Am. J. Hum. Genet. 70:905–919, 2002

The Scope of Genetic and Genomic Testing Today and Tomorrow: Identification of the Real Harms and Benefits

[illegible]

presence of these haplogroups in ancient populations of Northeast Asia would confirm this region as the homeland of North American colonizers.

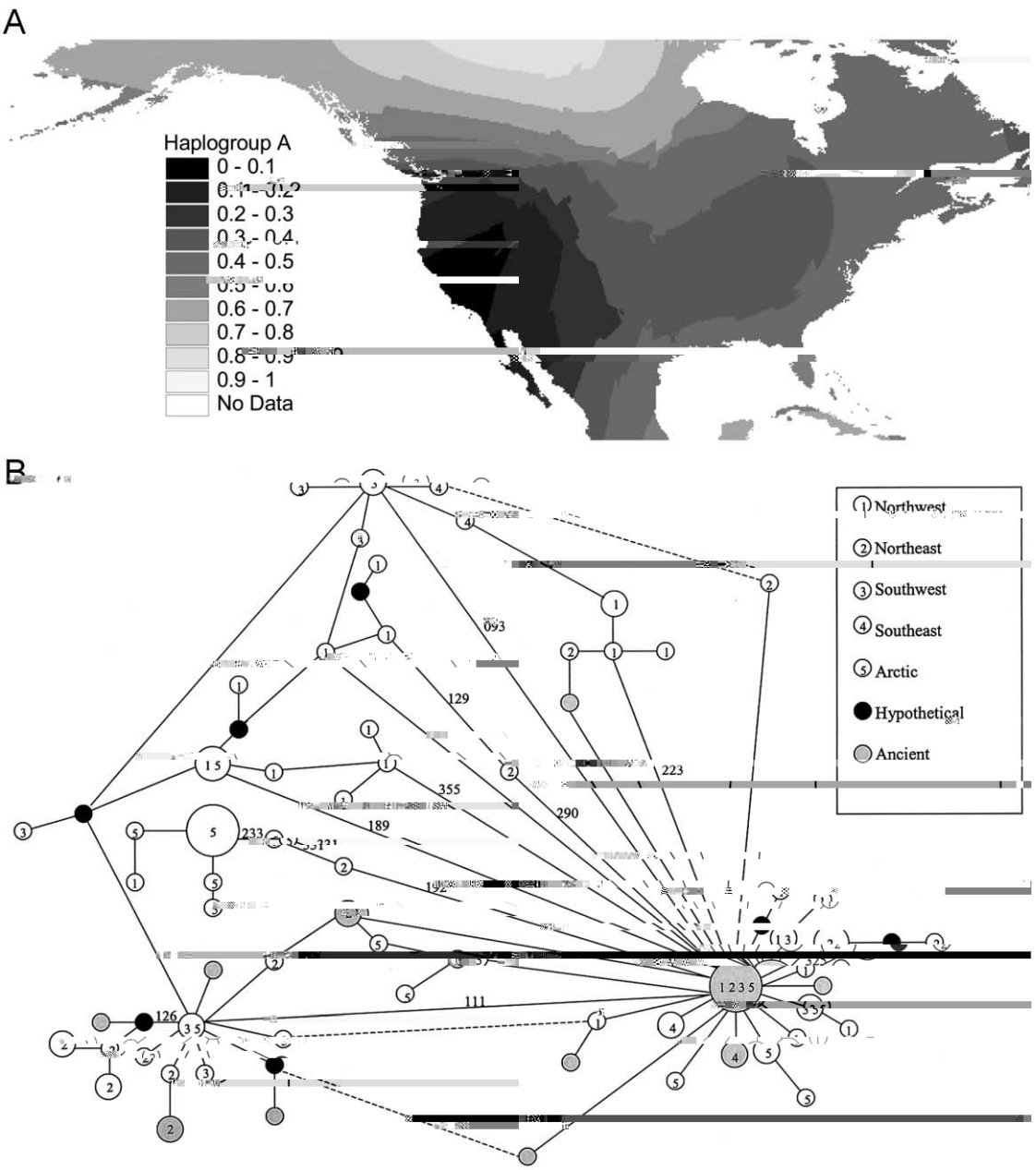
The timing of initial entry into the Americas is un-

Materials and Methods

The haplogroup identities of 1,612 Native Americans and the mtDNA sequence of the hypervariable segment I (HVSI) region (nps 16,090–16,362) of 395 of these individuals were analyzed. Populations from which haplogroup data were collected are reported in table A1 in the Appendix.

Individuals whose mtDNAs did not belong to one of the five Native American haplogroups were not included in this analysis. Although it is possible that one or more of these individuals possess previously undocumented founding Native American mtDNA types, previous studies indicate that the frequency of "other" mtDNA types is very low and that most—or all—of these result from recent admixture (Torroni et al. 1993, 1994; Huoponen et al. 1997; Smith et al. 1999). O'Rourke et al. (2000) demonstrated that most modern Native American populations surveyed to date display a pattern of regional continuity in haplogroup-frequency distribution. These patterns of haplogroup frequencies across North America suggest that a model of isolation by distance is appropriate for the analysis of such data. Using the Kriging method in the ARC/INFO software package, we interpolated haplogroup frequencies among 36 groups from across North America. Since the use of an interpolating method introduces artificial spatial autocorrelation (Sokal et al. 1999), only results that were also supported by network analyses using haplotype data were interpreted.

Haplotype networks were constructed for each of the five haplogroups, through use of three different methods. Median-joining and reduced median networks were constructed using the NETWORK 2.0 program



Fg e 1 *A*, Map of haplogroup A frequencies calculated using the Kriging interpolation method. *B*, Two-step haplotype network for

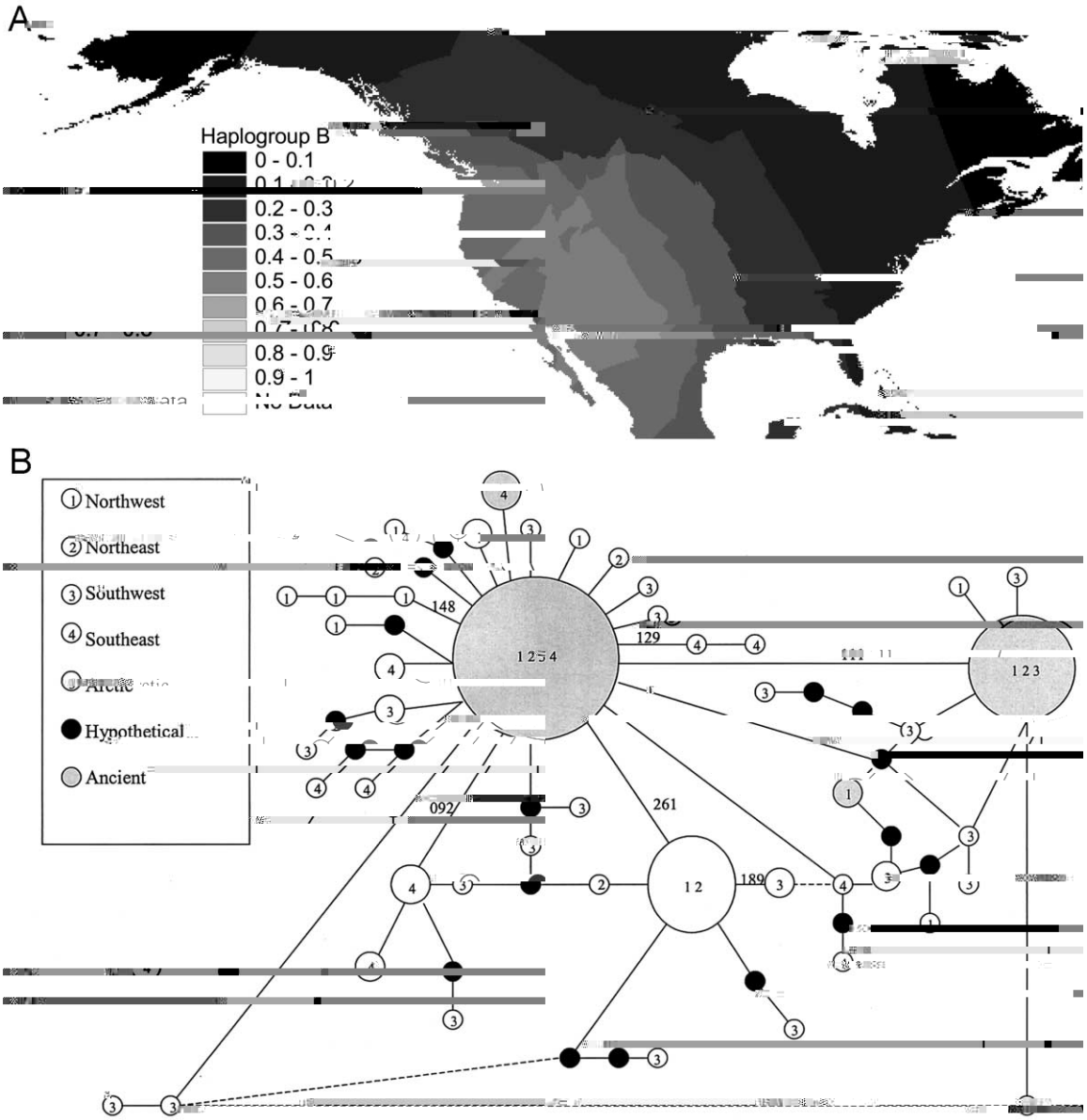


Fig e 2 *A*, Map of haplogroup B frequencies calculated using the Kriging interpolation method. *B*, Three-step haplotype network for haplogroup B. For an explanation of the diagram, see figure 1.

Estimate of Molecular Diversity

Estimates of within-haplogroup diversity are given in table 1. When the diversity of unique haplotypes only is compared (an unweighted estimate), haplogroups A, B, and C exhibit similar amounts of diversity for all measures analyzed. The lower values of θ found in haplogroups D and X are probably due to the smaller number of haplotypes used in the estimate of diversity—24 and 16, respectively. When diversity measures were weighted by frequency of haplotypes, both measures of diversity dropped. This decline was most noticeable in the calculation of π within haplogroups B and D, which appeared to have substantially lower diversity estimates under this condition.

Haplotype Distribution

An average of 29.6% of mtDNA sequences (haplotypes) are shared among Native American individuals (table 2). Ancient haplotypes were included in this estimate, even though it is possible that they are direct

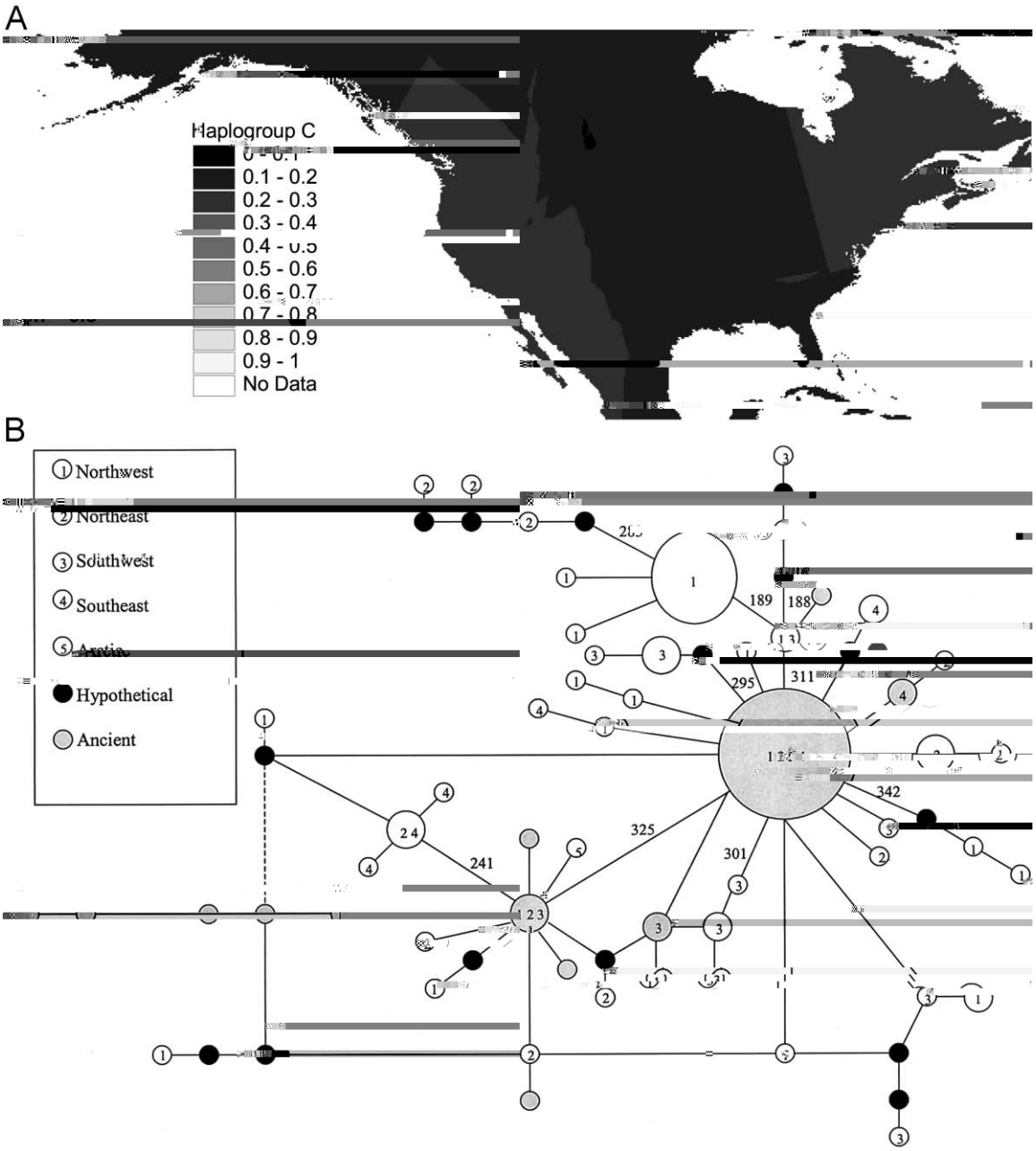
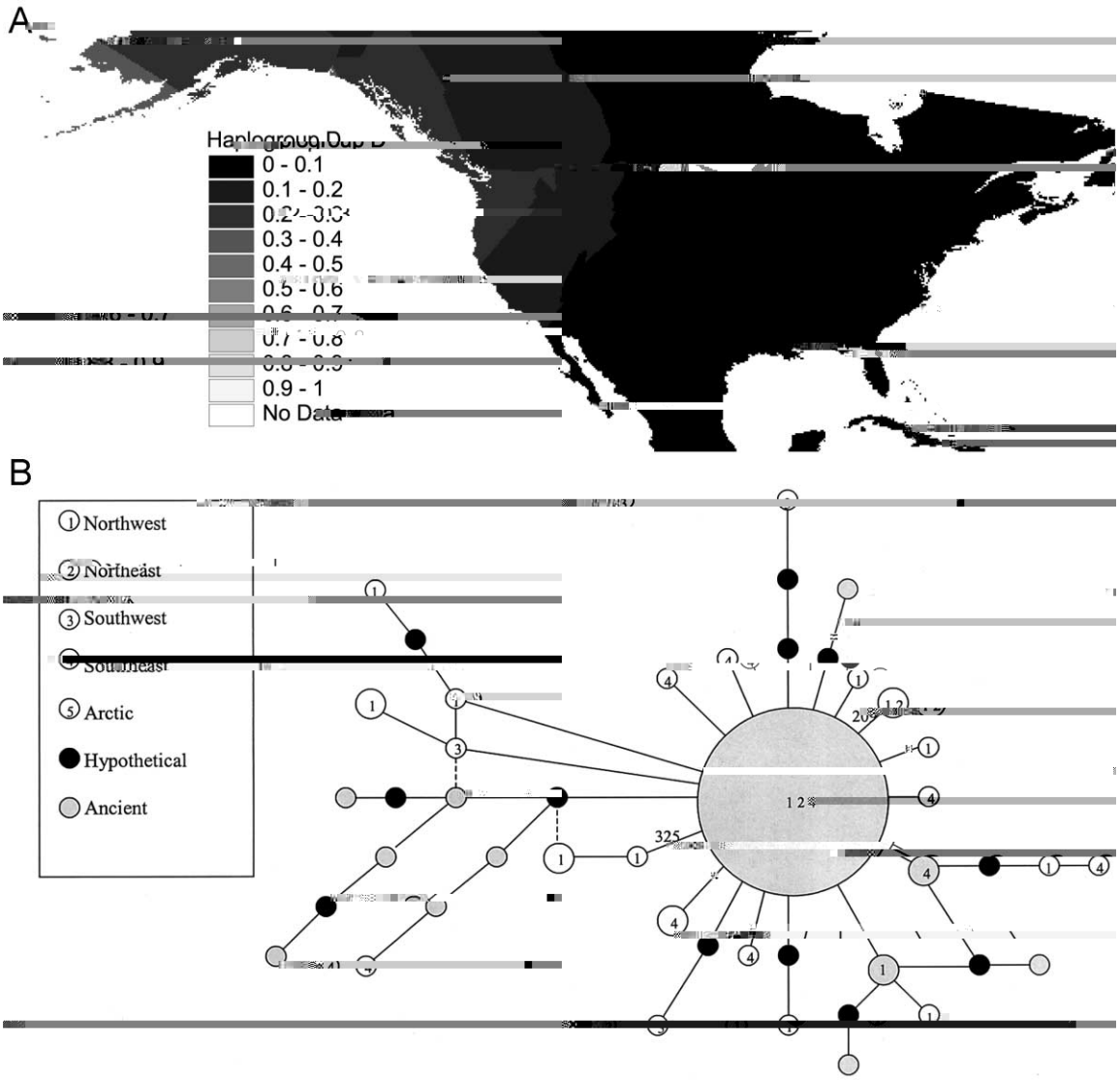


Fig e 3 *A*, Map of haplogroup C frequencies calculated using the Kriging interpolation method. *B*, Three-step haplotype network for haplogroup C. For an explanation of the diagram, see figure 1.

ancestors of modern haplotypes. However, 64% of ancient haplotypes in our sample are unique and therefore have left no known descendants. Of these shared haplotypes, 40.3% are shared among geographically distant individuals, 30.8% are shared among individuals within the same region, and 28.9% are private tribal polymorphisms.

Figures 1*B*, 2*B*, 3*B*, 4*B*, and 5*B* illustrate the haplotype networks. Not surprisingly, internal nodes are generally more widely distributed geographically than external nodes, but there is no other consistent pattern among the five haplotype networks. In many instances, external haplotypes are clustered among individuals belonging to the same tribe or region, but some external haplotypes do not follow this trend. The haplogroup C network exhibits more isolation by distance than do the



Fg e 4 *A,* Map of haplogroup D frequencies calculated using the Kriging interpolation method. *B,* Three-step haplotype network for ag0cga29J553g0cp. 9.9T*-312.

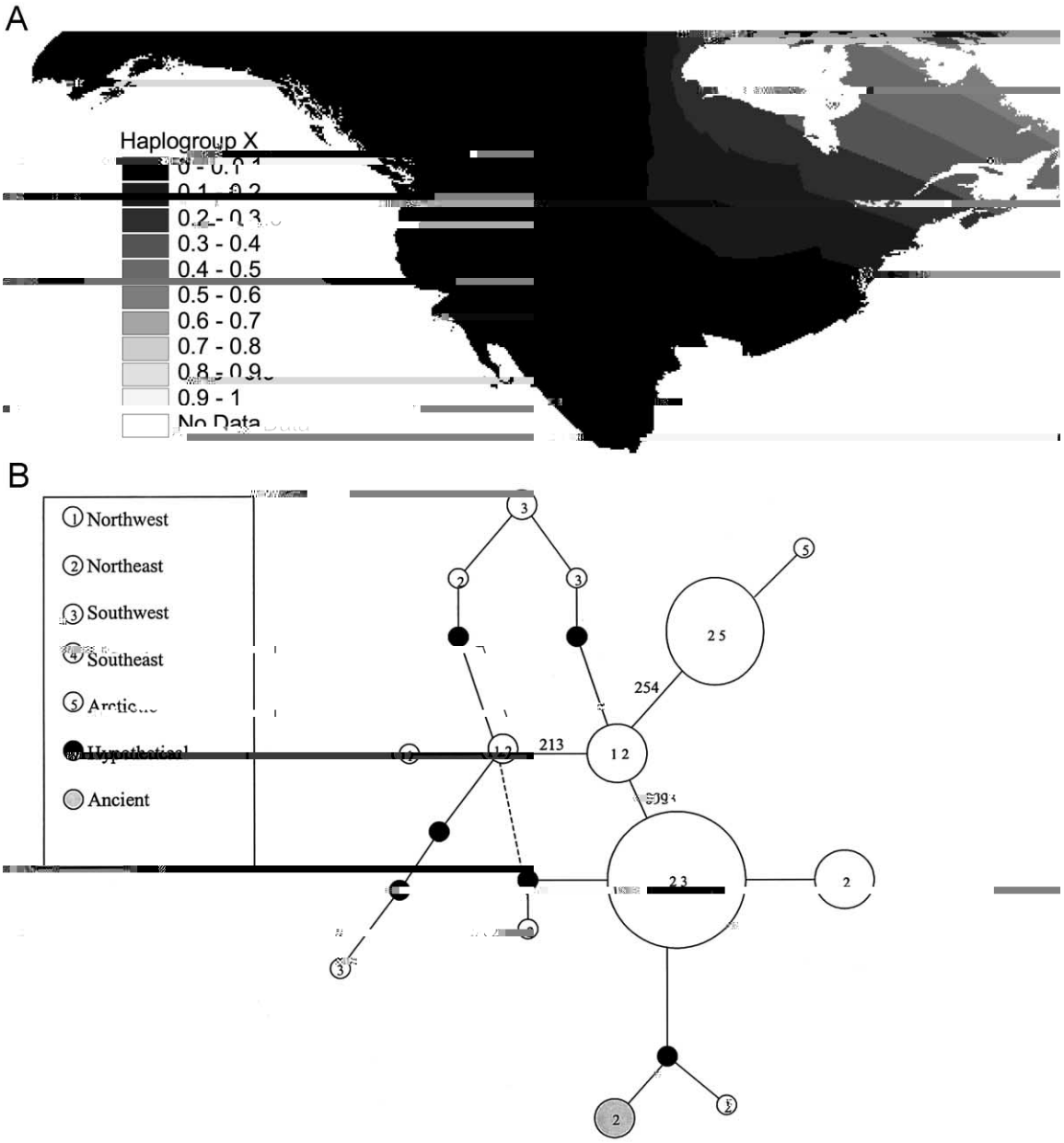


Figure 5 *A*, Map of haplogroup X frequencies calculated using the Kriging interpolation method. *B*, Three-step haplotype network for haplogroup X. For an explanation of the diagram, see figure 1.

terminal and might be extinct. The haplogroup X network exhibits haplotypes from four of the five geographic regions, but Algonquian-speaking individuals predominate in the network. The extent to which sampling contributes to this pattern is not clear.

In the present study, 36% of haplotypes from ancient samples are shared with modern Native Americans. If they are not shared, most ancient haplotypes cluster with the modern haplotypes, suggesting that European contact did not cause a significant change in haplotype structure in most regions of North America. In the Southeast, however, four different ancient haplotypes of haplogroup D form intermediate nodes that connect highly divergent modern external haplotypes with internal haplotypes found in ancient Native Americans. This suggests that, unlike other regions of North America, the Southeast experienced a high percentage of haplotype extinctions. This pattern is consistent with the random distri-

Table 1

Diversity Estimates

Haplogroup	*N*	π	θ_S	Sequence Divergence[a]
Unweighted:				
X	16	.183	5.123	1.201
A	60	.079	9.650	1.341
B	47	.071	10.415	1.252
C	45	.088	9.834	1.397
D	24	.089	8.569	1.201
Weighted:				
X	46	.143	3.868	.935
A	83	.072	9.018	1.215
B	80	.052	9.287	.910
C	75	.067	8.797	1.067
D	54	.048	7.242	.661

NOTE.—Ancient haplotypes are not included in the estimates.
[a] Based on 290 base pairs.

bution of haplogroup frequencies among closely related populations from the Southeast, which may indicate a recent bottleneck as a result of European contact (Weiss 2001).

Discussion

Hypervariable Sites and the Control Region

Analysis of the median joining, reduced median, and statistical parsimony networks for the 290-bp segment of HVSI revealed significant reticulation in all haplogroups that was not readily resolved using a coalescence method. Although the future discovery of new haplotypes might resolve some of these reticulations, the high degree of reticulation could be the result of the hypervariable nature of this region, since a number of sites have previously been reported to be hypermutable (Hasegawa et al. 1993; Wakeley 1993; Excoffier and Young 1999; Meyer et al. 1999; Gurven 2000; Sigurðardóttir et al. 2000; Stoneking 2000). Many sites appear to have mutated in multiple haplogroups; for example, the haplotype networks for haplogroups A and B both exhibit mutations at np 16111, np 16129, and np 16189. However, hypervariability need not be implicated in all instances of multiple hits at a nucleotide position. In the most conservative assessment of mutational positions, where mutations are distributed entirely at random throughout the 290-bp region, the probability that each mutational site among n such sequences is unique is 290 Permut n/(290^n); with 21 variable positions among the 290 bases analyzed, the probability of $\geqslant$1 position

Tab e 2

D b f Ha e a Ha g

	PERCENTAGE				
				Component of Shared Haplotypes[b]	
HAPLOGROUP (*N*)	Ancient Shared with Modern[a]	Basal Haplotype	Shared Haplotypes	Tribal Shared Polymorphisms	Distant Shared Polymorphisms
A (70)	38.46	7.22	24.29	17.65	35.29
B (56)	66.67	20.21	21.43	33.33	33.33
C (51)	45.45	16.67	23.53	25.00	50.00
D (35)	28.57	36.51	20.00	28.57	42.86
X (17)	1.00	3.51	58.82	40.00	40.00
Average	36.03	16.82	29.61	28.91	40.30
SD	24.06	12.93	16.41	8.44	6.60

[a] Indicates the percentage of nonunique ancient haplotypes.
[b] Includes ancient and modern samples.

(Lorenz and Smith 1996; Kaestle and Smith 2001; Malhi et al. 2001). It is po731le2blet1(s)-3h.1(iE4r3ibaliza343.05a)-215(Smpr)0(ibal)-337.6(Sha4t1(0npes.)uw5Smpr65(bf8vn498.4(3-31

The gene map suggests that haplogroup B experienced an expansion in the southwestern region of North America. This expansion broadened the distribution and increased the frequency of the haplogroup B clade, particularly one subclade of B in the Southwest. Defined by a T→C mutation at np 16261 (Malhi 2001), this subclade comprises 82% of haplogroup B mtDNA in North America. The high values of θ_S for haplogroups B and C, which predominate in the Southwest (Malhi 2001), also suggest an early population expansion in this region. There is archeological evidence of a population expansion in the Southwest during early Holocene times. The oldest Clovis-associated dates, ~11,550 radiocarbon years BP (rcbp) (13,400 years BP), come from Texas (the Aubrey site; Fiedel 1999) and Clovis sites that subsequently appear throughout North America and that undergo a stylistic transition in Central America into Fell's Cave fishtail points. These fishtail-type points are then carried throughout South America, reaching Tierra del Fuego 11,000 rcbp (Fiedel 1999).

Additional evidence of an early population expansion in the Southwest is provided by Fisher et al. (2001), whose phylogenetic analysis of microsatellite data demonstrated that the common ancestor for all variants of Valley Fever (*Coccidioides immitis*) in South America is located in the American Southwest. They suggest that Valley Fever spread from the American Southwest to Central and South America, some time before 9,000 years BP, as a commensal of humans. A much later population expansion associated with the development and use of agriculture in the Southwest, ~3,500–1,500 years BP (Fagan 2000), probably contributed the remainder of the variation in this clade of haplogroup B.

Our sample of haplogroup X consists of a large percentage of shared haplotypes among tribes speaking Chippewa/Ojibwa languages and dialects. The haplogroup X network and distribution of haplogroup frequencies suggest that populations with relatively high frequencies of haplogroup X experienced an expansion in the Great Lakes region. This expansion, which generated a value of θ_S only half that for haplogroup B, must have occurred much more recently in prehistory than the expansion of haplogroup B. Archeological, linguistic, and genetic evidence all strongly support the expansion of Algonquian-speakers from the Great Lakes region, ~2,500–3,000 years BP (Denny 1991; Malhi et al. 2001). Ancient-DNA studies of prehistoric populations from the Great Lakes region demonstrated that this Algonquian expansion probably occurred 700–3,000 years BP (Schultz et al. 2001).

Although the marked variation among haplogroups in the weighted estimates of π could be interpreted as evidence for multiple waves of colonization, these estimates are highly sensitive to sampling error. Specifically, a high percentage of the haplogroup B sequences is limited to the Southwest region, and many of these are basal sequences. Although this could arise from a more recent entry into the Southwest, as some have argued (Torroni et al. 1992), a recent re-expansion in this region could also result in an overrepresentation of basal lineages in the weighted sample. Indeed, under this condition, π would not reflect the diversity accumulated since colonization from Asia but rather a recent expansion resulting from the introduction of agriculture to this region

signed to haplogroup X lack a mutation at np 16213 in the HVSI that all Native Americans exhibit. However, the larger sample size of individuals assigned to haplogroup X in the present study reveals that a substantial number of Native Americans in multiple geographic regions also lack the np 16213G mutation and therefore have haplotypes identical to those of European (Brown et al. 1998) and Asian (Derenko et al. 2001) members of haplogroup X. A central X haplotype is shared among Native Americans in the Northwest and Northeast, suggesting that this haplotype might be the founding X haplotype in eastern North America. Smith et al. (1999) demonstrated that haplogroup X is present in a more linguistically diverse population in the Northwest, whereas in the Northeast this haplogroup is mainly limited to Algonquian speakers. This is consistent with the hypothesis that haplogroup X was first introduced to the eastern part of North America by Algonquians emigrating from northwestern North America (Malhi et al. 2001; Schultz et al. 2001).

The present study raises doubt about interpretations of previously reported evidence for the number of migrations to the Americas. If substantiated, the presence of additional founding haplotypes within haplogroups B and C in the New World would significantly reduce previous estimates of diversity accumulated since colonization within these haplogroups. Many researchers (Bonatto and Salzano 1997*a;* Lorenz and Smith 1997; Brown et al. 1998; Stone and Stoneking 1998) have interpreted similar estimates among at least four of the five Native American haplogroups as evidence that all haplogroups entered the Americas at the same time. In contrast, Torroni et al. (1994) employed exhaustive restriction analysis to argue that the lower diversity within haplogroup B suggests a later migration of this haplogroup to the New World. These estimates rely, at least in part, on knowing the level of diversity of haplotypes at the time of initial colonization. If multiple haplotypes within a haplogroup were successful colonizers of the New World, modern values of within-haplogroup diversity would overestimate the accumulated variation since contact.

In addition, the HVSI portion of the control region used to create estimates of genetic diversity in many studies exhibits a high percentage of polymorphic sites, suggesting that the entire region itself is hypervariable. If so, many nucleotide sites will experience multiple hits, resulting in back mutations, and certain mutational sites will mutate independently in separate lineages. Comparisons of mutational sites between haplogroups and of mutational sites within haplotype networks strongly suggest that both events have occurred. The level of diversity that accumulates after this saturation point has been reached will result in a nonlinear accumulation of mutations within a haplogroup, thereby further impairing the utility of molecular diversity for the dating of the colonization event. Diversity estimates are also strongly affected by both sampling and population historic and demographic events that have occurred since colonization. Previously reported lower diversity estimates for haplogroup B may well be a result of the more recent expansion of this haplogroup within the Southwest, an area that has been well represented—and sometimes overrepresented—in many studies of Native American mtDNA diversity. The lower levels of diversity within haplogroup B might actually be reflecting this expansion rather than an earlier colonization of the Americas. In this light, we believe that the wide distribution of haplogroups throughout North America is strong evidence for a single entry from Asia.

The high rate of mutation of HVSI suggests that control region data are useful for identifying diagnostic mutations that are specific to a tribe. However, older mutations that are potentially regionally specific or shared among all Native Americans are obscured by the high frequency of multiple hits at nucleotide positions in this region. Therefore, we are unable to distinguish whether shared mutations among geographically distant individuals are due to prehistoric Native Americans existing as one population with little substructure for an extended period of time prior to intensification of resource utilization during Archaic times or, alternatively, due to convergence. Analysis of polymorphic sites in a less mutable region of the mitochondrial genome, such as a coding region, in addition to the sites in the control region, might help resolve some of the reticulations in Native American haplotype networks, as was done with European haplotype networks (Finnila and Majamaa 2001). This would allow us to better date the time of tribalization of prehistoric Native American groups and to identify additional founding haplotypes.

Ac ,edg- e

We would like to thank Dr. Robert Bettinger and two anonymous reviewers, for their suggestions and insights. We are indebted to numerous personnel of Indian Health Service Facilities, where most of the samples studied were obtained, as well as to individuals who provided samples used in this analysis and to the Native Americans who authorized their use. This study was supported by National Institutes of Health grants RR00169 and RR05090, by National Science Foundation grants GER9255683 and SBR9630926, and by a National Science Foundation dissertation improvement grant (to R.S.M.).

A e d

Tab e A1

P a f - W c Ha g Da a We e C ec ed

Population	Sample Size	Geographic Location	Reference for RFLP Data
Navajo	64	Southwest/haplogroup A: Northwest	Malhi 2001
Northern Paiute	98	Northwest	Kaestle and Smith 2001
Cheyenne/Arapaho	35	Northeast	Malhi et al. 2001
Washo	38	Northwest	Lorenz and Smith 1996
Yokut	17	Southwest	Lorenz and Smith 1996
Havasupai	18	Southwest	Lorenz and Smith 1996
Quechan	23	Southwest	Lorenz and Smith 1996
Kumeyaay	16	Southwest	Lorenz and Smith 1996
Apache	38	Southwest/haplogroup A: Northwest	Malhi 2001
Pima	43	Southwest	Malhi 2001
Hopi	4	Southwest	Lorenz and Smith 1996
Sioux	45	Northeast	Malhi 2001
Mohawk	123	Northeast	Merriwether and Ferrell 1996
Ojibwa	33	Northeast	Scozzari et al. 1997
T.M. Chippewa	28	Northeast	Malhi et al. 2001
Pawnee	5	Southeast	Malhi et al. 2001
Stillwell Cherokee	37	Southeast	Malhi et al. 2001
Zuni	26	Southwest	Malhi et al. 2001
Jemez	36	Southwest	Malhi 2001
Eskimo	115	Arctic	Merriwether et al. 1995
Aleut	72	Arctic	Merriwether et al. 1995
Creek	35	Southeast	WeissCoNC6hoctaw6Aleut

72 Arc

Ogg A (eds) Proceedings of the 32d Algonquian Conference. Carleton University Press, Ottawa, pp 470–492

Schurr TG, Ballinger SW, Gan YY, Hodge JA, Merriwether DA, Lawrence DN, Knowler WC, Weiss KM, Wallace DC (1990) Amerindian mitochondrial DNAs have rare Asian mutations at high frequencies, suggesting they derived from four primary maternal lineages. Am J Hum Genet 46:613–623

Scozzari R, Cruciani F, Santolamazza P, Sellitto D, Cole DE, Rubin LA, Labuda D, Marini E, Succa V, Vona G, Torroni A (1997) mtDNA and Y chromosome–specific polymorphisms in modern Ojibwa: implications about the origin of their gene pool. Am J Hum Genet 60:241–244

Shields GF, Schmeichen AM, Frazier BL, Redd A, Voevoda MI, Reed JK, Ward RH (1993) mtDNA sequences suggest a recent evolutionary divergence for Beringian and northern North American populations. Am J Hum Genet 53:549–562

Sigurðardo