



## **GUIDE TO UNDERSTAND THE SCIENCE AND SIGNIFICANCE OF YOUR DNA ANCESTRY TEST**

- **INFORMATION ABOUT THE ANCESTRY TESTING**
- **ANSWERS TO YOUR MOST FREQUENTLY ASKED QUESTIONS**
- **GLOSSARY OF TERMS**

### **1. What is the Relationship between Race and Genetics?**

First, it is important to realize there are far more genetic similarities between all human beings than there are differences, regardless of the racial heritage. At the level of DNA, only one base in a thousand differs between the genomes of any two people. The term race, in general usage, includes both a cultural and biological feature of a person or group of people. Given the fact that physical differences between populations are often accompanied by cultural differences, it has been difficult to separate these two elements of race. Over the past few decades there has been a movement in several fields of science to oversimplify the issue declaring that race is "merely a social construct". While, indeed this may often be true, depending on what aspect of variation between people one is considering, it is also true that there are biological differences between the populations of the world. One clear example of a biological difference is skin color. There is a strong genetic component to the level of pigmentation in a person's skin and there are dramatic differences across populations. Pigmentation is literally only skin deep, a heritable trait under natural selection across the complex environments in which we all live. For example pigmentation is driven by the contrasting need to protect our skin from UV-light damage while allowing sufficient exposure to sunlight to catalyze the light-driven conversion of vitamin D precursors into the active form of this essential vitamin. Many other examples of genetic selection in human populations are under active scientific investigation.

The biological component of race is largely based on the genetic structure of human populations. This structure is a nested hierarchy from East to West where populations in the Americas and the South Pacific are a subset of the genetic diversity found in Eurasia which itself is a subset of the diversity found in Africa. It is clear that the human species is relatively young. As a species, we most likely originated in east Africa, according to most archeologists, 100,000 to 300,000 years ago, and diverged as groups, expanded, moved, and settled the globe. During these migrations, and in the time since, there has been some degree of independent evolution of the populations that settled the various continents of the world. The simplest evidence of this evolution can be seen in the differences in allele frequencies at genetic markers. Generally, we see that alleles found in one population are also found in all populations and the alleles that are the most common in one are also common in others. These similarities between populations highlight the recent common origin of all populations and strong connections between populations throughout human history. However, there are examples of genetic markers which are different between populations and it is these markers, called Ancestry Informative Markers (AIMs), which can be used to estimate the ancestral origins of a person or population.

It should be clear by now that race is a complex and multivariate construct. In our procedures we focus on a person's genetic ancestry and not their race. Your DNA has no recorded history of your political, social, personal or religious beliefs. It is a simple four letter code that records all of the changes in the DNA from one generation to the next. We identify and report those key changes that reflect your genetic ancestry most succinctly. They are like finger prints and snow flakes, unique and wildly complex.

### **2. What is BioGeographical Ancestry (BGA)?**

BioGeographical Ancestry (BGA) is the term given to the biological or genetic component of race.

BGA is a simple and objective description of the Ancestral origins of a person, in terms of the major population groups. (e.g. Native American, East Asian, European, sub-Saharan African, etc.) BGA estimates are able to represent the mixed nature of many people and populations today. In the US, as in many other countries across the globe, there has been extensive mixing among populations that had initially been separate. In the fields of human genetics and anthropology, this mixing is referred to as admixture. BGA estimates can also be understood as individual admixture proportions, which take the form of a series of percentages that add to 100%. For example, a person in question may be found to have: 75% European; 15% African; 10% Native American ancestry, or they may be found to have 100% European ancestry.

### **3. How is BioGeographical Ancestry estimated?**

The AncestryByDNA test uses an especially selected panel of Ancestry Informative Markers (AIMs) that have been characterized in a large number of well-defined population samples. These markers are selected on the bases of showing substantial differences in frequency between population groups and, as such, can tell us about the origins of a particular person whose ancestry is unknown. For example, the Duffy Null allele (FY\*o) is very common (approaching fixation or an allele frequency of 100%) in all sub-Saharan African populations. Thus, a person with this allele is very likely to have some level of African ancestry. After the analysis of these AIMs, in a sample of a person's DNA, the likelihood (or probability) that a person is derived from any of the parental populations and any of the possible mixes of parental populations is calculated. The population (or combination of populations) where the likelihood is the highest is then taken to be the best estimate of the ancestral proportions of the person. Confidence intervals on these point estimates of ancestral proportions are also being calculated.

### **4. How can BGA estimates be used?**

We believe that an objective assessment of the biological component of human ancestry is possible and that such research could enrich our lives in a number of ways such as understanding health disparities. Are there genetic contributions to the higher rates of hypertension and diabetes in African Americans or the higher rates of dementia in European Americans? If not, then what are the cultural or environmental differences that underlie the prevailing differences? Studies of these and other diseases require independent, objective measures of BioGeographical Ancestry (BGA). Estimates of BGA can help reconnect individuals separated by adoption, or some other event, with their ancestral populations.

Even if a person is not particularly motivated to reconnect with ancestors, he or she can uncover the past of their family either to verify family legends or to search for forgotten roots.

In the near future, we hope to allow customers to compare their ancestral proportions to others in their family, town, city, or state who have chosen to participate. Because it is based on DNA, and unlike the census, this new tool will provide the most accurate demographics data that is possible. We will call this our “personal demographics” tool.

### **5. What is the medical significance of BGA estimates?**

The medical significance of BGA estimate is negligible. Although some diseases are found at different frequencies in populations across the globe, hardly any are restricted to one group. The usefulness of BGA estimates, in biomedical research, comes from epidemiological analyses where many individuals are analyzed together to make very general statements about differences in risk. Even though these results can be very significant, they have almost no meaning regarding the level of risk for any one person in the population.

### **6. How is BGA analysis different from mtDNA and Y-chromosomal ancestry analysis?**

There are several commercially available tests of mtDNA (female contribution) and Y-chromosomal (male contribution) markers, which have been promoted as a means of learning one's ancestral origins. Although these tests could provide information regarding the provenance of some of a person's ancestors, they are very limited. For example, one generation ago a person has two ancestors, one mother and one father; five generations ago, a person has 32 ancestors; while 10 generations ago, a person has 1024 ancestors. Ten generations is roughly 250 years and within the

time frame of genealogical interest, especially when we are considering the settlement of North America, because they only look at two (2) chromosomes. Y-chromosomal analysis and mtDNA analysis each could only provide information on a very small proportion of a person's ancestors. Our test relies on sequences throughout your genome, so we can say more about a greater number of your ancestors.

### **7. Can BGA provide more specific information about ancestry?**

AncestryByDNA 1.0 is the first version of the test, and it is specifically designed to provide information on the proportions on ancestry on the continental level. In other words, this test allowed us to uncover the levels of Native American, European, and African ancestry, as three component groups. The current BGA tests that is offered, AncestryByDNA 2.0 and 2.5, are expanded to provide information on the proportions of ancestry on the continental level for most continents, Native American, European (which includes European, Middle Eastern and South Asian groups such as Indians), African, but we distinguish ancestries within Asia and the Pacific Rim by adding East Asian, (which includes the Pacific Islanders) as an additional group. Since there will also be interest in defining the levels of ancestry within continents (such as distinguishing Japanese from Chinese, or Northern European from Middle Eastern), we are in the process of developing a new series of Ancestry Informative Markers that will provide more insight into where within a particular continent a persons' ancestors were most likely derived.

### **8. Have there been any recent changes to the test?**

As of recently, the software for generating and presenting results was modified. The current version of the software builds on the knowledge gained during the development of AncestryByDNA 2.5. Prior to the change, we reported the results in terms of a single triangle plot, showing the confidence contours in the three dimensions corresponding to the three most likely ancestries for the individual. The new software shows BGA estimates using a more complex triangle plot and a bar graph. The new triangle plot and bar graph show the confidence contours in all four dimensions (Western Sub Saharan African, European, Native American and East Asian).

### **9. My proportions were 85% European and 15% East Asian, and the East Asian part surprised me. How reliable is the 15% part of the profile?**

The 15% East Asian means you most likely share sequence identity with East Asians at some markers. Studies have indicated that, although there is a measurable level of noise in the test, we do see a pattern of minor contribution among certain populations. These findings will be published at a later date.

Your range bars and confidence contours are also part of your answer. The values within these determined ranges represent other possible outcomes that are statistically significant but are less likely. Therefore, if your range bar on the bar graph includes zero, you should consider that possibility.

### **10. How can I confirm the significance of a low percentage of admixture, such as 4% Native American or 3% African?**

There are two ways for you to confirm the value of this estimate:

You may have access to historical records or other provenance that leads you to confirm or refute the admixture. For example, if your records suggest that you have a grandparent of East Asian heritage and you register with the test as of 5% East Asian, the two observations combined make a stronger case for East Asian ancestry than either on their own.

You can obtain the admixture proportions for your father and mother. Let's say you register with 4% African and you want to know whether this 4% is in error or is accurate. You obtain the admixture proportions from your parents and each is 100% European. Chances are the 4% was a result of statistical noise. However, if your mother was 15% African and your father was 100% European, your non-zero percentage of African is likely to be an accurate indicator of African ancestry.

**11. I think I have American Indian heritage, but my test results show that I am 100% European.**

Barring adoption and paternity issues, your range bars and confidence contours are also part of your answer. The values within these determined ranges represent other possible outcomes that are statistically significant but are less likely. Therefore, if your range bar on the bar graph includes values greater than zero, you should consider that possibility.

In the “results for simulated matings” section of the website, the distribution of scores around expected ideals are illustrated. This Demonstrates that a range of results are possible for a specific pedigree pattern (i.e. 1 Native American Grandparent and 3 European Grandparents)

Without direct genetic testing, it is may be uncertain that your ancestor was 100% Native American. If you only had one such ancestry in your family this would make detection all that much more difficult. In addition, the number of generations between a limited number of ancestors with Native American markers and your self will also influence the probability of detecting such a connection.

**12. I think I have American Indian heritage, but my test results show more East Asian than Native American admixture. Am I wrong or is the test wrong?**

Your results are derived to how well you compare to our reference populations. Your range bars and confidence contours are also part of your answer. The values within these determined ranges represent other possible outcomes that are statistically significant but are less likely. It is probable that your intervals and ranges encompass a number of values for both Native American and East Asian populations and this should be considered when questioning these types of results. 5 Generations ago, you had 32 Great Grandparents, all of whom contributed to your genetic makeup. Understanding the contributions from each of these individuals is difficult. It is possible that some of these people had some minor East Asian component and their contributions are influencing your results.

**13. I thought I was purely of Scandinavian origin, but my results show minor East Asian admixture. How is this result possible?**

In our testing we’ve observed a number of samples from Scandinavia and Eastern Europe that exhibit some East Asian admixture. This is probably due to past migrations and is possibly telling us something about the interaction of various groups in those regions. In addition, cultures from these areas traveled to region of Asia and it does not seem unreasonable that children may have been produced from some of these travels introducing new genetic markers into the Scandinavian population.

**14. How accurate are the minor (<10%) admixture scores?**

Pedigree studies have shown that the levels of admixture follow known inheritance patterns, supporting the validity of these low values. However, as stated above, the confidence intervals are important for understanding your results and the scores are the most likely estimates from your DNA. If you confidence interval overlaps zero your value may be within the statistical noise threshold of the test and should be considered a possible result. As the number of markers tested goes up the statistical noise should decrease as evidence in the change in confidence interval size between the 2.0 and 2.5 tests.

**15. Has this technology been published in the scientific literature?**

1. Parra, E., Marcini, A., Akey, J., Martinson, J., Batzer, M., Cooper, R., Forrester, T., Allison, D., Deka, R., Ferrell, R. and M. Shriver. 1998. Estimating African American Admixture Proportions by Use of Population Specific Alleles. *Am. J. Hum. Genet.* 63:1839-1851.

2. Pfaff, C., Parra, E., Bonilla, C., Hiester, K., McKeigue, P., Kamboh, M., Hutchinson, R., Ferrell, R., Boerwinkle, E., and M. Shriver. 2001. Population Structure in Admixed Populations: Effect of Admixture Dynamics on the Pattern of Linkage Disequilibrium. *Am. J. Hum. Genet.* 68:198-207.

3. Parra, E., Kittles, R., Argyropoulos, G., Pfaff, C., Hiester, K., Bonilla, C., Sylvester, N., Parrish-

Gause, C., Garvey, W., Jin, L., McKeigue, P., Kamboh, M., Ferrell, R., Pollitzer, W., and M. Shriver. 2001. Ancestral Proportions and Admixture Dynamics in Geographically Defined African Americans Living in South Carolina. *American Journal of Physical Anthropology* 114:18-29.

4. Frudakis, T., V Kondragunta, M Thomas, Z Gaskin, S Ginjupalli, S Gunturi, V Ponnuswamy, S Natarajan, and P Nachimuthu. 2002. A Classifier for SNP-Based Racial Inference. In Review, *Journal of Forensics Sciences*.

## **GLOSSARY OF TERMS**

**Admixture:** The action of mixing, the fact of being mixed, something added by mixing or a product of mixing.

**Allele:** Alternate sequences for a particular position in the genome. For example, a common variation in the genome is for some forms of the sequence to have Cytosine (C) while other forms have Thymidine (T). Thus, since we have two copies of each chromosome, there are three genotypes at this position CC, CT, and TT.

**Ancestry:** Line of descent.

**Ancestry Informative Marker (AIM):** AIMs are the subset of genetic markers that are different in allele frequencies across the populations of the world. Most polymorphism is shared among all populations and for most loci the most common allele is the same in each population.

**Antecedent:** A preceding event, condition, or cause. The conditional element in a proposition. Used herein to denote individuals that came before.

**Anthropology:** The science of human beings; especially : the study of human beings in relation to distribution, origin, classification, and relationship of races, physical character, environmental and social relations, and culture.

**Biogeography:** A science that deals with the geographical distribution of animals and plants.

**Chromosome:** The physical units of heredity: long linear strands of DNA. Humans have 22 autosomal chromosome pairs, plus two sex chromosomes, X and Y. Men have two copies of each autosome, 1, 2, ..., 22, X, Y. Women have two copies of each chromosome 1, 2, 3, ..., 22, X, X. Each person thus has a total of 46 chromosomes.

**Demography:** The statistical study of human populations especially with reference to size and density, distribution, and vital statistics.

**DNA: Deoxyribonucleic Acid.** Genetic information is encoded and transmitted from generation to generation in it. It is a coiled molecule organized into structures called chromosomes cells. Segments along the length of a DNA molecule form genes, the molecular laborers that carry out all life-supporting activities in the cell. Although all humans share the same set of genes, individuals can inherit different forms of a given gene, making each person genetically unique.

**Ethnic:** Of or relating to large groups of people classed according to common racial, national, tribal, religious, linguistic, or cultural origin or background.

**Exagomous:** Marriage outside of a specific group especially as required by custom or law. Meant herein to refer to admixture from outside of a group.

**Endagomous:** Marriage inside of a specific group especially as required by custom or law. Meant herein to refer to admixture from inside a group.

**Genealogy:** An account of the descent of a person, family, or group from an ancestor or from older forms. The study of family pedigrees.

**Genetics:** the study of the function and behavior of genes. Genes are bits of biochemical instructions found inside the cells of every organism from bacteria to humans. Genes direct the synthesis of proteins.

**Genome:** All of the genetic material in a species. The human genome is approximately 3,300,000,000 base pairs in length.

**Genomics:** The study of the complete compliment of genetic material in a species.

**Heterogeneous:** Consisting of dissimilar or diverse ingredients or constituents.

**Homogeneous:** Of uniform structure or composition throughout.

**Hypothesis:** A tentative assumption made in order to draw out and test its logical or empirical consequences.

**Locus (pl. loci):** The name for a physical position on the genome. Can either refer to a large region such as a complete gene or a very specific region, like a particular base pair position.

**MALD:** A mathematical algorithm that is used to determine population structure called Mapping by Admixture Linkage Disequilibrium and when used collectively with proprietary genomic maps and other algorithms it is called ADMIXMAP.

**Pedigree:** A register recording a line of ancestors.

**Pharmacogenomics:** Pharmacogenomics is the testing of individuals to predict their genetic predisposition to drug response. The field of study looks at how genetic variations among the population affect drug response. It involves the analysis of genomic data to develop a screening process for more efficient clinical trials and molecular diagnostic tests used to determine individualized drug responses. It also looks at how a new targeted drug therapy could be developed using genomic data and analysis.

**Polarized:** To break up into opposing factions or groupings. Used herein to refer to BioGeographical Ancestry admixture results such as 95% East Asian/5% Native American as opposed to a relatively even mix such as 50%/50%.

**Polymorphism:** The property of having more than one state or alternate sequence at a particular position. The alternate states are called alleles.

**Population genetics:** The study of the genetics of groups of individual organisms.

**Single Nucleotide Polymorphism (SNP; pronounced snip):** A precise base pair position where different people are found to vary in sequence. Generally two alternate alleles are found at a particular SNP. At least 2,000,000 SNPs are now known and there may be over 30,000,000 in the human genome.