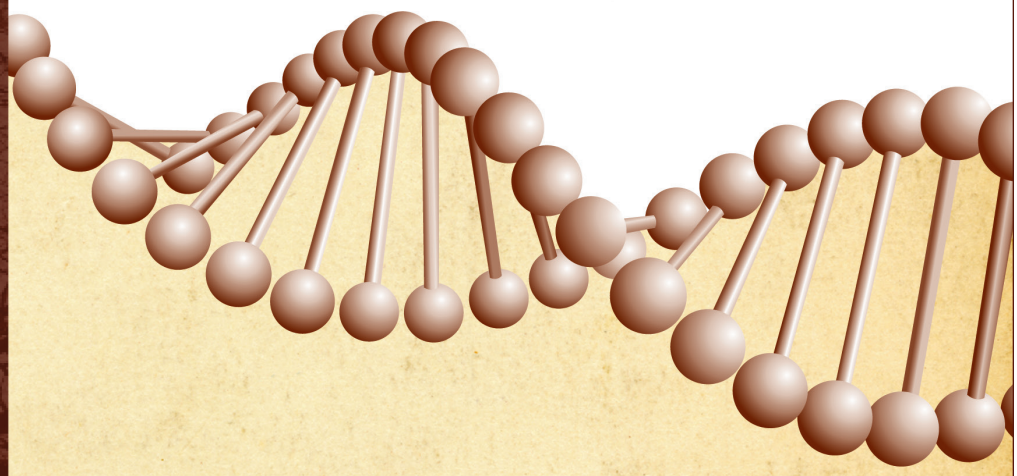
Ancestry

DNA Testing | Results Manual



Ancestry*by*DNA™

DNA Origins™ Test Manual

This Results Manual will walk you through the basics of DNA, what your results look like and what they mean, plus additional information and resources for you to use.

Ancestry*by*DNA™

Table of Contents

Introduction

Congratulations on completing your DNA Origins™ test.

We hope that you find your test results useful in learning more about your family history and ultimately, about you.

The purpose of this manual is to serve as a reference source for questions you may have about ancestry DNA testing in general, as well as the results that you received.

As you will learn in this manual, the DNA Origins™ test report you have received is a product of extensive scientific research combining works of geneticists, anthropologists, and social scientists. A lot of effort has been placed into ensuring that the methods used in the test are scientifically and statistically sound. DNA Origins™ is based on testing thousands of samples from populations around the world for DNA markers that can provide clues about your ancestral history.

In This Manual

We'll go over some DNA basics and information on human evolution and history, followed by a discussion of the ancestry testing services we currently offer. We also provide some resources at the end of this manual, should you wish to do more research on your own.

Each day, scientists around the world are further refining what we know today about ancestral genetics and worldwide human migrations. The results we give you are a great tool in your voyage as you learn more about your ancestors and where they came from.

Your Results Package

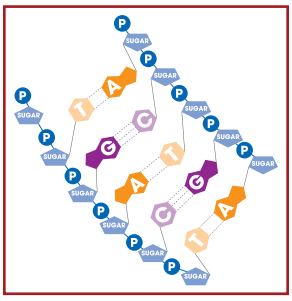
Your DNA Origins™ results package includes a certificate listing your biogeographical ancestry percentages, and a bar graph depicting the statistical confidence intervals (explained further in the Interpreting Your Results section). It also comes with this manual, which is available in both print and electronic editions.

DNA Basics

This chapter reviews some basic information about DNA biology and inheritance. As you may know, DNA is the genetic material found in all living things. Each cell in your body contains a full copy of the genetic material, which encodes all your body's structure and functions.

DNA is most often represented as a double helix. In the cell, the DNA helix is found in tightly coiled and packaged units called **chromosomes**. If all of the DNA inside a cell is stretched out and placed end to end, you would have a long, double-stranded helix that is about 3 meters in length.

The DNA helix looks like a twisted ladder. The two sides are composed of the four bases: adenine (A), thymine (T), guanine (G), and cytosine (C), and the rungs of the ladder represent hydrogen bonds that



connect specific pairs of these molecules together: A–T and G–C.

The arrangement of these molecules, called the DNA sequence, spell out the instructions for our physical characteristics and body functions. These instructions are found in units called **genes**. Not all of the DNA sequences code for genes. In fact, the majority of your cell's DNA is found in non-coding regions—they are thought to serve other purposes, which include regulating gene activity as well as providing structural support and protection. Many of these non-coding regions happen to have markers that are useful for human identification and ancestry studies.

Terms You Will Encounter in this Manual

In this manual, you will encounter terms specific to DNA testing and genealogy. A quick review of these terms will assist you in your understanding of your results.

Admixture (genetic)

In genetics, the result of interbreeding between two or more previously isolated populations within a species, resulting in the introduction of new genetic lineages into a population.

Ancestry Informative Marker (AIM)

The DNA Origins™ test examines AIMs—the subset of genetic markers that are distinctive of the founding populations of the world. These markers are found in all populations but in different forms (alternative sequences, also called alleles)—and for each marker, there is an identifying allele that is the same in each population. This allows our test to determine which of the founding populations have contributed to your genetic makeup today.

Allele

Alternate letters in the DNA sequence at a particular position in the genome. For example, a common variation in the genome is for some populations to have Cytosine (C) in a specific location on their DNA, while other populations would have Thymine (T). *See also entry on Single Nucleotide Polymorphisms.*

Biogeographical Ancestry
An estimation of your ancestral proportions based on the evolutionary and geographical history of the human race.

Chromosome
The physical units of heredity: long linear strands of DNA. Humans have 22 non-sex chromosome pairs (called autosomes), plus two sex chromosomes, X and/or Y (with men having X and Y, and women having two copies of X). Each person thus has a total of 46 chromosomes.

Genomics
The study of the entire genetic material in a species.

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

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.

Maximum Likelihood Estimate (MLE)
The most statistically probable estimate of your ancestral proportions.

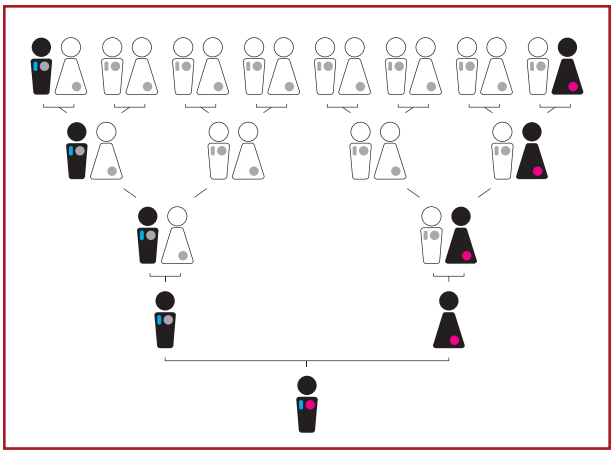
Polymorphism
The property of having more than one alternate sequence at a particular position on the chromosome. The alternate sequences are called alleles.

Single Nucleotide Polymorphism (SNP; pronounced snip)
A precise base pair position on the chromosome where 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.

Types of DNA Used in Ancestry Testing

When discussing ancestry DNA testing, we often refer to three types of DNA in your cells: the Y chromosome, used in direct paternal lineage testing, the mitochondrial DNA (mtDNA), used in direct maternal lineage testing, and the **autosomes**—non-sex chromosomes that make up the complement of your genome. The DNA Origins™ test scans 144 markers in your autosomes, called ancestry informative markers (AIMs), to determine your ancestral makeup. The three types of ancestry DNA tests are discussed briefly below.

As you may know, humans have a total of 46 chromosomes in each cell—23 are inherited from the mother



and 23 from the father. Two of these are the sex chromosomes—males inherit their Y chromosome from the father and their X chromosome from their mother. Because the Y chromosome is inherited strictly through the paternal line relatively unchanged through several generations, it is possible to perform a **Paternal Lineage Test** tracing direct male ancestry by examining the Y chromosome.

The 46 chromosomes are all found inside the cell nucleus. Another type of DNA, called mitochondrial DNA (mtDNA) is found outside the nucleus. Regardless of whether you are male or female, all of your mitochondrial DNA comes from your mother. During conception, only the male nuclear genetic material (i.e. the 23 chromosomes) enters the egg. As such, the fertilized egg only contains mtDNA from the mother. Due to this maternal mechanism of inheritance of mtDNA, we can perform a **Maternal Lineage Test** by examining markers on the mtDNA.

The chart to the left shows the direct paternal and maternal lines of a male individual (females cannot trace their paternal lines via their own DNA). The blue bar represents inheritance of the Y chromosome through the paternal line, and the pink circle represents the inheritance of the mitochondria through the maternal line. As you can see, among this individual's ancestors, his Y chromosome and mtDNA only give a slice of the full story of his ancestry.

As you can see from the illustration, much of your ancestral makeup is influenced by many other members of your family, and the paternal and maternal linage tests give you information about the direct lineages. The third test, DNA Origins™, fills in the gap by giving you the complete picture, taking into account contributions by relatives outside of the direct maternal and paternal lines—all those represented in white in the graphic above.

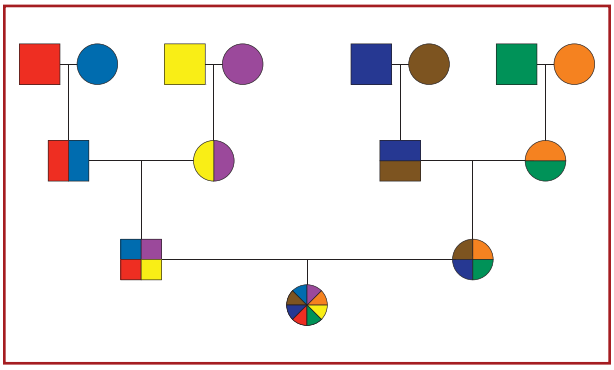
DNA Origins™: How it works

The DNA Origins™ test examines markers found across your 22 autosome pairs. As you may recall, you receive one copy of each pair of chromosomes from your mother and one copy from your father. In turn, your parents received a copy of each chromosome from both their parents. In effect, your maternal copy of chromosome 1, for example, could have been passed through your mother from your maternal grandmother OR your maternal grandfather, but which one you received was randomly determined at conception (you could not have received both).

Most of the time, this chromosome 1 copy that you receive from your mother is actually a chimeric (combination) chromosome that includes parts of chromosomes from her parents. The process by which these chimeric chromosomes are created is called recombination, and it occurs at least once on

each chromosome every time a new sperm or egg cell is made. **As such, our chromosomes are mixed together and so our genomes contain segments of DNA from all of our ancestors.**

The chart on the right shows how an individual, represented by the multicolored circle at the bottom, has inherited parts of her DNA from all her ancestors, represented by the different colors. This is a simplified diagram—in reality, all of this individual's relatives have multiple colors to pass down, and which colors she receives is randomly determined. In effect, the DNA Origins™ test looks at the full spectrum of colors, or DNA markers, that an individual has, and makes a statistical determination of his/her ancestral makeup.



To determine your ancestry, we have extracted DNA from your buccal sample (mouth swab) for the DNA Origins™ test to examine the sequence of your DNA at a large number of different positions across the 44 autosomes. The buccal sample you returned to us contained thousands of cells, and each of your cells contains exact copies of your DNA. Though all humans are 99.9% identical at the level of our DNA sequence, there are certain regions of each chromosome that are different from person to person. These regions are called genetic markers or **Single Nucleotide Polymorphisms (SNPs)**, and a small fraction of these SNPs have differences that are characteristic of the world's continental population groups. These types of markers are best termed **Ancestry Informative Markers (AIMs).** These AIMs constitute less than 0.5% of our genetic material.

Extensive research has been done to develop the DNA Origins™ test to ensure it includes the most informative and statistically relevant AIMs, and a discussion of the research and efforts that went into selecting the AIMs are found in the "Validation Studies" section.

By determining your sequence at these AIMs, we can calculate the relative contribution of the 4 main population groups to your ancestry makeup: European, sub-Saharan African, East Asian, and Indigenous American.

In contrast, the mitochondrial DNA or Y-chromosome test can only provide data on a single lineage of ancestors each generation into the past. For example, 10 generations ago (year 1802 at 20 years per generation), a baby born today has 1024 ancestors. By measuring your ancestral proportions throughout your DNA (a method commonly referred to as "genome scanning"), we are actually measuring the average population affiliations of all of these 1024 ancestors. Since random processes (recombination and independent assortment) during the production of sperm and egg cells as well as during conception determine the mixings and pairings you harbor, even siblings who share both parents can have different ancestral proportions (due to receiving different sets of chromosomes), even though they were the product of the same male-female union.

Until the DNA Origins™ test, genetic tests for genealogy or personal interest have been restricted to just one chromosome, such as the Y chromosome or the mitochondrial DNA. As such, these other tests offer information that is very different from the DNA Origins™ test. It is not that this information is incomplete or defective, it is simply different information that is useful for certain types of genealogy research. For example, while the Y and mtDNA tests give you information about specific lines of ancestry, the DNA Origins™ test give you a broader picture, and **will not tell you which side of your family** the ancestral contributions of African, East Asian, European, and Indigenous American came from.

Concepts of Race

We encounter concepts of race in our everyday lives. There are many situations in which you could be asked to identify your race—for example, as part of your medical history during a doctor's visit, in an application for entry into college, and in various government service forms.

It is important to recognize that there is no widely accepted and standardized definition of race. To date, there is no evidence that current social classifications accurately capture biological and genetic similarities. In addition, peoples' perceptions of their own race and ethnicity can change over time, and even self-reported categories of race could change[1]. Please consider the following concepts of race that are commonly accepted by scientific scholars today:

- Race is not a biological concept. There is not enough genetic differentiation among human populations to consider them distinct zoological races.
- Race is a social construct. This means that these classifications (black, white, Hispanic, Jewish) are defined (and redefined) by the prevailing sociopolitical structure.
- Ethnicity is a term that is often interchanged with race, although it has a more social connotation. It incorporates social, religious, linguistic, dietary, and other variables to identify individual persons and populations. Like race, ethnic boundaries are dynamic and imprecise[2].
- There is more genetic variation within races than there is between them[3].

BioGeographical Ancestry

In contrast to "defining your racial background," the DNA Origins™ test provides a research-grade estimation of a person's BioGeographical Ancestry (BGA). BGA is a means of expressing the proportional ancestry of a person that is devoid of the ethnic labels and the dichotomous grouping of persons into racial categories. There are important uses of this in epidemiological and complex disease mapping research and in forensic science. BGA estimates provide a description of a person in terms of ancestral proportions that are based on the evolutionary and geographical history of our species. Our recommended book, "The Great Human Diasporas: The History of Diversity and Evolution," written by one of the leaders in the field of Evolutionary Anthropology, Dr. Luca Cavalli-Sforza (Stanford University), details a broadly

[1]Comstock, RD; Castillo, EM; and Lindsay, SP. 2004. "Four-year review of the use of race and ethnicity in epidemiologic and public health research." American Journal of Epidiology 159:611-619.

[2]Witzig, R. 1996. "The medicalization of race: scientific legitimization of a flawed social construct." Annals of Internal Medicine 125:675-679.

[3]Williams, DR. 1994. "The concept of race in health services research: 1966 to 1990." Health Services Research 29:261-274.

accepted model of human evolution. It is within this scientific framework of human origins that the BGA estimation can be understood as a description of a person's placement on a ***Multi-Dimensional Continuum of Ancestry™***.

Ancestry percentages obtained with DNA Origins™ reports genetic affiliations, not necessarily recent genealogical histories in the way you are accustomed to thinking of them. They are anthropology-driven AND genealogy-driven: a 10% East Asian result for a European population, or person, could have a very simple, genealogical interpretation (i.e. your recent ancestors were Chinese) or a more complex, anthropological explanation (i.e. you are from an ethnic group with a historical connection to East Asians). Please refer to the "Interpreting the Results" section to understand how genetic affiliations arise in human populations and how to responsibly interpret genome base population affiliation test results such as those provided by DNA Origins™.

A brief discussion on human migration history follows in the next section.

The Story of the Human Race

Where did we come from? Today we are in an era of rapid intercontinental travel, where people can reach the other side of the world in less than 30 hours of air travel. However, the story of human migration goes back about 125,000 years ago, when the first modern humans made their journey out of Africa, into Asia, Europe, and beyond.

These migrants established founder groups that gave rise to present-day Europeans, Indigenous Americans, Africans, and East Asians. Many people from places such as the United States, Southeast Asia, or Latin America are admixtures of these founder groups, while many people from places such as Nigeria, Ireland and Japan are of unmixed heritage. The DNA Origins™ test reveals whether or not your heritage has been derived from one or more of the founder groups.

Wherever they settled, human populations accumulated DNA mutations that marked their presence at these times and places—these are the DNA markers that we bear today.

Pre-Human History

Archaeologists and anthropologists generally accept that the first hominids, or proto-humans, existed as far back as 7 million years ago in Africa. It was not until 2 million years ago when a new species emerged in the same place. Dubbed *Homo erectus*, members of this species were about as tall as modern humans, with a more muscular build. They made axes and cleavers out of stone, and they could communicate. However, their facial features were quite different from that of modern humans: flat noses, a thick bone ridge above their eyes. They were in the right place at the right time—one of the Earth's periodic ice ages had come to an end, and the world became a warmer, damper place where life could abundantly flourish. The wide availability of food enabled this species to migrate quickly, and groups of *H. erectus* traveled in

all directions, most notably into Asia, where a member of this species, the "Java Man," is found. Further fossils have been found in China, Spain, and Georgia (north of the Caucasus mountains).

Because the range of *H. erectus* was widespread, it has been argued that modern humans, *Homo sapiens*, evolved from this species. However, DNA evidence suggests otherwise. It is now believed that the *H. erectus* eventually died out within the past 500,000 years. Although there is some evidence that a member of *H. erectus*, the Neanderthal Man, may have coexisted with modern humans, it is believed that all humans today genetically trace back to the first *H. sapiens* who evolved in Africa sometime between 400,000 and 130,000 years ago.

Homo sapiens, the "Knowing Man"

Like its predecessors, *H. sapiens* was confined to its ancestral homeland for a long time. The Sahara had once again become a desert, forming an impenetrable barrier to the outside world. Then, about 125,000 years ago, earth temperatures warmed and the Sahara became green once more, this time for several thousands of years. Humans migrated north, possibly crossing the river Nile to the Egyptian coast into modern-day Israel and Lebanon, following the Mediterranean coast. As another cycle of cold and dry temperatures took place, food and shelter became scarce for the humans who left Africa, and it is possible that these humans died out themselves.

A second wave of migration out of Africa took place when favorable conditions returned some 40,000 years later. These humans reached Arabia from the northern edge of the Horn of Africa. The DNA of modern-day humans can be traced back to these individuals—they are the group from which all of today's non-African population descends.

Timeline of Human Migration

The timeline of human migration presented below corroborates with DNA evidence. Your very own DNA contains evidence of this migration, and your results give you an insight into your ancestors' role in human history.

170,000 B.C.E.	Modern humans arise in East Africa
160,000 B.C.E.	Humans spread into southern and western Africa
125,000 B.C.E.	The first modern humans leave Africa, arriving on the Mediterranean shore.
90,000 B.C.E.	*Homo sapiens* settle China.
85,000 B.C.E.	A second wave of migration of humans spread along the coastal regions and reached Java within 10,000 years.
65,000 B.C.E.	Humans migrate northward into Europe
50,000 B.C.E.	The Cro-Magnon people are the earliest *H. sapiens* to settle Europe.
40,000 B.C.E.	Ancestors of Australian aborigines arrive on the Australian continent.
25,000 B.C.E.	*Homo sapiens* in Europe show artful skill with their cave paintings.
20,000 B.C.E.	The first humans cross the Bering land bridge into North America.
18,000 B.C.E.	The peak of the last Ice Age, when Europeans retreated to 4 refuges areas: Iberia, Ukraine, Siberia, and the Balkans.
15,000 B.C.E.	Humans may have reached South America by boat.
10,000 B.C.E.	The Earth warms up, and Europe is repopulated. Agriculture spreads there.

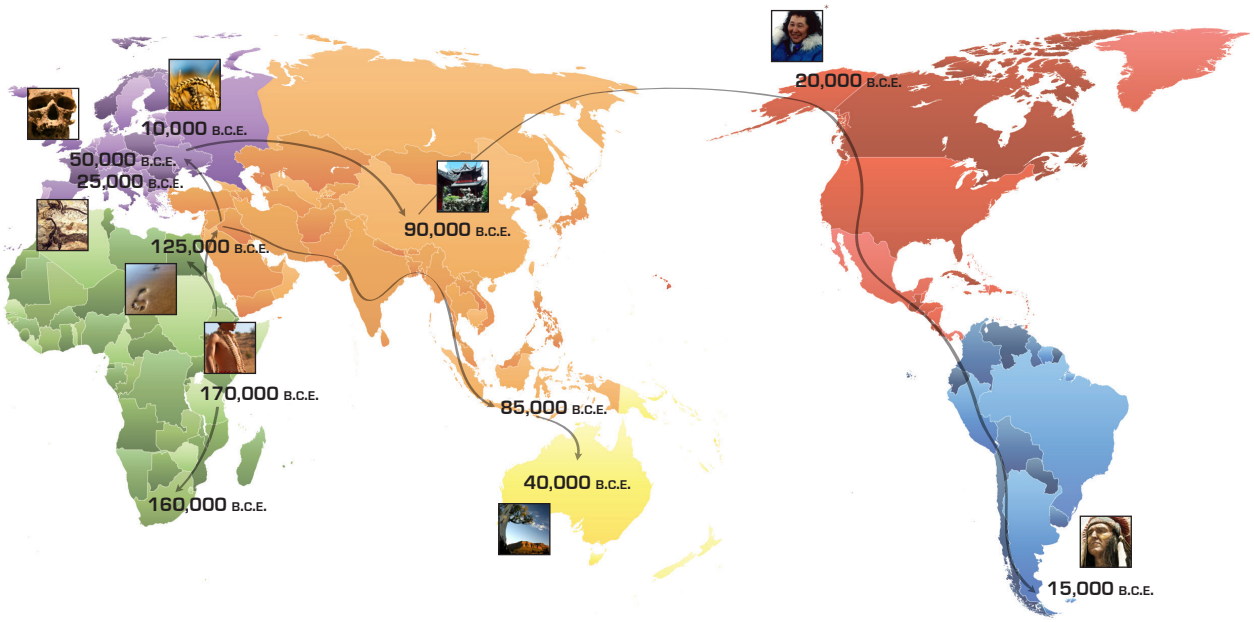


*Photo credit: Inuit grandmother by Angsar Walk, Creative Commons Attribution ShareAlike 2.5.

For much of our history as a species, we were more mobile than today. The advent of agriculture, in at least four separate global regions, about 10,000 years ago changed this for many people, but did not stop the process of migration. Indeed, the largest migrations in human history started only 500 years ago with the European colonial period, the trade in enslaved West Africans, and the colonization of the New World. However, prior to this time and for millennia, people have moved about and particular regions of the world show traces of these migrations back and forth, into and out of continental and sub-continental regions. Some examples of such regions are East Africa, North Africa, Central Asia, South Asia, and Insular Southeast Asia.

Although these populations are distinct groups today with languages, cuisines and cultures that identify them as such, their genetic makeup reflects the long-term history of migrations from more than one region.

Interpreting Your Results

The DNA Origins™ test reports estimates of your biogeographical ancestry (see the Biogeographical Ancestry discussion on page 6 of this manual). Your results come in two formats: a table report listing your ancestral percentages of the four founder groups (European, Sub-Saharan African, East Asian, and Indigenous American), and a bar graph showing the confidence intervals for your percentage ratios. The concept of confidence intervals is explained in the Bar Graph section.

Maximum Likelihood Estimate

An important concept to understand when interpreting your results is the **maximum likelihood estimate,** or MLE. The percentages listed in your certificate represent the most statistically likely ancestry mix. Other percentages are possible, though less likely, and these are shown in terms of confidence intervals in your bar graph (see Bar Graph section for details).

With a genetic test, ancestry can only be estimated in a statistical sense, much like the track of a hurricane. Anyone who lives in the Southeast United States knows that while it is impossible to project the track of a hurricane exactly, the "cones" of most likely migration are extremely accurate. The same is true with DNA Origins™.

Therefore, is it not possible to determine what your proportions are *exactly*. When drawing conclusions from DNA, one must use statistics. To determine ancestral ratios without using statistics, we would have to go back in time 200,000 years and keep track of every one of your ancestors since the origin of our species in Africa. Clearly, this is impossible. Since you inherited your DNA from your ancestors, your ancestral proportions are written in your DNA, but this information must be statistically inferred from the DNA sequence. If we measured 1,000 or 10,000 genetic sites in your DNA as opposed to the 200 or so we measure today, the confidence intervals shown in your bar graph would be smaller but the MLE would probably be very similar. The costs of measuring more genetic sites can quickly become out of hand, making the price to the consumer unrealistic.

In conclusion, the result of our test is your MLE. Though the MLE is a statistical estimate, and there is a small chance your true proportions are slightly different from that of the MLE, the MLE is the best estimate. If you want to keep it simple, use this MLE when describing your genetic heritage. An alternative to understanding more in terms of your family's heritage is to have more persons in your family (parents, grandparents, siblings) take the DNA Origins™ test.

Certificate/Data Table

The results of the MLE are reported in the data table, as in this example:

Estimate	Ancestry
90%	European
10%	Indigenous American
0%	Sub-Saharan African
0%	East Asian

The MLE percentages of European, Indigenous American, Sub-Saharan African and East Asian. In the example above, the person was determined to be 90% European and 10% Indigenous American.

Founding Populations

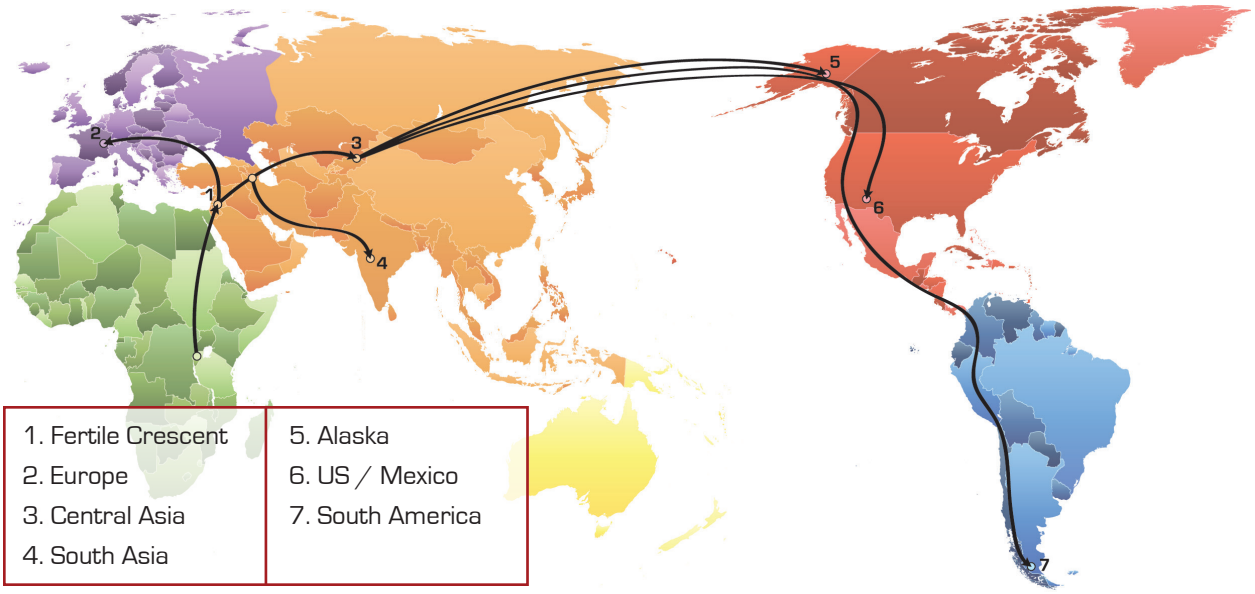
The test results provide ancestral estimates for the four founding populations: European, Indigenous American, Sub-Saharan African, and East Asian. These population groups are further defined as follows:

- **European.** This people group includes Europeans, Middle Easterners, and South Asians.
- **Indigenous American.** This group is composed of people who migrated to inhabit North, South and Central America.
- **Sub-Saharan African.** This group includes people with roots in the Sub-Saharan region of Africa
- **East Asian.** This people group includes the Japanese, Chinese, Koreans, and Pacific Islanders.

As you can tell from the brief descriptions, the names of the four founding populations listed in your test results are used in a simplistic sense. It is important to remember that these "founding populations" really refer to a group of people with shared ancestry who occupy certain geopolitical areas with "blurred boundaries."

For example, the "European" population denotes people of shared "proto-European" ancestry, which includes not only populations residing in the European continent, but also the Middle East and South Asia. This is based on evolutionary and anthropologic studies which show that these peoples' common ancestor arose from anatomically modern humans who travelled out of Africa about 50,000 years ago to colonize the Fertile Crescent area of the Middle East—what today encompasses the countries of Lebanon, Israel, Syria, Palestine, Jordan and Iraq, and including the land between the Tigris and Euphrates rivers.

Approximately 40,000 years ago this Fertile Crescent population branched to Europe and also likely mixed with South Asians, while founding populations in Central Asia (modern-day Kazakhstan, Kyrgyzstan, Tajikistan, Turkmenistan and Uzbekistan). About 10,000 years ago, Middle Eastern farmers spread to blend with the (by then-) indigenous Europeans, and much of the modern-day European gene pool is derived from this more recent source population. The same farming population also likely migrated in the opposite direction, contributing to modern-day South Asian Indian populations. In other genetic genealogical data, Y chromosome and mtDNA haplogroups (genetic populations of shared markers on the Y and mtDNA) are shared among the abovementioned populations.



Thus, people residing in Europe, Middle East, and South Asia (India) share common ancestral markers dating 10 to 50,000 years ago. In testing various world populations, for example, South Asian Indians are found to have a substantial, but lower, level of "European" markers while Middle Easterners exhibit a higher level of "European" markers. In the DNA Origins™ test, the average South Asian Indian exhibits 58% European ancestry, while Middle Easterners have about 80-90% European ancestry. Additional examples of test results from various world populations are provided in the chapter, "Average Results for Various Populations."

The Indigenous American founding population also calls for more explanation, because of this group's migration history.

Indigenous Americans came into the New World from Central Asia over many thousands of years in three major migrational waves: The first started possibly as long as 30,000 years ago and brought in the Amerind speakers. This is the largest group of Indigenous Americans in the Western Hemisphere, and over the ages they have intermingled with other populations like the Spanish invaders and other Indigenous Americans. They came across the Bering land bridge and migrated down the west coast all the way to the southern tip of South America—all the way to Patagonia—and it is thought that many did so by boat. Another migrational Indigenous American wave that crossed the Bering land bridge brought in the

Na-Denes (pronounced 'nah-dinnay'). They ultimately contributed to the Indigenous Americans found in Central America, Mexico, the United States, and Canada. This occurred around 16,000 years ago. The last group of Indigenous Americans to come over the Bering straits via the now-defunct land bridge were the Eskimo-Aleut speakers and they came around 6,000 years ago. There is no definitive break between any of these groups, and genetic markers are shared among these three Indigenous American groups.

Because of the migration pattern of Indigenous Americans, many people of Italian, Greek or Turkish heritage (and some Middle Easterners) may show "Indigenous American" ancestry—as much as ten percent—because of migrations from Central Asia south and west into those regions. Turkey was the passageway into Europe from Asia (the region occupied by Uzbekistan, Kazhakstan, Afghanistan, Southwestern Siberia, etc.). Central Asians do show overlapping markers with Indigenous Americans who migrated North and East to the Bering land bridge and into the New World.

Further, many people with European heritage should take into consideration the fact that the Roman (Italian) armies conquered and occupied much of Europe—including England—for more than 1,000 years, resulting in an intermingling of the genetics markers from that source. (Ireland, which is an island, was never occupied by the Romans, so many of the descendants of earlier "Erse" ancestors approach 100% in their European genetic portraits.) Thus, low levels of Indigenous American heritage could be detected in Europeans whose ancestors may have never set foot in the New World.

Percentages and Physical Appearance

Individuals exhibiting physical characteristics of a population group generally have at least 30-35% identity with that group. **For example, persons with an 85% European and 15% African generally exhibit few, if any, physical features characteristic of the African group, such as darker skin.**

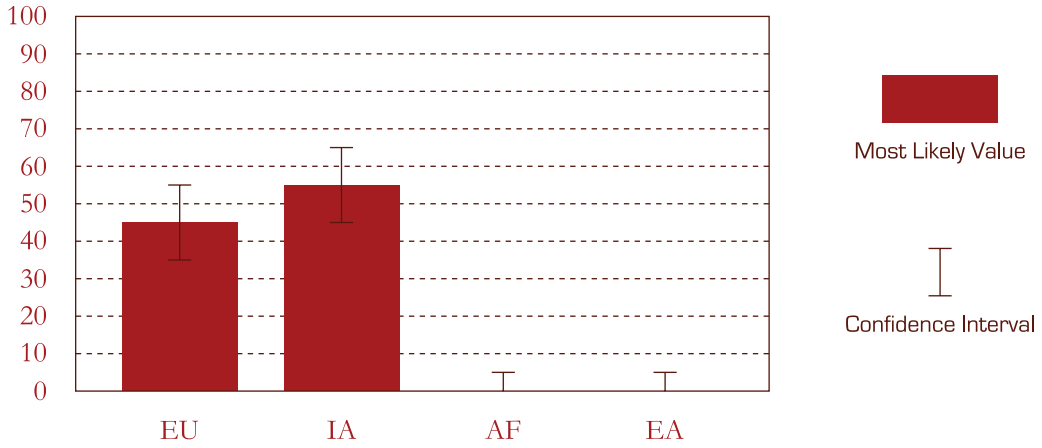
This is because the genes that determine physical appearance comprise a very small percentage of the total number of genes in the genome. Thus, for all of these genes to have sequences characteristic of one group, the person would need to be of relatively high proportions for that group. The higher the percentage of African a person is, the more likely the areas of the genome that determines physical appearance will be of African origin.

Bar Graph

In the sample percentage data table shown previously, the bar graph would show solid red bars at 90% for European and 10% Indigenous American. Unlike the table however, the bar graph will show the confidence regions around the MLE, which is, as we have mentioned, important for properly interpreting your results.

To generate the bar graph, we plotted the MLE values including the values within the 2-fold confidence range—that is, a 98% confidence level—one group at a time. The bar graph is a useful presentation because it provides a separate, objective view of your possible ancestry percentages for any one particular group—one group at a time.

For example, the bar graph below shows the results for a person of 2-population admixture. This person's most likely ancestry mix (MLE) is 45% European, 55% Indigenous American. There are confidence ranges extending above and below the tip of each bar.



These ranges show the other percentages this person could be, though any percentage is up to two times less likely than the percentage indicated by the red bar. Though this person is most likely to be 45% European, and 55% Indigenous American, they could also possibly be 35% European, 65% Indigenous American—though it is two times less likely than the MLE value of 45% European, 55% Indigenous American. You will notice that there are confidence bars for the East Asian (EA) and African (AF) groups too. Thus, it is possible that this person is 44% European, 54% Indigenous American, 1% Sub-Saharan African and 1% East Asian, though again, it is significantly less likely than the most likely estimate of 45% European, 55% Indigenous American.

Customers seeking to confirm a great-great-grandparent of Indigenous American ancestry who have obtained an MLE showing 100% European find the bar graph useful in understanding the statistical meaning of their results and how it may still be possible (though less likely) that there is a small amount of Indigenous American ancestry.

People of 1-population or 2-population admixture would most likely show one or two red bars, respectively. People with 3-population or 4-population mixture would most likely show 3 or 4 bars, respectively (For example, a 4-population admixture, such as 30% European, 20% Indigenous American, 30% African and 20% East Asian, might be obtained from a person with a Dominican father and a Philippine mother).

The confidence ranges are established for each ancestry group by searching the likelihood calculation and finding the highest and lowest values for each group that fall within the 2-fold likelihood range (0.3 Log base 10) of the MLE.

A person that shows 3 bars may still be a 4-way mix if there is a confidence range for that fourth group, its simply less likely (2 times less likely) that they are a 4-population mix incorporating a value for the fourth group within the range than the indicated 3-population mix.

Validation Studies

The science behind the DNA Origins™ test has been published in the scientific literature. We have determined the frequency of DNA sequence variants in the various human populations, and by determining your sequence for each; we can determine the probability that you identify with each group.

The test has been evaluated using a large number of people from a wide range of ancestral groups, and the estimates correspond well to what is known from anthropological and historical data. For example, Hispanics are known to have arisen as an ethnic group from the blending of colonial Europeans with Indigenous Americans, and the hundreds of Hispanics we have tested align with these two groups almost exclusively, as expected. As another example, though most Nigerians' results indicate unmixed African BioGeographical Ancestry (BGA), African Americans show more of a mixture between this group and Europeans, which is also what would be expected from what we know about the admixture between Africans and Europeans in the United States.

The method has also been validated through pedigree challenge; when the BGA is determined from a mother and father, that of their children should plot somewhere between the two. To date, we have tested numerous family pedigrees, and the ancestral proportions of offspring always plot somewhere among those of their parents. When outside agencies blindly test the MLE estimates, they prove to be excellent estimates of ancestral proportions. Below are some test cases that demonstrate the reliability of DNA Origins™ test results.

Experiment: DNA Origins™ Blind trials on samples from families.

Purpose: To determine how well the test results agree with expectations formed from appreciation of a family pedigree.

Results: When tested against known pedigrees, the DNA Origins™ test performs quite well.

Family 1

The data for a test individual, whom we will call **individual A**, is presented below. His wife, **individual B**, is Hispanic and she was determined to be of mostly Indigenous American ancestry but with some European and African heritage. This was also expected based on what we know from anthropological origin of the Hispanics (which were derived from the union of Spanish explorers, Indigenous Americans, and West Africans in Colonial Caribbean and Latin America).

Individual A	EUROPEAN	93	AFRICAN	0	INDIGENOUS AMERICAN	7
Individual B	EUROPEAN	7	AFRICAN	22	INDIGENOUS AMERICAN	71

Each of their 3 children **(C1, C2, and C3)** is plotted roughly half way amongst both parents, as expected. None of the children exhibit East Asian ancestry. The results of the children were consistent with those of the parents, and the MLE's are accurate estimates when tested against what is known from biographical data.

C1	EUROPEAN	47	AFRICAN	15	INDIGENOUS AMERICAN	38
C2	EUROPEAN	60	AFRICAN	2	INDIGENOUS AMERICAN	38
C3	EUROPEAN	57	AFRICAN	6	INDIGENOUS AMERICAN	37

Family 2

The father is F and mother is M. Both exhibit some Indigenous American admixture (8% and 12%, respectively). Their child, S, also exhibits some Indigenous American admixture (14%). Due to the law of independent assortment (the test markers span 22 chromosomes), these results are reasonable. For example, if one of the children measured with 75% Indigenous American, or 20% East Asian, these results would be unreasonable.

M	EUROPEAN	81	EAST ASIAN	7	INDIGENOUS AMERICAN	12
F	EUROPEAN	92	EAST ASIAN	0	INDIGENOUS AMERICAN	8
S	EUROPEAN	86	EAST ASIAN	0	INDIGENOUS AMERICAN	14

Simulations

In developing an admixture test such as DNA Origins™, we carefully considered how many AIMs and what quality of AIMs are required to minimize the incidence of individuals with MLEs of artificial or erroneous admixture. AIMs that are too low in quantity or too poor in quality could result in an inaccurate estimate of the MLE. To achieve a robust test, we subjected our AIMs to population simulations.

If we assume that the AIMs locations in the genome are unlinked (that is, are inherited independently, which they have been determined to be) and the mating between individuals of a given population are random, the multi-locus genotype frequencies in each population are determined from the allele frequencies and an equation known as the product rule. Using these allele frequencies, we simulate a population of 100,000 European, East Asian, African and Indigenous American individuals, draw 10,000 from each population and used our initial 71 and 144 marker DNA Origins™ tests (2.0 and 2.5, respectively) to calculate the BGA proportions for each simulated sample selected.

With a perfect test, using discretely distributed alleles, each simulated individual would type as 100% affiliation with their own group. In the real world, using markers that are continuously distributed, there is a level of statistical noise. The purpose of the simulations is to define what that level of noise should be expected.

Admixture Levels in Simulated Populations

Below we show the average percentages the DNA Origins™ test showed for each type of simulated population sample. The total admixture average is the sum of the average admixture percentages for samples of a particular simulated population (represented in the rows). For example, for the 71-marker test, we calculate the average level of European, East Asian and Indigenous American admixture in simulated Africans to be 0.96%, 0.1% and 0.87%, respectively.

DNA Origins™ 2.0 (71 Aim's)						
	AFR	EUR	EAS	IAM	Total	Total Admixture Average
African	98.07%	0.96%	0.1%	0.87%	100%	1.93%
European	0.08%	95.63%	2.25%	2.04%	100%	4.37%
East Asian	0.03%	2.45%	92.98%	4.54%	100%	7.02%
Indigenous American	0.01%	1.83%	3.63%	94.53%	100%	5.47%
						AVERAGE = 4.70%

DNA Origins™ 2.5 (144 Aim's)						
	AFR	EUR	EAS	IAM	Total	Total Admixture Average
African	98.21%	0.93%	0.71%	0.15%	100%	1.79%
European	0.4%	96.36%	1.5%	1.74%	100%	3.64%
East Asian	0.08%	1.43%	95.48%	3.01%	100%	4.52%
Indigenous American	0%	1.16%	2.08%	96.76%	100%	3.24%
						AVERAGE = 3.30%

From this table one can see that the average level of artificial admixture using the 71 marker test is about 5% and the level using the 144 marker test is lower, at about 3%. Looking at the European row in the 144-marker table, we can see that the average simulated European sample exhibits 1.5% East Asian ancestry and 1.74% Indigenous American ancestry. This suggests that the average (but of course, not every) real person who is 100% European will show 1.5% East Asian ancestry as statistical noise, and 3.64% non-European admixture in total as statistical noise. Some will show higher levels, and some will show 0%. One can use these values as a guide for interpreting their result.

For example, if a European suspects a small amount of IAM admixture using the 144 marker test and obtains a reading of 1%, they can see that the average simulated European has the same level, and so the data does not support (nor refute) the IAM admixture.

Variation of Results in Simulated Samples

We can look at the simulation data in another way – by looking at the variation of results in simulated samples. How common is it that a simulated European (or other) individual exhibits 15% or greater African (or other) ancestry? The results are shown below.

71-Marker Test					
>15%	AFR	EUR	EAS	IAM	Average Outside Group
African	100%	0.12%	0%	0.08%	0.07%
European	0.06%	100%	3.42%	3.06%	2.18%

East Asian	0%	2.9%	100%	12.25%	5.05%
Indigenous American	0%	1.77%	8.73%	100%	3.5%
					AVERAGE = 2.7%

144-Marker Test					
>15%	AFR	EUR	EAS	IAM	Average Outside Group
African	100%	0.048%	0%	0%	.016%
European	0.096%	100%	0.5268%	1.58%	.734%
East Asian	0%	0.165%	100%	5%	1.722%
Indigenous American	0%	0.1%	2.1%	100%	0.733%
					AVERAGE = 0.81%

We see that the about 2.7% of simulated samples show 15% or greater artificial admixture using the 71 marker test, and that less than 1% of simulated samples show 15% or greater artificial admixture using the 144 marker test. Reading across the European row, for the 144-marker test, we see that .5% of European individuals registered with 15% or greater East Asian admixture. One can use these values also as a guide for interpreting their result. For example, if a European suspects a small amount of IAM admixture using the 144 marker test and obtains a reading of 17%, this table shows that only 1.5% of simulated Europeans exhibited IAM affiliation this high, and so based on the average MLE, there is about a 98.5% chance that this result indicates real IAM admixture (of course this is based on the average MLE, and a customer should also refer to their individualized confidence contours on their triangle plot).

We can make tables like this for all of the possible values X>5%, X>10%, X>15% . . . all the way to X>50%. All 40,000 samples were completely affiliated with their own group at levels of 50% or greater. Down to X=5%, we see it is not uncommon for simulated samples to show affiliation with another group at this level of X. From all of these tables, we can compute the rough percentage value necessary to conclude with 95% certainty that partial group affiliation means real affiliation and not statistical noise.

Affiliation Percentage Thresholds

The tables below show the threshold of 95% confidence for each type of admixture in simulated individuals of homogeneous ancestry. A reading at or above X (where X is a % value in a cell of this table) means with 95% certainty that the reading is caused by affiliation with that group in the column, as opposed to the alternative, that the individual is really homogeneously affiliated with their group and the partial affiliation is the result of statistical noise. For example, using the 144 marker test, admixture greater than or equal to 10% Indigenous American is required for an individual of polarized (i.e. mainly) European ancestry to conclude with 95% confidence that there really is Indigenous American admixture as opposed to there being none (and no other admixture). Using the 71 marker test, one must see 12.5% IAM affiliation to conclude with 95% confidence that there really is Indigenous American admixture as opposed to there really being none (and no other admixture).
Of course, a customer could get a reading of 8% Indigenous American with the 71 marker test, which

is below the 12.5% level threshold, and it still be true that there is Indigenous American admixture. Such a person should type other members of their family. If the level of IAM increases going up the family tree, and if the level is above the threshold in some of the ancestors, it is probably a real indication of IAM admixture for the customer, even though the 8% is below the 95% confidence threshold. In fact, 8% falls near the 90% threshold (we don't show the 90% threshold on the website), not the 95% threshold, meaning that on its own (regardless of values in family members), the 8% is an indicator of IAM ancestry with 90% (not 95%) confidence.

Table SIMSUM144
Threshold of affiliation percentages for samples of polarized, binary affiliation, above which results indicate fractional affiliation with a $p < 0.05$, using the 144-marker admixture test.

	AFR	EUR	EAS	IAM
African	<3.0%	7%	5%	<3%
European	3.50%	<3.0%	9%	10%
East Asian	<3.0%	8%	<3.0%	12.50%
Indigenous American	<3.0%	7.50%	11.50%	<3.0%

Table SIMSUM71
Threshold of affiliation percentages for samples of polarized, binary affiliation, above which results indicate fractional affiliation with a $p < 0.05$, using the 71 AIM admixture test.

	AFR	EUR	EAS	IAM
African	<2%	8%	<2%	8%
European	<2%	<2%	13%	12.5%
East Asian	<2%	11.5%	<2%	17.5%
Indigenous American	<2%	9%	17.5%	<2%

Of course, there is nothing magical about a 95% vs. 90% confidence interval - and one might argue that genealogists rely on much lower levels of confidence considering non-genetic data. The threshold values required to conclude a bona-fide affiliation with 90% certainty are about 2/3 of those shown above.

Accuracy

The genotypes (nucleotide letters) we have determined for you are quite accurate. Because we use the latest genetic reading equipment available, we routinely achieve a greater than 99.99% accuracy for each site. In some cases, an accurate value could not obtained for you at a particular site, in which case we do not take that location into account when calculating your biogeographical ancestry. Having a few of these does not prevent us from making a good ancestry estimate, but of course having too many would—our laboratory maintains a standard threshold and will not release ancestry results if there are too many of these sites, which we call "failed loci (FL). Some reasons you may have an "FL" for a site include:

a) A small region of your chromosome around this site is missing or is of different sequence character than for most. This result is not uncommon given the highly variable nature of the chromosomal positions we measure, and it certainly does not imply you have any sort of defect in any way whatsoever (in fact, it may be an indication of your uniqueness).

b) We did not get enough DNA from your swab. Some markers are more sensitive to this than others. If there are too many "FLs" for your read-out, we will not be able to determine your ancestry proportions to a degree of accuracy that we would like, and in this case we will have to ask you to submit another sample for a second try.

Experiment: Repeated estimation from the same samples with DNA Origins™ 2.0(ABD).

Purpose: To determine how reproducible the results are by measuring the proportions in the same individuals on different occasions.

Results: Variation in percentages is a result of failed markers. This test shows a 5-6% variation for the absolute percentage in any one group. Since this experiment we have been using 5 samples on each run as internal controls. The average variation is 2-3% for these controls. Another group of 11 have also been tested repeatedly, and these show an average 2-3% variation for 10 samples, and an average 5% variation for the other sample. Best estimate from all of the data on repeated measurements is on order of 3-4% variations for most determinations if the individual tested has failed markers. Thus, if your profile came back as 96% European and 4% East-Asian, it is debatable whether the 4% East Asian is significant and would also be addressed by the confidence contours.

plate3-BD101-Data.INP	EUROPEAN	100	EAST ASIAN	0	INDIGENOUS AMERICAN	0
plate5-BD101-Data.INP	EUROPEAN	100	EAST ASIAN	0	INDIGENOUS AMERICAN	0
plate3-BD304-Data.INP	EUROPEAN	85	EAST ASIAN	15	INDIGENOUS AMERICAN	0
plate5-BD304-Data.INP	EUROPEAN	86	EAST ASIAN	14	INDIGENOUS AMERICAN	0
plate3-BD316-Data.INP	EUROPEAN	72	EAST ASIAN	27	INDIGENOUS AMERICAN	1
plate5-BD316-Data.INP	EUROPEAN	79	EAST ASIAN	20	INDIGENOUS AMERICAN	1
plate3-BD3162-Data.INP	EUROPEAN	100	EAST ASIAN	0	INDIGENOUS AMERICAN	0
plate5-BD3162-Data.INP	EUROPEAN	89	EAST ASIAN	5	INDIGENOUS AMERICAN	6
plate3-BD317-Data.INP	EUROPEAN	79	EAST ASIAN	21	INDIGENOUS AMERICAN	0
plate5-BD317-Data.INP	EUROPEAN	84	EAST ASIAN	16	INDIGENOUS AMERICAN	0

Special Considerations

Due to the statistical nature of the test, as mentioned earlier, your results are provided as maximum likelihood estimates. This comes from our inability to go back in time and measure precisely how and when admixture occurred in various parts of Europe.

In addition, there may be imperfections in the model we use to estimate admixture, because substantial

and directional genetic drift may have taken place between modern-day populations and the populations that admixed thousands of years ago. Genetic drift is a concept in evolution where a subset of a population attains more reproductive success, and therefore pass their genes/DNA markers on to succeeding generations more so than the rest of the original population. Scientists debate these issues all the time, and there is no one answer that is guaranteed to be correct. It would seem that the only way to estimate these errors is to compare expected and observed results for people with carefully documented genealogy, but even this is not possible. We cannot use genealogy information from the past few generations to evaluate results from an anthropological test that is looking back (potentially) thousands of years.

Since most genealogists do not have reliable information going back that far, we simply do not have access to the reference data with which we would need to compare performance against expectations and measure this error, or modify the test to eliminate it. When we run this test in particular, we are doing what meteorologists are doing when they calculate a hurricane track projection cone. The meteorologist cannot know for certain exactly where the storm will go, but he/she understands how the storms respond to major weather features (like fronts) well enough to form probability statements predicting the storm track. Historically, these predictions are usually quite impressive—they are fairly close to where the storms actually go. The same is true for the DNA Origins™ test—the results suggest that the estimates are fairly close to true values.

Additional Resources

The following resources are provided for further research in the field of genetic genealogy research.

Books

Molecular Photofitting: Predicting Ancestry and Phenotype Using DNA

Tony Frudakis PhD, 2007. 712 pages. Published by Academic Press.

Trace Your Roots with DNA: Use Your DNA to Complete Your Family Tree

Megan Smolenyak and Ann Turner, 2004. 256 pages. Published by Rodale Books.

The Seven Daughters of Eve

Bryan Sykes, 2002. 320 pages. Published by W.W. Norton & Co.

The Journey of Man: A Genetic Odyssey

Spencer Wells, 2004. 240 pages. Published by Random House Trade Paperbacks.

Deep Ancestry: Inside the Genographic Project

Spencer Wells, 2006. 256 pages. Published by National Geographic

Websites

The International Society of Genetic Genealogy

http://www.isogg.org/

Beginners can check out their "For Newbies" section on the left menu and join a listserve to ask questions from fellow enthusiasts in genetic genealogy. Local meetings and events are also held around the country by ISOGG speakers.

Journal of Genetic Genealogy

http://www.jogg.info/

An online journal that publishes articles on topics of general interest to the genealogical community, including mutation rates, geographic patterns in genetic data, information about haplogroups, and mtDNA and Y-chromosomal topics as well as new ancestry DNA testing tools.

The Genetic Genealogist

http://www.thegeneticgenealogist.com

An informative blog discussing current topics in genetic genealogy and ancestry testing; also offers an e-book on interpreting the results of genetic genealogy tests.

The National Genealogical Society

http://www.ngsgenealogy.org/

Find tutorials, research tips, conferences, and publications about the methods used in conventional family research.

Society Hall Directory

http://www.familyhistory.com/societyhall/search.asp

This directory allows you to search for your local genealogical societies by name, city, state, or zip.

Mitosearch

http://www.mitosearch.org/

An public-access online database of mtDNA sequences where you can find matches and potential relatives.

Ysearch

http://www.ysearch.org/

An public-access online database of Y-STR profiles where you can find matches and potential relatives.

Retrace your ancestors' footsteps.

Ancestry by DNA™