with few events. This is best illustrated by coin flips where as the number of flips increase, the cumulative frequency reaches 50%, but when few flips are involved, the frequency of heads to tails is seldom equal. The Bernouli principle is applicable to population samples.

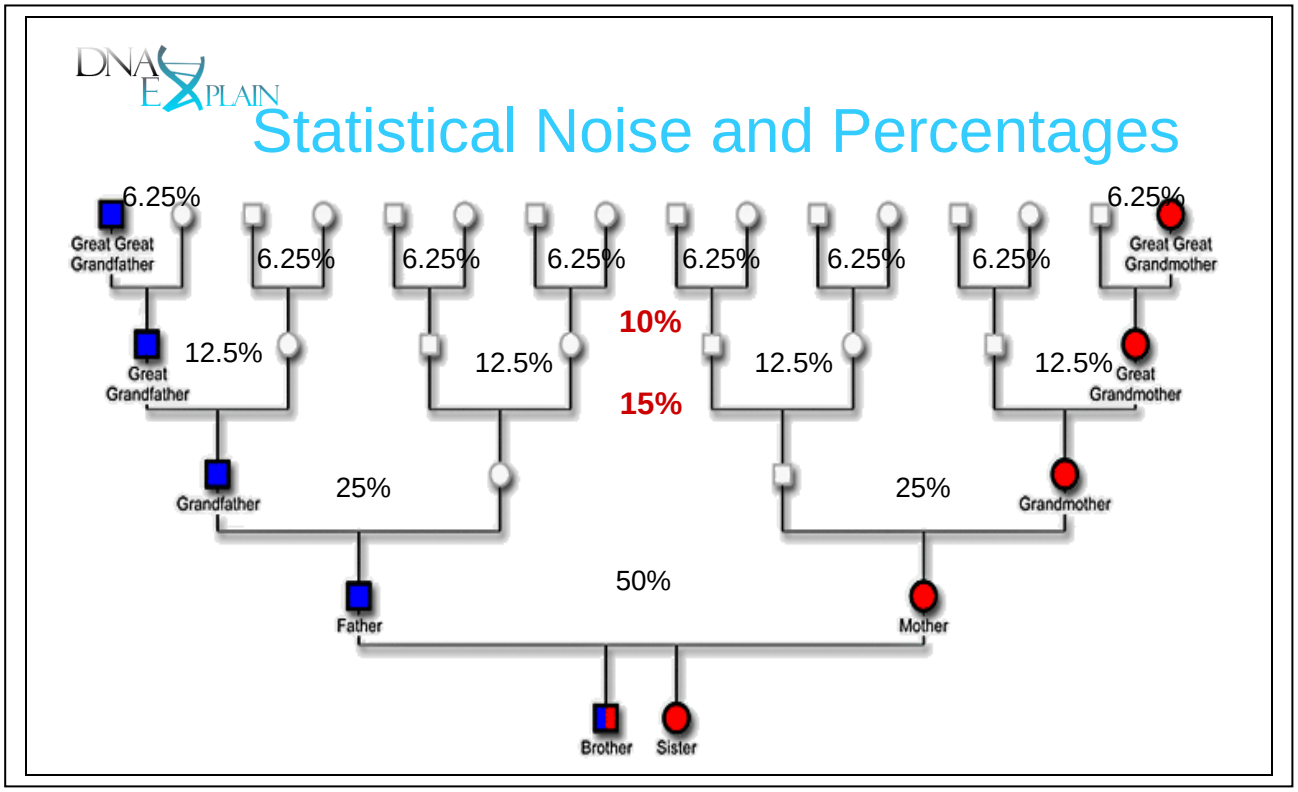

Figure 5: Statistical Noise

.

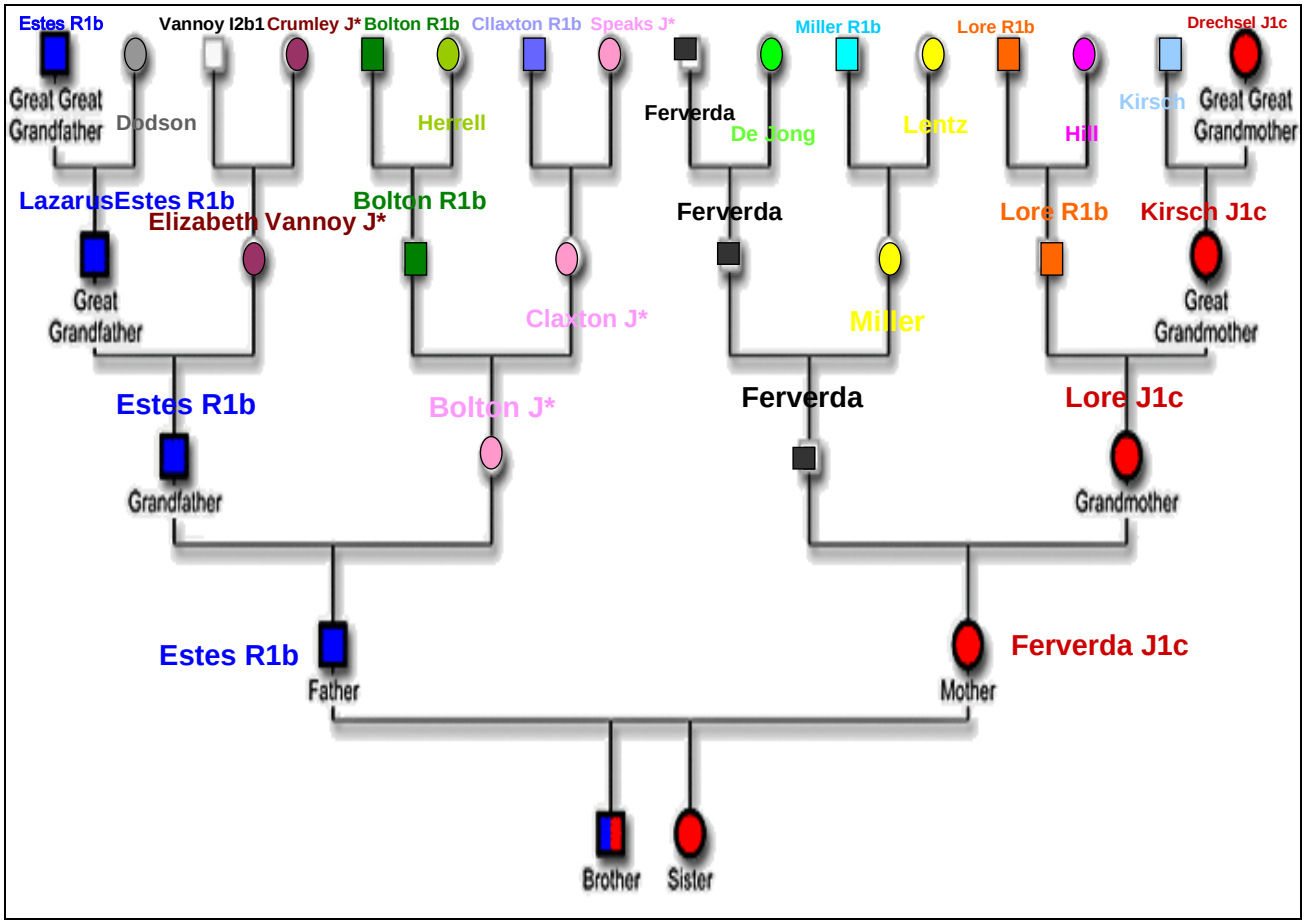

Figure 6: Example Pedigree Chart

Looking at this another way using generation length and history puts this situation into better perspective.

Assuming an average age of 50 for a genealogist, and an average generation length of 30 years, this extrapolates to the following information:

- Current genealogist –born about 1960
- Parents – born about 1930
- Grand-parents – born about 1900 (the genealogist probably knew them)
- Great-grandparents – born about 1870 (the genealogist probably didn't know them, but their parents and grandparents talked about them)
- Great-grandparents – born about 1840 (the genealogist definitely didn't know them, but probably knows who they were genealogically as their grandparents knew them)

By 1870, there were very few, if any, American Indian people from tribes originally located east of the Mississippi who were not admixed with European or African ancestors, although they might have been unaware of that fact if the admixture occurred several generations previously. The least admixed individuals were on reservations by 1870 or living in the west. Some Indians or their descendants living in the east were admixed enough to have eluded or avoided removal in the 1830s, 2 generations earlier.

By the great-grandparents generation, we are now at the 6.25% level, well within the statistical noise range, but also within the timeframe where we should receive at least some fragments of oral history. The 1850 census is beneficial at this point, as we can determine where relevant ancestors lived, if they were on a reservation, and if they were considered anything other than "white", such as black or mulatto, as many admixed American Indians were labeled.

Most people who seek to discover their Native American ancestry are by necessity looking back before the "Trail of Tears", often to the tribes that were exterminated by the colonists, wars and disease before the Revolutionary War[20].

Remnants of those tribes intermarried with whites and free people of color as well as joining the tribes still existent, such as the Cherokees and Creeks who were later removed in the 1830s.

Unfortunately, on the genealogy chart, this takes us back another two generations to ancestors born in 1810 and in 1780. Respectively, we carry an average of 3.125% and 1.56% of their DNA. The next generation back, born in 1750 before the Revolutionary War, we carry less than 1%, on average, of their individual DNA.

It's no wonder that the autosomal tests have such a difficult time finding traces of our Native ancestors. Unfortunately, because of the way DNA is recombined and transmitted generation to generation, we simply can't unlock those secrets easily without focused Y-line and mtDNA testing.

Lazarus Estes and his wife Elizabeth Vannoy are shown in Figure 7 in blue and burgundy. Elizabeth was originally given as the participant's Native ancestor through oral history, but additional information from Elizabeth's mother's family line indicated that Elizabeth's Native ancestor was through her mother, Phoebe Crumley's, line. This new information seemed particularly credible, since this same "Native heritage" story emerged from two independent lines of the family with no knowledge of each other, diverging in Phoebe Crumley's parents' generation by moving to different states. Phoebe Crumley's parents contributed 3.125% DNA each to our participant.

An appropriate mitochondrial DNA candidate to represent Elizabeth Vannoy was eventually found, and Elizabeth Vannoy's mitochondrial DNA is haplogroup J*. Haplogroup J is not a founding Native American mitochondrial haplogroup[21].

Elizabeth's Vannoy genealogy is relatively well proven, leaving little room for a Native ancestor. Therefore, we have focused on her mother's Crumley side where there are women without surnames that have potential to be

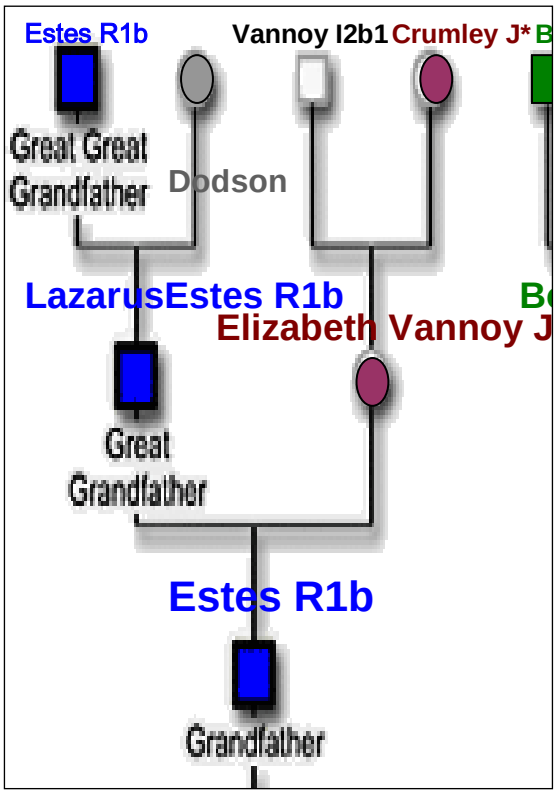

Figure 7: Pedigree Chart

Native based on the time and location, although an appropriate DNA candidate is yet to be identified.

If in fact Elizabeth's grandmother is of Native heritage, she would only have provided 1.56% of the DNA carried by our participant (assuming 50% transmission in every generation), a number that hardly approaches any number in the range of 10-40% or the most likely 25% reported by DNAPrint.

Obviously, the percentage our participant's Native heritage could be and probably is a combination of contributions from several ancestors.

During this discovery process, our participant's mother[22] who is of primarily German heritage also took the DNA Print test and was surprised that the results indicated some level of Native heritage. Her results were:

- Indo European: 91%
- East Asian: 7%
- Native American: 2%
- African: 0%

Given the mother's German heritage, it was initially assumed that her Asian and Native percentages were

[20] *Where Have All the Indians Gone? What We Know and What We Don't about Native American Eastern Seaboard Dispersal, Genealogy and DNA*, by Roberta Estes (2009), Journal of Genetic Genealogy, fall 2009, Vol 5 #2, http://www.jogg.info/52/index.html

[21] Native American mitochondrial founder haplogroups are A, B, C, D and X2a.

[22] Participant's father is deceased

either noise or possibly relics of the Hunic[23] heritage in Germany circa 400 AD or Magyar invasion of Hungary and parts of Germany[24] circa 900 AD. However, her elusive Lore (surname) brick wall fell and the Lore line turns out to be the Acadian Lord family and indeed, there is one female line genetically proven to be Native and another documented as such, but as yet genetically unproven. However, the proven line is 11 generations back in the participant's pedigree chart, contributing less than 1% to her DNA, .1954% actually. Together the two lines combined represent .2443% of her DNA.

There are four possible scenarios that explains these results:

- The Mother had significantly more Native ancestry than we are aware of which would have all been confined to her Acadian line which in total contributed no more than 6.25% of her ancestry. The rest of her ancestry is 19th century German and Dutch.
- The Mother's German ancestry is providing some amount of Asian ancestry.
- "False positive" DNA readings are being received from the DNA Print test.
- All or part of her Native American/East Asian results is statistical noise.

DNA Print Genomics is no longer a functioning company, having been sold and the subsequent company then becoming bankrupt, but the test is now being remarketed by another firm. [25]

As we discussed, DNA Print uses a proprietary set of autosomal markers known as Ancestry Informative Markers (AIMS)[26]. An ancestry-informative marker (AIM) is a set of marker values which exhibits substantially different frequencies between populations from different geographical regions. By using a number of AIMs one can estimate the geographical origins of the ancestors of an individual and ascertain what proportion of ancestry is derived from each geographical region, although for any measure of accuracy, both a very large number of markers would need to be used and the reference populations would need to be very balanced worldwide and substantive in number. Another type of autosomal DNA testing exists as well which uses a different set of markers.

CODIS Markers

CODIS (Combined DNA Index System) markers are a standardized set of 15 autosomal markers[27] established by the FBI. They are used for paternity and siblingship testing and additionally by police departments and forensics labs for identification. The markers employed in these tests are selected specifically to differentiate between people in order to identify them individually, not to find common markers to place them in ethnic groups.

Figure 8 provides an example of what raw Codis test results look like. They are very similar from any lab.

Codis marker testing is available at Family Tree DNA as a stand-alone test and at DNATribes as part of their product offering.

Analysis of Codis Markers

Unless you're using the Codis marker results to determine siblingship or some other personal reason, these numbers are fairly useless genealogically without additional analysis.

There are currently two avenues to analyze Codis results. The first is to use a free tool, OmniPop, created by Brian Burritt of the San Diego police department as a tool designed to differentiate between people, not to compare them for similarities. The second is to purchase the analysis service from DNATribes.

OmniPop

Brian Burritt created the Omnipop spreadsheet from 225 police and forensic articles that had been published and referenced Codis marker information about people from specific populations[28]. How these populations were identified, how the individuals were identified as members of that particular population, and by who are all questions that remain unanswered and probably vary depending on the article and situation in question.

Omnipop provides you with a list of closest matches in ascending order, where the first match is your best match. Our participant's Omnipop version 200.1 results are as shown in Table 6.

The number following the population description in brackets is the article number in Brian Burritt's reference data base, not the number of people in the study or the number of people matched.

[23] http://en.wikipedia.org/wiki/Huns

[24] http://www.geocities.com/egfrothos/magyars/magyars.html also *Medieval Germany* by John M. Jeep

[25] The test, originally called the DNA Print test is currently called the "Ancestry DNA Test" and is currently (July 2010) being sold by DDC, the DNA Diagnostics Center http://www.ancestrybydna.com/ancestry-by-dna.php

[26] http://en.wikipedia.org/wiki/Ancestry-informative_marker

[27] Originally 13 markers were specified, plus Amel for gender identification. Today most labs test for 15 markers. DNATribes uses 21 markers for their enhanced product.

[28] Version 200.1 of Omnipop uses 225 references. An earlier version, 150.5 used only 64.

Location	Mother	Child
CSF1PO	10, 12	10, 12
D2S1338	17, 25	17
D3S1358	17, 18	17, 18
D5S818	11, 12	11, 12
D7S820	8	8, 9
D8S1179	12, 14	12, 13
D13S317	12, 13	13
D16S539	11, 12	11, 12
D18S51	12, 13	12, 20
D19S433	12, 14	14, 15
D21S11	30, 31.2	31, 31.2
FGA	20, 24	20, 24
TH01	6, 9.3	6, 9.3
TPOX	11	8, 11
vWA	17	17, 19

Figure 8: CODIS Marker Results

Remembering that these population descriptions reflect the article from which the data was taken, we are still left with questions. For example, how might Michigan Native Americans vary from other Native Americans? Is the Michigan Native American a tribal member, and if so, which tribe, or is this one individual who self-identified themselves as Native and happened to live or be visiting in Michigan? Is the individual admixed, and if so, how much and with which other groups? For that matter, who categorizes these individuals, the person being arrested, the booking officer, a doctor, and using what criteria? Is this published reference reflective of one individual or several hundred people?

Listed in Table 6 are the participant's top 20 matches. Not shown here is a Lumbee listing. Given that the Lumbee are not a federally recognized tribe, who designated the individual as a Lumbee? People who claim Lumbee descent are known to be highly admixed including European, Native and African ancestry[29], so what exactly does a Lumbee match mean?

Furthermore, entering results into Omnipop in different ways based on the different markers used by various testing agencies produces conflicting results. The DNA test kits used by police agencies are products marketed by different companies. Each product uses a specific set of markers that differs from other similar products. The

Caucasian (64)
FBI Caucasian (1)
RCMP Combined Caucasian (56)
Mexicans (2)
Podlasie (NE Poland) (50)
Belgian (99)
Norwegian (224)
Azores (82)
Michigan Nat.Am. (2)
PBSO Caucasian (4)
Swiss Caucasian (3)
NCSBI Caucasian (4)
Bhumihar Brahmin (India) (72)
ABI Caucasian (14)
PC/BT Caucasian (4)
ABI-ID Caucasian (23)
Florida Caucasian (2)
Scottish (11)
Alabama Caucasian (2)
Southern Spain (Andalusia) (9)
Portuguese (6)

Table 6

results in the articles are reported using either the Cofiler or Profiler standard product markers which are different subsets of the entire Codis set. Some articles report the entire Codis set of markers. The Omnipop spreadsheet is programmed to report only on either the full set of markers or the two subsets, without overlap between them. This means that they only report on complete data sets and their data does not overlap. Therefore, if a particular marker is used in both tests, the data is only

[29] *Where Have All the Indians Gone? What We Know and What We Don't about Native American Eastern Seaboard Dispersal, Genealogy and DNA*, by Roberta Estes (2009), Journal of Genetic Genealogy, fall 2009, Vol 5 #2, http://www.jogg.info/52/index.html and http://www.huxford.com/Genetics_Lumbee_Results_YDNA/Lumbee_Results_YDNA.htm

pulled from the subset you are using (such as Cofiler), and not the data from the second subset (Profiler).[30] The Profiler data uses only 9 markers and references 202 populations in the data base. The Cofiler data utilizes only 6 markers and references 120 populations in the data base. Using all of the markers in Omnipop references 120 populations.

The results in Table 7 are produced using all of the markers, not just the Codis subset. Fewer populations are listed because these are the only populations having data in the full marker subset of data.

The next set of results, shown in Table 8, is using the Profiler subset of markers.

Which of these three methods of using OmniPop is right and which is wrong? The answer is that neither is right or wrong, and the differing results are a function of the different products (which make use of different markers) used by the various agencies and the reported ethnic or geographical heritage of the individuals who were tested, the results of which were then reported in the reference literature.

Shown in Figure 9 are the three sets of results side by side, with the matches color coded. The only result that was consistent in both appearance and ranking between the three is Caucasian, which in our participant's case is visually evident and needs no DNA testing to ascertain majority ancestry. Belgian also appears on all 3 lists, but in significantly different ranking order. Belgian is not a known genealogical origin for the participant whose majority ancestry is German (23%), British Isles (22%), Dutch (14%) and French (6%).

Caucasian (64)
Podlasie (NE Poland) (50)
Belgian (99)
ABI-ID Caucasian (23)
Serbian (157)
Byelorussian (163)
Hispanic (64)
Venezuelan (124)
Kosovo Albanian (155)
ABI-ID Hispanic (23)
ABI-ID Minnesota Native American (23)
ABI-ID African American (23)
African American (64)

Table 7

Caucasian (64)
Andalusians (Spain) (6)
Michigan Nat.Am. (2)
Minorcan (Spain) (34)
Garo (India) (19)
FBI Caucasian (1)
Canadian Caucasians (2)
RCMP Combined Caucasian (56)
Mexicans (2)
Bhumihar Brahmin (India) (72)
Azores (82)
Norwegian (224)
Northern Portugal (41)
Golla (India) (18)
Belgian (99)
Greek Cypriot (37)
Indiana Caucasians (21)
Saharawis (North Africa) (31)
S. Paulo (Brazil) (22)
Portuguese (Centre) (66)
ABI-ID Caucasian (23)

Table 8

Brian's comment regarding genealogists using OmniPop for genealogical comparisons is that they were using a tool not created for this purpose and were over-analyzing the results[31].

Omnipop is not being updated as new papers are released, so the tool is "as is".

My findings relative to Omnipop are that individuals, including some who analyze results for others, tend to select the result that best fits the desired outcome.

OmniPop and DNATribes both use compiled published data, yet their results are significantly different.

DNATribes

DNATribes also uses the Codis markers, but they use a proprietary analysis tool instead of Omnipop[32]. DNATribes also offers an enhanced 21 marker test.

DNATribes has been compiling population data on these genetic markers for several years now and compares Codis markers with their data base. Because DNATribes is a private company, we don't know much about their

[30] Dr. John Butler, National Institute of Standards and Technology, *Statistics and Population Genetics* (2006) http://www.cstl.nist.gov/strbase/pub_pres/NJSP2006_Statistics.pdf

[31] http://archiver.rootsweb.ancestry.com/th/read/GENEALOGY-DNA/2007-03/1173830117

[32] It appears both from Brian Burritt's comments and the 2006 DNATribes results that they originally used Omnipop or a similar tool. The exact tool they use today is unknown.

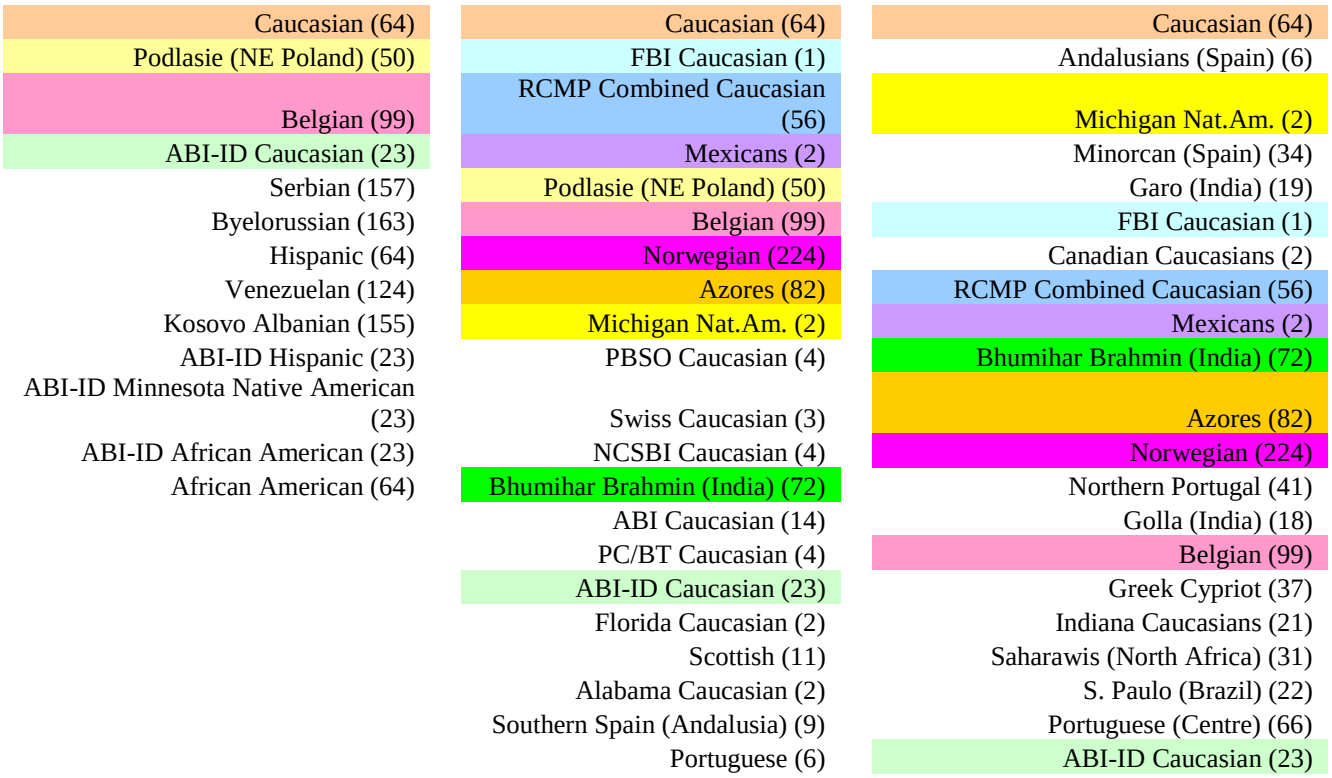

Figure 9: OmniPop Results

population data, whether it's widely representative of the world population distribution and whether it has been normalized or not.

Tribes' early population tables did not include data from the British Isles (neither did Omnipop), so their results were highly skewed towards other world populations.

Our participant had their DNA analyzed by DNATribes when they opened for business in January 2006 and again in March of 2009.

Tribes provides three sets of match results[33]. The first is a Native Population Match which are the participants' "top 20 matches in a data base of 652 native populations that have experienced minimal movement and admixture in the modern history (approximately the past 500 years). Native Population matches identify populations where your DNA is the most common reflecting deep ancestral origins".

The second set of match data is the "Global Population Match which is the top 20 matches in a data base of 896 global populations including native peoples as well as diaspora groups that expanded from their homelands and sometimes admixed with other populations in recent history. This distribution matches your closest genetic relatives today."

The third is the "World Region Match which represents the most comprehensive portion of your genetic ancestry analysis. These regions are the product of long term patterns of interactions between peoples within major geographic and cultural zones over hundreds and often thousands of years. World regions provide a broader, more general view of where your genetic ancestry is found among the major regions of the world."

The 2006 and 2009 comparison shown in Figures 10 and 11 are of the participant's Native Matches which represent their most ancient ancestry.

The results were significantly different. Given the similar naming conventions of the 2006 results with Omnipop's results, I suspect that their original product was based on Omnipop or a similar tool. Their 2009 results when compared with the 2006 results only share one country, Italy, and it appears to be from a different study since the geographic names are different. I question how in 2006 the best match was in Italy and in 2009, it's in Poland when the participant has no proven

[33] Description of the three match categories are taken from the DNATribes General Introduction to Results provided to customers with the delivery of their results (3-11-2009).

or suspected genealogy in either location. Germany, the Netherlands, England, and Scotland make sense, but Italy and Poland as the best match do not.

Does cumulative autosomal recombination over multiple generations resulting in autosomal convergence cause matches to inappropriate non-admixed or lesser-admixed groups? In this case, the participant is highly admixed with multiple European ancestors, plus a few of Native and probably a few of African ancestry as well. Does the DNA "soup" become so highly admixed that genes appear to be Polish or Italian when in fact they are simply a result of extensive admixture? Our participant's single largest European contributor is, by far, Germany at 23% followed by the British Isles at 22%.

2006	2009
Italian	Podlasie, Northeast Pc
Turkish	Austria
Greek Cypriot	Northeast Spain
Sicilian (Italy)	North and Central Pola
Portuguese	Veneto, Italy
Turkish	Norway
Swiss	Iceland
Greek Cypriot	Dundee, Scotland
Portuguese (Central Portugal)	Denmark
Belgian	Slovenia
	Central Poland
	Basque Country, Spair
	Sweden
	Vienna, Austria
	Budapest, Hungard
	Norway
	Netherlands
	Northern Italy
	Southern Russia
	Austria

In 2006, the Global Matches were exactly the same as the Native Matches, but the 2009 matches vary significantly from the 2006 matches.

In 2006, DNATribes provided a Continent Match that is similar to the 2009 World Region match and it showed the participant's ancestry primarily from Europe but with a small (ambiguous) amount of Native American and East/South Asian. SubSaharan African was zero.

In 2009, the Global Population match was similar to the Native match, but not identical, as shown in Figure 12.

In addition to graphs showing results, participants also receive a map that shows diaspora matches for both the Native and World Population matches, as shown in Figure 13. I'm only showing the Native Matches as the maps are very similar.

The DNATribes World Region Matches are shown in Figure 14. These were similar to the Continent matches in 2006, but more regions are represented.

Tribes also offers more granular tests that break down results by continent. Our participant does have a few African matches that they did not have in 2006, but they lost their Native American matches entirely in the 2009 results, although 50 different tribes or groups are individually listed.

Their African matches are as shown in Figure 15.

Several other tribes are shown with zero values, but of note, the Biaka, Mbuti, and Bantu are listed as well as Kenya. These zero values are in contrast to tribes with similarities reported in the participant's results by deCODEme.

Lastly, DNATribes offers a Europa breakdown by region which was quite interesting, especially since the participant's best matches in 2009 were in Poland, Austria and Spain, but the Europa breakdown delivered at the same time is inconsistent, showing their best matches as Germanic, Spanish and Polish, as shown in Figure 16.

I suspect that the primary difference between 2006 and 2009 is that DNATribes has amassed a much better database, but I'm left wondering what happened to the participant's Native American results, how they disappeared entirely, and if the participant will receive entirely different results again were the test to be repeated in another 3 years. I do feel much more confident with more than 900 populations represented, although the number of samples, quality control and other questions already discussed remain.

While the challenges inherent with autosomal testing using either the free OmniPop tool (or derivatives) or Tribes' services are obvious, there is one autosomal test that is definitive.

D9S919 Autosomal Allele

The D9S919[34] allele does not fall into the Low Marker Resolution Test category. It is a standalone test of a single SNP, the results of which are definitive.

[34] The D9S919 marker is the same as marker D9S1120 as published in the 2007/2008 papers as referenced. The D9S1120 label was officially depreciated as the D9S919 marker name was already in use, per Thomas Krahn at Family Tree DNA.

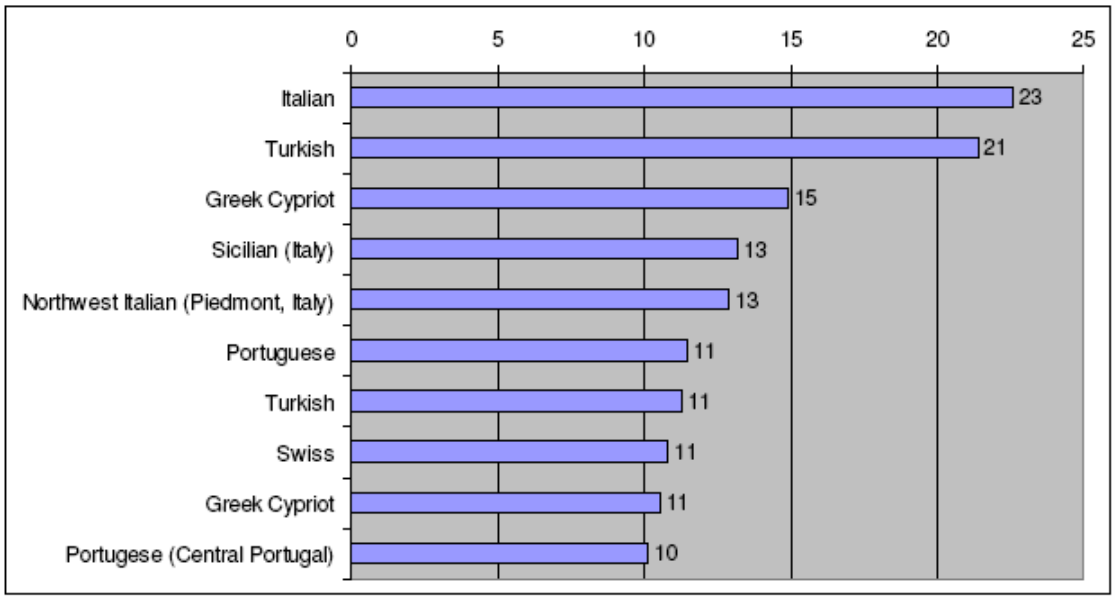

Figure 10: 2006 Native Matches

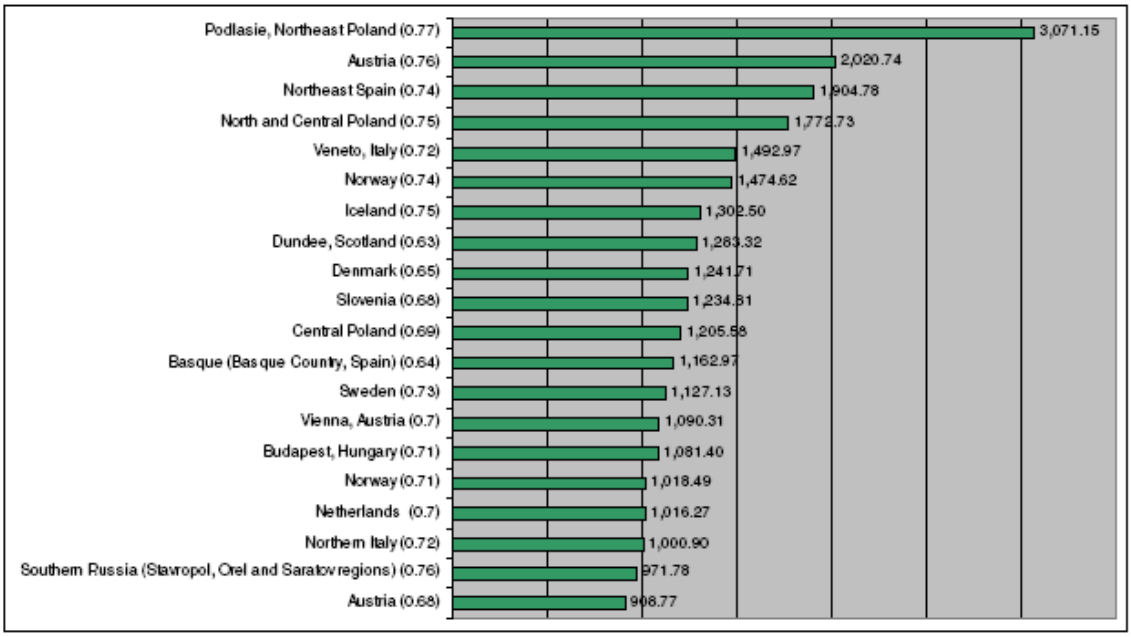

Figure 11: 2009 Native Matches

A paper was published in 2007[35] that indicated that about 30% of the Native Americans tested carry a specific value range for this particular autosomal marker[36]. These values are not known to occur in other populations. This is the only marker value to occur exclusively in the Native American population making this particular marker extremely useful in determining whether an individual carries Native American admixture.

[35] *A private allele ubiquitous in the Americas* by Schroeder et al, Biol Lett. 2007 April 22; 3(2): 218–223.

[36] C. Phillips, et al., D9S1120, a simple STR with a common Native American-specific allele: Forensic optimization, locus characterization and allele frequency studies, Forensic Sci. Int. Gene. (2008), doi:10.1016/j.fsigen.2008.07.002 and Haplotypic background of a private allele at high frequency n the Americas, Schroeder et al, Mol Biol Evol 26: 995-1016

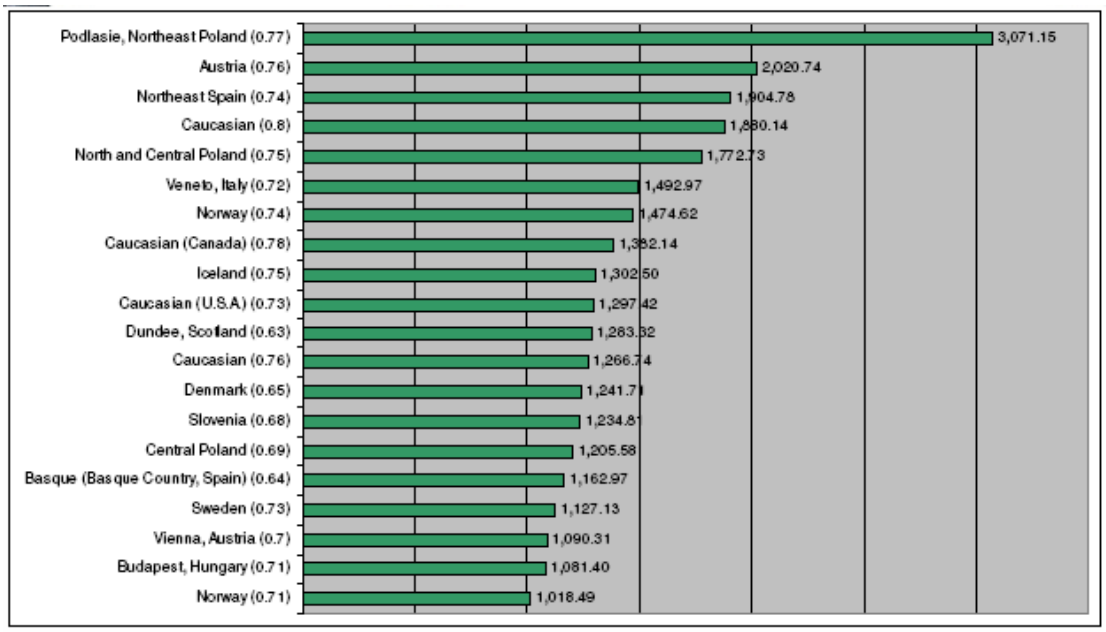

Figure 12: 2009 Global Population match

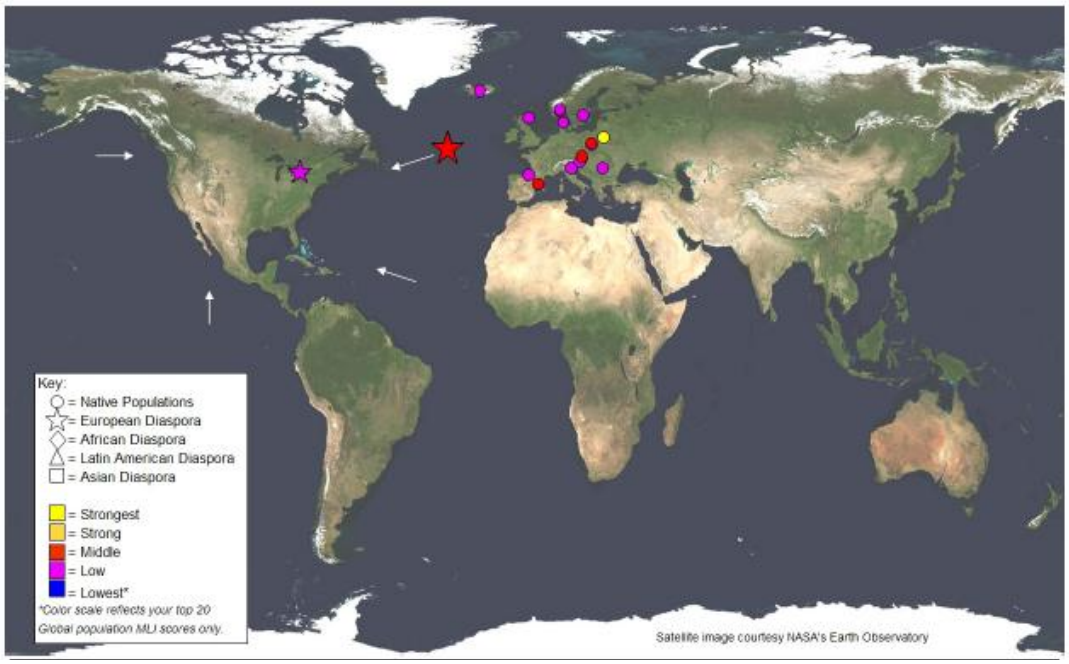

Figure 13: Native and World Population matches

A value of 9-10 confirms Native admixture, but a value of anything else does NOT disprove Native admixture. Our participant's values were not 9-10, which neither confirms nor eliminates Native Ancestry.

This test is only available at Family Tree DNA for existing customers.

High Resolution Array Tests

The next category of autosomal tests is new generation of tests that use a specialized chip that allows the rapid scanning and sequencing of over half a million DNA locations. The three companies offering genetic genealogy products in this new field are 23andMe, deCode genetics with their deCodeme product and Family Tree DNA with their Family Finder offering.

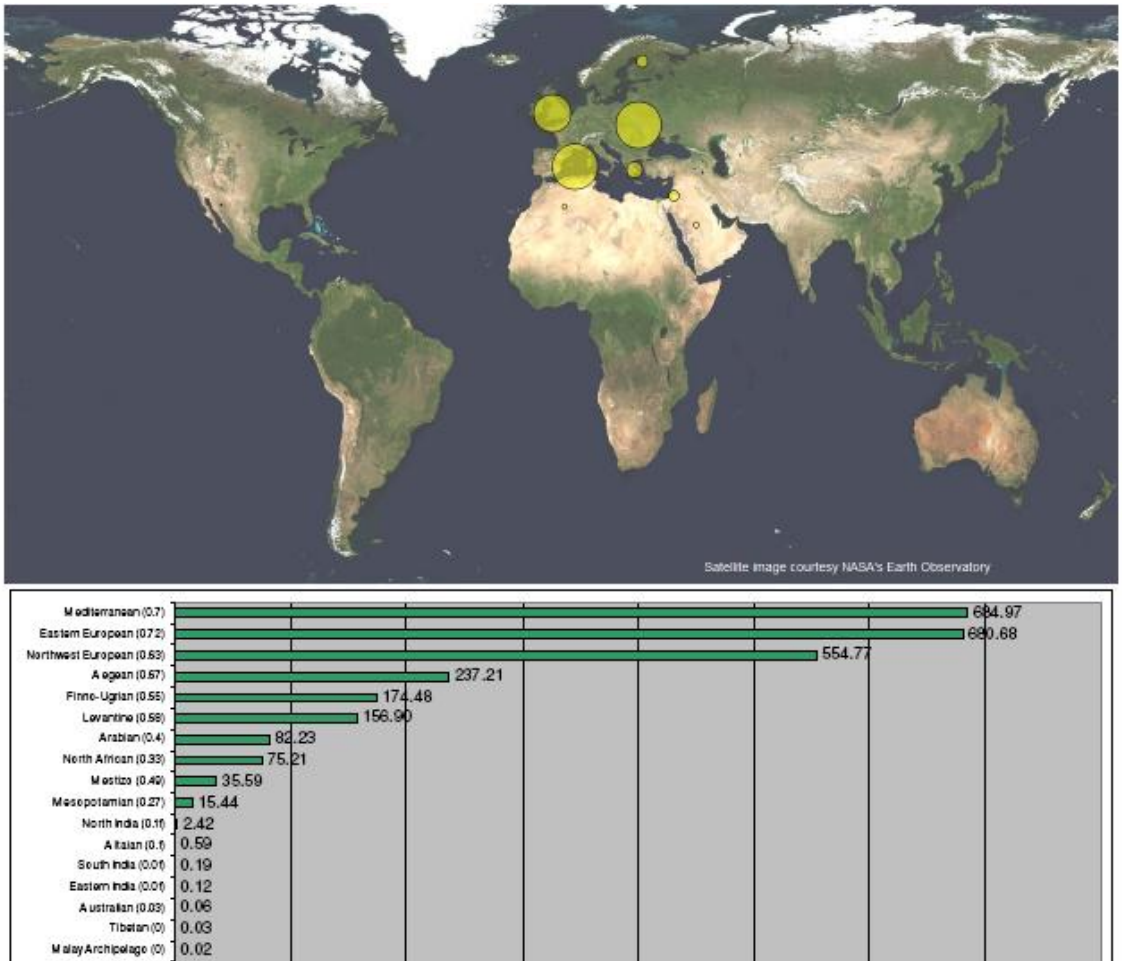

Figure 14: World Region Matches

Sample Name (TribeScore)	MLI Score
Sudan (0.03)	0.75
Sudan (0.03)	0.45
Guinea-Bissau (0.01)	0.21
Somalia (0)	0.01
Guinea-Bissau (0)	0.01

Figure 15: African Matches

23andMe Ancestry Testing

A recent entry into the field of consumer genetic testing is the firm 23andMe. Their primary focus is on medical testing and conversational"[37] genetic information. Later, they added genealogical aspects of their products which includes ethnicity percentages, haplogroup assignments and a product called Relative Finder which matches people with their close and distant relatives. Although Relative Finder is certainly interesting and does use autosomal DNA, it's not relevant to the discovery of minority ancestry.

[37] Conversational items include things like eye color, baldness, bitter taste and alcohol flush reaction. They are interesting but not terribly useful.

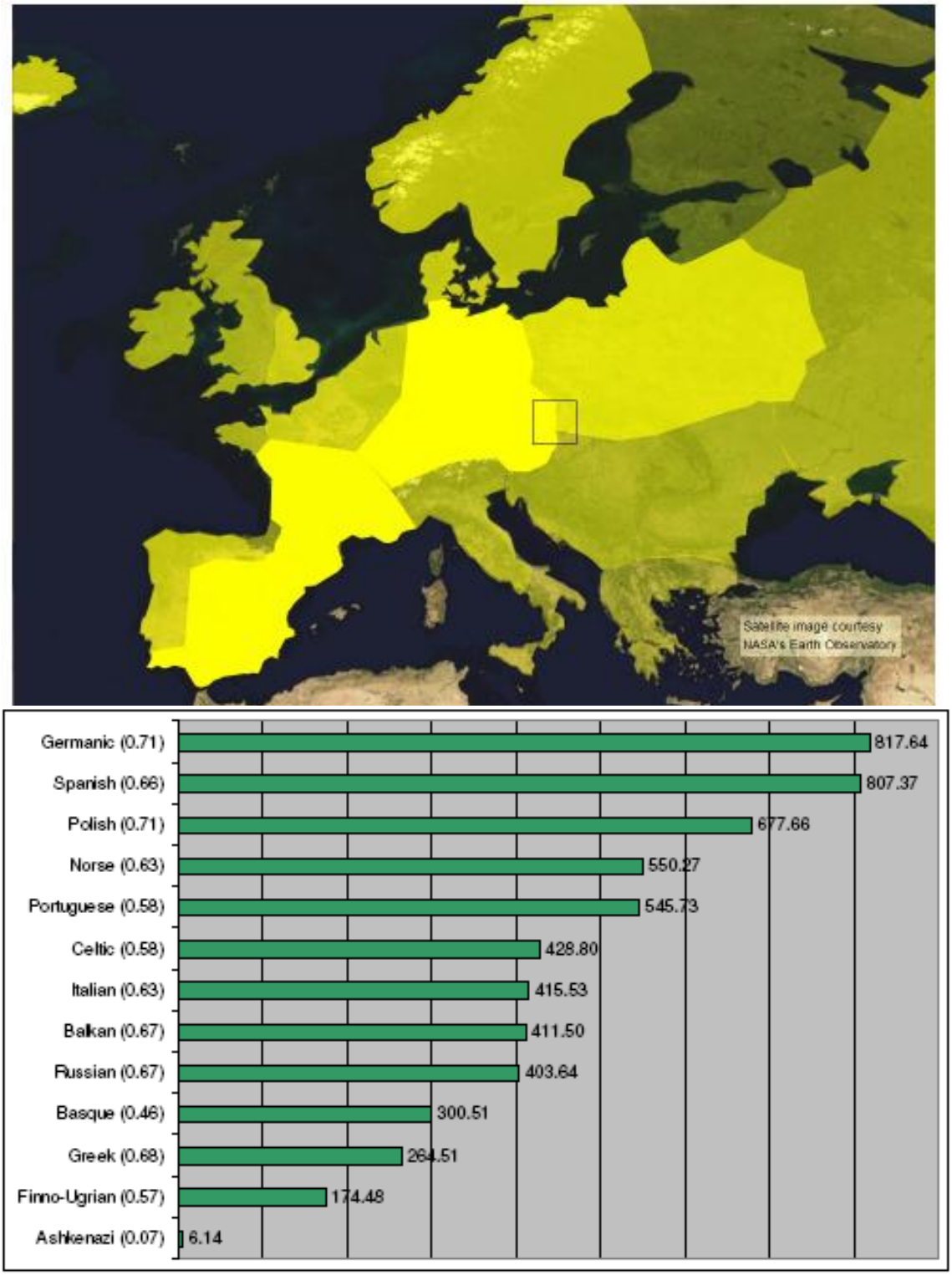

Figure 16: Europa Matches

The haplogroup assignments provided as part of their testing can be used to determine Native heritage, but it is not equivalent to traditional genetic genealogy testing, as there is no Y-line STR marker testing[38] or results returned, no mitochondrial results returned, no matching with others by surname (although they do have a search function), no projects, no mitochondrial full sequence, no mtDNA insertion or deletion testing and sometimes marginal haplogroup assignments. These areas are not their focus.

What they do offer is a wide array of testing on 580,000 locations on your genome, some of which provide ancestry information. You can now order only the genealogical testing without the medical tests, although that was not always the case. However, to have access to your raw data file, you must order the complete test, including the medical information.

In addition to the Y-Line haplogroup assignment (for males) and the mtDNA haplogroup assignment, they also provide consumers with an estimation of their percentage of ancestry, called Ancestry Painting, although they only include European, Asian and African. Unfortunately their sample size is small, less than 50 individuals in four representative populations, which is much too small to be reliably extrapolated to all populations[39]. Furthermore, Native American has to be inferred from Asian which is determined by a group of 45 individuals from Bejing, China.

23andMe introduced a feature called Native Ancestry Finder that evaluates your mtDNA and Y-line haplogroups, plus your percent of Asian heritage and tells you whether or not you're likely to have Native Ancestry. For our participant, it says Native Heritage is unlikely, but that it's possible beyond 5 generations. Given our participant's ancestral findings at their lab of 99% European and less than 1% Asian, their analysis is not incorrect, but their evaluation adds nothing that is not immediately discerned by looking at the haplogroup and Chromosome View ethnic percentages information.

The participant's 23andMe Ancestry results are shown in Figure 17.

According to 23andMe, our participant has Asian heritage only in one relatively large block on chromosome 1.

They also provide a map that shows majority ancestry, as shown in Figure 18.

A recent addition to their features is the Ancestry Finder function which is based on the results of other participants whose grandparents' heritage is confined to one country, and who have segments of DNA that matches the participant's, as shown in Figure 19.

Not shown on the "Country" legend (would be shown if one scrolled down), but plotted graphically on the chromosomes are Germany at 3%, France, Denmark, Slovenia, Greece and the Czech Republic all at 2%.

23andMe's medical/conversational genetic results are being omitted as they are irrelevant to genetic genealogy.

One benefit of the 23andMe test is that your raw data is available to browse or download. This is not useful to most genealogists, as advanced knowledge is required to be able to use or understand this data. However, raw data can be contributed to multiple individuals who are engaged in genetics and genealogy research. As this field develops, so will additional analysis tools.

deCode genetics

DeCode genetics from Iceland is another entrant into the consumer genetics arena about the same time as 23andMe with a similar product, deCODEme. DeCode genetics is a biomedical company well known for their research into heart disease and related genetics. Unfortunately, with the Icelandic government's collapse following the banking industry crisis (2008/2009), deCode Genetics is now in bankruptcy, but is still functioning.

The ancestry portion of their offering is only available if you purchase the entire deCodeme testing package[40]. Their package is similar to 23andMe in that they offer primarily medical informational testing.

The participant's deCODEme ancestry results are particularly interesting. They display the entire X chromosome, information that 23andMe does not provide. In the case of our particiant, this is a critical piece of information as their "Asian" ancestry is pronounced on the X chromosome. The X chromosome has a particular inheritance pattern and this limits the possibility of who, on the pedigree chart, contributed that "Asian" DNA. We'll discuss the X chromosome separately.

[38] The STR (Short Tandem Repeats) are the markers that cumulatively comprise the traditionbal Y-line genetic genealogy tests offered by several testing companies.
http://en.wikipedia.org/wiki/Genealogical_DNA_test

[39] 23andMe uses the HapMap reference populations which consists of only four sets: Utah Americans of Northern and Western European descent (30 adult and both parent trios), Japanese from Tokyo (44 individuals), Northern Han Chinese from Beijing (45 individuals), and Yoruba from Nigeria (30 adult and both parent trios).

[40] This offering is currently priced at $2000, 4 times that of the 23andMe product at $499, the most similar test and approximately 7 times that of the Family Tree DNA Family Finder test at $289 which provides genealogical information, but no medical information.

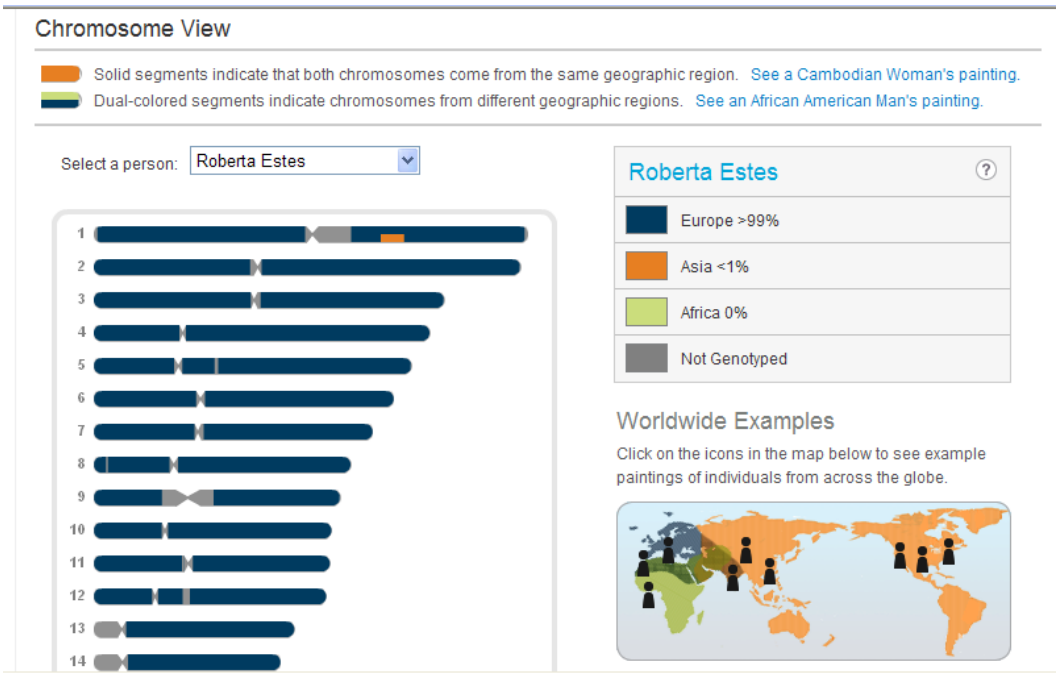

Figure 17: 23andMe Ancestry Results

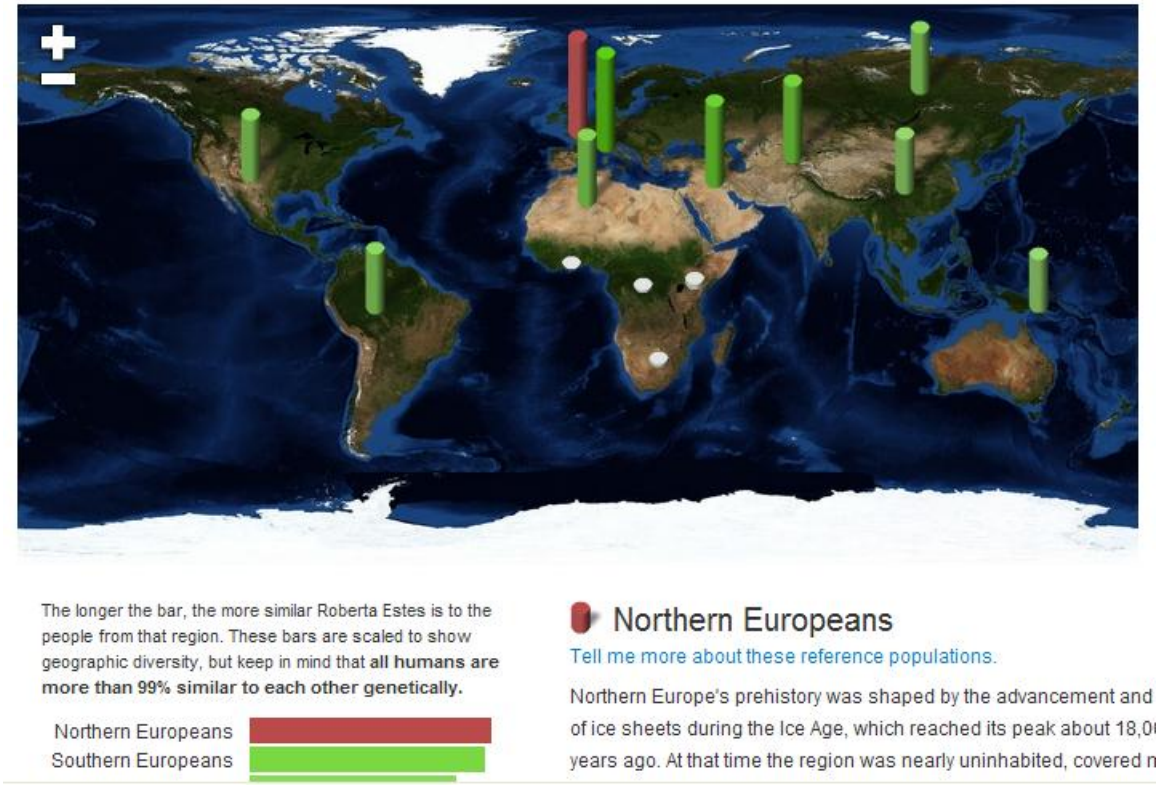

Figure 18: Majority Ancestry

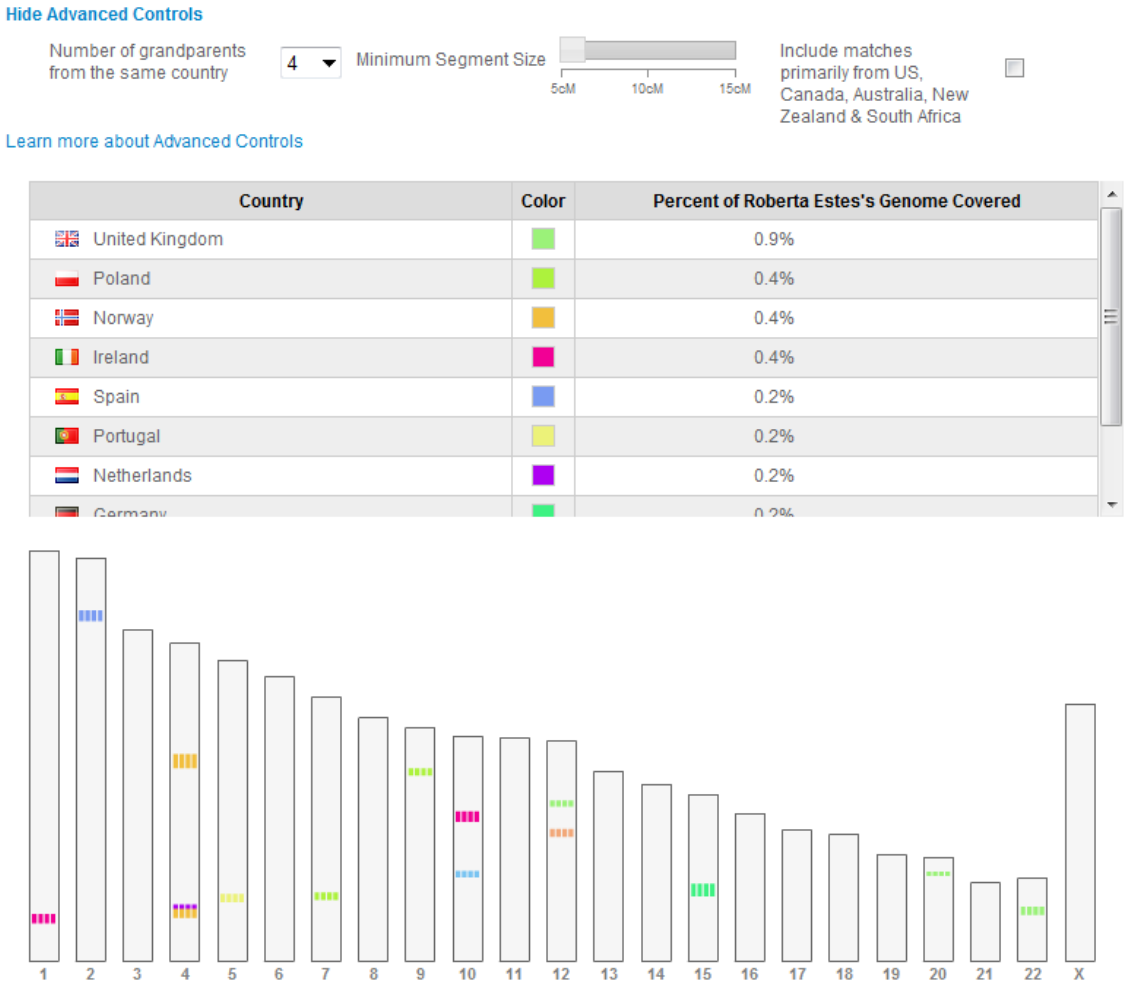

Figure 19: Country Matches

Unfortunately, deCode doesn't separate Native American from Asian either, so one must again extrapolate. They too use the HapMap references to derive ethnic percentages, although they indicate that they use additional reference data for about 1200 individuals for data displayed on the Continents tab.

Another very interesting aspect of these results is African admixture. Our participant had suspected this admixture for many years and had photographic and genealogical hints, but never any documented proof and was surprised when African did not show up in the original DNA Print results.

Not only does deCode show African admixture, on the continent tabs, they show the breakdown by tribe, as shown in Figure 21.

As shown in Figure 21, the participant's strongest African match is with the Bantu Tribe in Kenya, followed the Madenka and Yoruba.

Somewhat more confusing is the European continent tab which shows the participant's strongest matches in Iceland and the Orkneys in Figure 22. This would appear to be a case of skewed population data, meaning that they have more matches there because deCode has more population data from those areas as opposed to other European or world regions.

DeCODEme also provides a genome browser and the ability to download your raw data.

The X Chromosome

The X chromosome is a special case. In men, the X chromosome is inherited entirely from the mother with no admixture from the father. Women receive an X chromosome from both of their parents. However, the inheritance pattern for the X chromosome is dramatically different for males and females.

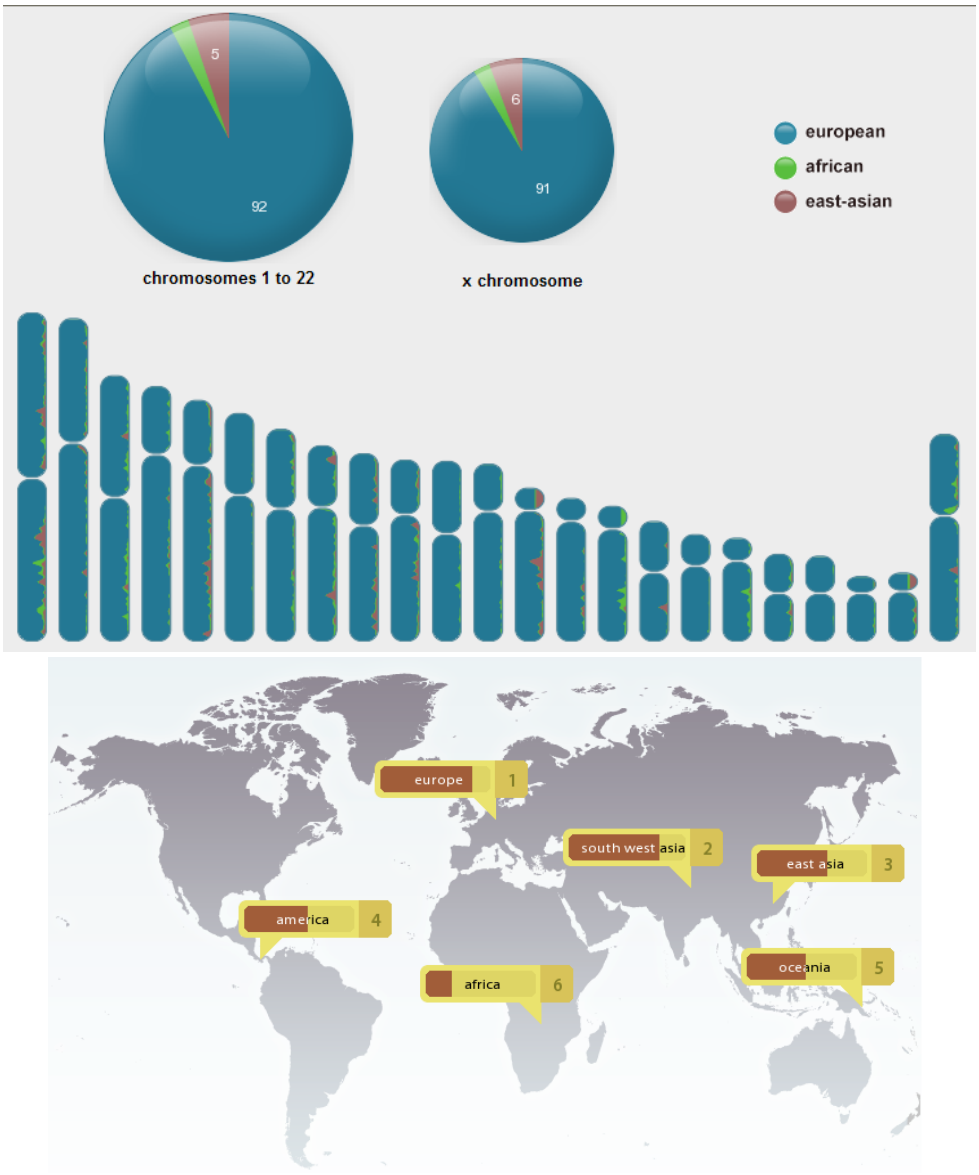

Figure 20: deCODEme Results

Blaine Bettinger posted color coded X chromosomal inheritance pedigree charts on his blog for both males and females[41].

The male's chart is shown in Figure 23 with the pink squares being X chromosome contributors.

The chart in Figure 24 shows the same inheritance pattern but for women who inherit an X chromosome from both their mother and father.

The participant's X chromosome shown in Figure 20 using the deCODEme results shows a significant amount of Asian DNA. The X chromosomal inheritance chart was plotted over the participant's pedigree chart in Figure 25 to determine who might be a candidate for Asian (i.e. extrapolated Native American) and African ancestry. The results were not what was expected and have proven very useful in terms of eliminating some possibilities and providing a tool to focus on others.

Using the X Chromosomal chart in combination with genealogy, we can immediately eliminate a few lines. On the mother's side, the German lines are completely eliminated. They are soundly back in Germany and are not candidates for American Indian or African ancestry.

[41] Bettinger, Blaine, *X Chromosome Charts*, The Genetic Genealogist, http://www.thegeneticgenealogist.com

Figure 21: African matches

Figure 22: European Matches

This leaves only three individuals on the mother's side as candidates for Native ancestry.

5 – Naby (probably short for Abigail), last name unknown but may be Curtis, born in Connecticut in about 1793.

7 – Capt. Samuel Mitchell, born probably about 1800, possibly in Kittery, Maine or possibly in Europe, mother unknown. This line is probably eliminated.

8 – Captain Mitchell's wife, Elizabeth, last name unknown

Using the pedigree chart, we narrowed the mother's side from 21 possible slots to 5 with one more probably eliminated. Of these, mitochondrial DNA sampling of the descendants of the two women whose last name is unknown would produce the answer to the question of maternal Native or African ancestry.

The father's side is more complex because many of his ancestors immigrated in the colonial era. Candidates for Native ancestry are as follows:

20 – Mary, wife of John Harrold (Herrald, later Harrell), born about 1750, died in 1826 in Wilkes County, NC. She was rumored to have been Irish.

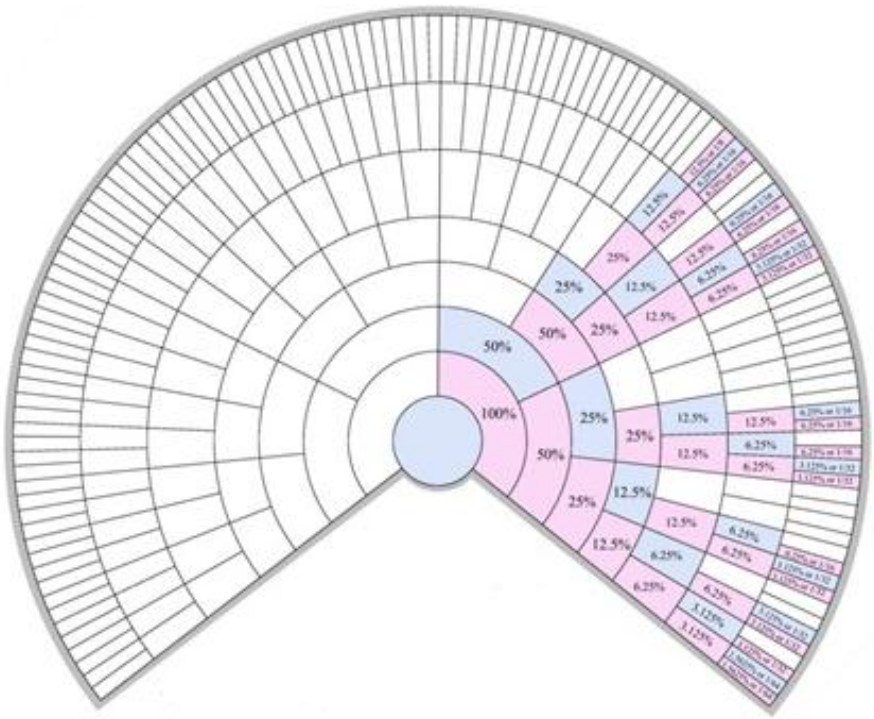

Figure 23: X Chromosomal Inheritance - Males

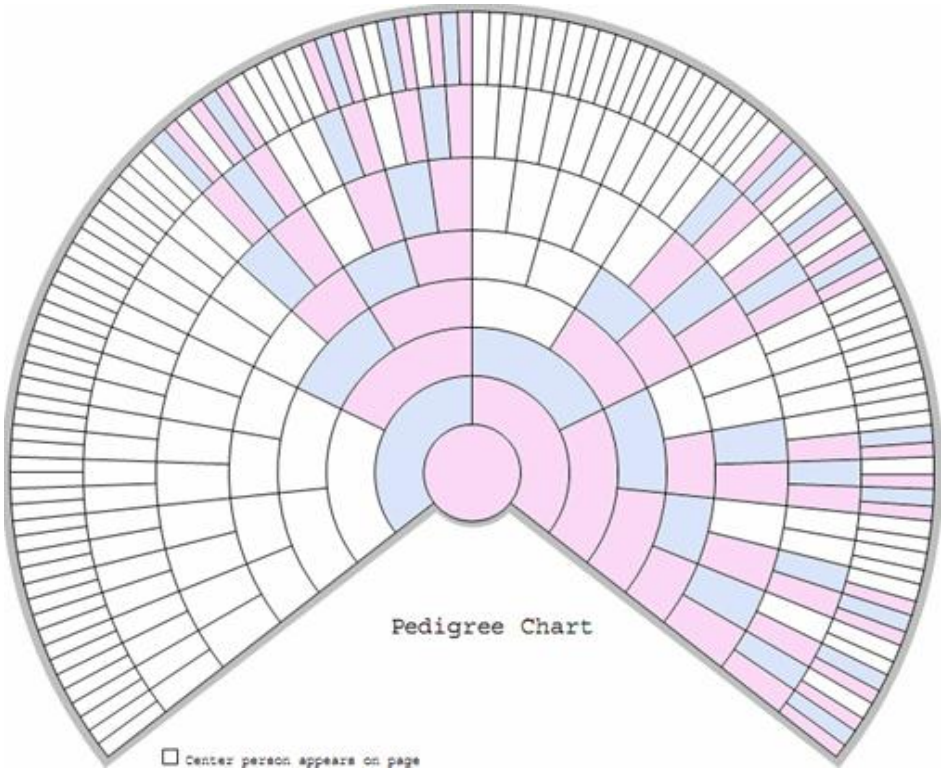

Figure 24: X Chromosomal Inheritance - Females

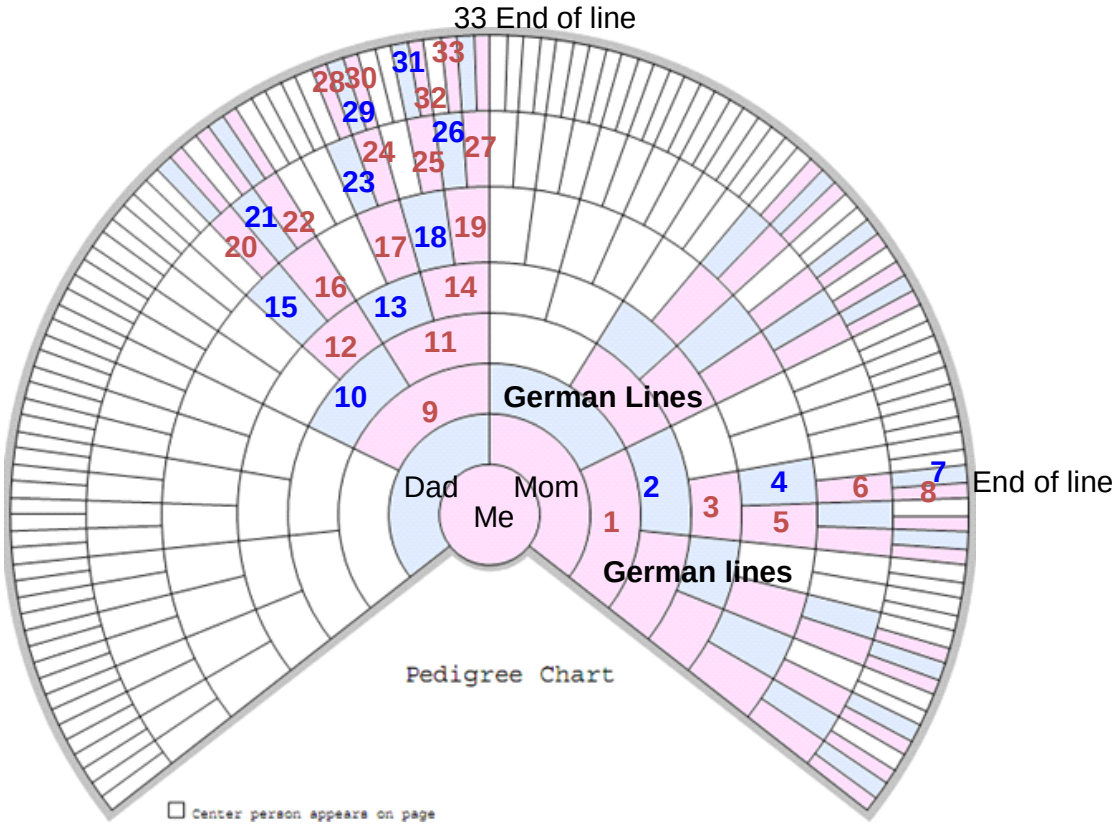

Figure 25: Participant's X chromosome

21 – Michael McDowell, born 1747 in Bedford Co., Va. – his mother is unknown. His father was a second generation immigrant who lived in Halifax and Bedford Counties in Virginia.

22 – Isabel, wife of Michael McDowell, probably born about 1750, surname unknown, located in Virginia.

27 – Elizabeth, born about 1765, wife of Andrew McKee of Virginia.

28 – Agnes Craven is the last slot on the chart, but not the last in the line. Her father was Col. Robert Craven born 1696 in Delaware and was well to do. His mother is unknown. Robert's wife was Mary Harrison, born in Oyster Bay, New York to Isaiah Harrison and Elizabeth Wright. These lines appear to reach back to Europe but are unconfirmed, probably eliminating these lines.

30 – Phoebe McMahon, wife of Joseph Workman, born 1745 York Co., Pa, daughter of Hugh McMahon, mother unknown.

31 – Gideon Faires' mother was Deborah, born 1734, possibly in Augusta Co., Va.

32 – Sarah McSpadden's father was Thomas McSpadden born 1721 in Ireland, eliminating this line. Sarah's mother was Dorothy Edmiston whose father was born in Ireland, eliminating that line. Dorothy's mother was named Jean and was born in 1696 but nothing further is known.

33 – Martha McCamm, born before 1743, wife of Andrew Mackie of Virginia, parents unknown.

On the father's side, we began with 13 slots, positively eliminating one and probably eliminating a second, leaving 11. Of these, 7 could be resolved on the maternal line by mitochondrial DNA testing. Taken together, this side of the pedigree chart is a much better candidate for both Native and African DNA sources.

While the X chromosomal pedigree chart analysis is not the perfect scenario, the pedigree chart has 128 slots. Using the X chromosome narrows the candidates to 34 slots. Genealogy narrowed the slots to 15 and focused mitochondrial DNA testing could narrow them to 6. Further genealogy research on those ancestors could potentially eliminate them by placing them "over the pond" or by discoveries which would facilitate DNA testing.

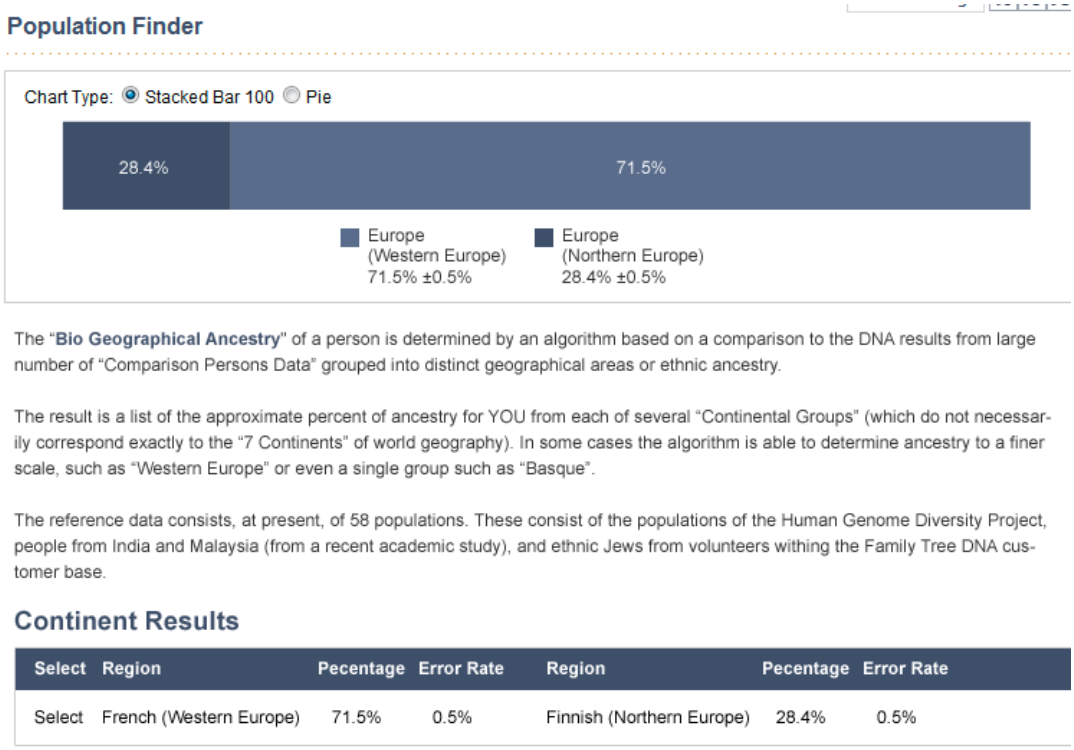

Figure 26: Population Finder

Family Tree DNA

As a part of Family Tree DNA's Family Finder product, they provide ethnicity information broken into seven major groups; Europe, Africa, Mideast, South Asia, the Americas, Oceania and East Asia. Breakdowns within continent are also provided, as shown in Figure 26.

Dr. Doug McDonald

Doug McDonald, a physical chemistry professor, is compiling contributed raw data and comparing the raw data locations with both reference populations and the contributor results. This is not a commercial endeavor and is a private research project. His analysis of our participant's raw data results from 23andMe showed that they are primarily European. His first analysis was without Middle Eastern populations and the results showed European except for a total of about 3% East Asian, Oceana and American. However, in a second run including the Pakistan and Middle Eastern populations, the results now showed 88% European, about 1% Oceanic and American and the balance Middle Eastern and Pakistani. Dr. McDonald indicated that this was slightly more, 1-2%, than most Europeans, and that our participant was generally planted firmly in the middle of the "English" area in his data. They showed no African.

His standard deviation (statistical noise) is about 1%. He can achieve these low deviation numbers by using such a large number of markers (536,904 to be exact)[42] for his comparison. I am grateful to Dr. McDonald for his contribution.

Mitochondrial DNA's Largest Challenge

Mitochondrial DNA is more difficult for genealogists to work with because one can't work with a specific surname. For a male Y-line, the first step to find a DNA match is to visit www.familytreedna.com and enter the surname in the search box to determine if a surname project exists and then to determine is if someone from the genealogical line has previously tested.

Because surnames change with every generation in the mitochondrial line, one can't use surnames or surname searches effectively, making both the genealogy and DNA much more difficult.

The most neglected tool in genealogy today is mitochondrial DNA. This is partially due to the lack of a searchable database, by ancestor, for mitochondrial DNA. If one wants to determine if any descendant of a particular woman has tested, today there is no place one can go to find this information.

[42] Genealogy-DNA Rootsweb posting by Doug McDonald on 7-26-09 and personal correspondence.

It's common today to discover that there is a project for a particular surname and that someone from an ancestral line has already tested. It's not unlikely that many female descendants from ancestral lines have already tested as well, but there is no centralized resource or methodology to find them.

If a mitochondrial pedigree search tool were readily available, more people would become excited about mitochondrial DNA testing as matches would be more easily discernable.

Testing Summary

What did we learn about the participant's ethnic mixture, in particular their Native heritage, by taking the various tests?

DNA Print said that the participant was 75% European, 15% Asian and 10% Native American. The Asian and Native can be interpreted as one group.

OmniPop says that the participant matches Michigan Native Americans (2 versions), Mexicans (2 versions) and Hispanics (the third version).

DNATribes in 2006 indicated a small amount of Native American and Asian but no African. In 2009, they indicated no Native American and some African.

D9S919 shows non-Native values which fails to confirm Native ancestry but also does not eliminate it.

DeCODEme shows 5% Asian, 3% African on autosomal DNA and 6% Asian and 3% African on the X chromosome.

A pedigree analysis shows confirmed Native ancestry on the Mother's side, 11 generations upstream, accounting for less than 1% of the participant's DNA (0.0489%). A second line which is not confirmed would account for .1954%. Cumulatively, they represent .2443%, less than a quarter of 1%.

Oral history of Native heritage on father's side is unproven in the line where it was supposed to exist, but upstream testing is continuing. These are not the lines indicated by the X chromosome on the inheritance chart, which remains unidentified.

A second line rumored to have Native Ancestry on the father's side would represent .049% but has no avenue for proof. This would have been a Native wife to a Jamestown settler and there are only two known children, both males, and the participant descends from both if the genealogy is accurate.

The cumulative Native heritage known or speculated surrounding specific individuals in both of the parents line is .2933%, about one third of one percent. This figure does not approach the 25% combined Native/Asian provided by DNAPrint. To reach the 25% number, all of the participant's unknown lines (24.5%) would have to be Native.

A summary of the various test results is shown in Table 9.

Test/Company	European	Asian	Native	African	Unknown
DNA Print	75%	15%[43]	10%	0	
Omnipop	Yes	No	Yes	Yes	
DNATribes[44]	Yes/Yes		Yes/No	No/Yes	
D9S919			No[45]		
deCodeme	92%	5%	Inferred[46]	3%	
deCodeme - X	91%	6%	Inferred	3%	
Dr. McDonald	97-99%	1-3%	Inferred	0	
Family Tree DNA	100%[47]	0	0	0	
Pedigree Analysis	75%	0	~1%	0	24%

Table 9: Summary of Results

[43] DNA Print Genomics indicated that the Asian and Native should be combined and interpreted cumulatively as Native.
[44] DNATribes results presented with the 2006 value followed by the 2009 value.
[45] While these results do not confirm Native heritage, the absence of this marker does not disprove Native heritage.
[46] Inferred that Asian is actually Native in an American with no history of Asian ancestry.
[47] 71.5% western European, 28.4% Northeastern European

The X chromosome inheritance chart shows that the Asian on the X chromosome is not inherited from any known or suspected Native individuals in either parents' lines.

The X chromosome chart combined with genealogy eliminated a significant number of lines and provided focus on 9 mitochondrial DNA candidates, leaving 6 lines requiring more genealogical research.

This testing cost about $4000 in total. Pricing has changed dramatically today, with deCode doubling in price, 23andMe halving theirs and Family Tree DNA priced below both, but without the medical test results. Some testing (DNAPrint, Codis and DNA Tribes) was duplicated as the participant's mother was tested as well to understand the genesis of certain values and to put the results in perspective relative to the source of the participant's various genetic elements.

Parental testing can yield important information about the source of DNA.

Conclusions

When dealing with the search for Native or minority ancestry, we are dealing with two distinct types of tests. The Y-line and mtDNA tests are absolutes. They absolutely confirm or deny Native (African, Asian or European) heritage in the particular line being tested, but only that line, removing any doubt whether that ancestral line was (or was not) Native. Aside from Y-line and mtDNA testing, we enter the realm of preponderance of evidence.

Is there enough evidence to suggest, or prove, that Native ancestry exists in a given family line, and is that evidence reliable? To date, the only autosomal DNA test that can prove that Native ancestry exists is the D9S919 test, although it cannot identify the line in the family tree where the Native ancestry arose. That must be done with genealogy, Y-line and mtDNA testing.

We have shown that in the Low Marker Resolution Tests, and in particular the DNAPrint test, the statistical noise barrier of 10-15% arises just at the point where most of us are seeking help, around the generation of our great-grandparents who contributed 12.5% each (on average) to our DNA.

Beyond that timeframe, autosomal testing has extreme difficulty in ascertaining minority ancestry and sorting it from statistical noise although Doug McDonald is making significant inroads in this arena and is collaborating with Family Tree DNA. 23andMe references the well known 5 generation barrier which occurs with our great-great-great grandparents who contribute 3.125% each (on average) of our genetic ancestry. Using a 30 year generation, and beginning with a 50 year old individual, our great-great-great-grandparents would have been born about 1810 and would likely have lived into the late 1800s, possibly within memory of our great-grandparents or even our grandparents. This provides a one or at most 2 link oral connection to us today, which should provide us with relatively correct information about their heritage, especially if it were remarkable. The 1850 census also provides us with the location where they lived, a birth state and an ethnic designation, typically white, black or mulatto. Indians where typically enumerated as white if they were quite admixed, or mulatto if they were not.

Most genealogists are looking beyond the 1800 horizon for Native ancestry, whether they realize it or not. The removal known as the "Trail of Tears" occurred in the mid-1830s. By that time, most Native descendants were admixed and were not required to move. Those that were not significantly admixed and were tribally affiliated were required to remove west of the Mississippi. Records of the families who removed generally exist. Records of those who died in route do not. We can presume that those who did not remove were already significantly admixed, probably 75% or more, causing them to be light skinned enough to escape the removal. This pushes the full blooded Native ancestor back at least 2 additional generations, or born in about 1760, into the 7th generation or further where each ancestor contributed less than 1% of our DNA. To show 10% Native ancestry today, we would need ten Native 6th or 7th generation grandparents.

Autosomal tests today fall into three main groupings. Codis and other Low Marker Resolution Tests, the D9S919 SNP test and the new chip based High Resolution Array Tests.

The D9S919 test from Family Tree DNA is reliable and has the ability to confirm but not disprove Native heritage. That test is the least expensive of all the tests at $15 and is not duplicated nor obsoleted by any of the other tests.

Codis and Other Low Marker Resolution Tests

Of the three tests in this category, two are proprietary and one is publicly available.

Test/Company	Markers	Results
OmniPop	Codis markers - 15 markers or subset of 9 (Profiler) or 6 (Cofiler)	Reference data base of 120 or 202 populations leads to highly irrelevant results which can be affected or manipulated by omission of markers or inclusion/exclusion of results in the comparison database
DNATribes	Codis markers - 21	Improvements since 2006 include over 900 reference populations
DNAPrint	V 2.0, 71 markers, V 2.5, 175 markers	Highly skewed, traditional research and pedigree analysis shows the reported results are impossible

Table 10: Codis and Other Low Marker Resolution Tests

Of the proprietary tests, DNATribes shows the most promise with their large reference database. DNAPrint and Omnipop are generally considered to be fatally flawed within the genetic genealogy community.

OmniPop is rather unreliable for genealogical purposes (Brian Burritt stated that it was not created to this purpose), especially given that the three universally accepted ways to run the program produce such different results. Equally unfortunately, many individuals are provided with one (of the three) versions of these results and are convinced of the absolute truth in the resulting report.

Because of its public availability, Omnipop had been adopted by entrepreneurial companies who use it as a basis to offer an ethnic evaluation product. Some have "enhanced" Omnipop in unknown ways. Results can be easily manipulated or skewed by the inclusion or exclusion of marker values. Additionally, a firm with a particular interest in a specific ethnic group could easily, intentionally or inadvertently, skew the Omnipop results by focusing on data relative to their interest group to the exclusion of other world populations. Of course, the result would be that their customers would all appear to fall within their ethnic group of interest because the comparative data base would be biased towards that group and not balanced.

The rapidly changing face of autosomal DNA testing is illustrated by the fact that the 2006 and 2009 DNATribes results bear almost no resemblance to each other. The 2006 results seem to be based on the Omnipop product where Tribes' current product draws from over 900 populations. Most troubling in those results is the inconsistency of the Native and African matches, the Native disappearing altogether in 2009 and the African appearing.

Recently on one of the lists I frequent, a lady reported that she was a white woman from Appalachia and her family had been there "forever". She then posted her results from an Omnipop derived autosomal test which did not include one "Caucasian" category. She interpreted this discrepancy as 'finding her hidden ancestry". It is imperative to be an educated consumer, ask informed questions and evaluate the answers in light of the current technological state of the genetics industry.

Cultural/religious ethnicity such as Jewish heritage can never be determined by autosomal testing of any type nor can conclusions regarding Jewish heritage be drawn from autosomal testing results. Some Jewish heritage can be confirmed by Y-line and mtDNA testing[48].

With the advent of the chip based High Resolution Array Tests, it's likely that the CODIS based and other Low Marker Resolution Tests will quickly become obsolete.

High Resolution Array Tests

Both 23andMe and Family Tree DNA also provide products, Relative Finder and Family Finder, respectively, that, based on autosomal matches find your relatives within their client data bases. While this is not specifically relevant to the determination of minority admixture, it is certainly a consideration that a consumer would want to consider when purchasing these products.

Both of these products have the potential to be extremely useful for individuals where a siblingship test has been returned with inconclusive results or close relationships need to be confirmed. Both individuals would test and the results would show approximately how closely the two individuals are related.

The resolution of more than 500,000 markers is exponentially greater than with the typical Codis 15 marker panel (used for siblingship and paternity testing) or the DNA Tribes panel of 21 markers. Increasing the markers to the half million level also allows the

[48] http://www.familytreedna.com/pdf/nature97385.html and A New Subclade of Y Haplogroup J2b, Athey and Schrack, 2008, Journal of Genetic Genealogy , V4#1 and A Mosaic of People: The Jewish Story and a Reassessment of the DNA Evidence, by Ellen Levy-Coffman, Journal of Genetic Genealogy, Vol 1 #1

Company	# Locations	Ethnicity %	X Chromosome	Raw Data	Other
23andMe	500,000+	Yes	No	Yes[49]	Optional Medical
deCodeme	1,000,000+	Yes	Yes	Yes	Medical
Family Tree DNA	500,000+	Yes	In development	Yes	Traditional tests[50]

Table 11: High Resolution Array Tests

[49] Access to raw data file requires the purchase of the full medical/genealogical version of their product.
[50] Y-line STR, SNP and mitochondrial testing

reduction of the margin of error to between 1 and 3%, from the 10%-15% inherent in the DNAPrint product, thereby dramatically increasing the reliability of the results of the testing.

DeCodeme has the added benefit of showing the X chromosome separately from the rest of the autosomal data which provides genealogists with a tool that can be very important when combined with pedigee analysis. In addition, they show the actual tribal match affiliations in Africa, an unexpected bonus, although none of the African tribes shown by deCODEme and DNATribes are the same.

It's unfortunate that with both 23andMe and deCODEme one has to extrapolate Native ancestry based upon an Asian heritage designation as compared with the HapMap grouping of 45 Han Chinese. This group is exceptionally small and could lead to at least partially incorrect ethnicities, especially for people with German, Polish or eastern European ancestry that might incorporate the Mongol, Magyar or Hun peoples. Both of these companies have the potential to provide a much better ethnic analysis with the size and diversity of their data bases. The new 23andMe Ancestry Finder function may be a first step in this direction.

This comparison with Han Chinese could be the source of at least part of the participant's Asian heritage, especially in the mother's heavily German lineage.

Family Tree DNA's new ethnicity test uses the largest reference population of 58 different groups with a minimum of 50 unrelated individuals per group, augmented by their own research data base. They have been able to reduce the margin of error to about half of a percentage point by utilizing a very high number of markers combined with the large data base of over 2900 reference individuals. They have collaborated with Dr. Doug McDonald on this product.

Autosomal DNA testing when used for minority admixture with the exception of D9S919 and the Duffy Null allele is never absolute, it is suggestive. It's only a tool, not gospel. Mitochondrial and Y-line testing are DNA gospel.

The watchword for autosomal testing today is common sense and caution. If autosomal results from the Low Marker Resolution Tests seem suspicious, treat them as such and continue searching using other DNA and genealogy tools.

The results of the various tools when taken together seem to create a preponderance of evidence that our participant does have Native heritage aside from the proven Native line on the Mother's side. That line is so far upstream as to be nearly undetectable at 0.0489%. The African heritage is less certain, as several tests did not pick up the African component, although genealogical evidence suggests strongly that it does exist and a photograph of an ancestor born in 1818 suggests African ancestry. Other men whose Y-line DNA and surname match that of the participant's ancestor with the suggested African ancestry match his European haplotype and extended haplogroup, but are clearly, visibly of African descent. These families both descend from colonial Virginia although cannot connect via traditional genealogical research methods due to the institution of slavery.

These several years and many tests have enabled a more complete understanding and analysis of the genetic landscape.

I expect that the quality and granularity will continue to improve in the autosomal arena. In order for this to occur, research needs to continue. Individuals who have had any of the High Resolution Array Tests can assist in this endeavor by contributing their data to Dr. Doug McDonald and other genetic genealogy researchers who are making inroads outside of the traditional academic circles. Genetic genealogists often lead the pack today.

My recommendation to people searching for minority ancestry is first and foremost to utilize Y-line and mitochondrial testing. Create a Personal DNA Pedigree chart as shown and utilized in this paper and use the other tools and procedures demonstrated, particularly a pedigree analysis. Autosomal testing can be interesting

and provide valuable insights. I don't discourage individuals from pursuing autosomal testing, but encourage education before the purchase so that expectations will not be unrealistic.

Several genetic genealogy tools can be utilized in the discovery and identification of Native or other minority admixture. For example, the haplogroup designation is a powerful piece of information. When making purchase decisions, especially between the High Resolution Array Testing companies, other factors may come into play as well, such as the desire for medical information or the desire for additional genealogical tools, tests and functions. Table 12 provides a summary of the companies and options they provide to the consumer.

In essence, Family Tree DNA is a genealogical genetics company. Their extremely large customer base reflects that focus as do their products and services. Customers who utilize both the 23andMe and the Family Tree DNA match contact services associated with Relative Finder and Family Finder report a significantly greater number of successful contacts through Family Tree DNA. For some individuals, especially adoptees seeking close family members, there is a significant benefit to being in both data bases as they will reach different populations segments.

Family Tree DNA will shortly provide the option of uploading 23andMe raw data results to their data base to be compared to Family Finder results, although the results of comparing the Illumina chip to the Affymetric chip results won't be as detailed as comparing data from the same chip.

Aside from the D9S919 test, which is definitive, autosomal results relative to minority ancestry are only suggestive and best viewed cumulatively and in conjunction with genealogical data after taking several tests and using a pedigree analysis as a "reality check" when possible. Autosomal DNA results should be interpreted and treated only as tools, data, hints and pieces of evidence, not as direct answers. Definitive answers lie in the combination of Y-line and mitochondrial DNA testing combined with historical and genealogical research taken together with autosomal results.

	Family Tree DNA	**23andMe**	**deCodeme**
Traditional Y-line testing/results	Yes	No	No
Y-line SNPS for extended haplogroups	Yes	No	No
Y-line haplogroup assignment	Yes	Yes	Very basic[51]
Traditional mtDNA testing/results	Yes	No	No
MtDNA full sequence for extended haplogroups	Yes	No[52]	No
MtDNA haplogroup assignment	Yes	Yes	Very basic
Projects	Yes	No	No
D9S919	Yes – separate	No	No
Raw Data	Yes	Yes	Yes
Locations Tested	500,000+	500,000+	1,000,000+
Data Base Size	300,000+[53]	40,000[54]	N/A[55]
DNA Matching to Relatives	Yes	Yes	No
Match Notification	Yes	Yes	N/A
E-mail Matches	Yes	No – contact through their webpage only	N/A
Ethnicity %	Yes - Europe, Africa, Mideast, South Asia, the Americas, Oceania, East Asia using 58 populations with minimum 50 unrelated individuals each	Yes – Europe, African, Asia - using HapMap (4 populations, approx 50 people each)	Yes – Europe, African, Asia - using HapMap (4 populations, approx 50 people each)
Chromosome Browser for Individual Data	Yes	Yes	Yes
Chromosome Browser Compare with Others	Yes	Yes	No compare except to base population
Compare to Family	Yes	Yes	No
Other	Relationship Confirmation, Continental Breakdown	Ancestry Finder, Global Similarity	Map of Kinship, Genetic Atlas

Table 12: Summary

[51] A very basic haplogroup designation would be the main haplogroup, such as haplogroup C (Y-line) or J (mtDNA). Typically this level of haplogroup assignment can be predicted from traditional STR testing (Y-line) or HVR1 testing (mtDNA). A more detailed haplogroup assignment, just as C3 (Y-line) or J1 (mtDNA) can sometimes be obtained or predicted by STR testing (Y-line) or HVR1+HVR2/3 testing (mtDNA), but a complete haplogroup assignment such as C3b (Y-line) or J1c (mtDNA) requires a deep clade SNP test (Y-line) or a full sequence test (mtDNA).

[52] 23andMe tests specific SNP location s for haplogroup assignment, but they do not check for insertions or deletions which are defining mutations in some haplogroups, leading to occasional incorrect or incomplete haplogroup designation (for example, J1a vs J1c), although the major haplogroup designation tends to be correct.

[53] Statistics available on the Family Tree DNA website home page. Only a subset of the 300,000 have taken the Family Finder test.

[54] 23andMe does not provide this information, but recently they announced the 40,000 number up from 30,000 last year. The comparative data base numbers are not an apples to apples comparison. While all of the 23andMe customers have Relative Finder results, experience indicates few are interested in genealogy and reply to "cousin connection requests". At Family Tree DNA, a small percentage of their total data base has ordered the Family Finder test, so participants at this time have fewer matches, but those who do match nearly universally reply to e-mails from their genetic cousins. This disparity reflects the difference in focus of the two companies involved. 23andMe has attracted many individuals interested in their genetic medical predispositions and Family Tree DNA attracts solely genealogists.

[55] This is not relevant as DeCode Genetics does not offer a product similar to the Relative Finder or Family Finder test which compares participant data against their data base.

Resources

Family Tree DNA – www.familytreedna.com – Family Finder autosomal, Y-Line, mtDNA, Codis markers and D9S919 autosomal test

DNAexplain – www.dnaexplain.com - DNA Analysis and Personalized Reports

DNAPrint – the company is now defunct and this test is now being marketed by another firm

DNATribes – www.dnatribes.com – Codis marker autosomal population matches

OmniPop – http://en.wikipedia.org/wiki/OmniPop download available

23andMe – www.23andMe.com – medical and ancestry testing

deCode genetics – www.decode.com – medical and ancestry testing

Doug McDonald - mcdonald@scs.uiuc.edu

Blaine Bettinger's X Chromosome Pedigree Chart for download:
The male version is available for download at: http://www.thegeneticgenealogist.com/wp-content/uploads/2008/12/1male.png
The female version is available for download at: http://www.thegeneticgenealogist.com/wp-content/uploads/2008/12/1b.png