

Geography Influences Microsatellite Polymorphism Diversity in Amerindians

Fabiana B. Kohlrausch,¹ Sidia M. Callegari-Jacques,² Luiza T. Tsuneto,³ M. Luiza Petzl-Erler,³ Kim Hill,⁴ A. Magdalena Hurtado,⁴ Francisco M. Salzano,¹ and Mara H. Hutz^{1*}

¹*Departamento de Genética, Universidade Federal do Rio Grande do Sul, 91501-970 Porto Alegre, Rio Grande do Sul, Brazil*

²*Departamento de Estatística, Universidade Federal do Rio Grande do Sul, 91540-000 Porto Alegre, Rio Grande do Sul, Brazil*

³*Departamento de Genética, Universidade Federal do Paraná, 81531-990 Curitiba, Brazil*

⁴*Department of Anthropology, University of New Mexico, Albuquerque, New Mexico 87131*

KEY WORDS South American Indians; genetic variability; geography; language

ABSTRACT Data related to 15 short tandem repeat polymorphisms (STRPs) are reported for four South American Indian populations, and integrated with previous Brazilian Indian results. Overall heterozygosities varied significantly among groups (Kruskal-Wallis test, $P = 0.002$). The lowest levels of heterozygosity were observed in the Aché, Ayoreo, and Surui, an expected finding considering their isolation and ethnohistory. Genetic distance

and gene diversity analyses suggested that geography was a good predictor of genetic affinity among these Native Americans. New evidence from this study supports the hypothesis that the Aché population descends from a Gê group that preceded the Guarani colonization of Paraguay. *Am J Phys Anthropol* 126:463–470, 2005.

© 2004 Wiley-Liss, Inc.

South American Indians present a remarkable number of populations spread over a vast territorial area. They have different demographic, linguistic, and cultural patterns, and exhibit considerable biological diversity. Although numerous studies based on protein polymorphisms and DNA markers (reviewed in Salzano, 2002) have already been performed, only the progressive accumulation of data can provide a better understanding of the evolutionary history of these groups.

Genetic diversity has been correlated with cultural, linguistic, and geographic differences. Linguistic affinities were considered clues to population history, but geographic patterns of genetic diversity also allowed inferences about population histories. Cavalli-Sforza et al. (1992) stated that linguistic evolution has paralleled the genetic differentiation in our species. Barbujani (1997) showed that language and geographic proximity were equally good predictors of genetic affinity. However, Rosser et al. (2000) suggested that populations are related primarily on the basis of geography, rather than linguistic affinity. Although linguistic boundaries in Europe frequently coincide with gene frequency discontinuities (Barbujani, 2000), in the Americas the relationship between genetic differentiation and linguistic affiliation is less clear-cut. The Amerind-speaking tribes of South America generally fail or exhibit weak correspondence between their genetic rela-

tionship and linguistic affiliation (Salzano and Callegari-Jacques, 1988; Ward et al., 1993).

As an example of the type of problems that can be considered in these studies, attention can be focused on the Aché (or Guayaki) Indians of Paraguay. They are physically and genetically dissimilar to most other South Amerindian group; although they do not show evidence of European or African admixture, they present light skin and thick beards (not common in other Native American groups; Salzano and Callegari-Jacques, 1988). The data available suggest that the Aché were hunters and gatherers who lived genetically and socially isolated from all neighboring populations for a long period of time prior to contact with non-Indians in the late twentieth cen-

Grant sponsor: Programa de Apoio a Núcleos de Excelência; Grant sponsor: Conselho Nacional de Desenvolvimento Científico e Tecnológico; Grant sponsor: Financiadora de Estudos e Projetos; Grant sponsor: Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul; Grant sponsor: Pró-Reitoria de Pesquisa da Universidade Federal do Rio Grande do Sul.

*Correspondence to: Mara H. Hutz; Departamento de Genética, UFRGS, Caixa Postal 15053, 91501-970 Porto Alegre, Rio Grande do Sul, Brazil. E-mail: mara.hutz@ufrgs.br

Received 25 March 2003; accepted 26 August 2003.

DOI 10.1002/ajpa.20042

Published online 5 August 2004 in Wiley InterScience (www.interscience.wiley.com).

TABLE 1. Characteristics of four populations investigated

Characteristics	Populations			
	Aché	Ayoreo	Caingang	Guarani
Localities	Arroyo Bandera Chupa-pou	Maria Auxiliadora Pozo Verde Tobité	Nonoai	Amambai Limão Verde Porto Lindo
Geographic location	55°W, 23°S 56°W, 24°S	58°W, 16°S 63°W, 22°S	52°W, 27°S	55°W, 23°S 55°W, 23°S 54°W, 23°S
Country	South Paraguay	South Bolivia and north Paraguay	South Brazil	Central Brazil
Linguistic group ¹	Tupi	Zamuco	Gê	Tupi
Non-Indian influence	Minimal	Minimal	Extensive	Extensive

¹ According to Greenberg (1987).

tury (Hill and Hurtado, 1996). The Aché are one of the southerly representatives of the tropical forest Tupi linguistic stock. Their origin has been controversial, some authors considering them as a differentiated Guarani group, while others claim that they descend from a Gê population that preceded the Guarani colonization of Paraguay (Hill and Hurtado, 1996).

Several studies about the biological classification of the Aché were recently performed, but the results are not conclusive. Clariá et al. (1998) analyzed electrophoretic polymorphisms at loci coding for erythrocyte and plasma proteins in two populations from the Gran Chaco and the Aché from the Paraguayan forest, observing that the Aché differ from the Chaco tribes. Goicoechea et al. (2001) obtained data on 17 blood group and protein genetic systems among two Indian groups of Paraguay, and integrated this information with previous results available for other four populations (including the Aché). They found that the Chaco tribes clearly separated from the Aché, the only forest group. Battilana et al. (2002) studied 12 *Alu* insertion polymorphisms and 20 protein and blood group markers in four South Amerindian groups, two of the Tupi and two of the Gê linguistic families, and could not decide between the two hypotheses concerning their classification, but the view that they are a differentiated Guarani group seemed more likely. Tsuneto et al. (2003) described an HLA DRB1 allele in the Aché previously found only among the Guarani-M'bya from the Brazilian state of Paraná. Finally, Gaspar et al. (2002), in a study with genes related to cancer susceptibility in seven Amerindian populations, supported the second hypothesis, since the Aché showed a closer relationship with the Xavante, a Gê-speaking population.

The present study investigates variability in 15 short tandem repeat polymorphisms (STRPs) in the Aché, Guarani, Caingang, and Ayoreo populations, and integrates the results with those already available for five other Brazilian Indian groups, trying to obtain a comprehensive picture of the diversity of these markers in South American Natives. These polymorphisms, with their high heterozygosity, vast number of alleles, and high mutation rates, are par-

ticularly useful in investigating recent evolutionary events. The following specific questions were also addressed: 1) What are the relative contributions of language and geography for the differentiation of these markers among Amerindian groups? 2) Since the origin of the Aché is still obscure, would STRPs provide a clue about it?

SUBJECTS AND METHODS

Populations

General information about the four populations studied is presented in Table 1. The Caingang and Guarani are the two major Amerindian tribes living in southern Brazil, and speak languages of the Gê and Tupi stocks, respectively. Despite living side by side for many centuries, the Guarani and Caingang populations still differ in many aspects of culture and biology (Petzl-Erler et al., 1993). Both tribes are in an advanced stage of acculturation. The Aché, who live in the tropical forest, and the Ayoreo, of the Gran Chaco area, are more isolated, presenting minimal non-Indian influence. The Aché are linguistically related to the Guarani and geographically close to the Ayoreo. Figure 1 shows the geographic location of the four tribes studied herein, together with the five Amerindian populations previously investigated by Hutz et al. (2002), which were included in the joint analysis. The present sample consisted of 198 individuals from the Aché (50), Guarani (50), Caingang (50), and Ayoreo (48) populations.

Laboratory analyses

Samples were collected with anticoagulant, refrigerated shortly afterwards, and transported in this condition to Porto Alegre (Aché, Ayoreo, and Caingang) or Curitiba (Guarani), where DNA was isolated using standard procedures.

In total, 15 STR loci (all repetitions of a four-base-pair motif), D2S1338, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D19S433, D21S11, FGA, vWA, CSF1PO, TPOX, and TH01, were determined using the AmpFISTR® Identifier™ amplification kit (Applied Biosystems, Foster City, CA). Amplifications were performed following the instructions provided in the kit user manual,



Fig. 1. Geographic location of four populations investigated, and of five other Brazilian Indian groups previously studied.

with the recommended DNA amount (1.0–2.0 ng). Electrophoresis of the amplified fragments was performed in a 310 ABI PRISM™ Genetic Analyzer (Applied Biosystems), according to the manufacturer's recommended protocol. The GeneScan™ version 3.2.1 analysis software was used to track lanes and measure fragment sizes, while the Genotyper™ version 2.5.2 software was employed to automatically designate alleles by comparison with locus-specific allelic ladders.

Statistical analyses

Allele frequencies were obtained directly by gene counting. Hardy-Weinberg equilibrium and heterogeneity among populations were tested by the exact test of Guo and Thompson (1992), using GENEPOP version 3.1d software (Raymond and Rousset, 1995). Average variances and their standard errors across loci were estimated using the MICROSAT program (Minch, 2001). Mean heterozygosities and their standard errors were calculated as described in Nei (1987), using DISPAN software (Ota, 1993). Since these estimates do not follow a normal distribution, they were compared across populations with the Kruskal-Wallis test. Dunn's method (Zar, 1999) was further used to identify statistical significant differences between populations at least at the 0.05 level.

To test for population subdivision, 2n analysis of molecular variance (AMOVA) (Weir and Cockerham, 1984; Excoffier et al., 1992; Weir, 1996) was performed, using language and geography as subdivision criteria. These analyses were done with Arlequin software version 2001 (Schneider et al., 2001).

Genetic distances among populations were estimated with the $(\delta\mu)^2$ distance (Goldstein et al., 1995), which is independent of population size and incorporates the stepwise mutation model. The neighbor-joining method (Saitou and Nei, 1987) was used to construct the tree, based on the genetic distance matrices with the POPTREE program (Takezaki, 2001). The corresponding dendrograms were visualized with the TREEVIEW version 1.5.2 program (Page, 1996). The reliability of the tree was tested by 1,000 bootstrap replications (Hedges, 1992).

Geographic distances between the approximate locations of each tribe were computed as great-circle distances calculated from their latitudinal and longitudinal coordinates, using a facility provided at the site <http://www.geocities.com/TheTropics/Shores/3098/distance.html>. Two sets of linguistic distances were determined, as based on classification by Loukokta (1968) and Greenberg (1987) of South American native languages. In both matrices, a zero distance was assigned to populations belonging to the same linguistic group, with values increasing as tribes had more distinct linguistic features until reaching a value of three, attributed to tribes that speak languages classified as different stocks.

Bivariate and partial correlation coefficients between the genetic, geographic, and linguistic distances were calculated (the partial correlation provides a measure of association between two quantitative measures, controlling for the variation of a third variable related to the other two). The statistical significance of these coefficients was obtained by the Mantel test (Mantel, 1967; Smouse and Long, 1992), using Arlequin software version 2001 (Schneider et al., 2001).

RESULTS AND DISCUSSION

Number of alleles, mean variance, and heterozygosity

Allele frequency distributions for the 15 STR loci found in the four populations investigated are shown in Table 2. The most common allele frequencies ranged from 0.89 (*D18S51*, allele 14, Aché) to 0.20 (*D19S433*, allele 13, Caingang). The most frequent allele was the same in the four populations for only two systems: *D3S1358* (*15) and *D18S51* (*14). The 60 tests for departure from Hardy-Weinberg equilibrium showed that seven were statistically significant at the 0.05 significance level. However, after Bonferroni's correction for multiple tests, only one case remained significant for heterozygote deficiency (*D16S539* locus in the Guarani). This deficiency could be due to the presence of null alleles

TABLE 2. Allele frequency distributions for 15 STRP systems in four South Amerindian populations

Systems and alleles	Populations				Systems and alleles	Populations			
	Aché (2n = 100)	Ayoreo (2n = 96)	Caingang (2n = 100)	Guarani (2n = 100)		Aché (2n = 200)	Ayoreo (2n = 90)	Caingang (2n = 200)	Guarani (2n = 200)
CSF1PO					D18S51				
*9	0.000	0.000	0.020	0.000	*12	0.040	0.031	0.040	0.200
*10	0.020	0.281	0.220	0.190	*13	0.000	0.094	0.200	0.100
*11	0.080	0.188	0.330	0.390	*14	0.890	0.344	0.240	0.320
*12	0.710	0.188	0.390	0.400	*15	0.060	0.000	0.140	0.150
*13	0.190	0.333	0.040	0.020	*16	0.000	0.094	0.140	0.030
*14	0.000	0.010	0.000	0.000	*17	0.000	0.281	0.110	0.160
D2S1338					*18	0.010	0.135	0.040	0.040
*16	0.000	0.000	0.000	0.010	*19	0.000	0.021	0.030	0.000
*17	0.080	0.000	0.260	0.200	*21	0.000	0.000	0.050	0.000
*18	0.740	0.083	0.050	0.070	*23	0.000	0.000	0.010	0.000
*19	0.000	0.302	0.170	0.100	D19S433				
*20	0.000	0.000	0.040	0.140	*12	0.050	0.000	0.060	0.010
*21	0.000	0.000	0.060	0.010	*12.2	0.270	0.000	0.120	0.030
*22	0.000	0.292	0.080	0.190	*13	0.180	0.031	0.200	0.300
*23	0.180	0.323	0.230	0.230	*13.2	0.010	0.031	0.160	0.120
*24	0.000	0.000	0.040	0.020	*14	0.000	0.198	0.170	0.230
*25	0.000	0.000	0.050	0.030	*14.2	0.030	0.010	0.010	0.010
*26	0.000	0.000	0.020	0.000	*15	0.230	0.386	0.150	0.120
D3S1358					*15.2	0.230	0.271	0.090	0.120
*14	0.000	0.000	0.020	0.040	*16	0.000	0.063	0.040	0.050
*15	0.560	0.542	0.370	0.470	*16.2	0.000	0.010	0.000	0.010
*16	0.000	0.177	0.340	0.310	D21S11				
*17	0.440	0.073	0.190	0.140	*28	0.000	0.000	0.270	0.000
*18	0.000	0.208	0.060	0.030	*29	0.130	0.000	0.120	0.080
*19	0.000	0.000	0.020	0.010	*30	0.490	0.438	0.180	0.400
D5S818					*30.2	0.040	0.000	0.000	0.000
*7	0.000	0.052	0.130	0.320	*31	0.000	0.010	0.040	0.020
*9	0.010	0.000	0.050	0.070	*31.2	0.000	0.469	0.140	0.090
*10	0.000	0.000	0.020	0.020	*32.2	0.270	0.083	0.150	0.200
*11	0.670	0.239	0.390	0.560	*33.2	0.070	0.000	0.090	0.210
*12	0.320	0.667	0.290	0.020	*37	0.000	0.000	0.010	0.000
*13	0.000	0.042	0.080	0.010	TH01				
*14	0.000	0.000	0.040	0.000	*6	0.420	0.281	0.190	0.570
D7S820					*7	0.190	0.489	0.500	0.280
*7	0.000	0.000	0.010	0.000	*8	0.000	0.000	0.080	0.020
*8	0.010	0.000	0.200	0.040	*9	0.000	0.000	0.050	0.000
*9	0.000	0.010	0.050	0.000	*9.3	0.390	0.230	0.180	0.130
*10	0.010	0.177	0.360	0.140	FGA				
*11	0.410	0.563	0.170	0.470	*18	0.000	0.000	0.000	0.010
*12	0.570	0.240	0.200	0.310	*19	0.120	0.010	0.060	0.170
*13	0.000	0.010	0.010	0.040	*20	0.000	0.000	0.080	0.000
D8S1179					*21	0.000	0.010	0.110	0.000
*8	0.000	0.000	0.010	0.000	*22	0.050	0.000	0.140	0.050
*9	0.010	0.000	0.000	0.000	*23	0.320	0.375	0.140	0.130
*10	0.440	0.063	0.060	0.090	*23.2	0.000	0.021	0.000	0.000
*11	0.000	0.010	0.040	0.000	*24	0.030	0.375	0.260	0.140
*12	0.290	0.469	0.130	0.080	*25	0.430	0.052	0.160	0.240
*13	0.250	0.219	0.300	0.120	*26	0.050	0.073	0.020	0.210
*14	0.010	0.177	0.230	0.490	*27	0.000	0.084	0.030	0.050
*15	0.000	0.010	0.180	0.210	TPOX				
*16	0.000	0.052	0.030	0.010	*7	0.000	0.000	0.010	0.000
*17	0.000	0.000	0.020	0.000	*8	0.450	0.323	0.480	0.310
D13S317					*9	0.000	0.000	0.040	0.000
*7	0.000	0.000	0.000	0.050	*10	0.000	0.000	0.050	0.020
*8	0.000	0.000	0.020	0.010	*11	0.270	0.219	0.240	0.520
*9	0.210	0.761	0.250	0.260	*12	0.280	0.458	0.180	0.150
*10	0.000	0.000	0.010	0.010	vWA				
*11	0.340	0.010	0.200	0.320	*14	0.000	0.000	0.070	0.010
*12	0.230	0.010	0.380	0.170	*15	0.010	0.021	0.080	0.000
*13	0.220	0.209	0.130	0.130	*16	0.420	0.281	0.390	0.290
*14	0.000	0.010	0.010	0.050	*17	0.280	0.687	0.370	0.340
D16S539					*18	0.280	0.010	0.040	0.270
*9	0.490	0.187	0.320	0.510	*19	0.010	0.000	0.030	0.080
*10	0.000	0.240	0.100	0.020	*20	0.000	0.000	0.020	0.010
*11	0.000	0.552	0.310	0.190					
*12	0.200	0.021	0.210	0.230					
*13	0.300	0.000	0.060	0.050					
*14	0.010	0.000	0.000	0.000					

TABLE 3. Total number of alleles, number in parentheses of those present in that population only, mean number of alleles, and mean heterozygosity for 15 STRPs in four South Amerindian populations¹

Locus	Populations				Total
	Aché	Ayoreo	Caingang	Guarani	
CSF1PO	4 (0)	5 (1)	5 (1)	4 (0)	6
D2S1338	3 (0)	4 (0)	10 (1)	10 (1)	11
D3S1358	2 (0)	4 (0)	6 (0)	6 (0)	6
D5S818	3 (0)	4 (0)	7 (1)	6 (0)	7
D7S820	4 (0)	5 (0)	7 (1)	5 (0)	7
D8S1179	5 (1)	7 (0)	9 (2)	6 (0)	10
D13S317	4 (0)	5 (0)	7 (0)	8 (1)	8
D16S539	4 (1)	4 (0)	5 (0)	5 (0)	6
D18S51	4 (0)	7 (0)	10 (2)	7 (0)	10
D19S433	7 (0)	8 (0)	9 (0)	10 (0)	10
D21S11	5 (0)	4 (0)	8 (2)	6 (0)	9
FGA	6 (0)	8 (1)	9 (1)	8 (1)	11
TH01	3 (0)	3 (0)	5 (1)	4 (0)	5
TPOX	3 (0)	3 (0)	6 (2)	4 (0)	6
vWA	5 (0)	4 (0)	7 (0)	6 (0)	7
Mean number of alleles per locus	4.1 ± 1.30	5.0 ± 1.69	7.3 ± 1.76	6.3 ± 1.95	7.9 ± 2.05
Mean heterozygosity ± SE	0.58 ± 0.04 ^b	0.63 ± 0.03 ^b	0.77 ± 0.02 ^a	0.71 ± 0.02 ^{a,b}	0.74 ± 0.02

¹ Comparison among heterozygosities: Kruskal-Wallis $\chi^2 = 19.910$; $df = 3$; $P < 0.001$. Heterozygosities indicated by same letter do not differ significantly by Dunn's test.

private to this group, but the small sample sizes involved in the comparisons, and eventual deviations from random mating that may occur in these relatively small populations, could also account for this finding.

The mean number of alleles per locus is considered a reasonable indicator of genetic variation within a population, given the assumption of mutation-drift equilibrium and similar sample sizes among populations (Nei, 1987). In total, 119 different alleles were detected in the four populations (Table 3), with a range of 11 (*D2S1338* and *FGA*) to five alleles per locus (*TH01*). The mean number of alleles per locus varied from 7.3 ± 1.8 in the Caingang to 4.1 ± 1.3 in the Aché. Probably some of the alleles observed in the Caingang are not of Amerindian origin. This population has the highest degree of genetic admixture with non-Indians among those investigated here (4.5–6.6%; Salzano et al., 1997; Callegari-Jacques and Salzano, 1999). Overall, the mean number of alleles per locus detected in these four Amerindian groups (7.9 ± 2.0) was very similar to that described for five other South American native populations, using 13 of the 15 STRPs included in the present study (7.0 ± 2.9 ; Hutz et al., 2002).

The mean heterozygosity (Table 3) ranged between 0.77 ± 0.02 (Caingang) and 0.58 ± 0.04 (Aché). Overall heterozygosities varied significantly among groups (Kruskal-Wallis test, $P < 0.001$). After pairwise comparisons (Dunn test), two subgroups could be discerned: Caingang (higher values), and Ayoreo plus Aché (lower heterozygosities). Guarani heterozygosity did not differ significantly from both groups.

Table 4 shows the average heterozygosity and mean allele length variance values considering the 13 STRPs for which data were available for the five previously investigated Amerindian populations (Hutz et al., 2002). Mean heterozygosities were sig-

TABLE 4. Mean variance in repeat number and mean heterozygosity, considering 13 STRPs in nine Amerindian populations¹

Populations	Variance in repeat number		Heterozygosity	
	Mean	SE	Mean	SE
Aché	2.09	0.50	0.58 ^a	0.04
Ayoreo	1.91	0.29	0.61 ^a	0.03
Caingang	2.89	0.54	0.76	0.02
Guarani	2.71	0.61	0.69	0.02
Gavião ²	3.19	0.76	0.73	0.02
Suruí ²	2.09	0.39	0.64 ^a	0.03
Xavante ²	2.34	0.53	0.69	0.03
Zoró ²	2.74	0.67	0.68	0.03
Wai Wai ²	2.79	0.60	0.71	0.03

¹ Tests for statistical significance of the differences: 1) Heterozygosity: Kruskal-Wallis $\chi^2 = 24.060$; $df = 8$; $P = 0.002$. Values indicated by same letter do not differ significantly by Dunn's test. 2) Variance in repeat number: Kruskal-Wallis $\chi^2 = 3.754$; $df = 8$; $P = 0.879$.

² Variances were calculated from data of Hutz et al. (2002). *D2S1338* and *D19S433* were not included.

nificantly different when the nine populations were considered (Kruskal-Wallis test, $P = 0.002$). The Suruí, Ayoreo, and Aché have the lowest heterozygosity levels, which is expected considering their isolation and ethnohistory (Salzano et al., 1978; Battilana et al., 2002; Hutz et al., 2002). The mean allele length variance ranged from 3.19 (Gavião) to 1.91 (Ayoreo), but these differences were nonsignificant.

Genetic affinities among nine South Amerindian populations

Figure 2 presents the neighbor-joining tree obtained when the four populations studied here were compared with the five other uniformly studied for the same 13 microsatellite loci. In general, although the bootstrap values are not very high, these tribes

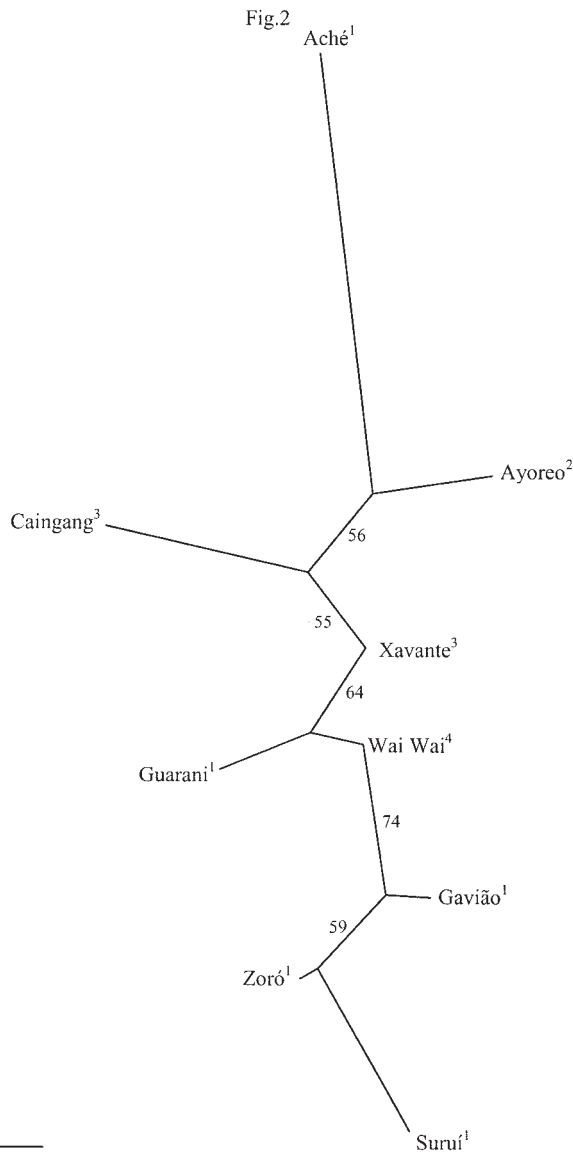


Fig. 2. Neighbor-joining tree based on $(\delta\mu)^2$ distance for 13 STRPs in nine South Amerindian populations. Superscript numbers refer to language groups. 1, Tupi; 2, Zamuco; 3, Gê; 4, Carib. Numbers associated with internal nodes of tree are bootstrap values.

are primarily distributed according to their geographic location. The Aché associated first with the Ayoreo and then with the Caingang. These populations live in Paraguay and in south Brazil (Fig. 1). This more southerly cluster joined the groups from central Brazil (Xavante and Guarani) and then the Amazonian tribes (Wai Wai, Gavião, Zoró, and Suruí).

The Gavião, Zoró, and Suruí, besides living in close geographic proximity (Fig. 1), have a common language, and share many cultural traits (Bevilaqua et al., 1995; Hutz et al., 1997, 1999, 2002; Salzano et al., 1998); the Xavante and Caingang, who speak languages from the same linguistic group, were also close to each other, suggesting that similarity in

TABLE 5. Microsatellite diversity analysis using different subdivision criteria in South Amerindian populations¹

Source of variation	Variance	% total	<i>P</i>	ϕ -statistics
I. Geographic criterion				
Among groups	0.139	2.89	0.022	0.029
Among populations	0.469	9.77	<0.0001	0.101
within groups				
Within populations	4.192	87.34	<0.0001	0.127
II. Language criterion				
Among groups	0.011	0.22	0.312	0.002
Among populations	0.278	5.70	<0.0001	0.057
within groups				
Within populations	4.592	94.08	<0.0001	0.059

¹ Geographic criterion: group 1 (Paraguayan): Aché and Ayoreo; group 2 (Amazonian): Gavião, Suruí, Zoró, and Wai Wai. Language criterion: group 1 (Tupi linguistic group): Gavião, Suruí, Zoró, and Guarani; Group 2 (Gê linguistic group): Caingang and Xavante.

language may have also contributed to the genetic relatedness of these populations. These findings were further explored by a gene diversity analysis (Table 5). The importance of geography for gene differentiation was further corroborated by the AMOVA tests ($P = 0.022$), while no heterogeneity between the two language groups established was disclosed by this analysis ($P = 0.312$). These results were also confirmed by partial correlations between genetic and geographic distances, controlling for the language variable, using classifications by Loukocota (1968) ($r = 0.414$; $P = 0.039$) or Greenberg (1987) ($r = 0.423$; $P = 0.037$). The partial correlations between genetic and language distances were nonsignificant.

Overall, genetic distances and the corresponding neighbor-joining tree, as well as the gene diversity analyses and partial correlations, suggest that geography influences the genetic relationships observed among these groups. Rosser et al. (2000) and Zerjal et al. (2001) analyzed Y-chromosome haplogroups in European populations and suggested that these populations are related primarily by geography, indicating, however, that linguistic differences have also been important for their genetic differentiation. Sanchez-Mazas (2001) observed that in relation to two HLA class II loci, *DRB1* and *DPB1*, genetic distances were significantly correlated with geography between continents, while genetic diversity within Africa was mostly explained by linguistic differences. As pointed out by Barbujani (1997, 2000), weak geographic barriers and weak linguistic barriers may together form strong barriers to gene flow.

The Gm haplotype distribution in Native Americans reveals clear and clinal differences between North, Central, and South American Indian populations (Callegari-Jacques et al., 1993). Salzano et al. (1991), Callegari-Jacques et al. (1994), and Hutz et al. (2000) suggested, based on protein and DNA markers, that the Amazon river could constitute a barrier to north-south gene flow. Eight of the nine

tribes included in the present study are situated south of the Amazon river, but the Plata-Paraná-Paraguay hydrographic system in the south could also have been a geographic barrier to gene flow. At the tribal level, groups exchange genes with their neighbors in varying degrees, and diversification may occur more rapidly in one region than another due to the fission-fusion processes which occur within a tribe (Salzano and Callegari-Jacques, 1988).

How can the data obtained in the present investigation help elucidate the Aché classification? The STRP markers add evidence to support the hypothesis that the Aché population may have descended from a Gê group that preceded the Guaraní colonization of Paraguay, since they show a closer relationship with the Caingang and Xavante, Gê-speaking populations, than with the Tupi groups considered in this analysis. These results are more similar to those obtained by Gaspar et al. (2002) than to those described by Battilana et al. (2002) and Tsuneto et al. (2003) mentioned in the introduction. Nevertheless, the long branch observed in Figure 2 shows their marked distinctiveness from the other Native South American populations, as consistently shown in previous studies (Battilana et al., 2002; Gaspar et al., 2002; Tsuneto et al., 2003).

ACKNOWLEDGMENTS

Thanks are due to the Fundação Nacional do Índio for permission to study the Indians and for help in the field, and to the Fundación Bertoni for logistical assistance. The Indian leaders and the subjects of the investigation were adequately informed about the aims of the study and gave their approval, which is also acknowledged. Our research program was approved by the Brazilian Ethics National Committee (Resolution 123/98). We are very grateful to Fernando Pagés and Ricardo Moreno for the collection, and to Jaqueline Battilana and Loreta B. Freitas for assistance in the retrieval, of Ayoreo DNA samples.

LITERATURE CITED

- Barbujani G. 1997. DNA variation and language affinities. *Am J Hum Genet* 61:1011–1014.
- Barbujani G. 2000. Geographic patterns: how to identify them and why. *Hum Biol* 72:133–153.
- Battilana J, Bonatto SL, Freitas LB, Hutz MH, Weimer TA, Callegari-Jacques SM, Batzer MA, Hill K, Hurtado AM, Petzl-Erler ML, Salzano FM. 2002. *Alu* insertions versus blood group plus protein genetic variability in four Amerindian populations. *Ann Hum Biol* 29:334–347.
- Bevilaqua LRM, Mattevi VS, Ewald GM, Salzano FM, Coimbra CEA Jr, Santos RV, Hutz MH. 1995. Beta-globin gene cluster haplotype distribution in five Brazilian Indian tribes. *Am J Phys Anthropol* 98:395–401.
- Callegari-Jacques SM, Salzano FM. 1999. Brazilian Indian/non-Indian interactions and their effects. *Cienc Cult* 51:166–174.
- Callegari-Jacques SM, Salzano FM, Constans J, Maurieres P. 1993. Gm haplotype distribution in Amerindians: relationships with geography and language. *Am J Phys Anthropol* 90:427–444.
- Callegari-Jacques SM, Salzano FM, Weimer TA, Hutz MH, Black FL, Santos SE, Guerreiro JF, Mestriner MA, Pandey JP. 1994. Further blood genetic studies on Amazonian diversity—data from four Indian groups. *Ann Hum Biol* 21:465–481.
- Cavalli-Sforza LL, Minch E, Mountain JL. 1992. Coevolution of genes and languages revisited. *Proc Natl Acad Sci USA* 89:5620–5624.
- Clariá DM, Demarchi DA, Moreno Azorero R, Gardenal CN. 1998. Protein polymorphism in three South Amerindian populations. *Ann Hum Biol* 25:581–588.
- 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.
- Gaspar PA, Hutz MH, Salzano FM, Hill K, Hurtado AM, Petzl-Erler ML, Tsuneto LT, Weimer TA. 2002. Gene polymorphisms of *CYP1A1*, *CYP2E1*, *GSTM1*, *GSTT1*, and *TP53* genes in Amerindians. *Am J Phys Anthropol* 119:249–256.
- Goicoechea AS, Carnese FR, Dejean C, Avena AS, Weimer TA, Estalote AC, Simões MLMS, Palatnik M, Salamoni SP, Salzano FM, Callegari-Jacques SM. 2001. New genetic data on Amerindians from the Paraguayan Chaco. *Am J Hum Biol* 13:660–667.
- Goldstein DB, Ruiz-Linares A, Cavalli-Sforza LL, Feldman MW. 1995. Genetic absolute dating based on microsatellites and the origin of modern humans. *Proc Natl Acad Sci USA* 92:6723–6727.
- Greenberg JH. 1987. *Language in the Americas*. Stanford: Stanford University Press.
- Guo SW, Thompson EA. 1992. Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics* 48:361–372.
- Hedges SB. 1992. The number of replications needed for accurate estimation of the bootstrap P value in phylogenetic studies. *Mol Biol Evol* 9:366–361.
- Hill K, Hurtado AM. 1996. *Aché life history*. New York: Aldine de Gruyter.
- Hutz MH, Mattevi VS, Callegari-Jacques SM, Salzano FM, Coimbra Junior CE, Santos RV, Carnese RF, Goicoechea AS, Dejean CB. 1997. *DIS80* locus variability in South American Indians. *Ann Hum Biol* 24:249–255.
- Hutz MH, Callegari-Jacques SM, Bortolini MC, Salzano FM. 1999. Variability in nDNA, mtDNA, and proteins: a test case. In: Papiha SS, Deka R, Chakraborty R, editors. *Genomic diversity: applications in human population genetics*. New York: Plenum. p 23–32.
- Hutz MH, Almeida S, Coimbra CE Jr, Santos RV, Salzano FM. 2000. Haplotype and allele frequencies for three genes of the dopaminergic system in South American Indians. *Am J Hum Biol* 12:638–645.
- Hutz MH, Callegari-Jacques SM, Almeida SEM, Armborst T, Salzano FM. 2002. Low levels of STRP variability are not universal in Amerindians. *Hum Biol* 74:791–806.
- Loukotka C. 1968. *Classification of South American Indian languages*. Los Angeles: Latin American Center, University of California.
- Mantel N. 1967. The detection of disease clustering and a generalized regression approach. *Cancer Res* 27:209–220.
- Minch E. 2001. <http://hpgl.stanford.edu/projects/microsat/>.
- Nei M. 1987. *Molecular evolutionary genetics*. New York: Columbia University Press.
- Ota T. 1993. *DISPAN: genetic distance and phylogenetic analysis*. University Park: Institute of Molecular Evolutionary Genetics, Pennsylvania State University.
- Page RDM. 1996. *TREEVIEW: an application to display phylogenetic trees on personal computers*. *Comput Appl Biosci* 12:357–358.
- Petzl-Erler ML, Luz R, Sotomaior VS. 1993. The HLA polymorphism of two distinctive South-American Indian tribes: the Kaingang and the Guaraní. *Tissue Antigens* 41:227–237.
- Raymond M, Rousset F. 1995. *GENEPOP: population genetics software for exact tests and ecumenicism*. *J Hered* 86:248–249.
- Rosser ZH, Zerjal T, Hurles ME, Adojaan M, Alavantic D, Amorim A, Amos W, Armenteros M, Arroyo E, Barbujani G, Beckman G, Beckman L, Bertranpetit J, Bosch E, Bradley DG, Brede G, Cooper G, Corte-Real HB, de Knijff P, Decorte R,

- Dubrova YE, Evgrafov O, Gilissen A, Glisic S, Golge M, Hill EW, Jeziorowska A, Kalaydjieva L, Kayser M, Kivisild T, Kravchenko SA, Krumina A, Kucinkas V, Lavinha J, Livshits LA, Malaspina P, Maria S, McElreavey K, Meitinger TA, Mikelsaar AV, Mitchell RJ, Nafa K, Nicholson J, Norby S, Pandya A, Parik J, Patsalis PC, Pereira L, Peterlin B, Pielberg G, Prata MJ, Previdere C, Roewer L, Rootsi S, Rubinsztein DC, Saillard J, Santos FR, Stefanescu G, Sykes BC, Tolun A, Villemers R, Tyler-Smith C, Jobling MA. 2000. Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. *Am J Hum Genet* 67:1526–1543.
- Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 6:539–545.
- Salzano FM, Callegari-Jacques SM. 1988. South American Indians: a case study in evolution. Oxford: Clarendon Press.
- Salzano FM, Pages F, Neel JV, Gershowitz H, Tanis RJ, Moreno R, Franco MH. 1978. Unusual blood genetic characteristics among the Ayoreo Indians of Bolivia and Paraguay. *Hum Biol* 50:121–136.
- Salzano FM, Black FL, Callegari-Jacques SM, Santos SE, Weimer TA, Mestriner MA, Kubo RR, Pandey JP, Hutz MH. 1991. Blood genetic systems in four Amazonian tribes. *Am J Phys Anthropol* 85:51–60.
- Salzano FM. 2002. Molecular variability in Amerindians: widespread but uneven information. *Ann Braz Acad Sci* 74:223–263.
- Salzano FM, Callegari-Jacques SM, Weimer TA, Franco MHL, Hutz MH, Petzl-Erler ML. 1997. Electrophoretic protein polymorphism in Kaingang and Guarani Indians of southern Brazil. *Am J Hum Biol* 9:505–512.
- Salzano FM, Weimer TA, Franco MHL, Callegari-Jacques SM, Mestriner MA, Hutz MH, Santos RV, Coimbra CEA Jr. 1998. Protein genetic systems among the Tupi-Mondé Indians of the Brazilian Amazonia. *Am J Hum Biol* 10:711–722.
- Sanchez-Mazas A. 2001. African diversity from the HLA point of view: influence of genetic drift, geography, linguistics, and natural selection. *Hum Immunol* 62:937–948.
- Schneider S, Roessli D, Excoffier L. 2001. Arlequin (ver. 2001)—a software for population genetics data analysis. Geneva: University of Geneva.
- Smouse PE, Long JC. 1962. Matrix correlation analysis in anthropology and genetics. *Yrbk Phys Anthropol* 35:187–213.
- Takezaki N. 2001. <http://www.bio.psu.edu/People/Faculty/Nei/Lab/Programs.html>.
- Tsuneto LT, Probst CM, Salzano FM, Hutz MH, Petzl-Erler ML. 2003. HLA class I and class II alleles in five Amerindian populations: Guarani-Nandeva, Guarani Kaiowá, Guarani-M'Bya, Kaingang and Ache. *Tissue Antigens* 62:512–526.
- Ward RH, Redd A, Valencia D, Frazier B, Pääbo S. 1993. Genetic and linguistic differentiation in the Americas. *Proc Natl Acad Sci USA* 90:10663–10667.
- 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.
- Zar J. 1999. Biostatistical analysis. 4th ed. Upper Saddle River, NJ: Prentice-Hall.
- Zerjal T, Beckman L, Beckman G, Mikelsaar AV, Krumina A, Kucinkas V, Hurles ME, Tyler-Smith C. 2001. Geographical, linguistic, and cultural influences on genetic diversity: Y-chromosomal distribution in Northern European populations. *Mol Biol Evol* 18:1077–1087.