Virologic and genetic studies relate Amerind origins to the indigenous people of the Mongolia/Manchuria/southeastern Siberia region
(human T-cell lymphotropic virus type II in Siberia/Amerindian origins/mitDNA in Siberia)

JAMES V. NEEL*, ROBERT J. BIGGAR†, and REM I. SUKERNIK‡

*Department of Human Genetics, University of Michigan Medical School, Ann Arbor, MI 48109-0618; †Viral Epidemiology Section, Division of Cancer Etiology, National Cancer Institute, Rockville, MD 20852; and ‡Institute of Cytology and Genetics, Siberian Branch, Russian Academy of Sciences, Novosibirsk, Russia

Contributed by James V. Neel, July 14, 1994

ABSTRACT A commonly held theory is that the first wave of migrants into the New World was derivative from the ethnic groups then inhabiting eastern Siberia. However, these ethnic groups lack a mtDNA haplogroup (B) that is well represented in Amerindian tribes. Also, the time depth of the other three mtDNA haplogroups found in Amerindians (A, C, and D) appears to be greater in the Amerindians than in the eastern Siberian ethnic groups. In this communication we demonstrate that the human T-cell lymphotropic virus type II, present in 11 of the 38 Amerindian tribes thus far examined, is not present in any of the 10 ethnic groups of eastern Siberia that we have studied. However, the virus has just been reported in the indigenous population of Mongolia, and mtDNA haplogroup B is also represented in this region. On the basis of these facts, we propose that the ancestors of the first migrants to the New World were not derived from north and central Siberia but from populations to the south, inhabiting the regions of Mongolia, Manchuria, and/or the extreme southeastern tip of Siberia.

The question of the origin of the first wave of migrants to enter the Americas by way of the Bering Land Bridge and the route they took in East Asia has intrigued scholars of many persuasions. Recently, two new approaches to investigating this question have become available. The first has been the demonstration that a type of human T-cell lymphotrophic retrovirus (HTLV-II) is widely distributed throughout American Indian tribes and could be presumed to have its origins in the Old World. The second is the recognition that the relatively rapid evolution of mitochondrial DNA (mtDNA) might offer insights into the recent genetic relationships of human populations more readily than the more slowly evolving nuclear DNA. In this communication, we report the results of surveying 10 of the ethnic groups of eastern Siberia for infection with the HTLV-II virus. Synthesizing the results of this study with other recent developments regarding the distribution of HTLV-II and of mtDNA types in the Americas and northeastern Asia, we suggest that the direct Asian ancestors of the first wave of Amerindians to enter the New World originated in eastern Central Asia in the region now designated Mongolia/Manchuria/extreme southeastern Siberia.

Background

Current Relevant Data Regarding HTLV-II. The HTLV family of viruses first came to light in 1980 when type-C retrovirus particles were isolated from a patient with a type of adult T-cell leukemia/lymphoma (1), a rare type of lymphoma that had first been well delineated on the basis of patients from Southern Japan (2). This virus was subsequently termed human T-cell lymphotrophic virus type I (HTLV-I). In 1982, a related virus, HTLV-II, was recovered from a Caucasoid male patient with hairy cell leukemia (3). The initial recovery of the virus from a patient with this diagnosis may have been coincidental, and the clinical consequences of HTLV-II infection, if any, have been difficult to document.

Serologically, the differentiation between HTLV-I and HTLV-II is generally on the basis of sensitive enzyme-linked immunosorbent assay (ELISA) tests and immunoblot analyses, but the identification of either virus is definitively made by polymerase chain reaction (PCR) tests with type-specific primers and probes (see ref. 4). In this communication, we will accept the presence of HTLV-II infection in a tribe/ethnic group on the basis of convincing immunoblot data but note that the majority of groups in which serology indicates the presence of HTLV-II have culture and/or PCR-confirmed diagnoses in at least some members of the group. Both HTLV-I and HTLV-II can be transmitted by fresh blood and blood products, but they are probably normally transmitted by mother-to-infant (especially through breast milk) or sexual modes.

With respect to Amerindians, many tribes have now been described as having endemic HTLV-II. Because there is considerable cross-reactivity between antibodies raised by HTLV-I and -II, early serological studies, which used antigens from HTLV-I, did not clearly distinguish the type of HTLV found in various populations. This confusion has been rectified by establishing conventions for differentiating HTLV-I from -II (5) (which we follow) and by obtaining definitive results from PCR testing, not earlier available. Through use of these approaches, HTLV-II has been demonstrated in the Navajo and the Pueblo tribes of Arizona (6–8), the Seminole of Florida (9), the Guaymi of Panama (10–12), the Tunebo and Wayuu of Columbia (13, 14), the Cayapo and Kraho of Brazil (4), and the Tobas and Matacos of Argentina (15, 16). In all, 11 (29%) of the 38 North and South American tribes so far examined have had endemically spread HTLV-II. Although the North American tribes and perhaps others are characterized by Caucasoid and Negroid admixture, there has been little or no admixture of the South American groups in whom infection has been detected. Therefore, we accept that the virus is endemic in Amerindians, and the wide distribution of HTLV-II in these groups implies a great antiquity of this infection.

Abbreviations: HTLV-I and HTLV-II, human T-cell lymphotrophic virus types I and II; YBP, years before present.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.
Evidence is accumulating that there have been several waves of colonizers that entered the New World (17-19), although the details remain quite controversial (20). All but one of the tribes in which the HTLV-II virus has been encountered are classified as Paleo-Indians or Amerinds, who are seen as the descendants of the early wave or waves of migration, as contrasted with the later waves, which resulted in the Na-dene-speaking American Indians and the Eskimos; the sole exception are the Navajo, the southern-most representatives of the Na-dene-speaking American Indians, whose ancestors may have acquired the virus subsequent to arrival in the New World through intermarriage with tribes in which the virus was present. For present purposes we are concerned only with the occurrence of the virus in Amerinds and suggest that it was endemic among the ancestors of the Amerinds when they reached the Americas.

Recently, there have appeared reports of the occurrence of HTLV-II in other ethnic groups. The first such reports involved African natives, especially Ituri forest (Zaire) pygmies (21-23). Accordingly, human infection with the virus may be quite ancient, a postulate strengthened by the apparently benign nature of the infection. However, the key report from the standpoint of this presentation is the recent demonstration of HTLV-II in 3 among 96 natives of Mongolia (formerly the Mongolian Peoples Republic; ref. 26). Given the failure to demonstrate HTLV-II in Siberia, which we will shortly describe, these findings constitute one of several pivotal facts in formulating a theory of Amerind origins.

**Current Relevant Data Regarding mtDNA.** Because of its apparently relatively rapid rate of evolutionary change, studies of mtDNA are currently very much in vogue in evolutionary biology. With respect to Amerinds, the appropriate studies of a select set of 20 Amerind tribes have been interpreted as indicating that all of the indigenous mtDNA variants recognized by a battery of 14 restriction enzymes fall into four different haplogroups, termed A, B, C, and D (27, 28). Recently, evidence for a fifth minor haplogroup (E) in Amerinds has been forthcoming (29). (A haplogroup is a collection of haplotypes all of which share at least one common general variant. Haplogroups are also referred to as mtDNA lineages.) A phylogeny can be constructed for each of the four dominant haplogroups that traces the constituent haplotypes back to one or more haplotypes still present in the population (28). (This demonstration does not imply that only the apparent founding haplotypes were present in the early wave of migrants; these haplotypes are the survivors among whatever haplotypes were present in the earliest migrants.) Studies of 10 of the aboriginal ethnic groups of northeastern Siberia have revealed the presence of four of these haplogroups (A, C, D, and E), but thus far haplogroup B has not been encountered (30).

Haplotypes certainly or probably belonging to the missing haplogroup B have been reported, however, in samples from Mongolia (31, 32) as well as in the Southern Altai people, who are residents of Uebes-Kan, a village in the Russian Altai Mountains just to the northwest of Mongolia (33), and in the Russian Buryats, who are situated just to the northeast of Mongolia (34); the Russian Buryats are commonly regarded as an extension of the ethnic groups of Mongolia. Members of haplogroup B are also widely distributed in East Asia (35-39). These findings have led to the suggestion that the ancestry of the Amerindian is multipartite, derived in part from northeastern Siberian groups (the source of haplogroups A, C, and D) and in part from groups to the south (the source of haplogroup B and possibly an additional source of the other four haplogroups) (30, 34).

On the basis of estimates of the rate of mtDNA evolution of 2.0% to 4.0% per million years (see refs. 40-42), Torroni et al. (28) have estimated the ages of the four chief mtDNA haplogroup lineages in the Amerinds as follows: A, 22,750-45,500 years before present (YPB); B, 6,000-12,000 YPB; C, 24,000-48,000 YPB; D, 13,250-25,500 YPB. The range in these estimates reflects the uncertainty concerning the rate of evolution of mtDNA and does not include the large statistical uncertainties inherent in estimates of this type (43). With the same statistical approach, Torroni et al. (30) estimated the time depth for haplogroups C and D in Siberia at 15,000-30,000 YPB and 10,000-20,000 YPB, respectively. Since the B haplogroup is missing in the ethnic groups studied in Siberia, there is no time depth to be calculated.

The treatment by Torroni et al. (30) of their findings as regards time depth with respect to the A haplogroup deserves scrutiny. They write in ref. 30 on pages 604-605, "Haplogroups C and D were the only haplogroups containing sufficient aboriginal Siberian and Native American mtDNA haplotypes to permit estimation of the continent-specific diversity. Haplogroup B was absent in Siberia, and Siberian haplogroup A was represented by only two complete haplotypes observed in the Evenks." This statement requires examination. For only 3 of the 10 Siberian populations that they examined were the blood samples sufficiently well preserved to permit complete haplotyping. In these 3 populations (Niviks, Evenks, and Udgeys; total of 153 samples), there were nine haplogroup C haplotypes, seven haplogroup D haplotypes, but only the two aforementioned haplogroup A haplotypes. While to be sure two haplotypes is scarcely the basis for an estimate of continent-specific diversity, nevertheless, this value of two was arrived at by the same sampling procedure that yielded values of nine and seven for the other two haplogroups. The age of a haplogroup is proportional to the number of variants it encompasses. If, as postulated, northern and central Siberia are the origin of the haplogroups A, C, and D of the Amerinds and if these groups were accordingly in this area when the ancestors of the Amerinds left the region, then the paucity of haplogroup A haplotypes thus far is in contrast to the findings as regards haplogroups C and D. Although on the face of it haplogroup A would appear to be a relatively recent intrusion into Siberia, there is no logical basis for excluding it from the calculation of mtDNA time depth in this region. We note, however, that even when the argument is restricted to haplogroups C and D, the Siberian lineages appear to be younger than the Amerind, whereas in a derivative situation, the reverse should be true. We will return to the interpretation of this observation later.

**Data Concerning the Absence of HTLV-II in Northeastern Siberia**

The demonstration of widespread seropositivity to HTLV-II in Amerindians created an obvious interest in evidences for HTLV-II infection in Siberia. We now present the results of studies examining the occurrence of HTLV-II in 10 ethnic groups of northeastern Siberia. Of these 10 Siberian populations, 9 have been studied with respect to mtDNA types, and a summary of these groups will be found in Torroni et al. (30). The 10th group is the Nymylan Koryaks, living on the eastern coast of the Kamchatka Peninsula. There are clear dialectical and cultural differences between this population and the Chavchuven Koryaks studied earlier. In particular, the principal subsistence of the Chavchuven Koryaks is through reindeer herding, whereas the Nymylan Koryaks subsist mainly by fishing and hunting small sea mammals. We judge the differences between these two groups to be equivalent to those that would distinguish tribes among the American Indians.

Samples were collected between 1974 and 1993. All subjects were adults and included approximately equal numbers of males and females. The populations studied, living in small and isolated villages, represent the remnants of indigenous
tribes, each with a local language and identifying itself as a separate group. However, some groups living in close proximity, particularly the Chukchis and Koryaks, have intermarried with each other. Natives with known Russian admixture have been excluded from the study.

The breakdown by ethnic group for the 473 samples examined was as follows: Siberian Eskimos (three villages, 41 subjects), Chukchis (six villages, 92 subjects), Chavchuvren Koryaks (two villages, 47 subjects), Nymylan Koryaks (three villages, 107 subjects), Ngangasans (three villages, 46 subjects), Yukagirs (two villages, 24 subjects), Evens (two villages, 38 subjects), Udegeys (one village, 46 subjects), and Sel’kups (one village, 18 subjects). Cell pellets, but no sera or plasma, were available for a 10th group, the Nivkhi of northern Sakhalin Island. Under the assumption that the latter group the oldest persons would be those most likely to have HTLV, we obtained the cellular material from the 14 oldest members of this group (ages ranged from 58 to 91 years), extracted the DNA, and used the PCR method with nested primers capable of amplifying HTLV-I and HTLV-II to obtain the material for testing. The locations of the groups studied are shown in Fig. 1.

The sera or plasma were tested with whole HTLV-I virus [Cambridge HTLV ELISA (Cambridge, MA) and Genetic Systems (Seattle)] and purified HTLV envelope protein assays (Cambridge Biotech). Both assays are capable of detecting HTLV-I and HTLV-II, but in our experience with the same assays on Amerinds of Central and South America, the envelope assay is more sensitive for HTLV-II than the whole-virus assay.

Of the 459 samples tested, 4 were positive in the whole-virus assays (all weakly so), and none was positive in the envelope assay. The 4 possibly positive samples were immunoblotted (Cambridge Biotech immunoblots enhanced with p21e, a purified HTLV envelope protein). The immunoblots of 3 samples lacked any envelope bands (1 was completely negative, 1 had a weak p24 band only, and 1 had a weak p19 band and a trace p24 band but no other band). The immunoblot from sample 4, from a 37-year-old Nymylan Koryak female, had a trace p21e (env) band and a weak p24 (gag) band. On rebloating with an immunoblot containing type-specific antigens for HTLV-I and -II (Cellular Products), no reactivity was seen at any other bands, including either of the type-specific bands. Even though this latter sample met the technical criteria for positivity (reactivity against gag and env bands), we consider the results likely to be a nonspecific reaction because the observed reactivity was barely present and the type-specific markers were not observed. In addition, this 37-year-old female was the only reactive person in a village where almost all members were tested, many of whom must have been related to her. In other tribes in which HTLV-II has been described, from 5% to 37% of the population has been seropositive (average 13%). Thus, none of nine tribes tested serologically presents evidence of infection with HTLV-II. In addition, a 10th group from Sakhalin Island was tested by PCR because no sera/plasma remained. In the 14 oldest individuals (ages from 58 to 90 years), none was reactive. Thus, there were 10 groups in which we failed to detect convincing evidence of HTLV-II.

The failure to detect the virus with certainty in any of these 10 ethnic groups is of course not absolute proof the virus is not present in any of them. However, the Siberian samples are generally of a size adequate to detect virus if present. The designation "ethnic group" in Siberia is probably more encompassing than the designation "tribe" in the Americas—i.e., an ethnic group may in the past have contained subdivisions that would have been given tribal status if encountered in Native Americans. If, however, we equate American Indian tribes with Siberian ethnic groups, we can proceed with a simple calculation. If the proportion of positive ethnic groups was similar in eastern Siberia to the findings among tribal American Indians reviewed earlier, the probability of not encountering a positive ethnic group among the 10 groups sampled would be (26/37)10, or 0.029. Inasmuch as this probability is not a priori but empirically derived, it is only approximate. Because, as noted above, ethnic group is a larger population unit than tribe, this is a conservative calculation.

HYPOTHESIS: The most proximal Asian ancestors of all the Amerinds share a common origin with the indigenous people of the general region now designated Mongolia/Manchuria/extreme southeastern Siberia.

The foregoing review and data suggest three points vital to understanding the provenance of the ancestors of the Amerinds—namely, (i) infection with HTLV-II has not been detected in the ethnic groups of Siberia but has been detected in the ethnic groups of the present Mongolia; (ii) mtDNA haplotypes belonging to haplogroup B are absent in northeastern Siberia but are present in southern Siberia and Mongolia; and (iii) mtDNA diversity in haplogroups A, C, D, and E is less in the combined ethnic groups of northern and central Siberia than in the combined Amerind tribes, suggesting that the latter as a group have more time depth. While no one of these points can be regarded as conclusively established, in the sense that future studies could alter the picture, they all raise questions concerning the currently favored hypothesis that the direct ancestors of the Amerinds were primarily drawn from northern and central Siberia. Rather, these data suggest that the immediate ancestors of the earliest trans-Beringia wave(s) of migrants into the Americas (the Amerinds) have a common origin with the ethnic groups of the area now designated Mongolia or to the east (Manchuria, extreme southeastern Siberia), areas whose populations should now be extensively studied with respect to the mtDNA haplogroups and the HTLV-II virus. This hypothesis holds that the most probable route to Beringia followed by these ancestors was the area adjacent to the Siberian Pacific Coast, where they left the archaeological sites dated (not without controversy) to some 30,000 YBP (review in ref. 44). Under this hypothesis, the entry into Siberia of the ancestors of the humans now in northeastern Siberia was generally later, and those entrants were drawn from groups lacking both HTLV-II and the mtDNA haplogroup B.

With reference to timing, we point out that human entry into the New World demands both a land bridge and an ice-free corridor by which to travel south once North America was reached. Such a situation apparently existed several thousand years to either side of 13,000 YBP and of 30,000 YBP (44–46). Earlier in this treatment we presented comparable estimates for the age of the mitochondrial haplogroup C and D lineages in the Americas and Siberia. More recently,
we have improved the estimate for the Americas, to an average for the four haplogroups in Amerinds of 22,000–29,000 YBP (see ref. 47). Again, the range in the estimate reflects the uncertainties in the rate of mtDNA evolution but does not incorporate the statistical error in the estimate. This estimate, with its large but essentially indeterminate error (43), is an ambiguous result with reference to an “early” vs. a “late” arrival but favors the former. However, what is important in the present context is not the precise estimate of the time of entry of the Amerind into the Americas, but the apparently relatively younger age of the A, C, and D mtDNA lineages in the populations of Northern and Central Siberia than in Amerinds.

Torrioni et al. (30) have previously noted that the mtDNA haplogroups C and D of Siberians and Amerinds can be traced back to a common ancestral haplotype present in both Siberia and the Americas. However, this finding does not necessarily imply derivation of any of the Amerind tribes from one or more of the Siberian ethnic groups. Rather, it is at least equally possible that the root haplotypes occurred in lineages further south, in a population ancestral both to the Amerinds and some of the Siberian ethnic groups. As well as group B haplotypes, haplotypes that belong to haplogroups A, C, and D are present in southeast Asia (31, 32). We note, incidentally, that in the Siberian ethnic groups (with the exception of Eskimos) the proportion of mtDNA types not falling into haplogroups A, C, and D (and B is absent) ranges from 2% in the Evenks to 82.6% in the Udegeys (30).

Completing the DNA typings for these populations, in conjunction with expanded knowledge of the mtDNA types of southeastern Asians, might substantially clarify the provenance of the Siberian ethnic groups.

Discussion

The principal weakness of our hypothesis is the failure thus far to demonstrate the presence of representatives of the B haplogroup or unquestionable infection with HTLV-II in modern populations in the far east of Siberia, along the presumed route of migration into the New World. A second apparent weakness of the hypothesis is the seeming relative youth of the B haplogroup in Amerinds, which lends itself to consideration of a second, later and independent wave of Amerind migration into the New World (28, 48). However, the error that characterizes these time-depth estimates is so large that no great significance can be attached to the apparent youth of the B haplogroup. Furthermore, the representation of the four major mtDNA haplogroups in the tribes of the Americas, summarized in Table 1, does not support the hypothesis that the B haplogroup was introduced by a later wave of migration. Sample sizes are still relatively small, and some tribes in which a particular haplogroup has not yet been detected will undoubtedly be found to harbor members of the group when larger numbers of subjects are studied. Nevertheless, there is a commingling in the various tribes of representatives of the four basic haplogroups in which it is difficult to detect any kind of pattern, with missing haplogroups readily explained by random loss or small population samples.

An adequate test of the hypothesis we have advanced depends on the accumulation of more data with respect to the ethnic distributions of both the HTLV-II virus and the mtDNA characteristics of these same ethnic groups, especially in Siberia and the areas to the south, coupled with an expanded molecular characterization of both the virus and the mtDNA. In addition, the evidences provided by nuclear DNA must be extended. Given the current level of investigative activity in these fields, the critical data should be available within the next decade.

J.V.N. acknowledges the financial support of Department of Energy Grant 87ER05533, and R.I.S. acknowledges the support of International Science Foundation Grant RB9000 and Wenner-Gren Foundation Grant 5680.