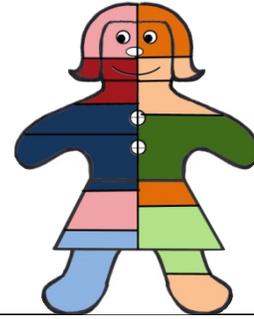


DNA Ethnicity Estimates (BioGeographical and Admixture Estimates): Why the Differences?

Debbie Parker Wayne, CG[®], CGLSM



INTRODUCTION

Researchers who keep up with mass media articles on Direct to Consumer (DTC) genetic tests may be confused as to how useful DNA tests are for genealogy. Articles that show triplets receive different ethnicity predictions¹ (also known as biogeographical and admixture estimates; ethnicity may be the least accurate term, but it is the one most used by the general public so it will be used here). Other articles indicating that the same person gets different ethnicity predictions from different companies receive much publicity.²

Different estimates are to be expected. By its very nature, ethnicity estimates will vary between companies and **even between siblings**, but the testing companies do not explain this before convincing a person to buy a test kit. This misunderstanding causes many people to question how accurate DNA is as evidence to answer *any* genealogical question. Shared segment analysis is very accurate, ethnicity estimates are less accurate for answering some questions.

Good researchers develop logical analysis skills and dig in to find the truth. We read the contradictory articles so we understand the issues. On the face of it, some statements seem to make sense. Instead of accepting these

¹All URLs accessed 17 November 2017.

²“How Reliable Are Home DNA Ancestry Tests? Investigation Uses Triplets to Find Out,” *Inside Edition*, 21 February 2017 (<http://www.insideedition.com/investigative/21784-how-reliable-are-home-dna-ancestry-tests-investigation-uses-triplets-to-find-out>).

² Any person who has tested at multiple companies can display the differing estimates received from each. A quick search on *Google* or *Facebook* DNA groups will find many such posts.

Original publication:

Wayne, Debbie Parker. “DNA Ethnicity Estimates (BioGeographical and Admixture Estimates): Why the Differences?” *Stirpes* 56 (Dec 2017): 18-22; online (http://debbiewayne.com/pubs/pub_stirpes_201712_admix.pdf).

© 2017, All Rights Reserved, Debbie Parker Wayne.

statements, good researchers investigate to learn all of the facts so that we can make an informed decision on whether or how to use evidence from all of the records that we consult, whether documentary or DNA sources.

THE LURE OF ETHNICITY ESTIMATES

Every television viewer has seen the ubiquitous commercials encouraging people to order a DNA test to learn about ethnicity. No work is required of the test-taker and the concept is easy for non-researchers to understand—until they see different numbers from different companies or the predictions do not exactly match the expected ethnicity.

The testing companies all feature ethnicity results on their home pages. *23andMe*³ advertises “a breakdown of your global ancestry by percentages, connect with DNA relatives and more.” *AncestryDNA*⁴ promises users will make discoveries from “their ethnicity to connecting with distant relatives.” *Family Tree DNA*⁵ will help “find family, break through brick walls and trace your lineage through time.” *LivingDNA*⁶ helps “find out where you really come from.” *MyHeritage*⁷ tells users they can “uncover your ethnic origins and find new relatives.” Notice that *Family Tree DNA* is the only company that mentions finding family first before ethnicity or tracing your lineage through time.

So we should not be surprised that the genetic genealogy databases, while rapidly growing, are filled with many who know little of their genealogy at the time a test is taken. We can encourage these test-takers to develop a well-researched family tree since they already have shown an interest in their heritage. We also may be able to help those who are interested in their family tree, but have little information to begin with due to adoption, donor-conception, family secrets, and other matters. All of these, all of our

3 *23andMe* (<https://www.23andme.com/>).

4 *AncestryDNA* (<https://www.ancestry.com/dna/>).

5 *Family Tree DNA* (<https://www.familytreedna.com/>).

6 *LivingDNA* (<https://www.livingdna.com/en-us>). As of this writing, *LivingDNA* does not yet offer a list of matches to other test-takers in their database. This company compares your DNA to determine which area of the British Isles your DNA most closely matches using the database of the People of the British Isles (POBI) Project (http://isogg.org/wiki/People_of_the_British_Isles).

7 *MyHeritage* (<https://www.myheritage.com/dna>).

Original publication:

Wayne, Debbie Parker. “DNA Ethnicity Estimates (BioGeographical and Admixture Estimates): Why the Differences?” *Stirpes* 56 (Dec 2017): 18-22; online (http://debbiewayne.com/pubs/pub_stirpes_201712_admix.pdf).

© 2017, All Rights Reserved, Debbie Parker Wayne.

interaction with others in any situation, should be handled with diplomacy, tact, and an understanding that every person may have differing ideas of what is right than we have. And each of these test-takers may have had a different reason for testing.

LIMITATIONS OF ETHNICITY ESTIMATES

Ethnicity estimates are assigned based on autosomal DNA (atDNA) analysis—analysis of some of the DNA on the chromosomes numbered one through twenty-two. There is no ethnicity “tree” based on specific mutations as there are mitochondrial DNA and Y-DNA trees.

Genetic versus Genealogical Trees

Our genetic family tree may not exactly match our genealogical tree. Because atDNA is recombined before being passed down, the ethnicity estimates may not exactly match the documentary evidence—DNA from some ancestors is lost during recombination.⁸ See figure 1. In this hypothetical example, no atDNA is inherited from any of the ancestors in the white boxes. If a Native American ancestor is the ancestor in the slot pointed to by the arrow the test-taker at the center of this chart will show no Native American ethnicity even though it exists in the genealogical family tree.

Siblings inherit different segments of DNA from their parents; therefore, the siblings have different genetic trees even though they have the same genealogical tree. This results in cousins who match some siblings, but not all siblings. This can also result in siblings receiving different ethnicity estimates.

8 Blaine T. Bettinger, “Q&A: Everyone Has Two Family Trees - A Genealogical Tree and a Genetic Tree,” *The Genetic Genealogist*, 10 November 2009 (<http://thegeneticgenealogist.com/2009/11/10/qa-everyone-has-two-family-trees-a-genealogical-tree-and-a-genetic-tree/>).

Original publication:

Wayne, Debbie Parker. “DNA Ethnicity Estimates (BioGeographical and Admixture Estimates): Why the Differences?” *Stirpes* 56 (Dec 2017): 18-22; online (http://debbiewayne.com/pubs/pub_stirpes_201712_admix.pdf).

© 2017, All Rights Reserved, Debbie Parker Wayne.

The total inherited from the paternal grandparents will be 50% and the total from the maternal grandparents will be 50%. Random recombination often results in a person inheriting more DNA from one grandparent than another. For example, the 50% of atDNA a person inherits from her mother may consist of 28% from the maternal grandmother and 22% from the maternal grandfather. If the grandparents are of different ethnicities, the grandchild may show more of the grandmother's ethnicity.

Recombination also affects ethnicity estimates once the amount of DNA received from any specific ancestral line drops below a few percent. It can be difficult to determine if a small amount of a specific ethnicity is real or is just statistical "noise" (irregular fluctuations that may obscure information). The statistical average amount of DNA inherited from a fourth-great-grandparent is 1.56%. That small amount of DNA can also be noise. Stronger, more credible evidence is needed to support conclusions made from low estimates.

Timeframe of Estimates

Most of the ethnicity calculators used by genealogists are predicting ancestry approximately five hundred years back in time. Very few researchers have family trees where every ancestral line is complete going so far back. Even if recombination resulted in inheritance of exactly one-half of each ancestor's atDNA, most of us could not determine the accuracy of an ethnicity prediction because our trees are not complete going back far enough.

Accuracy of Predictions

Ethnicity estimates are most accurate at the continental level. See figure 2. Estimates are less accurate at regional and country levels. These lower levels have been affected by human migrations and shifting political borders. Estimates are also less accurate where fewer reference samples exist in the databases (Sephardic Jews, Native American tribes, and some others).

Original publication:

Wayne, Debbie Parker. "DNA Ethnicity Estimates (BioGeographical and Admixture Estimates): Why the Differences?" *Stirpes* 56 (Dec 2017): 18-22; online (http://debbiewayne.com/pubs/pub_stirpes_201712_admix.pdf).

© 2017, All Rights Reserved, Debbie Parker Wayne.

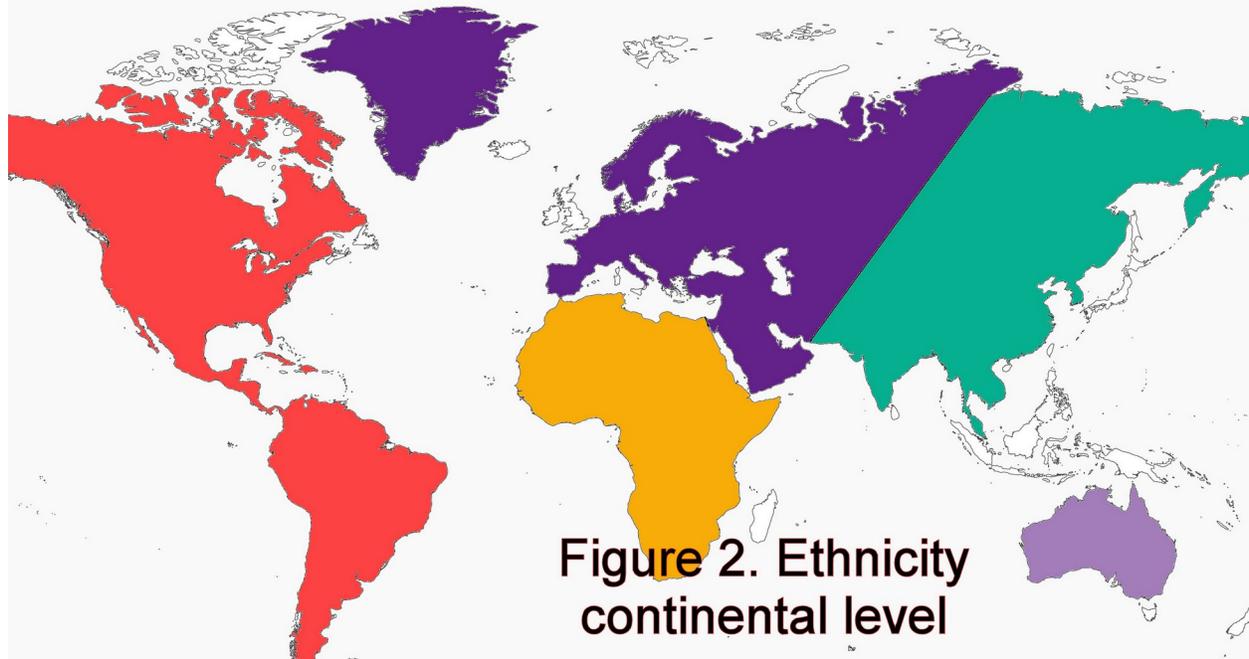


Figure 2. Ethnicity at continental level.

Human migrations and intermarrying complicate breaking the ethnicity estimates down further than to a continental level. Limitations of using the DNA of living persons to estimate what the DNA of a group was like 500, 1,000, or 5,000 years ago is also problematic. Always remember the admixture or ethnicity **predictions** are **estimates**.

Ideally, phased chromosomes should be used for ethnicity analysis. Phasing assigns each allele to a maternal or paternal chromosome. One copy of each atDNA chromosome is inherited from our mother and one from our father. Current technology can read both chromosomes, but cannot determine which allele came from which parent. The atDNA inherited from a mother and father is all “jumbled” together in our DNA data files. Using unphased data for ethnicity comparisons can result in some anomalies when the maternal and paternal alleles are incorrectly linked during analysis.

23andMe phases a person’s data when parents have also been tested there, but many of us do not have living parents to allow phasing our DNA data.

Original publication:

Wayne, Debbie Parker. “DNA Ethnicity Estimates (BioGeographical and Admixture Estimates): Why the Differences?” *Stirpes* 56 (Dec 2017): 18-22; online (http://debbiewayne.com/pubs/pub_stirpes_201712_admix.pdf).

© 2017, All Rights Reserved, Debbie Parker Wayne.

Some companies, *AncestryDNA* being one, use algorithms to pseudo-phase the DNA data. Other companies have not implemented such procedures.

HOW ETHNICITY ESTIMATES ARE CREATED

*Reference Populations*¹⁰

A test-taker's DNA is compared to the DNA samples in a reference population database. The portion of DNA being compared is then assigned the ethnicity of the most closely matching reference sample.

In order to achieve accurate ethnicity estimates, reference populations must be large and diverse. If a population database does not include a specific population, then that "ethnicity" will not be listed even when comparing a person whose ancestry comes from that population. If the number of tested persons from a reference population is small or restricted, then others with the same ancestry may not match the reference population as closely as expected.

Each testing company and third-party analysis tool uses differing reference populations and differing mathematical algorithms to predict ethnicity back to a different timeframe. This is one reason that a test-taker receives different numbers from each company. Each company pulls from public scientific sources (*Human Genome Diversity Project, 1000 Genomes Project, HapMap*) what the scientists advising the company think are the best reference population samples. Some of the companies have also promoted testing of living individuals in targeted populations (persons in a location whose four grandparents were all from that same area, Africans, African Americans) to test to augment the public sources. Some companies also use customer data.

¹⁰ Testing companies often offer access to scientific papers or white papers describing the reference populations. "Reference Populations," *23andMe* (<https://customercare.23andme.com/hc/en-us/articles/212169298-Reference-Populations>). "Determining Your DNA Ethnicity Estimates," *AncestryDNA* (<https://blogs.ancestry.com/ancestry/2016/10/12/determining-your-dna-ethnicity-estimate/>). "Population Clusters in myOrigins," *Family Tree DNA Learning Center* (<https://www.familytreedna.com/learn/ftdna/myorigins-population-clusters/>).

Original publication:

Wayne, Debbie Parker. "DNA Ethnicity Estimates (BioGeographical and Admixture Estimates): Why the Differences?" *Stirpes* 56 (Dec 2017): 18-22; online (http://debbiewayne.com/pubs/pub_stirpes_201712_admix.pdf).

© 2017, All Rights Reserved, Debbie Parker Wayne.

Process and Algorithms

The testing company divides the DNA data into blocks, segments, or SNPs which are then compared to the reference population database. Some companies make multiple passes comparing the data. If a company randomly selects the segments or blocks of DNA to compare, it is possible that one person's DNA data can result in different ethnicity predictions each time a calculator is run. The basic comparison process includes

- 1. Dividing the test-taker's data into segments or blocks**
- 2. Selecting a DNA segment or block and comparing the test-taker's DNA to the reference population database**
- 3. Assigning the segment to the closest matching reference population (even if it means shoving a square peg into a round hole— the "best" fit is assigned)**
- 4. Repeating the process multiple times using different blocks (done by some companies)**
- 5. Compiling all of the calculations into an overall percentage prediction**

USING ETHNICITY ESTIMATES TO SOLVE GENEALOGICAL PROBLEMS

Third-party tools allow comparisons to a testing company's predictions to determine the accuracy of the predictions. *GEDmatch*¹¹ offers multiple admixture calculators some of which map the predicted ethnicities to specific chromosome segments; many of the calculators offer options using different reference populations. *Interpretome*¹² is a scientific chromosome browser tool that genealogists can use. To date, *23andMe* is the only testing company that indicates which DNA segment is assigned to each ethnicity.

11 *GEDmatch* (<https://www.gedmatch.com/>).

12 Karczewski, Tirrell, Kim, and the Interpretome team, *Interpretome* (<http://interpretome.com/>).

Original publication:

Wayne, Debbie Parker. "DNA Ethnicity Estimates (BioGeographical and Admixture Estimates): Why the Differences?" *Stirpes* 56 (Dec 2017): 18-22; online (http://debbiewayne.com/pubs/pub_stirpes_201712_admix.pdf).

© 2017, All Rights Reserved, Debbie Parker Wayne.

Even with the limitations of ethnicity estimates, there are ways this evidence can help answer genealogical questions. Some are described below.

- When multiple ethnicity calculators provide the same or similar percentages, those results are likely to be more accurate.
- When multiple calculators indicate the same segment of atDNA is assigned to the same ethnicity, those results are likely to be more accurate.
- Once a test-taker has mapped specific chromosome segments to particular ancestral lines, the ethnicity assignments for the segment can be compared to the documented lineage. For example, an ethnicity calculator may indicate that a segment at the beginning of chromosome 7 is predicted to be Native American. See figure 3. If no prior research has been done, this ethnicity estimate indicates one of the ancestral lines includes Native American ancestry. Alternatively, prior research and “chromosome mapping” may indicate the ancestral line from which that portion of chromosome 7 was inherited is also Native American. In this case, the documentary evidence and the DNA evidence support the same conclusion.
- Ethnicity estimates can also be useful when one ancestral line is distinctly different from others. For example, when documentary research indicates a second-great-grandmother is Native American, but all other ancestral lines are all European. We cannot predict a Native American tribe, but Native American DNA can be distinguished from European. If the ethnicity estimate indicates 5.8% of your DNA is Native American that is very close to the 6.25% statistical average predicted to be inherited from a second-great-grandparent. The possibility that an ancestor has mixed Native American, African, and European ancestry can make the analysis more complicated. That 5.8% of Native American DNA might have been inherited from multiple ancestral lines and not just one in which case perhaps the second-great-grandmother is only part Native American. Other hypothetical possibilities should be investigated before reaching a genealogical conclusion.

Original publication:

Wayne, Debbie Parker. “DNA Ethnicity Estimates (BioGeographical and Admixture Estimates): Why the Differences?” *Stirpes* 56 (Dec 2017): 18-22; online (http://debbiewayne.com/pubs/pub_stirpes_201712_admix.pdf).

© 2017, All Rights Reserved, Debbie Parker Wayne.

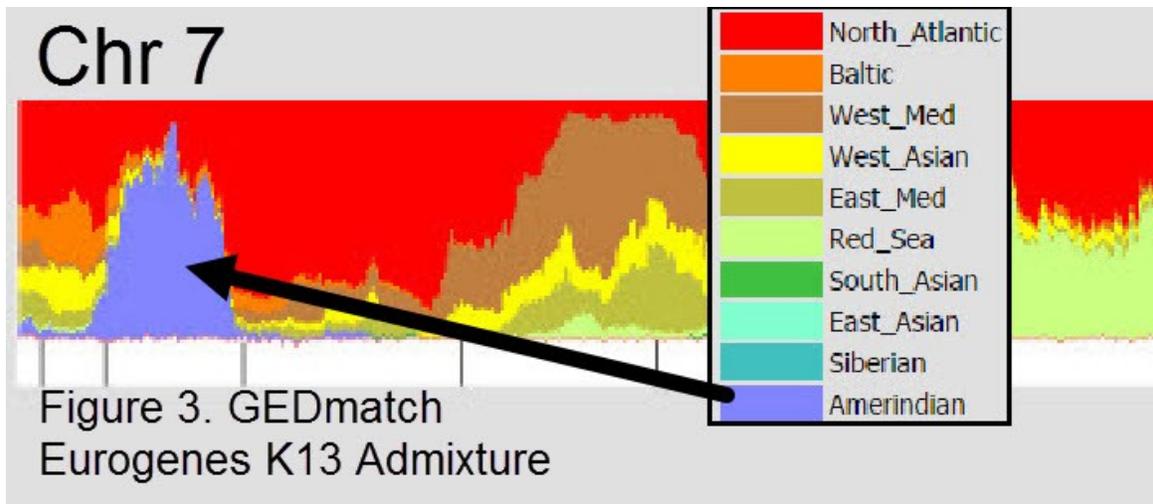


Figure 3. GEDmatch Eurogenes K13 Admixture indicating a Native American segment on Chromosome 7.

CONCLUSION

While ethnicity has some application to genealogical questions, it is not as useful in most situations as more detailed analysis of the matching DNA segments between test-takers. Many factors, such as the contents of the reference database, the matching threshold, and the matching algorithms, result in different estimates from the same DNA sample. Future articles will explore some of the ways ethnicity predictions can be most useful to answer a genealogical question.

Understanding how testing companies and third-party tools actually interpret the DNA test results helps researchers understand the clues in our DNA that can help us answer genealogical questions. It also helps us convince others to test to provide evidence to answer those questions.

For additional information on ethnicity or admixture tools see “Admixture analyses,” *ISOGG Wiki* (https://isogg.org/wiki/Admixture_analyses).

Original publication:

Wayne, Debbie Parker. “DNA Ethnicity Estimates (BioGeographical and Admixture Estimates): Why the Differences?” *Stirpes* 56 (Dec 2017): 18-22; online (http://debbiewayne.com/pubs/pub_stirpes_201712_admix.pdf).

© 2017, All Rights Reserved, Debbie Parker Wayne.

Debbie Parker Wayne, CG, CGL, is experienced using DNA analysis as well as more traditional techniques for genealogical research in Texas, the South and the West. She coordinates the genetic genealogy courses at several genealogy institutes and is the Texas State Genealogical Society's DNA Project director. She wrote the NGS *Continuing Genealogical Studies: Autosomal DNA* course and co-authored the first genetic genealogy workbook, *Genetic Genealogy in Practice*, published by NGS. See <http://debbiewayne.com/> for more information.

The words Certified Genealogist and designation CG are registered certification marks with the United States Patent and Trademark Office, and the designations Certified Genealogical Lecturer and CGL are service marks of BCG, used under license by certificants after periodic competency evaluations (and only during the current five-year period for which they are certified).

Original publication:

Wayne, Debbie Parker. "DNA Ethnicity Estimates (BioGeographical and Admixture Estimates): Why the Differences?" *Stirpes* 56 (Dec 2017): 18-22; online (http://debbiewayne.com/pubs/pub_stirpes_201712_admix.pdf).

© 2017, All Rights Reserved, Debbie Parker Wayne.