

[see full issue: May-June 2000](#)

Mitochondrial DNA and the Peopling of the New World

Genetic variations among Native Americans provide further clues to who first populated the Americas and when they arrived

[Theodore G. Schurr](#)

On the eve of Christopher Columbus's arrival on San Salvador (now part of the Bahamas) in 1492, there were perhaps several tens of millions of people inhabiting the Americas. Once it became evident that the inhabitants of this New World were not, in fact, East Indians (as Columbus had at first supposed), the existence of the Native American population became a huge puzzle to the Renaissance Europeans. Just who were these people across the ocean, and where did they come from? Various theories were proposed in the centuries that followed, including the notion that the Native Americans (now often called Amerindians) were the descendants of the "lost tribes of Israel." Some scholars even attempted to draw parallels between the Amerindians and the contemporary European Jews of the era. It was not until the 18th century that scholars hit on the notion, now well established, that the Amerindians originated on the Asian continent. (Recent claims—including the putative "Caucasian" characteristics of the Kennewick skeleton—that European stock may have been present in pre-Columbian America do not deny the overwhelming contribution of Asiatic peoples to the ancestry of modern Amerindians.)

Anthropologists have been struggling with the question ever since, attempting to explain the diversity of cultures, languages and biological traits among Native Americans. Were there several waves of migrations from Asia, or merely one large movement? Did the ancestral populations reach the Americas only 12,000 years ago (as traditionally believed) or much earlier?

[click for full image and caption](#)



[click for full image and caption](#)



Many recent studies have been guided by a tripartite migration model, which envisions three waves of migration corresponding to a three-part division of Native American languages: Amerind, Na-Dene and Eskaleut.

However, not all scientists agree that these subgroups represent an accurate partition of the Amerindian languages. Furthermore, the dental and nuclear-DNA studies that were used to support a tripartite migration are not unequivocal.

The other major question concerning Native American origins asks when people first might have arrived in the Americas. Archeological excavations in North America date the oldest human artifacts to between 11,000 and 14,000 years ago, whereas South American artifacts appear to be at least contemporaneous, if not older. Linguistic analyses also suggest entry dates ranging between 12,000 to 35,000 years ago. Studies of nuclear DNA genes further suggest an arrival time of 30,000 years ago, whereas research on dental variation indicates an emergence of the ancestral Amerindian groups about 18,000 to 20,000 years ago. There is indeed little agreement and much confusion.

Over the past 10 years, however, new methodologies have provided a number of important insights into the peopling of the New World. Molecular genetic studies of the variation of mitochondrial DNA (mtDNA) in Siberian and Amerindian populations have allowed further inferences to be made about the timing of the colonizations, the number of migrations that reached the New World, and possible regions from which ancestral Native Americans might have originated. Most notably, the new mtDNA data suggest not only a very early movement of peoples into the New World but also the genetic contributions of populations originating outside of Siberia, from other parts of Asia. Overall, the mtDNA research implies that the colonization of Siberia and the Americas was more complex than previously supposed—that there were, in fact, multiple expansions of ancient peoples that contributed to the genetic diversity in aboriginal Siberian and Amerindian populations.

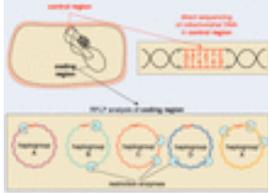
The Beauty of mtDNA Analysis

Mitochondrial DNA is a gift to the molecular anthropologist seeking to sort out the genetic relations between people. Since mtDNA is maternally inherited, the analysis of the mitochondrial genome is effectively the study of female genetic history within human populations. Unlike nuclear DNA, mtDNA does not appear to undergo a recombination of its nucleotide bases during division and reproduction of the mitochondria. The lack of recombination permits mutations to accumulate in a more or less "linear" or chronological fashion within maternal lineages. As a result, there is minimal ambiguity in reconstructing the branching of female lineages along the accumulated changes in their mtDNA.

Mitochondrial DNA can also provide a record of human migration. The presence of identical mutations in mtDNA from geographically separated human groups is a good indication that there might have been migrations or contacts between the groups. The mtDNA also records stochastic processes that can alter the genetic composition of a population. Processes such as *genetic drift*—the natural accumulation of unique genetic changes in a population that is isolated from other populations (and so cannot share its genes)—and the *founder effect*—in

which a newfound population of migrants contains a purely happenstance sample of genetic variants that are not representative of the parent population—were common events in human history that are recorded in the mtDNA of living peoples in terms of the frequencies and types of mtDNAs present in them.

[click for full image and caption](#)



The mtDNA has two major regions that are the focus of genetic studies of Native American origins (*Figure 2*). About 94 percent of the mtDNA sequence consists of coding regions, which include genes for ribosomal RNAs (rRNAs), transfer RNAs (tRNAs) and proteins involved in oxidative phosphorylation (the major biochemical process within mitochondria). The remainder of the mtDNA genome (about 1,100 nucleotide base pairs) consists of the non-coding *control region*, which initiates and regulates the replication of mtDNA. The control region mutates about two- to ten times faster than the coding region; two *hypervariable segments* (HVS-I and HVS-II) are the most rapidly evolving portions of it.

Two different molecular methods have been employed to take advantage of mtDNA's genetic features. Restriction fragment length polymorphism (RFLP) analysis surveys individual mtDNAs for sequence variation using a series of DNA-cutting enzymes (restriction endonucleases) that cleave the mtDNA molecule at specific nucleotide sequences known as recognition sites. If there are genetic differences between two individuals such that restriction enzymes cut their DNA at different points in the mtDNA, the resulting fragments will be of different lengths. In RFLP analysis, the unique combination of fragments detected by a set of restriction enzymes represent genetic polymorphisms within a mtDNA *haplotype*. In turn, each grouping of related haplotypes that is defined by a specific set of shared RFLPs is called a *haplogroup* or *mtDNA lineage*.

Considerable genetic variation within and between human populations has been detected with RFLP analysis. Although a significant portion of this variation is shared among all human populations, a certain amount is only found within geographically circumscribed populations or ethnic groups.

The other method of studying mtDNA variation involves the direct sequencing of one or both of the hypervariable segments (HVS-I and HVS-II) of the control region. In contrast to RFLP analysis, which can be likened to a general scan of the mtDNA genome, direct sequencing provides a nucleotide-by-nucleotide reading of a portion of the mtDNA. Because the mutation rate is high in the control region, direct sequencing provides a detailed look at small genetic changes that may have taken place quite recently. Mutations in the control region help to define specific mtDNA lineages in human populations (as do certain RFLPs), and they also reveal the genetic differentiation of these lineages in geographically circumscribed areas. By statistically assessing the amount of variation in the control regions within and between mtDNA lineages, it is possible to estimate the relative age of genetic variants in

a particular geographic region. This process can also be carried out with haplotypes defined by RFLP analysis.

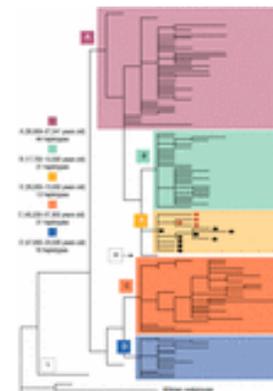
Because the mtDNA control region mutates rapidly, a single RFLP haplotype may be associated with several control region sequences. Such variations in the control region serve to distinguish subtypes of haplotypes (and may indicate a relatively recent divergence). In addition, the same control region sequence may be observed among different RFLP haplotypes, indicating the evolutionary convergence of control region sequences, the recent accumulation of new RFLPs or the loss of RFLP variation. Consequently, control region sequencing and RFLP haplotyping may not describe the same genotypes. However, in practice this does not necessarily pose a problem because each haplogroup usually has unique RFLPs and control region sequences that distinguish them from other mtDNA lineages. The identity of the haplogroups is therefore usually fairly robust, and the relative level of diversity within each mtDNA lineage is roughly equivalent.

MtDNA in the Founding Americans

What does the mtDNA genome tell us about the diversity of the original migrations into the Americas? Many studies have now established that the vast majority of modern Native American haplotypes belong to merely four mtDNA lineages, the haplogroups designated A, B, C and D. When subjected to the methods of phylogenetic analysis, the RFLP haplotypes usually segregate into their respective haplogroups, revealing their integrity as genealogical units (*Figure 3*). Moreover, ancient Amerindian samples obtained from different locations in the New World reveal the same general pattern of mtDNA diversity. Sequencing studies of the HVS-I region also provide an independent confirmation of four primary mtDNA haplogroups in Native Americans.

A comparison of Native Americans, Siberians and Asians reveals that the same mtDNA lineages in all groups share mutations in the control region that are specific to the haplogroups. The simplest explanation is that the control region mutations arose in Asia in the founding mtDNA lineages and were carried to the New World by the ancestral Native Americans. Most of the remaining variants in control region sequences of Native Americans appear to be unique (much like their RFLPs). This suggests that Native American and Asian groups diverged rapidly after the founding New World populations separated from their Asian parental populations.

[click for full image and caption](#)



Although most Native American mtDNAs fall into four distinct haplogroups, there is less of a consensus about the number of founding haplotypes that accompanied each haplogroup to the New World. To be considered a founding haplotype, candidate mtDNAs have to meet several criteria. First, founding haplotypes should be widespread within Amerindians and shared between tribes, since they would have

preceded tribal differentiation. Second, founding haplotypes should be central to the branching of their haplogroup in phylogenetic analyses because all new haplotypes would have originated from them. Third, founding haplotypes should be present in East Asian and Siberian populations because they originated in those regions. Conversely, mtDNA haplotypes that are derived from the founding haplotypes should have a limited distribution.

[click for full image and caption](#)



When these criteria are applied to Native American mtDNA, only four to seven RFLP haplotypes clearly meet the standards, and only one (or possibly two) control region sequences are likely to have been in the founding migrants. Thus, both sets of data suggest that a limited number of founding mtDNAs were brought to the New World by the founding populations. Another implication is that either some degree of reduction in genetic diversity took place during the colonization of the New World, or a geographically specific subset of Asian mtDNAs was brought to the Americas.

The geographic and linguistic distribution of haplogroups A-to-D in the Americas also suggests that all four of them were present in the original migration(s). All four haplogroups are observed in populations throughout the Americas and are also found in the three proposed Native American linguistic groups (Amerind, Na-Dene, Eskaleut). However, the original Na-Dene Indians and Eskimo-Aleuts appear to have lacked haplogroup B.

Although mtDNAs from haplogroups A-to-D are often found together in a single population, many tribes lack haplotypes from at least one of these haplogroups. This trend reflects the fact that genetic drift and founder events have played a significant role in shaping the distribution of mtDNA haplotypes in these populations. Such an interpretation is also supported by the high frequency of *private* mtDNA haplotypes in different Amerindian tribes. These results support the idea that early tribal isolation and founder effects led to the divergence of tribal gene pools in different regions (although not all scientists agree on this point).

It's also important to mention that the genetic composition of an ancient population may not be the same as the population currently occupying the same geographic region, because of migrations, genetic drift or other stochastic processes. For example, based on haplogroup frequencies, the ancient Stillwater Marsh population in the Great Basin region does not appear to be ancestral to any modern Amerindian population in the same area. On the other hand, ancient Eskimo and Aleut samples have nearly the same haplogroup frequencies as their modern descendents, and the same is true for the ancient Anasazi and Fremont cultures and the modern Pueblo Indian groups. Such results imply that once they became genetically distinct from surrounding groups, many regional Amerindian populations maintained their genetic

integrity over a considerable period of time.

When Did They Arrive?

The antiquity of the founding Native American haplogroups can also provide a temporal yardstick by which to measure human occupancy in the New World. Assuming a particular rate of mutational change for mtDNA, it's possible to estimate the number of years that must have passed since two haplogroups or haplotypes diverged. Based on analyses of RFLP haplotypes and control region sequences, haplogroups A, C and D appear to be the oldest mtDNA lineages in the New World (averaging about 47,650 to 23,535 years old). The ages of these haplogroups in Siberia are generally comparable to those for the Americas. The considerable age of these haplogroups suggests that the genetic links between Siberian and Native American populations are indeed quite ancient.

The RFLP data further suggest that haplogroup B in the Americas is considerably younger (17,700 to 13,500 years old) than American haplogroups A, C and D. However, recent analyses of variation in control region sequences in Native Americans have indicated that haplogroup B may be as diverse (and therefore as old) as haplogroups A, C and D. In addition, previous work suggested that the haplotypes of haplogroup B were present in Central-East Asia at least 24,000 to 30,000 years ago. Overall, it appears that the four primary haplogroups in Native Americans were brought to the New World well before the last glacial maximum (about 18,000 years ago).

[click for full image and caption](#)



The age of haplogroup A in Siberia is considerably younger than that in the Americas (averaging only 12,727 to 9,655 years old). This discrepancy arises because the Siberian estimate is based on RFLP data from only two populations, the Chukchi and Siberian Eskimos, whereas the Native American estimate is based on data from 15 to 20 populations. On the other hand, a relatively recent age for haplogroup A has also been estimated for North American Eskimos and Na-Dene Indians based on control region sequences. Thus, it appears that the ancient Beringian populations that gave rise to the Chukchi, the Eskimo-Aleuts and the Na-Dene Indians underwent a recent bottleneck in which variation was reduced (perhaps 13,000 to 7,000 years ago), followed by an expansion of haplogroup A in the Arctic and Subarctic regions of North America.

"Other" MtDNA Lineages

A number of mtDNAs found in Native Americans do not fall into the haplogroups A-to-D. These were originally designated as "other" (OTHER) haplotypes, and the majority were attributed to non-native admixture because of their apparent affinities to European mtDNAs. In particular, the OTHER haplotypes detected in the Ojibwa and Navajo resembled the haplogroup X mtDNAs seen in French Canadians and other European

groups. A single haplotype in the Maya also appeared to belong to the European haplogroup H, the most commonly observed mtDNA lineage in Caucasian American and European populations. These findings suggested that most of the OTHER haplotypes seen in Native Americans were likely to be of European origin and, hence, of more recent derivation in New World populations.

[click for full image and caption](#)



However, recent work has shown that the OTHER haplotypes that looked similar to the European haplogroup X mtDNAs actually belong to a divergent branch of this particular mtDNA lineage. All the Amerindian haplogroup X haplotypes share a common core set of RFLP and control region sequence mutations with European haplogroup X mtDNAs, but otherwise differ from them by at least several control region mutations. Also, four distinct sublineages of haplogroup X have been identified in Amerindian populations, implying that it has been in the New World long enough to have undergone considerable genetic diversification.

In contrast with the distribution of haplogroups A to D, haplogroup X is found nearly exclusively in North American populations. Interestingly, it occurs at its highest frequencies among Algonkian-speaking groups such as the Ojibwa, with all four of the Amerindian sublineages being present with this population. This mtDNA lineage has also been detected in two pre-Columbian North American populations and may be present in a few ancient Brazilian samples. These data imply that haplogroup X was present in the New World long before Europeans first arrived in the New World. Indeed, in Amerindians haplogroup X ranges between 35,000 to 13,000 years old, depending on the number of founding haplotypes assumed to have been present in the ancestral population and whether RFLP or control region sequence data are used to make the estimates. Haplogroup X is now considered to be a fifth but minor founding mtDNA lineage in Native American populations.

Non-Native Admixture

Aside from haplogroup X, additional OTHER mtDNAs have been detected in Amerindian tribes, all of which appear to have different genetic affinities. Because many Native populations exhibit European admixture in nuclear genetic studies, it seems likely that some of their OTHER mtDNAs were acquired through gene flow with modern Europeans. Indeed, this does appear to be the case, as European haplogroup H and T mtDNAs are found in the Ojibwa, and European haplogroup H and J mtDNAs are seen in the Cherokee. It is not yet clear whether the OTHER haplotypes appearing in South American populations such as the Argentinean Mapuche were acquired through non-native admixture (because of limited analyses).

There has also been a non-negligible amount of African gene flow into certain Amerindian populations, because African slaves and their descendents intermixed with Native Americans in historical times. Evidence of African-Amerindian admixture has been detected in the

Seminoles of Florida and the Narragansett of Massachusetts, in the form of African haplogroup L mtDNAs. Similar observations have been made for the Mixatec and Zapotec Indians of Southern Mexico. Judging from these data, African mtDNA haplotypes may be present among other Native American groups.

The remaining OTHER mtDNAs exhibit mutational features similar to those seen in the majority of Asian haplotypes. The great majority of these "Asian" OTHER haplotypes have been observed in South American tribes from the Brazilian Amazon, such as the Makiritare and Yanomami, but a few have also been detected at very low frequencies in a handful of North American groups. These haplotypes possess two mutations that appear in all haplogroup C and D mtDNAs, but they lack other characteristic markers of these two mtDNA lineages. Because 50 to 75 percent of all Asian mtDNAs also have these mutations, these particular OTHER mtDNAs were thought to belong to a previously undetected Asian haplogroup that had been brought to the Americas by ancestral populations. However, when subjected to phylogenetic analysis, the control region sequences of these OTHER mtDNAs are interspersed with those from Amerindian haplogroups C and D, because they share specific nucleotide substitutions that define these haplogroups. These results suggest that OTHER mtDNAs having the two mutations are indigenous haplotypes that arose through reversion mutations after the entry and spread of haplogroups C and D in the Americas.

Where Did They Come From?

There has been considerable speculation about the areas from which ancestral Paleoindians emerged and expanded into the Americas. Recent studies have suggested that northern China, southeastern Siberia or Mongolia may have been sources because of haplogroup D being present in those regions. The highest frequencies of these four haplogroups occur in the Altai Mountain/Tuva/Lake Baikal region, implying that this general region gave rise to the founders of Native American populations. Otherwise, haplogroup B is absent in the vast majority of native Siberian populations, haplogroup A occurs at very low frequencies outside of Chukotka, and haplogroups C and D are the predominant mtDNA lineages in northern Asia.

However, the presence of a certain control region mutation in haplogroups C and D may point to alternative source areas for ancestral Native Americans. This mutation appears in the majority of both haplogroup C and D mtDNAs in Native American populations, suggesting it is part of the original sequence motifs for both of them. Among all Asian and Siberian mtDNAs, however, this mutation only appears in haplogroup C mtDNAs from Mongolia and the Amur River region and in haplogroup D mtDNAs in the Japanese, Korean and Ainu. This distribution suggests that East Asia as well as southeast Siberia or Mongolia might be source areas (or migration pathways) for these two haplogroups.

The origins of haplogroup X mtDNAs remain somewhat ambiguous. Judging from its age, haplogroup X could have arrived in the New World

either before or after the last glacial maximum (about 18,000 years ago). Regardless of when it was brought to the Americas, the lack of haplogroup X mtDNAs in Asian and Siberian groups (and their presence in certain European populations) suggests that these haplotypes originated outside of eastern Siberia, perhaps taken through Beringia by an ancient Eurasian migratory event distinct from the migration(s) that brought the other four mtDNA lineages to the Americas.

Bibliography

- Bonatto, S. L., and F. M. Salzano. 1997. Diversity and age of the four major mtDNA haplogroups, and their implications for the peopling of the New World. *American Journal of Human Genetics* 61:1413–1423. [[CrossRef](#)]
- Brown, M. D., S. H. Hosseini, A. Torroni, J. Allen, T. G. Schurr, R. Scozzari, F. Curciani and D. C. Wallace. 1998. Haplogroup X: An ancient link between Europe/Western Asia and North America? *American Journal of Human Genetics* 63:1852–1861. [[CrossRef](#)]
- Kolman, C. J., N. Sambuughin and E. Bermingham. 1996. Mitochondrial DNA analysis of Mongolian populations and implications for the origins of New World founders. *Genetics* 142:1321–1334.
- Lorenz, J. G., and D. G. Smith. 1997. Distribution of sequence variation in the mtDNA control region of Native North Americans. *Human Biology* 69:749–776.
- Merriwether, D. A., F. Rothhammer and R. E. Ferrell. 1994. Genetic variation in the New World: Ancient teeth, bone, and tissue as sources of DNA. *Experientia* 50:592–601.
- Nichols, J. 1994. The spread of language around the Pacific rim. *Evolutionary Anthropology* 1:206–215.
- Schurr, T. G., R. I. Sukernik, E. B. Starikovskaya and D. C. Wallace. 1999. Mitochondrial DNA diversity in Koryaks: Ancient and recent population expansions and dispersals in Okhotsk-Bering Sea region. *American Journal of Physical Anthropology* 108:1–40.
- Schurr, T. G., S. W. Ballinger, Y. Y. Gan, J. A. Hodge, D. A. Merriwether, D. N. Lawrence, W. C. Knowler, K. M. Weiss and D. C. Wallace. 1990. Amerindian mitochondrial DNAs have rare Asian variants at high frequencies, suggesting they derived from four primary maternal lineages. *American Journal of Human Genetics* 46:613–623.
- Starikovskaya, Y. B., R. I. Sukernik, T. G. Schurr and D. C. Wallace. 1998. Mitochondrial DNA diversity in Chukchi and Siberian Eskimos: Implications for the genetic prehistory of ancient Beringia. *American Journal of Human Genetics* 63:1473–1491. [[CrossRef](#)]
- Szathmary, E. J. E. 1993. Genetics of aboriginal North Americans. *Evolutionary Anthropology* 1:202–220.
- Torroni, A., J. V. Neel, R. Barrantes, T. G. Schurr and D. C. Wallace. 1994. A mitochondrial DNA "clock" for the Amerinds and its implications for timing their entry into North America.

Proceedings of the National Academy of Sciences, USA.
91:1158–1162.

Article Tools

 [printer
friendly](#)

 [request classroom
permission](#)

 [e-mail this
article](#)

Of Possible Interest

book review: [An Education in Irrationality](#)

feature article: [The Molecular Anatomy of an Ancient Adaptive Event](#)

book review: [Limits of the Genetic Lexicon](#)

book review: [Tracking Genes Across the Globe](#)

book review: [Making Hay with Straw Men](#)