
Genetic Evidence for the Phylogenetic Relationship between Na-Dene and Yeniseian Speakers

ROHINA RUBICZ, KRISTIN L. MELVIN, AND MICHAEL H. CRAWFORD¹

Abstract Ruhlen's hypothesis, based on linguistic evidence, for a common phylogenetic origin of Na-Dene and Yeniseian speakers is tested using genetic data. Gene frequency data for the Kets, the only surviving Yeniseian speakers, were collected during a field study in 1993. Data for several Na-Dene groups, as well as other New World and Siberian populations, were compiled from the literature. These data were analyzed using R-matrix, principal components analysis, and Mantel tests. In a comparison of 10 New World and Siberian populations using eight alleles, 55.8% of the variation was accounted for by the first principal component, and 22.1% of the variation was subsumed by the second principal component. Contrary to Ruhlen's interpretation of the linguistic data, analysis of the genetic data shows that the Na-Dene cluster with other Native American populations, while the Kets genetically resemble the surrounding Siberian groups. This conclusion is further supported by correlations that are higher when the Kets are considered unrelated to Na-Dene speakers, and an insignificant partial correlation between genes and language when geography is held constant, indicating that spatial patterning accounts for most of the variation present in these populations.

Considerable debate has occurred over the validity of treating linguistic and genetic data as analogous measures of human relationships. Some linguists argue that these data are not comparable (Bateman et al. 1990). While genes are transmitted in one direction only, from parent to offspring (vertical transmission), languages can be acquired not only vertically but also horizontally, with little or no exchange of genes (Chen et al. 1995; Cavalli-Sforza and Feldman 1981). However, anthropological geneticists argue that genetic and linguistic data reflect similar patterns of differentiation due to varying degrees of geographic isolation. A number of studies have demonstrated correlations between genetic and linguistic data, based on arguments that these are affected by similar evolutionary processes (Barbujani and Pilastro 1993; Barbujani and Sokal 1990; Chen et al. 1995; Crawford et al. 1997).

¹All authors are associated with the Laboratory of Biological Anthropology, Department of Anthropology, University of Kansas, Lawrence, KS 66045.

Ruhlen used linguistic data to investigate the origin of the Na-Dene language family of North America (Ruhlen 1998). Other linguists have grouped Na-Dene languages with the Caucasian or Sino-Tibetan families of Asia, rather than other New World languages (Bengston 1991; Nikolaev 1991). Based on 36 cognates, including words for body parts, tools, seasons, flora, and fauna, Ruhlen suggested that the Na-Dene linguistic family is most closely related to Yeniseian. This conclusion is controversial primarily because of the large geographic separation between these populations. The Na-Dene-speaking populations reside in Alaska, western Canada, the Pacific Northwest, and the American Southwest, while the Yeniseian speakers are located only in central Siberia.

The purpose of this study is to test Ruhlen's hypothesis of a phylogenetic relationship between Na-Dene and Yeniseian speakers, using nuclear gene frequencies and mitochondrial haplogroup frequencies. This research will determine whether it is possible to detect the coevolution of genes and language in these populations.

Materials and Methods

Populations. The Na-Dene language family includes Haida, Tlingit, and Eyak speakers, distributed along the Pacific Coast, and the Athabaskan languages. The Athabaskan branch consists of 30 languages, including Kutchin and Dogrib in Canada and Alaska, and Apache and Navaho in the southwestern United States (Ruhlen 1987). The Na-Dene language family has over 200,000 speakers.

The Yeniseian family has only one extant language, spoken by the Kets. Four other members of this family went extinct over the last two centuries (Grimes 2000). Today, the Kets live in two small communities (Sulamai and Kellogg) at the confluence of the Yenisey and Stony Tunguska Rivers in central Siberia, with fewer than 500 native speakers (Levin and Potapov 1964).

Samples. Gene frequency data including blood groups, immunoglobulins, and mitochondrial DNA (mtDNA) haplogroups were collected from the literature (see Appendix 1) for the following populations: Apache, Dogrib, Haida, Kutchin, Navajo, and Tlingit in the Na-Dene family; Blackfeet, Cree, Eskimo, Ojibwa, and Papago in non-Na-Dene of the Americas; and Altai, Asian Eskimo, Coastal Chukchi, Evenki, Forest Nentsi, Nganasan, Reindeer Chukchi, Sel'kup, and Yukaghir in Siberia. Blood samples were collected from the Kets during a previous field season by researchers from the University of Kansas. Blood typing was done at the Minneapolis War Memorial Blood Bank, and the immunoglobulins were characterized at the Analytical Genetic Testing Center, Denver, Colorado (Crawford et al. np). The genetic and molecular markers used in the analyses are shown in Table 1. The geographic locations of the populations described in this study are shown in Figure 1, and their language affiliations appear in Appendix 2.

Table 1. Genetic and Molecular Markers Used in Analyses, and Number of Populations for Which the Data Were Available

<i>Locus</i>	<i>Allele</i>	<i>n</i>	<i>Locus</i>	<i>Allele</i>	<i>n</i>
<i>ABO</i>	<i>B</i>	20	<i>GM</i>	<i>za;g</i>	10
	<i>O</i>	20		<i>zax;g</i>	10
<i>MNS</i>	<i>MS</i>	19	<i>KM</i> <i>MiDNA</i>	<i>Km1</i>	10
	<i>Ms</i>	19		<i>A</i>	20
	<i>M</i>	19		<i>B</i>	20
<i>RH</i>	<i>CDe (R1)</i>	20		<i>C</i>	20
	<i>cDE (R2)</i>	20		<i>D</i>	20
<i>FY</i>	<i>FY*A</i>	18			

Analytic Procedures. To estimate the genetic relationships between populations, nuclear and mitochondrial gene frequency data were converted to a relationship matrix (R-matrix) after the method of Harpending and Jenkins (1973) using ANTANA (Harpending and Rogers 1984). A matrix of distances between populations was calculated based on each allele, using the sample coefficient of kinship (r_{ij}) for each allele, such that

$$R_{ij} (p_i - \bar{p})(p_j - \bar{p})/\bar{p}(1 - \bar{p}), \tag{1}$$

where p_i and p_j are frequencies of allele p in populations i and j , and \bar{p} is the weighted mean gene frequency of allele p in the array. These allele distance matrices were averaged into the overall R-matrix, a covariance matrix of genetic distances between populations (Harpending and Jenkins 1973). The R-matrix was then used in a principal components analysis to provide a genetic map of relationships between groups. The R-matrix was not adjusted for census size, as this did not make a significant difference in the principal component analysis. An S-matrix of covariances among allele frequencies was calculated from the eigenvectors of the R-matrix to provide a representation of the allelic variation contributing to the spread of populations.

To test for the effects of gene flow and genetic drift on population relationships, a matrix of heterozygosity versus distance from the centroid of the gene frequency array (r_{ii}) was constructed using the method of Harpending and Ward (1982). According to this model, mean per locus heterozygosity is negatively correlated with genetic distance. Populations that fall below the theoretical regression line experience significantly more drift than those that are most proximal to the line. Those that have a higher rate of gene flow are found well above the theoretical regression. Heterozygosity and r_{ii} were calculated from frequency data for seven alleles in 17 populations using ANTANA v.1.1 (Harpending and Rogers 1984).

To test Ruhlen's hypothesis of a genetic relationship between Na-Dene and

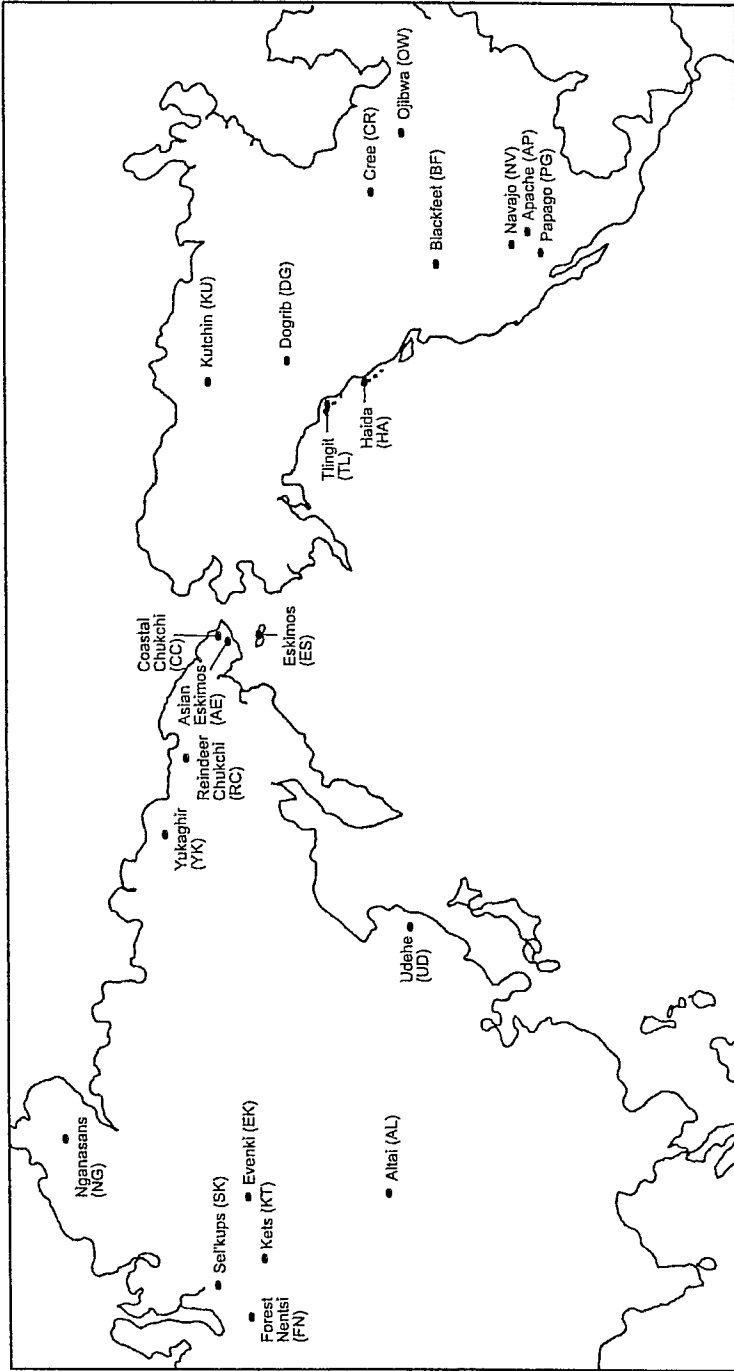


Figure 1. Map of population locations. The two-letter abbreviations presented here are used throughout the paper.

Yeniseian speakers, geographic, genetic, and linguistic distances were calculated for 17 populations. Great circle geographic distances were computed using GEOG 2.1 (Relethford 2000). Squared Euclidean genetic distances between the populations were calculated from the R-matrix using seven alleles. The genetic distance, d_{ij}^2 , between populations i and j is given as the relationship coefficient of populations i and j minus twice the difference in relationship coefficients between populations i and j (Harpending and Rogers 1973):

$$d_{ij}^2 = r_{ii} + r_{jj} - 2r_{ij}. \tag{2}$$

Two hierarchical linguistic distance matrices were constructed using scores based on Ruhlen's linguistic phylogeny (1987 and 1998). According to this classification, there are seven hierarchical levels within the Na-Dene language family. Therefore, in this analysis the most closely related Na-Dene languages, such as Navajo and Apache, were assigned a distance value of 2. The most distantly related languages in this family, for example Dogrib and Haida, were assigned a value of 7. Unrelated languages were assigned a value of 15, the minimum distance between linguistic phyla according to a worldwide analysis of the relationship between genes and languages by Chen et. al (1995). One linguistic matrix grouped the Kets with the Na-Dene, giving the Kets a distance value of 8 in relation to all Na-Dene languages. The second distance matrix treated the Kets as an isolated language, with a distance value of 15.

Mantel tests were performed to test for independence of the elements of all the distance matrices (genetic, linguistic, and geographic) and to calculate correlations between them. Because the distribution of gene frequencies and languages are both effected by geography, a partial correlation of genetic and linguistic distances holding geography fixed was calculated using the method of Smouse et al. (1986). In this method, both genetic (**A**) and linguistic (**B**₁ and **B**₂) distance matrices are regressed on the geographic distance matrix (**C**). The residual matrices (r_{AC} , r_{B1C} , r_{B2C}) from these regression analyses are then randomly permuted using the Mantel procedure. In this way, the relationship between genes and language (for both linguistic distance matrices) could be examined without the confounding effect of geography.

Results

Contrary to Ruhlen's hypothesis, genetic data show that the Na-Dene are most closely related to other Native American populations, while the Kets cluster with other Siberian groups. Figure 2 shows a plot of the first two principal components based on an R-matrix of 10 populations (Na-Dene and Siberians) and eight alleles (*ABO*B*, *ABO*O*, *MN*M*, *RH*RI*, *RH*R2*, *GM*AG*, *GM*AXG*, *KM*I*) from three blood group systems and immunoglobulins. The first principal component accounts for 55.8% of the total variation and separates the Na-Dene (AP, DG, HA, and NV) from the Siberian populations (AE, CC, EK, FN, KT, and

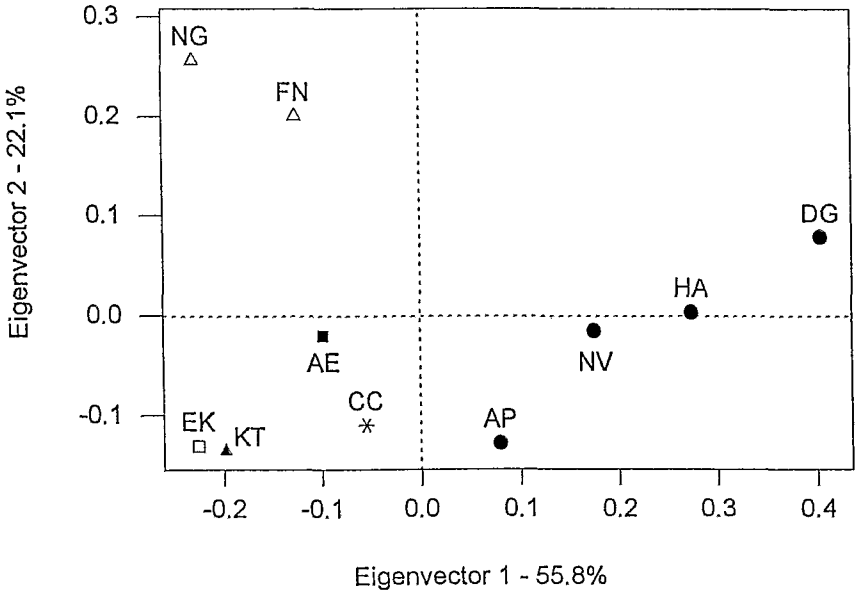


Figure 2. Plot of the relationship between six Siberian and four North American populations using eight alleles. The first two eigenvectors accounted for 77.9% of the total variance.

NG). The second principal component accounts for 22.1% of the total variation, and separates the Nganasan and Forest Nentsi from the remaining Siberian populations. The spread of alleles used in this analysis, shown in Figure 3, demonstrates that the Na-Dene groups share similar frequencies of *GM*AG*, *KM*I*, and *MN*M*. The Nganasan and Forest Nentsi differ from other Siberian groups with high frequencies of *RH*R2*.

Figure 4 is a plot of the first two principal components based on an R-matrix of 17 populations (Na-Dene, Siberians, and other Native American groups) using seven alleles (*ABO*B*, *ABO*O*, *MNS*MS*, *MNS*Ms*, *RH*RI*, *RH*R2*, *FY*A*). The first two principal components account for 72.9% of the total variation. The plot clearly shows that the Na-Dene (DG, HA, KU, NV, and TL) do not cluster with the Yeniseian speakers (KT). Likewise, the Kets (KT) are grouped with their central Siberian neighbors, the Evenki (EK) and Sel'kups (SK). The first eigenvector separates the Na-Dene from other Native American groups, while the second eigenvector separates Siberians from Native Americans. Figure 5 reveals that the Na-Dene populations cluster together based on similar frequencies of *ABO*O*, *FY*A*, *MNS*Ms*, and *RH*R2*, while the Siberians share similar frequencies of *ABO*B* and *RH*RI*. The Navajo differ from the other Na-Dene populations due to a higher frequency of *MNS*MS*.

Figure 6 is a plot of the first two principal components of an R-matrix of 17 populations and the four major New World mitochondrial DNA haplogroups (A,

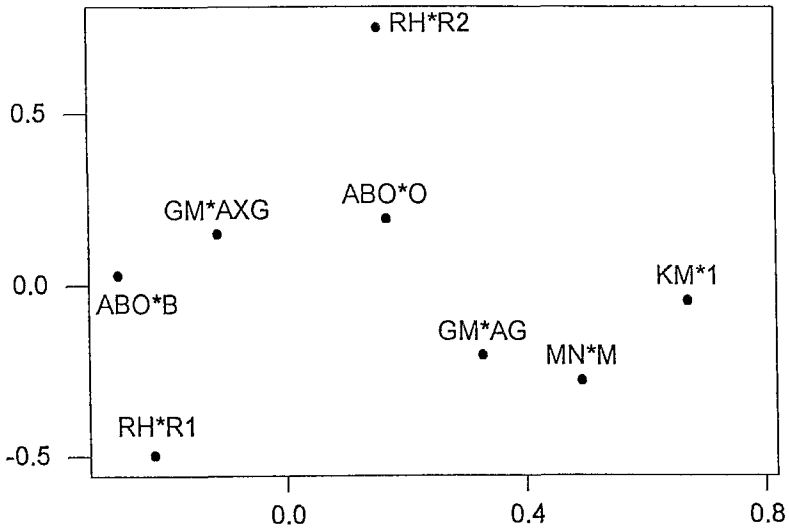


Figure 3. Plot of the S-matrix for eight alleles used in the genetic plot shown in Figure 2.

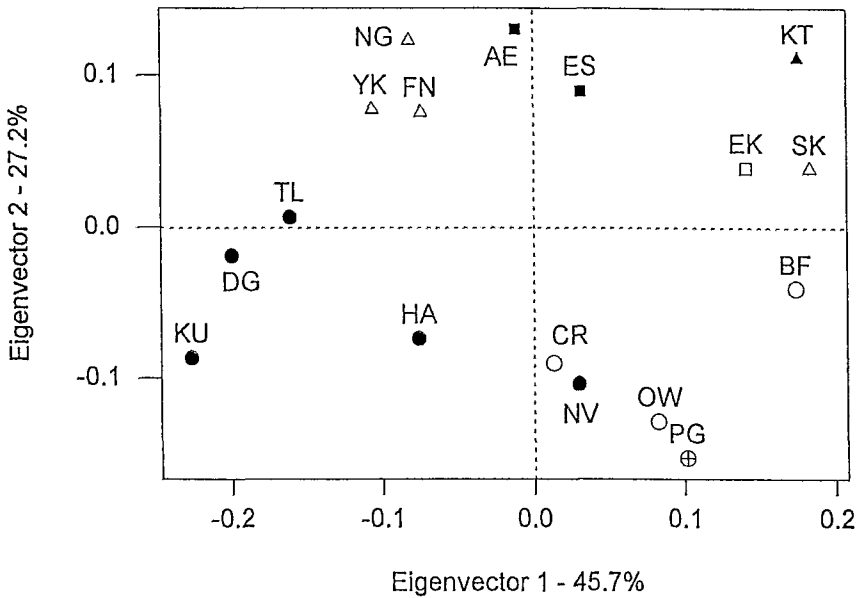


Figure 4. Plot of the relationship between seven Siberian and 10 North American populations using seven alleles. The first two eigenvectors accounted for 72.9% of the total variance.

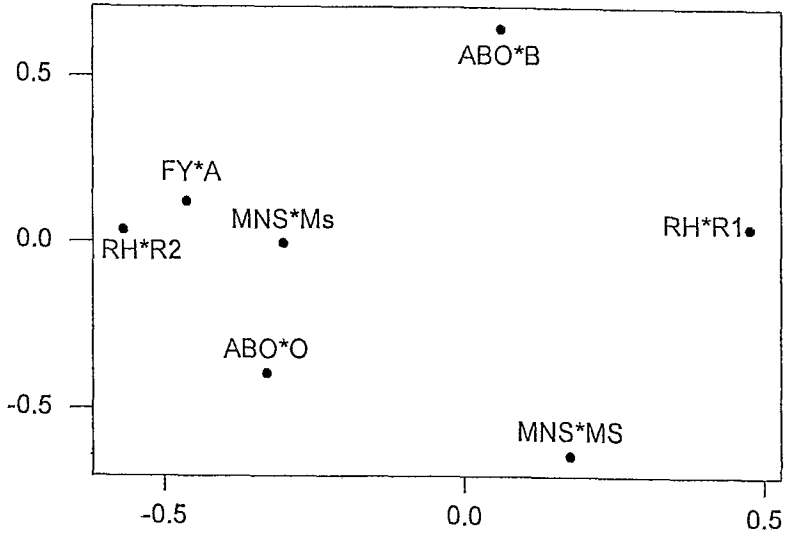


Figure 5. Plot of the S-matrix for seven alleles used in the genetic plot shown in Figure 4.

B, *C*, and *D*). The first two principal components account for 78.5% of the total variation in the sample. The first eigenvector separates the Native American and Siberian populations, with the exception of the Asian Eskimo who cluster with the northern Native Americans. The second eigenvector separates northern and southern Na-Dene groups. As shown in Figure 7, the Kets (KT) are separated from the Na-Dene (AP, DG, HA, and NV) by the presence of haplogroup *C* and absence of haplogroup *A*. *MTDNA***A* is highest in northern Native Americans, *MTDNA***B* is present in the southern Na-Dene groups but rare or absent in northern Native American and Siberian populations, and *MTDNA***C* and *MTDNA***D* are highest in Siberian populations.

Figure 8 provides a plot of mean per locus heterozygosity versus distance from the centroid (r_{ii}) for the populations used in Figure 4, relative to the theoretical regression line. The Kets, Blackfeet, and Navaho have the highest levels of heterozygosity, reflecting the most admixture with Russians and other nonnative groups, respectively. The Dogrib and Kutchin have the highest r_{ii} values, indicating that these populations have experienced the most genetic drift due to their isolation in northern Canada.

The results from pairwise and partial correlation analyses of the distance matrices are presented in Table 2. A significant correlation was found between genetic and geographic distances, and also between linguistic and geographic distances. There was no significant correlation ($r = 0.1102$, $p = 0.096$) between genes and language when the Kets were linguistically grouped with the Na-Dene. However, there was a significant correlation ($r = 0.2272$, $p = 0.015^*$) between genes and language when the Kets were considered linguistically unrelated to the

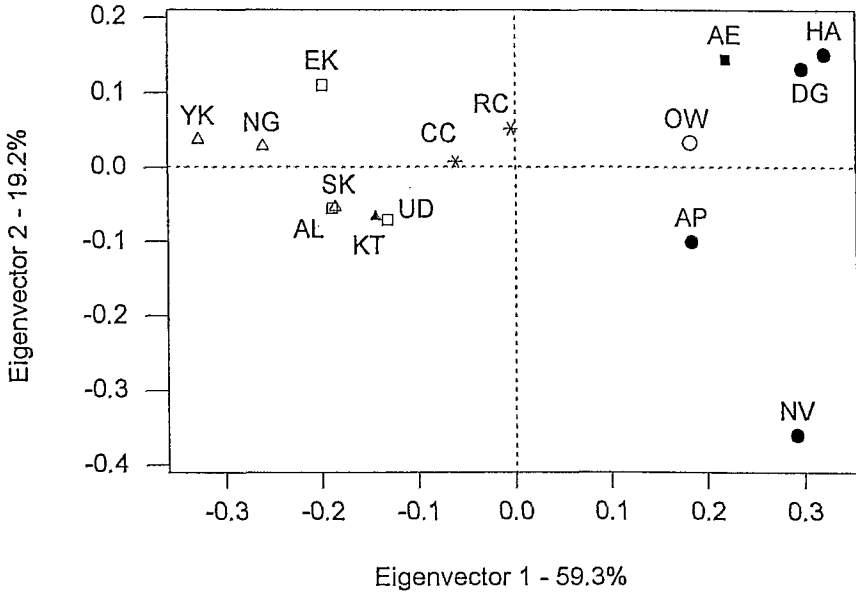


Figure 6. Plot of the relationship between 10 Siberian and seven North American populations using four mtDNA haplogroups. The first two eigenvectors accounted for 78.5% of the total variance.

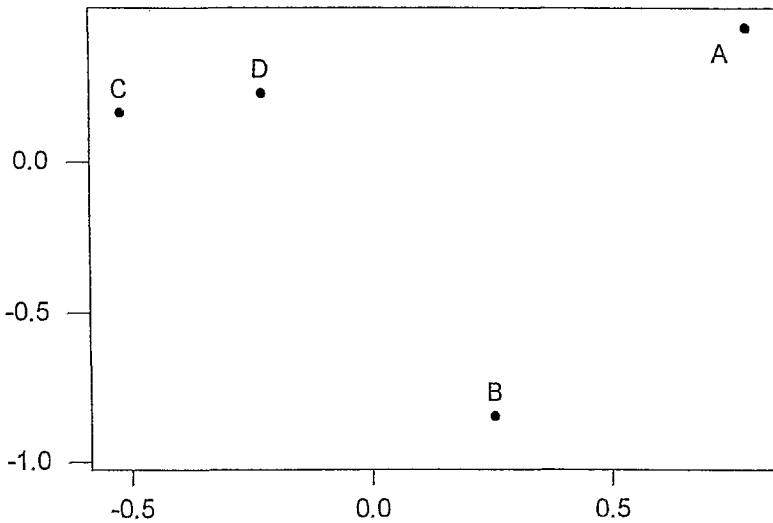


Figure 7. Plot of the S-matrix for four mtDNA haplogroups used in the genetic plot shown in Figure 6.

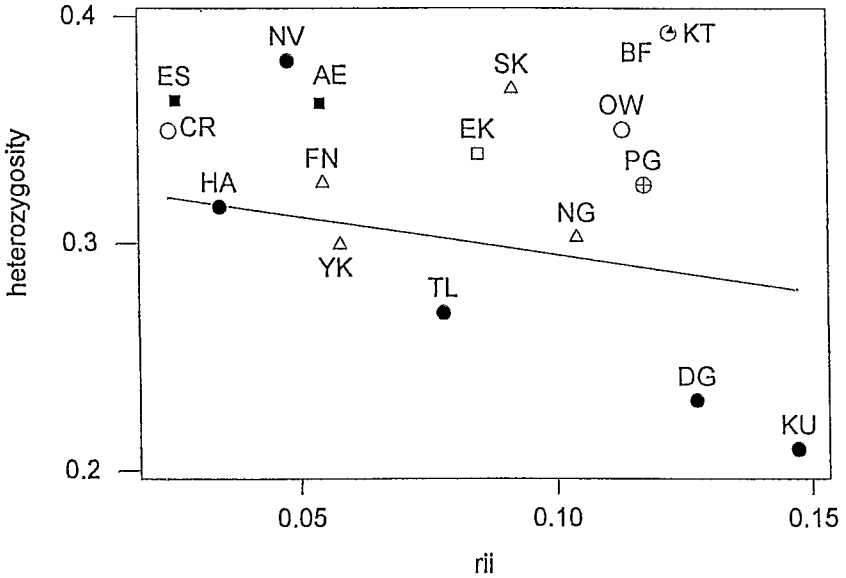


Figure 8. Plot of the regression of heterozygosity on the distance from the centroid of gene frequency distribution (r_{ii}). Line shows theoretical regression.

Table 2. Pairwise and Partial Correlations of Genetic (GENE), Geographic (GEOG), and Linguistic (LANG) Distance Matrices

<i>Test of Relationship</i>	<i>Correlation</i>	<i>p</i>
GENE*GEOG	0.2466	0.016*
GENE*LANG 1 ^a	0.1102	0.096
GENE*LANG 2 ^b	0.2272	0.015*
GEOG*LANG 1 ^a	0.3042	0.013*
GEOG*LANG 2 ^b	0.3909	0.002**
GENE*LANG 1 ^a (GEOG)	0.0381	0.345
GENE*LANG 2 ^b (GEOG)	0.1467	0.068

a. LANG 1 is the linguistic distance matrix with the Kets treated as related to the Na-Dene language family.

b. LANG 2 is the linguistic distance matrix with the Kets considered linguistically unrelated to the Na-Dene family.

Na-Dene. A partial correlation was run between genes and language, holding geography constant, in order to examine the relationship between genes and language without the confounding effect of geography. Regardless of which linguistic distance matrix was used, the partial correlations were lower than their respective pairwise correlations. However, the partial correlation that linguistical-

ly grouped the Kets with the Na-Dene was not significant ($r = 0.0381$, $p = 0.345$), while the partial correlation in which the Kets were considered linguistically distinct was slightly higher than convention ($r = 0.1467$, $p = 0.068$).

Discussion

Contrary to Ruhlen's proposed relationship between Na-Dene and Yeniseian speakers, this study did not find genetic evidence for a recent common origin of these groups. The results of the R-matrix analyses demonstrate that the Na-Dene are more closely related to other Native American groups, while the Kets, the only extant member of the Yeniseian language family, are similar to other Siberians, irrespective of whether standard genetic markers (blood group markers, immunoglobulins) or molecular markers (mtDNA haplogroup frequencies) are used. Other markers may give different results. An analysis of Y chromosomes, which included 20 Native Americans and 10 Kets, suggests that the most common Native American haplotype was descended from a haplotype that originated in central Siberia (Santos et al. 1999). However, only one of these samples was Na-Dene.

Despite the linguistic cognates shared between Na-Dene and Yeniseian language families, as reported by Ruhlen (1998), there was no significant correlation between genes and language when these languages were treated as part of the same family. The only significant correlation between genes and language occurred when Yeniseian was considered a separate language family. On a global scale, genetics and language affiliation are shown to be correlated. In a study by Chen et al. (1995) using 130 populations and 11 genetic systems, the correlation between genetic and linguistic distance data was $r = 0.361$, $p = 0.004$. In Europe, Sokal et al. (1992) found significant relationships between several genetic systems (including ABO, RH, FY, GM, and KM) and language, with an average correlation of $r = 0.141$. Siberian populations demonstrate a comparable correlation to that of populations worldwide, with $r = 0.35$, $p = 0.001$ (Crawford et al. 1997). Similar correlations between genes and language have not been reported for South America, however. In a reanalysis of genetic and linguistic distance data taken from a study of eight South American populations (Rothhammer et al. 1979), a Mantel test showed the correlation between genes and language was $r = 0.1653$, $p = 0.2070$. The authors suggest that the discordance between genes and language may be the result of recent linguistic evolution that is not accounted for by the classification used (Rothhammer et al. 1979).

In the present study, significant correlations were also found between genes and geography, and language and geography. The correlation between language and geography was higher, however, when the Kets were considered a linguistic outgroup ($r = 0.3909$, $p = 0.002$). These results are not surprising considering that both genetic and linguistic distances are spatially autocorrelated (Sokal et al. 1992). Partial correlations between genetic distance and both linguistic distance

matrices, holding geographic distance fixed, were not significant, indicating that geography accounts for the majority of the variation. Similar spatial patterning of genetic and linguistic distances has been reported previously. In Europe, Barbu-jani and Sokal (1990) demonstrated that zones of rapid genetic change were associated with linguistic differentiation, suggesting that not only are genes and languages similarly affected by geographic boundaries, but that language may also play a role in maintaining genetic differentiation among populations.

Given the discordance between the results of Ruhlen's analysis (1998) and the present study, a number of scenarios may explain how linguistic similarities could occur between such geographically isolated groups: (1) Na-Dene and Yeniseian speakers do not share common genetic ancestry, but they are linguistically related due to horizontal language transmission. According to Renfrew (1992), there are situations in which language replacement may occur, such as demic expansion or dominance by a small elite. Either the Na-Dene or the Yeniseian speakers may have at one time adopted their language from a genetically unrelated group. (2) Na-Dene and Yeniseian groups do share a common origin, but language and genes are no longer concordant because they evolved at different rates. This is more likely to have happened if the populations were separated for a long period of time. Analysis of language distribution in the Americas has revealed that the Na-Dene are not recent migrants from Asia, but rather had settled in North America prior to the end of the last ice age. Higher language densities were found in areas that were ice-free during the last glacial maximum, between 13,000 and 25,000 years ago (Rogers et al. 1990). For the Na-Dene, the highest language density, or the greatest number of different languages, occurs along the southern coast of Alaska, suggesting not only that this may be the homeland for Na-Dene languages, but also that the Na-Dene speakers were present in the New World during the last glaciation (Rogers et al. 1994). This would indicate that the split between Na-Dene and Yeniseian speakers happened before 13,000 years ago. Given this time frame, the likelihood that the genes and languages of these groups have changed at different rates is increased, and it is not surprising that the genetic and linguistic distances are not correlated. (3) The Na-Dene and Yeniseian speakers share common linguistic and genetic origins, but recent genetic differentiation has obscured this relationship. Genetic similarity may have been swamped by gene flow from surrounding populations. For example, the Navajo have a high heterozygosity value, which reflects significant gene flow into the population. This is not surprising, considering the Navajo are the largest Native American tribe in the United States. The Kets have also experienced admixture, and likewise have a high heterozygosity value, probably due to admixture with Russians over the last two generations, and/or admixture with other surrounding Siberian populations such as the Evenki and Sel'kups, who cluster with the Kets in the R-matrix analyses (Figures 2 and 4). It is also possible that genetic drift has resulted in the differentiation of populations that once shared genetic similarity. The high distance from the centroid (r_{ii}) values found in the Dogrib (DG) and Kutchin (KU) indicate that these two Na-Dene groups have undergone genetic drift (Fig-

ure 8), likely due to their geographic isolation in the Northwest Territories and the Yukon, respectively, and small population sizes (Appendix 1).

However, this study does show a significant correlation between genes and language when Na-Dene and Yeniseian are treated as distinct language families. These results suggest that the Na-Dene and Yeniseian populations are both genetically and linguistically unrelated, and the 36 cognates used by Ruhlen (1998) are insufficient evidence for a common recent origin of these two language families. Examined more closely, the majority (68%) of the Na-Dene languages included in Ruhlen's study had five cognates or less in common with Yeniseian. Of these, 77% shared two or fewer cognates for basic vocabulary. When investigating linguistic relationships, it is necessary to have not only phonetic correspondences but also a sufficient number of words for comparison, and of these, an adequate number of basic vocabulary words (Starostin 1991). While it appears that Ruhlen has met the first condition, we contend that he has not met the last two.

Appendix 1. Populations Used in Analyses, Location, Size, Language Family, and References

<i>Population</i>	<i>Location</i>	<i>Population Size</i>	<i>Language Family</i>	<i>References</i>
Altai	Mendur-Sokkon, Siberia	1,000	Turkic	5, 17
Apache	New Mexico	2,000	Na-Dene	1, 23, 25
Asian Eskimo	Chaplino, Chukotka	1,200	Eskimo-Aleut	15, 17, 24
Blackfeet	Montana	1,000	Algonquian	12
Coastal Chukchi	Siryeniki, Chukotka	4,000	Chukotko-Kamchatkan	13, 17
Cree	Manitoba	35,000	Algonquin	9
Dogrib	Northwest Territories	2,900	Na-Dene	10, 19, 20
Eastern Evenks	Udehe, Siberia	1,600	Tungus-Manchu	5, 17
Eskimo	St. Lawrence Island, Alaska	900	Eskimo-Aleut	7
Evenki	Poligus, Siberia	500	Tungus-Manchu	4, 24
Forest Nentsi	Varyogan, Siberia	1,500	Samoyedic	16
Haida	Queen Charlotte Islands, Canada	2,000	Na-Dene	6, 21, 23
Kets	Sulamai, Siberia	150	Yeniseian	5
Kutchin	Yukon Territory, Canada	2,600	Na-Dene	8
Navajo	Navajo Reservation, Arizona	200,000	Na-Dene	3, 22, 25
Nganasan	Avamskaya, Siberia	1,300	Samoyedic	14, 17
Ojibwa	Northern Ontario, Canada	8,000	Algonquin	18, 23
Papago	Sacaton, Arizona	20,000	Uto-Aztecan	11
Reindeer Chukchi	Amgyema, Siberia	12,000	Chukotko-Kamchatkan	15, 17
Sel'kup	Tolka-upon-Taz, Siberia	3,600	Samoyedic	14, 24
Tlingit	Sitka, Alaska	1,200	Na-Dene	2
Yukaghir	Northeast Siberia	1,100	Uralic-Yukaghir	14, 24

References: (1) Brown et al. 1958; (2) Corcoran et al. 1959; (3) Corcoran et al. 1962; (4) Crawford et al. 2002; (5) Crawford, Schanfield et al. unpublished; (6) Field et al. 1988; (7) Ferrell et al. 1981; (8) Lewis et al. 1961; (9) Lucciola et al. 1974; (10) Merriwether et al. 1995; (11) Niswander et al. 1978; (12) Rokala et al. 1977; (13) Sheremetyeva and Rychkov 1978; (14) Sukernik et al. 1978; (15) Sukernik, Lemsä et al. 1981; (16) Sukernik, Osipova et al. 1981; (17) Sukernik et al. 1996; (18) Szathmary et al. 1975; (19) Szathmary 1983; (20) Szathmary et al. 1983; (21) Thomas et al. 1964; (22) Torroni et al. 1992; (23) Torroni, Schurr et al. 1993; (24) Torroni, Sukernik et al. 1993; (25) Williams et al. 1985.

Appendix 2. Language Affiliations of Populations Used in Analysis (based on Ruhlen 1998 and Grimes 2000)

North American

I. Na-Dene (●)

A. Haida (HA)

B. Continental Na-Dene

1. Tlingit (TL)

2. Athabaskan-Eyak

a) Athabaskan

(1) Canadian

(a) Hae-Chipewayan

(i) Dogrib (DG)

(b) Han-Kutchin

(i) Kutchin (KU)

(2) Apachean

(a) Navajo-Apache

(i) Western Apache-Navajo

(a) Navajo (NV)

(b) Apache (AP)

II. Algonquian (○)

A. Central

1. Cree (CR)

2. Ojibwa (OW)

B. Plains

1. Blackfeet (BF)

III. Uto-Aztecan (⊕)

A. Southern Uto-Aztecan

1. Sonoran

a) Tepiman

(1) Papago (O'Odham) (PG)

Siberian

I. Yeniseian (▲)

A. Ket (KT)

II. Altaic (□)

A. Turkic

1. Common Turkic

a) Northern

(1) Altai (AL)

B. Mongolian-Tungus

1. Tungus

a) Northern

(1) Evenki (EK)

b) Southern

(2) Udehe (UD)

III. Uralic-Yukaghir (△)

A. Yukaghir (YK)

B. Uralic

1. Samoyed

a) North

(1) Forest Nentsi (FN)

(2) Ngansan (NG)

b) South

(1) Sel'kup (SK)

Appendix 2. (Continued)

Siberian (continued)

IV. Eskimo-Aleut (■)

A. Eskimo

1. Yupik

a) Alaskan (St. Lawrence Island) (ES)

b) Siberian

(1) Chaplino-Naukan

(a) Chaplino (AE)

V. Chukchi-Kamchatkan (*)

A. Northern

1. Chukchi (CC, RC)

Note: Symbols in parentheses indicate language family affiliation in Figures 2, 4, 6, 8. Abbreviations in parentheses indicate populations used in analyses (Figures 2, 4, 6, 8).

Acknowledgments This research was supported in part by the National Science Foundation through grants BSR 9101571 and OPP 990590.

Received 20 February 2002; revision received 30 August 2002

Literature Cited

- Barbujani, G., and A. Pilastro. 1993. Genetic evidence on origin and dispersal of human populations speaking languages of the Nostratic macrofamily. *Proc. Natl. Acad. Sci. USA* 90:4670–4673.
- Barbujani, G., and R.R. Sokal. 1990. Zones of sharp linguistic change in Europe are also linguistic boundaries. *Proc. Natl. Acad. Sci. USA* 87:1816–1819.
- Bateman, R., I. Goddard, R. O'Grady et al. 1990. Speaking of forked tongues: The feasibility of reconciling human phylogeny and the history of language. *Curr. Anthropol.* 31(1):1–24.
- Bengston, J.B. 1991. Notes on Sino-Caucasian. In *Dene-Sino-Caucasian Languages*, V. Shevoshkin, ed. Bochum, Germany: Brockmeyer, 67–129.
- Brown, K.S., B.L. Hanna, A.A. Dahlberg et al. 1958. The distribution of blood group alleles among the Indians of southwest North America. *Am. J. Hum. Genet.* 10:175–195.
- Cavalli-Sforza, L.L., and M.W. Feldman. 1981. *Cultural Transmission and Evolution*. Princeton, NJ: Princeton University Press.
- Chen, J., R.R. Sokal, and M Ruhlen. 1995. Worldwide analysis of genetic and linguistic relationships of human populations. *Hum. Biol.* 67:595–612.
- Corcoran, P.A., F.H. Allen, A.C. Alison et al. 1959. Blood groups of Alaskan Eskimos and Indians. *Am. J. Phys. Anthropol.* 17:187–193.
- Corcoran, P.A., D.L. Rabin, and F.H. Allen. 1962. Blood groups of 237 Navajo school children at Pinon Boarding School, Pinon, Arizona. *Am. J. Phys. Anthropol.* 20:389–390.
- Crawford, M.H., J. McComb, M.S. Schanfield et al. 2002. Genetic structure of Siberian populations: The Evenki of central Siberia and the Kizhi of Gorno Altai. In *Human Biology of Pastoralist Populations*, W.R. Leonard and M.H. Crawford, eds. Cambridge, UK: Cambridge University Press, 10–49.

- Crawford, M.H., M.S. Schanfield, M. Mount et al. Genetic structure of Siberian populations: Standard blood markers. Unpublished.
- Crawford, M.H., J. T. Williams, and R. Duggirala. 1997. Genetic structure of the indigenous populations of Siberia. *Am. J. Phys. Anthropol.* 104:177–192.
- Ferrell, R.E., R. Chakraborty, H. Gershowitz et al. 1981. The St. Lawrence Island Eskimos: Genetic variation and genetic distance. *Am. J. Phys. Anthropol.* 55:351–358.
- Field, L.L., J.P. Gofton, and T.D. Kinsella. 1988. Immunoglobulin (GM and KM) allotypes and relation to population history in Native Peoples of British Columbia: Haida and Bella Coola. *Am. J. Phys. Anthropol.* 76:155–163.
- Grimes, B.F., ed. 2000. *Ethnologue*. Summer Institute of Linguistics, Dallas, Texas.
- Harpending, H., and T. Jenkins. 1973. Genetic distance among Southern African populations. In *Methods and Theories of Anthropological Genetics*, M.H. Crawford, and P.L. Workman, eds. Albuquerque, NM: University of New Mexico Press.
- Harpending, H., and A. Rogers. 1984. *ANTANA: A Package for Multivariate Data Analysis* v.1.1. Distributed by the authors.
- Harpending, H., and R. Ward. 1982. Chemical Systematics and Human Populations. In *Biochemical Aspects of Evolutionary Biology*, M. Nitecki, ed. Chicago, IL: University of Chicago Press, 213–256.
- Levin, M., and L. Potapov. 1964. *The Peoples of Siberia*. Chicago, IL: University of Chicago Press.
- Lewis, M., J.A. Hildes, H. Kaita et al. 1961. The blood groups of the Kutchin Indians at Old Crow, Yukon Territory. *Am. J. Phys. Anthropol.* 19:383–389.
- Lucciola, L., H. Kaita, J. Anderson et al. 1974. The blood groups and red cell enzymes of a sample of Cree Indians. *Can. J. Genet. Cytol.* 16:691–695.
- Merriwether, D.A., F. Rothhammer, and R.E. Ferrell. 1995. Distribution of the four-founding lineage haplotypes in Native Americans suggests a single wave of migration for the New World. *Am. J. Phys. Anthropol.* 98:411–430.
- Nikolaev, S.L. 1991. Sino-Caucasian languages in America. In *Dene-Sino-Caucasian Languages*, V. Shevoroshkin, ed. Bochum, Germany: Brockmeyer, 42–66.
- Niswander, J.D., K.S. Brown, B.Y. Iba et al. 1970. Population studies on Southwestern Indian tribes. I. History, culture, and genetics of the Papago. *Am. J. Hum. Genet.* 22:7–23.
- Relethford, J. 2000. *GEOG*. Distributed by the author.
- Relethford, J. 1996. Mantel for Windows, v.3. Department of Anthropology, SUNY College at Oneonta, Oneonta, NY.
- Renfrew, C. 1992. Archaeology, genetics and linguistic diversity. *Man* 27:445–478.
- Rogers, R.A., L.D. Martin, T.D. Nicklas. 1990. Ice-Age geography and the distribution of native North American languages. *Journal of Biogeography* 17:131–143.
- Rogers, R.A., L.A. Rogers, L.D. Martin. 1994. The distribution of Nadene speakers and needle-leaf evergreen forests: A western North American parallel. *Institute of Tertiary-Quaternary Studies—TER-QUA Symposium Series* 2:107–117.
- Rokala, D.A., H.F. Polesky, and G.A. Matson. 1977. The genetic composition of reservation populations: The Blackfeet reservation, Montana, U.S.A. *Hum. Biol.* 49:19–29.
- Rothhammer, F., R. Chakraborty, and E. Liop. 1979. Dermatoglyphic variation among South American tribal populations and its association with marker gene, linguistic, and geographic distances. In *Dermatoglyphics Fifty Years Later*, W. Wertelecki and C.C. Plato, eds. New York, NY: Liss, 169–179.
- Ruhlen, M. 1998. The origin of the Na-Dene. *Proc. Natl. Acad. Sci. USA* 95:13994–13996.
- Ruhlen, M. 1987. *A Guide to the World's Languages*. Stanford, CA: Stanford University Press.
- Santos, F.R., A. Pandya, C. Tyler-Smith et al. 1999. The central Siberian origin for Native American Y chromosomes. *Am. J. Hum. Genet.* 64:619–628.
- Sheremet'yeva, V.A., and Y.G. Rychkov. 1978. *Population Genetics of the People of Northeastern Asia*. Moscow University (in Russian).
- Smouse, P.E., J.C. Long, and R.R. Sokal. 1986. Multiple regression and correlation extensions of the Mantel test of matrix correspondence. *Syst. Zool.* 35(4):627–632.

- Sokal, R.R., N.L. Oden, and B.A. Thomson. 1992. Origins of the Indo-Europeans: Genetic evidence. *Proc. Natl. Acad. Sci. USA* 89:7669-7673.
- Starostin, S.A. 1991. On the hypothesis of a genetic connection between the Sino-Tibetan languages and the Yeniseian and North-Caucasian languages. In *Dene-Sino-Caucasian Languages*, V. Shevoroshkin, ed. Bochum, Germany: Brockmeyer, 12-41.
- Sukernik, R.I., T.M. Karaphet, and L.P. Osipova. 1978. Distribution of blood groups, serum markers and red cell enzymes in two human populations from northern Siberia. *Hum. Hered.* 28:321-327.
- Sukernik, R.I., S.V. Lemza, T.M. Karaphet et al. 1981. Reindeer Chukchi and Siberian Eskimos: Studies on blood groups, serum proteins and red cell enzymes with regard to genetic heterogeneity. *Am. J. Phys. Anthropol.* 55:121-128.
- Sukernik, R.I., L.P. Osipova, S.V., T.M. Karaphet et al. 1981. Studies on blood groups and other genetic markers in Forest Nentzi: Variation among the subpopulations. *Hum. Genet.* 53:397-404.
- Sukernik, R.I., T.G. Schurr, E.B. Starikovskaya et al. 1996. Mitochondrial DNA variation in Native Siberians, with special reference to the evolutionary history of American Indians. *Genetika* 32:432-439.
- Szathmary, E.J.E. 1983. Dogrib Indians of the Northwest Territories, Canada: Genetic diversity and genetic relationship among the subarctic Indians. *Ann. Hum. Biol.* 10:147-162.
- Szathmary, E.J.E., R.E. Ferrell, and H. Gershowitz. 1983. Genetic differentiation in Dogrib Indians: Serum protein and erythrocyte enzyme variation. *Am. J. Phys. Anthropol.* 62:249-254.
- Szathmary, E.J.E., J.F. Mohn, H. Gershowitz et al. 1975. The northern and southeastern Ojibwa: Blood group systems and the causes of genetic divergence. *Hum. Biol.* 47:351-368.
- Thomas, J.W., M.A. Stuckey, H.S. Robinson et al. 1964. Blood groups of the Haida Indians. *Am. J. Phys. Anthropol.* 22:189-192.
- Torrioni, A., T.G. Schurr, M.F. Cabell et al. 1993a. Asian affinities and continental radiation of the four founding native American mitochondrial DNAs. *Am. J. Hum. Genet.* 53:563-590.
- Torrioni, A., T.G. Schurr, C.C. Yang et al. 1992. Native American mitochondrial DNA analysis indicates that the Amerind and the Nadene populations were founded by two independent migrations. *Genetics* 130:153-162.
- Torrioni A., R. Sukernik, T. Schurr et al. 1993b. Mitochondrial DNA variation of aboriginal Siberians reveal distinct genetic affinities with Native Americans. *Am. J. Hum. Genet.* 53:591-608.
- Williams R.C., A.G. Steinberg, H. Gershowitz et al. 1985. GM allotypes in Native Americans: Evidence for three distinct migrations across the Bering Land Bridge. *Am. J. Phys. Anthropol.* 66:1-19.
- Zolotareva, I.M., and A.G. Bashlay. 1968. Études sérologiques en Yakoutie (in Russian). *Sovet. Ethnogr.* 1:46-55.