Contrasting Patterns of Nuclear and mtDNA Diversity in Native American Populations

Ning Ning Yang[1*], Stephane Mazières[2*], Claudio Bravi[3], Nicolas Ray[4], Sijia Wang[5], Mari-Wyn Burley[1], Gabriel Bedoya[6], Winston Rojas[6], Maria V. Parra[6], Julio A. Molina[7], Carla Gallo[8], Giovanni Poletti[8], Kim Hill[9], Ana M. Hurtado[9], Maria L. Petzl-Erler[10], Luiza T. Tsuneto[10], William Klitz[11], Ramiro Barrantes[12], Elena Llop[13], Francisco Rothhammer[13], Damian Labuda[14], Francisco M. Salzano[15], Maria-Cátira Bortolini[15], Laurent Excoffier[16,17], Jean Michel Dugoujon[2] and Andrés Ruiz-Linares[1$]

[1]*Department of Genetics, Evolution and Environment, University College London, UK*
[2]*Laboratoire AMIS, Anthropologie Moléculaire et Imagerie de Synthèse, CNRS FRE 2960/Université Paul Sabatier, Toulouse, France*
[3]*Instituto Multidisciplinario de Biología Celular, La Plata, Argentina*
[4]*EnviroSPACE Laboratory, Institute for Environmental Sciences, University of Geneva, Switzerland*
[5]*Harvard FAS Center for Systems Biology, Cambridge, MA*
[6]*Laboratorio de Genética Molecular, Universidad de Antioquia, Medellín, Colombia*
[7]*Centro de Investigaciones Biomédicas de Guatemala, Ciudad de Guatemala, Guatemala*
[8]*Laboratorios de Investigación y Desarrollo, Facultad de Ciencias y Filosofía, Universidad Peruana Cayetano Heredia, Lima, Perú*
[9]*School of Human Evolution and Social Change, Arizona State University, Tempe, AZ*
[10]*Departamento de Genética, Universidade Federal do Paraná, Curitiba, Brazil*
[11]*University of California, School of Public Health, Berkeley, CA and Public Health Institute, Oakland, USA*
[12]*Escuela de Biología, Universidad de Costa Rica, San José, Costa Rica*
[13]*Facultad de Medicina, Universidad de Chile, and Instituto de Alta Investigación, Universidad de Tarapacá, Chile*
[14]*CHU Sainte-Justine, Département de Pédiatrie, Université de Montréal, Montréal, Canada*
[15]*Departamento de Genética, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil*
[16]*Computational and Molecular Population Genetics Laboratory, Institute of Ecology and Evolution, University of Berne, Switzerland*
[17]*Swiss Institute of Bioinformatics, Lausanne, Switzerland*

Summary

We report an integrated analysis of nuclear (autosomal, X– and Y-chromosome) short tandem repeat (STR) data and mtDNA D-loop sequences obtained in the same set of 22 Native populations from across the Americas. A north to south gradient of decreasing population diversity was observed, in agreement with a settlement of the Americas from the extreme northwest of the continent. This correlation is stronger with "least cost distances," which consider the coasts as facilitators of migration. Continent-wide estimates of population structure are highest for the Y-chromosome and lowest for the autosomes, consistent with the effective size of the different marker systems examined. Population differentiation is highest in East South America and lowest in Meso America and the Andean region. Regional analyses suggest a deviation from mutation–drift equilibrium consistent with population expansion in Meso America and the Andes and population contraction in Northwest and East South America. These data hint at an early divergence of Andean and non–Andean South Americans and at a contrasting demographic history for populations from these regions.

Keywords: Ameriindians, mtDNA, Y-chromosome, X-chromosome, STRs

*Joint first authors
$Corresponding author: Andrés Ruiz-Linares, Department of Genetics, Evolution and Environment, University College London, UK. Tel: +44–20-3108–4003; Fax: +44–20-7679–5052; E-mail: a.ruizlin@ucl.ac.uk

Introduction

Contrasting patterns of diversity between different genetic marker systems can provide refined insights into aspects of human population evolution. For instance, such analyses have been used to evaluate possible differences in migration rates

between men and women throughout human evolution (reviewed in Wilkins, 2006 and Segurel et al., 2008). In another application, a recent study of autosomal and X chromosome data has led to the proposal of a more severe male than female bottleneck during the initial migration of anatomically modern humans out of Africa (Keinan et al., 2009). In the Americas, continent-wide Native population surveys have been carried out independently with mtDNA (Torroni et al., 1993; Merriwether et al., 1995; Forster et al., 1996; Bonatto & Salzano, 1997; Fagundes et al., 2008), Y-chromosome (Lell et al., 2002; Bortolini et al., 2003), autosomal (Salzano & Callegari-Jacques, 1988; Cavalli-Sforza et al., 1994; Wang et al., 2007) and, more recently, X-chromosome markers (Bourgeois et al., 2009; Wang et al., 2010). However, other than regional studies (Mesa et al., 2000) or continental studies combining data for different population samples (Bortolini et al., 2003) there has not been a systematic, continent-wide, analysis of genetic diversity across marker systems on the same Native American population samples.

We recently performed a genome-wide survey of autosomal diversity in the Americas with data for 678 STRs typed in 22 Native populations from North, Central, and South America (Wang et al., 2007). Here we expand that study with data for 38 X-chromosome STRs, Y-chromosome markers (16 STRs and three biallelic polymorphisms), and mtDNA D-loop sequences obtained in the same population samples. These data emphasize the existence of a north to south gradient in population genetic diversity, consistent with the serial founder model of human population expansion (Ramachandran et al., 2005) and with a colonization of the Americas from the extreme northwest of the continent; the coasts acting as facilitators during this process. Contrasting diversity across marker systems points to a differing demographic history across South America, with Andean populations diverging early in the settlement of the region and having larger long-term effective sizes than non-Andean populations.

Methods

Samples

A total of 396 individuals from 22 Native American populations were examined (Fig. 1). These were the same individuals and populations examined in Wang et al. (2007), including 114 males and 282 females. For several analyses these populations were divided into five groups based mainly on geographic/linguistic grounds: North America (Cree and Ojibwa), Meso America (Mixtec, Mixe, Zapotec, and Kaqchikel), Northwest South America (including Lower Central America: Embera, Guaymi, Arhuaco, Zenu, Wayuu, Waunana, Kogi, and Cabecar),

the Andes (Inga, Quechua, Huilliche, and Aymara), and East South America (Aché, Guaraní, Ticuna, and Kaingang).

mtDNA Sequencing

DNA amplification and sequencing were based on the methods described in Brandstatter et al. (2004). The entire mtDNA control region (D-loop) was amplified with primers: F15878 (5′-AAATGGGCCTGTCCTTGTAG-3′) and R 599 (5′-TTG AGG AGG TAA GCT ACA TA-3′). Each 20 μl PCR reaction consisted of 2 μl Bioline 10× reaction buffer, 0.8 μl 50 mM MgCl$_2$, 0.4 μl 10 μM dNTP mix, 0.4 μl 10 μM F/R primers, and 0.4 μl BiolineTaq (Bioline Ltd., London, UK). Thermal cycler conditions were: 95°C for 10 min and then 36 cycles of 94°C for 1 min, 56°C for 1 min, and 72°C for 2 min. PCR products were purified and sequenced with BigDye^{TM} Terminator v3.1 cycle sequencing kit (Applied BioSystems, Foster City, CA). Each sequencing reaction, made up to a total volume of 11 μl, contained 2.07 μl of 5× sequencing buffer, 0.25 μl of BigDye^{TM} Terminator v3.1 ready reaction mix, 0.18 μl of the sequencing primer, and 2 μl of purified PCR product. Sequencing was carried out with 25 cycles of 96°C for 10 sec, 55°C for 5 sec, and 60°C for 4 min. Both forward and reverse strands were sequenced, providing at least two independent readings for each nucleotide position. Primers used for sequencing included the two initial PCR primers as well as primers: F 16190 (5′-CCC CAT GCT TAC AAG CAA GT-3′), F15 (5′-CAC CCT ATT AAC CAC TCA CG-3′), R274 (5′-TGT GTG GAA AGT GGC TGT GC-3′), R16400 (5′-GTC AAG GGA CCC CTA TCT GA-3′), and R16175 (5′-TGG ATT GGG TTT TTA TGT A-3′). After raw data editing, the nucleotide sequences were analyzed from position 16021 to 16569 and from position 1 to 499 of the mtDNA control region (D-loop). All sequence coordinates used followed the revised Cambridge Reference Sequence (rCRS) (Anderson et al., 1981; Chinnery et al., 1999).

X-Chromosome STRs

The following 38 X chromosome STRs were genotyped by the Marshfield Mammalian Genotyping Service (http://research.marshfieldclinic.org/genetics/home/index.asp): GATA52B03, GATA124B04, AGAT144, AFMA184WF1P, GATA175D03, TATC052, ATA28C05, GATA124E07, GATA186D06, ATCT057M, GATA027M, AFMA124XD9, GATA69C12, GATA144D04, AFM276XF5P, GATA72E05M, AGAT104M, GATA31D10M, GATA31F01P, TAGA017, ATA31E12, GATA10C11, GATA172D05, GATA48H04, AFM248WE5, GATA165B12P, GATA198A10P, ATCT003, AFMA046WB9, AAAT112P, GATA31E08, TATC043, 224ZG11, AFMA121ZE5, TTTA062, GGAT3F08, GATA189B04P, and ATA71D03M.
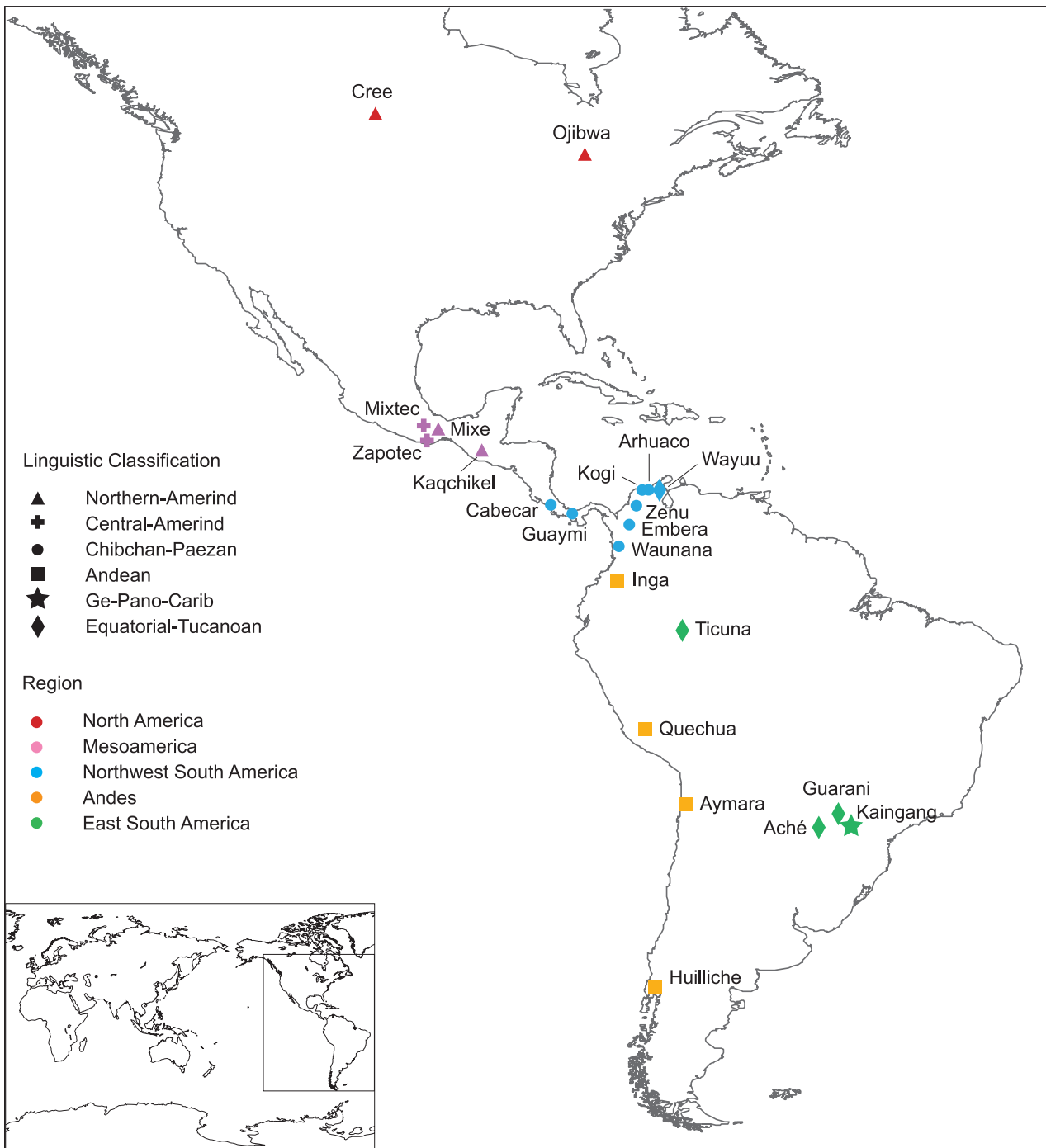
Figure 1 Approximate location of the sampling site for the populations examined here. Color codes indicate for each population its affiliation to a major Native American linguistic stock, following the classification of Ruhlen (Ruhlen, 1991).

Y-Chromosome Markers

For male individuals, three Y-SNPs (M3, M19, and M242) were genotyped as described previously (Karafet et al., 1999; Ruiz-Linares et al., 1999; Seielstad et al., 2003). In addition, data for 18 Y STRs were also obtained. The following seven markers were genotyped at the Marshfield Mammalian Genotyping Service: DYS19, DYS388, DYS389 (GATA30F10L, GATA30F10LA),

DYS390, DYS391, and DYS395. An additional 11 markers were genotyped in two sets by multiplex PCR. The first set comprised markers: DYS447, DYS448, DYS450, DYS456, and DYS458 and the second set comprised markers: DYS437, DYS438, DYS439, DYS426, DYS460 (A7.1), and H4. PCRs were performed using the Hotstart Taq system (Promega, London, UK), at total volumes of 20 μl (for the 5-plex panel) or 20.6 μl (for the 6-plex panel). For the 5-plex panel, each PCR reaction consisted of 10 μl Hotstart Taq Master Mix, 8 μl Primer mix (final concentration for each primer was 0.4 μM), 1 μl water, and 1μl template DNA. For the 6-plex panel, each PCR reaction consisted of 10 μl Hotstart Taq Master Mix, 9.6 μl Primer mix, and 1μl genomic DNA. After denaturing for 10 min at 95°C, amplification was carried out with 28 cycles of: 94°C for 1 min, 55°C for 1 min, then 72°C for 1 min. The sequence of the PCR primers used for each marker is provided in Table S1.

Data Analysis

Population diversity, structure and demography
The average unbiased gene diversity across loci for each population and each geographic region was computed using ARLEQUIN v3.1 (Excoffier et al., 2005). The correlation between the ranks of genetic diversities calculated from the different marker systems was evaluated with Kendall's coefficient of concordance (Siegel, 1956). Continental and regional F_{ST} estimates were obtained using POWERMARKER (Liu & Muse, 2005) with standard errors calculated by bootstrap and significance of differences between estimates evaluated using a *t*-test. STRUCTURE 2.2 was used to evaluate clustering of individuals with the Bayesian approach of Pritchard et al. (2000) and Falush et al. (2003) and employing the same analysis parameters as in Wang et al. (2007). Results were displayed using DISTRUCT 1.1 (Rosenberg, 2004). Based on the X-STR data, a principal component analysis (PCA) was performed, using the SPSS package, on the pairwise population distance matrix, calculated with POWERMARKER, using the D_A distance (Nei & Chesser, 1983). Fu's *Fs* test was applied to the mtDNA D-loop sequence data using ARLEQUIN v3.1 (Excoffier et al., 2005). The program BOTTLENECK (Cornuet & Luikart, 1996) was run on the X-STR data with 1000 iterations per locus.

Geographic computations
Geographic coordinates for the populations were taken from Wang et al. (2007). Initially, distances between populations were computed using simple great arc routes, with obligatory way-points as specified in Wang et al. (2007). In addition, least-cost distances between the Bering Strait and population locations were computed using PATHMATRIX (Ray, 2005) following the approach described in detail in Wang et al. (2007). Briefly, these distances are based on least-cost paths computed on the basis of a spatial cost map incorporating landscape components. For instance, at a coastal/inland relative cost of 1:1 (i.e., a "uniform" cost) the only spatial constraints are the boundaries of the continental landmasses. Since we wanted to evaluate the role of

coastlines as corridors of migration, we also compared the following coastal/inland relative costs: 1:2, 1:5, 1:10, 1:20, 1:30, 1:40, 1:50, 1:100, 1:200, 1:300, 1:400, and 1:500. Computations were performed on a Lambert azimuthal equal area projection of the American landmass (central meridian 80°W, reference latitude 10°N) divided into a grid of 100 km² square cells. Pearson's correlation coefficient (*r*) was estimated between the levels of gene diversity (mean expected heterozygosity (He)) and the great arc (or least cost) population distances from the Bering Strait.

Results

Within-Population Diversity

Table 1 shows the genetic diversity for the different marker systems tested in the set of Native American populations examined here and by Wang et al. (2007). A positive correlation is observed for all pairwise comparisons across marker systems, with the strongest correlation seen between autosomal and X-chromosome STRs (Kendall's tau = 0.77) and the weakest between mtDNA and Y-chromosome STRs (Kendall's tau = 0.01). A negative correlation (*r* = −0.43) of population genetic diversity with (great arc) distance from the Bering Strait has been previously reported for the autosomal STRs (Wang et al., 2007). Diversity at the additional marker systems examined here shows evidence of the same trend, with similarly negative correlations with distance from the Bering Strait being observed for X-STRs and mtDNA: ($r^{(X)} = -0.413$; $r^{(mtDNA)} = -0.421$). A somewhat lower negative correlation is observed for Y-STRs ($r = -0.128$).

Consistent with the analyses of autosomal STRs (Wang et al., 2007), the geographic distance-genetic diversity correlations increase for mtDNA and the X-chromosome when using least-cost paths that consider the coasts as facilitators of migration (Fig. 2). Interestingly, the highest correlation is seen for a coastal/inland cost ratio similar to the one observed previously with the autosomal STR dataset (∼1:10) (Wang et al., 2007).

Table 2 contrasts the mean genetic diversity across markers in different parts of the Americas. As expected, based on differences in ploidy levels, diversity is highest for the autosomes, intermediate for the X-chromosome, and lowest for the Y-chromosome. When comparing the different geographic regions, Eastern South America consistently shows the lowest genetic diversity across all markers systems.

Population Structure

Continent-wide F_{ST} estimates considering the set of 15 populations for which data were available for at least four Y-chromosomes are: $F_{ST}^{(A)} = 0.068$, $F_{ST}^{(X)} = 0.094$, $F_{ST}^{(mtDNA)} = 0.256$, and $F_{ST}^{(Y)} = 0.390$ (all pairwise comparisons are

Table 1 Native American population diversity estimated with nuclear STRs and mtDNA D-loop sequences

Population	mtDNA D-Loop			678 A-STRs			38 X-STRs			18 Y-STRs		
	Sample size (N)	Nucl div	SE	Sample size ($2N$)	H_A	SE	Sample Size (N)	H_X	SE	Sample Size (N)	H_Y	SE
Aché	11	0.0015	0.0011	38	0.482	0.205	14	0.442	0.206	9	0.068	0.232
Arhuaco	16	0.0035	0.0021	34	0.619	0.145	13	0.502	0.168	6	0.333	0.000
Aymara	17	0.0090	0.0048	36	0.661	0.132	12	0.634	0.128	9	0.438	0.266
Cabecar	15	0.0093	0.0050	40	0.622	0.146	14	0.609	0.155	9	0.332	0.244
Cree	11	0.0104	0.0058	36	0.695	0.115	10	0.670	0.109	2	N/A	N/A
Embera	9	0.0053	0.0032	22	0.616	0.155	16	0.576	0.171	4	0.667	0.000
Guaraní	8	0.0056	0.0034	20	0.644	0.144	18	0.606	0.178	1	N/A	N/A
Guaymi	13	0.0082	0.0045	36	0.583	0.174	13	0.523	0.196	9	0.395	0.306
Huilliche	20	0.0093	0.0050	40	0.667	0.119	15	0.638	0.134	7	0.257	0.247
Inga	16	0.0092	0.0050	34	0.640	0.140	12	0.595	0.154	3	N/A	N/A
Kaingang	2	N/A	N/A	14	0.623	0.178	6	0.543	0.180	1	N/A	N/A
Kaqchikel	16	0.0087	0.0047	24	0.662	0.138	14	0.686	0.120	4	0.287	0.298
Kogi	16	0.0081	0.0044	34	0.560	0.175	12	0.517	0.179	0	N/A	N/A
Mixe	20	0.0101	0.0054	40	0.642	0.136	16	0.631	0.117	7	0.105	0.180
Mixtec	17	0.0096	0.0051	40	0.646	0.141	15	0.639	0.152	6	0.272	0.282
Ojibwa	16	0.0095	0.0051	40	0.689	0.115	17	0.645	0.152	4	0.407	0.262
Quechua	18	0.0098	0.0052	40	0.671	0.123	18	0.646	0.132	1	N/A	N/A
Ticuna	12	0.0092	0.0051	70	0.603	0.156	27	0.554	0.161	16	0.293	0.256
Waunana	18	0.0080	0.0043	40	0.610	0.156	14	0.559	0.148	4	0.213	0.277
Wayuu	18	0.0114	0.0060	34	0.670	0.125	7	0.653	0.158	7	0.560	0.288
Zapotec	19	0.0114	0.0060	38	0.668	0.138	17	0.658	0.105	1	N/A	N/A
Zenu	15	0.0054	0.0031	36	0.639	0.142	13	0.585	0.152	4	0.343	0.291

N/A not estimated due to sample size.

significantly different at p < 0.05). The corresponding regional F_{ST} values are shown in Table 3. In all regions F_{ST} is highest for the Y-chromosome followed by mtDNA. In Northwest and East South America $F_{ST}^{(X)}$ is higher than $F_{ST}^{(A)}$, while in the Andes and Meso America the opposite is observed. Comparing regions, the highest population differentiation for all types of markers is seen in Eastern South America. Northwest South America also shows an increased differentiation, relative to Meso America and the Andes (except for the Y-chromosome, which shows a higher differentiation in Meso America).

The X-chromosome STR data were used to examine population structure with the model-based approach implemented in the program STRUCTURE (Pritchard et al., 2000; Falush et al., 2003). In a continent-wide analysis (Fig. 3), the model with $K = 4$ identifies two Eastern South American populations (Ticuna and Aché) and the Northwest South American populations as separate clusters. Increasing K to 5 identifies an additional Northwest South American component (predominant in the Kogi and Arhuaco). At $K = 6$, a North American component is apparent, which is also slightly predominant in Meso Americans relative to other populations. In addition, the component that predominates in Andeans is

also seen at a relatively high frequency in two Northwestern populations (Wayuu and Zenu) and in two of the Eastern South American populations (Guaraní and Kaingang). Previous STRUCTURE analysis of autosomal STRs produced similar patterns (Wang et al., 2007), the least variable groups forming distinctive clusters (particularly for East and Northwest South Americas). Autosomal STRs show somewhat less evidence of a differentiation between Meso Americans and Andeans than seen with X-STRs. An analysis restricted to the South American populations shows similar patterns (Fig. 4). At $K = 3$, predominant components are identified for the Eastern Aché and the Northwestern Kogi and Arhuaco. Increasing K to 4 identified the Eastern Ticuna as a separate cluster while at $K = 5$ an Andean component is defined, which is also predominant in the Northwestern Wayuu and Zenu, and in the Eastern Guarani and Kaingang.

Population Relatedness

Principal component analysis (PCA) was used to examine the relatedness of populations based on the X-chromosome STR data. Figure 5A shows results excluding two East South
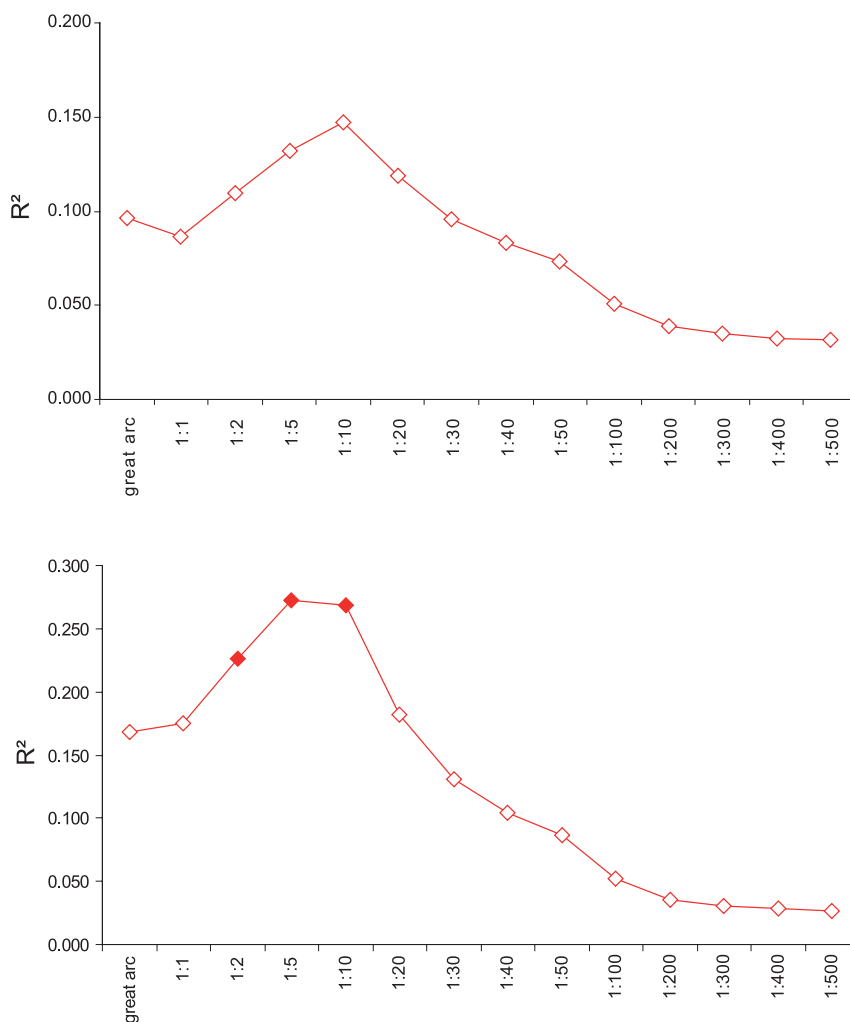
Figure 2 R^2 (square of the correlation) between population diversity (estimated as gene diversity for X-STRs– top and as nucleotide diversity for mtDNA– bottom) and distance from the Bering Strait. Correlations were evaluated for great arc distances as well as for least-cost distances at a range of coastal/inland relative costs (1:1 to 1:500). Correlations significant at the 0.05 level are indicated by filled symbols.

American populations (the Ache and Kaingang) that represent extreme outliers (Figure S1 shows results for the full set of 22 populations). Figure 5B shows results after excluding four outliers seen in Figure 5A (the Northwest South American: Arhuaco, Kogi, Embera, and Guaymi). These PCA plots are consistent with the population diversity and structure analyses presented above in showing greatest differentiation among East and Northwest South American populations. Populations from these regions cluster at opposite ends relative to the two North American populations (Cree and Ojibwa), with the Andean and Meso-American populations occupying intermediate positions. Figure 5B shows a closer relatedness of Eastern and Northwest South American populations to the

Andeans than to Meso Americans. Similar broad patterns of relatedness were observed in the phylogenetic tree reported by (Wang et al., 2007) based on the autosomal STR data collected in these same population samples.

Population Demography

We used three independent approaches to evaluate changes in population size with the different marker data obtained here. First, we applied tests based on the allele frequency distribution of X-STRs, as implemented in the program BOTTLENECK (Cornuet & Luikart, 1996). This program implements a test for departure from the expected relationship,

Table 2 Genetic diversity across regions in the Americas estimated with different marker systems

	A-STRs			X-STRs			Y-STRs			mtDNA		
	N	H_A	SE	N	H_X	SE	N	H_Y	SE	N	Nucl div	SE
North Am.	76	0.699	0.105	27	0.657	0.127	6	0.465	0.275	27	0.0100	0.0052
Meso Am.	142	0.665	0.123	62	0.663	0.110	18	0.367	0.250	72	0.0108	0.0055
Northwest S. Am.	276	0.661	0.121	94	0.626	0.131	43	0.482	0.234	120	0.0102	0.0052
Andes	150	0.672	0.114	57	0.645	0.117	20	0.445	0.218	71	0.0099	0.0050
East S. Am.	142	0.633	0.135	56	0.603	0.144	27	0.335	0.273	33	0.0093	0.0048

N = number of chromosomes

North America: Cree, Ojibwa.

Meso America: Mixtec, Mixe, Zapotec, Kaqchikel.

Northwest South America (including Lower Central America): Embera, Guaymi, Arhuaco, Zenu, Wayuu, Waunana, Kogi, Cabecar.

Andes: Inga, Quechua, Huilliche and Aymara.

East South America: Aché, Guaraní, Ticuna, Kaingang.

Table 3 F_{ST} within regions in the Americas estimated with different marker systems

	A-STRs	X-STRs	Y-STRs	mtDNA
Meso Am.	0.025	0.017	0.506	0.144
Northwest S. Am	0.069	0.117	0.402	0.300
Andes	0.021	0.019	0.235	0.160
East S. Am.	0.149	0.202	0.585	0.548

Calculations were done across all marker sets using the same set of 15 populations for which data for at least 4 Y-chromosomes were available. All estimates across marker systems within a geographic region are significantly different (p < 0.05).

at mutation–drift equilibrium, between the observed heterozygosity and the number of alleles. It has been shown that population contraction and expansion lead, respectively, to a transient excess or deficiency (respectively) of heterozygosity, relative to that expected based on the observed number of alleles (Nei & Chakraborty, 1975; Maruyama & Fuerst, 1984). This is due to allelic diversity varying more rapidly than heterozygosity with changes in population size, rare alleles disappearing after a contraction or becoming more common after an expansion. BOTTLENECK results are summarized in Table 4 and illustrative X-STR allele frequency distributions are shown in Figure 6 (graphs for all populations are shown in Figure S2). Several populations, mainly those from Eastern and Northwest South America (e.g. the Aché) show few low-frequency alleles, consistent with a recent population contraction. Populations from Meso America and the Andes mostly show an opposite trend, with an apparent excess of rare alleles, consistent with population expansion. Overall, seven of the 18 populations tested deviate from mutation–drift equilibrium based on the "standardized difference" test implemented in BOTTLENECK (Table 4, Cornuet & Luikart, 1996). Second, we applied Fu's *Fs* test to the mtDNA

D-loop sequences obtained here (Table 4). This test evaluates whether the number of different sequences found exceeds that expected for a population at equilibrium, based on the observed nucleotide diversity, and is characterized by negative values in expanding populations (Fu, 1997; Excoffier & Schneider, 1999). Among the Native American populations examined negative *Fs* values were observed in all the Andean populations and in some of the Meso and North Americans. The Northwestern South and Eastern South American populations all show positive *Fs* values (Table 4). Third, we examined the X to autosomal STR diversity ratio (H_X/H_A). It has been shown that due to differences in ploidy this ratio is sensitive to changes in population size (Pool & Nielsen, 2007). Values of the H_X/H_A ratio in the Native American populations examined are shown in Table 4. The mean H_X/H_A values for each region are: 0.907 (Northwestern South America), 0.918 (East South America), 0.950 (North America), 0.952 (Andean South America), and 0.998 (Meso America). There is a highly significant negative correlation between the *Fs* values obtained from the mtDNA D-loop sequences and the H_X/H_A ratios ($r^{Spearman} = -0.592$; p = 0.005).

Discussion

The mtDNA, X-chromosome, and Y-chromosome data obtained here are globally consistent with the previously reported autosomal STR data in showing a north to south gradient in genetic diversity (and an increase in population differentiation) in the Americas. This trend extends the pattern of decreasing population diversity seen worldwide as a function of distance from Africa (Ramachandran et al., 2005; Handley et al., 2007; Wang et al., 2007; Jakobsson et al., 2008; Li et al., 2008). In the Americas, this gradient is consistent with an initial settlement from Northeast Asia into the extreme Northwest of the continent, followed by a southward
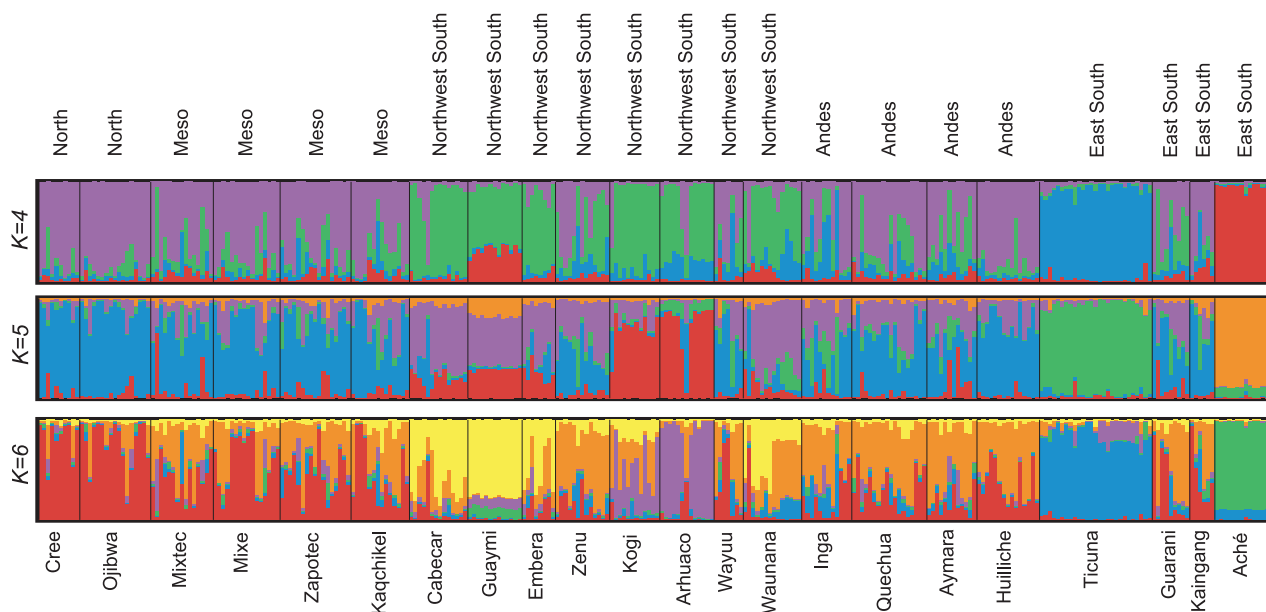
Figure 3 Unsupervised analysis of population structure in America based on data for 38 X-STRs typed in 22 Native American populations, obtained using the STRUCTURE program. The number of clusters in a given plot is indicated by the value of K on the left (plots are shown only for $K = 4$–6). The geographic regions considered in other analyses and the name of each individual population, are indicated at the top and bottom of the figure, respectively. The left-to-right order of the individuals is the same in all plots.
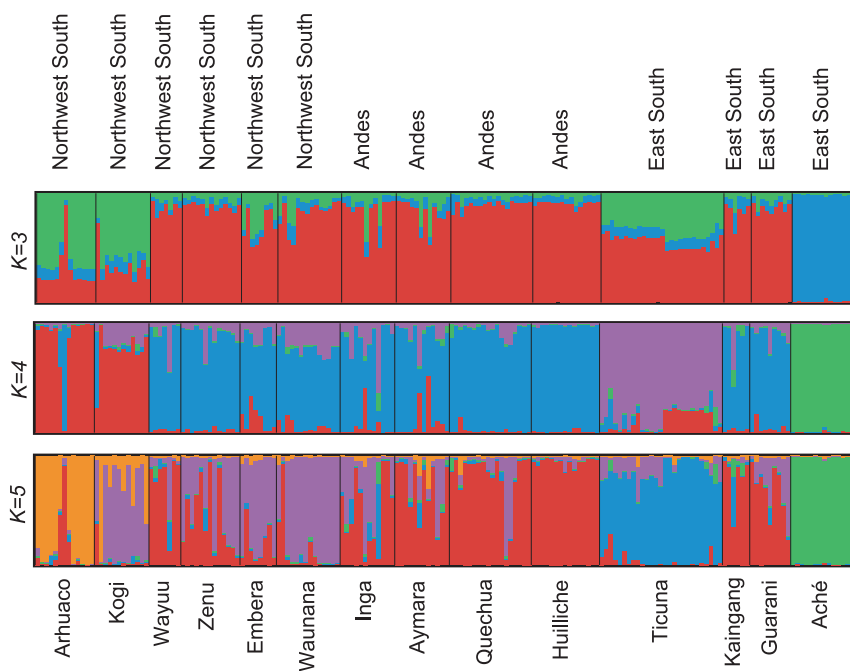

Figure 4 Unsupervised analysis of South American population structure based on 38 X-STRs, obtained using the STRUCTURE program. The number of clusters in a given plot is indicated by the value of K on the left. The geographic regions considered in other analyses and the name of each individual population, are indicated at the top and bottom of the figure, respectively. The left-to-right order of the individuals is the same in all plots.

A)

B)
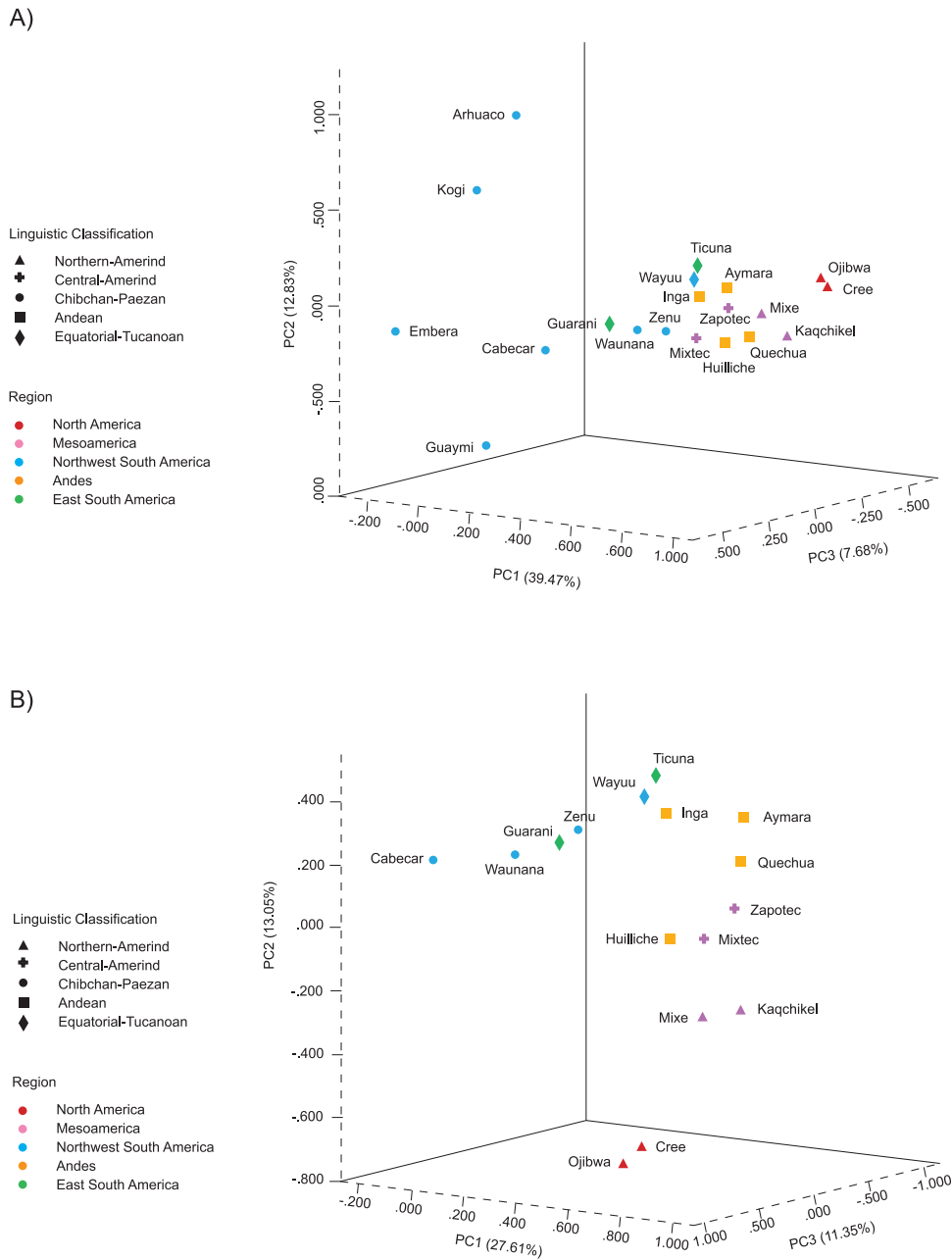
Figure 5 First three components obtained by PCA of a pairwise genetic distance matrix obtained from allele frequency data for 38 X-chromosome STRs for 20 (A) and 16 (B) of the Native American populations examined here (results for the full 22 populations are shown in Figure S1). Population symbols and coloring are as in Figure 1.

colonization (Wang et al., 2007). Interestingly, as reported based on autosomal STRs (Wang et al., 2007), the mtDNA and X-STR data obtained here (in the same population samples) are also consistent with the coasts playing an important role during the initial human population dispersal in the American continent. For these various sets of markers, the correlation of population diversity with distance from the Northwest increases when considering "least cost" distances incorporating the coastal outline, coinciding in a maximum correlation around a 1:10 coast/inland cost ratio. The role of the coasts in the initial human population dispersal outside of Africa has become the subject of increased scrutiny in the

Population	Region	X-STRs	mtDNA	H_X/H_A
Aymara	Andes	0.180	**−5.77**	0.960
Huilliche	Andes	**0.002**	**−10.44**	0.956
Inga	Andes	**0.035**	3.73	0.929
Quechua	Andes	**0.005**	**−8.29**	0.963
Aché	East S. Am.	**0.044**	5.01	0.917
Ticuna	East S. Am.	**0.006**	2.54	0.918
Kaqchikel	Meso Am.	0.499	**−4.83**	1.036
Mixe	Meso Am.	0.381	1.95	0.983
Mixtec	Meso Am.	0.440	−1.2	0.989
Zapotec	Meso Am.	**0.005**	**−1.74**	0.985
Cree	North Am.	0.370	−0.1	0.963
Ojibwa	North Am.	0.418	2.98	0.936
Arhuaco	NorthWest S Am.	**0.000**	4.62	0.810
Cabecar	NorthWest S Am.	0.283	0.34	0.979
Guaymi	NorthWest S Am.	0.181	1.95	0.896
Kogi	NorthWest S. Am.	0.280	3.51	0.924
Waunana	NorthWest S. Am.	0.377	2.3	0.915
Zenu	NorthWest S. Am.	0.182	4.98	0.916

Table 4 Evaluation of deviation from mutation–drift equilibrium with different marker systems in the Native American populations examined

Only populations with data for > 10 samples are included. Results for the standardized difference test (Cornuet & Luikart, 1996) and Fu's Fs are shown for X-STRs and mtDNA, respectively. Values shown in bold are significant at the 5% level.
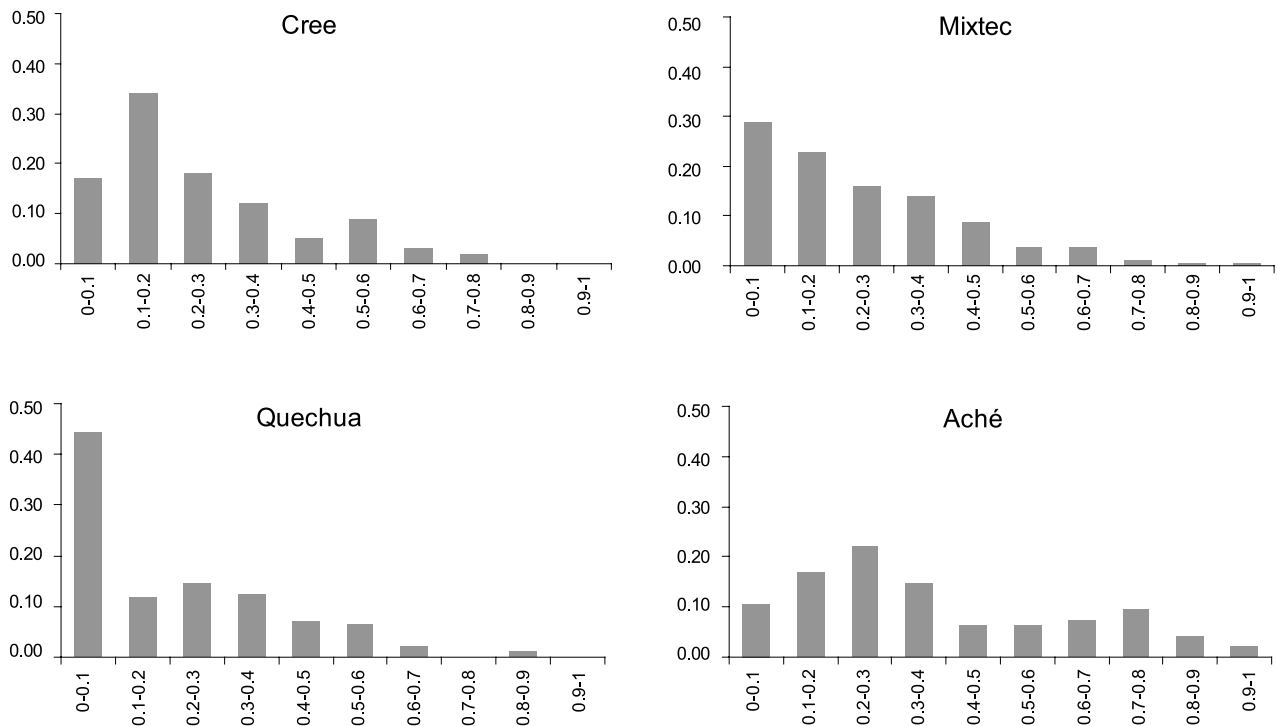

Figure 6 Histograms showing the proportion of alleles within frequency bins for 38 X-STRs in four Native American populations. Plots for the full set of 22 populations examined are displayed in Figure S2.

last few years, with a growing body of evidence pointing to the coasts as facilitators of migration (Liu et al., 2006). In the Americas, although classical models of settlement posited an inland migration (Fiedel, 2000; Dixon, 2001) there is growing interest in the possibility of a coastal migration route (Dixon, 2001; Fix, 2002, 2005). Recent geological research indicates the existence of ice-free coastal areas west of the Cordilleran ice sheet ~14,000 years ago, considerably earlier than the proposed time of opening of an inland ice-free corridor (~11,000 years ago). Computer simulations indicate that coastal migration models could be more consistent with patterns of mtDNA and classical marker data (Surovell, 2003; Fix, 2005). Recently, full mtDNA genome sequence data have also been interpreted as supporting a coastal population dispersal in the Americas (Fagundes et al., 2008).

Previous research has evidenced contrasting patterns of genetic variation in Western (mostly Andean) versus Eastern South America, characterized by lower within-population diversity (and greater differentiation) in the East relative to the West (Luiselli et al., 2000; Tarazona-Santos et al., 2001; Wang et al., 2007). Craniometric analyses also indicate an important differentiation between populations of East and West South America (Pucciarelli et al., 2006). The data examined here for uniparental and biparental markers support this view in showing the lowest within-population diversity and the highest differentiation in Eastern South America. By contrast, Andean populations show the highest diversity and lowest differentiation in the region, while populations from Northwest South America show intermediate values (Tables 2–3). These observations point to a differing demographic history across the Americas, with lower genetic drift in Meso America and in the Andes relative to the rest of South America. The availability of data for different genetic marker systems provides further opportunities to explore possible differences in the demographic history of these populations. The various marker systems examined here allow additional, independent, assessments of demographic scenarios. For instance, Fu's F test has been previously applied to worldwide mtDNA D-loop sequence data and has been shown to reveal a signal of population expansion (as evidenced by negative Fs) in all human populations, except certain hunter gatherers and Native Americans (Excoffier & Schneider, 1999). The lack of a signal of expansion in these populations has been interpreted as resulting from recent bottlenecks that have erased the signal of past population growth; or that effective sizes during initial spatial population expansions were smaller for populations with positive Fs (Excoffier & Schneider, 1999; Ray et al., 2003). In our survey, we find that Fs, calculated from mtDNA D-loop sequences, are negative in populations from Meso America and the Andes, while populations from East and Northwest South America are characterized by positive F values (Table 4). The mtDNA data are thus consistent with population ex-

pansion in Meso America and the Andes, but not in North West and East South America. Independent evidence of a contrasting demographic history in the Americas is provided by the tests applied to the X chromosome STR data shown in Table 4. These are also suggestive of bottlenecks in East and Northwest South America and of population expansion in Meso America and the Andes. The data collected here also allow a (rough) evaluation of the variable impact that changes in population size could have on the relative diversity of different marker systems in the Americas. It has been shown that population contraction can lead to relatively lower ratios of X to autosomal diversity (Pool & Nielsen, 2007), as observed in populations from East and Northwest South America (Table 4; equilibrium values being dependent on the mutational properties of STRs and population effective sizes). Similarly, estimates of population differentiation obtained with various marker systems will be affected differently by the unequal impact of changes in population size on estimates of intrapopulation diversity. At a continent-wide level we observe increasing population structure with data from autosomes, the X-chromosome, mtDNA, and the Y-chromosome (F_{ST} of 0.068, 0.094, 0.256, and 0.390, respectively). However, there is considerable regional variation in F_{ST} values between marker systems (Table 3). It is interesting to note that F_{ST} estimates based on autosomal STRs for Meso America and the Andes are higher than those obtained with X-STRs for the same regions. The opposite is seen in North=west and East South America. As mentioned above, after a change in population size, diversity on the X-chromosome is expected to reach new equilibrium values more rapidly than that on the autosomes. Consequently, although after a population contraction the difference between $F_{ST}^{(X)}$ and $F_{ST}^{(A)}$ will be increased over equilibrium values, the opposite will occur if there is a population expansion. Consistent with this scenario is the observation that higher F_{ST} estimates are observed with the Y-chromosome than with mtDNA across all regions of the Americas (Table 3). A number of previous studies have examined population structure based on mtDNA and Y-chromosome data in order to evaluate differences in migration rates between men and women (Cavalli-Sforza & Bodmer, 1971; Mesa et al., 2000; Bortolini et al., 2002; Segurel et al., 2008), and it has been shown that such differences in migration rates can also reduce the expected difference between $F_{ST}^{(X)}$ and $F_{ST}^{(A)}$ (Segurel et al., 2008). Nevertheless, population genetic theory indicates that $F_{ST}^{(A)}$ is not expected to exceed $F_{ST}^{(X)}$. For this to happen, additional demographic factors need to be invoked, such as a greater variance in reproductive success in men than in women (Cavalli-Sforza & Bodmer, 1971; Segurel et al., 2008). Overall, our results point to larger long-term effective population sizes in Andean relative to the rest of South America, possibly associated with

events of population expansion in the Andes and contraction elsewhere. Further assessment of this scenario will require further analyses, accouting for the specific mutational properties of the markers examined and the fine-grained population structure patterns in each region (Stadler et al., 2009).

From a population history perspective, our results point to Andean populations separating from other South Americans early in the colonization of the subcontinent and subsequently maintaining larger population sizes relative to other South American populations (Rothhammer & Dillehay, 2009). This early split of Andeans and non-Andeans is also consistent with the geographic distribution of Amerind languages in that the Andean stock extends furthest South, along the western side of the subcontinent, and as such is likely to represent the initial settlers of the region (Cavalli-Sforza et al., 1994). Archaeological evidence suggests that human population sizes were relatively large along the Pacific coast, even prior to the development of agriculture, sustained by the rich local fishing resources (Bellwood, 2001). Within South America, agriculture developed first along the central Andes resulting in the establishment in this region of a dense population, culminating in the development of the Inca Empire (Bellwood, 2001). Populations in Western South America therefore seem to have had relatively large sizes for a considerable amount of time and also to have undergone a relatively recent expansion, possibly associated with the use of intensive agriculture. Archaeological information indicates that developments in the Northwestern, and particularly in Eastern South America, were more incipient than in the Andes and were associated with relatively lower population densities.

In conclusion, the nuclear and mtDNA data analyzed here are consistent with an important role for the coasts in the initial settlement of the American continent. Although our analyses did not aim at estimating specific evolutionary parameters, our observations suggest a differentiated demographic history between regions in South America, perhaps associated with an early divergence of Andean from non-Andean populations. Examining a larger number of Native populations (particularly increasing geographic coverage) should allow a more detailed evaluation of the routes of dispersal during the settlement of the American continent. Other than evaluating more precisely the role of the coasts in this dispersal, it will be interesting to examine the impact of other major geographic features of the continent, including the main mountain ranges and river valleys. The feasibility of genome-wide population surveys of increasing resolution will soon culminate in a full description of the genetic diversity of human population samples (i.e. individual genome sequences). The application of explicit methods for testing alternative evolutionary scenarios (and estimating their associated parameters) to such data should enable a refined assessment of the initial settlement and routes of population dispersal in the American continent.

References

Anderson, S., Bankier, A. T., Barrell, B. G., De Bruijn, M. H., Coulson, A. R., Drouin, J., Eperon, I. C., Nierlich, D. P., Roe, B. A., Sanger, F., Schreier, P. H., Smith, A. J., Staden, R. & Young, I. G. (1981) Sequence and organization of the human mitochondrial genome. *Nature* **290**, 457–465.

Bellwood, P. (2001) Early agriculturalist population diasporas? Farming, languages, and genes. *Annu Rev Anthropol* **30**, 181–207.

Bonatto, S. L. & Salzano, F. M. (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, M. C., Salzano, F. M., Bau, C. H., Layrisse, Z., Petzl-Erler, M. L., Tsuneto, L. T., Hillk, Hurtado, A. M., Castro-De-Guerra, D., Bedoya, G. & Ruiz-Linares, A. (2002) Y-chromosome biallelic polymorphisms and Native American population structure. *Ann Hum Genet* **66**, 255–259.

Bortolini, M. C., Salzano, F. M., Thomas, M. G., Stuart, S., Nasanen, S. P., Bau, C. H., Hutz, M. H., Layrisse, Z., Petzl-Erler, M. L., Tsuneto, L. T., Hill, K., Hurtado, A. M., Castro-De-Guerra, D., Torres, M. M., 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., Yotova, V., Wang, S., Bourtoumieu, S., Moreau, C., Michalski, R., Moisan, J. P., Hill, K., Hurtado, A. M., Ruiz-Linares, A., & Labuda, D. (2009) X-chromosome lineages and the settlement of the Americas. *Am J Phys Anthropol* **140**, 417–428.

Brandstatter, A., Peterson, C. T., Irwin, J. A., Mpoke, S., Koech, D. K., Parson, W., & Parsons, T. J. (2004) Mitochondrial DNA control region sequences from Nairobi (Kenya): inferring phylogenetic parameters for the establishment of a forensic database. *Int J Legal Med* **118**, 294–306.

Cavalli-Sforza, L. L. & Bodmer, W. F. (1971) *The genetics of human populations.* San Francisco, CA: Freeman.

Cavalli-Sforza, L. L., Menozzi, P., & Piazza, A. (1994) *The History and Geography of Human Genes.* Princeton, NJ: Princeton University Press.

Chinnery, P. F., Howell, N., Andrews, R. M., & Turnbull, D. M. (1999) Mitochondrial DNA analysis: polymorphisms and pathogenicity. *J Med Genet* **36**, 505–510.

Cornuet, J. M. & Luikart, G. (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics* **144**, 2001–2014.

Dixon, E. J. (2001) Human colonization of the Americas: timing, technology and process. *Quat Sci Rev* **20**, 277–299.

Excoffier, L. & Schneider, S. (1999) Why hunter-gatherer populations do not show signs of pleistocene demographic expansions. *Proc Natl Acad Sci USA* **96**, 10597–10602.

Excoffier, L., Laval, G., & Schneider, S. (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol Bioinform Online* **1**, 47–50.

Fagundes, N. J., Kanitz, R., Eckert, R., Valls, A. C., Bogo, M. R., Salzano, F. M., Smith, D. G., Silva, W. A., Jr., Zago, M. A., Ribeiro-Dos-Santos, A. K., Santos, S. E., Petzl-Erler, M. L., & Bonatto, S. L. (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.

Falush, D., Stephens, M., & Pritchard, J. K. (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164**, 1567–1587.

Fiedel, S. J. (2000) The peopling of the New World: present evidence, new theories, and future directions. *J Archaeol Res* **8**, 39–103.

Fix, A. G. (2002) Colonization models and initial genetic diversity in the Americas. *Hum Biol* **74**, 1–10.

Fix, A. G. (2005) Rapid deployment of the five founding Amerind mtDNA haplogroups via coastal and riverine colonization. *Am J Phys Anthropol* **128**, 430–436.

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

Fu, Y. X. (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147**, 915–925.

Handley, L. J., Manica, A., Goudet, J., & Balloux, F. (2007) Going the distance: human population genetics in a clinal world. *Trends Genet* **23**, 432–439.

Jakobsson, M., Scholz, S. W., Scheet, P., Gibbs, J. R., Vanliere, J. M., Fung, H. C., Szpiech, Z. A., Degnan, J. H., Wang, K., Guerreiro, R., Bras, J. M., Schymick, J. C., Hernandez, D. G., Traynor, B. J., Simon-Sanchez, J., Matarin, M., Britton, A., Van De Leemput, J., Rafferty, I., Bucan, M., Cann, H. M., Hardy, J. A., Rosenberg, N. A., & Singleton, A. B. (2008) Genotype, haplotype and copy-number variation in worldwide human populations. *Nature* **451**, 998–1003.

Karafet, T. M., Zegura, S. L., Posukh, O., Osipova, L., Bergen, A., Long, J., Goldman, D., Klitz, W., Harihara, S., De Knijff, P., Wiebe, V., Griffiths, R. C., Templeton, A. R., & Hammer, M. F. (1999) Ancestral Asian source(s) of new world Y-chromosome founder haplotypes. *Am J Hum Genet* **64**, 817–831.

Keinan, A., Mullikin, J. C., Patterson, N., & Reich, D. (2009) Accelerated genetic drift on chromosome X during the human dispersal out of Africa. *Nat Genet* **41**, 66–70.

Lell, J. T., Sukernik, R. I., Starikovskaya, Y. B., Su, B., Jin, L., Schurr, T. G., Underhill, P. A., & Wallace, D. C. (2002) The dual origin and Siberian affinities of Native American Y chromosomes. *Am J Hum Genet* **70**, 192–206.

Li, J. Z., Absher, D. M., Tang, H., Southwick, A. M., Casto, A. M., Ramachandran, S., Cann, H. M., Barsh, G. S., Feldman, M., Cavalli-Sforza, L. L., & Myers, R. M. (2008) Worldwide human relationships inferred from genome-wide patterns of variation. *Science* **319**, 1100–1104.

Liu, H., Prugnolle, F., Manica, A., & Balloux, F. (2006) A geographically explicit genetic model of worldwide human-settlement history. *Am J Hum Genet* **79**, 230–237.

Liu, K. & Muse, S. V. (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics* **21**, 2128–2129.

Luiselli, D., Simoni, L., Tarazona-Santos, E., Pastor, S., & Pettener, D. (2000) Genetic structure of Quechua-speakers of the Central Andes and geographic patterns of gene frequencies in South Amerindian populations. *Am J Phys Anthropol* **113**, 5–17.

Maruyama, T. & Fuerst, P. A. (1984) Population bottlenecks and nonequilibrium models in population genetics. I. Allele numbers when populations evolve from zero variability. *Genetics* **108**, 745–763.

Merriwether, D. A., Rothhammer, F., & Ferrell, R. E. (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, N. R., Mondragon, M. C., Soto, I. D., Parra, M. V., Duque, C., Ortiz-Barrientos, D., Garcia, L. F., Velez, I. D., Bravo, M. L., Munera, J. G., Bedoya, G., Bortolini, M. C., & 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.

Nei, M. & Chesser, R. K. (1983) Estimation of fixation indices and gene diversities. *Ann Hum Genet* **47**, 253–259.

Nei, M. M. T. & Chakraborty, R. (1975) The bottleneck effect and genetic variability in populations. *Evolution* **29**, 1–10.

Pool, J. E. & Nielsen, R. (2007) Population size changes reshape genomic patterns of diversity. *Evolution* **61**, 3001–3006.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000) Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959.

Pucciarelli, H. M., Neves, W. A., Gonzalez-Jose, R., Sardi, M. L., Rozzi, F. R., Struck, A., & Bonilla, M. Y. (2006) East-West cranial differentiation in pre-Columbian human populations of South America. *Homo* **57**, 133–150.

Ramachandran, S., Deshpande, O., Roseman, C. C., Rosenberg, N. A., Feldman, M. W., & Cavalli-Sforza, L. L. (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.

Ray, N. (2005) PATHMATRIX: a geographical information system tool to compute effective distances among samples. *Mol Ecol Notes* **5**, 177–180.

Ray, N., Currat, M., & Excoffier, L. (2003) Intra-deme molecular diversity in spatially expanding populations. *Mol Biol Evol* **20**, 76–86.

Rosenberg, N. A. (2004) DISTRUCT: a program for the graphical display of population structure. *Mol Ecol Notes* **4**, 137–138.

Rothhammer, F. & Dillehay, T. D. (2009) The late Pleistocene colonization of South America: an interdisciplinary perspective. *Ann Hum Genet* **73**, 540–549.

Ruhlen, M. (1991) *A guide to the world's languages.* Stanford, CA: Stanford University Press.

Ruiz-Linares, A., Ortiz-Barrientos, D., Figueroa, M., Mesa, N., Munera, J. G., Bedoya, G., Velez, I. D., Garcia, L. F., Perez-Lezaun, A., Bertranpetit, J., Feldman, M. W., & Goldstein, D. B. (1999) Microsatellites provide evidence for Y chromosome diversity among the founders of the New World. *Proc Natl Acad Sci USA* **96**, 6312–6317.

Salzano, F. M. & Callegari-Jacques, S. M. (1988) *South American Indians: a case study in evolution.* Oxford, Clarendon Press.

Segurel, L., Martinez-Cruz, B., Quintana-Murci, L., Balaresque, P., Georges, M., Hegay, T., Aldashev, A., Nasyrova, F., Jobling, M. A., Heyer, E., & Vitalis, R. (2008) Sex-specific genetic structure

and social organization in Central Asia: insights from a multi-locus study. *PLoS Genet* **4**, e1000200.

Seielstad, M., Yuldasheva, N., Singh, N., Underhill, P., Oefner, P., Shen, P., & Wells, R. S. (2003) A novel Y-chromosome variant puts an upper limit on the timing of first entry into the Americas 1. *Am J Hum Genet* **73**, 700–705.

Siegel, S. (1956) *Nonparametric statistics for the behavioural sciences.* Tokyo: McGraw-Hill.

Stadler, T., Haubold, B., Merino, C., Stephan, W., & Pfaffelhuber, P. (2009) The impact of sampling schemes on the site frequency spectrum in nonequilibrium subdivided populations. *Genetics* **182**, 205–216.

Surovell, T. A. (2003) Simulating coastal migration in New World colonization. *Curr Anthropol* **44**, 580–591.

Tarazona-Santos, E., Carvalho-Silva, D. R., Pettener, D., Luiselli, D., De Stefano, G. F., Labarga, C. M., Rickards, O., Tyler-Smith, C., Pena, S. D., & Santos, F. R. (2001) Genetic differentiation in South Amerindians is related to environmental and cultural diversity: evidence from the Y chromosome. *Am J Hum Genet* **68**, 1485–1496.

Torroni, A., Schurr, T. G., Cabell, M. F., Brown, M. D., Neel, J. V., Larsen, M., Smith, D. G., Vullo, C. M., & Wallace, D. C. (1993) Asian affinities and continental radiation of the four founding Native American mtDNAs. *Am J Hum Genet* **53**, 563–590.

Wang, S., Bedoya, G., Labuda, D., & Ruiz-Linares, A. (2010) Brief communication: patterns of linkage disequilibrium and haplotype diversity at Xq13 in six Native American populations. *Am J Phys Anthropol* **142**, 476–480.

Wang, S., Lewis, C. M., Jakobsson, M., Ramachandran, S., Ray, N., Bedoya, G., Rojas, W., Parra, M. V., Molina, J. A., Gallo, C., Mazzotti, G., Poletti, G., Hill, K., Hurtado, A. M., Labuda, D., Klitz, W., Barrantes, R., Bortolini, M. C., Salzano, F. M., Petzl-Erler, M. L., Tsuneto, L. T., Llop, E., Rothhammer, F., Excoffier, L., Feldman, M. W., Rosenberg, N. A., & Ruiz-Linares, A. (2007) Genetic Variation and Population Structure in Native Americans. *PLoS Genet* **3**, e185.

Wilkins, J. F. (2006) Unraveling male and female histories from human genetic data. *Curr Opin Genet Dev* **16**, 611–617.

Supporting information

Additional supporting information may be found in the online version of this article:

Table S1 Sequence of primers used for amplification of Y-STRs

Figure S1 First three components obtained by PCA of a pairwise genetic distance matrix obtained from allele frequency data for 38 X-chromosome STRs for the full set of 22 Native American populations examined here. Population symbols and colouring are as in Figure 1 of the main paper.

Figure S2 Histograms showing the proportion of alleles within particular frequency bins across 38 X-STRs for the 22 Native American populations examined here.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.