AMERICAN JOURNAL OF PHYSICAL ANTHROPOLOGY 140:417–428 (2009)

X-Chromosome Lineages and the Settlement of the Americas

Stephane Bourgeois,[1] Vania Yotova,[1] Sijia Wang,[2] Sylvie Bourtoumieu,[1] Claudia Moreau,[1] Roman Michalski,[3] Jean-Paul Moisan,[4] Kim Hill,[5] Ana M. Hurtado,[5] Andres Ruiz-Linares,[2] and Damian Labuda[1,6]*

[1]*Centre de Recherche de l'Hôpital Sainte-Justine, Montréal, QC H3T 1C5, Canada*
[2]*Galton Laboratory, University College London, London NW1 2HE, UK*
[3]*Prince Albert Parkland Health Region, Victoria Hospital, Prince Albert, SK S6V 5T4, Canada*
[4]*Centre Hospitalier Régional et Universitaire, 44093 Nantes, France*
[5]*School of Human Evolution and Social Change, Arizona State University, Tempe, AZ 85287-2402*
[6]*Département de Pédiatrie, Université de Montréal, Montréal, QC H3T 1C5, Canada*

KEY WORDS Amerindians; population genetics; genetic structure; migration

ABSTRACT Most genetic studies on the origins of Native Americans have examined data from mtDNA and Y-chromosome DNA. To complement these studies and to broaden our understanding of the origin of Native American populations, we present an analysis of 1,873 X-chromosomes representing Native American (n = 438) and other continental populations (n = 1,435). We genotyped 36 polymorphic sites, forming an informative haplotype within an 8-kb DNA segment spanning exon 44 of the dystrophin gene. The data reveal continuity from a common Eurasian ancestry between Europeans, Siberians, and Native Americans. However, the loss of two haplotypes frequent in Eurasia (18.8 and 7%) and the rise in frequency of a third haplotype rare elsewhere, indicate a major population bottleneck in the peopling of the Americas. Although genetic drift appears to have played a greater role in the genetic differentiation of Native Americans than in the latitudinally distributed Eurasians, we also observe a signal of a differentiated ancestry of southern and northern populations that cannot be simply explained by the serial southward dilution of genetic diversity. It is possible that the distribution of X-chromosome lineages reflects the genetic structure of the population of Beringia, itself issued from founder effects and a source of subsequent southern colonization(s). Am J Phys Anthropol 140:417–428, 2009.

In the last few decades, molecular biology and genetics have joined archaeology, anthropology, and linguistics in the quest for the understanding of the origins of Native Americans and the peopling of the Americas. Although it is generally accepted that the source of Native American migrations lies somewhere in North Eastern Asia and that the Bering Strait was the gateway to the Americas, there is no consensus as to the number and timing of the migrations, the number of migrants, and the routes they took.

The debate has principally been driven by studies of mtDNA and Y-chromosome diversity. Studies of the maternal lineages described five founding mtDNA haplogroups in the Native American population (Torroni et al., 1992, 1993; Brown et al., 1998), presently referred to as A2, B2, C1, D1, and X2a, respectively. All five mtDNA haplotypes were identified in ancient pre-Columbian Native American remains (Stone and Stoneking, 1998; Malhi et al., 2002; Jones, 2003). Recent mtDNA analyses suggest that ancestors of Native Americans paused when they reached Beringia, which served as a glacial refugium during last glacial maximum (LGM). At the end of LGM, the coastal route was presumably used to reach southern South America around 14,500 years ago, and this initial migration seems to have involved all five of the main mtDNA haplogroups (Tamm et al., 2007; Achilli et al., 2008; Fagundes et al., 2008). Amerindian Y-chromosome lineages consist of two major haplogroups, Q and C, also found among Asian populations. In the Americas, haplogroup Q is spread from north to south, while haplogroup C is restricted to the north and is especially frequent among NaDene speakers (Bortolini et al., 2003; Zegura et al., 2004; Malhi et al., 2008).

Until recently, data from other genetic systems were patchy with different markers studied in different population groups (Salzano, 2002). A recent study by Mateus Pereira et al. (2005) analyzed polymorphic L1 and Alu insertions in 11 Native American populations. They showed a strong affinity of Native Americans with East Asians and their clear discrimination from Europeans

Additional Supporting Information may be found in the online version of this article.

Present address of Sylvie Bourtoumieu: Université Paul Sabbatier, F-31062 Toulouse Cedex 9, France

Grant sponsor: Canadian Institutes of Health Research; Grant number: MOP-67150.

*Correspondence to: Damian Labuda, Centre de Recherche, de l'Hôpital Sainte-Justine, 3175 Côte Sainte-Catherine, Montréal, Québec H3T 1C5, Canada. E-mail: damian.labuda@umontreal.ca

Received 10 September 2008; accepted 27 February 2009

DOI 10.1002/ajpa.21084
Published online 7 May 2009 in Wiley InterScience (www.interscience.wiley.com).

and Africans, and suggested that South-American diversity may have been caused by a major population bottleneck and genetic drift. Wang et al. (2007) studied variations in 29 Native American populations by genotyping 678 microsatellites from throughout the genome. They observed a decrease in genetic diversity in terms of geographic distance from the Bering Strait and from west to east in South America. Their results suggested that coastal routes played an important role in the peopling of the Americas, and that most Native American genetic ancestry could be derived assuming a single wave of migration. Interestingly, using an approximate Bayesian computation framework and coalescence-based simulations, Ray et al. (submitted) reanalyzed the same data to evaluate statistically the relative probability of alternative migration scenarios. They found that a model in which an initial migration wave is followed by recurrent gene flow between Asia and America provides by far the best fit to levels of genetic diversity observed across the American continent. Apparently, a strong signal of a single migration wave is not inconsistent with subsequent gene flow (Tamm et al., 2007). Although these studies analyzed collections of markers from all over the genome, less attention has been given to individual lineages of single DNA segments with correlated genealogies.

Studies focusing on a single locus may reveal specific aspects of the genetic past of studied populations that might be lost, less clear, or "invisible" in global analysis of many markers taken from throughout the genome. mtDNA and Y-chromosomes are effectively single loci and capture only partially the history of the population. Our study of nuclear DNA segments both complements analyses and challenges these studies.

We focus on the genetic lineages of an 8-kb DNA segment around exon 44, called *dys44*, in the middle of a very large dystrophin gene on Xp21. Its mostly intronic sequence can be expected to evolve neutrally (Nachman and Crowell, 2000). Because of the location in a highly recombining region, far from other genes and other dystrophin exons, we can also assume that its evolution has not been affected by selection acting on neighboring loci. X-chromosome facilitates work with extended haplotypes directly obtained in males. Because of its reduced sensitivity to genetic drift, the X-chromosome covers genetic lineages with greater historical depth than mtDNA or Y-chromosome data (Schaffner, 2004). X-linked markers have additional advantage to be less sensitive to the genetic admixture than the autosomal markers when admixture is mostly male-driven. *Dys44* was investigated earlier in a number of human populations from different continents (Zietkiewicz et al., 1997, 1998, 2003; Xiao et al., 2004; Lovell et al., 2005). This study adds new information and further develops our understanding of the peopling of the Americas.

MATERIALS AND METHODS

Population samples

The data reported in this article consist of 474 newly genotyped X-chromosomes, including 205 Native American chromosomes representing Chipewyan (n = 35), Ojibwa (n = 22), Navajo (n = 24), Algonquian speakers (n = 33), Cree (n = 36), Ache (n = 40), and Ticuna (n = 15) populations as well as Siberian (n = 5), Spanish (n = 22), Basque (n = 21), French (n = 79), Breton (n = 60), German (n = 1), Italian (n = 28), and African American (n = 53) samples. Our analysis included 1,873 chromosomes (474 + 1,399 previously typed chromosomes) representing 36 populations listed in Supporting Information Table S1 and shown on the map in Supporting Information Figure S1. Our North American samples included Navajo from the South-West, Chipewyan from northern Saskatchewan and Cree from south-western Saskatchewan, Ojibwa from Ontario, and Algonquian speakers from northern Québec; Maya are from Central America, whereas Karitiana, Ticuna, and Ache represented South America. Siberian samples were primarily Tundra and Forest Nentsi from North-Western Siberia.

Genotyping

The genomic segment *dys44* consists of exon 44 (148 bp) and its surrounding intronic sequence (positions −2782 to −1 of intron 43 and positions 1–4987 of intron 44) of the human dystrophin gene at Xp21 (Zietkiewicz et al., 1998) (accession number: U94396). Thirty-five simple intronic polymorphisms resulting from 33 nucleotide substitutions (one site is three-allelic), two three-nucleotide deletions, and one eight-nucleotide duplication were previously ascertained by SSCP/heteroduplex analysis of a 7,622 bp within the *dys44* segment in 250 worldwide distributed chromosomes including 40 Native Americans (Zietkiewicz et al., 1997). The ancestral allele was determined by comparison with at least two great ape sequences (Zietkiewicz et al., 1998). In addition, this segment contains a poly-T microsatellite, $(T)_{14–24}$. A 7,917 bp distance between upstream and downstream flanking polymorphic sites determines the length of the haplotype.

Four overlapping DNA fragments covering *dys44*, i.e., an 8-kb DNA segment surrounding exon 44 of the dystrophin gene (Zietkiewicz et al., 1998), were obtained by PCR using the primers: P1F −2824/−2798 GGC AATTACCAATATGATGAGTGATTG; P1R −892/−865 CTAAACTATGCGATGCCTTCTTTAATCC; P1g −892/−865 GGATTAAAGAAGGCATCGCATAGTTTAG; P2R +1164/+1188 CCATTATGCTACATTAGCCCAAAAT; P3F +1164/+1188 ATTTTGGGCTAATGTAGCATAATGG; P3R +3351/+3375 GGAAAGAAATACCTTAACTGTT TGG; P4F +3351/+3375 CCAAACAGTTAAGGTATTTC TTTCC; P4R +5000/+5024 CTACTGGCAACTACAAACA TTTTCC. Amplification was in 20 μl containing 1.5 mM $MgCl_2$, 0.2 mM dNTPs (each), 0.5 μM of each primer, 10–20 ng DNA, and 0.2 U Platinum Taq Polymerase (Invitrogen) in the buffer supplied by the manufacturer. After 7 min at 95°C, the amplification was carried out for 45 cycles (95°C for 30 s, 55°C for 30 s, and 68°C for 2 min) and terminated by 10 min incubation at 68°C. Genotyping of simple segregating sites (Supporting Information Table S2) was performed as described by Zietkiewicz et al. (1997), using a 75–30°C temperature gradient according to the modified dynamic allele-specific oligonucleotide hybridization protocol (Bourgeois and Labuda, 2004). In addition, a $(T)_n$-microsatellite within the *dys44* segment, at position 2684, between ancestral T at position 2663 and eight-nucleotides duplication at position 3133 (Supporting Information Table S2), was typed by denaturing gel electrophoresis.

Haplotypes

In female samples, haplotypes were derived as previously described (Labuda et al., 2000; Zietkiewicz et al.,

2003) assisted by the PHASE software package (Stephens et al., 2001). Note that an upper case "B" in the haplotype name (Supporting Information Table S2) denotes haplotypes that were directly observed at least once (i.e., either in a male sample, or in a homozygous or single heterozygous female), whereas those denoted with a lower case "b" were inferred from a complex genotype. The $(T)_n$ polymorphism is represented in a haplotype's name by an extension describing the length variant, e.g., B001_14 or B001_15 (see Supporting Information Tables S3 and S4) referred to below as "extended" haplotypes.

Statistical analyses

Diversity indices were evaluated using ARLEQUIN v. 3.11 (http://cmpg.unibe.ch/software/arlequin3/) (Excoffier and Schneider, 2005). Gene (haplotype) diversity, which represents the probability that two haplotypes randomly chosen from a sample are different, was estimated as

$$G = [n/n-1)]\left(1 - \sum_{i=1}^{k} p_i^2\right) \quad (1)$$

where n is the sample size, k is the number of distinct haplotypes, and p_i is the frequency of each haplotype. Both k and G can be compared in terms of a common population mutation parameter θ that depends on the effective population size N_e and the rate/generation with which new haplotypes are generated and which here depends both upon mutation and recombination. The indices k and G, and their derived θ_k and θ_G estimates, are expected to be differentially affected by demographic processes causing departures from the mutation-genetic drift equilibrium (migration, founder effect, population growth, or admixture), justifying their joint use in describing population diversity. From the observed k and G, the corresponding θ_k and θ_G are obtained using the formulas

$$E(k) = \theta_k \sum_{i=0}^{n-1} 1/(\theta_k + i) \quad (2)$$

by Ewens (1972) and

$$E(G) = \theta_G/(1 + \theta_G) \quad (3)$$

by Kimura and Crow (1964), respectively, as implemented in the Arlequin Package, where θ_G, referred to as θ_{hom} in the package, is estimated according to Chakraborty and Weiss (1991).

For the timing of founder (colonization) events, we used the proportion P_A of ancestral, i.e., nonmutated or nonrecombined, chromosomes remaining in the sample. Provided that mutations or recombinations are rare, such that recurrent events can be neglected, the proportion $P_A = (1-c)^g$, after g generations, is such that $-\ln P_g = g \cdot c$, and

$$g = \frac{-\ln P_A}{c} \quad (4)$$

where c is either the mutation rate μ or the apparent recombination rate r_{app} that corresponds to the rate of nonsilent recombinations, i.e., those that result in new observable recombinants (Zietkiewicz et al., 2003). Alternatively, assuming star-like phylogeny, we evaluated the age of expansion g as

$$g = \frac{R_N}{c \cdot nR} \quad (5)$$

where R_N is the count of new mutations or recombinations (accumulated since the founder event) and n_R is the number of chromosomes. In the calculation of P_A, we considered the major, highest frequency allele of $(T)_n$-microsatellite to be the ancestral one. In the case of recombinations, as ancestral we considered a set of common haplotypes shared with other continents.

Arlequin was also used to obtain genetic diversity indices, F_{ST} estimates and to carry out the Mantel test for geographic, linguistic, and genetic distances. The language distance matrix was based on the Atlas of languages (Matthews et al., 1996), with a distance between languages directly proportional to the number of levels separating the languages. The Mantel test was performed twice, with increased value when languages belong to separate stocks (Supporting Information Table S5), all yielding similar results (data not shown). We used Matlab R14 (The MathWorks) to carry out principal component analysis (PCA), using haplotype frequencies, as well as multidimensional scaling (MDS), nonmetric stress method, stress value of 0.1, to display results of the pairwise F_{ST} distances (Supporting Information Table S6). In the F_{ST} analysis, negative F_{ST}s given by Arlequin were multiplied by -10^{-6}, i.e., a small negative number, to turn them positive, while keeping them nonsignificantly different from zero, and thus suitable for MDS.

Contour maps were obtained from haplotype frequencies using Surfer 8.02 from Golden Software (http://www.goldensoftware.com/products/surfer/surfer.shtml), using the Kriging gridding method (Yang et al., 2004). Recombination graphs were obtained using Recomb (http://www.iro.umontreal.ca/~mabrouk/) (El-Mabrouk and Labuda, 2004) (for more information, refer to ALFRED: http://alfred.med.yale.edu/alfred/; Ethnologue: http://www.ethnologue.com; Leadmix: http://www.zoo.cam.ac.uk/ioz/software.htm#LEADMIX; ISOGG [Y-DNA Haplogroup Tree]: http://www.isogg.org/tree/index.html.)

The probability of loss of an allele in a population bottleneck, defined by the number of chromosomes and its duration in generations, was computed using PopG 3.0 (http://evolution.gs.washington.edu/popgen/popg.html) by averaging the results of five simulations for each parameter combination.

Admixture analysis was carried out using the 678 autosomal STR dataset reported in Wang et al. (2007), including 123 Africans, 160 Europeans, 212 no-Siberian East Asians, 39 Siberians, and 436 Native Americans including population samples of Chipewyan, Cree, and Ojibwa, which partially overlapped with samples of these populations examined in this study. We used program STRUCTURE (http://pritch.bsd.uchicago.edu/structure.html) (Pritchard et al., 2000) to apply supervised clustering analysis reported by Wang et al. (2007) using his STR dataset. STRUCTURE runs used an admixture model with correlated allele frequencies. A burn-in period of 20,000 iterations was followed by an additional 10,000 iterations from which parameter estimates were obtained. Individuals from Europe, Africa, East Asia (excluding Siberia), and Siberia were forced into separate clusters, and supervised analysis of the Native

TABLE 1. Genetic diversity estimators

Population	*n*	*k*	Gene diversity ± SD	Θ_G ± SD	Θ_k (95% CI Θ_k)	Expected *k*	Chakraborty's *p* (*k* or more alleles)
Ojibwa	52	10	0.78 ± (0.04)	2.73 ± (0.64)	3.41 (1.65, 6.71)	8.69	0.35
Chipewyan	75	5	0.69 ± (0.03)	1.71 ± (0.24)	1.03 (0.40, 2.44)	7.03	0.88
Algonquian	33	5	0.66 ± (0.06)	1.48 ± (0.42)	1.39 (0.52, 3.44)	5.19	0.63
Cree	36	6	0.79 ± (0.04)	2.95 ± (0.70)	1.80 (0.72, 4.16)	8.10	0.89
Navajo	24	5	0.72 ± (0.05)	2.03 ± (0.57)	1.63 (0.59, 4.12)	5.67	0.73
Maya	80	9	0.78 ± (0.02)	2.81 ± (0.39)	2.40 (1.15, 4.72)	10.03	0.71
Ache	40	2	0.30 ± (0.08)	0.31 ± (0.12)	0.26 (0.06, 1.00)	2.20	0.72
Karitiana	83	5	0.52 ± (0.05)	0.80 ± (0.17)	1.00 (0.38, 2.36)	4.30	0.42
Ticuna	15	4	0.70 ± (0.08)	1.74 ± (0.70)	1.43 (0.46, 4.12)	4.44	0.72
Native Americans	438	20	0.75 ± (0.01)	2.31 ± (0.13)	4.18 (2.58, 6.56)	12.67	0.02
Kazakh	54	10	0.75 ± (0.04)	2.37 ± (0.59)	3.34 (1.63, 6.56)	8.03	0.25
Bait	29	7	0.80 ± (0.04)	3.35 ± (0.95)	2.61 (1.09, 5.90)	8.07	0.77
Derbet	31	8	0.74 ± (0.06)	2.15 ± (0.73)	3.16 (1.38, 6.87)	6.38	0.27
Kalkha	25	6	0.75 ± (0.06)	2.28 ± (0.81)	2.18 (0.86, 5.22)	6.16	0.62
Khoton	30	7	0.66 ± (0.07)	1.46 ± (0.50)	2.56 (1.07, 5.77)	5.00	0.19
Myangad	30	8	0.78 ± (0.05)	2.77 ± (0.94)	3.23 (1.41, 7.04)	7.33	0.45
Olet	30	8	0.80 ± (0.06)	3.27 ± (1.29)	3.23 (1.41, 7.04)	8.06	0.59
Uriankhai	32	6	0.75 ± (0.05)	2.35 ± (0.68)	1.91 (0.76, 4.45)	6.81	0.73
Zakhchin	29	7	0.70 ± (0.07)	1.80 ± (0.61)	2.61 (1.09, 5.90)	5.63	0.30
Mongolians	290	17	0.75 ± (0.02)	2.34 ± (0.23)	3.79 (2.23, 6.17)	11.81	0.06
Hiri	18	5	0.67 ± (0.08)	1.52 ± (0.59)	1.93 (0.69, 5.08)	4.39	0.45
Alores	16	4	0.68 ± (0.09)	1.58 ± (0.64)	1.37 (0.44, 3.92)	4.31	0.69
Flores	27	9	0.81 ± (0.05)	3.31 ± (1.18)	4.31 (1.92, 9.33)	7.80	0.36
Roti	17	3	0.63 ± (0.07)	1.30 ± (0.38)	0.77 (0.22, 2.46)	3.96	0.84
Ternate	13	3	0.62 ± (0.08)	1.21 ± (0.41)	0.89 (0.25, 2.92)	3.50	0.76
Indonesian Isl.	91	12	0.69 ± (0.03)	1.67 ± (0.27)	3.48 (1.82, 6.36)	7.25	0.04
Chinese	103	9	0.66 ± (0.03)	1.44 ± (0.20)	2.19 (1.06, 4.27)	6.72	0.20
Japanese	65	7	0.57 ± (0.05)	0.99 ± (0.22)	1.79 (0.78, 3.81)	4.72	0.15
PNG	67	12	0.67 ± (0.06)	1.51 ± (0.44)	4.00 (2.07, 7.42)	6.30	0.01
Siberian	53	9	0.81 ± (0.04)	3.29 ± (0.80)	2.86 (1.35, 5.76)	9.84	0.70
Asians	669	34	0.73 ± (0.01)	2.10 ± (0.13)	7.42 (5.11, 10.57)	12.63	<10-5
Spanish	55	14	0.81 ± (0.04)	3.38 ± (0.97)	5.73 (3.04, 10.48)	10.13	0.09
Basque	41	12	0.85 ± (0.03)	4.49 ± (1.22)	5.34 (2.67, 10.35)	10.87	0.39
French	129	13	0.81 ± (0.02)	3.33 ± (0.49)	3.41 (1.85, 6.03)	12.78	0.52
Breton	151	14	0.82 ± (0.02)	3.71 ± (0.48)	3.58 (1.99, 6.20)	14.34	0.59
German	73	14	0.80 ± (0.03)	3.18 ± (0.66)	4.88 (2.63, 8.72)	10.60	0.14
Italian	54	10	0.73 ± (0.04)	2.10 ± (0.50)	3.34 (1.63, 6.56)	7.43	0.17
Western Europeans	503	36	0.80 ± (0.01)	3.28 ± (0.25)	8.72 (6.03,12.36)	17.05	<10-5
African American	120	43	0.94 ± (0.01)	14.69 ± (3.34)	23.56 (15.91, 34.58)	32.99	0.02
Biaka	85	16	0.81 ± (0.03)	3.41 ± (0.75)	5.56 (3.12, 9.59)	11.60	0.08
Mbuti	58	11	0.76 ± (0.05)	2.43 ± (0.74)	3.75 (1.88, 7.17)	8.32	0.17
SubSaharan Africans	263	54	0.91 ± (0.01)	8.41 ± (1.14)	20.31 (17.69, 27.77)	29.71	<10-5

Letter *n* denotes the number of chromosomes analyzed for each population and *k* the number of distinct haplotypes (disregarding *T*-repeat variability); two estimators of theta, θ_G and θ_k are based on haplotype heterozygosity and *k*, respectively. Note that we were not looking for new mutations, but genotyped known polymorphisms, so that new haplotypes result from recombination events. However, the mutation rate and the recombination rate are of the same order within the haplotype. Chakraborty's test (1990) gives the probability *p* to observe *k* or more haplotypes in a population, given its haplotype heterozygosity.

American data was performed with $K = 5$ clusters. The same analysis was applied on the 29 X-chromosome STR dataset reported in Wang et al. (2008) to which we added new data on the three Native populations of interest, Chipewyan, Cree, and Ojibwa, using the same sample sets as in Wang et al. (2007) (Supporting Information Table S7).

RESULTS

Diversity and geographic distribution of Dys-44 haplotypes

In a worldwide sample of 1,873 X-chromosomes, we find 98 *dys44* haplotypes, consisting of 35 segregating sites, which were previously ascertained in a worldwide population sample (Table 1 and Supporting Information Table S2). These haplotypes can be extended by including the allelic status of its imbedded $(T)_n$-microsatellite (see Materials and Methods) (Supporting Information Table S3). Allelic variation of the $(T)_n$-microsatellite is bimodal, with two frequency peaks at allele $(T)_{15}$ and allele $(T)_{22}$ (Zietkiewicz et al., 2003). Usually only one of the length modes predominates against the background of a given haplotype. For example, 81 copies of the frequent haplotype B005 represent the form B005_22. Two rare variants B005_21 and B005_23 can be explained by a stepwise mutation because of a single deletion or addition starting from the ancestral $(T)_{22}$ allele. In contrast, B005_15 is most likely due to a recombination with another haplotype carrying $(T)_{15}$, such as a common haplotype B001_15 (581 copies, Supporting Information Table S3). Likewise, crossovers can also account for B001_22 and B001_24 singletons given the low probability of a stepwise mutation of size 7 or 9. For simplicity, in the description later, we will define haplotype $(T)_n$

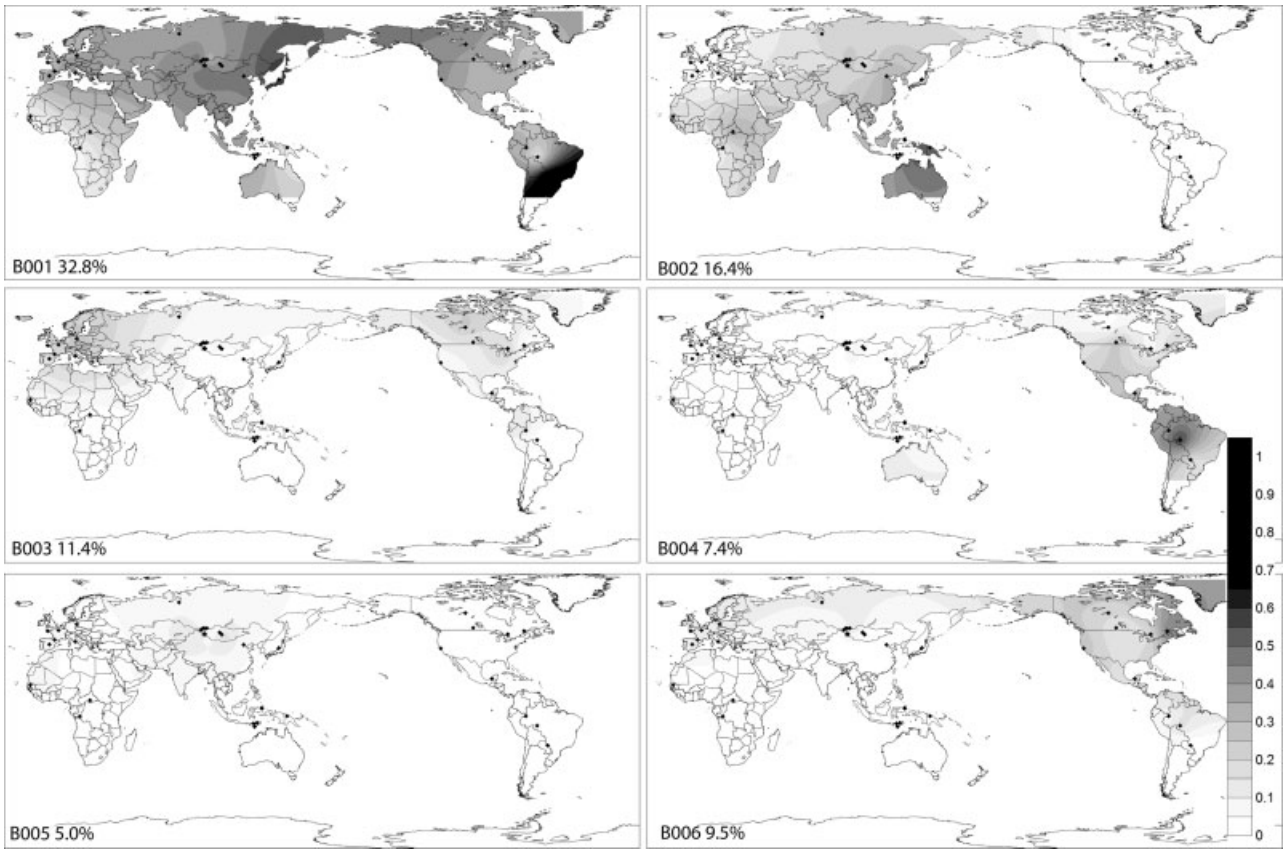

Fig. 1. Distribution of *dys44* haplotypes across continents. Contour maps have been obtained using frequencies of haplotypes B001–B006, all remaining haplotypes being grouped under the denomination "Other" (see Supporting Information Tables S2 and S3, and Fig. S3), in all populations. The same scale used for all maps is included on the right side, from zero frequency to one. Beside the haplotype name, its worldwide average frequency is indicated.

status only when relevant for the analysis. New haplotypes that are created by recombination and/or gene conversion usually result from crossovers between the most frequent haplotypes (see later). Except for $(T)_n$, no new haplotype-creating mutations are observed because the genotyping is limited to previously ascertained polymorphic sites (Zietkiewicz et al., 1997).

Worldwide, there are six common ($\geq\sim$5%) haplotypes, B001–B006, representing 83% of all X-chromosomes in our sample. They differ in their continental and intercontinental distribution (Figs. 1 and 2 and Supporting Information Fig. S1). B001 is the most frequent *dys44* haplotype outside of Africa ($\sim$37%). Its frequency is markedly increased in Japan (60%) and decreased in Papua New Guinea (PNG) (7.5%). In the Americas, only Karitiana (13.3%) and Aché (82.5%) differ substantially in their B001 frequency relative to the continental average (34%). B003 is another ubiquitous haplotype with an overall frequency of 23.7% in Western Europeans, 4.6% in Africans and 5.1% in Asians (but 11.3% in Siberians and 10% in Olet, Mongolia). In the Americas (11.2%) it is unevenly distributed: absent in Aché and Karitiana, at 5% in Navajo and Algonquian speakers, up to 21.2% in Ojibwa and 26.7% in Chipewyan populations.

Two haplotypes, B004 and B006, are characteristic of the Americas. B004, representing 27.6% of the Native American chromosomes, is rarely seen elsewhere. It is almost absent from Europe (a single chromosome in Basques), whereas one copy found in African Americans is presumably due to admixture. In Asians (2.2%) its occurrence is uneven, from 0 to 11.8% (Roti). Its relatively high frequency in Mongolians (Olet: 10%; Zakchin: 6.9%) and Siberians (3.8%) is suggestive of the Northern Asiatic origin of the Native American B004. In the Americas, B004 increases in frequency from north to south (15.3–43.5%), with intermediate frequency in Maya (31.3%). It occurs at 4% in Chipewyan and 5.8% in Ojibwa, up to 46.7% in Ticuna and 66.3% in Karitiana, thus displaying a north to south cline, like B003 above, but in the opposite direction (see Fig. 1). Haplotype B006 (overall 20.1%) also shows a frequency gradient similar to that of B003 (see Fig. 1), decreasing in frequency from 27.3% in native North Americans (51.5% in Algonquian speakers, 29.2% in Navajo) to 20% in Maya, down to overall 8.2% in native South Americans: 18.1% in Karitiana, absent from Aché and at 6.7% in Ticuna.

The bulk of the American B006 is represented by B006_16, but in three northern populations, Algonquian, Chipewyan, and Cree, the allelic B006_15 contributes a significant proportion (up to 40%) of this haplotype. With a $(T)_n$ mutation rate estimated at about 2.3×10^{-5} per generation (Zietkiewicz et al., 2003) all haplotypes within each of the two allelic groups, B006_15 and B006_16, can be considered identical by descent. B006 is relatively abundant in Europeans (1.0% B006_15 and 11.5% B006_16) and in Siberians where B006_16 also predominates (1.9 and 9.4%, respectively). It is found in Mongolians (5.5%), where, in contrast, 12 of 16 of its

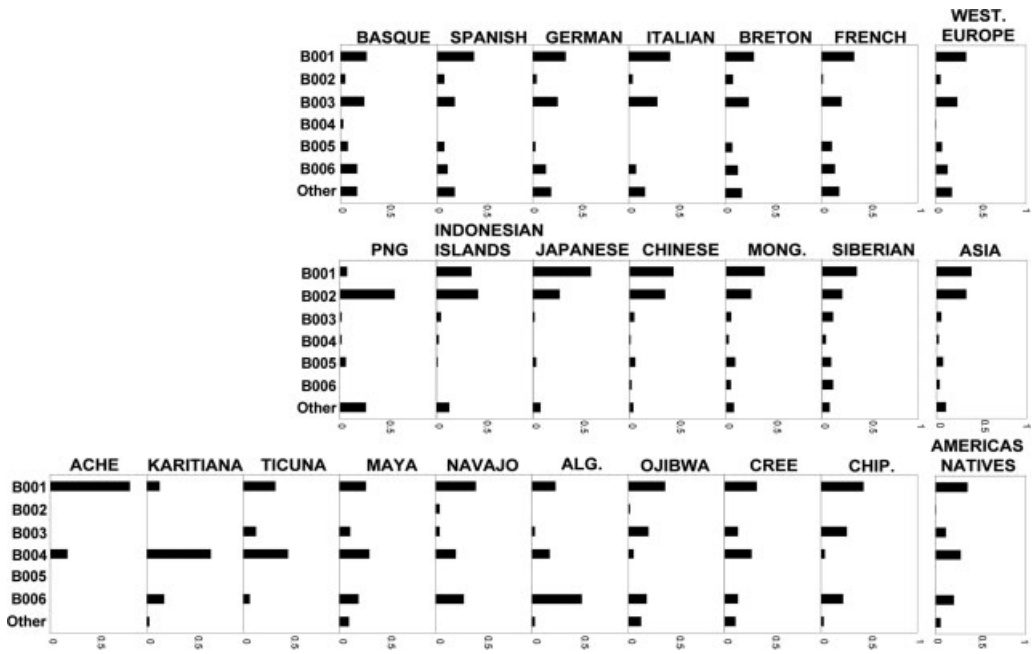

Fig. 2. Haplotypes distribution at the population level. Bar charts represent frequencies of haplotypes B001–B006 in all populations. All remaining haplotypes were grouped under the denomination "Other" (see Supporting Information Tables S2 and S3) Abbreviations: PNG, Papua New Guinea; MONG, Mongolians; ALG, Algonquians; CHIP, Chipewyans.

copies are due to B006_15. The distribution of B006 points to the Euro-Asiatic and Siberian connections of the early Native Americans. The occurrence of B006_15 restricted to North Americans may reflect its preferential loss in the southern populations. Alternatively, it could be due to its subsequent gain through more recent Asiatic contacts that would have preferentially influenced the northern populations (Bortolini et al., 2003).

Two other common haplotypes, B002 and B005, are virtually absent from Native Americans. B002 is found in one copy in Navajo and another was found in Ojibwa (overall 2/438). It occurs at a 5% level in Europeans and is very frequent in Asians, carried by one-third of the Asiatic chromosomes (33%). It is found at 16.7% in Olet and 45.2% in Derbet, Mongolia, 20.8% in Siberians, 42.9% in Indonesian Island populations, and 56.7% in people from PNG. B002 is also relatively frequent in Africans, where, in contrast to the non-African B002_15, it is associated with a broader spectrum of $(T)_n$-microsatellite alleles. Likewise, there is no haplotype B005 in Native Americans in spite of its substantial presence in Africans (5%), Europeans (overall 6.8% but notably 10.9% in French), and Asians (7%), where it is highly represented in Khalkha (20%) and Kazakh (14.8%). Of note, there is a third, rarer haplotype B008, but equally common in Europeans (5.2%) and in Mongolians (7%), which is also almost (one copy in Ojibwa) absent from the Americas.

In three Native American populations, Navajo, Aché, and the Ticuna, there are no other haplotypes besides the six common B001–B006 dealt with above. In contrast, in Ojibwa, Maya, and Cree more than 10% of chromosomes carry other haplotypes. The rare haplotypes B013, B016, B019, B024, B030, B035, b052, B053, and b086 are shared either with Europeans or Asians or both (Supporting Information Table S3 and data not shown), where they are equally rare. In addition, there are b064, b066, and b091 seen only in Americans. Recombination analysis (El-Mabrouk and Labuda, 2004) and direct inspection of Supporting Information Table S2 suggest that all these haplotypes originated from recombination and/or gene conversion events between common haplotypes, B001–B006, B008 or their derivatives. Some of the rare haplotypes enumerated above have a high probability of recurrent origin (see Appendix in Zietkiewicz et al., 2003) and thus, though identical by state, when seen on different continents could be considered endemic. In the Americas, this is likely to be the case for B019, B024, B035, b052, B053, and b086. Counting these, together with b064, b066, and b091, as de novo recombinants leads to the proportion of chromosomes specific to Native Americans of 3.0%.

Time estimates

Except for the B006 haplotype with its ubiquitous two variants B006-15 and B006_16, three other major haplotypes, B001_15, B003_15, and B004_15, were most likely introduced to the Americas as monomorphic in their $(T)_n$-repeats. Using the present proportion of these haplotypes carrying mutated alleles T_{14} and T_{16} (6/317 or 1.9%) and taking T_{15} as ancestral, we can evaluate [Eq. (4)] the age of these chromosomes in the Americas. On the basis of the mutation rate μ of 2.3×10^{-5} per generation (Zietkiewicz et al., 2003), we obtain 830 generations or 23,000 years. Here, we used a generation time of 28 years that averages the relative contribution of male and female meioses to the X-chromosome transmission (Yotova et al., 2007), based on the male and the female generation time of 31.5 and 25.6 years, respectively (Fenner, 2005). An independent age estimate can be obtained from the proportion of new recombinant haplotypes (13/438 or 3%), which were described in the previous section. On the basis of the apparent recombi-

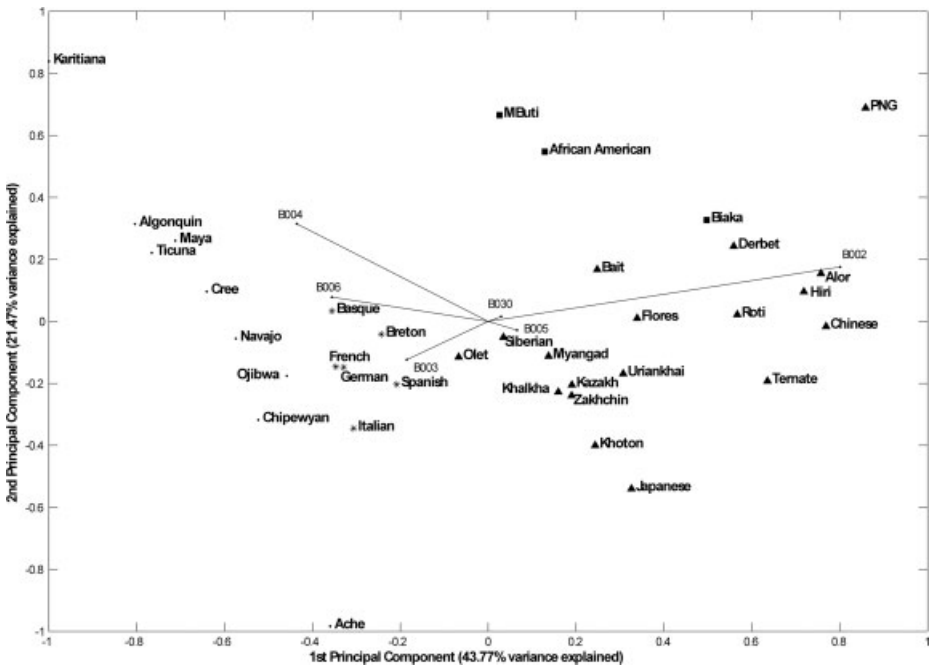

Fig. 3. Principal component analysis based on the frequencies of the haplotypes. Only the two first components are shown, representing a total of 64.96% of the variance. The six haplotypes contributing the most to the first component are represented as vectors, all originating at the center of the figure, whose lengths indicate how influential they were in defining the component, pointing toward the populations with the highest frequency of the haplotype.

nation rate of 3.2×10^{-5} per generation (Zietkiewicz et al., 2003), we obtained the age of the founder effect at 940 generations corresponding to about 26,000 years. In turn, assuming a star-like phylogeny [Eq. (5)] and the observation of two (2/317) distinct $(T)_n$ mutations, or nine (9/438) distinct recombinations, the age of the American founder effect can be evaluated, applying the same rates, at 275 generations (7,700 years) and 640 generations (16,500 years), respectively.

Population diversity and genetic structure

There is less genetic diversity in Native Americans than in Europeans and Africans. The haplotype diversity, G (Table 1), declines from sub-Saharan Africans (0.91) to Europeans (0.80), down to Asians (0.73) and Native Americans (0.75). The same trend is observed in the number of haplotypes, k, and population parameters Θ_k and Θ_G. When the two estimates of Θ are compared, they do not differ significantly at the level of individual populations. Here as well, the number of observed haplotypes does not exceed the expected number of haplotypes (Table 1). This is no longer the case when individual populations are considered together, as subpopulations of the continental groups or the entire world. Pooling of continental subpopulations leads to a significant excess in the number of the observed haplotypes that reflects the difference in the corresponding estimates of Θ_k and Θ_G. Interestingly, this effect, which can be explained by population amalgamation (Chakraborty, 1990), is least pronounced in Native Americans who display the highest population variance. In the Americas, the F_{ST} is 0.14 and compares with F_{ST} of 0.04 in Asia, 0.08 in Africa and 0 in Europe. The overall world F_{ST} is 0.13, whereas the average differentiation of populations within continents, F_{sc}, is 0.05 and that among the continental groups, F_{ct}, is 0.08. Clearly, genetic drift appears to have played a more important role in the genetic diversification of the American populations than it did on other continents where, in contrast, we observe a greater accumulation of new haplotypes that disproportionately inflate Θ_k with respect to Θ_G in the pooled samples (Table 1, Fig. 2; Supporting Information Figs. S2 and S3). It suggests that American populations are relatively younger and/or of smaller size and that their meridional distribution was less conducive to gene flow than the latitudinal locations of Eurasiatic populations.

Genetic relations among populations

PCA analysis based on haplotype frequencies (see Fig. 3) reveals a very similar pattern to that of the analysis of pairwise F_{ST}s displayed by MDS (Supporting Information Fig. S4). Native Americans form a widespread group with Karitiana and Aché as outliers. The cluster of populations closest to Native Americans consists of Europeans along with Siberians and Mongolian populations. Other Asian populations appear further apart, including outlier PNG, and are accompanied by African populations additionally removed along a second principal component axis. The first principal component accounts for 43% and the second for 22% of the haplotype frequency variance. The haplotypes contributing to the first component are as follows: i) B002, B005, and B030, with an increase in frequency from the center to the right, and ii) B003, B004, and B006, increasing in frequency from the center to the left of the plot, and responsible for European with Native American clustering (see Fig. 3). B001 does not significantly contribute to the distribution as its frequency is very similar among different populations.

DISCUSSION

To address the question of the genetic origins of Native Americans, we compiled data on 438 X-chromosomes from nine Native American populations (205 newly typed), genotyped for 35 biallelic markers and the $(T)_n$-microsatellite of the *dys44* segment. They were analyzed in the context of 1,435 X-chromosome samples (269

newly typed) representing 27 populations from other continents. When analyzed within a worldwide context (Fig. 3 and Supporting Information Fig. S4), Native Americans make a widespread but separate cluster close to Western Europeans that are followed by Northern and Central Asians, represented here by Siberians and Mongolians. This analysis reveals a deep ancestral connection between genetic lineages leading to Europe, Siberia, and the Americas. In addition to the most ubiquitous haplotype B001, West Eurasian, Siberians, and Amerindian populations retained high frequencies of haplotypes B003 and B006 (Figs. 1 and 2, and Supporting Information Tables S3 and S4). In contrast, the otherwise common haplotypes B005 and B002—especially frequent in Asia—are not present in the Americas. Importantly, the overall frequency of haplotype B006 in the Americas is almost twice that of Europe, whereas the frequency of B003 is half of that in Europe (though in Chipewayans it is even more frequent than in Europe). Haplotype frequencies that are higher or comparable with those found in Europe cannot be accounted for by the effect of post-Colombian admixture, particularly if these haplotypes are found throughout the Americas, North and South. Although such high haplotype frequencies do not exclude the possibility that Europeans could have contributed some copies of certain haplotypes, they do show that European influence did not play a dominant role in shaping current Native American genetic diversity. A signature of such admixture is seen in our data, but it is relatively weak.

We observed one copy of B002_16 in Navajo, where $(T)_{16}$ allele suggests African-American origin, and one copy of B002_15 in Ojibwa of presumably recent European origin. B002_15 chromosomes were also observed in admixed Mexicans and Colombians (Bedoya et al., 2006) (data not shown). A single copy of the haplotype B008 in Ojibwa could be due to European admixture as well. In contrast, no B005 haplotypes were observed in our Amerindian sample. This may be surprising, given the frequency of 6.7% of B005 in Europe (10.9% in France) and reported European admixture levels in Native North Americans of 17% (Zegura et al., 2004) or more (Bolnick et al., 2006). There are two plausible explanations to our observations.

First is that European admixture was principally male-driven as shown earlier (Mesa et al., 2000; Bedoya et al., 2006; Wang et al., 2008). This would lead to 1:1 and 1:3 relative admixture ratios for Y- and X-chromosomes, respectively, and to twice greater admixture levels on autosomes than on the X-chromosome. For example, in Ojibwa only two non-Native mtDNA sequences out of 35 (5.7%) analyzed were reported by Scozzari et al. (1997), consistent with such a scenario. We also combined the data on dys44 haplotypes with data from a linked, highly informative, compound interrupted microsatellite, DXS1238, localized 7 kb downstream of the *dys44* segment. DXS1238 was characterized recently (Yotova et al., 2007) in a subset of samples analyzed here that included 76 X-chromosomes (14 Cree, 22 Navajo, 24 Chipewayan, and 16 Karitiana) of which 27 were haplotype B001, 11 haplotype B003, and 17 haplotype B006. In one case, the extended haplotype B001-DXS1238 was identical by state with an extended haplotype only found in Eastern and Central Asia, three B001 chromosomes shared two such haplotypes with the Middle-East, and another three carried two ubiquitous haplotypes observed all over the world, including Europe. Out of 11 copies of Native American B003, nine shared the same DXS1238 haplotype, and only one chromosome shared the extended haplotype with European and Middle-Eastern samples. Moreover, half of the West Eurasiatic B003 haplotypes are linked to the DXS1238 haplogroup 0000×2Y that was not observed among our Native American samples. Likewise, in 13 samples of European descent B006 was found linked to nine distinct DXS1238 haplotypes, whereas our Native American B006 was linked with just three DXS1238 haplotypes, sharing only one of them with West-Eurasians. These data indicate again that post-Columbian admixture at the level of chromosome X is relatively modest.

Second might be the fact that northern Native populations show a considerable genetic affinity to Siberians that may mimic the effect of admixture and inflate the estimates of European admixture evaluated using classical autosomal markers (Field et al., 1988; Cavalli-Sforza et al., 1994). To revisit these possibilities, we used the datasets of Wang et al. (2007, 2008) enriched with 29 X-chromosome STR markers in Chipewyan, Cree, and Ojibwa (see Materials and Methods). We evaluated admixture levels in these three populations that were also analyzed in this article (Supporting Information Table S7). We find that the level of European admixture of 5.2, 24.4, and 18.4% evaluated using autosomal STR markers for Chipewayan, Cree, and Ojibwa sample goes down to 3.1, 6.8, and 6.9%, respectively, when the analysis is done for the X-chromosome markers. The level of "Siberian admixture" is much greater from 33 to 62% on autosomes down to about 30% on the X-chromosome. There are certain caveats here. We do not know the proportion of European or related alleles that are present in Siberians and could confound the analysis, especially given that our sample lacks populations from eastern Siberia. Because of genetic drift, which is stronger on the X-chromosome than on autosomes, X-chromosomes would also differentiate faster. In contrast, this long-term process is not expected to affect the outcomes of recent admixture. Finally, when compared with autosomal dataset, the number of X-chromosome STR markers was small. Nevertheless, this analysis suggests a possibility of a considerable Siberian influence that may confound autosomal estimates of post-Colombian European admixture in northern Native populations. Importantly, it is consistent with earlier findings pointing to preferentially male-driven admixture. This explains our observation of a relatively modest admixture effect in dys44 locus and suggests X-chromosome markers as a model of choice in studies of Native populations.

A dramatic shift in haplotype frequencies in the Americas relative to other continents includes a prominent rise in the haplotype B004, an increase in the frequency of B006, and the disappearance of B002 and B005 (Figs. 1 and 2). This shift can be explained by a founder effect because of the small number of the initial migrants. In a population bottleneck any rare allele can serendipitously rise to high frequencies. B004, although not exclusively American, can be compared to other frequent variants which define Native American lineages, such as the 275 allele of microsatellite D9S1120, the mitochondrial 16111T allele, or the Y-chromosome DYS199T allele (Schroeder et al., 2007). B004 differs from B002 by the three downstream alleles (Supporting Information Table S2), which suggests its likely origin in a recombination between haplotypes B002 and B001 (or any other haplotype sharing the downstream alleles with B001). The

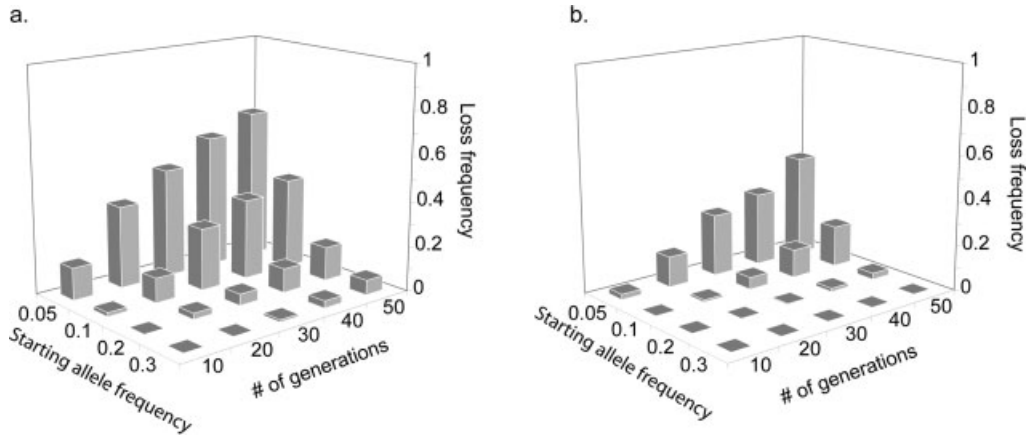

Fig. 4. Probability of loss of an allele given population bottleneck (in number of chromosomes) and its duration (in generation) 3D bar charts representing the results of simulations using PopG 3.0 for: (**a**) population of 100 diploid individuals; (**b**) population of 200 diploid individuals. The fitness of all genotypes was left to 1, with no possible mutation and no migration rate between populations, leaving only genetic drift as evolution force. Two hundred populations were evolved at a time (the maximum allowed by the software), replicated five times, for a total of 1,000 populations evolving during 10–50 generations (*y*-axis), with a starting minor allele frequency from 5 to 30% (*x*-axis). The frequency of loss of the minor allele was averaged between simulations and is reported in the figures on the *z*-axis.

same exchange between B001 and B002 but retaining the "complementary" recombinant leads to haplotype b054, found in a single copy in Japan (Supporting Information Tables S2 and S3). The absence of B002 in the Americas implies that B004 was imported presumably by the first migrants along with similarly frequent haplotypes B001, B003, and B006. The probability of the loss of an allele of frequency p after sampling n chromosomes can be evaluated as $P_{\text{loss}} = (1 - p)^n$ (for two haplotypes p would represent their cumulative frequency). In Figure 4, we calculated such probabilities in populations of 100 and 200 chromosomes (i.e., census size of about 200 and 400 individuals, respectively, considering X-chromosomes) for alleles of different initial frequencies and for a time period of up to 50 generations, i.e., about 1,400 years. In a population of 100 chromosomes, after 50 generations there is a 66% probability of losing an allele initially present at 5% (and 46% in a population of 200 chromosomes). Considering the selective neutrality of our haplotype (Zietkiewicz et al., 2003; Yotova et al., 2007), it suggests that the population bottleneck related to the peopling of the Americas was extremely severe (Hey, 2005) or lasted much longer (Tamm et al., 2007; Gonzalez-Jose et al., 2008) to account for the loss of B002 and B005, whose cumulative frequencies in Eurasian populations usually exceed 10%. A single population bottleneck seems also more likely, otherwise both B002 and B005 would have to be jointly lost twice. These findings argue in favor of a peopling scenario with a single major migration wave (Merriwether et al., 1995; Forster et al., 1996; Kolman et al., 1996; Merriwether and Ferrell, 1996; Underhill et al., 1996; Bonatto and Salzano, 1997; Bianchi et al., 1998; Stone and Stoneking, 1998; Santos et al., 1999; Silva et al., 2002; Hey, 2005; Wang et al., 2007; Fagundes et al., 2008). According to the archeological and environmental evidence, this could have occurred around 15,000 years ago, immediately after deglaciation of the Pacific coastal corridor (Goebel et al., 2008). Our crude estimates of the age of the founder effect of between 26,000 and 7,700 years ago fall within this time range in agreement with the genetic estimates based on mtDNA and Y-chromosome diversity (Bortolini et al., 2003; Zegura et al., 2004; Tamm et al., 2007; Achilli et al., 2008; Endicott and Ho, 2008; Fagundes et al., 2008). Although their accuracy can be disputed, it shows the potential of genetic systems other than mtDNA and the Y-chromosome to time past evolutionary events. Time estimates are important for evolutionary and historical inferences using population genetics data. However, they depend critically upon the population sampled as well as the mutation and recombination models, rate calibrations, and other necessary simplifying assumptions, such that the results of such inferences are often debatable (Fagundes et al., 2008; Ho and Endicott, 2008). Although we used very simple and robust population models [Eqs. (4) and (5)], all such estimates are associated with a large variance. In addition, the mutation and recombination rates we used to infer time are themselves estimates based on other data sets and represent point estimates over a range of values that may be different at different parts of the sequence segment. Large stochastic errors, particularly affecting our lowest estimate of 7,700 years, will eventually be improved by increasing the amount data available to us. Ultimately, different genetic systems representing different genetic models should provide congruent time estimates.

More than one migration wave was originally postulated to explain the introduction of distinct linguistic families, such as NaDene and Amerind, into the Americas (Greenberg, 1987; Ruhlen, 1996). In our analysis, the Mantel test did not reveal any correlation between *dys44* haplotype frequencies and the groups of Na-Dene or Amerind speakers (value 0.10, $P = 0.35$), corroborating other findings (Hunley and Long, 2005; Wang et al., 2007; Malhi et al., 2008). It yielded a significant correlation between these haplotype frequencies and geography (value 0.41, $P = 5.6 \times 10^{-3}$). In the Americas, F_{ST} analysis reveals greater differentiation and more genetic structure when compared with Europe and other continents. This can be a consequence of the relatively small population sizes of the Native American groups. In spite of their common origin from Beringia and low genetic diversity when compared with other continents (Wang et al., 2007), their distribution over all climatic zones

separated by numerous geographic barriers was conducive to genetic diversification, also noted in geographically less dispersed samples (Torroni et al., 1993; Mesa et al., 2000; Bortolini et al., 2002). At the same time, our data reveal noticeable genetic differences between the south and the north, which cannot be solely ascribed to a more general effect of the loss of diversity by genetic drift during southward colonization movements (Torroni et al., 1993; Lorenz and Smith, 1997; Malhi et al., 2001; Ramachandran et al., 2005; Wang et al., 2007). We notice that the two lineages represented by haplotypes B003 and B006 that are responsible for the affinity between Native American, Siberian, and European samples observed in PCA and MDS plots (Fig. 3 and Supporting Information Fig. S4) are concentrated in northern populations (Figs. 1 and 2). Furthermore, northern populations are enriched in the variant B006-15, which predominates among contemporary Central Asians (Supporting Information Table S3), whose distribution seems to overlap with that of the Y-chromosome haplogroup C3b (Karafet et al., 1999; Bortolini et al., 2003; Zegura et al., 2004; Malhi et al., 2008). Recently, Hellenthal et al. (2008), based on the novel analysis of a set of data from 32 autosomal regions (Conrad et al., 2006), proposed that the sources of ancestry differed between Native South and North Americans. Accordingly, this could explain the relationship between distance from Siberia and greater genetic similarity to Siberians of northern populations in the data of Wang et al. (2007), where this similarity was originally attributed to serial dilution (Ramachandran et al., 2005). Although, as discussed earlier, our data are consistent with uneven ancestry of southern and northern populations as proposed by Hellenthal et al. (2008), they also point to a strong population bottleneck (Schroeder et al., 2007) A series of founder effects resulting in a genetically structured population of Beringia (Tamm et al., 2007; Achilli et al., 2008) as a source of subsequent southern colonization(s) could provide a plausible explanation for this apparent discrepancy. The analysis of X-linked lineages of Arctic populations could help testing this hypothesis. In this context, recent analysis of Ray et al. (submitted) using the data of Wang et al. (2007) is very revealing by suggesting a model in which an initial major migration wave is followed by recurrent gene flow between Asia and America. This could explain differences between southern and northern populations in spite of a strong signal of a bottleneck and single major migration wave observed with most of the genetic data. Another view on this issue was recently offered by Perego et al. (2009). They provided mtDNA evidence of two separate migrations from Beringia, which could explain the different distributions of haplogroup X2a in Northern and Southern Native American populations, thus further substantiating our conclusions based on X-chromosome diversity.

ACKNOWLEDGMENTS

The authors are indebted to all the individuals who provided samples for analysis and to various colleagues who generously shared their DNA collections. They thank C. Piché, A. Lovell, F. Xio, D. Gehl, and R. Hamon for their help with the experiments and data. S. Bourgeois was a bursary of the Canadian Institutes of Health Research.

LITERATURE CITED

Achilli A, Perego UA, Bravi CM, Coble MD, Kong QP, Woodward SR, Salas A, Torroni A, Bandelt HJ. 2008. The phylogeny of the four pan-American MtDNA haplogroups: implications for evolutionary and disease studies. PLoS One 3:e1764.

Bedoya G, Montoya P, Garcia J, Soto I, Bourgeois S, Carvajal L, Labuda D, Alvarez V, Ospina J, Hedrick PW, Ruiz-Linares A. 2006. Admixture dynamics in Hispanics: a shift in the nuclear genetic ancestry of a South American population isolate. Proc Natl Acad Sci USA 103:7234–7239.

Bianchi NO, Catanesi CI, Bailliet G, Martinez-Marignac VL, Bravi CM, Vidal-Rioja LB, Herrera RJ, Lopez-Camelo JS. 1998. Characterization of ancestral and derived Y-chromosome haplotypes of New World native populations. Am J Hum Genet 63:1862–1871.

Bolnick DA, Bolnick DI, Smith DG. 2006. Asymmetric male and female genetic histories among Native Americans from Eastern North America. Mol Biol Evol 23:2161–2174.

Bonatto SL, Salzano FM. 1997. Diversity and age of the four major mtDNA haplogroups, and their implications for the peopling of the New World. Am J Hum Genet 61:1413–1423.

Bortolini MC, Salzano FM, Bau CH, Layrisse Z, Petzl-Erler ML, Tsuneto LT, Hill K, Hurtado AM, Castro-De-Guerra D, Bedoya G, Ruiz-Linares A. 2002. Y-chromosome biallelic polymorphisms and Native American population structure. Ann Hum Genet 66 (Part 4):255–259.

Bortolini MC, Salzano FM, Thomas MG, Stuart S, Nasanen SP, Bau CH, Hutz MH, Layrisse Z, Petzl-Erler ML, Tsuneto LT, Hill K. Hurtado AM, Castro-de-Guerra D, Torres MM, Groot H, Michalski R, Nymadawa P, Bedoya G, Bradman N, Labuda D, Ruiz-Linares A. 2003. Y-chromosome evidence for differing ancient demographic histories in the Americas. Am J Hum Genet 73:524–539.

Bourgeois S, Labuda D. 2004. Dynamic allele-specific oligonucleotide hybridization on solid support. Anal Biochem 324: 309–311.

Brown MD, Hosseini SH, Torroni A, Bandelt HJ, Allen JC, Schurr TG, Scozzari R, Cruciani F, Wallace DC. 1998. mtDNA haplogroup X: an ancient link between Europe/Western Asia and North America? Am J Hum Genet 63:1852–1861.

Cavalli-Sforza LL, Menozzi P, Piazza A. 1994. The history and geography of human genes. Princeton, NJ: Princeton University Press.

Chakraborty R. 1990. Mitochondrial DNA polymorphism reveals hidden heterogeneity within some Asian populations. Am J Hum Genet 47:87–94.

Chakraborty R, Weiss KM. 1991. Genetic variation of the mitochondrial DNA genome in American Indians is at mutation-drift equilibrium. Am J Phys Anthropol 86:497–506.

Conrad DF, Jakobsson M, Coop G, Wen X, Wall JD, Rosenberg NA, Pritchard JK. 2006. A worldwide survey of haplotype variation and linkage disequilibrium in the human genome. Nat Genet 38:1251–1260.

El-Mabrouk N, Labuda D. 2004. Haplotypes histories as pathways of recombinations. Bioinformatics 20:1836–1841.

Endicott P, Ho SY. 2008. A Bayesian evaluation of human mitochondrial substitution rates. Am J Hum Genet 82:895–902.

Ewens WJ. 1972. The sampling theory of selectively neutral alleles. Theor Popul Biol 3:87–112.

Excoffier LGL, Schneider S. 2005. Arlequin ver. 3.0: an integrated software package for population genetics data analysis. Evol Bioinf Online 1:47–50.

Fagundes NJ, Kanitz R, Eckert R, Valls AC, Bogo MR, Salzano FM, Smith DG, Silva WA Jr, Zago MA, Ribeiro-dos-Santos AK, Santos SE, Petzl-Erler ML, Bonatto SL. 2008. Mitochondrial population genomics supports a single pre-Clovis origin with a coastal route for the peopling of the Americas. Am J Hum Genet 82:583–592.

Fenner JN. 2005. Cross-cultural estimation of the human generation interval for use in genetics-based population divergence studies. Estimation of fecundability from survey data. Am J Phys Anthropol 128:415–423.

Field LL, Gofton JP, Kinsella TD. 1988. Immunoglobulin (GM and KM) allotypes and relation to population history in native peoples of British Columbia: Haida and Bella Coola. Am J Phys Anthropol 76:155–163.

Forster P, Harding R, Torroni A, Bandelt HJ. 1996. Origin and evolution of Native American mtDNA variation: a reappraisal. Am J Hum Genet 59:935–945.

Goebel T, Waters MR, O'Rourke DH. 2008. The late Pleistocene dispersal of modern humans in the Americas. Science 319: 1497–1502.

Gonzalez-Jose R, Bortolini MC, Santos FR, Bonatto SL. 2008. The peopling of America: craniofacial shape variation on a continental scale and its interpretation from an interdisciplinary view. Am J Phys Anthropol 137:175–187.

Greenberg JH. 1987. Language in the Americas. Stanford: Stanford University Press.

Hellenthal G, Auton A, Falush D. 2008. Inferring human colonization history using a copying model. PLoS Genet 4:e1000078.

Hey J. 2005. On the number of New World founders: a population genetic portrait of the peopling of the Americas. PLoS Biol 3:e193.

Ho SY, Endicott P. 2008. The crucial role of calibration in molecular date estimates for the peopling of the Americas. Am J Hum Genet 83:142–146; author reply 146-147.

Hunley K, Long JC. 2005. Gene flow across linguistic boundaries in Native North American populations. Proc Natl Acad Sci USA 102:1312–1317.

Jones M. 2003. Ancient DNA in pre-Columbian archaeology: a review. J Archaeol Sci 30:629–635.

Karafet TM, Zegura SL, Posukh O, Osipova L, Bergen A, Long J, Goldman D, Klitz W, Harihara S, de Knijff P, Wiebe V, Griffiths RC, Templeton AR, Hammer MF. 1999. Ancestral Asian source(s) of new world Y-chromosome founder haplotypes. Am J Hum Genet 64:817–831.

Kimura M, Crow JF. 1964. The number of alleles that can be maintained in a finite population. Genetics 49:725–738.

Kolman CJ, Sambuughin N, Bermingham E. 1996. Mitochondrial DNA analysis of Mongolian populations and implications for the origin of New World founders. Genetics 142:1321–1334.

Labuda D, Zietkiewicz E, Yotova V. 2000. Archaic lineages in the history of modern humans. Genetics 156:799–808.

Lorenz JG, Smith DG. 1997. Distribution of sequence variation in the mtDNA control region of Native North Americans. Hum Biol 69:749–776.

Lovell A, Moreau C, Yotova V, Xiao F, Bourgeois S, Gehl D, Bertranpetit J, Schurr E, Labuda D. 2005. Ethiopia: between Sub-Saharan Africa and western Eurasia. Ann Hum Genet 69 (Part 3):275–287.

Malhi RS, Eshleman JA, Greenberg JA, Weiss DA, Schultz Shook BA, Kaestle FA, Lorenz JG, Kemp BM, Johnson JR, Smith DG. 2002. The structure of diversity within New World mitochondrial DNA haplogroups: implications for the prehistory of North America. Am J Hum Genet 70:905–919.

Malhi RS, Gonzalez-Oliver A, Schroeder KB, Kemp BM, Greenberg JA, Dobrowski SZ, Smith DG, Resendez A, Karafet T, Hammer M, Zegura S, Brovko T. 2008. Distribution of Y chromosomes among native North Americans: a study of Athapaskan population history. Am J Phys Anthropol 137: 412–424.

Malhi RS, Schultz BA, Smith DG. 2001. Distribution of mitochondrial DNA lineages among Native American tribes of Northeastern North America. Hum Biol 73:17–55.

Mateus Pereira LH, Socorro A, Fernandez I, Masleh M, Vidal D, Bianchi NO, Bonatto SL, Salzano FM, Herrera RJ. 2005. Phylogenetic information in polymorphic L1 and Alu insertions from East Asians and Native American populations. Am J Phys Anthropol 128:171–184.

Matthews S, Polinsky M, Comrie B. 1996. The atlas of languages: the origin and development of languages throughout the world. New York: Facts on File.224 p.

Merriwether DA, Ferrell RE. 1996. The four founding lineage hypothesis for the New World: a critical reevaluation. Mol Phylogenet Evol 5:241–246.

Merriwether DA, Rothhammer F, Ferrell RE. 1995. Distribution of the four founding lineage haplotypes in Native Americans suggests a single wave of migration for the New World. Am J Phys Anthropol 98:411–430.

Mesa NR, Mondragon MC, Soto ID, Parra MV, Duque C, Ortiz-Barrientos D, Garcia LF, Velez ID, Bravo ML, Munera JG, Bedoya G, Bortolini MC, Ruiz-Linares A. 2000. Autosomal, mtDNA, and Y-chromosome diversity in Amerinds: pre- and post-Columbian patterns of gene flow in South America. Am J Hum Genet 67:1277–1286.

Nachman MW, Crowell SL. 2000. Contrasting evolutionary histories of two introns of the duchenne muscular dystrophy gene. Dmd, in humans. Genetics 155:1855–1864.

Perego UA, Achilli A, Angerhofer N, Accetturo M, Pala M, Olivieri A, Kashani BH, Ritchie KH, Scozzari R, Kong QP, Myres NM, Salas A, Semino O, Bandelt HJ, Woodward SR, Torroni A. 2009. Distinctive Paleo-Indian migration routes from Beringia marked by two rare mtDNA haplogroups. Curr Biol 19:1–8.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945–959.

Ramachandran S, Deshpande O, Roseman CC, Rosenberg NA, Feldman MW, Cavalli-Sforza LL. 2005. Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. Proc Natl Acad Sci USA 102:15942–15947.

Ruhlen M. 1996. The origin of language: tracing the evolution of the mother tongue. New York: Wiley. 239 p.

Salzano FM. 2002. Molecular variability in Amerindians: widespread but uneven information. An Acad Bras Cienc 74:223–263.

Santos FR, Pandya A, Tyler-Smith C, Pena SD, Schanfield M, Leonard WR, Osipova L, Crawford MH, Mitchell RJ. 1999. The central Siberian origin for native American Y chromosomes. Am J Hum Genet 64:619–628.

Schaffner SF. 2004. The X chromosome in population genetics. Nat Rev Genet 5:43–51.

Schroeder KB, Schurr TG, Long JC, Rosenberg NA, Crawford MH, Tarskaia LA, Osipova LP, Zhadanov SI, Smith DG. 2007. A private allele ubiquitous in the Americas. Biol Lett 3:218–223.

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.

Silva WA Jr, Bonatto SL, Holanda AJ, Ribeiro-Dos-Santos AK, Paixao BM, Goldman GH, Abe-Sandes K, Rodriguez-Delfin L, Barbosa M, Paco-Larson ML, Petzl-Erler ML, Valente V, Santos SE, Zago MA. 2002. Mitochondrial genome diversity of Native Americans supports a single early entry of founder populations into America. Am J Hum Genet 71:187–192.

Stephens M, Smith NJ, Donnelly P. 2001. A new statistical method for haplotype reconstruction from population data. Am J Hum Genet 68:978–989.

Stone AC, Stoneking M. 1998. mtDNA analysis of a prehistoric Oneota population: implications for the peopling of the New World. Am J Hum Genet 62:1153–1170.

Tamm E, Kivisild T, Reidla M, Metspalu M, Smith DG, Mulligan CJ, Bravi CM, Rickards O, Martinez-Labarga C, Khusnutdinova EK, Fedorova SA, Golubenko MV, Stepanov VA, Gubina MA, Zhadanov SI, Ossipova LP, Damba L, Voevoda MI, Dipierri JE, Villems R, Malhi RS. 2007. Beringian standstill and spread of Native American founders. PLoS One 2:e829.

Torroni A, Schurr TG, Cabell MF, Brown MD, Neel JV, Larsen M, Smith DG, Vullo CM, Wallace DC. 1993. Asian affinities and continental radiation of the four founding Native American mtDNAs. Am J Hum Genet 53:563–590.

Torroni A, Schurr TG, Yang CC, Szathmary EJ, Williams RC, Schanfield MS, Troup GA, Knowler WC, Lawrence DN, Weiss KM, Wallace, DC. 1992. Native American mitochondrial DNA analysis indicates that the Amerind and the Nadene populations were founded by two independent migrations. Genetics 130:153–162.

Underhill PA, Jin L, Zemans R, Oefner PJ, Cavalli-Sforza LL. 1996. A pre-Columbian Y chromosome-specific transition and its implications for human evolutionary history. Proc Natl Acad Sci USA 93:196–200.

Wang S, Lewis CM, Jakobsson M, Ramachandran S, Ray N, Bedoya G, Rojas W, Parra MV, Molina JA, Gallo C, Mazzotti G, Poletti G, Hill K, Hurtado AM, Labuda D, Klitz W, Barrantes R, Bortolini MC, Salzano FM, Petzl-Erler ML, Tsuneto LT, Llop E, Rothhammer F, Excoffier L, Feldman MW, Rosenberg NA, Ruiz-Linares A. 2007. Genetic variation and population structure in native Americans. PLoS Genet 3:e185.

Wang S, Ray N, Rojas W, Parra MV, Bedoya G, Gallo C, Poletti G, Mazzotti G, Hill K, Hurtado AM, Camrena B, Nicolini H, Klitz W, Barrantes R, Molina JA, Freimer NB, Bortolini MC, Salzano FM, Petzl-Erler ML, Tsuneto LT, Dipierri JE, Alfaro EL, Bailliet G, Bianchi NO, Llop E, Rothhammer F, Excoffier L, Ruiz-Linares A, Lewis CM, Jakobsson M, Ramachandran S, Labuda D, Feldman MW, Rosenberg NA. 2008. Geographic patterns of genome admixture in Latin American Mestizos. PLoS Genet 4:e1000037.

Xiao FX, Yotova V, Zietkiewicz E, Lovell A, Gehl D, Bourgeois S, Moreau C, Spanaki C, Plaitakis A, Moisan JP, Labuda D. 2004. Human X-chromosomal lineages in Europe reveal Middle Eastern and Asiatic contacts. Eur J Hum Genet 12:301–311.

Yang CS, Kao SP, Lee FB, Hung PS. 2004. Twelve different interpolation methods: a case study of Surfer 8.0. Proceedings of the XXth ISPRS Congress, Vol. 35. p 778–785.

Yotova V, Lefebvre JF, Kohany O, Jurka J, Michalski R, Modiano D, Utermann G, Williams SM, Labuda D. 2007. Tracing genetic history of modern humans using X-chromosome lineages. Hum Genet 122:431–443.

Zegura SL, Karafet TM, Zhivotovsky LA, Hammer MF. 2004. High-resolution SNPs and microsatellite haplotypes point to a single, recent entry of Native American Y chromosomes into the Americas. Mol Biol Evol 21:164–175.

Zietkiewicz E, Yotova V, Gehl D, Wambach T, Arrieta I, Batzer M, Cole DE, Hechtman P, Kaplan F, Modiano D, Moisan JP, Michalski R, Labuda D. 2003. Haplotypes in the dystrophin DNA segment point to a mosaic origin of modern human diversity. Am J Hum Genet 73:994–1015.

Zietkiewicz E, Yotova V, Jarnik M, Korab-Laskowska M, Kidd KK, Modiano D, Scozzari R, Stoneking M, Tishkoff S, Batzer M, Labuda D. 1997. Nuclear DNA diversity in worldwide distributed human populations. Gene 205:161–171.

Zietkiewicz E, Yotova V, Jarnik M, Korab-Laskowska M, Kidd KK, Modiano D, Scozzari R, Stoneking M, Tishkoff S, Batzer M, Labuda D. 1998. Genetic structure of the ancestral population of modern humans. J Mol Evol 47:146–155.