Ethnicity Estimates – A Deep Dive

Presentation for the Allen County Public Library

Steven P. Frank – steven@geneticgenealogist.net

I. Importance of ethnicity and genetic communities for family history.
   A. Connects you to distant ancestral populations
   B. Makes the historical personal
   C. Visualizes migrations
II. Ethnicity Matches
   A. Science and history behind ethnicity estimates
      1. NAT GEO Genographic Project reference panels
      2. Ancestry's improvements to reference panels and algorithms
      3. Ancestry ethnicity regions vs other companies
   B. Limitations of ethnicity estimates
      1. Genetic Diversity
         I. Genetic drift- The amount two given populations have separated genetically over time.
            I. Happens on the scale tens to hundreds of thousands of years. The more recent the separation, the less drift.
            II. More drift = easier to distinguish.
         II. Population isolation vs admixture
            I. The more populations interacted and mixed historically, the less differences there will be genetically. Prevents drift from occurring.
         III. Founder effect - Aka bottleneck effect.
            I. A small group started a community that then had a population explosion. Within that new population, genetic diversity will be lacking.
            II. All humans outside sub-Saharan Africa are the result of a bottleneck Subsequent bottlenecks occurred later, most notably in the Americas.
      2. Difficulties in defining and testing groups
         I. Mixed Origins of Populations
            I. Example: What does it mean to be English? Origins are German, Scandanavian, French, Celt, Roman, etc.
         II. Poor representation of some ethnicities (Native North American)
         III. Mixed modern populations may not correlate to ancient populations.
      3. Differences in algorithms
         I. Each country uses its own proprietary algorithms and groupings.
         II. Ethnicities/

III. In creating the algorithms, the companies incorporate certain weights and biases/assumptions about populations.
I. Ex: Assuming most test takers are American, of British decent. This might cause the algorithm to assume a segment is English, when it could easily be from another part of Europe.

C. Unexpected results and what they may/may not mean
1. 1% or less results are only reliable.

III. Genetic Communities
A. Science behind genetic communities
1. Genetic networks to genetic communities
2. Applying Tree data
3. Matching to historical migrations
B. Using Genetic Communities for family history research
1. Identifying matches with shared Genetic Communities/ethnicities
2. Using Genetic Communities to put context to your ancestor's journey

IV. Sideview
A. Splits your ethnicity estimate between your parents.
B. Cannot identify mother vs father without them testing
C. NOT a representation of your parents' ethnicity, just what you inherited from each.
D. Has potential to be a very useful genealogy tool in the future.

Final thoughts on Ethnicity Estimates:

They are ESTIMATES. Think of them as a rough sketch rather than a photograph. DNA is random. You don't get an even amount from all your ancestors, and it is normal for siblings to have different results.

All the companies frequently update their ethnicity results. Remember to check back later, your results may change!