ABSTRACT

A number of important insights into the peopling of the New World have been gained through molecular genetic studies of Siberian and Native American populations. While there is no complete agreement on the interpretation of the mitochondrial DNA (mtDNA) and Y chromosome (NRY) data from these groups, several generalizations can be made. To begin with, the primary migration of ancestral Asians expanded from south-central Siberia into the New World and gave rise to ancestral Amerindians. The initial migration seems to have occurred between 20,000–15,000 calendar years before present (cal BP), i.e., before the emergence of Clovis lithic sites (13,350–12,895 cal BP) in North America. Because an interior route through northern North America was unavailable for human passage until 12,550 cal BP, after the last glacial maximum (LGM), these ancestral groups must have used a coastal route to reach South America by 14,675 cal BP, the date of the Monte Verde site in southern Chile. The initial migration appears to have brought mtDNA haplogroups A-D and NRY haplogroups P-M45a and Q-242/Q-M3 to the New World, with these genetic lineages becoming widespread in the Americas. A second expansion that perhaps coincided with the opening of the ice-free corridor probably brought mtDNA haplogroup X and NRY haplogroups P-M45b, C-M130, and R1a1-M17 to North and Central America. Finally, populations that formerly inhabited Beringia expanded into northern North America after the LGM, and gave rise to Eskimo-Aleuts and Na-Dene Indians. Am. J. Hum. Biol. 16:420–439, 2004.

Over the past 7 years there has been a great proliferation of molecular anthropological studies of different human populations across the world. These studies have clarified various aspects of migrations into the major continental regions of the world, shed light on the emergence of linguistic and archeological traditions, and provided insights into differential male and female gene flow within local populations. Not surprisingly, many new data have also been obtained from Native American and Siberian populations. Such data have been used to address several major questions about the peopling of the New World, namely, when ancestral Native Americans first arrived in the Americas, how many population expansions or migrations were involved in this colonization process, and where in Asia/Eurasia that these ancestral groups came from. This genetic evidence has yielded new insights into the origins of Native Americans, while also raising a number of additional and intriguing questions about their prehistory.

The two genetic systems most commonly used in studies of Native American and Siberian population variation have been the mitochondrial DNA (mtDNA) and the non-recombining portion of the Y chromosome (NRY). Each of them possesses a series of different markers that help to define or identify specific genetic lineages (maternal and paternal, respectively) present in human groups. By analyzing the sequence variation in these two genomes, one can identify the genetic lineages that are present within populations and ascertain the manner in which they have been spread across geographic areas. Moreover, by measuring the sequence variation that has accumulated within them,
one can estimate their approximate ages in a particular part of the world, assuming certain assumptions about the initial population size, etc.

In this article, we explore the prehistory of the Americas from three different perspectives. The first examines the antiquity of genetic lineages in Native American groups and how it relates to the timing of the initial human entry in the New World. The second focuses on the demographic aspects of population expansions into the Americas—in particular, the way in which such expansions leave signatures in the genetic diversity of extant populations. The third considers the genetic evidence for the occurrence of multiple migrations into the New World. In the final section, we synthesize the molecular data in the context of other anthropological evidence to generate a current view of the peopling of the Americas.

GENETIC DIVERSITY OF NATIVE AMERICAN POPULATIONS

mtDNA haplogroups in the Americas

The vast majority of mtDNAs from modern Native American populations belong to primarily five different haplogroups, which have been designated A–D and X (Schurr et al., 1990; Torroni et al., 1992, 1993a, 1994a,b; Forster et al., 1996; Brown et al., 1998). Each of these is distinguished by a unique combination of coding region RFLPs and HVR-I sequence polymorphisms. Together, they comprise 95–100% of all mtDNAs in indigenous populations of the New World (Schurr, 2001, 2002, and references therein). The same pattern of variation is also observed in ancient Amerindian samples (Merriwether et al., 1994; Fox, 1996; Monsalve et al., 1996; Parr et al., 1996; Ribeiro-Dos-Santos et al., 1996; Lalueza et al., 1997; Stone and Stoneking, 1998; Carlyle et al., 2000; O’Rourke et al., 2000; Kaestle and Smith, 2001). Therefore, these five haplogroups are clearly the main founding mtDNA lineages in Native American populations.

However, a certain number of haplotypes not belonging to these five maternal lineages have been detected in different Native American groups (Bailliet et al., 1994; Easton et al., 1996; Merriwether et al., 1994, 1995; Lorenz and Smith, 1996, 1997; Ribeiro-Dos-Santos et al., 1996; Rickards et al., 1999; Santos et al., 1996; Torroni et al., 1993a; Ward et al., 1991). Most of these have been shown to belong to haplogroup X, derive from haplogroups A–D mtDNAs, or result from non-native admixture (Schurr and Wallace, 1999; Smith et al., 1999; Schurr, 2002, 2004a). The remaining haplotypes were not sufficiently analyzed to determine their haplogroup status (e.g., Bailliet et al., 1994; Ribeiro-dos-Santos et al., 1996).

Haplogroup A–D mtDNAs are observed in indigenous populations from North, Central, and South America (summarized in Schurr, 2001, 2002, and references therein). Among Amerindians, haplogroup A generally occurs at higher frequencies in North America relative to other regions, whereas haplogroups C and D generally occur at higher frequencies in South America. There does not appear to be a distinct clinal distribution for haplogroup B, but it is virtually absent from northern North America (Schurr et al., 1990; Torroni et al., 1992, 1993a, 1994a,b; Fox, 1996; Lorenz and Smith, 1996, 1997; Malhi et al., 2001, 2002). In contrast, haplogroup X is found nearly exclusively in North America (Torroni et al., 1992, 1993a; Scozzari et al., 1997; Brown et al., 1998; Smith et al., 1999; Malhi et al., 2001, 2002; Bolnick et al., 2003). These distributions likely reflect both the original pattern of settlement of the Americas and the subsequent genetic differentiation of Native American populations within these continental areas.

Haplogroup A–D mtDNAs have also been detected in populations representing the three Native American linguistic groups (Amerind, Na-Dené, Eskimo-Aleut) proposed by Greenberg (1987). Although haplogroups A–D usually appear together in Amerindian populations, many tribes lack haplotypes from at least one of these haplogroups (Torroni et al., 1992, 1993a, 1994a,b; Batista et al., 1995; Kolman et al., 1995; Easton et al., 1996; Lorenz and Smith, 1996, 1997; Kolman and Bermingham, 1997; Scozzari et al., 1997; Rickards et al., 1999; Moraga et al., 2000; Bert et al., 2001). This pattern likely reflects the extent to which genetic drift and founder events have influenced the distribution of mtDNA haplotypes in New World populations.

However, ancestral populations for the Na-Dené Indians and Eskimo-Aleuts may not have possessed all four of these haplogroups. These populations show different haplogroup profiles than Amerindians, which consist largely of haplogroup A and D mtDNAs. In
addition, they essentially lack haplogroup B and have very low frequencies of haplogroup C. Moreover, none of them have haplogroup X mtDNAs (Torroni et al., 1992, 1993a; Shields et al., 1993; Ward et al., 1993; Starikovskaya et al., 1998; Saillard et al., 2000; Rubicz et al., 2003). Thus, circumboreal groups appear to have experienced different population histories than Amerindians.

**NRY haplogroups in the Americas**

In characterizing NRY variation in Native Americans, researchers have employed a number of different single nucleotide (SNP) and short tandem repeat (STR) loci to define the paternal lineages present within them (Pena et al., 1995; Underhill et al., 1996, 1999, 2000; Bianchi et al., 1997, 1998; Hammer et al., 1997; Karafet et al., 1997, 1999; Lell et al., 1997, 2002). However, these research groups have not used the same combination of genetic markers in their studies, leading to alternative and sometimes confusing nomenclatures for NRY haplotypes and haplogroups. A recent synthesis of these data has resulted in a consensus nomenclature based on known single nucleotide polymorphisms (SNPs) (Y Chromosome Consortium 2002), and it will be used in this article. This system identifies a NRY haplogroup by a letter and the SNP that defines it (e.g., G-M201).

Native American NRY haplotypes derive from subsamples of the haplogroups present in Siberia. These include haplogroups Q-M3, R1a1-M17, P-M45, F-M89, and C-M130. Two of them, Q-M3 and P-M45, represent the majority of Native American Y-chromosomes. Q-M3 haplotypes appear at significant frequencies in most Native American populations and are distributed in an increasing north-to-south cline within the New World (Bianchi et al., 1996, 1998; Underhill et al., 1996; Karafet et al., 1997, 1999; Lell et al., 1997, 2002; Santos et al., 1999). The STR data from Q-M3 haplotypes also reveal significant differences in haplotype distributions between North/Central and South American populations, suggesting different population histories in the two major continental regions (Bianchi et al., 1997, 1998; Karafet et al., 1997, 1999; Lell et al., 1997, 2002; Ruiz-Linares et al., 1999; Santos et al., 1999).

Haplogroup P-M45 is also widely distributed among Native American populations and represents ~30% of their NRY haplotypes (Ruiz-Linares et al., 1999; Santos et al., 1999; Lell et al., 2002; Bortolini et al., 2003). In addition, phylogenetic analysis has revealed two distinct sets of P-M45 haplotypes in Native American populations. The first of these (M45a) is more broadly distributed in populations from North, Central, and South America, whereas the second (M45b) appears in only North and Central American groups (Lell et al., 2002).

Most of the remaining NRY haplotypes belong to one of several different haplogroups, and comprise only 5% of Native American Y chromosomes. In general, these haplotypes have limited distributions in the New World. For example, C-M130 haplotypes have only been detected in the Na-Dene-speaking Tanana, Navajo and Chipewyan, and the Amerindian Cheyenne (Bergen et al., 1999; Karafet et al., 1999; Lell et al., 2002; Bortolini et al., 2003). In addition, R1a1-M17 haplotypes have only been observed in the Guaymi (Ngöbe), a Chibchan-speaking tribe from Costa Rica (Lell et al., 2002). Neither of these haplogroups has been detected in South American Indian populations.

**TIMING OF THE INITIAL COLONIZATION OF THE AMERICAS**

Having defined the genetic lineages present in Native American populations, we can now turn to the question of when ancestral populations first entered the New World. Considerable attention has been paid to this question for several reasons. First, the Americas were the last continental regions to be settled by modern humans, who entered areas previously uninhabited by people. For this reason, genetic studies of Native American groups will illuminate the process and rate of human migration and adaptation to new environments. Second, there is great interest in knowing how much time was required to generate the biocultural diversity observed in contemporary Native American populations. Assigning a temporal starting point for the accumulation of this diversity will provide a calibration point for estimating into how fast languages and cultures can change, particularly if a single primary migration gave rise to all Amerindian populations. Third, estimates for the timing of the initial migration to the New World based on genetic studies have ranged from 30,000–14,000 calendar years before present (cal BP). However, this range encompasses the Last Glacial Maximum (LGM) (~24,000–13,050 cal BP),
a period of time when human movement into the New World through an interior route was not possible because of glacial barriers. Therefore, reconciling the genetic data with archeological and geological evidence will be crucial for determining whether the First Americans arrived before or after the LGM, and whether they followed an interior or coastal route into the New World during this time period.

**Antiquity of mtDNA haplogroups in Siberia and the Americas**

While there is little controversy about the number and type of founding haplogroups in the New World, the ages of these maternal lineages continues to be contested. Early studies of RFLP variation in Native American populations produced time depths for haplogroups A, C, D, and X of between 35,000–20,000 cal BP (Torroni et al., 1992, 1993a, 1994a). These estimates were viewed as reflecting the genetic diversity that had accumulated in the American branches of these mtDNA lineages, hence, the time at which modern humans first entered the Americas. Additional support for these findings came from the fact that Native American and Siberian populations did not share any specific haplotypes (Torroni et al., 1993b; Starikovskaya et al., 1998; Schurr et al., 1999). By contrast, the age for haplogroup B in the New World was estimated at 17,000–13,000 cal BP, suggesting that it was brought to the Americas in a later and separate migration from the earlier one that brought the other four lineages. The age of haplogroup X based on RFLP haplotype data was identical to that of haplogroup B, although it increased in age when estimated from HVS-I sequence data (Brown et al., 1998).

Subsequent analyses of HVS-I sequence variation in Native Americans provided somewhat different perspectives on the antiquity of these haplogroups. Several of them showed that haplogroups A, B, and C had roughly the same extent of genetic diversification in North America, and that haplogroup B could possibly have been present in the New World by 30,000–20,000 cal BP (Lorenz and Smith, 1997; Bonatto and Salzano, 1997; Stone and Stoneking, 1998). The older date also implied that haplogroup B arrived in the Americas around the same time that haplogroups A, C, D, and X did. In fact, most HVS-I studies have provided ages for the four major founding haplogroups that range between 35,000–15,000 cal BP, with the earliest dates being 14,000–12,000 cal BP (Shields et al., 1993). A recent analysis of coding region sequence variation in Native American populations has also generated average dates for haplogroups A–D of between 20,000–15,000 cal BP (Silva et al., 2002a). Thus, most molecular studies favor an entry time for these mtDNA lineages that is somewhat earlier than the dates associated with the Clovis lithic culture in North America.

Two other important issues about haplogroup ages have arisen in this debate. The first centers on the question whether the older haplogroup age estimates actually reveal the timing of human expansion(s) into the New World. Because the genetic divergence or coalescence times of genetic lineages do not necessarily correspond to the timing of population splits, it has been suggested that the older dates may reflect the emergence of these mtDNA lineages in Asia rather than their entry into the Americas (e.g., Shields et al., 1993). On the other hand, only the founding RFLP haplotypes for haplogroup A–D have been shown to be present in both Siberia and the Americas (Torroni et al., 1992, 1993a; Schurr et al., 1999). These data suggested that the temporal split between the ancestral Amerindian population and its Asian precursor mirrored the split in the branches of each respective haplogroup. This issue will be reexamined below when considering demographic aspects of New World colonization.

The second issue is the number of founding haplotypes that were brought with each founding haplogroup. The number of founders present in a genetic lineage will affect estimates of its age because a certain amount of the diversity present in that lineage will have accumulated from each founding type. If more than one founding haplotype was present within a haplogroup, then the age of this mtDNA lineage would be inflated due to the fact that the diversity of haplotypes that accumulated from each founding haplotype was not taken into account in the coalescence estimates. For Native Americans, the presence of multiple founding haplotypes would imply that the ages of haplogroups A–D would be less than 30,000–20,000 cal BP; hence, the colonization date of the Americas would be more consistent with a late entry (Clovis-first) migration model.

As noted above, there appears to be only one founding RFLP haplotype each for
haplogroups A–D and X (Torroni et al., 1992, 1993a, 1994a,b; Brown et al., 1998; Schurr et al., 1999). These founder haplotypes are the most widely distributed mtDNAs in the Americas, and central to the diversification of their respective haplogroups. However, other investigators have suggested that more than one founding haplotypes from haplogroups A–D were among the original set of founding Native American mtDNAs (Bailliet et al., 1994; Merriwether et al., 1994, 1995; Easton et al., 1996; Santos et al., 1996; Rickards et al., 1999). Unfortunately, none of these studies have provided additional RFLP or HVR-I sequence data to demonstrate that these are actually the same founding haplotypes defined in other studies (Schurr and Wallace, 1999; Schurr, 2004b).

At the same time, Malhi et al. (2002) argue that there could possibly be more than one founder HVS-I sequence for at least some of these haplogroups, due to these sequences being identical to ones present in Asian and Siberian populations. However, they also recognize the difficulty in delineating ancestral sequences from derivative forms that have lost or gained key polymorphisms that distinguish American from Asian sequence motifs, due to the recurrent mutations that typically occur in the mtDNA control region (Gurven, 2000; Stoneking, 2000). Thus, additional coding region data must be obtained from these mtDNAs to confirm their status as founder haplotypes.

**Antiquity of NRY haplogroups in Siberia and the Americas**

The methods for dating NRY haplogroups have employed both the SNPs that define them and the STR loci that occur on each Y chromosome. Because SNPs are rare, if not unique, evolutionary events, it is difficult to estimate when they evolved in a particular paternal lineage using only this kind of data. To get around this problem, Underhill et al. (2001) used an average mutation rate estimated from SNP variation in three NRY genes (Thomson et al., 2000) to date the various branches (haplogroups) of their phylogeny. This estimate of $1.24 \times 10^{-9}$ produced an age for the major expansion of modern humans out of Africa of $\sim 59,000$ cal BP. Using this date for the Most Recent Common Ancestor (MRCA) of their SNP phylogeny, Underhill et al. (2000) estimated an average SNP evolution rate of one per every 6,900 years. With this rate, it is possible to tentatively date the origins of the major branches of the phylogeny, as well as other points of SNP haplotype diversification.

An alternative strategy for dating the ages of NRY haplogroups is to analyze variation in the faster evolving STR loci that co-occur on each SNP haplotype. In this case, the extent of allelic diversity of a set of STR loci are measured and averaged over all loci, with the average then being multiplied by an STR mutation rate to determine the actual age of the NRY lineage. Recent mutation rates have been estimated across multiple generations of males (meiotic transmissions) in human families. Although these rates vary somewhat depending on the type of STR used for the estimates (di-, tri-, tetra-), most studies have found that the average mutation rate of NRY STRs is around $2.80 \times 10^{-3}$ per generation (Heyer et al., 1997; Kayser et al., 1997, 2000; Bianchi et al., 1998; Thomson et al., 2000). Considerable effort has been made to estimate the age of Q-M3 haplotypes, given that they appear to signal the initial entry of ancestral populations into the New World. Using the SNP mutation rate of Underhill et al. (2000), one obtains an age for haplogroup Q-M3 of $\sim 13,800$ cal BP (Schurr, 2004a). The estimates made with STR mutation rates have ranged from 30,000–7,600 cal BP (Underhill et al., 1996; Bianchi et al., 1998; Hammer et al., 1998; Karafet et al., 1999; Forster et al., 2000). Together, these analyses of NRY variation in Native American populations do not clearly point to an early or late entry of the Q-M3 lineage into the New World, but tend to favor the latter.

The P-M45 lineage is considerably older than the Q-M3 lineage, which derives from it. Using the SNP mutation rate of Underhill et al. (2000), P-M45 haplotypes are estimated to be at least 30,000 years old. This degree of antiquity is also reflected by their widespread distribution in Siberia and Eurasia (Underhill et al., 2000; Lell et al., 2002). In addition, the P-M45 lineage has been present in Siberia long enough to diversify into different sub-haplogroups. This is shown by the presence of two different types of NRY haplotypes in Native Americans, a central Siberian set (P-M45a) that is shared with all Native American populations, and an eastern Siberian set (P-M45b) that appears only in Native Americans from North and Central America (Lell et al., 2002).
The ages of several other NRY lineages present in Siberia and the Americas have also been estimated. One of the older lineages in Siberia, K-M9, has been dated at >50,000 cal BP (Karafet et al., 1999; Underhill et al., 2000). The antiquity of the K-M9 lineage is consistent with the presence of this SNP in a sizeable majority of Siberian Y chromosomes (Karafet et al., 1999; Santos et al., 1999; Underhill et al., 2000; Lell et al., 2002). The oldest SNP in the Eurasian branch of the NRY phylogeny, F-89, dates to ~62,000 cal BP (Schurr, 2004a), and predates the occurrence of the K-M9 lineage, since it appears in all haplotypes bearing the latter mutation. F-89 is an important SNP because it marks the initial diversification and spread of non-African NRY lineages into the rest of the world.

By contrast, the C-M130 lineage is somewhat younger than the F-M89 or K-M9 lineages, having been dated at ~30,000–25,000 cal BP (Karafet et al., 1999; Underhill et al., 2000). Its age is generally consistent with its broad distribution in East and Southeast Asia, in which it appears to have originated, and its haplotypic diversity in eastern Siberian and Asian populations (Su et al., 1999; Lell et al., 2002).

The estimated age of haplogroup R1a1-M17 is rather intriguing. Using the SNP evolution rate of Underhill et al. (2000), an age of 13,800 cal BP was obtained for this lineage, one that falls into the very end of the LGM (Schurr, 2004a). These haplotypes constitute a distinct branch within R1a, and are not especially common in Siberian populations, although occurring across a broad geographic area (Lell et al., 2002). The data suggest that R1a1-M17 haplotypes did not emerge in Siberia until after the Americas had already been colonized and were brought to the New World through a secondary expansion of ancient Asian populations, along with C-M130 and P-M45b haplotypes (Lell et al., 2002).

The most recent efforts to date the NRY haplotypes present in Native American populations have utilized a newly identified SNP, Q-M242, to make this estimate (Bortolini et al., 2003; Seielstad et al., 2003). The Q-M242 marker occurred within haplogroup P-M45 in Central Asia prior to the emergence of the Q-M3 SNP and the expansion of its haplotypes in the Americas. As such, the M242 marker may better define the initial entry into the Americas than M3. Using STR mutation rates, researchers have dated the age of M242 haplotypes in the New World at ~18,000–15,000 cal BP (Bortolini et al., 2003; Seielstad et al., 2003).

DEMOGRAPHY OF NATIVE AMERICAN POPULATIONS

The manner in which a population enters and establishes itself in a geographic area should be reflected in the pattern of variation in its genes. A sudden population expansion shows certain demographic features that distinguish it from a more uniform pattern of growth and dispersal. In the former case, a founding population rapidly grows from small numbers until reaching some plateau. Concomitantly, the extent of genetic diversity also increases substantially because of this rapid spread of human groups and their genetic divergence from each other because of being geographically isolated from one another. Such diversification appears as a multiplication of the branches and subbranches in a gene genealogy, with the major bifurcation points or nodes corresponding to the initial phase of expansion.

If the resulting haplotypes are compared for pairwise mutational differences, one can generate a frequency plot of these differences and determine the parameters of this expansion (Rogers and Harpending, 1992). The resulting curve should show the speed and nature of the expansion, based on the initial slope and shape of the curve, respectively. A steep upward trajectory of the curve suggests the rapid expansion of the population, whereas a more modest slope indicates a somewhat slower growth process. If the curve is unimodal and smooth, then the growth or expansion probably occurred as a single event. However, if it is bimodal or ragged (multiple peaks), then two or more expansions may have contributed to the genetic make-up of that population. Alternatively, the population may have undergone fluctuations in size because of various stochastic processes, such as population bottlenecks. Thus, we can distinguish several basic parameters of populations having undergone different demographic processes through this kind of sequence analysis.

Such signatures of population expansion have been examined through the analysis of mismatch and crossmatch distributions in mtDNA sequence data (Harpending and Rogers, 1992; Harpending et al., 1993, 1998; Sherry et al., 1994). Mismatch distributions
refer to pairwise comparisons between all members in a single population (or haplogroup), while crossmatch distributions refer to all pairwise comparisons between members of two separate populations (or haplogroups). Given these empirical distributions and the locus-specific generational mutation rate, \( \mu \), it is possible to estimate when an episode of population expansion began or when two populations diverged, both events being measured in units of mutational time.

In mtDNA studies, mismatch distributions can be estimated from either HR-RFLP haplotype or HVR-I sequence data, or even whole genome sequences. Ordinarily, when considering DNA sequence data, mismatch and crossmatch distributions are simply tabulations of the number of mutational differences between all pairs of individuals in a sample, and the estimation of the per-site mutation rate, \( \mu \), is simply the product of \( \mu \) and the number of nucleotides sequenced. However, when HR-RFLP data are being analyzed, a unit conversion has to be performed because haplotype scores (0 and 1 for the absence or presence of a RFLP, respectively) are not equivalent to single nucleotide substitutions, due to the fact that each restriction enzyme surveys variation across 4–6 nucleotides within each recognition site (Table 1). However, once converted into sequence format, the haplotype data can be analyzed with the mismatch and crossmatch techniques as though they were HVS-I sequences.

Using this analytical approach, we investigated whether all of the founding haplogroups in Native Americans showed the same pattern of haplotypic diversity. If these mtDNA lineages were brought to the Americas at the same time through a single expansion event, then the mismatch and crossmatch distributions should look essentially identical, i.e., the same shape of the curve, and the same average number of mutation differences between haplotypes or HVS-I sequences. By contrast, if one or more of them arrived at different times through independent expansions, then the distributions should vary between the mtDNA lineages. Furthermore, if any of the haplogroups were brought to the Americas more than once, then the mismatch distributions for those haplogroups could be bimodal or ragged, depending on how much time passed between the expansions. Thus, we are able to make some predictions as to what the mismatch profile of these mtDNA lineages might look like under different demographic or migration scenarios.

Two different kinds of mismatch analyses were performed with the HR-RFLP haplotype data from haplogroups A–D and X (Torroni et al., 1992, 1993a, 1994a,b). In the first analysis (Analysis I), the entire HR-RFLP dataset was analyzed, with expansion times for the haplogroups and divergence times between the haplogroups being computed. In the second (Analysis II), the dataset was analyzed after removing the RFLPs that define each haplogroup, as these sites are known to pre-date the entry of ancestral Native American populations into the New World. Comparison of the results of these two analyses would, therefore, provide approximate dates for the expansion of these mtDNA lineages in Asia/Siberia (Analysis I) and their expansions in the New World (Analysis II).

The results of Analysis I showed that haplogroups C and D were the first haplogroups

### Table 1. Unit conversion of RFLP haplotype data into sequence data

<table>
<thead>
<tr>
<th>RFLP study</th>
<th>Haplotype estimate</th>
<th>(2\mu) estimate for DNA sequence data*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cann et al. 1987</td>
<td>( \mu )</td>
<td>0.126</td>
</tr>
<tr>
<td>V = 236, ( r = 4 ), ( T = 370 )</td>
<td>( 2\mu )</td>
<td>0.252</td>
</tr>
<tr>
<td>Torroni et al. 1993a, 1994a,b</td>
<td>( \mu )</td>
<td>0.247</td>
</tr>
<tr>
<td>V = 149, ( r = 4 ), ( T = 459 )</td>
<td>( 2\mu )</td>
<td>0.494</td>
</tr>
</tbody>
</table>

*The \(2\mu\) value represents the nucleotide divergence rate. The data for each population consisted of haplotype frequencies and designations (AM01-AM06) for 149 polymorphic sites (Torroni et al. 1993a, 1994a,b). All of the polymorphisms included in these analyses were presumed to be the product of simple point mutations. As a result, the Region V 9-bp deletion, which is frequently used as a marker of Asian ancestry and almost completely defines haplogroup B, was excluded from analysis. This exclusion is justified by the fact that Region V deletions arise by a different mutational process than that generating RFLPs, one that has not been incorporated into the analytical theory underlying this method.
Two-parameter method of moments developed by Rogers (1994) was used to estimate the time of population growth, distributions, which define the distribution and relative divergence of a pair of samples, were constructed by comparing each haplotype and tallying the number of differences in a vector indexed by i number of differences, where the index value i

\( t\)

was

\( \theta \)

for these groups were quite variable (Tables 3

\( q\)

The mismatch distributions were constructed by comparing each haplotype against every other one, for a total of n (n − 1)/2 comparisons, and tallying the number of differences in a vector indexed by i number of differences, where the index value i = 0, 1, 2, … . Crossmatch distributions, which define the distribution and relative divergence of a pair of samples, were constructed by comparing each haplotype in one sample against every other haplotype, for a total of n comparisons, in the same manner as mismatch distributions. The two-parameter method of moments developed by Rogers (1994) was used to estimate the time of population growth, \( \tau \)

\( \theta \)

were

\( q\)

earlier expansion times for haplogroups C and D (~25,000 cal BP), with haplogroup X falling in between them (~20,000 cal BP). The Analysis I expansion times were proportionately similar, indicating that the inclusion of haplogroup-defining sites did not greatly bias the mismatch calculations towards older expansion times. It should also be pointed out that the haplogroup A expansion time estimate in this analysis was based solely on Native American HR-RFLP haplotypes, not those from the Chukchi and Siberian Eskimos, as was the case for a more recent divergence time estimate for this mtDNA lineage (Starikovskaya et al., 1998; Schurr et al., 1999).

To determine the consistency of the patterns obtained for these mtDNA lineages, an additional set of mismatch and crossmatch analyses were carried out with the HR-RFLP haplotype data. In this case, the HR-RFLP haplotypes were partitioned into the 24 Native American populations in which they were first characterized before being analyzed. In general, the mismatch distributions for these groups were quite variable (Tables 3 and 4, Fig. 2). Some populations appeared to

### Table 2. Parameter estimates for Amerindian mtDNA haplogroups

<table>
<thead>
<tr>
<th>Haplogroup</th>
<th>N</th>
<th>Analysis I (full data set)</th>
<th>Analysis II (haplogroup-defining sites removed)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>( \tau \pm SE \left( \frac{\tau}{C6} \right) )</td>
<td>( \theta )</td>
</tr>
<tr>
<td>A</td>
<td>158</td>
<td>1.94 ± 0.65 (0.74)</td>
<td>17,600</td>
</tr>
<tr>
<td>B</td>
<td>93</td>
<td>0.33 ± 0.40 (0.82)</td>
<td>3,000</td>
</tr>
<tr>
<td>C</td>
<td>61</td>
<td>3.58 ± 0.60 (0.00)</td>
<td>32,400</td>
</tr>
<tr>
<td>D</td>
<td>59</td>
<td>2.13 ± 0.48 (0.00)</td>
<td>19,300</td>
</tr>
<tr>
<td>X</td>
<td>8</td>
<td>2.11 ± 0.45 (0.00)</td>
<td>19,100</td>
</tr>
<tr>
<td>TOTAL</td>
<td>377</td>
<td>3.47 ± 0.80 (1.14)</td>
<td>31,400</td>
</tr>
</tbody>
</table>

Crossmatch analysis (population divergence)

| A × B     | 249        | 3.72 ± 2.00 (0.00)          | 33,600                                       | 1.04 ± 0.79 (0.85) | 9,800  |
| A × C     | 217        | 7.24 ± 3.10 (0.00)          | 65,600                                       | 2.89 ± 0.69 (0.00) | 27,200 |
| A × D     | 215        | 6.53 ± 3.42 (0.00)          | 59,000                                       | 2.18 ± 0.66 (0.00) | 20,600 |
| A × X     | 164        | 5.00 ± 2.93 (0.00)          | 45,400                                       | 2.18 ± 0.66 (0.00) | 21,800 |
| B × C     | 154        | 6.03 ± 2.61 (0.00)          | 54,800                                       | 2.18 ± 0.66 (0.00) | 16,400 |
| B × D     | 152        | 5.45 ± 3.16 (0.00)          | 49,400                                       | 2.18 ± 0.66 (0.00) | 26,200 |
| B × X     | 101        | 2.78 ± 1.24 (0.00)          | 25,200                                       | 2.18 ± 0.66 (0.00) | 24,200 |
| C × D     | 120        | 5.04 ± 2.19 (0.00)          | 45,600                                       | 2.18 ± 0.66 (0.00) | 33,600 |
| C × X     | 69         | 7.28 ± 3.38 (0.00)          | 65,800                                       | 2.18 ± 0.66 (0.00) | 25,400 |
| D × X     | 67         | 6.40 ± 3.31 (0.00)          | 58,000                                       | 2.18 ± 0.66 (0.00) | 25,400 |

All mismatch distributions were constructed by comparing each haplotype against every other one, for a total of n (n − 1)/2 comparisons, and tallying the number of differences in a vector indexed by i number of differences, where the index value i = 0, 1, 2, … . Crossmatch distributions, which define the distribution and relative divergence of a pair of samples, were constructed by comparing each haplotype in one sample against every other haplotype, for a total of n comparisons, in the same manner as mismatch distributions. The two-parameter method of moments developed by Rogers (1994) was used to estimate the time of population growth, \( \tau \)

\( \theta \)

were

\( q\)

earlier expansion times for haplogroups C and D (~25,000 cal BP), with haplogroup X falling in between them (~20,000 cal BP). The Analysis I expansion times were proportionately similar, indicating that the inclusion of haplogroup-defining sites did not greatly bias the mismatch calculations towards older expansion times. It should also be pointed out that the haplogroup A expansion time estimate in this analysis was based solely on Native American HR-RFLP haplotypes, not those from the Chukchi and Siberian Eskimos, as was the case for a more recent divergence time estimate for this mtDNA lineage (Starikovskaya et al., 1998; Schurr et al., 1999).

To determine the consistency of the patterns obtained for these mtDNA lineages, an additional set of mismatch and crossmatch analyses were carried out with the HR-RFLP haplotype data. In this case, the HR-RFLP haplotypes were partitioned into the 24 Native American populations in which they were first characterized before being analyzed. In general, the mismatch distributions for these groups were quite variable (Tables 3 and 4, Fig. 2). Some populations appeared to
Fig. 1. Pairwise mismatch analysis of Native American haplogroups. Each figure portrays a possible population history of Amerindian haplogroups. The overall structure provides haplogroup fissioning times (crossmatch data), while the shaded overlays indicate periods of population expansion (estimated from $\tau \pm 2\text{SE}$ values (Table 2). For Analysis II, the RFLPs defining each haplogroup were removed from the haplotypes being analyzed, since they were common to both Asian and Native American mtDNAs belonging to these lineages. This step allowed an estimate of the times at which new haplotypes within each haplogroup began to accumulate diversity or expand in the New World. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]
<table>
<thead>
<tr>
<th>Population</th>
<th>F_0</th>
<th>F_1</th>
<th>F_2</th>
<th>F_3</th>
<th>F_4</th>
<th>F_5</th>
<th>F_6</th>
<th>F_7</th>
<th>F_8</th>
<th>F_9</th>
<th>F_10</th>
<th>F_11</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dogrib</td>
<td>.448</td>
<td>.460</td>
<td>.092</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ojibwa</td>
<td>.161</td>
<td>.023</td>
<td>.013</td>
<td>.018</td>
<td>.008</td>
<td>.098</td>
<td>.230</td>
<td>.230</td>
<td>.071</td>
<td>.077</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Navajo</td>
<td>.175</td>
<td>.244</td>
<td>.173</td>
<td>.198</td>
<td>.086</td>
<td>.042</td>
<td>.056</td>
<td>.026</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Pima</td>
<td>.064</td>
<td>.110</td>
<td>.117</td>
<td>.126</td>
<td>.067</td>
<td>.113</td>
<td>.205</td>
<td>.166</td>
<td>.032</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Maya</td>
<td>.063</td>
<td>.122</td>
<td>.140</td>
<td>.143</td>
<td>.111</td>
<td>.222</td>
<td>.103</td>
<td>.032</td>
<td>.048</td>
<td>.008</td>
<td>.008</td>
<td>0</td>
</tr>
<tr>
<td>Teribe/Naso</td>
<td>.663</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>.257</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Guatuso</td>
<td>.637</td>
<td>0</td>
<td>0</td>
<td>.011</td>
<td>.263</td>
<td>.089</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Boruca</td>
<td>.505</td>
<td>0</td>
<td>.022</td>
<td>.110</td>
<td>.220</td>
<td>0</td>
<td>.110</td>
<td>.022</td>
<td>0</td>
<td>.011</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Briobri/Cabecar</td>
<td>.229</td>
<td>0</td>
<td>.254</td>
<td>0</td>
<td>.518</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Guaymi/Ngobe</td>
<td>.183</td>
<td>.267</td>
<td>.109</td>
<td>.192</td>
<td>.208</td>
<td>.042</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Kuna (San Blas)</td>
<td>.875</td>
<td>0</td>
<td>0</td>
<td>.125</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Yanomama</td>
<td>.156</td>
<td>.022</td>
<td>.043</td>
<td>.029</td>
<td>.355</td>
<td>.181</td>
<td>.087</td>
<td>.112</td>
<td>.014</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Makiritare</td>
<td>.222</td>
<td>0</td>
<td>.022</td>
<td>.111</td>
<td>.111</td>
<td>.133</td>
<td>.245</td>
<td>.089</td>
<td>.022</td>
<td>.045</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Wapishana</td>
<td>.364</td>
<td>0</td>
<td>0</td>
<td>.106</td>
<td>.152</td>
<td>.060</td>
<td>.318</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Makushi</td>
<td>.022</td>
<td>0</td>
<td>.067</td>
<td>.067</td>
<td>.067</td>
<td>.177</td>
<td>.244</td>
<td>.222</td>
<td>.067</td>
<td>.067</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Piaroa</td>
<td>.267</td>
<td>0</td>
<td>.400</td>
<td>0</td>
<td>.200</td>
<td>0</td>
<td>.133</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ticuna</td>
<td>.066</td>
<td>.048</td>
<td>.063</td>
<td>.188</td>
<td>.098</td>
<td>.119</td>
<td>.225</td>
<td>.153</td>
<td>.040</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Kraho</td>
<td>.132</td>
<td>0</td>
<td>.077</td>
<td>.176</td>
<td>.220</td>
<td>.186</td>
<td>.132</td>
<td>.077</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Marubo</td>
<td>.200</td>
<td>0</td>
<td>0</td>
<td>.155</td>
<td>.044</td>
<td>.156</td>
<td>.267</td>
<td>.156</td>
<td>.022</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Mataco/Wichi</td>
<td>.404</td>
<td>0</td>
<td>.080</td>
<td>0</td>
<td>.044</td>
<td>.516</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Note: Histogram values were normalized to unit area to permit comparisons between distributions drawn from unequal sample sizes.
have very recently expanded in the New World, such as some Chibchan-speaking groups of Central America and the Canadian Dogrib, while others seemed to have done so very early, such as the Ticuna and Yanomami of South America. A number of the populations also showed a bimodal distribution of pairwise sequences, suggesting they could have experienced a population bottleneck or reduction after the Americas had been colonized, or perhaps as a consequence of contact with Europeans, beginning in the 16th century.

The overall expansion time for this set of populations using an Analysis I approach was 31,454 cal BP, a date consistent with early divergence times for the mtDNA haplogroups in the Americas. However, when haplogroup-specific RFLPs were excluded from the mismatch estimates (Analysis II), the expansion time dropped by half, to 16,275 cal BP (Fig. 3). This date is concordant with similar estimates for the five haplogroups when the same RFLP sites were excluded. These results point to expansion times for Native American populations that are somewhat more recent than those inferred from previous estimates for haplogroup divergence based on the same data sets (e.g., Torroni et al., 1993a, 1994b).

While these results are intriguing, one must also consider the effects of sampling on these mismatch expansion estimates. For these analyses, relatively small sample sizes from each Native American population (10–24 individuals) were analyzed. These numbers are clearly not representative of the overall sizes of the populations being analyzed, nor are these estimates fully reflective of the genetic diversity present within these groups. In addition, the vast majority of Native American mtDNAs examined for variation has not been analyzed by the HR-RFLP method and, hence, is not available for comparable mismatch analyses. These facts, along with the apparent sensitivity of the mismatch method to demographic effects occurring in human populations (Marjoram and Donnelly, 1994), suggests that the shallowness of the expansion times for both haplogroups and populations could possibly underestimate their actual expansions times.

Similar mismatch studies have been carried out with HVS-I sequence data from Native Americans. In their analysis of ancient and modern Native American samples, Stone

---

**TABLE 4. Amerindian population expansion times estimated from RFLP haplotype data**

<table>
<thead>
<tr>
<th>Population</th>
<th>N</th>
<th>Mean</th>
<th>Variance</th>
<th>$\tau$</th>
<th>$\theta$</th>
<th>&quot;Expansion Time&quot;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dogrib</td>
<td>30</td>
<td>0.62</td>
<td>0.41</td>
<td>0.62</td>
<td>0.00</td>
<td>5,643</td>
</tr>
<tr>
<td>Ojibwa</td>
<td>28</td>
<td>5.17</td>
<td>8.49</td>
<td>3.35</td>
<td>1.82</td>
<td>30,365</td>
</tr>
<tr>
<td>Navajo</td>
<td>48</td>
<td>2.21</td>
<td>3.39</td>
<td>1.13</td>
<td>1.08</td>
<td>10,223</td>
</tr>
<tr>
<td>Pima</td>
<td>30</td>
<td>4.06</td>
<td>5.90</td>
<td>2.70</td>
<td>1.36</td>
<td>24,496</td>
</tr>
<tr>
<td>Mixe</td>
<td>16</td>
<td>4.20</td>
<td>5.68</td>
<td>2.99</td>
<td>1.21</td>
<td>27,108</td>
</tr>
<tr>
<td>Mixtec-Alta</td>
<td>15</td>
<td>3.64</td>
<td>7.62</td>
<td>1.65</td>
<td>2.00</td>
<td>14,959</td>
</tr>
<tr>
<td>Mixtec-Baja</td>
<td>14</td>
<td>2.39</td>
<td>1.60</td>
<td>2.39</td>
<td>0.00</td>
<td>21,656</td>
</tr>
<tr>
<td>Zapotec</td>
<td>16</td>
<td>4.50</td>
<td>5.94</td>
<td>3.30</td>
<td>1.20</td>
<td>29,905</td>
</tr>
<tr>
<td>Maya</td>
<td>27</td>
<td>3.62</td>
<td>5.23</td>
<td>2.36</td>
<td>1.27</td>
<td>21,371</td>
</tr>
<tr>
<td>Teribe</td>
<td>20</td>
<td>1.60</td>
<td>5.44</td>
<td>-0.36</td>
<td>1.96</td>
<td>-3,261</td>
</tr>
<tr>
<td>Guatuso</td>
<td>20</td>
<td>1.46</td>
<td>4.10</td>
<td>-0.17</td>
<td>1.63</td>
<td>-1,548</td>
</tr>
<tr>
<td>Boruca</td>
<td>14</td>
<td>2.01</td>
<td>5.72</td>
<td>0.08</td>
<td>1.93</td>
<td>753</td>
</tr>
<tr>
<td>Bribri/Cabecar</td>
<td>24</td>
<td>2.47</td>
<td>2.80</td>
<td>1.90</td>
<td>0.58</td>
<td>17,192</td>
</tr>
<tr>
<td>Guaymi</td>
<td>16</td>
<td>1.97</td>
<td>2.50</td>
<td>1.24</td>
<td>0.73</td>
<td>11,251</td>
</tr>
<tr>
<td>Kuna</td>
<td>16</td>
<td>0.35</td>
<td>0.93</td>
<td>-0.41</td>
<td>0.76</td>
<td>-3,716</td>
</tr>
<tr>
<td>Yanomami</td>
<td>24</td>
<td>3.78</td>
<td>5.09</td>
<td>2.64</td>
<td>1.14</td>
<td>23,298</td>
</tr>
<tr>
<td>Makiritare</td>
<td>10</td>
<td>3.74</td>
<td>7.95</td>
<td>1.69</td>
<td>2.05</td>
<td>15,306</td>
</tr>
<tr>
<td>Wapishana</td>
<td>12</td>
<td>2.88</td>
<td>6.72</td>
<td>0.91</td>
<td>1.96</td>
<td>8,290</td>
</tr>
<tr>
<td>Macushi</td>
<td>10</td>
<td>5.08</td>
<td>6.35</td>
<td>3.95</td>
<td>1.13</td>
<td>35,838</td>
</tr>
<tr>
<td>Paroa</td>
<td>10</td>
<td>2.00</td>
<td>4.00</td>
<td>0.59</td>
<td>1.41</td>
<td>5,313</td>
</tr>
<tr>
<td>Ticuna</td>
<td>28</td>
<td>4.31</td>
<td>5.26</td>
<td>3.33</td>
<td>0.98</td>
<td>30,180</td>
</tr>
<tr>
<td>Kraho</td>
<td>14</td>
<td>3.55</td>
<td>4.66</td>
<td>2.50</td>
<td>1.05</td>
<td>22,674</td>
</tr>
<tr>
<td>Marubo</td>
<td>10</td>
<td>3.86</td>
<td>7.28</td>
<td>2.01</td>
<td>1.85</td>
<td>18,235</td>
</tr>
<tr>
<td>Mataco</td>
<td>28</td>
<td>2.64</td>
<td>5.77</td>
<td>0.87</td>
<td>1.77</td>
<td>7,899</td>
</tr>
<tr>
<td>All-Amerindians*</td>
<td>377</td>
<td>4.61</td>
<td>5.92</td>
<td>3.47</td>
<td>1.14</td>
<td>31,454</td>
</tr>
<tr>
<td>All-Amerindians**</td>
<td>377</td>
<td>2.22</td>
<td>2.47</td>
<td>1.72</td>
<td>0.50</td>
<td>16,274</td>
</tr>
</tbody>
</table>

*All RFLPs were used for these estimates.

**Haplogroup-defining RFLPs (HaeIII 663, HincII 13259, Alul 13262, Alul 5176, Ddel 10394, and Alul 10397) were excluded for these estimates.
Fig. 2. Mismatch distributions for individual Native American populations using RFLP haplotypes. The distributions are presented as normalized values to permit comparisons between populations of unequal sample sizes. These values are indicated along the y-axis, while the number of site differences between haplotypes is shown along the x-axis, with the scale indicated in units of five.
and Stoneking (1998) estimated that haplogroups A–D and X began expanding, on average, between 37,000–22,000 cal BP, with haplogroups A and D being the oldest and haplogroups B and C being slightly younger. These estimates fall within analogous coalescence dates estimated from equivalent HVR-I sequence data by Bonatto and Salzano (1997). In contrast to the RFLP mismatch results, however, neither of these studies generated later expansion dates for haplogroups A and B. Instead, the HVS-I sequences gave relatively smooth unimodal distributions that revealed similar dates of expansion of haplogroups A–D in the Americas. Furthermore, neither study excluded the HVS-I sequence polymorphisms that define the sequence motif of each haplogroup when making their coalescence estimates.

MIGRATIONS TO THE NEW WORLD

There has been considerable discussion about the number of migrations that reached the New World and gave rise to ancestral Native Americans. Based on nonmolecular data, this number has ranged from one to eight or more, depending on the dataset being used (craniometric, dental, classical blood group markers, GM allotypes, HLA haplotypes). There is general agreement that the Eskimo-Aleuts and Na-Dene Indians represent the last significant population expansion into the New World (Shields et al., 1993; Starikovskaya et al., 1998; Schurr et al., 1999; Saillard et al., 2000; Rubicz et al., 2003); and, for this reason, the emergence of circumpolar genetic lineages will not be addressed here.

The primary question is the number of migrations responsible for the genetic diversity of Amerindian groups. The alternative migration scenarios can be evaluated by scrutinizing the evidence for multiple entries of genetic lineages into the New World. By identifying the number of founding lineages and haplotypes present in Siberia and the Americas, one can determine whether or not multiple founder types were brought to the New World at different times. In addition, by examining the patterns of mutations present in these haplotypes, it may be possible to find specific mutations that demarcate the movement of certain of these haplotypes or genetic lineages, hence, human groups, into the New World.

Number of migrations based on mtDNA haplogroup data

Several different models for the peopling of the New World have been proposed based on mtDNA data. Most of them have suggested a region extending from the Altai Mountains to South-Central Siberia and northern China as the potential source area(s) for ancestral Native American populations (Kolman et al., 1996; Merriwether et al., 1996; Sukernik et al., 1996; Santoro et al., 1997). However, there is no complete agreement on the numbers of migrations that left this region and entered the New World.

Many researchers have asserted that haplogroups A–D were brought to the New World in a single migratory event. This view is based on the fact that all four founding haplogroups are present throughout the Americas (Merriwether et al., 1995, 1996; Forster et al., 1996; Kolman et al., 1996), and that statistical and pairwise mismatch analyses of their HVR-I sequences indicate similar levels of diversity in each mtDNA lineage (Bonatto and Salzano, 1997; Stone and Stoneking, 1998; Silva et al., 2002). According to this view, the pattern of genetic variation seen in modern Native American groups is largely attributed to in situ differentiation and population movements occurring after the initial colonization of the New World, rather than a consequence of sequential expansions.

Other investigators have argued for the occurrence of two or more migrations to the Americas. It has been proposed that ancestral Amerindian populations brought haplogroup A, C, and D mtDNAs from Siberia during the initial colonization(s) of the New World, with haplogroup B possibly representing a second independent migration. This view was based on the fact that haplogroup B appeared to be
younger than the other founding lineages, was absent from most of northern Asia from East Asia to the Americas, and was widely distribution in East and Southeast Asia (Torrioni et al., 1993a,b; Starikovskaya et al., 1998). Others have asserted that each haplgroup represents a separate migratory wave to the New World (Horai et al., 1993). However, evidence indicating similar levels of diversity in haplogroups A–D in the Americas complicates these interpretations.

On the other hand, haplogroup X may still represent a separate migration from somewhere in Eurasia. It is absent in nearly all indigenous Siberian populations (Torrioni et al., 199b; Starikovskaya et al., 1998; Schurr et al., 1998, 2000; Derenko et al., 1998, 2001; Schurr and Wallace, 2003), and appears only in North American Amerindian populations (Torrioni et al., 1992, 1993a; Scozzari et al., 1997; Huoponen et al., 1997; Brown et al., 1998; Smith et al., 1999; Malhi et al., 2001, 2003; Bolnick et al., 2003). Its estimated age in Eurasia ranges from 35,000–20,000 cal BP (Torrioni et al., 1998; Richards et al., 1998, 2000; Simoni et al., 2001; Reidla et al., 2003), and is somewhat younger in the Americas (17,000–13,000 cal BP based on RFLP data; Brown et al., 1998), suggesting that it could have been brought to the New World within a relatively broad time range.

It also remains possible that mtDNAs from haplogroups A–D were introduced into the Americas in separate expansion events. This scenario receives tentative support from the fact that, in addition to the consensus founder haplotypes for these haplogroups, there are other HVS-I sequences shared between Asian/Siberian and Native American populations that could potentially be additional founder haplotypes (see above). These putative founder haplotypes are not widespread in Asia and the Americas, but appear in populations living in the vicinity of the hypothesized source area for ancestral Native Americans (Schurr and Wallace, 1999; Malhi et al., 2002; Schurr, 2002). However, further investigation of HVR-I sequence and coding region variation in both Siberian and Native American populations will be necessary to resolve this issue.

**Number of migrations based on NRY haplogroup data**

At least six different paternal haplotypes have been identified in Siberia and the Americas using SNP markers (DE-M1, Q-M3, R1a1-M17, P-M45, N3-M46, and C-M130). Only two of these, P-M45 and Q-M3, fundamentally contributed to the initial peopling of the New World, either through single (Underhill et al., 1996; Bianchi et al., 1998; Santos et al., 1999) or multiple (Karafet et al., 1999; Lell et al., 2002, Bortolini et al., 2003) migration events. The founding Q-M3 haplotype is the most frequent haplotype in Native American populations, and is widely distributed throughout the New World (Bianchi et al., 1998; Karafet et al., 1999; Lell et al., 2002; Bortolini et al., 2003). In addition, the ancestral P-M45a haplotype, the direct ancestor to the Q-M3 founder haplotype, has the widest geographic distribution of all of those present in the Americas, occurring in populations from central Siberia to South America (Lell et al., 2002; Bortolini et al., 2003).

A second and later expansion(s) of human groups from Beringia into the Americas brought with it a different set of P-M45 haplotypes. These P-M45b haplotypes show a different array of STR alleles than the Q-M3/P-M45a haplotypes that came with the initial expansion into the New World, as well as the M173 SNP. This second set of P-M45b haplotypes is also shared between eastern Siberian and North and Central American groups, but are absent in those from central Siberia and South America (Lell et al., 2002). The secondary expansion may also have contributed the R1a1-M17 and C-M130 haplotypes to Amerindian populations. Based on their distribution in Siberia, P-M45b and C-M130 haplotypes were proposed to have come from the Amur River region (Karafet et al., 1999; Lell et al., 2002).

Bortolini et al. (2003) supports the proposal that there were two major expansions of NRY haplogroups into the New World. However, they argue that both NRY migrations came from southern/central Siberia to the Americas. In their view, this interpretation is more consistent with the generally high frequency of haplogroup K-M9 in eastern Siberia and its absence in the Americas.

**DISCUSSION**

As attested by the new molecular data, the diversity of mtDNA and NRY haplogroups in Siberia and the Americas points to the dynamic quality of human movements in
these regions over the course of the last 30,000 years. However, there continues to be discussion about the incongruity between the late (14,675–13,050 cal BP) archeological visibility of the early colonizers of the Americas and the greater antiquity (35,000–15,000 cal BP) of the genetic lineages brought to these regions by the founding populations of Native Americans. The initial problem with the older age estimates for mtDNA and NRY haplogroups was that some were equal to the oldest known human occupation sites in southeastern Siberia (40,000–30,000 cal BP), and much older than the archeological sites in northeastern Siberia (Goebel, 1999; Goebel et al., 2001). However, the expanded datasets from Siberian and Native American populations, and the refinement of methods used to estimate mutation rates, effective population sizes, and genetic diversity now give ages for these genetic lineages that are more consistent with archeological data from these regions. Furthermore, with the recent discoveries of pre-Clovis cultures in the Americas, the dates for the earliest archeological sites in the Americas and the expansion times for these genetic lineages are slowly drawing closer together.

Both mtDNA and NRY data generally provide an initial entry time of ancestral Native Americans of between 20,000–15,000 cal BP. While this expansion time would seem to favor a late entry of ancestral Native Americans and is more consistent with archeological data from the New World, it presents other interesting problems. For one thing, these dates fall in the middle of the LGM, before the earliest time at which an ice-free corridor was available for passage by modern human population (~13,050–12,550 cal BP). In addition, the ages of the earliest archeological sites in the New World are somewhat older than the time at which the ice-free corridor became available for human movement. The Clovis sites long held as the benchmark of human colonization span a range of 13,350–12,895 cal BP (Haynes, 1992, 1993; Boldurian and Cotter, 1999; Fiedel, 1999), the pre-Clovis Monte Verde site has been dated at ~14,675 cal BP (Dillehay, 1997), and other pre-Clovis sites in South America are only slightly younger than Monte Verde (Dillehay, 1999; Dixon, 2001, 2002). Thus, one could conclude that ancestral Native American populations were already present in the New World before the LGM.

If glacial coalescence prevented access to North America via an interior route until ~12,550 cal BP, and if sites in South America were occupied by 14,675 cal BP, then how did human populations arrive in the New World prior to, or during, the LGM? Despite no clear evidence for human occupation of northeastern Siberia before 22,000–20,000 cal BP (Goebel, 1999; Goebel et al., 2001), it is possible that ancestral Amerindians moved from Beringia into North America before the beginning of the LGM around 24,000 cal BP. In this case, the expansion across Beringia and down into northern North America would likely to have been relatively quick, given the periglacial conditions in this ice-free corridor.

An alternative explanation for these data is that ancestral Amerindians followed a coastal route into the New World. This idea was put forward a number of years ago based on both geological and linguistic evidence (Fladmark, 1979, 1983; Gruhn, 1987, 1992a,b). However, only recently has geological evidence of deglaciation of the Northwest Coast of North America by between 16,800–14,850 cal BP clearly indicated that human occupation of this area during the LGM was possible (Blaise et al., 1990; Bobrowsky et al., 1990; Mann and Hamilton, 1995; Jackson and Duk-Rodkin, 1996; Josenhans et al., 1997; Mandryk et al., 2000; Fedje and Christiansen, 1999; Fedje et al., 2001; Fedje, 2002). In addition, computer simulations of a colonization process by interior and coastal routes using demographic data from hunter-gatherer groups has suggested that the patterns of genetic diversity in Amerindian populations are more consistent with a coastal model that allowed an earlier and rapid expansion into the Southern hemisphere (Fix, 2002).

Therefore, based on the existing anthropological data, one can generate the following picture of the peopling of the New World. There was a pre-Clovis entry of ancestral Asian groups into the Americas during the LGM. These immigrants used a coastal route to reach the areas below the glaciated areas of northern North America somewhere between 18,000–15,000 cal BP. They apparently brought mtDNA haplogroups A–D and NRY haplogroups P-M45a and Q-242/Q-M3 haplotypes with them to the Americas, with these being dispersed throughout all continental areas of the New World. A subsequent expansion probably brought mtDNA haplogroup X...
and NRY haplogroups P-M45b, C-M130, and R1a1-M17, with these being disseminated in only North and Central America. This expansion may have coincided with the opening of the ice-free corridor around 12,550 cal BP. A somewhat later expansion likely involved the emergence of circumarctic populations, such as Eskimos, Aleuts, and Na-Dené Indians (Shields et al., 1993; Starikovskaya et al., 1998; Schurr et al., 1999; Saillard et al., 2000; Rubiecz et al., 2003).

Of course, this model, like others before it, will continue to evolve as more archæological, genetic, and geological data are gathered from Siberia and the Americas. However, the increasing congruence between these different lines of anthropological evidence would seem to indicate that we are gaining a much better understanding of the processes that gave rise to the diversity of Native American peoples and cultures.

ACKNOWLEDGMENT

The authors thank the two anonymous reviewers for helpful comments on the manuscript.

LITERATURE CITED


Gurven M. 2000. How can we distinguish between mutational “hot spots” and “old sites” in human mtDNA samples? Hum Biol 72:455–471.


Merriwether DA, Hall WW, Vahne A, Ferrell RE. 1996. mtDNA variation indicates Mongolia may have been the source for the founding population for the New World. Am J Hum Genet 59:204–212.


Underhill PA, Jin L, Zemans R, Oefner PJ, Cavalli-Sforza L-L. 1996. A pre-Columbian Y chromosome-


