

Phylogenetic analysis was conducted on the lineages identified using PAUP (version 3.1.1; Swofford 1993). Maximum parsimony trees were generated through the simple addition of sequences using the tree bisection and reconnection algorithm. A maximum of 2000 trees were saved for each search, and a 50% consensus tree was constructed.

Sequence differences between pairs of the studied 225 individuals were computed from matrices produced by PAUP. Intragroup mean pairwise sequence differences were calculated for members of each of the three mtDNA haplogroups. Both intra- and intergroup mean pairwise sequence differences were computed for all individuals representing each of three language taxa (Eskimo, Na-Dene, Amerind). Similar estimates were made for members of haplogroup A that were present in the three language taxa and for each of the 5 ethnic groups represented by 12 or more individuals in this study. Although the unity of Amerind as a valid language entity is contentious, the distributional properties of pairwise differences between other clusters of tribes whose languages have been hypothesized to constitute valid language entities were not studied. However, the clustering of tribes representing alternative hypothetical language groups can be visually inspected in the tree phylogeny. This is important because language spreads (e.g., the recent spread of the Numic languages into and throughout the Great Basin), discrete migrations (e.g., that of Athapaskans to the American Southwest), and the reduced ability to accurately reconstruct phylogenies of languages that split more than 10,000 years ago or so might obscure correlations between genes and language. The last source of confusion is especially cogent because some linguists have argued that the level of diversity among native American languages would have required 30,000–35,000 years to accumulate (Nichols 1990), a time period that probably predates the settlement of the New World.

Historical events leading to admixture and/or stochastic effects on the size distribution of lineages within a language group or tribe can distort estimates of pairwise sequence divergence from their expected distribution and linearity in time (Marjoram and Donnelly 1994; Rogers and Harpending 1992; Slatkin and Hudson 1991). For example, recent gene flow can pose as shared ancestry between two groups, whereas subsequent rapid increases in the size of groups or any other lineages as a result of variance in reproductive success in groups of limited size can obscure close common ancestry to other groups. Because several of the studied groups have inhabited regions through which continuous migration probably occurred, contact among unrelated or very distantly related groups might have led to admixture, causing overestimates of genetic similarities among dissimilar language groups. The potential impact of these influences were evaluated by estimating pairwise sequence differences, in a separate analysis, using only one member of each tribal-specific or language-specific lineage represented in each group. Thus, only the variety, not the frequency distribution, of the lineages in each group could

influence the differences in pairwise divergence, and the influence of population substructure on divergence estimates was minimized.

The modalities of the distributions of pairwise differences were also studied to evaluate the plausibility of panmixia and founder effects, both of which are associated with unimodal distributions of pairwise sequence differences. Population expansion leads to the increased retention of lineages, causing a starlike phylogeny and a Poisson distribution of pairwise sequence differences (Rogers and Harpending 1992). We also analyzed the rate of coalescence of lineages over time, measured as the number (or percentage) of gene substitutions, to assess whether or not expansion was constant (Nee et al. 1996), as assumed when gene divergence is calibrated to time.

Results

The ethnic affiliations of the 225 samples are shown in Table 1 for each of the three haplogroups. The 85 distinct mtDNA lineages represented by these individuals were defined by 68 polymorphic sites, all of which were transitions; these are shown in Table 2. Of the 34 individuals that we sequenced in this study, 22 belonged to 1 of 18 previously undescribed lineages. Three additional individuals belonged to two lineages that have not been described previously in North America but have been found elsewhere: 2 Kumiiai belonged to lineage 120, which has also been found among Indonesians, Papua New Guineans, Samoans, Hawaiians, Malaysians, and Chinese (Lum et al. 1994; Redd et al. 1995), and 1 Nahuatl individual belonged to lineage 171, which is also found among the Mapuche of South America (Ginther et al. 1993). The remaining nine individuals belonged to two previously described North American lineages (lineages 89 and 180).

Seventy-three of the 85 lineages were tribal specific. Approximately equal portions (85–88%) of the lineages within each of the three haplogroups were tribal specific. The 12 lineages that were shared between 2 or more tribes (identified as “multiple” in Table 2) are listed in Table 3. Six of these 12 lineages were shared by tribal groups within the same language family, as defined by Greenberg (1987) (i.e., Eskimo, Na-Dene, Amerind), with 2 of these 6 lineages (lineages 63 and 64) being shared between 2 of the 3 Eskimo groups, 1 being shared between 2 of the 5 Na-Dene groups (lineage 30), and 3 being shared among 2 or more of the 14 Amerind groups (lineages 86, 89, and 180). The most common of these 3 (lineage 89) was shared by 7 of the 14 different Amerind groups. Of the six remaining lineages that were shared by two of the language groups (none was shared by all three of the language groups), two (lineages 66 and 67) were shared between at least one of the two Eskimo groups and one Na-Dene group (the Alaskan Athapaskans), one (lineage 65) was shared between one Eskimo and one Amerind group, and three (lineages 15, 18, and 195) were shared between at least one Na-Dene