

Where the Heck is my German DNA Estimate?

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

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

Ethnicity estimates – also called admixture or biogeographical estimates – are likely the single most common reason that non-genealogists take an autosomal DNA test. Since ethnicity testing appeals to the broadest swath of people (both genealogists and non-genealogists), advertising for autosomal DNA tests necessarily revolves around ethnicity testing. However, few people examine *how* these ethnicity estimates are obtained. Accordingly, few people understand the benefits and limitations of ethnicity estimates. Are these estimates “accurate”? Are they useful?

What is an Ethnicity Estimate?

An ethnicity estimate is the assignment of a test taker’s DNA to populations around the world based on computerized comparisons of each of those segments to “reference populations.” A segment is assigned to the reference population that it most closely matches.

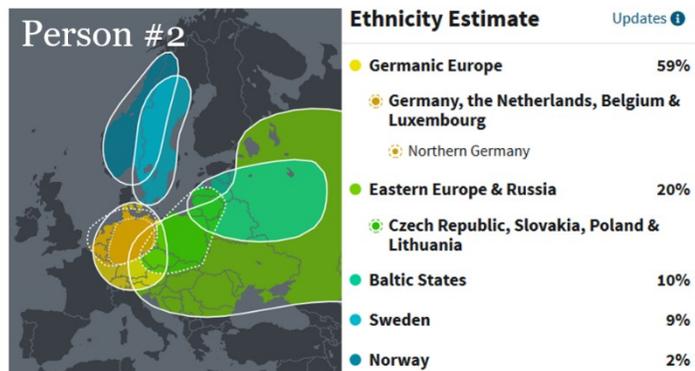
Depending on a wide variety of factors (some of which are discussed below) people with similar ancestry can have very different ethnicity estimate. For example, the two people in the images to the right both self-identify as “100% German,” but they obvious have very different estimates at AncestryDNA!

What are the Limitations of Ethnicity Estimates?

Although ethnicity estimates have many beneficial uses, some of which we’ll examine, they also have many limitations that are important to understand when interpreting results. Below are some of the most important and impactful limitations:

1. Reference Populations are Relatively Small and Regionally Limited

Ethnicity estimates are generated by comparing a test taker’s DNA to DNA obtained from multiple different “reference populations” from around the world. A reference population is a group of people with a long history of ancestry in a specific geographic location, long enough to have DNA that is distinctively associated with that specific geographic location. Accordingly, ethnicity estimates are fundamentally limited by the size and diversity of their reference populations.



Although the testing companies have relatively good diversity of populations from geographic locations such as Europe, many other locations such as Asia and Africa are not adequately represented. As a result, people with Asian ancestry will receive a very high-level ethnicity estimate rather than specific sub-regions within Asia.

Additionally, few populations have been consistently stable or isolated. Germany, for example, has been a cross-road for millennia, and thus DNA from many different regions and backgrounds has consistently entered into and left the region.

2. What the Heck Does “German” Even Mean?

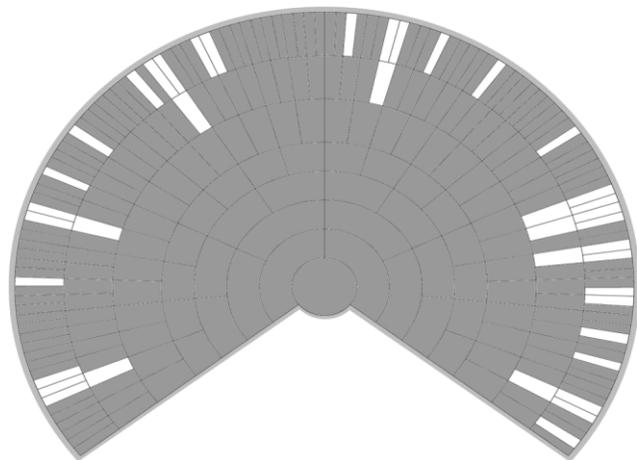
Another limitation of ethnicity estimates is that while they typically provide an estimate of a person’s ancestry 500 to 1,000 years ago, current geopolitical boundaries and country names are very different than they were 500 or 1,000 years ago. An ethnicity estimate requires *some* geographic association for test takers to understand them, so most of the testing companies use modern geopolitical boundaries. Obviously, this is a limitation of these estimates.

3. You Inherit Random DNA From Your Ancestors

Your genome (the complete set of DNA that you have inside you) is the same size as your great-great-great-grandparent’s genome. But you have 32 great-great-great-grandparents, so you get only about 3% of your genome from *each* of those great-great-great-grandparents.

Since you inherit that 3% randomly from those great-great-great-grandparents, you don’t know what ethnicity you are inheriting if there is any diversity among those ancestors. On the other hand, if those ancestors are all solidly from a single location for many generations, we would expect to see a single ethnicity inherited from that single location. The vast majority of test takers, however, do not have ancestry consistently from a single location.

Additionally, the further back in time you go, the fewer ancestors you receive DNA from. Around 7 generations (give or take 2 generations), you start to lose ancestors from your genealogical family tree. The image to the right shows an example Genetic Family Tree in which ancestors represented in the test taker’s DNA are shown in gray, and ancestors that did NOT pass down DNA are shown in white. Accordingly, if the ancestors in white had a unique ethnicity, that unique ethnicity will not be shown in the test taker’s results!



How Are Ethnicity Estimates Created?

Each of the “Big Five” DNA testing companies (23andMe, AncestryDNA, Family Tree DNA, Living DNA, and MyHeritage) provide an ethnicity estimate to the test taker. There are both similarities and differences in how these ethnicity estimates are obtained.

In general, an ethnicity estimate is most commonly obtained by:

1. Obtaining/receiving DNA file;
2. Optionally phasing the DNA (i.e., separate into ‘Mom DNA’ and ‘Dad DNA’);
3. Breaking the DNA sequence into short sequences;
4. Compare each short sequence to the reference populations, and find the reference population(s) that are most similar to the short sequence;
5. Assign the short sequence to that most similar reference population;
6. Optionally repeat multiple times; and
7. Add the assignments together to generate a full ethnicity estimate.

We will examine this process and some of these variations during the lecture.

Ethnicity Estimates at GEDmatch

GEDmatch has a variety of different ethnicity calculators which analyze the raw data you’ve uploaded to the service. Each of these ethnicity calculators has a slightly different focus. Like other calculators, each of these use different reference populations and have different algorithms for processing the data.

1. **MDLP Project** (Magnus Ducatus Lituaniae Project) (<http://magnusducatus.blogspot.com>)
2. **Eurogenes** (Eurogenes Genetic Ancestry Project) (<http://bga101.blogspot.com>)
3. **Dodecad** (Dodecad Ancestry Project) (<http://dodecad.blogspot.com>)
4. **HarappaWorld** (Harappa Ancestry Project) (<http://www.harappadna.org>)
5. **Ethiohelix** (Intra African Genome-Wide Analysis) (<http://ethiohelix.blogspot.com>)
6. **puntDNAL** (Ancient DNA) (puntdnalking@gmail.com)
7. **Gedrosia DNA** (Ancient DNA, Indian, Asian) (Dilawerkh4@gmail.com)

GEDmatch ethnicity calculators can be a good way to compare a company’s estimate to the estimates obtained from a testing company. Additionally, since only 23andMe provides the location within the genome where certain ethnicities are located (for example, African on chromosome 12), GEDmatch ethnicity calculators can provide this location information for DNA tests obtained from the other testing companies.

Can We Use Ethnicity Estimates for Genealogical Research?

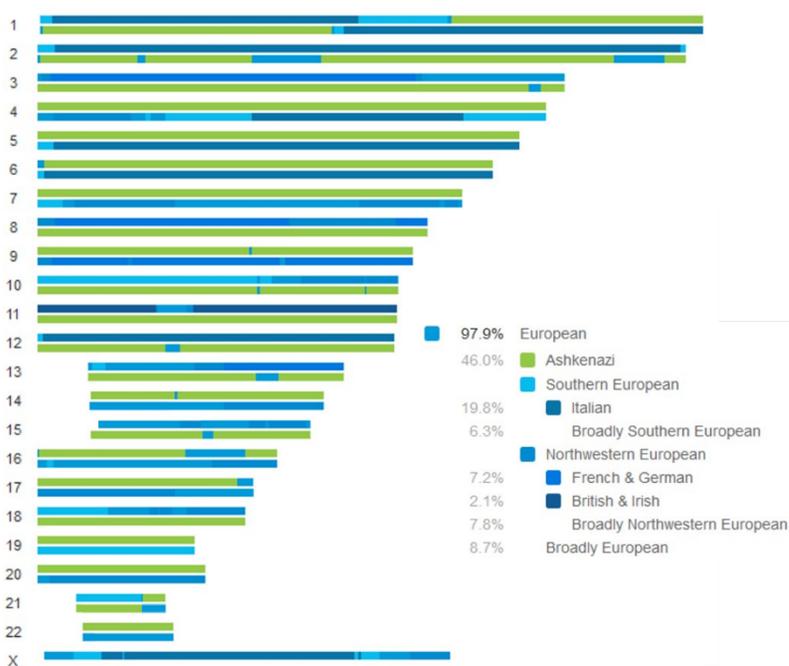
Despite their limitations and detractors, ethnicity estimates can be an extremely powerful tool for genealogical research. Obviously, if your ancestry is solidly from one location such as Northern Europe for many, many generations, you shouldn’t expect to learn much from your ethnicity estimate (unless, of course, your results indicate that you’re actually NOT 100% Northern European due to a misattributed parentage event!). However, there are many situations where an ethnicity estimate provides very useful and actionable genealogical information! Discussed below are just a few of these situations.

1. Unknown Parentage and Unexpected

Test takers who do not have any information about a biological parent or grandparent, or discover that their biological parent or grandparent is not who they believed it to be, can use ethnicity

estimates to potentially learn about that ancestor. For example, as shown in the image to the right, the test taker was expecting to find 100% Northern/Italian European ancestry. However, the results showed that the test taker was 50% Northern/Italian European and 50% Ashkenazi.

And since this 50% was nearly always one full chromosome, it appeared to come from one parent (and was not on the X, suggesting the Ashkenazi parent was the FATHER). Although the cousin matching made it clear, the ethnicity estimate made clear that the test taker's father was not his biological father.



2. Chromosome Mapping

Another potential use of an ethnicity estimate is for chromosome mapping. For example, in the following image of chromosome 2, there is a Native American segment of DNA in orange:



If the test taker knows where that segment came from, they can assign that segment of DNA in a chromosome map to that ancestor!

Resources:

- Crossan, Andrea. *You took a DNA test and it says you are Native American. So what?* PRI's the World, November 24, 2016 (<https://www.pri.org/stories/2016-11-24/you-took-dna-test-and-it-says-you-are-native-american-so-what>).
- Genealogical Musings. *Finally! A Gedmatch Admixture Guide!* April 6, 2017 (<http://genealogical-musings.blogspot.com/2017/04/finally-gedmatch-admixture-guide.html>).
- Russell, Judy. *Admixture: not soup yet.* The Legal Genealogist, May 18, 2014 (<https://www.legalgenealogist.com/2014/05/18/admixture-not-soup-yet/>).
- Southard, Diahan. *Results may vary - one family's DNA ethnicity percentages.* Lisa Louise Cooke's Genealogy Gems, 28 February 2016 (<https://lisalouisecooke.com/2016/02/28/dna-ethnicity-percentages-may-vary/>).