



A New Genetic Map of Living Humans in Interconnected World Regions

E. Valaitis¹ and L. Martin
Revised 03/10/2008

Table of Contents

Introduction.....	2
About DNA Tribes	2
What are DNA Tribes® World Regions?	3
How are DNA Tribes® World Regions Identified?	3
How are World Regions Related?	4
Genetic Affiliations of Individuals within World Regions	6
American Indian Regions.....	7
European and Near Eastern Regions.....	8
Sub-Saharan African and Central and South Asian Regions.....	9
East Asian and Pacific Populations.....	10
DNA Tribes® Europa: A More Detailed Analysis of Europe.....	11
A Genetic Family Tree of European Sub-Regions.....	12
Individual Genetic Affiliations within European Sub-Regions	13
Score Ranges for Three Major U.S. Ethnic Groups by Percentile.....	14
Caucasian (United States) (World Regions).....	15

¹ Dr. Eduardas Valaitis received his Doctorate in Statistics from Yale University, New Haven, CT and has been an Assistant Professor in the Department of Mathematics and Statistics at American University, Washington, D.C. He is currently a Senior Associate with PricewaterhouseCoopers.



Caucasian (United States) (Europa Sub-Regions)	16
African-American (United States).....	17
Hispanic (United States).....	18
Estimated Admixture in Latin American and Hispanic Populations	19
Further Discussion: Regions, Not Races.....	20

Introduction

DNA Tribes® offers a genetic ancestry analysis that measures the likelihood of an autosomal STR DNA profile of an individual in world genetic regions (displayed in the map below). These world regions have been identified by Dr. Eduardas Valaitis using proprietary statistical analysis of our global DNA database of over 800 living populations. This study provides an overview of these world regions and genetic affiliations for individuals within them.

About DNA Tribes

DNA Tribes® is a private firm specializing 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 has been an Assistant Professor in the Department of Mathematics and Statistics at American University in Washington, D.C., and is currently a Senior Associate at PricewaterhouseCoopers. 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.

DNA *what's your tribe?* TRIBES

What are DNA Tribes® World Regions?

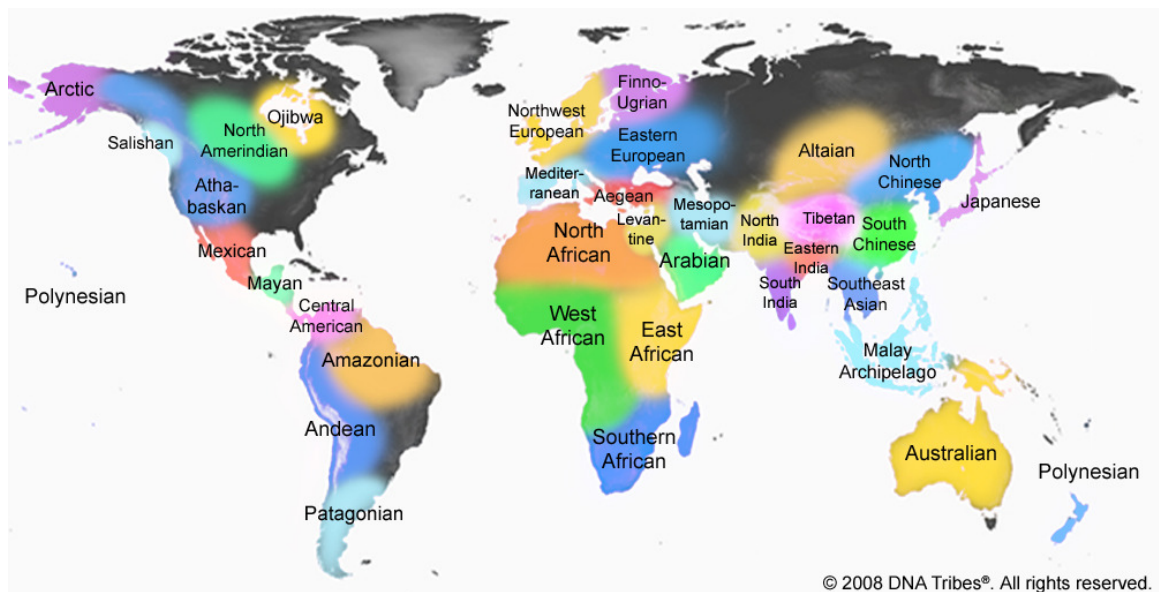
Each world region represents a unique genetic family within the human species shaped by shared history and geography. Each region is characterized by a distinctive pattern of allele frequencies across the STR loci studied. Although all humans are connected by ancient common origins, each of these genetic families shares a unique relationship due to more intense and persistent contacts within a geographical area. DNA Tribes® offers the first and only analysis able to distinguish 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 presumed racial or ethnic divisions, DNA Tribes® defines world regions by objective mathematical criteria. A proprietary statistical method is applied to over 800 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.

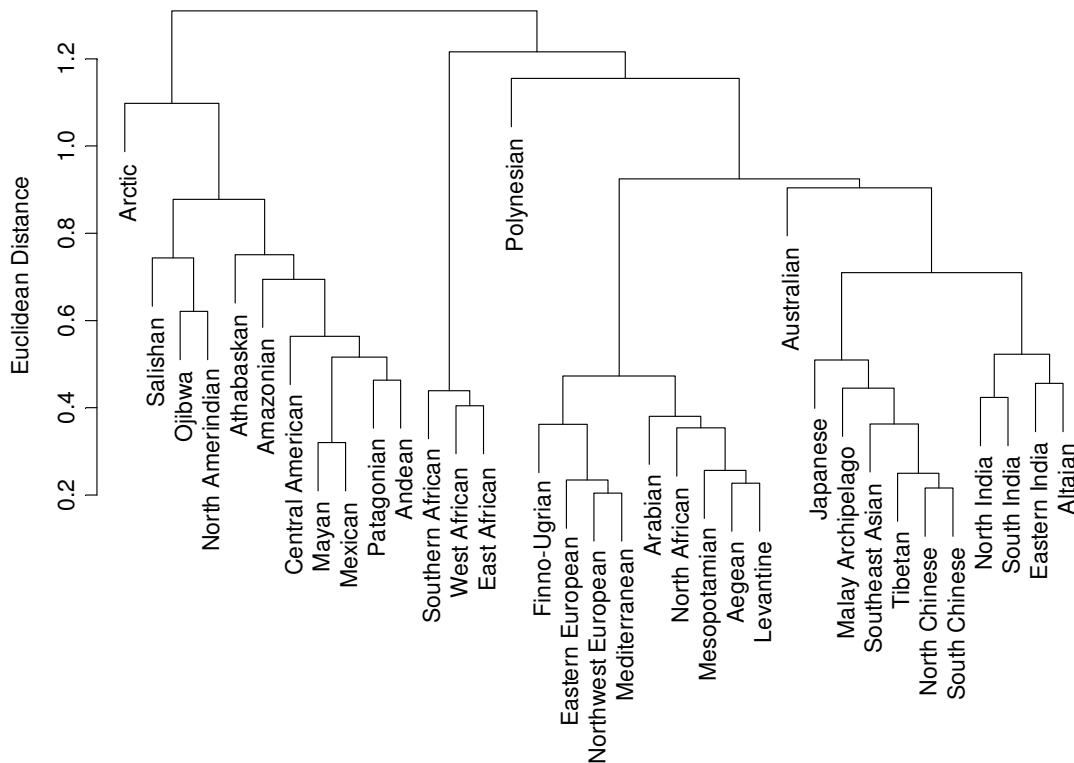
Map of DNA Tribes® World Regions

This map illustrates the geographical boundaries of these world regions. Within the borders of nearly all regions, individuals can be found with genetic ties to neighboring and sometimes distant regions.



How are World Regions Related?

Method: Hierarchical clustering was performed on world region clusters with the distance metric as the sum of absolute differences. In this plot, the distance between clusters is the average of the distances between the points in one cluster and the points in the other cluster.

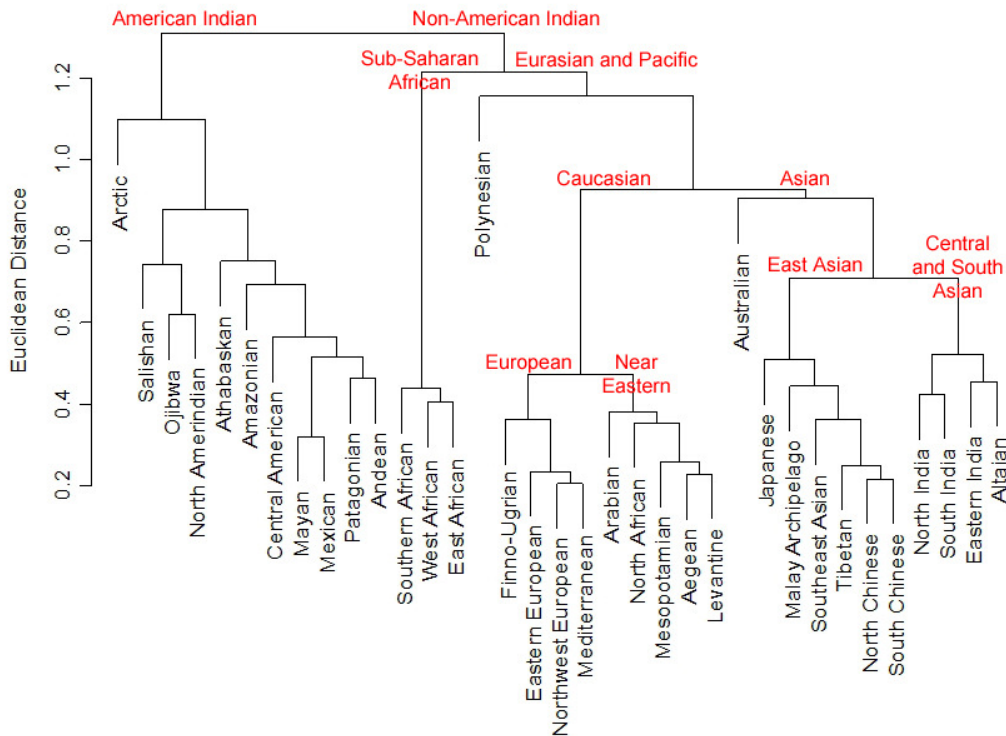


Discussion: This diagram illustrates the relationships between world regions identified by DNA Tribes. Closely related regions connect towards the bottom of the diagram. For instance, the Mediterranean and Northwest European regions are the two most closely related regions. The largest divisions appear at the top (root) of the tree. For instance, the Polynesian region is only distantly related to other world regions and branches off alone towards the top of the tree diagram.

Individual regions group together 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 Asian super-family that also includes Central and South Asian regions. Similarly, all American

DNA *what's your tribe?* TRIBES

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.



The relationships illustrated by this diagram 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

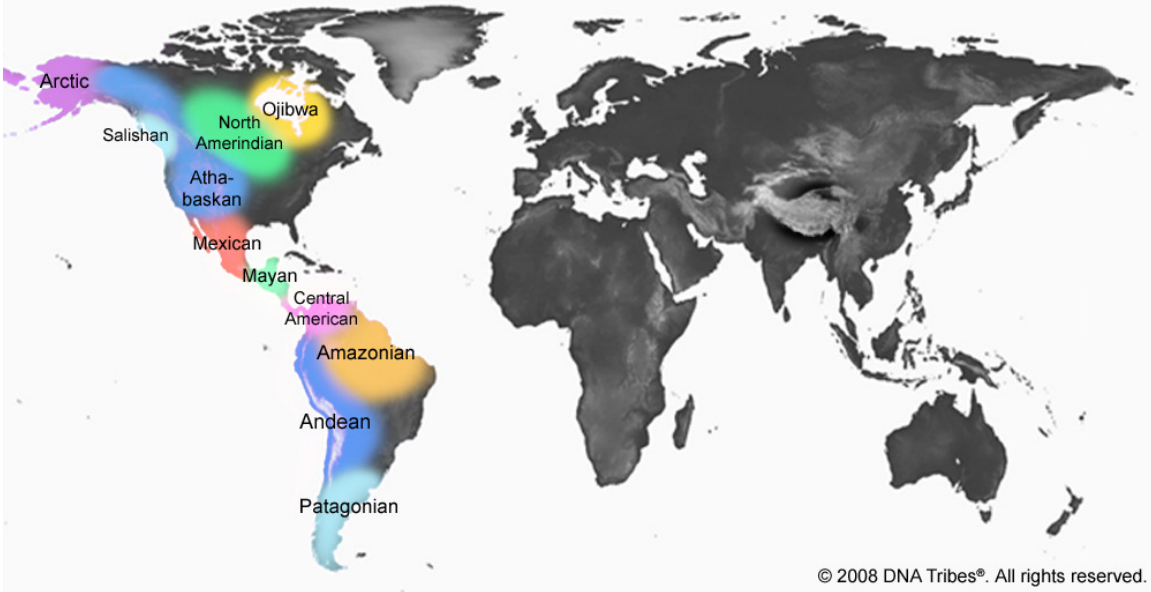
***Method:** Observed allele frequency data was used to simulate 2,000 individual genetic profiles for the studied world regions. Each simulated profile was processed by the DNA Tribes® algorithm, which measured the simulated individual's occurrence frequency in each world region. The strongest regional 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:

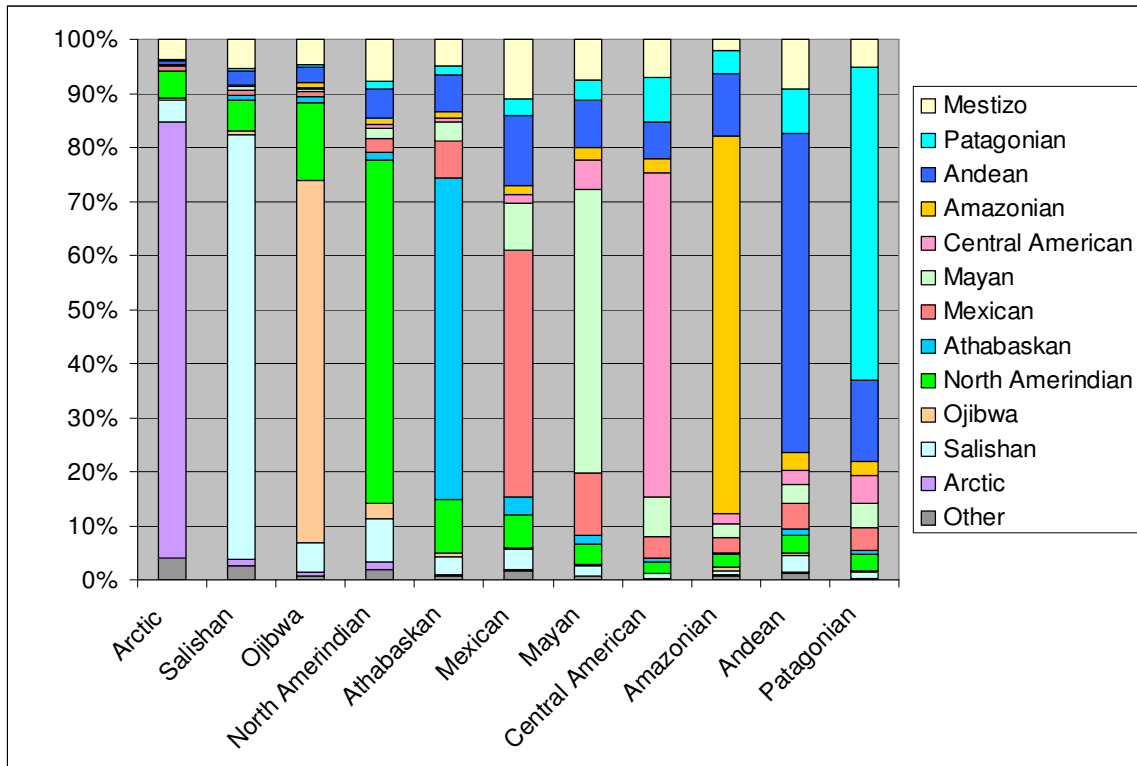
American Indian Regions.....	7
European and Near Eastern Regions.....	8
Sub-Saharan African and Central and South Asian Regions.....	9
East Asian and Pacific Regions.....	10

DNA *what's your tribe?* TRIBES

American Indian Regions



© 2008 DNA Tribes®. All rights reserved.



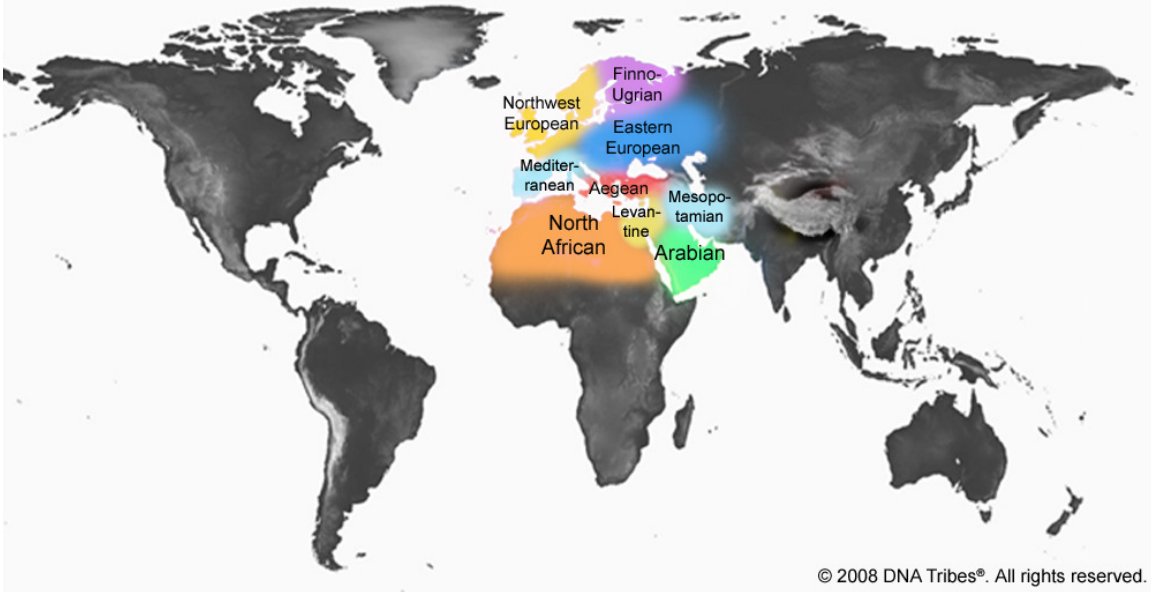
March 10, 2008

www.dnatribes.com

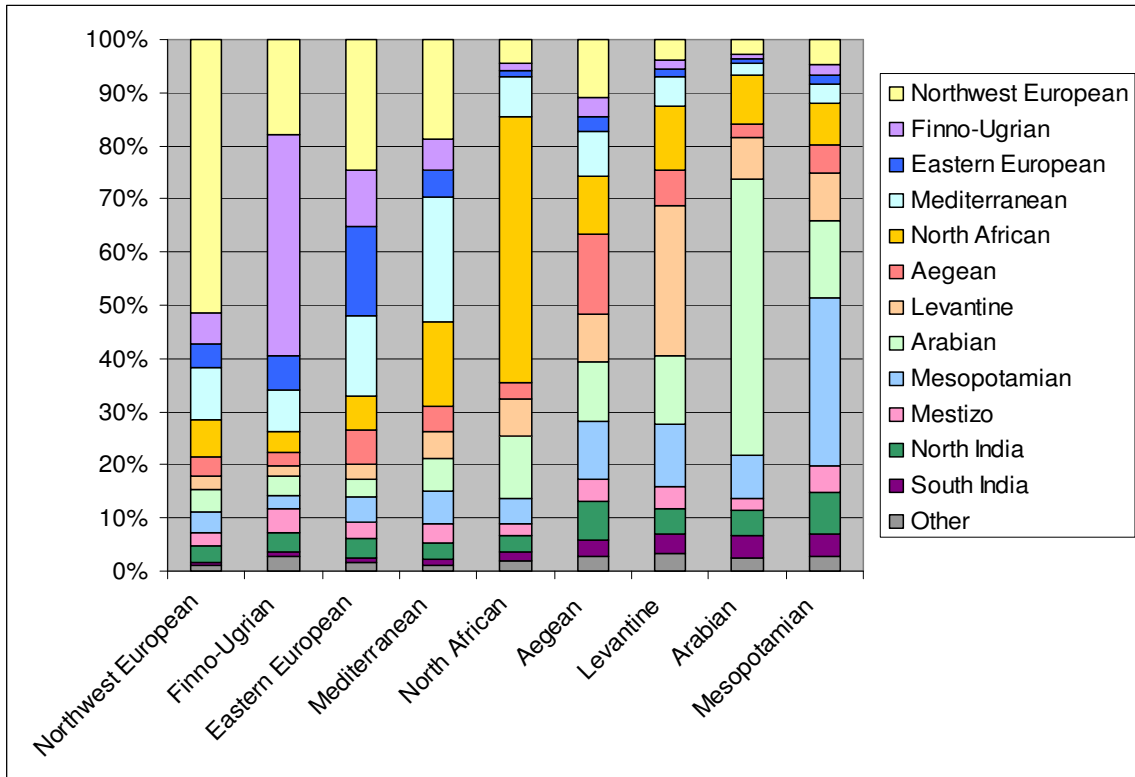
7

DNA Tribes *what's your tribe?*

European and Near Eastern Regions

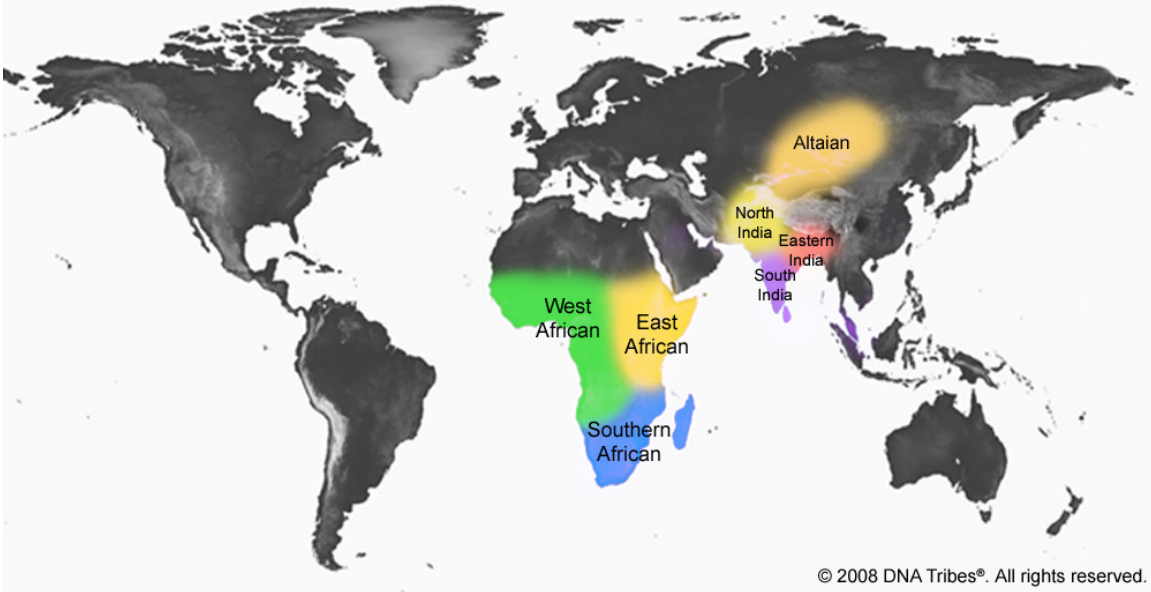


© 2008 DNA Tribes®. All rights reserved.

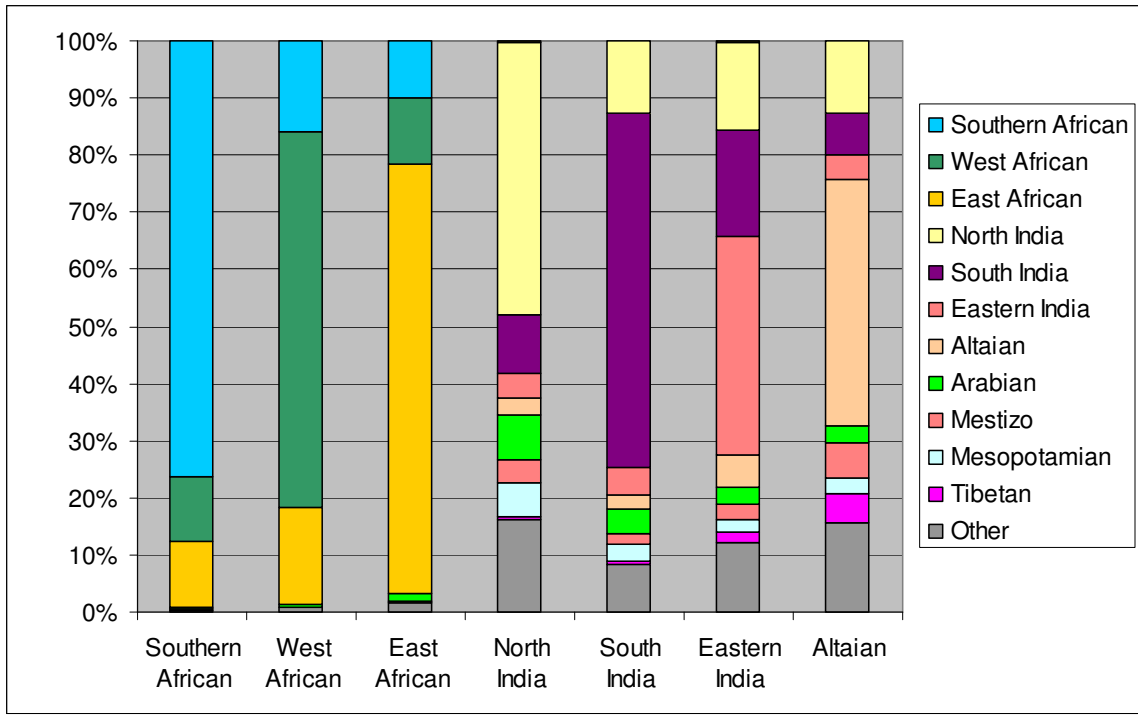


DNA Tribes *what's your tribe?*

Sub-Saharan African and Central and South Asian Regions

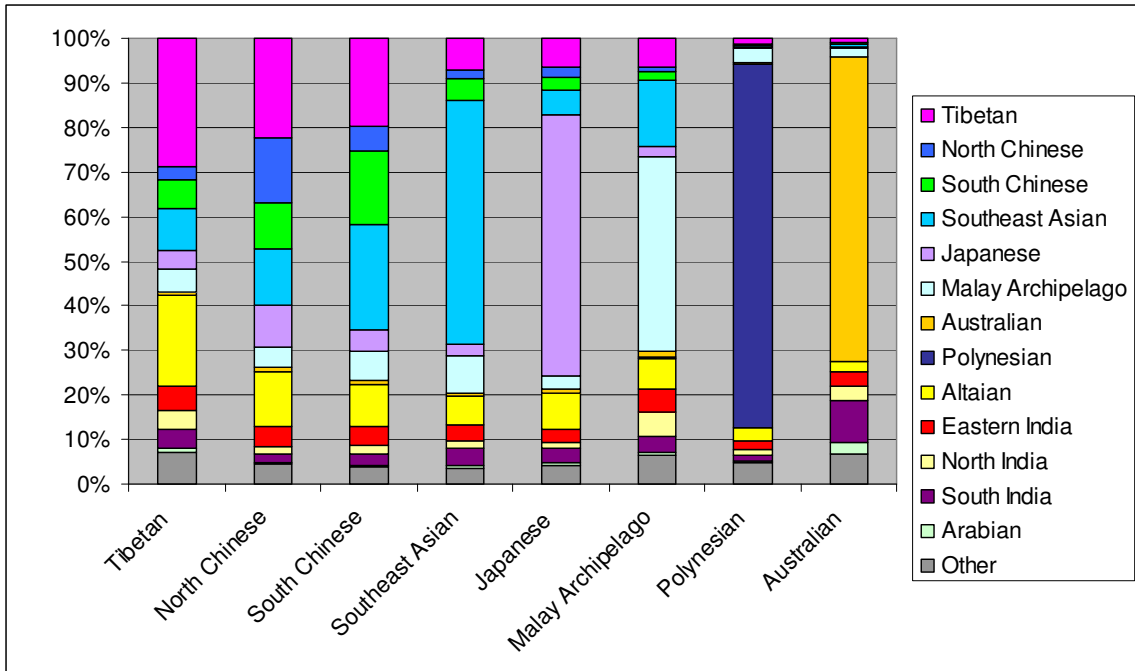
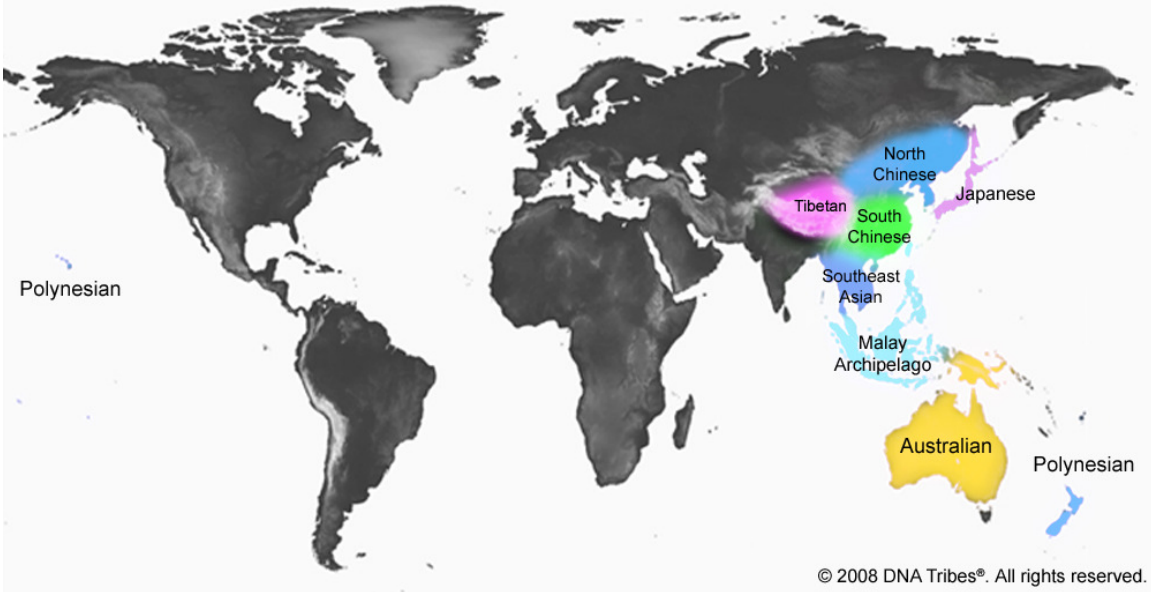


© 2008 DNA Tribes®. All rights reserved.



DNA *what's your tribe?* TRIBES

East Asian and Pacific Populations





DNA Tribes® Europa: A More Detailed Analysis of Europe

DNA Tribes® has also developed DNA Tribes® Europa, a more detailed comparison to genetic sub-regions of Europe. 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 on an objective mathematical basis, and reflect long-term geographical and/or ethnic relationships within Europe. Some of these genetic territories correspond to modern political boundaries, but 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.

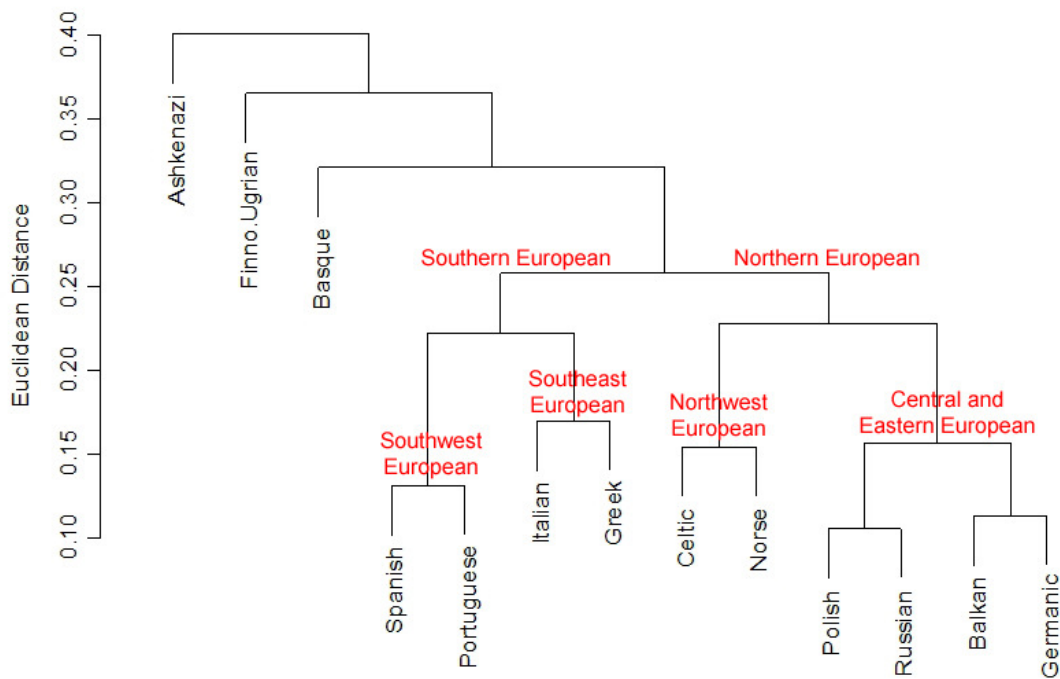




A Genetic Family Tree of European Sub-Regions

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

Method: Hierarchical clustering was performed on European sub-region clusters with the distance metric as the sum of absolute differences. In this plot, the distance between clusters is the average of the distances between the points in one cluster and the points in the other cluster.



Discussion: This diagram illustrates the relationships between European sub-regions identified by DNA Tribes®. Closely related sub-regions appear towards the bottom of the diagram. For instance, the Polish and Russian regions are the two most closely related regions. The largest divisions appear at the top (root) of the tree. For instance, the Ashkenazi, Finno-Ugrian, and Basque sub-regions are only distantly related to other European sub-regions and branch off alone towards the top of the tree diagram. These unique genetic patterns found within these outlier sub-regions reflect relative isolation from other sub-regions due to cultural endogamy and/or geographical isolation.

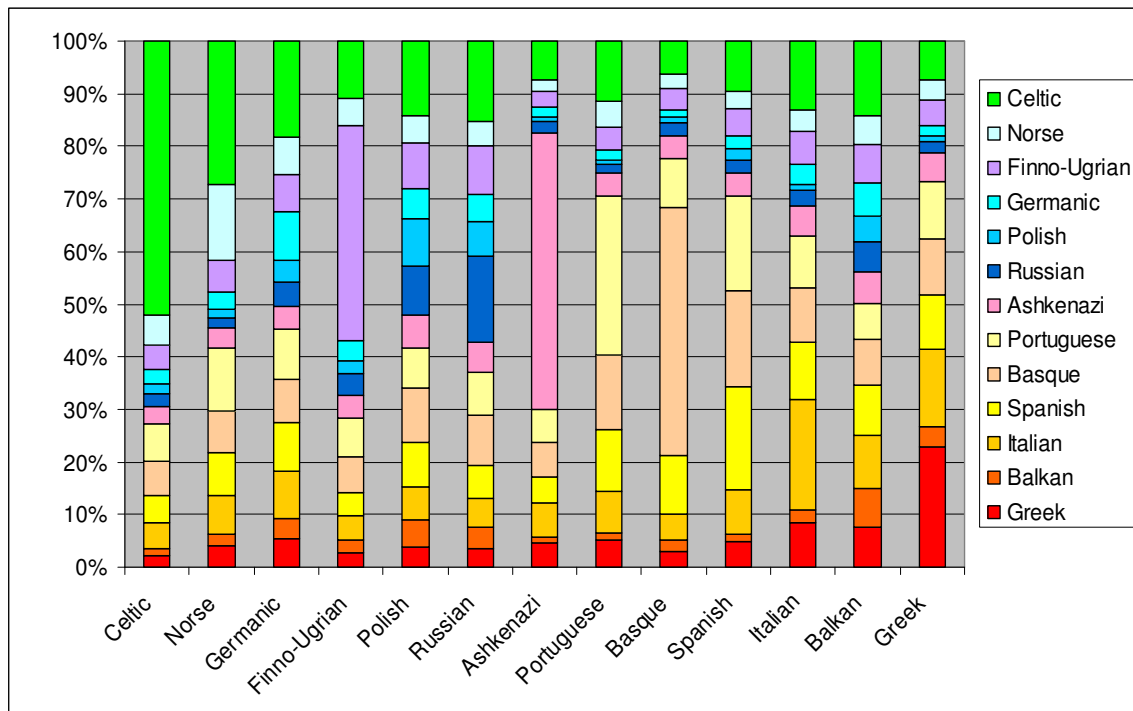


Individual Genetic Affiliations within European Sub-Regions

Method: Observed allele frequency data was used to simulate 2,000 individual genetic profiles for the studied Europa sub-regions. Each simulated profile was processed by the DNA Tribes® algorithm, which measured the simulated individual's occurrence frequency in each 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 Ashkenazi, Basque, or Celtic 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 Europa genetic 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 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)

	5%	25%	50%	75%	95%
	(very low)	(ordinary)	(ordinary)	(ordinary)	(very high)
Northwest European	1.90	26.31	182.19	1,423.14	25,825.32
Mediterranean	1.43	21.15	136.15	975.53	17,564.77
Eastern European	0.96	15.28	98.03	794.92	12,872.07
Aegean	1.93	17.26	84.68	468.46	5,950.63
Levantine	0.86	9.19	50.68	293.52	4,245.86
Finno-Ugrian	0.35	6.27	45.22	321.29	6,272.69
Mesopotamian	0.94	8.18	39.86	218.51	2,623.37
North African	0.32	4.83	35.58	280.34	5,664.06
Arabian	0.40	4.45	26.37	163.74	2,198.13
Mestizo	0.48	3.99	16.32	73.37	636.78
North India	0.57	3.51	12.45	45.16	283.76
South India	0.04	0.28	1.03	4.28	36.01
Altaian	0.02	0.18	0.66	2.23	11.47
Eastern India	0.02	0.14	0.48	1.52	8.55
East African	0.00	0.00	0.01	0.27	16.74
Tibetan	0.00	0.00	0.01	0.06	0.58
Malay Archipelago	0.00	0.00	0.01	0.03	0.29
Australian	0.00	0.00	0.00	0.01	0.16
West African	0.00	0.00	0.00	0.05	4.24
Southeast Asian	0.00	0.00	0.00	0.01	0.13
South Chinese	0.00	0.00	0.00	0.00	0.08
Southern African	0.00	0.00	0.00	0.01	1.58
Andean	0.00	0.00	0.00	0.00	0.18
Salishan	0.00	0.00	0.00	0.00	0.08
North Amerindian	0.00	0.00	0.00	0.00	0.08
Mexican	0.00	0.00	0.00	0.00	0.06
North Chinese	0.00	0.00	0.00	0.00	0.03
Arctic	0.00	0.00	0.00	0.00	0.03
Mayan	0.00	0.00	0.00	0.00	0.02
Japanese	0.00	0.00	0.00	0.00	0.02
Patagonian	0.00	0.00	0.00	0.00	0.03
Athabaskan	0.00	0.00	0.00	0.00	0.01
Central American	0.00	0.00	0.00	0.00	0.01
Polynesian	0.00	0.00	0.00	0.00	0.00
Ojibwa	0.00	0.00	0.00	0.00	0.00
Amazonian	0.00	0.00	0.00	0.00	0.00

For instance, half of all U.S. Caucasians obtain a Northwest European score between 26.31 and 1,423.14, with a median of 182.19. Northwest European scores within this range could be considered ordinary for Caucasian individuals. Only five percent of Caucasians obtain a score either below 1.90 or above 25,825.32. These could be considered very low and very high scores within this ethnic group, respectively.



Caucasian (United States) (Europa Sub-Regions)

	5.00%	25.00%	50.00%	75.00%	95.00%
	(very low)	(ordinary)	(ordinary)	(ordinary)	(very high)
Norse	1.79	25.02	192.90	1,350.58	22,996.43
Celtic	1.38	23.26	183.23	1,352.04	26,615.71
Germanic	1.44	22.02	170.02	1,187.72	18,656.73
Italian	1.72	23.61	169.77	1,096.99	13,686.63
Spanish	1.31	20.37	155.08	1,091.29	18,701.38
Portuguese	1.34	18.51	146.18	1,077.58	18,472.18
Balkan	1.19	17.88	131.88	974.70	13,581.44
Polish	0.78	12.15	94.24	711.43	11,324.42
Greek	0.97	13.35	82.90	545.47	6,717.89
Russian	0.64	10.19	77.88	572.29	10,102.97
Basque	0.28	6.38	63.90	583.43	11,913.49
Ashkenazi	0.36	6.20	53.14	403.26	7,495.74
Finno-Ugrian	0.46	6.75	47.97	341.42	6,421.41

For instance, half of all U.S. Caucasians obtain a Norse score between 25.02 and 1,350.58, with a median of 192.90. Norse scores within this range could be considered ordinary for Caucasian individuals. Only five percent of Caucasians obtain a score either below 1.79 or above 22,996.43. These could be considered very low and very high scores within this ethnic group, respectively.



African-American (United States)

	5%	25%	50%	75%	95%
	(very low)	(ordinary)	(ordinary)	(ordinary)	(very high)
West African	5.93	1,150.84	53,876.48	2,792,606.95	1,621,469,999.23
East African	9.30	1,237.82	45,452.08	1,613,934.30	450,063,793.37
Southern African	0.81	265.59	17,383.57	880,528.46	450,255,906.57
Arabian	1.28	17.54	108.66	713.26	13,696.12
North African	0.36	10.23	94.59	833.15	24,259.09
Levantine	0.24	3.85	24.14	146.13	2,000.81
Mestizo	0.28	2.79	12.43	57.32	592.13
Aegean	0.11	1.45	9.27	53.45	748.59
Northwest European	0.04	0.81	6.25	51.32	1,060.66
Mesopotamian	0.05	0.74	4.44	25.64	331.69
Mediterranean	0.03	0.61	4.36	34.92	712.25
North India	0.04	0.31	1.58	8.19	77.68
Eastern European	0.00	0.11	0.89	8.58	200.30
Finno-Ugrian	0.00	0.04	0.37	3.38	77.60
South India	0.01	0.06	0.35	1.99	18.25
Altaian	0.00	0.01	0.03	0.16	1.35
Eastern India	0.00	0.00	0.01	0.04	0.41
Tibetan	0.00	0.00	0.00	0.01	0.08
Australian	0.00	0.00	0.00	0.00	0.06
Malay Archipelago	0.00	0.00	0.00	0.00	0.02
Arctic	0.00	0.00	0.00	0.00	0.26
Salishan	0.00	0.00	0.00	0.00	0.16
Andean	0.00	0.00	0.00	0.00	0.05
North Amerindian	0.00	0.00	0.00	0.00	0.08
Mexican	0.00	0.00	0.00	0.00	0.04
Southeast Asian	0.00	0.00	0.00	0.00	0.00
South Chinese	0.00	0.00	0.00	0.00	0.00
Mayan	0.00	0.00	0.00	0.00	0.02
Athabaskan	0.00	0.00	0.00	0.00	0.01
Patagonian	0.00	0.00	0.00	0.00	0.01
Polynesian	0.00	0.00	0.00	0.00	0.00
Central American	0.00	0.00	0.00	0.00	0.00
North Chinese	0.00	0.00	0.00	0.00	0.00
Ojibwa	0.00	0.00	0.00	0.00	0.00
Japanese	0.00	0.00	0.00	0.00	0.00
Amazonian	0.00	0.00	0.00	0.00	0.00

For instance, half of all African-Americans obtain a West African score between 1,150.84 and 2,792,606.95, with a median of 53,876.48. 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 5.93 or above 1,621,469,999.23. These could be considered very low and very high scores within this ethnic group, respectively.



Hispanic (United States)

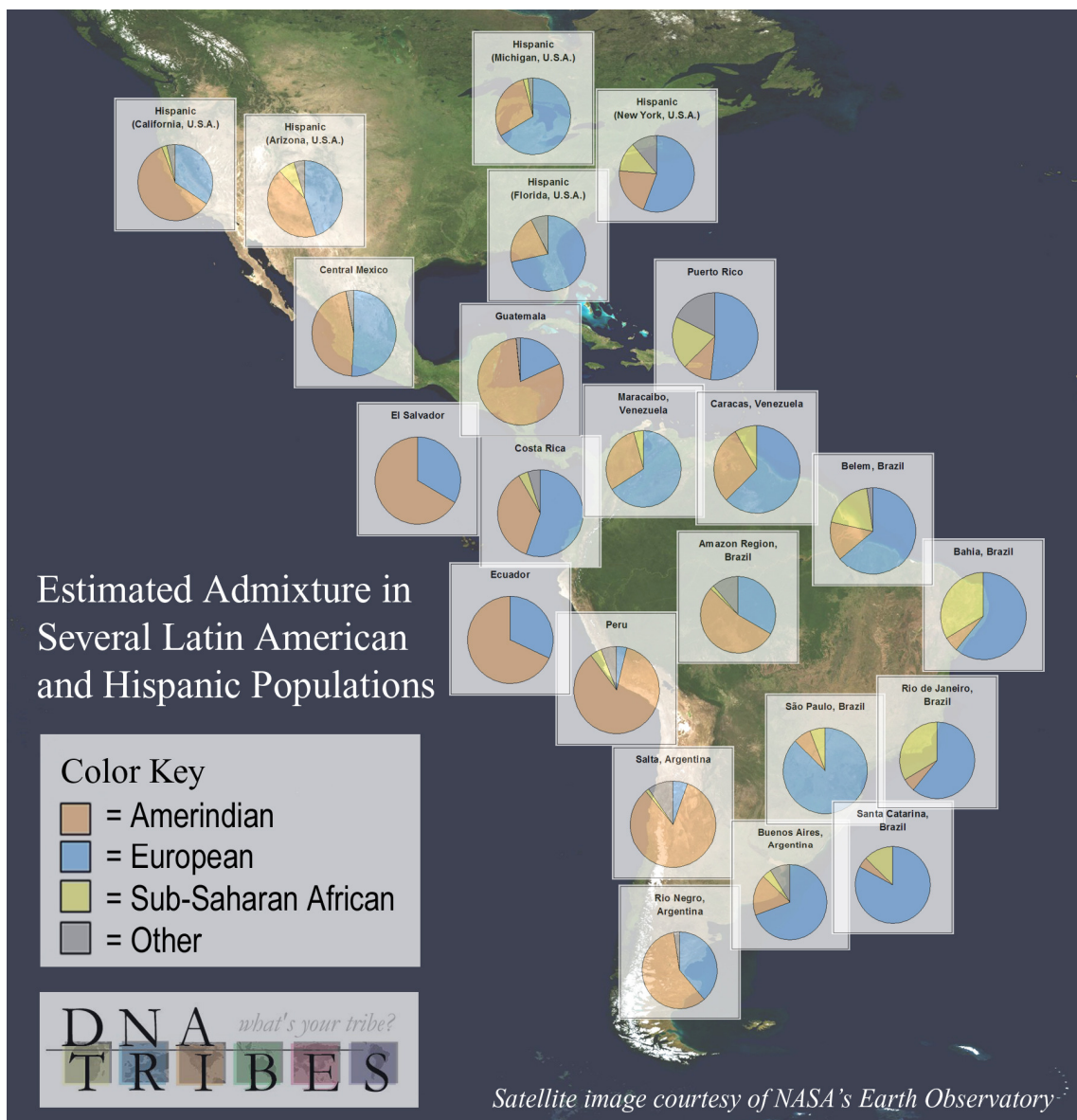
	5%	25%	50%	75%	95%
	(very low)	(ordinary)	(ordinary)	(ordinary)	(very high)
Mestizo	0.87	8.47	38.02	199.42	2,623.65
Aegean	0.14	1.25	6.36	34.55	504.01
Levantine	0.10	0.97	5.41	31.78	441.33
Mesopotamian	0.11	0.95	5.38	27.77	300.75
Northwest European	0.05	0.66	4.68	35.46	672.72
Arabian	0.07	0.72	4.09	23.60	339.28
Mediterranean	0.04	0.55	3.82	27.46	591.60
North India	0.13	0.86	2.84	11.27	72.23
Eastern European	0.02	0.33	2.41	17.60	332.84
Finno-Ugrian	0.02	0.25	1.94	14.74	272.57
North African	0.02	0.25	1.90	14.54	320.18
Altaian	0.02	0.17	0.67	2.35	16.60
South India	0.02	0.12	0.47	1.76	12.18
Eastern India	0.01	0.06	0.22	0.76	3.51
Andean	0.00	0.00	0.03	0.73	45.98
Tibetan	0.00	0.01	0.03	0.13	1.25
Mexican	0.00	0.00	0.01	0.23	17.83
North Amerindian	0.00	0.00	0.01	0.20	18.67
East African	0.00	0.00	0.01	0.16	10.16
Malay Archipelago	0.00	0.00	0.01	0.04	0.40
Mayan	0.00	0.00	0.00	0.10	10.52
Salishan	0.00	0.00	0.00	0.07	6.95
Patagonian	0.00	0.00	0.00	0.11	14.53
Australian	0.00	0.00	0.00	0.01	0.15
Southeast Asian	0.00	0.00	0.00	0.01	0.40
South Chinese	0.00	0.00	0.00	0.01	0.29
West African	0.00	0.00	0.00	0.03	2.36
Central American	0.00	0.00	0.00	0.03	3.78
Athabaskan	0.00	0.00	0.00	0.02	1.61
Arctic	0.00	0.00	0.00	0.01	0.85
Southern African	0.00	0.00	0.00	0.01	1.13
North Chinese	0.00	0.00	0.00	0.00	0.10
Japanese	0.00	0.00	0.00	0.00	0.07
Ojibwa	0.00	0.00	0.00	0.00	0.97
Amazonian	0.00	0.00	0.00	0.00	1.07
Polynesian	0.00	0.00	0.00	0.00	0.00

For instance, half of all U.S Hispanics obtain a Mestizo score between 8.47 and 199.42, with a median of 38.02. 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 0.87 or above 2,623.65. These could be considered very low and very high scores within this ethnic group, respectively.

DNA *what's your tribe?* TRIBES

Estimated Admixture in Latin American and Hispanic Populations

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. Scientists have also 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 over-simplified 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. At a theoretical level, this approach adds nothing to the popular or scientific understanding of human relationships and bestows an air to scientific legitimacy on outdated ideas of race. Technological applications of these over-simplifying models present potential for neglect or harm: for instance, to the Subcontinental Indian or Samoan administered a drug regimen based on a racial admixture model that does not take into account their world region of origins.

DNA *what's your tribe?* TRIBES

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 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 proceeding 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 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.

As DNA Tribes® continues to incorporate more data in a refined statistical analysis, this map will likely be clarified. However, a number of basic points have become apparent:

- American Indians, traditionally described as a homogeneous group or as a minor offshoot of a putative “Mongoloid race,” are instead a diverse family of autonomous world regions. The genetic gap between American Indian and all non-American Indian populations (geographically corresponding to the Bering Sea) represents the most significant division in world populations.
- Mongolians proper are affiliated with an Altaian world region that bears only a distant relation to East Asians.
- Intermediate regions within Eurasia are not equivalent to hypothetical admixtures between far Western Europeans and far Eastern Asians. Anatolians, Mongolians, North Indians and others possess unique genetic characteristics not explained by a simple racial model.
- South Asia is the home of at least several unique world regions not consistent with a simple model of East-West contact. Each of the North India, South India, and

DNA *what's your tribe?* TRIBES

Eastern India regions is characterized by distinct allele frequencies.

- The cultural term “Hispanic” as used in the United States describes genetically disparate populations descended from American Indian, European and African ancestors, most of which share substantial affiliation with a Mestizo world region.
- Many diaspora ethnic groups retain traces of their origins as well as their current homes. For instance, Polish Tatars are descended from Central Asians and retain Mongolian genetic affiliations as well as affiliations with European populations.
- The Australian and Polynesian peoples are genetic outliers within the broader family or Eurasian regions, with no close relatives outside of their territorial homelands.