



All contents © 2006-2010 DNA Tribes. DNA Tribes®. DNA Tribes® patent pending analysis is available exclusively from DNA Tribes. All rights reserved.

A New Genetic Map of Living Humans in Interconnected World Regions

E. Valaitis¹ and L. Martin
Updated 4/29/2010

Table of Contents

Introduction.....2

About DNA Tribes®2

What are DNA Tribes® World Regions?2

How are DNA Tribes® World Regions Identified?2

How are World Regions Related?3

Genetic Affiliations of Individuals within World Regions5

American Indian Regions6

Caucasian (European and Near Eastern) Regions7

Sub-Saharan African and Central and South Asian Regions8

East Asian and Pacific Populations9

DNA Tribes® Europa: A More Detailed Analysis of Europe.....10

A Genetic Family Tree of European Sub-Regions11

Individual Genetic Affiliations within European Sub-Regions12

Score Ranges for Three Major U.S. Ethnic Groups by Percentile.....13

Caucasian (United States) (World Regions)14

Caucasian (United States) (Europa Sub-Regions)15

African-American (United States)16

Hispanic (United States)17

**Estimated Admixture in Latin American and Hispanic Populations
(Compiled May 2007).....18**

Further Discussion: Regions, Not Races.....19

¹ Dr. Eduardas Valaitis received his Doctorate in Statistics from Yale University, New Haven, CT.



All contents © 2006-2010 DNA Tribes. DNA Tribes®.
DNA Tribes® patent pending analysis is available
exclusively from DNA Tribes. All rights reserved.

Introduction

DNA Tribes® offers a genetic ancestry analysis that measures the likelihood of an autosomal STR DNA profile of an individual occurring among world genetic regions. These world regions have been identified by Dr. Eduardas Valaitis using proprietary statistical analysis of our global DNA database of over 950 living populations. This study provides an overview of these world regions and genetic affiliations for individuals within them.

About DNA Tribes®

DNA Tribes® specializes in genetic ancestry analysis, including both geographical analysis of world populations and the comparison of individuals to living populations and world regions. DNA Tribes®' proprietary analysis incorporates statistical techniques developed by Dr. Eduardas Valaitis, who received his Doctorate in Statistics from Yale University in 2005. Dr. Valaitis' background includes extensive work in multivariate analysis and classification, which involves identifying mathematical structure present within large and complex datasets. This expertise allows DNA Tribes® to perform a uniquely detailed and comprehensive analysis of world populations to identify genetic structure on an objective mathematical basis.

What are DNA Tribes® World Regions?

Each world region represents a unique genetic family within the human species shaped by shared history and geography. These regions are characterized by distinctive patterns of allele frequencies across the studied STR loci. Although all humans are connected by ancient common origins, each of these genetic groupings shares a unique relationship due to more intense and persistent contacts within a geographical area. DNA Tribes® offers a detailed analysis that distinguishes these genetic families on an objective mathematical basis without relying on presumed racial or ethnic categories.

How are DNA Tribes® World Regions Identified?

Instead of relying on socially constructed racial or ethnic divisions, DNA Tribes® defines world regions using objective mathematical criteria. A proprietary statistical method is applied to over 950 individual population samples around the world to identify groups of populations with shared genetic characteristics. These genetic groups are then plotted on a map and named according to the geographical regions they occupy.

Map of DNA Tribes® World Regions

Figure 1 illustrates the geographical boundaries of the world regions identified by DNA Tribes® analysis. Within the borders of nearly all regions, individuals can be found with genetic ties to neighboring and sometimes distant regions.

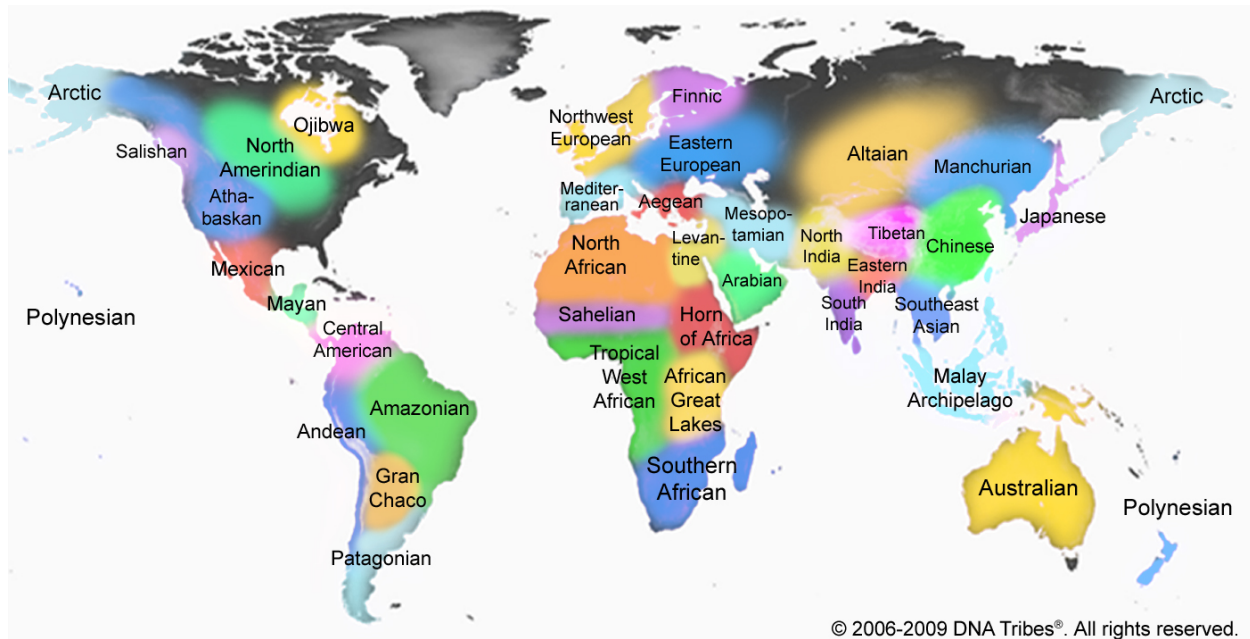


Figure 1: Map of world regions identified by DNA Tribes® analysis as of February 2010. More detailed information is available at: <http://dnatribes.com/populations.html>.

How are World Regions Related?

Method: Hierarchical clustering was performed on world region clusters using Euclidean distance, illustrated in **Figure 2** below.

Discussion: **Figure 2** illustrates the relationships among world regions identified by DNA Tribes®. Closely related regions connect towards the bottom of the diagram. For instance, the Mediterranean and Aegean regions are two closely related regions. The largest divisions appear towards the top (root) of the tree: for instance, the Australian region is only distantly related to other Eurasian regions and branches off alone towards the top of the tree diagram.

Individual regions group to form families and super-families of regions. Most of these larger groupings correspond to major continents. For instance, all East Asian regions form their own family. This East Asian family is part of a larger Eurasian super-family that also includes Caucasian (European and Near Eastern) regions (sometimes called “West Eurasian”). Similarly, all American Indian regions are part of their own super-family that is distinct from the other super-family (labeled “Non-American Indian”) that includes all other world regions.

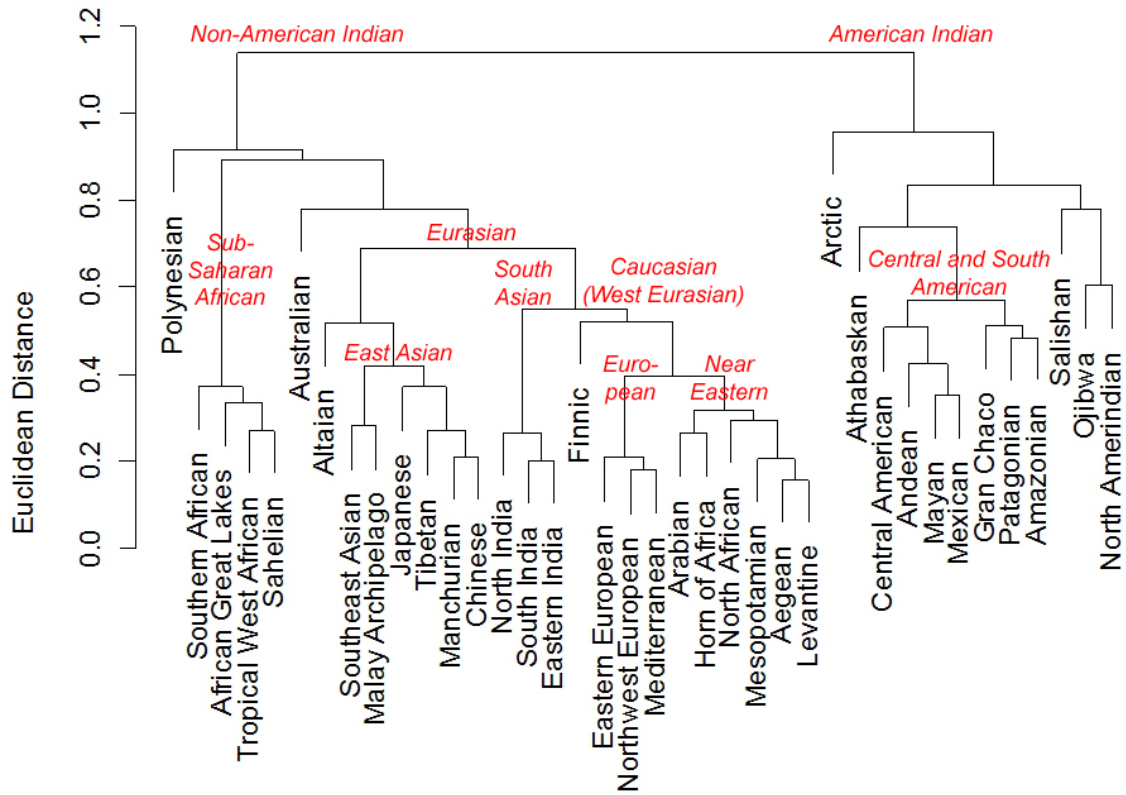


Figure 2: Neighbor joining tree illustrating the relationships among world regions based on Euclidean distance.

The relationships illustrated in **Figure 2** are the cumulative product of (1) genetic contact within each region created by migrations, intermarriage, and gradual diffusion; and (2) relative isolation from other regions. Natural features that make these contacts easier or more difficult can determine regional genetic relationships: waterways, mountain regions, fertile plains, and continental borders shape the pathways of human interactions that create both cultural areas and genetic regions. For instance, the historical difficulty of travel between Asia and North America corresponds to the great distance between the American Indian super-family and all other world regions.



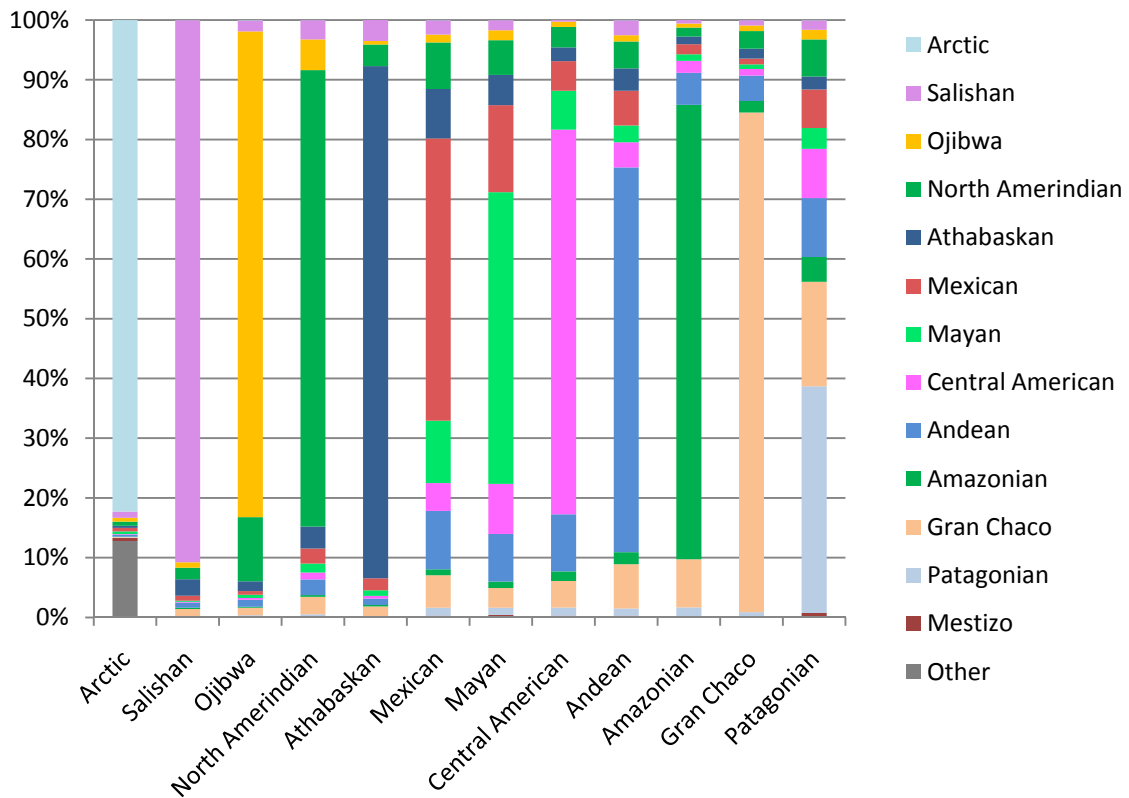
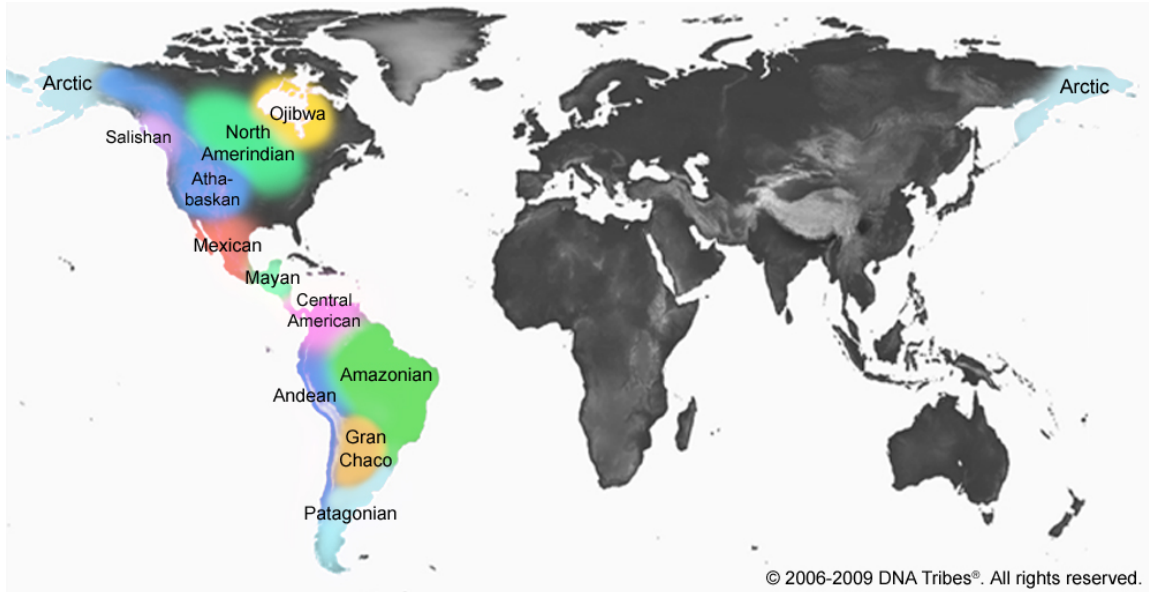
Genetic Affiliations of Individuals within World Regions

Method: Observed allele frequency data was used to simulate 5,000 individual genetic profiles for the studied world regions. Each simulated profile was processed by the DNA Tribes® algorithm, which measured the occurrence frequency of each simulated individual's 27 marker STR profile in each world region. The strongest regional match was then identified for each individual. These primary matches were then tallied for all simulated profiles to produce regional affiliation proportions.

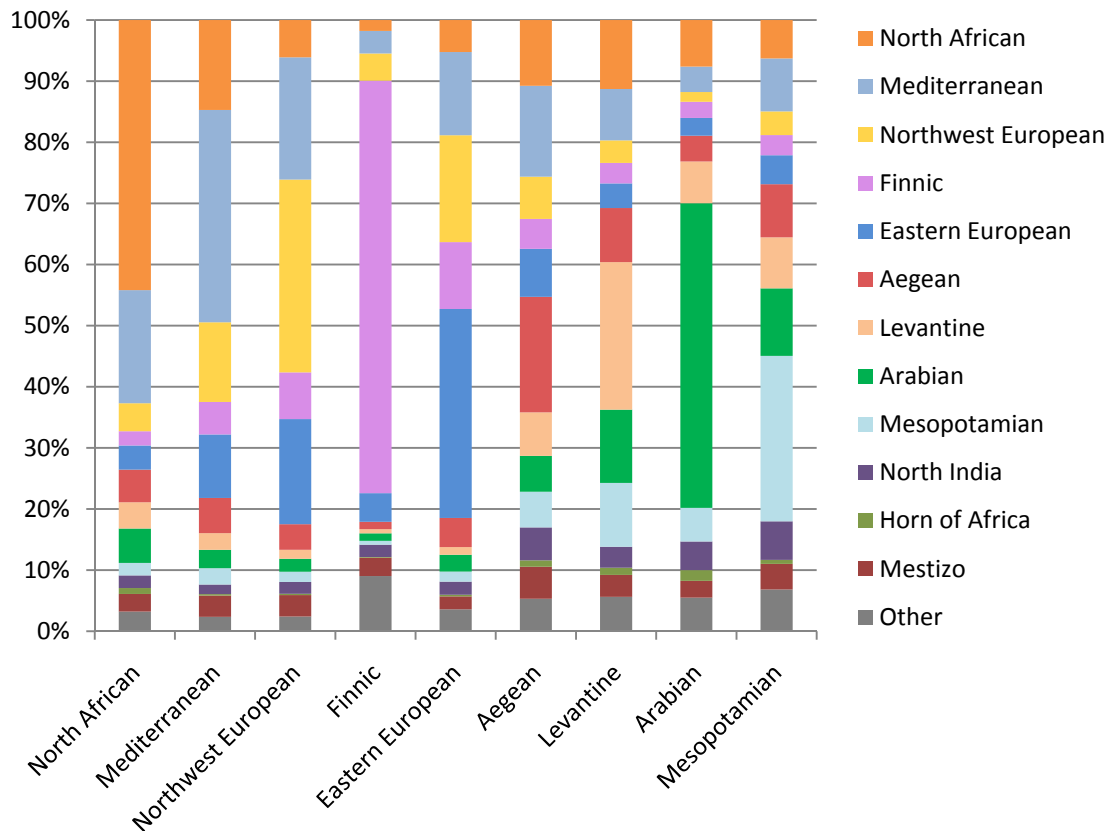
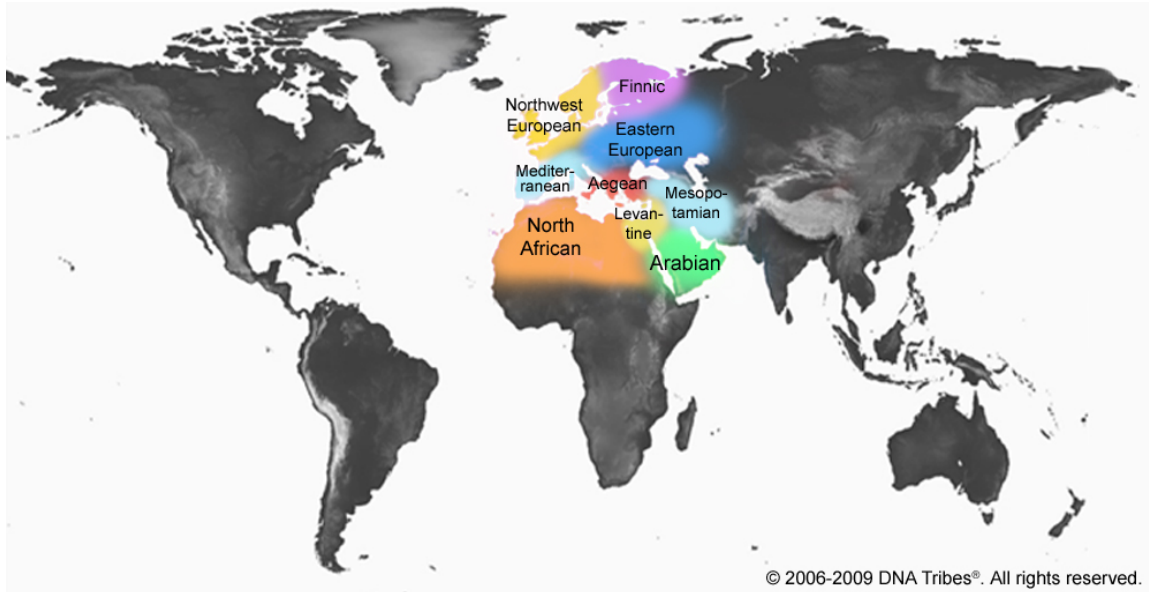
Individuals within each genetic region obtain a spectrum of regional genetic affinities. The study on the following pages illustrates the composition of individuals within each genetic world region:

American Indian Regions.....	7
European and Near Eastern Regions.....	8
Sub-Saharan African and Central and South Asian Regions.....	9
East Asian and Pacific Regions.....	10

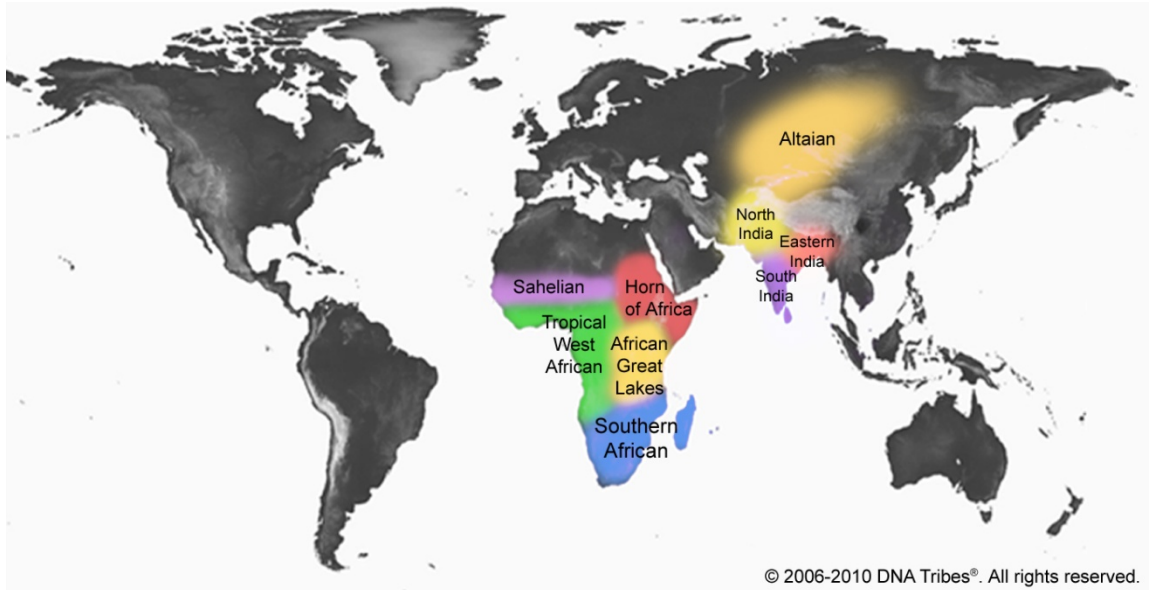
American Indian Regions



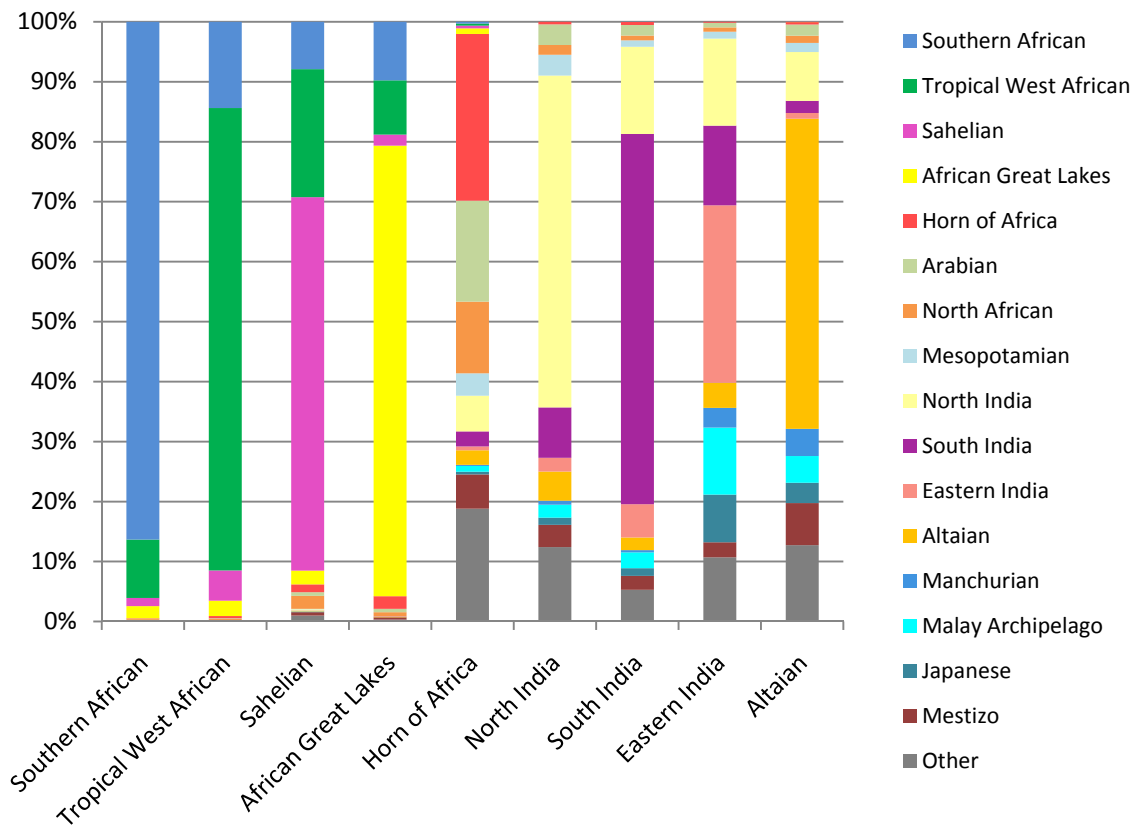
Caucasian (European and Near Eastern) Regions



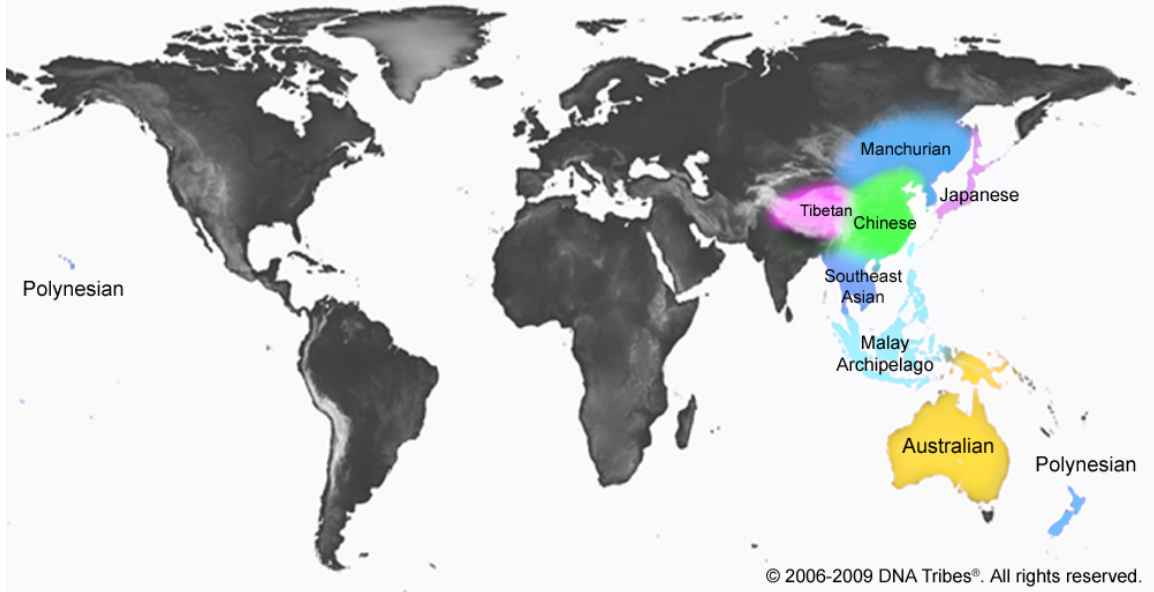
Sub-Saharan African and Central and South Asian Regions



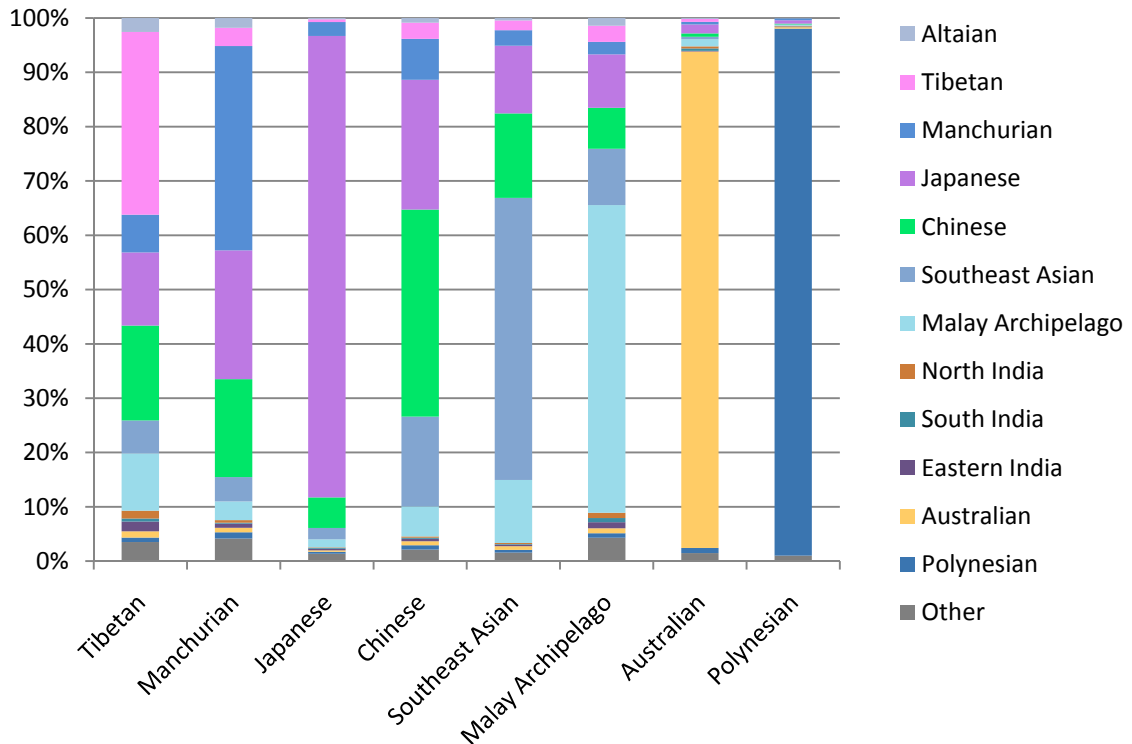
© 2006-2010 DNA Tribes®. All rights reserved.



East Asian and Pacific Populations



© 2006-2009 DNA Tribes®. All rights reserved.



DNA Tribes® Europa: A More Detailed Analysis of Europe

DNA Tribes® Europa add-on analysis performs a more detailed comparison of a person's own DNA to genetic sub-regions of Europe (illustrated in **Figure 3**). This allows a more specific identification of genetic relationships in Europe, including the identification of local or private genetic characteristics not otherwise typical of major European world regions.

As with world regions, these Europa sub-regions are distinguished using an objective mathematical algorithm, and reflect long-term geographical and/or ethnic relationships within Europe. Some of these genetic territories correspond to modern political boundaries, while others also reflect more ancient geographical relationships. For instance, the Greek genetic region includes not only modern Greece, but also provinces of Southern Italy that were once part of *Magna Græcia* (Greater Greece) in the classical world.



Figure 3: Map of genetic sub-regions identified by DNA Tribes® Europa analysis. More information about DNA Tribes® Europa analysis is available at: <http://dnatribes.com/dnatribes-europa.html>.

A Genetic Family Tree of European Sub-Regions

The following diagram illustrates the genetic relationships within DNA Tribes® Europa sub-regions.

Method: Hierarchical clustering was performed on European sub-region clusters based on Euclidean distance, illustrated in **Figure 4**.

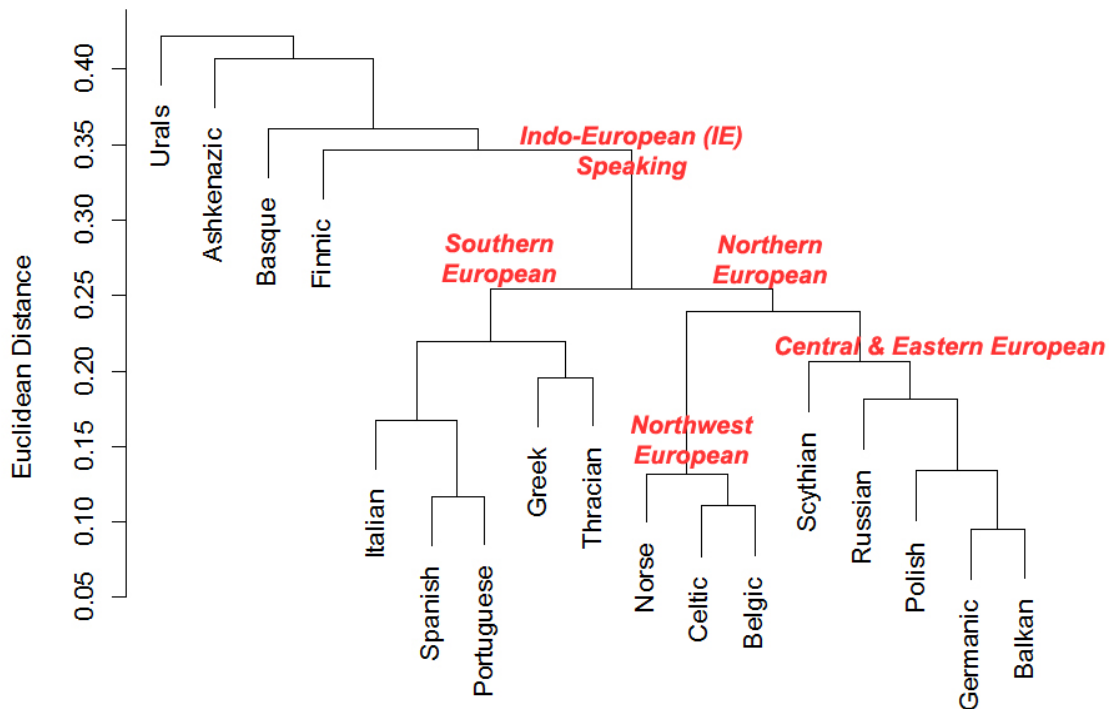


Figure 4: Neighbor joining tree illustrating the relationships among Europa sub-regions based on Euclidean distance.

Discussion: **Figure 4** illustrates the relationships between European sub-regions identified by DNA Tribes® Europa analysis. Closely related sub-regions appear towards the bottom of the diagram. For instance, the Germanic and Balkan regions are the two most closely related regions. The largest divisions appear at the top (root) of the tree. For instance, the Finnic, Basque, Ashkenazic, and Urals sub-regions are each only distantly related to each other and to the large family of more closely related Indo-European (IE) speaking sub-regions. The distinctive genetic patterns found within these outlier sub-regions reflect relative isolation from other sub-regions due to cultural endogamy and/or geographical isolation.

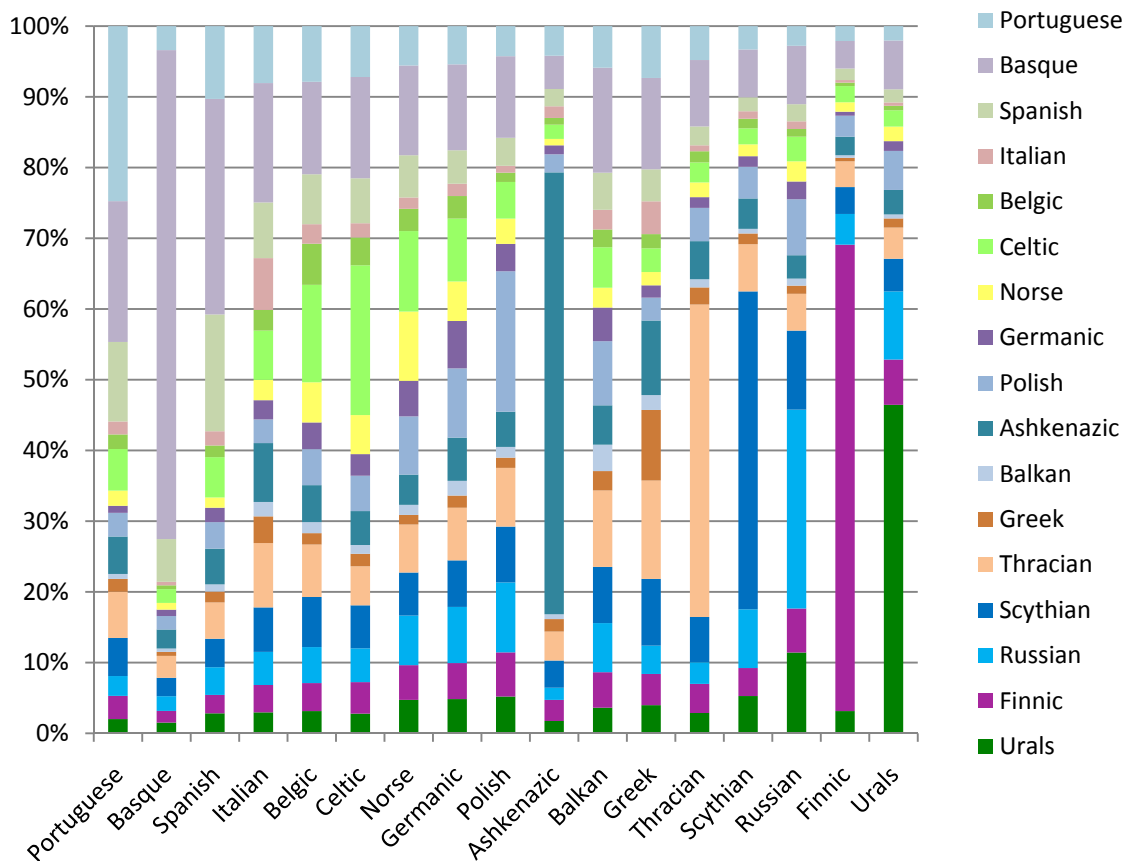
Individual Genetic Affiliations within European Sub-Regions

Method: Observed allele frequency data was used to simulate 5,000 individual genetic profiles for the studied Europa sub-regions. Each simulated profile was processed by the DNA Tribes® algorithm, which measured the occurrence frequency of each simulated individual's 27 marker STR profile in each European sub-region. The strongest regional match was then identified for each individual. These primary matches were then tallied for all simulated profiles to produce regional affiliation proportions.

Individuals within each Europa sub-region obtain a spectrum of regional genetic affinities. Because of close genetic relationships within Europe, individuals in a sub-region can inherit genetic material that is most common in other sub-regions.

Individuals from sub-regions with a history of ethnic endogamy or geographic isolation (such as the Ashkenazic, Basque, or Finnic sub-regions) exhibit higher frequencies of primarily in-group genetic affiliation. Individuals from centrally located sub-regions, such as the Balkan or Germanic regions, exhibit more variety in genetic affiliations and lower frequencies of primarily in-group genetic affiliation.

The follow study illustrates the range of individual affiliations observed within each European sub-region:





Score Ranges for Three Major U.S. Ethnic Groups by Percentile

Individuals within each population exhibit a characteristic range of world region scores. The following tables list 27 marker MLI scores by percentile for Caucasians, African-Americans and Hispanics living in the United States. These percentiles provide the range of expected scores for individuals within each group.

Caucasian (U.S.A.) (World Regions).....	15
Caucasian (U.S.A.) (Europa Sub-Regions).....	16
African-American (U.S.A.).....	17
Hispanic (U.S.A.).....	18

Caucasian (United States) (World Regions)

World Region	5% (very low)	25% (ordinary)	50% (ordinary)	75% (ordinary)	95% (very high)
Northwest European	12.6	285.0	2,433.7	19,534.6	501,431.2
Mediterranean	10.8	248.5	1,927.6	16,033.7	369,051.5
Eastern European	9.6	224.4	1,868.5	14,207.3	365,274.3
Aegean	11.5	155.0	1,022.1	6,806.9	119,631.9
North African	3.4	69.8	617.6	5,592.8	136,881.2
Finnic	3.0	64.8	471.6	3,413.2	84,207.1
Levantine	3.7	52.0	381.7	2,873.7	51,689.7
Mesopotamian	3.4	45.1	303.4	2,076.1	37,248.1
Mestizo	1.7	23.3	133.0	764.6	10,147.9
Arabian	0.9	16.5	131.1	1,059.2	26,680.3
North India	1.0	10.8	52.7	245.5	2,240.6
Altaian	1.1	8.6	37.0	155.4	1,208.6
Horn of Africa	0.3	4.3	29.0	198.2	3,212.1
South India	0.1	0.6	2.9	12.1	125.8
Eastern India	0.1	0.5	1.7	4.9	25.8
Malay Archipelago	0.0	0.0	0.1	0.8	10.0
Sahelian	0.0	0.0	0.0	0.7	36.5
Tibetan	0.0	0.0	0.0	0.3	3.2
African Great Lakes	0.0	0.0	0.0	0.2	18.9
Arctic	0.0	0.0	0.0	0.1	2.5
Australian	0.0	0.0	0.0	0.0	0.8
Manchurian	0.0	0.0	0.0	0.0	1.3
Chinese	0.0	0.0	0.0	0.0	0.9
Southeast Asian	0.0	0.0	0.0	0.0	0.9
Tropical West African	0.0	0.0	0.0	0.0	2.5
Japanese	0.0	0.0	0.0	0.0	0.6
Polynesian	0.0	0.0	0.0	0.0	0.2
Southern African	0.0	0.0	0.0	0.0	1.0
Mayan	0.0	0.0	0.0	0.0	0.3
Mexican	0.0	0.0	0.0	0.0	0.3
Patagonian	0.0	0.0	0.0	0.0	0.3
Central American	0.0	0.0	0.0	0.0	0.2
Andean	0.0	0.0	0.0	0.0	0.1
Ojibwa	0.0	0.0	0.0	0.0	0.1
North Amerindian	0.0	0.0	0.0	0.0	0.0
Gran Chaco	0.0	0.0	0.0	0.0	0.0
Salishan	0.0	0.0	0.0	0.0	0.0
Athabaskan	0.0	0.0	0.0	0.0	0.0
Amazonian	0.0	0.0	0.0	0.0	0.0

For instance, half of all U.S. Caucasians obtain a Northwest European score between 285.0 and 19,534.6, with a median of 2,433.7. Northwest European scores within this range could be considered ordinary for Caucasian individuals. Only five percent of Caucasians obtain a score either below 12.6 or above 501,431.2. These could be considered very low and very high scores within this ethnic group, respectively.

Caucasian (United States) (Europa Sub-Regions)

Europa Sub-Region	5%	25%	50%	75%	95%
	(very low)	(ordinary)	(ordinary)	(ordinary)	(very high)
Germanic	14.0	310.4	2,621.3	21,090.3	523,082.6
Celtic	11.9	294.8	2,540.1	21,329.3	604,227.1
Belgic	12.2	298.5	2,475.8	20,737.7	519,898.4
Norse	13.2	288.5	2,432.5	20,486.9	514,214.0
Italian	13.4	262.9	1,989.0	15,079.6	336,216.3
Balkan	11.1	242.8	1,958.7	14,771.3	382,214.7
Polish	8.5	208.7	1,802.3	14,423.3	419,997.3
Spanish	7.5	195.9	1,725.4	14,936.3	384,810.7
Portuguese	7.6	181.2	1,640.0	14,092.8	383,517.9
Russian	4.5	137.1	1,168.8	10,026.1	291,489.8
Greek	9.4	144.6	1,124.9	8,190.2	159,330.2
Thracian	6.2	106.2	932.1	6,591.8	151,651.1
Scythian	4.2	77.5	788.6	6,047.4	156,271.8
Basque	1.5	63.6	729.3	9,113.2	336,800.6
Ashkenazic	2.1	58.8	613.2	6,710.2	177,482.0
Finnic	3.0	64.8	471.6	3,413.2	84,207.1
Urals	0.7	29.9	338.2	3,541.5	114,709.5

For instance, half of all U.S. Caucasians obtain a Norse score between 310.4 and 21,090.3, with a median of 2,621.3. Celtic scores within this range could be considered ordinary for Caucasian individuals. Only five percent of Caucasians obtain a score either below 14.0 or above 523,082.6. These could be considered very low and very high scores within this ethnic group, respectively.

African-American (United States) (World Regions)

World Region	5% (very low)	25% (ordinary)	50% (ordinary)	75% (ordinary)	95% (very high)
Tropical West African	16.9	4,297.0	368,386.8	26,120,830.0	18,153,554,035.9
Sahelian	22.8	2,688.8	101,953.9	3,847,128.4	1,186,692,424.8
Southern African	0.8	437.7	37,140.4	3,353,569.0	2,978,268,492.9
African Great Lakes	3.5	696.3	33,225.5	1,368,404.4	484,979,001.5
Horn of Africa	21.5	460.9	3,820.5	38,187.1	1,055,845.0
North African	1.4	42.2	390.0	4,327.2	116,693.8
Levantine	0.6	17.3	135.2	1,088.1	23,022.5
Arabian	0.5	13.5	134.2	1,350.9	45,222.1
Aegean	0.3	6.1	48.7	375.7	7,082.3
Mestizo	0.4	6.4	41.5	281.7	4,148.6
Mediterranean	0.1	2.6	25.3	238.6	6,701.4
Mesopotamian	0.1	2.2	16.1	130.2	3,074.3
Northwest European	0.0	0.7	8.2	84.7	2,469.3
Eastern European	0.0	0.4	4.6	58.1	1,942.0
North India	0.0	0.3	1.9	12.5	154.9
Finnic	0.0	0.1	1.2	14.4	405.0
Altaiian	0.0	0.1	0.7	4.6	54.8
South India	0.0	0.1	0.6	3.5	43.1
Eastern India	0.0	0.0	0.0	0.2	1.7
Malay Archipelago	0.0	0.0	0.0	0.0	0.5
Tibetan	0.0	0.0	0.0	0.0	0.3
Polynesian	0.0	0.0	0.0	0.0	1.2
Arctic	0.0	0.0	0.0	0.0	0.4
Australian	0.0	0.0	0.0	0.0	0.1
Manchurian	0.0	0.0	0.0	0.0	0.1
Chinese	0.0	0.0	0.0	0.0	0.0
Southeast Asian	0.0	0.0	0.0	0.0	0.1
Mexican	0.0	0.0	0.0	0.0	0.1
Japanese	0.0	0.0	0.0	0.0	0.0
Mayan	0.0	0.0	0.0	0.0	0.1
Patagonian	0.0	0.0	0.0	0.0	0.0
Central American	0.0	0.0	0.0	0.0	0.0
Andean	0.0	0.0	0.0	0.0	0.0
Ojibwa	0.0	0.0	0.0	0.0	0.0
Salishan	0.0	0.0	0.0	0.0	0.0
North Amerindian	0.0	0.0	0.0	0.0	0.0
Athabaskan	0.0	0.0	0.0	0.0	0.0
Gran Chaco	0.0	0.0	0.0	0.0	0.0
Amazonian	0.0	0.0	0.0	0.0	0.0

For instance, half of all African-Americans obtain a Tropical West African score between 4,297.0 and 26,120,830.0, with a median of 368,386.8. Tropical West African scores within this range could be considered ordinary for African-American individuals. Only five percent of African-Americans obtain a score either below 16.9 or above 18,153,554,035.9. These could be considered very low and very high scores within this ethnic group, respectively.

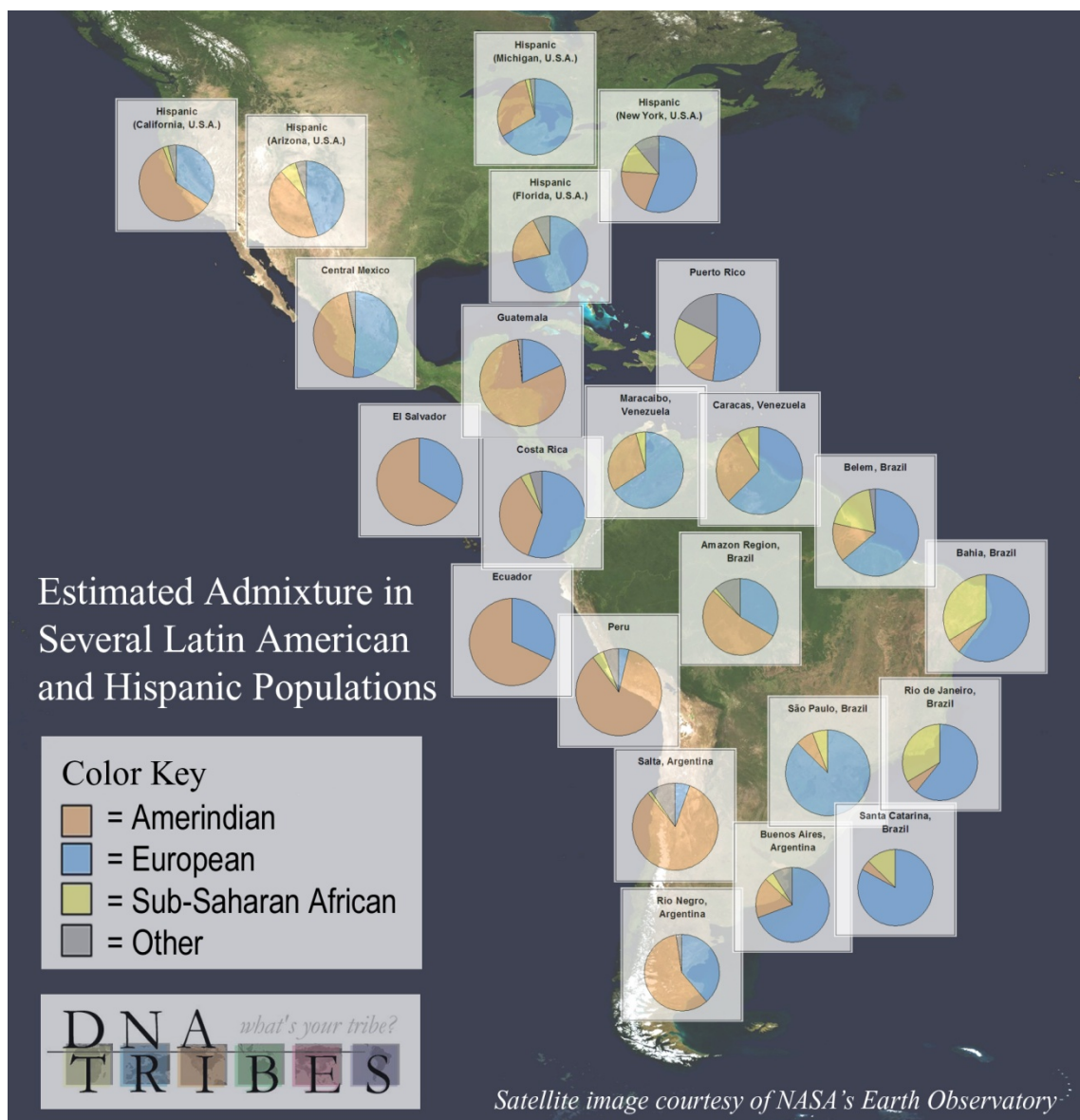
Hispanic (United States)

World Region	5%	25%	50%	75%	95%
	(very low)	(ordinary)	(ordinary)	(ordinary)	(very high)
Mestizo	8.5	90.1	534.6	3,419.3	53,698.7
Aegean	0.3	3.8	22.4	151.6	2,526.9
Mediterranean	0.1	1.7	15.2	118.2	3,088.9
Northwest European	0.1	1.7	14.8	117.7	2,901.3
Eastern European	0.1	1.4	12.0	90.4	2,156.1
Finnic	0.1	1.6	12.0	89.7	1,958.6
Altaian	0.3	2.7	11.5	45.9	469.2
Levantine	0.1	1.7	11.1	84.4	1,592.6
North African	0.1	1.0	8.5	74.5	2,109.2
Mesopotamian	0.1	1.3	8.4	56.1	1,023.1
North India	0.1	1.0	4.7	20.7	168.8
Arabian	0.0	0.6	4.3	32.8	766.6
Horn of Africa	0.0	0.6	3.6	22.8	407.8
South India	0.0	0.2	0.9	3.9	33.1
Mexican	0.0	0.0	0.7	28.1	4,495.9
Eastern India	0.0	0.2	0.6	2.0	10.0
Mayan	0.0	0.0	0.5	24.1	2,735.1
Malay Archipelago	0.0	0.0	0.2	1.0	13.8
Central American	0.0	0.0	0.2	7.8	1,423.2
Arctic	0.0	0.0	0.1	1.8	63.4
Patagonian	0.0	0.0	0.1	4.7	875.7
Tibetan	0.0	0.0	0.1	0.6	7.1
Andean	0.0	0.0	0.1	4.6	1,279.1
Manchurian	0.0	0.0	0.0	0.2	7.4
North Amerindian	0.0	0.0	0.0	1.3	647.2
Chinese	0.0	0.0	0.0	0.1	4.8
Sahelian	0.0	0.0	0.0	0.1	7.5
Ojibwa	0.0	0.0	0.0	0.5	276.9
Southeast Asian	0.0	0.0	0.0	0.1	4.7
African Great Lakes	0.0	0.0	0.0	0.1	12.1
Gran Chaco	0.0	0.0	0.0	0.4	174.1
Japanese	0.0	0.0	0.0	0.1	3.8
Australian	0.0	0.0	0.0	0.0	0.7
Athabaskan	0.0	0.0	0.0	0.1	119.2
Salishan	0.0	0.0	0.0	0.1	65.4
Polynesian	0.0	0.0	0.0	0.0	0.8
Tropical West African	0.0	0.0	0.0	0.0	1.6
Southern African	0.0	0.0	0.0	0.0	0.6
Amazonian	0.0	0.0	0.0	0.0	1.0

For instance, half of all U.S Hispanics obtain a Mestizo score between 90.1 and 3,419.3, with a median of 534.6. Mestizo scores within this range could be considered ordinary for U.S. Hispanic individuals. Only five percent of U.S. Hispanics obtain a score either below 8.5 or above 53,698.7. These could be considered very low and very high scores within this ethnic group, respectively.

Estimated Admixture in Latin American and Hispanic Populations (Compiled May 2007)

Introduction: Latin American is the historic meeting place of indigenous Amerindian, European, and African populations. Contribution from each of these sources varies between Latin American populations. Amerindian contribution is greatest along the Pacific coastline, peaking in the Central American isthmus and the Andes. European contribution is greatest along the Atlantic coastline, peaking in southern Brazil. African contribution is also greatest along the Atlantic coast, peaking in northeastern Brazil. The map below displays estimated admixture from these major sources in living populations.



Further Discussion: Regions, Not Races

For several decades, scientists have known that geographical variation exists within the human species. People around the world share genetic traits with their neighbors that distinguish them from peoples living further away. However, scientists have also known that 3-5 simple racial categories are insufficient to describe the genetic diversity of living human populations. Moreover, scientists have documented that genetic variation within ethnic groups or races exceeds the differences between groups. The complex empirical reality of world populations that are overlapping and yet distinct implies the question: how can the genetic landscape of mankind be described, and how can individuals meaningfully be compared to entire populations? DNA Tribes® provides a novel answer to this question by describing the world not as an oversimplified checkerboard of racial divisions, but as a complex mosaic of overlapping world regions.

Traditional anthropology once classified four races corresponding to four major continents: African, European, Asian and American. This simple system of classification dates back to the 18th century taxonomist Carolus Linnaeus and is still commonly used when describing ethnic groups and individuals. Certain areas of each continent are traditionally designated as representatives of each race, and other regions are assumed to be mixed between these presumed pure areas. Since molecular DNA has provided a more objective way of charting human relationships, scientists have known that while inter-population differences do exist, traditionally defined racial groups do not exist as exclusive or pure genetic units.

However, interpretations of molecular DNA continue to use a traditional racial scheme in the absence of a more fully descriptive global model. The genetic differences between peoples traditionally identified as Black, White, Asian, and American Indian in North America are great enough to allow a rough estimate of an individual's "percentage" membership in each racial group. This approach has been used for medical and police applications as well as for individuals interested in learning more about their genetic ancestry.

However, this racial scheme creates problems when used outside of the core regions ancestral to modern North Americans. For instance, a Subcontinental Indian or Samoan can be classified as some percentage of American Indian, European, East Asian and Sub-Saharan Africa, but the result would lack meaning. More significantly, technical reliance on a simplified racial admixture model of ancestry potentially misses an opportunity to expand the scientific and popular understanding of human genetic relationships on an empirical basis.

In addition, gene flow between human populations has been continuous throughout most of history. Many ethnic crossings now thought of as admixture are reiterating crossings that have taken place in history and prehistory. For instance, gene flow between Sub-Saharan African and Caucasian (European and Near Eastern) populations has taken place in areas in and near the Sahara Desert (including the North African and East African genetic regions) throughout history. The blending of African and European peoples that has taken place since the colonial era in North, Central and South America is usually thought of as a novel crossing, but from a broader perspective is recapitulating an ancient pattern of trans-Saharan gene flow. In contrast, the



All contents © 2006-2010 DNA Tribes. DNA Tribes®.
DNA Tribes® patent pending analysis is available
exclusively from DNA Tribes. All rights reserved.

combination of American Indian and European and/or African populations is a novel type of crossing without a known historical precedent, and characterizes the relatively young Mestizo world region.

As the global survey in the preceding pages illustrates, the four-fold racial division is insufficient to fully describe the diversity of the human species. The world regions identified by DNA Tribes® offer a novel approach to answering the question by objectively identifying groups of populations based on a statistical analysis of neutral genetic markers. The result is a network of world regions each characterized by shared history and genetic patterns. The geographical outlines of these regions echo the borders of countless empires, trade networks and kin groups. Most regions reflect processes that predate all known ethnic, linguistic, or political forms, and do not reflect any single historical event, such as a single empire or migration. Instead, they describe the cumulative outcome of all historical and prehistoric interactions.

DNA Tribes® Digest articles discuss DNA Tribes® original empirical analysis of world genetic relationships on an ongoing basis. Past issues of DNA Tribes® Digest are archived online at: <http://dnatribes.com/library.html>.