AMERICAN JOURNAL OF PHYSICAL ANTHROPOLOGY 134:190–197 (2007)

The β-Globin Gene Cluster Distribution Revisited—Patterns in Native American Populations

Sidia M. Callegari-Jacques,[1,2]* Shaiane G. Crossetti,[1] Fabiana B. Kohlrausch,[1] Francisco M. Salzano,[1] Luiza T. Tsuneto,[3] Maria Luiza Petzl-Erler,[3] Kim Hill,[4] A. Magdalena Hurtado,[4] and Mara H. Hutz[1]

[1]*Departamento de Genética, Universidade Federal do Rio Grande do Sul, Caixa Postal 15053, 91501-970 Porto Alegre, RS, Brazil*
[2]*Departamento de Estatística, Universidade Federal do Rio Grande do Sul, Campus do Vale, 91509-900 Porto Alegre, RS, Brazil*
[3]*Departamento de Genética, Universidade Federal do Paraná, Caixa Postal 19071, 81531-990 Curitiba, PR, Brazil*
[4]*Department of Anthropology, University of New Mexico, Albuquerque, NM 87131*

KEY WORDS genetic diversity; South American gene geography; language and genetics

ABSTRACT New frequencies for the β-globin gene cluster haplotypes are presented for the Aché ($N = 82$ individuals), Guarani ($N = 76$), and Kaingang ($N = 54$), three Native South American populations that live in an area between parallels 20°S and 30°S not covered by previous studies at this locus. The haplotype frequencies obtained for the three populations are within the interval observed for 28 other Native American populations. The Aché show much less haplotypes (five) than the other two populations (9–10), the haplotype prevalences being more similar to those of the Guarani than to the Kaingang. The Native American total heterozygosity was about half (0.41) that obtained for the African populations (0.71), but was not much different from those obtained for other continents. A geographical pattern was disclosed in South America by mapping the frequencies of the most common haplotype (haplotype 2), and by means of spatial correlation analysis. The analysis of molecular variance (AMOVA) and pairwise F_{ST} data suggest three distinct sectors for the genetic landscape of Native South America: the Andes, the Center/Southeast region, and the Amazon. Am J Phys Anthropol 134:190–197, 2007.

Although many studies tried to understand the spatial distribution of genetic variation in native South America, only one feature seems clearly supported by different approaches: a west-east division, separating the Pacific coast plus Andes highlands, *versus* the lowlands of South America. This pattern was observed using craniofacial data (Rothhammer and Silva,1989), blood group and protein markers (Rothhammer and Silva, 1992; Luiselli et al., 2000; Simoni et al., 2000), HLA frequencies (Rothhammer et al., 1997), Y chromosome variation (Tarazona-Santos et al., 2001), mtDNA results (Fuselli et al., 2003; Lewis et al., 2004), and linguistic data (Nichols, 2002).

Different results however were obtained by Salzano and Callegari-Jacques (1988), who analyzed the genetic variability in South America using principal components (PC) based on allele frequencies of 13 blood groups and protein loci of 19 Native American populations (from now on, we will use the terms "Native Americans" or "American Natives" to refer to indigenous peoples of all the Americas). In a tridimensional graphical representation of 54% of the total variation, they observed that populations that speak languages of Loukotka's (1968) Tropical Forest division clearly separated from those who speak Paleo-American (Xavante, Kayapó, Krahó, Mataco, Ayoreo, Choroti) or Andean (Quechua, Aymara) languages. Paleo-American speakers live in Central Brazil or in the Chaco region (southeast of South America), and based on the simple west/east division, they would be expected to cluster with Amazonian Tropical Forest speakers.

Other studies showed still different patterns. Using the first three PC maps based on 11 classical systems for 70 populations, O'Rourke and Suarez (1985) obtained a patchy distribution, which they interpreted as being produced by local isolation and drift (although the PC2 map had a general west/east differentiation). Callegari-Jacques et al. (1993) also observed a patchy distribution when mapping axis 1 scores (57% of variation) in a correspondence analysis of gammaglobulin (Gm) haplotypes from 36 South American native populations.

The synthetic maps obtained by Cavalli-Sforza et al. (1994) using principal components based on 72 "classical" gene frequencies depicts different scenarios according to the principal component considered. The first (representing 33% of all America's variation) supports the differentiation between the western and the eastern halves of South America, whereas the second (13%) presents a patchy pattern.

The usefulness of the β-globin gene cluster haplotypes for studying human evolutionary relationships has already been demonstrated. Its DNA can be conveniently assessed by five restriction enzymes, and the different haplotypes observed in relation to the corresponding

Grant sponsor: Institutos do Milênio, Programa de Apoio a Grupos de Excelência, Conselho Nacional de Desenvolvimento Científico e Tecnológico, Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul.

*Correspondence to: Sidia Maria Callegari-Jacques, Departamento de Genética, Universidade Federal do Rio Grande do Sul, Caixa Postal 15053, 91501-970 Porto Alegre, RS, Brazil.
E-mail: sidia.jacques@ufrgs.br

Received 15 January 2007; accepted 13 April 2007

DOI 10.1002/ajpa.20653
Published online 27 June 2007 in Wiley InterScience (www.interscience.wiley.com).

restriction sites can be hierarchically ordered, showing also extensive interethnic variability. The split between Africans and the rest of the world was clearly seen using these markers (Wainscoat et al., 1986; Chen et al., 1990; Long et al., 1990), as well as the structuring according to continental regions. Similarities of East Asians, Pacific populations, and Native Americans were also observed, but no specific investigation about its pattern of distribution in Native Americans is available. This fact is unfortunate, since these haplotypes have already been investigated in 28 Native American populations, three of them located in the northern/central subcontinent (Guerreiro et al., 1992, 1994; Bevilaqua et al., 1995; Kaufman et al., 1998; Mattevi et al., 2000; Mesa et al., 2000; Villalobos-Arámbula et al., 2000; Mousinho-Ribeiro et al., 2003; Shimizu and Harihara, 2004). Twenty-four of the 25 South American populations live north of latitude 14°S, the Mapuche inhabiting a more southern location (around 39°S, 68°W). A large area between 14°S and 39°S have populations not studied for this locus; therefore, we decided to study three populations that live between 20–30°S and 50–60°W. The results were integrated with the published information, with the following questions in mind: (a) How do the Native American frequencies and variability compare with those obtained in other continents? (b) Can distinctive geographical patterns be observed within the American continent? and (c) What inferences can thus be obtained about the continent's early colonization?

SUBJECTS AND METHODS

Populations

Three Native American populations were studied. The Guarani and Kaingang are the two major Native American groups living in southern Brazil, and speak languages of the Tupian and Jean stocks, respectively (Campbell, 1997). The Guarani ($N = 76$) were sampled in Amambaí, Limão Verde and Porto Lindo, Mato Grosso do Sul (average geographical coordinates: 23°30′ S, 55° W, Tsuneto et al., 2003). The Kaingang ($N = 54$) live in Nonoai, Rio Grande do Sul (27°20′ S, 52°45′ W). Both populations are in an advanced stage of acculturation but still keep their language, and the fact that they live in reservations condition a relative isolation. Although living in nearby areas for centuries, the Guarani and Kaingang still differ in many cultural and biological aspects (Petzl-Erler et al., 1993; Salzano et al., 1997; Tsuneto et al., 2003). The Aché from Paraguay ($N = 82$) have a series of morphological and genetic distinctive peculiarities (Hill and Hurtado, 1996; Tsuneto et al., 2003; Schmitt et al., 2004). Linguistically, they are related to the Guarani (Tupian stock, Loukotka, 1968; Campbell, 1997), but there are also suggestions that they may have descended from a Jean group that preceded the Guarani colonization of Paraguay (Hill and Hurtado, 1996). The individuals were sampled in Arroyo Bandera and Chupa-Pou (average coordinates: 24° S, 56° W). The geographic location of these and the other populations considered can be visualized in Figure 1. The interviews and sampling collections were always made after appropriate informed consent. Materials from all volunteers were obtained, but the tests were conducted duly considering their biological relationships.

Laboratory tests

The blood samples were obtained with anticoagulant, refrigerated shortly after collection, and sent by air to Porto Alegre or Curitiba (Brazil), where DNA was extracted by standard salting out procedures. Five polymorphic restriction sites of the β-globin gene cluster (*Hinc*II-5′ε, *Hind*III-^{G}γ, *Hind*III-^{A}γ, *Hinc*II-ψβ, and *Hinc*II-3′ψβ) were investigated. PCR amplification conditions and primers used were as described in Bevilaqua et al. (1995).

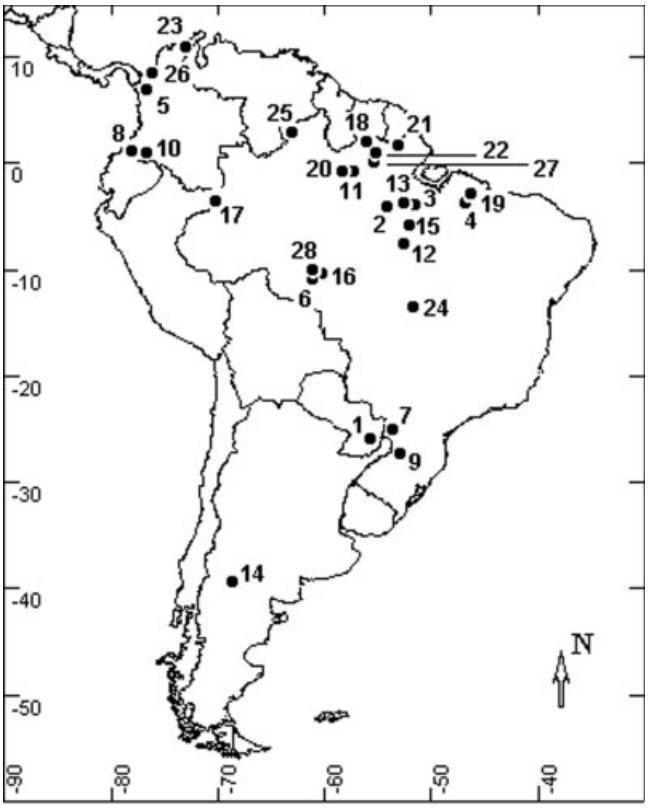

Fig. 1. Geographic location of 28 South Amerindian populations for which β-globin gene cluster haplotype data are available. New data: (1) Aché; (7) Guarani; (9) Kaingang. Names for the other populations are given in Table 2.

Statistical analyses

Three softwares were employed to determine the β-globin haplotypes and to estimate their frequencies: the Multiple Locus Haplotype Analysis program (MLOCUS, Long et al., 1995, 1999; Petterson et al., 1999), and the ARLEQUIN v. 3.1 program (Excoffier et al., 2005), both producing maximum-likelihood estimates; and the PHASE v.2.1.1 software, which implements a Bayesian estimation method (Stephens et al., 2001; Stephens and Donnelly, 2003).

Measures of genetic diversity (heterozygosity, G_{ST}, and G'_{ST}) were calculated according to Nei (1987). The DISPAN program (Ohta, 1993) was used to obtain these estimates as well as modified Cavalli-Sforza D_A genetic distances (Nei et al., 1983). The analysis of molecular variance (AMOVA) (Weir and Cockerham, 1984; Excoffier et al., 1992; Weir, 1996) was employed to test for population subdivision, using the ARLEQUIN program. F_{ST} estimates were also obtained with this software.

Geographical distances were great-circle distances. Linguistic distances between the Native South American populations were coded using the following scores: 2 for languages classified in different major categories as pro-

TABLE 1. Frequencies (%) of β-globin gene cluster haplotypes in Aché, Guarani, and Kaingang Amerindians, as well as variability observed in five continental groups

	Aché	Guarani	Kaingang	Native America[a]	Africa[b]	Europe[b]	Asia[b]	Pacific Islands[b]
No. of populations	1	1	1	28	6	5	7	13
No. of chromosomes	164	152	108	1741	192	258	720	972
				Mean (range)	Mean (range)	Mean (range)	Mean (range)	Mean (range)
Haplotypes								
1 (− − − − −)	0.6	2.6	5.7	1.3 (0–12)	0 (0–0)	0 (0–0)	0.7 (0–2)	1.0 (0–5)
2 (+ − − − −)	79.9	74.9	58.9	74.3 (46–100)	6.3 (0–13)	60.8 (43–72)	66.2 (56–90)	79.2 (51–94)
3 (− − − − +)	0.0	0.7	7.6	1.1 (0–17)	45.9 (26–62)	0.8 (0–6)	0 (0–0)	0.2 (0–3)
4 (− + − − +)	1.2	3.9	1.5	0.9 (0–7)	20.8 (11–40)	0.4 (0–3)	1.8 (0–3)	2.8 (0–19)
5 (− + − + +)	0.0	0.0	12.0	4.6 (0–18)	13.5 (0–50)	23.3 (15–40)	17.9 (2–24)	8.8 (0–20)
6 (− + + − +)	16.5	9.1	5.5	11.6 (0–30)	7.3 (0–16)	11.2 (6–26)	9.3 (0–13)	0.7 (0–4)
7 (− + + − −)	0.0	0.7	0.9	1.2 (0–8)	0 (0–0)	2.7 (1–12)	0.4 (0–2)	0.4 (0–2)
8 (+ − − + +)	0.0	0.0	0.0	0.3 (0–8)	0 (0–0)	0 (0–0)	0.3 (0–3)	2.6 (0–8)
9 (− + + + +)	0.0	0.0	0.0	0.7 (0–10)	0 (0–0)	0 (0–0)	0 (0–0)	1.2 (0–5)
10 (+ + − + +)	0.0	0.0	0.0	0.9 (0–7)	0 (0–0)	0.4 (0–3)	0.7 (0–3)	0 (0–0)
11 (− − − + +)	0.0	0.0	0.0	0.8 (0–4)	0.5 (0–2)	0 (0–0)	0.8 (0–3)	0.9 (0–5)
12 (+ + − − −)	1.8	0.7	2.2	0.4 (0–6)	0 (0–0)	0 (0–0)	0.7 (0–2)	1.1 (0–8)
13 (+ − − − +)	0.0	6.7	1.1	0.4 (0–4)	2.1 (0–10)	0 (0–0)	0 (0–0)	0.6 (0–2)
14 (− − + + −)	0.0	0.6	0.0	0.1 (0–2)	0.5 (0–3)	0 (0–0)	0 (0–0)	0.1 (0–1)
15 (+ + + − +)	0.0	0.0	0.0	0.8 (0–9)	1.0 (0–7)	0 (0–0)	0.4 (0–1)	0 (0–0)
16 (− + − − −)	0.0	0.0	4.6	0.2 (0–2)	2.1 (0–4)	0.4 (0–3)	0.6 (0–2)	0.3 (0–1)
17 (− − + − −)	0.0	0.0	0.0	0.1 (0–1)	0 (0–0)	0 (0–0)	0 (0–0)	0 (0–0)
18 (+ − + − −)	0.0	0.0	0.0	0.3 (0–2)	0 (0–0)	0 (0–0)	0.1 (0–<1)	0 (0–0)
19 (+ − + + +)	0.0	0.0	0.0	0 (0–0)	0 (0–0)	0 (0–0)	0.1 (0–<1)	0 (0–0)
21 (− − + − −)	0.0	0.0	0.0	<0.1 (0–1)	0 (0–0)	0 (0–0)	0 (0–0)	0 (0–0)

[a] Does not include the new data presented herein. References in Table 2.
[b] References: Oehme et al. (1985), Wainscoat et al. (1986), Ramsay and Jenkins (1987), Chen et al. (1990), Long et al. (1990), Shimizu and Harihara (2004).

posed by Loukotka (1968), namely: Paleo-American, Tropical Forest, and Andean; 1 for languages of different stocks belonging to the same major category; 0.5 for the distance between the Tupi-Mondé and the Tupi-Guarani speakers; and 0 for languages belonging to a group with no subdivision. There is no consensus among linguists about the relationships among South American Indian languages. We have chosen Loukotka's approach because our group showed in a previous study (Salzano et al., 2005) that this language classification correlates better with South American genetic data than Greenberg's (Greenberg, 1987) proposal. Other classifications (e.g. that presented by Campbell, 1997) do not present sufficient hierarchical levels to allow an appropriate distinction by means of a score.

Correlation coefficients between genetic, geographic, and linguistic distance matrices were obtained using Mantel's (1967) approach for partial correlation, which allows for the estimation and testing of the correlation between two quantitative measures, controlling for the variation of a third variable associated to the other two. The ARLEQUIN program was used for these calculations.

Spatial correlation analysis was performed with the GENALEX software (Peakall and Smouse, 2006), which allows a global analysis of all haplotypes and yields a single correlogram for the locus.

Map representations of the populations' geographic locations and the isolines for the most frequent haplotype (haplotype 2) frequencies were produced with the IDRISI©v.32.02 software (Eastman, 2000).

RESULTS

The β-globin gene cluster haplotypes and respective frequencies for the Aché, Guarani, and Kaingang estimated with the three softwares were essentially the same. All differences in haplotype frequencies were below 0.01 except in three instances, the maximum difference being 0.03, between the MLOCUS and PHASE estimates. We used the ARLEQUIN estimates for the analyses that follow.

Table 1 presents the haplotype frequencies obtained in the Aché, Guarani, and Kaingang samples, as well as the average and range of values observed in previously studied native populations of America and other continents. Ten haplotypes were observed in the Kaingang, nine in the Guarani, and only five among the Aché. Their frequencies generally lie within the ranges observed for Native Americans. Haplotype 2 was the most frequent in the three samples, other first order haplotypes being also observed with moderate frequencies. Overall, the Aché differences in frequencies were two times higher in relation to the Kaingang than to the Guarani.

Estimates of genetic diversity for the Aché, Guarani, and Kaingang, and for 28 other Native American populations are presented in Table 2. The heterozygosity (H) values ranged from 0.00 (Awá-Guajá) to 0.76 (Emberá), these extreme values being observed in South America; the three North/Central Amerindian populations presented values within this range. The number of haplotypes varied from only 1 (Awá-Guajá) to 13 (Kamsá). Aché and Guarani have intermediate heterozygosities (0.34 and 0.43, respectively) and five and nine haplotypes, respectively. Both the H value (0.63) and the number of haplotypes encountered (10) in the Kaingang are the highest fourth in the range for South American Native populations. A tendency for higher heterozygosity values at the extremes (north, south) of the geographical distribution was observed.

Table 3 presents measures of genetic diversity in Native American and other world populations. H_T is higher in Africa (0.72), followed by Europe (0.62) and North/Central

TABLE 2. Linguistic affiliation, geographic location, and β-globin gene cluster diversity for 33 Amerindian populations (listed in alphabetic order)

Population numbers	Populations	No. of chromosomes	Linguistic group[a]	Geographic location		Heterozygosity (*H*)	No. of haplo types	Reference numbers[b]
				Latitude	Longitude			
South America								
1	Aché	164	Tupí-Guarani	24° S	56° W	0.336	5	1
2	Arara	30	Cariban	4° S	54° W	0.302	4	2
3	Asuriní	14	Tupí-Guarani	3° 43′ S	52° 27′ W	0.394	4	3
4	Awá-Guajá	86	Tupí-Guarani	3° 30′ S	46° 40′ W	0.000	1	3
5	Emberá	24	Chocó	7° N	76° 30′ W	0.757	6	4
6	Gavião	58	Tupí-Mondé	10° 10′ S	61° 08′ W	0.225	5	5
7	Guarani	152	Tupí-Guarani	23° 30′ S	55° W	0.427	9	1
8	Ingano	66	Quechuan	1° 12′ N	78° W	0.649	9	4; 6
9	Kaingang	108	Jean	27° 20′ S	52° 45′ W	0.629	10	1
10	Kamsá	110	Chibchan	1° N	76° 30′ W	0.523	13	6
11	Katuena	56	Cariban	0° 40′ S	57° 30′ W	0.139	3	3
12	Kayapo-Kokraimôro	44	Jean	10° 49′ S	55° 27′ W	0.334	2	2
13	Kayapo-Xikrin	40	Jean	3° 46′ S	53° 21′ W	0.229	3	3
14	Mapuche	86	Mapudungu	39° 10′ S	68° 37′ W	0.605	8	7
15	Parakanã	28	Tupí-Guarani	5° 45′ S	51° 52′ W	0.378	4	3
16	Suruí	44	Tupí-Mondé	10° 50′ S	61° 10′ W	0.323	4	5
17	Ticuna	46	Ticuna	3° 30′ S	70° 30′ W	0.243	4	4
18	Tiriyó	50	Cariban	2° N	56° W	0.504	4	3
19	Urubu-Kaapor	94	Tupí-Guarani	2° 48′ S	46° 10′ W	0.490	8	3
20	Wai-Wai	56	Cariban	0° 40′ S	57° 55′ W	0.223	2	5
21	Wayampi	30	Tupí-Guarani	1° 30′ N	53° W	0.250	4	2
22	Wayana-Apalaí	34	Cariban	1° N	55° W	0.308	3	2
23	Wayuu	260	Maipurean	11° N	73° W	0.526	12	4; 6
24	Xavante	60	Jean	13° 20′ S	51° 40′ W	0.591	5	5
25	Yanomámi	34	Yanomaman	3° N	63° W	0.169	3	8
26	Zenu	64	Chocó	8° 30′ N	76° W	0.396	8	4
27	Zoé	60	Tupí-Guarani	0° 18′ S	55° 18′ W	0.370	3	3
28	Zoró	50	Tupí-Mondé	10° 20′ S	60° 20′ W	0.127	2	5
Central America								
	Huichol	97	Uto-Astecan	19° 26′ N	99° 07′ W	0.362	7	9
North America								
	Carrier-Sekani	50	Eyeak-Athabaskan	52° 20′ N	123° 15′ W	0.600	5	10
	Mvskoke	70	Muskogean	35° 26′ N	97° 28′ W	0.700	11	10

[a] According to Loukotka (1968), Rodrigues (1994), and Campbell (1997).
[b] References: 1. Present investigation; 2. Guerreiro et al. (1994); 3. Mousinho-Ribeiro et al. (2003); 4. Mesa et al. (2000); 5. Bevilaqua et al. (1995); 6. Shimizu and Harihara (2004); 7. Kaufman et al. (1998); 8. Guerreiro et al. (1992); 9. Villalobos-Arámbula et al. (2000); 10. Mattevi et al. (2000).

America (0.58) values. All other figures are lower than 0.43. G_{ST}' (the relative magnitude of interpopulation differentiation corrected for the number of populations considered) is also higher in Africa (0.11). The lowest figures are observed in Europe and the Americas (both 0.07).

Genetic relationships among Native Americans can be blurred if recent admixture with Euro- or African-descendants are not accounted for. Therefore, we decided to drop haplotype 3 (of indisputable African origin; see Table 1) from the analysis, adjusting the remaining frequencies to sum up one. All the following results are based on these "corrected" haplotype prevalences. But it should be pointed out that only minor adjustments had to be made.

Because of the scarce information on North/Central American natives (three samples), we continued the analyses using the 28 South American Native populations only (Fig. 1). Mantel's matrix correlation coefficients between genetic and geographic ($r = 0.29$; $P = 0.018$) and genetic and linguistic distances ($r = 0.19$; $P = 0.043$) were statistically significant, but low. The partial correlation coefficient for D_A and geographic distances controlling for language was $r = 0.24$ ($P = 0.032$), while the coefficient for D_A and linguistic distances adjusting for geography was statistically non significant ($r = 0.09$; $P = 0.193$).

Figure 2 presents the spatial correlation analysis for the 19 β-globin gene cluster haplotype frequencies observed in South American Native populations. The coefficients were positive at short distances, decreasing almost linearly to a negative coefficient at the 2,000–4,000 km interval, suggesting a cline. However, an unexpected positive correlation between genetic and geographic distances was observed in the last distance interval (above 4,000 km).

To better understand the pattern observed, we mapped the frequencies of the most frequent haplotype (haplotype 2) according to the geographic location of the populations studied (Fig. 3). Values generally ≥ 0.8 (mean: 0.84; range: 0.68−1.00) were observed in the Amazon area, while lower frequencies (mean: 0.67; range: 0.55−0.80) were found both in the Andean and the Central/SE regions of South America, suggesting that lowland South America is genetically heterogeneous in respect to the frequencies of the β-globin haplotypes.

Previous suggestions of geographical and linguistic clustering were tested using the molecular analysis of variance (Table 4). In general, over 94% of the variation was observed within populations, about 3–5% being observed among populations within divisions, and 1–2% among the linguistic and geographical divisions established. All except the North/Central America *versus* South America comparisons were statistically significant. The difference between the Andean and lowland areas

TABLE 3. Genetic diversity for the β-globin gene cluster in five continents

	No. of populations	No. of haplotypes	All haplotypes				Excluding haplotype 3			
			H_T	H_S	G_{ST}	G_{ST}'	H_T	H_S	G_{ST}	G_{ST}'
Africa	6	10	0.715	0.651	0.090	0.106	–	–	–	
Europe	5	8	0.622	0.588	0.055	0.068	–	–	–	
Asia	7	14	0.429	0.399	0.071	0.081	–	–	–	
Pacific Islands	8	14	0.425	0.394	0.074	0.084	–	–	–	
Americas	31	19	0.413	0.383	0.073	0.075	0.401	0.374	0.068	0.070
North/Central	3	13	0.583	0.545	0.065	0.095	0.583	0.545	0.065	0.095
South	28	19	0.391	0.365	0.067	0.069	0.378	0.355	0.060	0.062

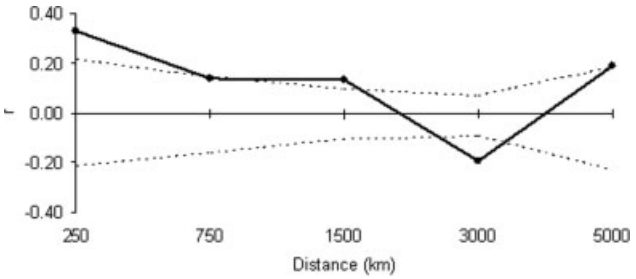

Fig. 2. Spatial correlation for the β globin gene cluster haplotypes. The dotted lines indicate a 95% probability interval around zero obtained with 1,000 permutations of the data, for a null hypothesis of no correlation between genetic and geographic distances.

observed in Figure 3 was confirmed ($P = 0.026$), but besides that, a tripartite geographic division (Andes, Amazon, Center/SE) was also disclosed ($P = 0.010$). Subsequent F_{ST} estimation between these three geographic regions yielded the following results: Andes *versus* Central/SE: $F_{ST} = 0.003$ ($P = 0.061$); Andes *versus* Amazon: 0.045 ($P < 0.001$); Amazon *versus* Central/SE: 0.024 ($P < 0.001$).

DISCUSSION

The first point to be stressed, in relation to the new data reported here, is that they confirm the generally low Aché genetic variability and their closeness to the Guarani than to the Kaingang, as found by Tsuneto et al. (2003) and Schmitt et al. (2004), respectively for HLA and mtDNA results. On the other hand, the genetic diversity observed for the Native American populations is not low as compared to those of other continents. Even dropping haplotype 3 from the data, the average intrapopulation diversity (H_S) is 0.37 in America, not much different from the estimates for the Pacific Islands (0.39) and Asia (0.40). The G'_{ST} estimate is 0.07, which falls within the non-African world range (0.07–0.08). These results are in agreement with Hutz et al. (2002) and Salzano and Callegari-Jacques (2006) conclusions that although some genetic variability may have been lost in the colonization process of the Americas, a non-negligible amount can still be detected both within and between populations.

Mantel's partial correlation between genetic and language distances adjusted for geography ($r = 0.09$) was nonsignificant, so language does not seem to explain much of the geographical distribution of β-globin gene cluster haplotypes in America. Supporting this conclusion, an also nonsignificant result was obtained in an AMOVA comparing the four language groups represented by more than one population (Carib: 5 populations; Je: 4; Tupi: 11; Chocó: 2; $P = 0.173 \pm 0.004$). Although this may not be always true for other regions of the globe or in the Americas (review in Schurr, 2004), evidence from different sources is accumulating on a stronger association between geography and genetics than between language and genetics in South America (e.g., Zegura et al., 2004; Kohlrausch et al., 2005; but see also Fagundes et al., 2002). We did find significant differences between Tropical Forest Languages speakers and the remaining populations. This result, however, can be explained, since Loukotka's (1968) major groups of languages are associated with South America's major geographic areas.

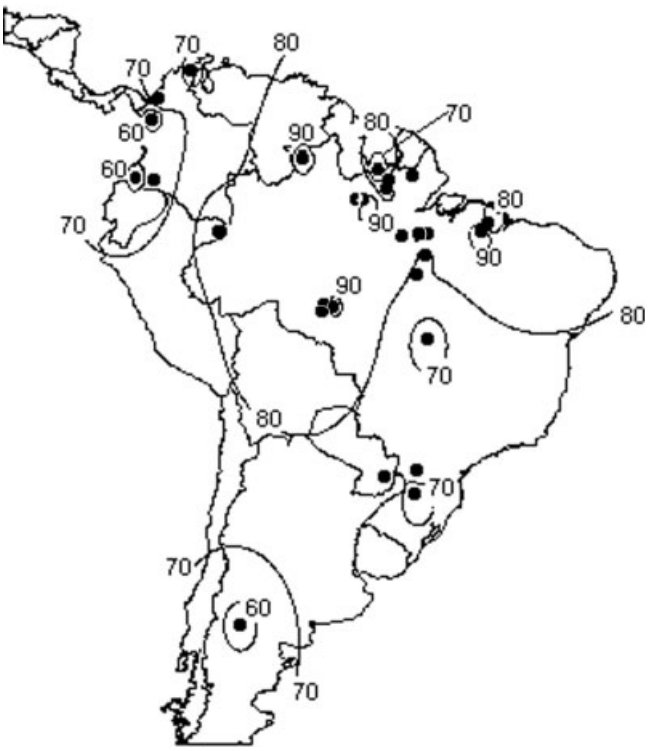

Fig. 3. Isolines delimiting similar South American β-globin gene cluster haplotype 2 frequencies (%).

An interesting finding is the closeness of the β-globin haplotype frequencies observed in the Andean region and in the center/SE of South America, as opposed to the Amazonian area. This pattern, suggested by the correlogram and the haplotype 2 frequencies map, was extended to verify if the Andean could be differentiated from the Center/SE frequencies. An AMOVA, now considering three geographic areas, and F_{ST} comparison analyses indicated that the South American Native genetic landscape should be considered as encompassing the Amazon, Andes, and

TABLE 4. Molecular analysis of variance using 18 β-globin gene cluster haplotypes

Divisions (and number of populations) tested	Number of groups	Number of populations (and chromosomes)	Percentage of variation			Among divisions *P*-value
			Within populations[b]	Among populations within divisions[b]	Among divisions	
North/Central America (3) *versus* South America (28)	2	31 (2,136)	94.2	4.5	1.3	0.131 ± 0.003
Paleo-American/Andean (9) *versus* Tropical Forest (19) Languages[a]	2	28 (1,919)	94.8	3.0	2.2	0.009 ± 0.001
Andean region (6) *versus* Lowland South America (22)	2	28 (1,919)	94.9	3.1	2.0	0.026 ± 0.002
Andean region (6) *versus* Amazon (18) *versus* Center/SE of South America (4)	3	28 (1,919)	95.3	2.6	2.1	0.010 ± 0.001

[a] According to Loukotka (1968).
[b] In all cases, $P < 0.001$.

Center/SE regions. This proposal is supported by the pairwise F_{ST} values between these areas: the Amazon is clearly distinctive from the other two; Andes and Center/SE regions, although more similar, present a statistically non-negligible F_{ST} value.

Two possible alternatives for the interpretation of how this pattern was formed are:

1. A first group of inhabitants carrying lower frequencies of haplotype 2 would colonize South America following mainly the eastern and western coastal routes, then moving inland; a second group, with higher frequencies for this haplotype and maybe coming from the Caribbean islands, would colonize the Amazonian region, displacing the first colonizers to the south.
2. After the crossing of the Isthmus of Panamá, the first migrants split in two groups that expanded into South America. One of them, characterized by low frequencies of haplotype 2, followed the western coast/Andean route to the south, also moving east to Argentina and Brazil. The other, with higher prevalences for this haplotype, moved east and southeastward into the Amazonian area. This hypothesis has already been previously considered by Salzano and Callegari-Jacques (1988), and by Keyeux et al. (2002).

Inferences about past migrations based on a single system are always dangerous. However, the analysis of single loci may identify patterns that could be blurred when several of them are used simultaneously. This approach has been successfully used with the Gm system to reject Native Americans homogeneity (Callegari-Jacques et al., 1993). On the other hand, maps of a single gene or haplotype can be done on the basis of the observed frequencies, without making use of synthetic variables that summarize the information of several loci with various degrees of information loss. Cavalli-Sforza et al. (1994) emphasized the utility of gene-frequency maps and devoted a large part of their impressive analysis of the history and geography of human genes to maps of single selected alleles.

CONCLUSIONS

The β-globin gene cluster can furnish a large amount of information on human history. The new data presented here, integrated with previous results, lead to the following conclusions: (1) There is significant heterogeneity in the Native American's β-globin haplotype frequencies; (2) The Aché, a peculiar Paraguayan population, presented a reduced intrapopulation variability and showed affinities with the Guarani. This information is important for the understanding of their history; (3) Native Americans as a whole show levels of variability not much different from those observed in other continents, with the exception of Africa, which displayed a much higher variation; (4) Geography is more important than language in shaping the pattern of β-globin variability in Native South Americans; (5) The Andes/non-Andes dichotomy for the South America genetic landscape should be replaced by a tripartite division (Amazon, Andes and Center/SE).

ACKNOWLEDGMENTS

Thanks are due to the Fundação Nacional do Índio and Fundación Moisés Bertoni for logistical support in the field. Our research program was approved by the Brazilian Ethics National Committee (Resolution 123/98).

LITERATURE CITED

Bevilaqua LRM, Mattevi VS, Ewald GM, Salzano FM, Coimbra CEA Jr, Santos RV, Hutz MH. 1995. β-Globin gene cluster haplotype distribution in five Brazilian Indian tribes. Am J Phys Anthropol 98:395–401.

Callegari-Jacques SM, Salzano FM, Constans J, Maurieres P. 1993. Gm haplotype distribution in Amerindians: relationship with geography and language. Am J Phys Anthropol 90:427–444.

Campbell L. 1997. American Indian languages: the historical linguistics of Native America. Oxford: Oxford University Press.

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

Chen LZ, Easteal S, Board PG, Kirk RL. 1990. Evolution of β-globin haplotypes in human populations. Mol Biol Evol 7:423–437.

Eastman JR. 2000. IDRISI software. Clark Labs, Clark University, Worcester MA. Software available at http://www.clarklabs.org/.

Excoffier L, Laval G, Schneider S. 2005. Arlequin ver. 3.0: an integrated software package for population genetics data analysis. Evol Bioinformatics Online 1:47–50. Software available at http://cmpg.unibe.ch/software/arlequin3/.

Excoffier L, Smouse PE, Quattro JM. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131:479–491.

Fagundes NJR, Bonatto SL, Callegari-Jacques SM, Salzano FM. 2002. Genetic, geographic, and linguistic variation among South American Indians: possible sex influence. Am J Phys Anthropol 117:68–78.

Fuselli S, Tarazona-Santos E, Dupanloup I, Soto A, Luiselli D, Pettener D. 2003. Mitochondrial DNA diversity in South America and the genetic history of Andean highlanders. Mol Biol Evol 20:1682–1691.

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

Guerreiro JF, Figueiredo MS, Santos SEB, Zago MA. 1992. β-Globin gene cluster haplotypes in Yanomama Indians from the Amazon region of Brazil. Hum Hered 89:629–631.

Guerreiro JF, Figueiredo MS, Zago MA. 1994. Beta-globin gene cluster haplotypes of Amerindian populations from the Brazilian Amazon region. Hum Hered 44:142–149.

Hill K, Hurtado AM. 1996. Ache life history: the ecology and demography of a foraging people. New York: Aldine de Gruyter.

Hutz MH, Callegari-Jacques SM, Almeida SEM, Armborst T, Salzano FM. 2002. Low levels of STRP variability are not universal in American Indians. Hum Biol 74:791–806.

Kaufman L, Carnese FR, Goicoechea A, Dejean C, Salzano FM, Hutz MH. 1998. β-globin gene cluster haplotypes in the Mapuche Indians of Argentina. Gen Mol Biol 21:435–437.

Keyeux G, Rodas C, Gelvez N, Carter D. 2002. Possible migration routes into South America deduced from mitochondrial DNA studies in Colombian Amerindian populations. Hum Biol 74:211–233.

Kohlrausch FB, Callegari-Jacques SM, Tsuneto LT, Petzl-Erler ML, Hill K, Hurtado AM, Salzano FM, Hutz MH. 2005. Geography influences microsatellite polymorphism diversity in Amerindians. Am J Phys Anthropol 126:463–470.

Lewis CM Jr., Tito RY, Lizárraga B, Stone AC. 2004. Land, language, and loci: mtDNA in Native Americans and the genetic history of Peru. Am J Phys Anthropol 127:351–360.

Long JC. 1999. Multiple Locus Haplotype Analysis, version 2.0. [software and documentation distributed by the author]. Section on Population Genetics and Linkage, Laboratory of Neurogenetics, NIAAA, National Institutes of Health, Bethesda, MD.

Long JC, Chakravarti A, Boehm CD, Antonarakis S, Kazazian HG. 1990. Phylogeny of human β-globin haplotypes and its implications for recent human evolution. Am J Phys Anthropol 81:113–130.

Long JC, Williams RC, Urbanek M. 1995. An E-M algorithm and testing strategy for multiple locus haplotypes. Am J Hum Genet 56:799–810.

Loukotka C. 1968. Classification of South American Indian languages. Los Angeles: University of California Latin American Center.

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.

Mantel N. 1967. The detection of disease clustering and a generalized regression approach. Cancer Res 27:209–220.

Mattevi VS, Fiegenbaum M, Salzano FM, Weiss KM, Moore J, Monsalve MV, Devine DV, Hutz MH. 2000. β-Globin gene cluster haplotypes in two North American indigenous populations. Am J Phys Anthropol 112:311–317.

Mesa NR, Mondragón MC, Soto ID, Parra MV, Duque C, Ortíz-Barrientos D, Garcia LF, Velez ID, Bravo ML, Múnera 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.

Mousinho-Ribeiro RC, Pante-de-Souza G, Santos EJM, Guerreiro JF. 2003. Genetic relationships among native Americans based on β-globin gene cluster haplotype frequencies. Gen Mol Biol 26:229–234.

Nei M. 1987. Molecular evolutionary genetics. New York: Columbia University Press.

Nei M, Tajima F, Tateno Y. 1983. Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. J Mol Evol 19:153–170.

Nichols J. 2002. The first American languages. In: Jablonski NG, editor. The first Americans. San Francisco, CA: California Academy of Sciences. p 273–293.

O'Rourke DH, Suarez BK. 1985. Patterns and correlates of genetic variation in South Amerindians. Ann Hum Biol 13:13–31.

Oehme R, Kohne E, Horst J. 1985. DNA polymorphic patterns linked to β-globin genes in German families affected with hemoglobinopathies and thalassemias: a comparison to other ethnic groups. Hum Genet 71:219–222.

Ohta, T. 1993. DISPAN: genetic distance and phylogenetic analysis. University Park: Institute of Molecular Evolutionary Genetics, Pennsylvania State University. Software available at http://www.bio.psu.edu/people/faculty/nei/lab/.

Peakall R, Smouse PE. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol Ecol Notes 6:288–295. Software available at http://www.anu.edu.au/BoZo/GenAlEx/.

Peterson RJ, Goldman D, Long JC. 1999. Nucleotide sequence diversity in non-coding regions of ALDH2 as revealed by restriction enzyme and SSCP analysis. Hum Genet 104:177–187.

Petzl-Erler ML, Luz R, Sotomaior VS. 1993. The HLA polymorphism of two distinctive South-American Indian tribes: the Kaingang and the Guarani. Tissue Antigens 41:227–237.

Ramsay M, Jenkins T. 1987. Globin gene-associated restriction-fragment-length polymorphisms in Southern African peoples. Am J Hum Genet 41:1132–1144.

Rodrigues AD. 1994. Línguas brasileiras: para o conhecimento das línguas indígenas, 2nd ed. São Paulo: Edições Loyola.

Rothhammer F, Silva C. 1989. Peopling of Andean South America. 1989. Am J Phys Anthropol 78:403–410.

Rothhammer F, Silva C. 1992. Gene geography of South America: testing models of population displacement based on archeological evidence. Am J Phys Anthropol 89:441–446.

Rothhammer F, Silva C, Callegari-Jacques SM, Llop E, Salzano FM. 1997. Gradients of HLA diversity in South American Indians. Ann Hum Biol 24:197–208.

Salzano FM, Callegari-Jacques SM. 1988. South American Indians: a case study in evolution. Oxford: Clarendon Press.

Salzano FM, Callegari-Jacques SM. 2006. Amerindian and non Amerindian autosome molecular variability—a test analysis. Genetica 126:237–242.

Salzano FM, Callegari-Jacques SM, Weimer TA, Franco MHLP, Hutz MH, Petzl-Erler ML. 1997. Electrophoretic protein polymorphisms in Kaingang and Guarani Indians of southern Brazil. Am J Hum Biol 9:505–512.

Salzano FM, Hutz MH, Salamoni SP, Rohr P, Callegari-Jacques SM. 2005. Genetic support for proposed patterns of relationship among lowland South American languages. Curr Anthropol 46(Suppl):S121–S129.

Schmitt R, Bonatto SL, Freitas LB, Muschner VC, Hill K, Hurtado AM, Salzano FM. 2004. Extremely limited mitochondrial variability among the Aché natives of Paraguay. Ann Hum Biol 31:87–94.

Schurr TG. 2004. The peopling of the New World: perspectives from molecular anthropology. Annu Rev Anthropol 33:551–583.

Shimizu K, Harihara S. 2004. Re-evaluation of the β-globin gene cluster haplotypes in Japanese, Koreans, and three Colombian Amerinds (Wayuu, Kamsa, and Inga tribes). Anthropol Sci 112:109–114.

Simoni L, Tarazona-Santos E, Luiselli D, Pettener D. 2000. Genetic differentiation of South American native populations inferred from classical markers: from explorative analyses to a working hypothesis. In: Renfrew C, editor. America past, America present: genes and languages in the Americas and beyond. Cambridge: McDonald Institute for Archaeological Research. p 125–138.

Stephens M, Donnelly P. 2003. A comparison of Bayesian methods for haplotype reconstruction from population genotype data. Am J Hum Genet 73:1162–1169. Software available at http://www.stat.washington.edu/stephens/phase/.

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

Tarazona-Santos E, Carvalho-Silva DR, Pettener D, Luiselli D, De Stefano GF, Labarga CM, Rickards O, Tyler-Smith C, Pena SD, Santos FR. 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.

Tsuneto LT, Probst CM, Hutz MH, Salzano FM, Rodriguez-Delfin LA, Zago MA, Hill K, Hurtado AM, Ribeiro-dos-Santos AKC, Petzl-Erler ML. 2003. HLA class II diversity in seven Ameridian populations. Clues about the origins of the Aché. Tissue Antigens 62:512–526.

Villalobos-Arámbula AR, Rivas F, Sandoval L, Perea FJ, Casas-Castañeda M, Cantu JM, Ibarra B. 2000. β^A Globin gene haplotypes in Mexican Huichols: genetic relatedness to other populations. Am J Hum Biol 12:201–206.

Wainscoat JS, Hill AVS, Boyce AL, Flint J, Hernandez M, Thein SL, Old JM, Lynch JR, Falusi AG, Weatherall DJ, Clegg JB. 1986. Evolutionary relationships of human populations from an analysis of nuclear DNA polymorphisms. Nature 319:491–493.

Weir BS. 1996. Genetic data analysis II: methods for discrete population genetic data. Sunderland, MA: Sinauer.

Weir BS, Cockerham CC. 1984. Estimating F-statistics for the analysis of population structure. Evolution 38:1358–1370.

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.