

Polymorphisms of *CYP1A1*, *CYP2E1*, *GSTM1*, *GSTT1*, and *TP53* Genes in Amerindians

Pedro A. Gaspar,¹ Mara H. Hutz,¹ Francisco M. Salzano,¹ Kim Hill,² A. Magdalena Hurtado,² M. Luiza Petzl-Erler,³ Luiza T. Tsuneto,³ and Tania A. Weimer^{1,4*}

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

²Department of Anthropology, University of New Mexico, Albuquerque, New Mexico 87131

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

⁴Departamento de Biologia, Universidade Luterana do Brasil, Canoas, RS, Brazil

KEY WORDS gene diversity; South American Indians; molecular markers

ABSTRACT Polymorphisms at the *TP53*, cytochrome P-450 (*CYP*), and glutathione S-transferase (*GST*) genes are related to cancer susceptibility and present high diversity in allele frequencies among ethnic groups. This study concerns the *CYP2E1*, *GSTM1*, and *GSTT1* polymorphisms in seven Amerindian populations (Xavante, Guarani, Aché, Wai Wai, Zoró, Surui, and Gavião). Polymorphic sites at *CYP1A1* and *TP53* were also studied in the Aché and Guarani tribes and compared with previous results about these systems already obtained in the other populations. The *CYP2E1**5B haplotype showed, respectively, the highest and the lowest frequencies already observed in human groups. High frequencies of *CYP1A1**2A and *CYP1A1**2C alleles and mostly low values of *GSTM1**0/*0 and *GSTT1**0/*0 genotypes were observed. These data may be interpreted as being due to genetic drift or selection for these high-frequency *CYP1A1*

alleles and against *GST* null genotypes during America's colonization. Intrapopulation diversity varied from 0.19 (Guarani) to 0.38 (Surui), and 90% of the total diversity was due to the variability within populations. The relationships between these Amerindians and with other ethnic groups were evaluated based on D_A distances and the neighbor-joining method. Low correlation was observed between genetic relationships and geographic distances or linguistic groups. In the *TP53* comparison with other ethnic groups, Amerindians clustered together and then joined Chinese populations. The cluster analysis seems to indicate that the Aché tribe might descend from a Gê group that could have first colonized that Paraguayan region, but had also assimilated some amount of the Guarani gene pool, maybe through intertribal admixture. *Am J Phys Anthropol* 119:249–256, 2002.

© 2002 Wiley-Liss, Inc.

The cancer process is usually a multistep phenomenon, during which consecutive somatic cell mutations occur. Genes involved in cell-cycle control, genetic repair systems, or codifying enzymes for the biotransformation of environmental carcinogens have important roles in this process (Indulski and Lutz, 2000).

TP53 is a tumor-suppressor gene with a critical role in cell-cycle control, and is frequently mutated in many human cancers (Sansom and Clarke, 2000). At least three *TP53* polymorphisms have been reported as involved in cancer, showing also a high interpopulation heterogeneity in allele and haplotype frequencies: a 16-bp duplication in intron 3, an amino-acid change in exon 4 (72 Arg→Pro), and an *MspI* restrictive fragment length polymorphism (RFLP) in intron 6 (Själänder et al., 1996; Khaliq et al., 2000; Gaspar et al., 2001).

Cytochrome P-450 (*CYP*) comprises a superfamily of enzymes that act on phase I of xenobiotic metabolic transformation. During these reactions, toxic metabolites are generated which might be processed by phase II enzymes (Indulski and Lutz, 2000). The *CYP1A1* gene encodes for the CYP1A1 enzyme that

catalyzes the bioactivation of polycyclic aromatic hydrocarbons (Indulski and Lutz, 2000). Two *CYP1A1* gene polymorphisms have been extensively studied in relation to cancer susceptibility: a 462 Ile→Val substitution at exon 7 (*CYP1A1**2C allele), and an associated 6235 T→C mutation at the 3' noncoding region (*CYP1A1**2A allele; Hayashi et al., 1991b). The frequencies of these mutations also exhibit significant interethnic differences (Aynacioglu et al., 1998; Kvitko et al., 2000). *CYP2E1* metabolizes sev-

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.

*Correspondence to: Tania de Azevedo Weimer, Departamento de Genética, UFRGS, Caixa Postal 15053, 91501-970 Porto Alegre, RS, Brazil. E-mail: weimer@ufrgs.br

Received 8 October 2001; accepted 16 April 2002.

DOI 10.1002/ajpa.10128
Published online in Wiley InterScience (www.interscience.wiley.com).

TABLE 1. Amerindian groups investigated in the present study: their location and languages

Population	Country	Localities	Geographical coordinates	Language ¹	Linguistic group ¹	No. of individuals investigated
Xavante	Brazil	Etéñitépa	51°40'W, 13°20'S	Chavante	Gê-Kaingang	33
Guarani	Brazil	Amambai	55°12'W, 23° 6'S	Guarani	Tupi	51
		Limão Verde	55°6' W, 23°12' S			
		Porto Lindo	54°30'W, 23°48'S			
Aché	Paraguay	Arroyo Bandera	55°50'W, 23°30'S	Guayaki	Tupi	67
		Chupa-pou	56°30'W, 24°10'S			
Wai Wai	Brazil	Mapuera village	57°55'W, 0°40'S	Parukoto-Charumã	Carib	26
Zoró	Brazil	Aripuanã Park	60°20'W, 10°20'S	Mondé	Tupi	28
Surui	Brazil	Sete de Setembro	61°10'W, 10°50'S	Mondé	Tupi	21
Gavião	Brazil	Igarapé Lourdes	61°8'W, 10°10'S	Mondé	Tupi	31

¹ According to Rodrigues (1986) and Greenberg (1987).

eral occupational and environmental carcinogens (Indulski and Lutz, 2000). Two RFLPs in linkage disequilibrium at the regulatory region of the *CYP2E1* gene have been described (Hayashi et al., 1991a). Due to linkage disequilibrium, two main arrangements are usually found: *CYP2E1*5B* and *CYP2E1*1A*. Although rare in many populations (Hamada et al., 1995; Griese et al., 2001), the *CYP2E1*5B* arrangement is common in Asians (19–30%; Morita et al., 1997; Tan et al., 2000) and Amerindians (25%; Muñoz et al., 1998).

Glutathione S-transferases (*GST*) are a group of phase II enzymes that detoxify endogenous and exogenous electrophiles, determined by a gene family. Genetic polymorphisms at these loci seem to be related to a higher risk of cancer development (Indulski and Lutz, 2000). Two deletions of the *GSTM1* or *GSTT1* loci which result in no enzymatic activity have been described: both *GSTM1*0/*0* and *GSTT1*0/*0* genotype frequencies present interethnic variability (Rebbeck, 1997).

This study furnishes data on *CYP2E1*, *GSTM1*, and *GSTT1* polymorphisms in seven South Amerindian populations. Polymorphic sites at *CYP1A1* and *TP53* were also studied in the Aché and Guarani tribes. Previous results about these systems from other populations were compiled to allow genetic relationship analyses. A comparison of these Amerindian populations with other ethnic groups was also made, using *TP53* haplotypes. The specific questions posed were as follows: 1) Would the population relationships and genetic variability obtained with these markers, which may be influenced by selection, present the same pattern as those found with polymorphisms in which this process may not have acted as strongly? 2) Could we confirm some unusual frequencies that had been previously obtained in Amerindians for the *CYP1A1*, *GSTM1*, and *GSTT1* systems? 3) Since the origin of the Aché, a recently contacted tribe of Paraguay, is still obscure, would these systems provide a clue about it?

SUBJECTS AND METHODS

Samples of 257 individuals were obtained from seven South American Indian tribes living in Brazil and Paraguay (Table 1 and Fig. 1). More details

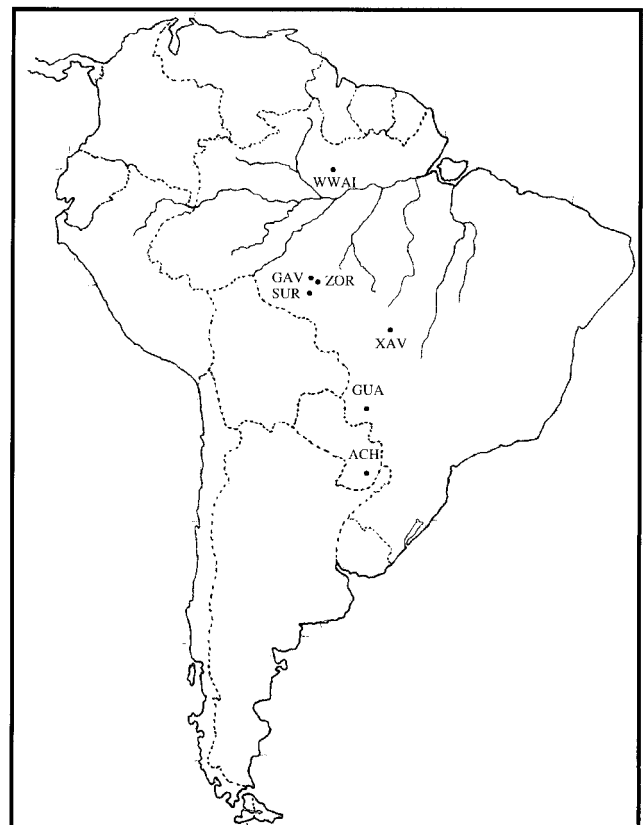


Fig. 1. Map showing locations of populations sampled. ACH, Aché; GAV, Gavião; GUA, Guarani; SUR, Surui; WAI, Wai Wai; XAV, Xavante; ZOR, Zoró; dashed lines, national borders.

about these populations can be found in Hill and Hurtado (1996) and Hutz et al. (1999).

Genomic DNA was isolated from whole blood by the salting-out method of Miller et al. (1988) or by the procedure described by Lahiri and Nurnberger (1991). This latter methodology was employed for the Aché and Wai Wai samples.

Table 2 presents the characterization of the *CYP1A1*, *CYP2E1*, *GSTM1*, *GSTT1*, and *TP53* polymorphisms investigated here. *CYP1A1* and *TP53* were PCR-RFLP analyzed, *CYP1A1* according to Hayashi et al. (1991b) and Cascorbi et al. (1996),

TABLE 2. Characterization of CYP1A1, CYP2E1, GSTM1, GSTT1, and TP53 polymorphisms

Loci	Gene location	Allele or genotype	Phenotypes ¹	Mutation
<i>CYP1A1</i> ²	3' flanking	<i>CYP1A1</i> *2A	<i>MspI</i> +	6325T→C
	Exon 7	<i>CYP1A1</i> *2C	<i>BsrDI</i> -	462A→G (Ile→Val)
<i>CYP2E1</i> ²	5' flanking	<i>CYP2E1</i> *1A	<i>PstI</i> -; <i>RsaI</i> +	-1293 G; -1053 C
		<i>CYP2E1</i> *5B	<i>PstI</i> +	-1293 G→C; -1053 C→T
<i>GSTM1</i>	Whole gene	<i>GSTM1</i> *0/*0	No amplification	Gene deletion
<i>GSTT1</i>	Whole gene	<i>GSTT1</i> *0/*0	No amplification	Gene deletion
<i>TP53</i>	Intron 3	<i>TP53</i> 16bp*A2	Duplication	16bp duplication
	Exon 4 (codon 72)	<i>TP53</i> BstUI*A1	<i>BstUI</i> -	72G→C (Arg→Pro)
	Intron 6	<i>TP53</i> MspI*A1	<i>MspI</i> -	A→G

¹ Plus & minus signs indicate presence or absence of indicated restriction site.

² Allele nomenclature as recommended in <http://www.imm.ki.se/CYPalleles>.

TABLE 3. Genotype frequencies (in %) for GST loci, allele frequencies (in %) for CYP and TP53, and gene diversity values (×100)

Characteristic	Populations						
	Xavante	Guarani	Aché	Wai Wai	Zoró	Surui	Gavião
Genotype or allele							
<i>GSTM1</i> *0/*0	18.2	3.9	35.8	26.9	14.3	43.0	12.9
<i>GSTT1</i> *0/*0	30.3	11.8	17.9	0.0	14.3	0.0	6.5
<i>CYP1A1</i> *2A ¹	95.0	96.1	100.0	81.0	87.0	96.0	72.0
<i>CYP1A1</i> *2C ¹	97.0	90.2	100.0	81.0	76.0	54.0	59.0
<i>CYP2E1</i> *5B	3.0	18.6	42.5	1.9	32.1	33.3	29.0
<i>TP53</i> 16bp*A1 ¹	100.0	100.0	100.0	100.0	100.0	100.0	98.1
<i>TP53</i> BstUI*A1 ¹	28.0	8.8	36.6	11.9	6.8	25.0	19.2
<i>TP53</i> MspI*A1 ¹	0.0	0.0	0.0	7.1	0.0	0.0	1.9
Sample size	33.0	51.0	67.0	26.0	28.0	21.0	31.0
Gene diversity²							
Intrapopulation (H_S)	24.4	18.8	34.7	24.4	29.8	38.2	32.4
Interpopulation (G_{ST})							
				$G_{ST} \pm SE: 10.3 \pm 1.2$			

¹ Data from Kvitko et al. (2000) and Gaspar et al. (2001), except for the Aché and Guarani.

² According to Nei (1987).

and *TP53* following Sjalander et al. (1995). For *GSTM1*, *GSTT1*, and *CYP2E1* typing a multiplex PCR protocol was developed which consisted briefly of the following procedures: an initial denaturation at 94°C for 5 min, 6 touchdown cycles including 1 min at 94°C, 2 min at 59–54°C with a decrease of 1°C in each cycle, and 1 min at 72°C, followed by 30 cycles at 94°C for 1 min, 1 min at 55°C, and 1 min at 72°C, plus a final extension of 5 min at 72°C. The reaction mixture consisted of 100 ng of genomic DNA, 15 pmol of *GSTM1* and *GSTT1* primers, 7.5 pmol of *CYP2E1* primers, 10 mM Tris HCl, 4 mM MgCl₂, 50 mM KCl, 150 mM dNTPs, and 1.0 U of Taq DNA polymerase. An aliquot of the amplification product was electrophoresed in horizontal agarose gel (4%) to verify the presence/absence of *GSTM1* and *GSTT1* fragments, the *CYP2E1* product being used as an internal control for this reaction. A second aliquot of that amplified product was then tested with the *PstI* and *RsaI* enzymes to establish the *CYP2E1* haplotypes. Primer sequences were those reported by Kato et al. (1992), Bell et al. (1993), and Pemble et al. (1994).

Allele frequencies were estimated by gene counting. Haplotypes were derived using the Multiple Locus Haplotype Analysis program (Long, 1999), which uses the E-M algorithm (Long et al., 1995; Peterson et al., 1999). Hardy-Weinberg equilibrium fit was evaluated by exact tests using the Markov chain (Guo and Thompson, 1992) through Arlequin

software version 2.0 (Schneider et al., 2000). Phenotype differences among populations were tested by means of the χ^2 test of Roff and Bentzen (1989), using the PEPI computer program (Gahlinger and Abramson, 1995). Linkage disequilibrium was calculated on basis of estimated haplotype frequencies and using the Arlequin software. The D' value (D/D_{max}) was obtained as suggested by Lewontin (1988).

Genetic affinities among populations were evaluated through D_A distances (Nei, 1987), and the neighbor-joining clustering method (Saitou and Nei, 1987), using the NJBAFD computer program (Takezaki, 1999). This latter software was also employed to estimate gene diversity values (Nei, 1987). The reliability of the trees was tested by 2,000 bootstrap replications (Hedges, 1992). The dendrogram comparing Amerindians with the other ethnic groups was based on *TP53* haplotype frequencies only, since data for the other markers were not available for the other populations.

RESULTS

Table 3 presents *GSTM1**0/*0 and *GSTT1**0/*0 genotype frequencies, allele frequencies for the other genetic systems, and the gene diversity values.

Genotype distribution was consistent with Hardy-Weinberg expectations for all loci and populations. For the majority of markers, genotype and allele distributions were highly heterogeneous among pop-

TABLE 4. Estimated CYP1A1 haplotype frequencies (%) and linkage disequilibrium (D') values

Populations ¹	Haplotypes ²				D'	P	2n
	1A	2A	2C	2B			
Xavante	3.0	3.0		94.0	1.00	*	42
Guarani	2.9	6.9	1.1	89.1	0.72	*	102
Aché				100.0			134
Wai Wai	16.0	16.0	5.0	63.0	0.60	*	52
Zoró	8.0	16.0	5.0	71.0	0.50	*	60
Surui	4.0	42.0		54.0	1.00	NS	48
Gavião	27.0	14.0		59.0	1.00	*	60

¹ With the exception of the Aché and Guarani, haplotype frequencies were previously published by Kvitko et al. (2000).

² Haplotypes can be characterized as follows: 1A, $MspI$ -/ $BsrDI$ +; 2A, $MspI$ +/ $BsrDI$ +; 2C, $MspI$ -/ $BsrDI$ -; 2B, $MspI$ +/ $BsrDI$ -. NS, $p > 0.05$.

* $P < 0.001$.

ulations, regardless of whether they were of the same geographic region or linguistic group. The only exception was *TP53*, in which *16bp**A1 and *MspI**A2 alleles were fixed in almost all samples. The intra-population diversity varied from 0.19 (Guarani) to 0.38 (Surui); most of the total diversity (90%) was due to heterogeneity within populations. Gene diversity values per locus/per tribe always showed positive figures, and the interpopulation differences (G_{ST}) per locus were not significantly different (range, 7 (*TP53*) to 13 (*CYP1A1*)).

As for the *CYP2E1* gene, as previously described (Hayashi et al., 1991a), a complete linkage disequilibrium between the *PstI* and *RsaI* sites was observed and resulted in only two haplotypes: *CYP2E1**1A (*PstI*-, *RsaI*+) and *CYP2E1**5B (*PstI*+, *RsaI*-). The prevalence of *CYP2E1**5B (and of its complementary arrangement) also showed wide variation among populations: from 2% in the Wai Wai, to 42% in the Aché. But these extreme values are restricted to three populations only (the two above-indicated and the Xavante (3%)), while the other values (19–33%) are more in accordance with previous Asian or Asian-derived groups.

Table 4 presents the estimated haplotype frequencies for the *CYP1A1* polymorphisms and the linkage disequilibrium (D') values. The Aché population was monomorphic for *CYP1A1**2B, the most frequent haplotype in the six other populations. Highly significant linkage disequilibrium was observed in five of these groups.

The estimated *TP53* haplotypes are given in Table 5. Only two haplotypes were observed in the Aché and Guarani, as was found for 3 of the 5 other Amerindian tribes previously considered by Gaspar et al. (2001).

The relationships among the seven populations obtained with the D_A distances and the neighbor-joining method, using the five loci studied, are shown in Figure 2. The extremes of this unrooted tree are occupied on the one hand by the Wai Wai (a Carib group from northern Brazil) and Surui (a Tupi-Mondé tribe from southwestern Amazonia), while on the other, the Xavante (a Gê-speaking population from Central Brazil) and Aché (a Tupi-Guarani-speaking group from the Paraguayan forest)

TABLE 5. Estimated TP53 haplotype frequencies (%) in seven South Amerindian populations

Populations	Haplotypes ¹				2n
	1-1-2	1-2-2	1-2-1	2-1-1	
Xavante	28.0	72.0			50
Guarani	7.8	92.2			102
Aché	38.1	61.9			134
Wai Wai	11.9	81.0	7.1		42
Zoró	6.8	93.2			44
Surui	25.0	75.0			40
Gavião	17.3	80.8		1.9	52

¹ Data from Gaspar et al. (2001), except for the Aché and Guarani.

² Haplotypes can be characterized as follows: 1-1-2; duplication (dupl.)-/*BstUI*-/*MspI*+; 1-2-2; dupl.-/*BstUI*+/*MspI*+; 1-2-1; dupl.-/*BstUI*+/*MspI*-; 2-1-1; dupl.+/*BstUI*-/*MspI*-.

occur. Therefore, although 3 of the 4 Tupi tribes group together in the center of the tree, no clear correlation could be observed between the genetic relationships and geographic distances or linguistic group.

Figure 3 shows the *TP53* relationships obtained considering the Amerindians and other ethnic groups. It is well-known that dendrograms based on one marker only should be viewed with caution. Absence of comparative information for the other loci, however, forced us to restrict the analysis to *TP53*. Despite this limitation, the results agreed with the expectations: three clusters can be observed, one involving Asian or Asian-derived (Amerindian) groups, another composed of the two European or European-derived populations, and a third including Africans, African Brazilians, and Pakistani.

DISCUSSION

Polymorphisms at the *TP53*, *CYP*, and *GST* loci have been described as related to cancer susceptibility. For example, the *GSTM1**0/*0 and *GSTT1**0/*0 genotypes, *CYP1A1**2A, *CYP1A1**2C, and *TP53* *BstUI**A1 alleles, and the *CYP2E1**5B haplotype have been associated with different tumors in several human populations (Weston et al., 1997; Indulski and Lutz, 2000). These polymorphisms also present high diversity among ethnic

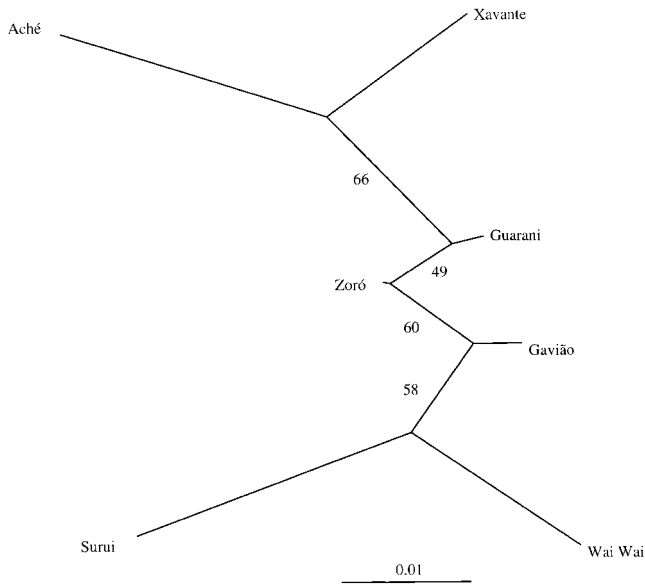


Fig. 2. Dendrogram obtained using neighbor-joining method and D_A distance, considering *CYP11A1*, *CYP2E1*, and *TP53* haplotypes, as well as *GSTM1* and *GSTT1* genotype frequencies. Numbers indicate bootstrap values.

groups. Trying to explain the differences of allele frequencies observed in genes which influence the phase I and phase II detoxifying process, Nebert (1997) proposed an effect of natural selection in response to population differences in diet, while Beckman et al. (1994), and Sjalander et al. (1996) had already suggested that population differences in *TP53* allele frequencies might result from selective effects on ecological adaptation to climatic conditions.

Any hypothesis that uses diet as a basis is confronted by the quantification problem. It is very difficult to accurately establish the amount of energy received in any given meal, and the type and quantity of food ingested may vary widely on a daily and/or seasonal basis. The amount of dietary information available for the groups studied here is generally indirect and of a fragmentary nature. In terms of outside influences, the Guarani have already adopted the neo-Brazilian way of living, and their diet is not much different from those of rural non-Indians, but no scientific investigation of food consumption has been performed among them. The Xavante have more than half a century of contact with non-Indians, and is the group for which the most information is available. The relative importance of game, fish, gathered plant products, and purchased foods varied markedly during this period; details are provided in Coimbra et al. (2002). The degree of isolation of the Wai Wai and Tupi-Mondé groups (Gavião, Surui, and Zoró) is still high, and they maintain much of their traditional way of living. No quantitative nutritional data is available for the Wai Wai (reviewed in Gallois and Ricardo, 1983; Callegari-Jacques et al. 1996), while for the Tupi-Mondé only indirect information exists (e.g., see

Santos and Coimbra, 1994). Finally, the Aché only started more permanent contact with non-Indians in 1970; until recently, therefore, their diet would resemble those of forest peoples. But again, no specific food consumption investigation had been performed among them (Hill and Hurtado, 1996).

Even considering the scarcity of data, it is clear that the relationships obtained in Figure 2 for the seven tribes cannot be explained on the basis of diet differences, since the Guarani, who should have the most distinctive diet from the others, occupy a central position in the tree.

This investigation confirmed that most Amerindians are monomorphic for both the *16bp**A1 allele in intron 3, and for the presence of the *MspI**A2 allele in intron 6 of the *TP53* gene. They are, therefore, not informative in relation to the second hypothesis put forward to explain variability in this system. Beckman et al. (1994) and Sjalander et al. (1996) found a correlation between the variability at codon 72 of the *TP53* gene and latitude, suggesting a possible role for natural selection involving climatic adaptation. The Guarani and Aché live in the southern part of South America and speak a Tupi-Guarani language, but show *TP53 BstUI**A1 gene frequencies of respectively 9–37%. On the other hand, the Zoró, Gavião, and Surui, who live in the north of South America in about the same area and speak languages classified as a Tupi subgroup, presented frequencies of the same allele varying from 7–25%. Therefore, our data do not show any correlation between the frequencies of this allele and latitude.

Despite extensive changes that most Amerindians have suffered due to contact with other groups, most of the populations studied here, as mentioned, retain many of their previous ecological conditions, such as nomadism, hunting, gathering, and horticulture (Hill and Hurtado, 1996; Santos et al., 1996), and are not continuously exposed to environmental chemicals. All tribes presently investigated presented high levels of diversity in most of the systems (averages range from 19–38). Therefore, although they live in small groups, evolutionary factors such as selection or drift seem not to have significantly decreased their genetic diversity. These data confirm the degree of variability previously verified in the same populations for other systems (eight nuclear DNA markers, 41–57; mtDNA, 5–17; 12 *Alu* loci, 15–29; 23 protein systems, 13–18; Hutz et al., 1999; Battilana et al., 2002).

As previously indicated, the *CYP2E1**5B distribution presented remarkable (2–42%) differences. These are the most extreme values reported so far for any other ethnic group, except the 0.02% detected among Australian Aborigines (Griese et al., 2001). These extreme values are, however, restricted to the Wai Wai, Xavante, and Aché. When an inspection is made of the prevalences observed in the other systems studied here, the peculiarities of the Aché stand out clearly. The figures displayed in Tables 3–5 are not completely independent, but in

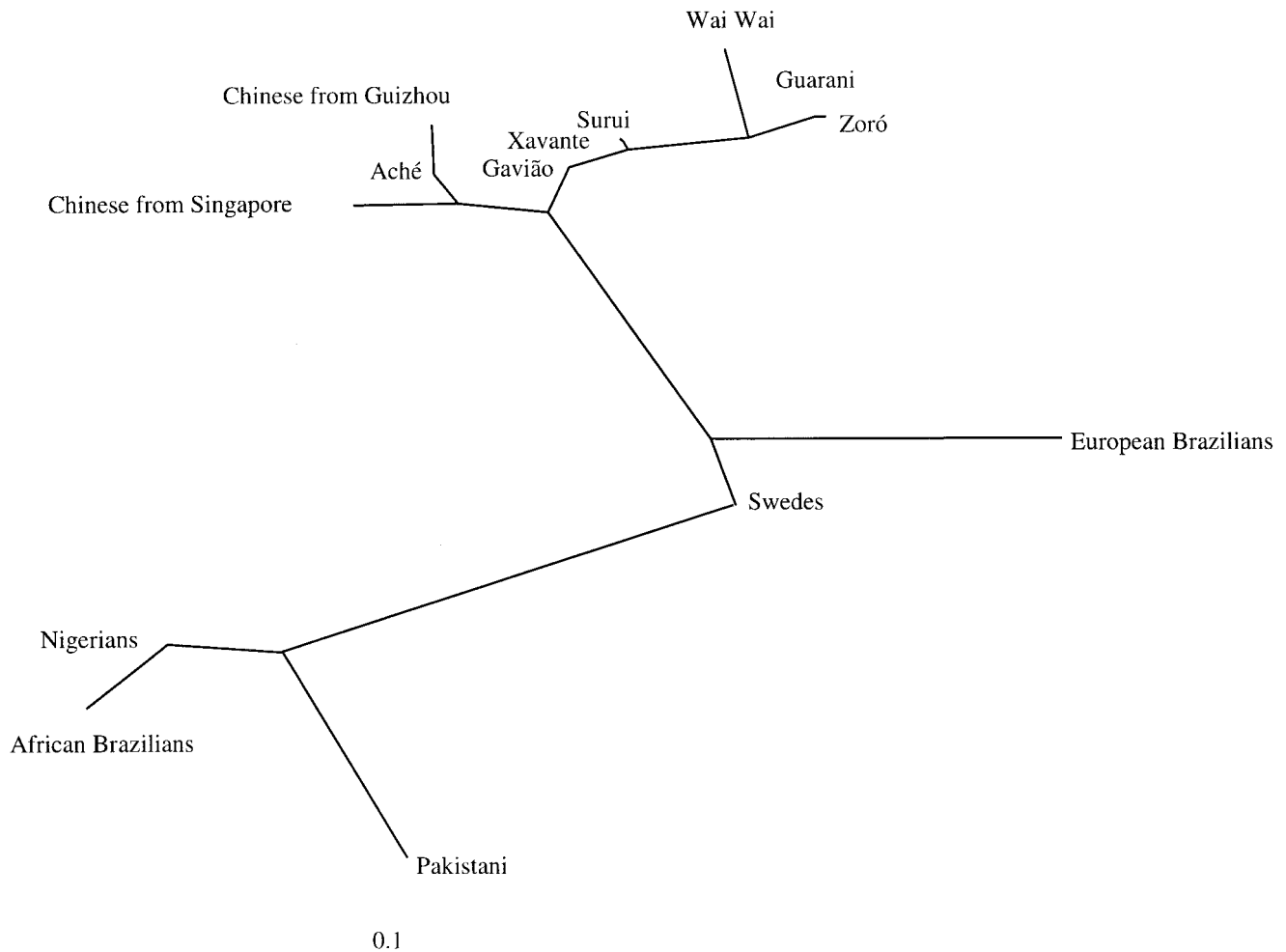


Fig. 3. Dendrogram obtained using neighbor-joining method and D_A distances, based on *TP53* haplotype frequencies. Data on Nigerians, Chinese, and Swedes were obtained from Sjalander et al. (1996); on Pakistani, from Khaliq et al. (2000); and on African and European Brazilians, from Gaspar et al. (2001).

most cases the most extreme values were provided by the Aché. The Xavante also presented some extreme but less numerous differences, while the Wai Wai conformed more to the average. Similar results were obtained considering blood group plus protein and other nDNA and mtDNA systems (Callegari-Jacques et al., 1996; Battilana et al., 2002; Coimbra et al., 2002). Detailed discussion about the interrelationships obtained for 5 of the 7 groups studied here (Gavião, Surui, Zoró, Wai Wai, and Xavante) was presented by Hutz et al. (1999). The peripheral position observed here for the Surui was also observed when 23 protein genetic systems were considered, and may be due to the fact that this group was, of the seven, that which suffered most the first impact of contact with the surrounding populations (Coimbra, 1989).

Also extremely interesting are the very high values of the *CYP1A1**2A and *CYP1A1**2C alleles and the mostly low frequencies of the *GSTM1**0/*0 and *GSTT1**0/*0 genotypes. The highest frequencies so far described in non-Amerindians for the first two

markers are 33% (Japan; Aynacioglu et al., 1998) and 35% (Siberia; Duzhak et al., 2000), respectively, which are much lower than the 81–100% and 54–100% observed here. Our values, however, are similar to those observed in the Mapuche of Chile (respectively, 83% and 77%; Muñoz et al., 1998). The *GSTM1**0/*0 frequencies vary from 22% in Africa to 100% in Oceania; most of these populations present values above 30% (reviewed in Rebbeck, 1997). The values observed here (4–43%) are in some cases lower, in agreement with a 20% frequency obtained previously among the Amazonian Parakanã Indians (Arruda et al., 1998). The *GSTT1**0/*0 frequency, on the other hand, is about 16% in Caucasians and above 38% in the other ethnic groups (Rebbeck, 1997), while the interval observed here was 0–30%. The Parakanã studied by Arruda et al. (1998) showed a prevalence of 11% of this genotype.

The enzymes codified by *CYP1A1**2A and *CYP1A1**2C alleles have higher catalytic activities than the products of the wild alleles, producing a larger amount of toxic metabolites which are mainly

detoxified by both the GSTM1 and GSTT1 proteins (Indulski and Lutz, 2000). As the *GSTM1* and *GSTT1* null genotypes have no enzyme activity, toxic products induced by the action of CYP enzymes could accumulate in these individuals.

It is possible, therefore, that *CYP1A1*2A* and *CYP1A1*2C* allele frequencies had increased during America's colonization, either by genetic drift (Cavalli-Sforza et al., 1994) or by selection in response to new environmental challenges (Nebert, 1997). These high *CYP* prevalences could have acted as selective factors reducing *GSTM1*0/*0* and *GSTT1*0/*0* frequencies, since the ratio between *CYP* and *GST* activities is critical to avoid the accumulation of toxic reactive intermediates (Rebeck, 1997).

The origin of Aché population has been controversial: some authors consider them as a differentiated Guaraní group, while others claim that they descended from a Gê group that preceded the Guaraní colonization of Paraguay (Hill and Hurtado, 1996). Although the possible action of selection, as outlined above, should be also taken into consideration, the cluster analysis given here seems to support the second hypothesis. The Aché showed a closer relationship with the Xavante, a Gê-speaking population, than with the four Tupi-Guaraní groups included in the study. But, as previously stressed, they were clearly differentiated from all the other populations studied, showing their genetic distinctiveness. Battilana et al. (2002), who examined the variability of 20 blood group plus protein systems and 12 *Alu* insertions in the same Aché, Xavante, and Guaraní samples studied here, also obtained indications of Aché distinctiveness, although in their data, the closest group to cluster with the Aché was the Guaraní. It is therefore possible that the Aché might have descended from a Gê group that had first colonized Paraguay, but that they had also assimilated some amount of the Guaraní gene pool, maybe through intertribal admixture. New data are needed to clarify this point.

CONCLUSIONS

In conclusion, this study found that: 1) the patterns of population relationships obtained show distinctive features (but not the levels of genetic variability) which may in some undefined way be due to selective processes; 2) the unusual Amerindian *CYP1A1*, *GSTM1*, and *GSTT1* frequencies have been confirmed; and 3) due to the Aché genetic distinctiveness, their origin is an unresolved issue.

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 Ricardo V. Santos and Carlos E.A. Coimbra Jr. for the collection of four of the Amerindian samples.

LITERATURE CITED

- Arruda VR, Grignolli CE, Gonçalves MS, Soares MC, Menezes R, Saad STO, Costa FF. 1998. Prevalence of homozygosity for the deleted alleles of glutathione S-transferase mu (*GSTM1*) and theta (*GSTT1*) among distinct ethnic groups from Brazil: relevance to environmental carcinogenesis? *Clin Genet* 54:210–214.
- Aynacioglu AS, Cascorbi I, Mrozikiewicz PM, Roots I. 1998. High frequency of *CYP1A1* mutations in a Turkish population. *Arch Toxicol* 72:215–218.
- Battilana J, Bonatto SL, Freitas LB, Hutz MH, Weimer TA, Callegari-Jacques SM, Batzer MA, Hill K, Hurtado AM, Tsuneto LT, Petzl-Erler LT, Salzano FM. 2002. *Alu* insertion vs. blood group plus protein genetic variability in four Amerindian populations. *Ann Hum Biol* (in press).
- Beckman G, Birgander R, Sjölander A, Saha N, Holmberg PA, Kivelä A, Bekman L. 1994. Is p53 polymorphism maintained by natural selection? *Hum Hered* 44:266–270.
- Bell DA, Taylor JA, Paulson DF, Robertson CN, Mohler JL, Lucier GW. 1993. Genetic risk and carcinogen exposure: a common inherited defect of the carcinogen-metabolism gene glutathione S-transferase M1 (*GSTM1*) that increases susceptibility to bladder cancer. *J Natl Cancer Inst* 85:1159–1164.
- Callegari-Jacques SM, Salzano FM, Weimer TA, Franco MHL, Mestriner MA, Hutz MH, Schüller L. 1996. The Wai Wai Indians of South America: history and genetics. *Ann Hum Biol* 23:189–201.
- Cascorbi I, Brockmüller J, Roots I. 1996. A C4887A polymorphism in exon 7 of human *CYP1A1*: population frequency, mutation linkages, and impact on lung cancer susceptibility. *Cancer Res* 56:4965–4969.
- Cavalli-Sforza LL, Menozzi P, Piazza A. 1994. The history and geography of human genes. Princeton: Princeton University Press.
- Coimbra CEA Jr. 1989. From shifting cultivation to coffee farming: the impact of change on the health and ecology of the Surui Indians of the Brazilian Amazon. Ph.D. dissertation, Indiana University at Bloomington.
- Coimbra CEA Jr, Flowers NM, Salzano FM, Santos RV. 2002. The Xavante in transition: health, ecology and bioanthropology in Central Brazil. *Ann Arbor: Michigan University Press* (in press).
- Duzhak T, Mitrofanov D, Ostashevskii V, Gutkina N, Chasovnikova O, Posukh O, Osipova L, Lyakhovich VV. 2000. Genetic polymorphisms of *CYP2D6*, *CYP1A1*, *GSTM1* and p53 genes in a unique Siberian population of tundra. *Pharmacogenetics* 10: 1–7.
- Gahlinger PM, Abramson JH. 1995. Computer programs for epidemiologic analysis: PEPI. Stone Mountain: USD.
- Gallois D, Ricardo CA. 1993. Povos indígenas no Brasil. 3. Amapá/Norte do Pará. São Paulo: Centro Ecumênico de Documentação e Informação.
- Gaspar PA, Hutz MH, Salzano FM, Weimer TA. 2001. TP53 polymorphisms and haplotypes in South Amerindians and neo-Brazilians. *Ann Hum Biol* 28:184–194.
- Greenberg JH. 1987. Language in the Americas. Stanford: Stanford University Press.
- Griese EU, Ilett KF, Kitteringham NR, Eichelbaum M, Powell H, Spargo RM, LeSouef PN, Musk AW, Minchin RF. 2001. Allele and genotype frequencies of polymorphic cytochromes P450D6, 2C19 and 2E1 in Aborigines from Western Australia. *Pharmacogenetics* 11:69–76.
- Guo S, Thompson E. 1992. Performing the exact test of Hardy-Weinberg proportions for multiple alleles. *Biometrics* 48:361–372.

- Hamada GS, Sugimura H, Suzuki I, Nagura K, Kiyokawa E, Iwase T, Tanaka M, Takahashi T, Watanabe S, Kino I, Tsugane S. 1995. The heme-binding region polymorphism of cytochrome P450IA1 (Cyp1A1), rather than the RsaI polymorphism of IIE1 (CypIIE1), is associated with lung cancer in Rio de Janeiro. *Cancer Epidemiol Biomarkers Prev* 4:63–67.
- Hayashi SI, Watanabe J, Kawajiri K. 1991a. Genetic polymorphisms in the 5'-flanking region change transcriptional regulation of the human cytochrome P450IIE1 gene. *J Biochem (Tokyo)* 110:559–565.
- Hayashi SI, Watanabe J, Nakachi K, Kawajiri K. 1991b. Genetic linkage of lung cancer-associated *MspI* polymorphisms with amino acid replacement in the heme binding region of the human cytochrome P450IA1. *J Biochem (Tokyo)* 110:407–411.
- Hedges SB. 1992. The number of replications needed for accurate estimation of the bootstrap P value in phylogenetic studies. *Mol Biol Evol* 9:366–369.
- 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, 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: Kluwer Academic/Plenum. p 23–32.
- Indulski JA, Lutz W. 2000. Metabolic genotype in relation to individual susceptibility to environmental carcinogenesis. *Int Arch Occup Environ Health* 73:71–85.
- Kato S, Shields PG, Caporaso NE, Hoover RN, Trump BF, Sugimura H, Weston A, Harris CC. 1992. Cytochrome P450IIE1 genetic polymorphisms, racial variation, and lung cancer risk. *Cancer Res* 57:6712–6715.
- Khaliq S, Hameed A, Khaliq T, Aybur Q, Qamar R, Mohyuddin A, Mazhar K, Mehdi Q. 2000. p53 mutations, polymorphisms, and haplotypes in Pakistani ethnic groups and breast cancer patients. *Genet Test* 4:23–29.
- Kvitko K, Nunes JCB, Weimer TA, Salzano FM, Hutz MH. 2000. Cytochrome P450IA1 polymorphisms in South American Indians. *Hum Biol* 72:1039–1043.
- Lahiri DK, Nurnberger JL. 1991. A rapid nonenzymatic method for the preparation of HMW DNA from blood for RFLP studies. *Nucleic Acids Res* 19:5444.
- Lewontin RC. 1988. On measures of gametic disequilibrium. *Genetics* 120:849–852.
- Long JC. 1999. Multiple locus haplotype analysis, version 2.0. Bethesda, MD: Section on Population Genetics and Linkage, Laboratory of Neurogenetics, National Institute on Alcohol Abuse and Alcoholism, National Institutes of Health. Software and documentation distributed by the author.
- 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.
- Miller AS, Dykes DD, Polesky HF. 1988. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* 16:1215.
- Morita S, Yano M, Shiozaki H, Tsujinaka T, Ebisuti C, Morimoto T, Kishibuti M, Fujita J, Ogawa A, Taniguchi M, Inoue M, Tamura S, Yamazaki K, Kikkawa N, Mizunoya S, Monden M. 1997. CYP1A1, CYP2E1 and GSTM1 polymorphisms are not associated with susceptibility to squamous-cell carcinoma of the esophagus. *Int J Cancer* 71:192–195.
- Muñoz S, Vollrath V, Vallejos MP, Miquel JF, Covarrubias C, Raddatz A, Chianale J. 1998. Genetic polymorphisms of CYP2D6, CYP1A1 and CYP2E1 in the South-American population of Chile. *Pharmacogenetics* 8:343–351.
- Nebert WN. 1997. Polymorphisms in drug-metabolizing enzymes: what is their clinical relevance and why do they exist? *Am J Hum Genet* 60:265–271.
- Nei M. 1987. *Molecular evolutionary genetics*. New York: Columbia University Press.
- Pemble S, Schroeder R, Spencer SR, Meyer DJ, Hallier E, Bolt HM, Ketterer B, Taylor JB. 1994. Human glutathione S-transferase theta (GSTT1): cDNA cloning and the characterization of a genetic polymorphism. *Biochem J* 300:271–276.
- 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.
- Rebeck TR. 1997. Molecular epidemiology of the human glutathione S-transferase genotypes GSTM1 and GSTT1 in cancer susceptibility. *Cancer Epidemiol Biomarkers Prev* 6:733–743.
- Rodrigues AD. 1986. *Linguas brasileiras. Para o conhecimento das línguas indígenas*. São Paulo: Loiola.
- Roff DA, Bentzen P. 1989. The statistical analysis of mitochondrial DNA polymorphisms: χ^2 and the problem of small samples. *Mol Biol Evol* 6:539–545.
- Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 6:539–545.
- Sansom OJ, Clarke AR. 2000. P53 null mice: damaging the hypothesis? *Mutat Res* 452:149–162.
- Santos RV, Coimbra CEA Jr. 1994. Contato, mudanças socioeconômicas e a bioantropologia dos Tupi-Mondé da Amazônia brasileira. In: Santos RV, Coimbra CEA Jr, editors. *Saúde e povos indígenas*. Rio de Janeiro: FIOCRUZ. p 189–211.
- Santos RV, Flowers NM, Coimbra CEA Jr, Gugelmin SA. 1996. Human ecology and health in the context of change: the Xavante Indians of Mato Grosso, Brazil. In: Follér ML, Hansson L, editors. *Human ecology and health adaptation to a changing world*. Göteborg: Göteborg University. p 94–117.
- Schneider S, Kueffer JM, Roessli D, Excoffier L. 2000. Arlequin version 2.000: a software for population genetic data analysis. Geneva: Genetics and Biometry Laboratory, University of Geneva.
- Själänder A, Birgander R, Kivelä A, Beckman G. 1995. p53 polymorphisms and haplotypes in different ethnic groups. *Hum Hered* 45:144–149.
- Själänder A, Birgander R, Saha N, Beckman L, Beckman G. 1996. p53 polymorphisms and haplotype show distinct difference between major ethnic groups. *Hum Hered* 46:41–48.
- Takezaki N. 1999. NJBAFD: neighbor-joining tree constructions from allele frequency data. <ftp://ftp.nig.ac.jp/pub/Bio/njbafd/dos/>
- Tan W, Song N, Wang G, Liu Q, Tang H, Kadlubar FF, Lin D. 2000. Impact of genetic polymorphisms in cytochrome P450 2E1 and glutathione S-transferase M1, T1, and P1 on susceptibility to esophageal cancer among high-risk individuals in China. *Cancer Epidemiol Biomarkers Prev* 9:551–556.
- Weston A, Pan C, Ksieski B, Wallenstein S, Berkowitz GS, Tarter PI, Bleiweiss IJ, Brower ST, Senie RT, Wolff MS. 1997. p53 haplotype determination on breast cancer. *Cancer Epidemiol Biomarkers Prev* 6:105–112.