AMERICAN JOURNAL OF HUMAN BIOLOGY 16:264–277 (2004)

Original Research Article

Polymorphic *Alu* Insertions in Six Brazilian African-Derived Populations

NELSON HENDERSON COTRIM,[1] MARIA TERESA B.M. AURICCHIO,[1] JOÃO PEDRO VICENTE,[2] PAULO A. OTTO,[1] AND REGINA CÉLIA MINGRONI-NETTO[1]*
[1]*Centro de Estudos do Genoma Humano, Departamento de Biologia, Instituto de Biociências, Universidade de São Paulo, São Paulo, Brazil*
[2]*Departamento de Pediatria, Hospital das Clínicas, Faculdade de Medicina, Universidade de São Paulo, São Paulo, Brazil*

ABSTRACT At least 25 African-derived populations (*quilombo* remnants) are believed to exist in the Ribeira River Valley, located in the southern part of São Paulo State, Brazil. We studied four *Alu* polymorphic loci (APO, ACE, TPA25, and FXIIIB) in individuals belonging to six quilombo remnants in addition to individuals sampled from the city of São Paulo. The allelic frequencies observed in the quilombo remnants were similar to those previously observed in African-derived populations from Central and North America. Genetic variability indexes (F_{st} and G_{st} values) in our quilombos were higher than the reported values for the majority of other populations analyzed for the same kind of markers, but lower than the variability usually observed in Amerindian groups. The observed high degree of genetic differentiation may be due to genetic drift, especially the founder effect. Our results suggest that these populations behave genetically as semi-isolates. The degree of genetic variability within populations was larger than among them, a finding described in other studies. In the neighbor-joining tree, some of the Brazilian quilombos clustered with the African and African-derived populations (São Pedro and Galvão), others with the Europeans (Pilões, Maria Rosa, and Abobral). Pedro Cubas was placed in an isolated branch. Principal component analysis was also performed and confirmed the trends observed in the neighbor-joining tree. Overall, the quilombos showed a higher degree of gene flow than average when compared to other worldwide populations, but similar to other African-derived populations. Am. J. Hum. Biol. 16:264–277, 2004.

Short interspersed elements (SINEs), composed mainly of sequences originated by retrotransposition, are a class of repetitive DNA found in the genome of mammals. The *Alu* family, exclusive to primates (Zietkiewicz et al., 1998), is the most frequent SINE found in the human genome. With ~1 million copies, *Alu* insertions are found on average once every 4 kb interval and therefore correspond to about 11% of the human genome (International Human Genome Sequencing Consortium, 2001).

Alu insertions are retrotransposable elements that are roughly 281 nucleotides in length. They are composed of two smaller monomeric units, united by a poli-A tract and with a poli-A tail in the 3′ flank (Novick et al., 1996; Cooper, 1999). *Alu* insertions do not code for proteins and are believed to be derived from a retrotransposed copy of the 7SL RNA gene, in which a series of duplications and deletions has occurred (Cooper, 1999). The 7SL RNA is part of the signal recognition particle (SRP), a ribonucleoprotein whose function is to target secreted or membrane-bound proteins to the endoplasmic reticulum.

Besides the presence of internal RNA polymerase III promoters, effective transcription of *Alu* insertions is thought to also be influenced by external promoters (Novick et al., 1996). Thus, it is widely believed that there are only a few *Alu* master genes capable of efficient transcription and therefore capable of retrotransposition (Batzer et al., 1990; Shen et al., 1991; Deininger et al., 1992).

Contract grant sponsor: CEPID/FAPESP; Contract grant number: 98/14254-2; Contract grant sponsor: FAPESP; Contract grant number: 99/11698-0; Contract grant sponsor: PRONEX/CNPq.

*Correspondence to: Regina Célia Mingroni-Netto, Departamento de Biologia, Universidade de São Paulo, São Paulo, Brazil, Rua do Matão, 277 São Paulo, SP, Brazil 05508-900. E-mail: renetto@ib.usp.br

Received 16 July 2003; Revision received 14 November 2003; Accepted 25 November 2003

Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/ajhb.20024

Recent models hypothesize that the effectiveness of the retrotransposition is influenced by the length of the poli-A tail (Roy-Engel et al., 2002). Whenever one of these master genes suffers a mutation, its subsequent copies will present the same mutation, thereby originating a new subfamily of *Alu* elements.

Some of these insertions have occurred so recently that they have not been fixed, and therefore represent new polymorphic loci that can be used in population studies (Batzer et al., 1996; Roy et al., 1999, Comas et al., 2001; Hollies et al., 2001). Because there are so many polymorphic *Alu* insertions, they constitute a set of highly informative markers for the study of human populations. Thus, *Alu* insertions can be used in a broad variety of applications: to study the origin and dispersal of modern humans (Batzer et al., 1996; Sherry et al., 1997; Watkins et al., 2001), to understand the colonization of the Americas (Novick et al., 1998), and to test hypotheses regarding the origins of populations and the genetic relationships among specific populations (Parra et al., 1998; Majumder et al., 1999; Comas et al., 2000; de Pancorbo et al., 2001; Bamshad et al., 2001; Jorde et al., 2001; Nazidse et al., 2001).

Most *Alu* insertion studies that have focused on Brazilian populations have dealt with Amerindian populations (Battilana et al., 2002; Mateus-Pereira et al., 2004; Oliveira, pers. commun.). Some of these studies have compared the results obtained using *Alu* insertions to those obtained with other markers, such as LINEs and classical protein polymorphisms (Mateus-Pereira et al., 2004; Oliveira, pers. commun.). Battilana et al. (2002) analyzed the affinities between four Amerindian populations. One of them, the Ache, of rather controversial origin, presented very distinct genetic characteristics from other Amerindian populations, thus far thought to be related to it. These studies have shown that Amerindians usually present higher indexes of genetic differentiation than other human populations.

Before the abolition of slavery in Brazil (1888), many communities (*quilombos*) were founded by fled or abandoned African slaves, presently referred to as quilombo remnants. Interestingly, they still remain at least partially genetically isolated. It is estimated that there are at least 700 such communities within Brazilian territory. They can be regarded as relics of the original African contribution to the Brazilian population. Some reports have focused on the molecular variability of Brazilian African-derived populations (Bortolini et al., 1997, 1999; Guerreiro et al., 1999; Silva et al., 1999; Arpini-Sampaio et al., 1999; Oliveira et al., 2001; Mingroni-Netto et al., 2002; Ribeiro-dos-Santos et al., 2002), but none of them have studied *Alu* insertions.

We analyzed the allelic frequencies of four *Alu* polymorphic loci (APO, ACE, TPA25, and FXIIIB) in six different African-derived populations (quilombo remnants) from the Ribeira River Valley (Vale do Ribeira) in the southern part of São Paulo State, Brazil (Fig. 1), comparing our results to a sample collected in the city of São Paulo. The aim of our study was to infer genetic relationships between quilombos and other populations (Africans, Amerindians, and Europeans). We also evaluated the degree of genetic isolation and gene flow experienced by these quilombo remnants in comparison to other populations.

SUBJECTS AND METHODS

Populations studied

The geographical location of the populations, total number of inhabitants, and number of individuals analyzed is summarized in Table 1. A map is presented in Figure 1. A sample of 41 unrelated individuals from the city of São Paulo was also analyzed. The sample is composed mainly of individuals of European ancestry and a few individuals from other ethnic groups. This study was approved by the ethics committee of the Instituto de Ciências Biomédicas da Universidade de São Paulo. Informed consent was obtained from all participants in the study.

PCR amplification of polymorphic loci

We analyzed the frequencies of four *Alu* polymorphic insertions (APO, ACE, TPA25, and FXIIIB). The primer sequences for amplification are described in Batzer et al. (1996).

Amplification of DNA samples for the APO, ACE, and TPA25 loci was carried out in 25 μl reactions using 100–200 ng of target DNA, 45 pmol of each oligonucleotide, 200 μM dNTPs, 1.5 mM $MgCl_2$, 20 mM Tris-HCl, pH 8.4, 50 mM KCl, and 1.25 U Taq DNA polymerase. Each sample was subjected to the following amplification conditions: 1 min at 94°C (denaturation), annealing at 50°C (APO), and 58°C (ACE and TPA25) for 2 min, and 2 min at 72°C (extension) for 35 cycles.

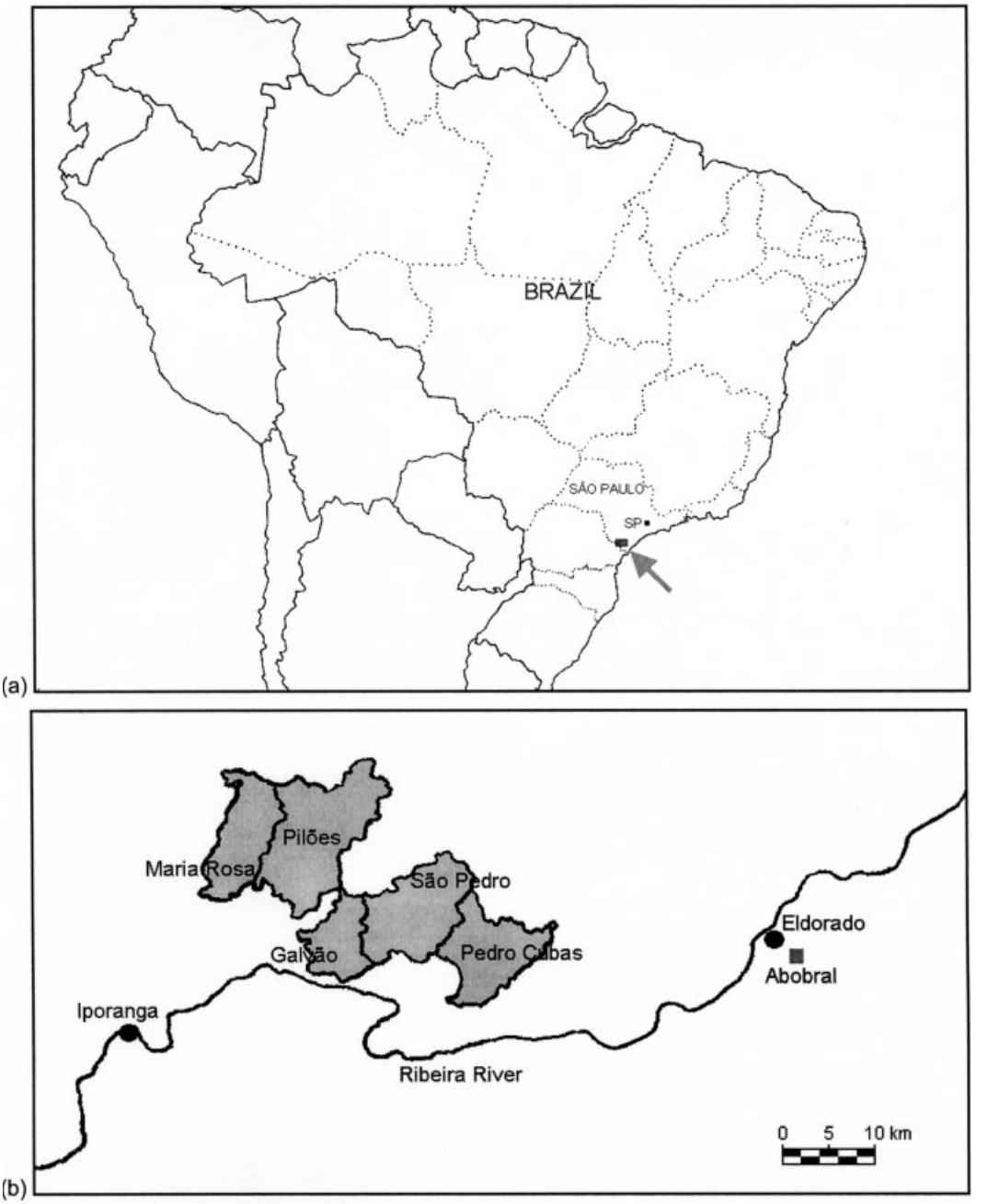

Fig. 1. Location of the populations studied. **a:** The city of São Paulo and, indicated by the arrow, the area of the Ribeira River Valley corresponding to **b**. **b:** The six quilombo remnants and the cities of Eldorado and Iporanga. The definitive area of Abobral has not yet been determined.

Amplification of DNA samples for the FXIIIB locus was carried out in 25 μl reactions using 100–200 ng of target DNA, 45 pmol of each oligonucleotide, 400 μM dNTPs, 2.6 mM $MgCl_2$, 20 mM Tris-HCl, pH 8.4, 50 mM KCl, and 2 U Taq DNA polymerase. Each sample

TABLE 1. Quilombos *location, total inhabitants and numbers of individuals analyzed*

Population	Abobral	Galvão	São Pedro	Pedro Cubas	Pilões	Maria Rosa
Location	24°28′S 48°04′O	24°32′S 48°26′O	24°31′S 48°24′O	24°34′S 48°16′O	24°29′S 48°29′O	24°28′S 48°30′O
Total inhabitants	397	134	132	286	128	56
Individuals analyzed	123 (31%)	50 (38%)	51 (38%)	117 (41%)	39 (31%)	22 (39%)

was subjected to the following amplification conditions: 1 min at 94°C (denaturation), annealing at 55°C for 2 min, and 4 min at 72°C (extension) for 40 cycles, followed by a final extension of 6 min at 72°C.

PCR products were analyzed by electrophoresis through a 2% agarose gel stained with ethidium bromide and the reaction products were directly visualized using UV fluorescence. The PCR process for *Alu* insertions is known to preferentially amplify the smaller allele (i.e., the lack of insertion). Therefore, for every individual identified as homozygous for the smaller allele, a second amplification was performed to confirm the results.

Data analysis

Due to the small size of the populations, our samples necessarily contained related individuals. A contingency table analysis was performed to ascertain whether the quilombo samples presented any significant difference when the related individuals (with a relationship coefficient equal or larger than one-fourth) were taken into account. Since we did not observe significant differences in the genotypic frequencies between the two groups, we decided to use the total samples in the final analysis in order to have larger sample sizes.

The allelic frequencies, heterozygosities, and fixation indexes were calculated according to the equations in Nei (1987), and G_{st} and F_{st} values were calculated according to the equations of Weir and Cockerham (1984) and Nei (1987) by means of computer programs that we prepared. The neighbor-joining tree was built using the DISPAN package (Ota, 1983).

Principal component analysis (PCA) was performed using the allelic frequencies observed for the four *Alu* insertions using SPSS 10.1 (Chicago, IL).

Both the neighbor-joining tree and the PCA analyses included previously reported worldwide samples: Karitiana, Surui, Wayuu, Arhuaco, Chimila, Ingano, Guambiano, Guayabero, Kogui, Paez, Inca, Ngobe, Waunana, Quechua, Toba, Navajo, Moskoke, Zuni, Sioux, Cree-Ojibwa, Maya-Campeche, Maya-Buctotz, Alaska-Aleut, Greeks, Turks, Nigerians, Pygmies from Zaire and Central African Republic (Zaire+RCA Pygmies), and European-Americans (Novick et al., 1998); Cainang, Guarani, Xavante and Ache (Battilana et al., 2002); Maya, Alaska natives, Greenland natives, Chinese, Taiwan Chinese, Javanese, Philippine, Indonese-Mollucas, Indonese-Nusa Tengarras, Malayan, Australians (mixed), Papua New Guinea coastal natives (PNG coast), Papua New Guinea interior natives (PNG interior), Nguni, Sotho, !Kung, Pygmies from Central African Republic (RCA Pygmies), and Pygmies from Zaire (Stoneking et al., 1997); African-Americans, British Afro-Caribbeans, Greenland natives, Bretons, French, French Acadians, and Swiss (Batzer et al., 1996); some Taiwanese native samples: Ami, Atayal, Bunun, and Paiwan (Melton et al., 1998); Comunidad Autonoma Vasca (Bascs-CAV), Goiherri, Arratians, Arabans, and Bascs from Bilbao (de Pancorbo et al., 2001).

To assess the relative amount of gene flow experienced by each population, the expected Hardy-Weinberg proportions of heterozygotes of each population were plotted against the distance from the centroid, as described by Harpending and Ward (1982), where the distance from the centroid r_i for a population i is given by the formula:

$$r_i = (p_i - P)^2 / [P\,(1 - P)]$$

where p_i and P are the frequency of the *Alu* insertion in population i and in the set of all populations, respectively. According to Harpending and Ward (1982), under an island model of population structure there exists a linear relationship between heterozygosity and the distance from the centroid:

$$h_i = H\,(1 - r_i)$$

where h_i and H are the heterozygosities of population i and all populations agglutinated,

respectively. Populations that have experienced more gene flow than average will fall above the theoretical prediction given by the regression line, whereas populations with less gene flow than average will fall below.

Genetic admixture estimates were obtained using ADMIX (Long, 1991). This analysis included as parental samples the Bantu Sotho (Stoneking et al., 1997), a Guarani population (Battilana et al., 2002), and a sample of French individuals (Batzer et al., 1996) or the sample of the city of São Paulo. The Bantu Sotho were selected as representatives of African populations because most Brazilian slaves probably belonged to this ethnic group (Zago et al., 1992; Gonçalves et al., 1994; Figueiredo et al., 1994; Wagner et al., 1996; Pante-de-Sousa et al., 1998). The Guarani population was selected because this Amerindian group is supposed to have been present in the Ribeira River Valley region at the time the quilombos were founded. The French were selected to represent a Western European population because published data on Portuguese populations analyzed for the four loci studied here are not available. The São Paulo sample was alternatively used as representative of European ancestry, since it is composed predominantly by individuals of European ancestry.

RESULTS

Genetic variation within populations

Allelic frequencies, fixation indexes (F), and the observed and expected heterozygosities for the four *Alu* insertions analyzed in all the populations are summarized in Table 2. The four *Alu* insertions were polymorphic in all populations. Twenty-eight Hardy-Weinberg equilibrium tests were performed and only two significant departures from Hardy-Weinberg equilibrium were found (Pedro Cubas for APO and São Pedro for ACE). This is expected since ~2 of the 28 tests should be significant at the 5% level based on chance events alone.

The observed heterozygosities were overall higher in the quilombo remnants than in the São Paulo sample. The observed heterozygosities averaged across the four loci were also high, ranging from 0.399 in São Pedro to 0.500 in Maria Rosa.

Gene flow and genetic differentiation among populations

F_{st} and G_{st} values are summarized in Table 3. The F_{st} values ranged from 0.110–0.044, respectively, for the APO and FXIIIB loci, with an average value of 0.073. The G_{st} values ranged from 0.110–0.043, respectively, for the APO and ACE loci, with an average of 0.067.

A neighbor-joining tree that displays all the populations analyzed and the other worldwide populations is presented in Figure 2. Although only four polymorphic loci were employed in the analysis, Figure 2 shows clearly four main clusters: cluster A groups the Asian and Amerindian populations; cluster B groups the populations from Oceania and two Native American populations; cluster C contains the African and African-derived populations, including São Pedro and Galvão; cluster D is divided in European and European-derived populations, including São Paulo, and in another small group comprised of Pilões, Maria Rosa, and Abobral. Pedro Cubas and the Nigerians fell between cluster A and clusters B, C, and D, with Pedro Cubas closer to clusters B, C, and D than to the Nigerians.

Batzer et al. (1996) argued that since the direction of mutation for *Alu* is the insertion rather than the deletion of each *Alu* element, the root of the tree could be derived by the inclusion of a hypothetical ancestor which did not contain any of the polymorphic *Alu* insertions (i.e., the allele frequencies for each locus were set to zero). When another tree was built using this ancestral population, the ancestor was placed between the two main clusters: that of the Asian and Amerindian populations and that of the other populations (data not shown). The topology of the remaining populations was the same as the observed in the tree of Figure 2.

The main characteristics of the principal components generated after PCA can be observed in Table 4. The first two principal components scores explained 81% of the total variance. The principal component scores generated for each population are presented in Table 5. These scores were used to generate the two-dimensional graph of Figure 3. The Asian and Amerindian populations clustered together, as expected, with the exception of the Incas, which clustered with the populations from Oceania. The Europeans clustered tightly together, with the São Paulo sample among them. The Africans also formed a separate cluster, along with São Pedro and Galvão. Abobral and Pedro Cubas were placed halfway between the Africans and the Asian/Amerindian populations, and Pilões and

TABLE 2. Distribution of polymorphic Alu *insertions*

	APO					ACE				
Population	n	Frequency of Alu	F	Expected heterozygosity	Observed heterozygosity	n	Frequency of *Alu*	F	Expected heterozygosity	Observed heterozygosity
São Paulo	41	0.902	−0.108	0.176 ± 0.052	0.195 ± 0.062	41	0.305	0.022	0.424 ± 0.040	0.415 ± 0.077
Abobral	74	0.750	0.171	0.375 ± 0.036	0.311 ± 0.054	75	0.447	−0.052	0.494 ± 0.010	0.520 ± 0.058
Pedro Cubas	78	0.423	**0.265**	0.488 ± 0.013	0.359 ± 0.054	78	0.641	0.109	0.460 ± 0.022	0.410 ± 0.056
Galvão	50	0.590	0.049	0.484 ± 0.019	0.460 ± 0.070	50	0.360	−0.215	0.461 ± 0.027	0.560 ± 0.070
São Pedro	51	0.392	0.177	0.477 ± 0.022	0.392 ± 0.068	51	0.382	**0.294**	0.472 ± 0.023	0.333 ± 0.066
Pilões	37	0.770	0.007	0.354 ± 0.053	0.351 ± 0.078	37	0.392	0.036	0.477 ± 0.026	0.459 ± 0.082
Maria Rosa	22	0.705	0.017	0.416 ± 0.057	0.409 ± 0.105	22	0.432	−0.204	0.491 ± 0.025	0.591 ± 0.105
Quilombos average	312			0.486 ± 0.007	0.372 ± 0.027	313			0.497 ± 0.003	0.466 ± 0.028

	TPA25					FXIIIB					Total	
Population	n	Frequency of Alu	F	Expected heterozygosity	Observed heterozygosity	n	Frequency of Alu	F	Expected heterozygosity	Observed heterozygosity	Expected heterozygosity	Observed heterozygosity
São Paulo	41	0.524	−0.027	0.499 ± 0.010	0.512 ± 0.078	41	0.476	0.169	0.499 ± 0.010	0.415 ± 0.077	0.399 ± 0.153	0.384 ± 0.134
Abobral	75	0.307	0.059	0.425 ± 0.029	0.400 ± 0.057	75	0.527	0.224	0.499 ± 0.006	0.387 ± 0.056	0.448 ± 0.059	0.404 ± 0.087
Pedro Cubas	78	0.641	0.109	0.460 ± 0.022	0.410 ± 0.056	75	0.493	0.147	0.500 ± 0.005	0.427 ± 0.057	0.477 ± 0.020	0.402 ± 0.029
Galvão	50	0.290	0.077	0.412 ± 0.038	0.380 ± 0.069	50	0.260	0.168	0.385 ± 0.042	0.320 ± 0.066	0.435 ± 0.045	0.430 ± 0.104
São Pedro	43	0.326	0.047	0.439 ± 0.036	0.419 ± 0.075	51	0.304	−0.066	0.423 ± 0.036	0.451 ± 0.070	0.453 ± 0.026	0.399 ± 0.050
Pilões	37	0.432	−0.211	0.491 ± 0.018	0.595 ± 0.081	38	0.474	−0.056	0.499 ± 0.011	0.526 ± 0.081	0.455 ± 0.068	0.483 ± 0.104
Maria Rosa	22	0.523	−0.002	0.499 ± 0.017	0.500 ± 0.107	22	0.477	−0.002	0.499 ± 0.017	0.500 ± 0.107	0.476 ± 0.040	0.500 ± 0.074
Quilombos *average*	305			0.488 ± 0.006	0.433 ± 0.028	311			0.490 ± 0.006	0.421 ± 0.028	0.490 ± 0.005	0.423 ± 0.039

Significant F values are shown in bold.

TABLE 3. Differentiation indexes estimated in the quilombo *remnants*

	Weir and Cockerham (1984)			Nei (1987)		
	f[Fis] (Intra-populational)	theta[Fst] (Inter-populational)	*F*[Fit] (Total)	Hs (Intra-populational)	Gst (Inter-populational)	Ht (Total)
APO	0.160	0.110	0.252	0.432	0.110	0.486
ACE	0.027	0.046	0.072	0.476	0.043	0.497
TPA25	0.042	0.092	0.131	0.454	0.069	0.488
FXIIIB	0.110	0.044	0.149	0.468	0.046	0.490
Total	0.084	0.073	0.151	0.457	0.067	0.490

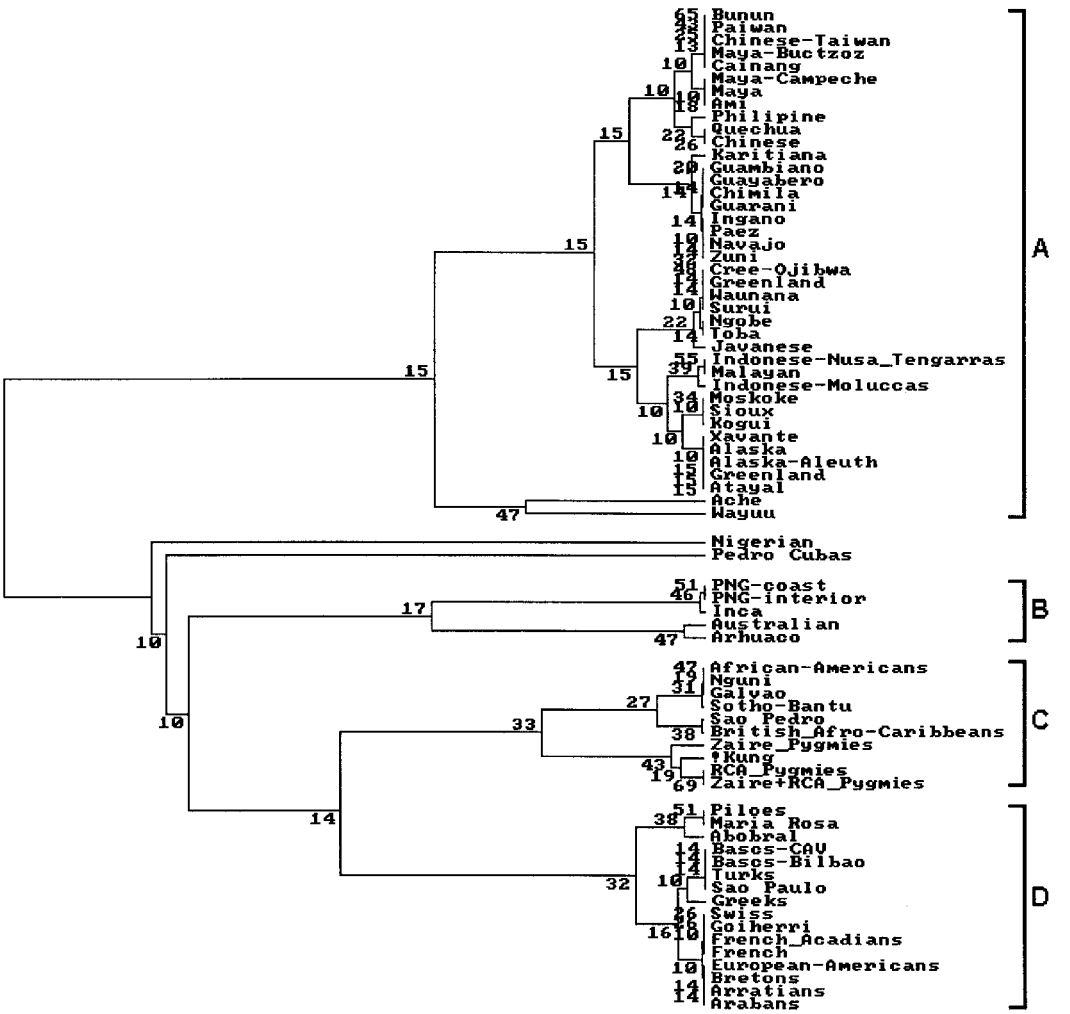

Fig. 2. Neighbor-joining tree of population relationships. This tree was derived directly from the allele frequencies of four polymorphic *Alu* repeats (APO, ACE, TPA25, and FXIIIB) of the populations presented in Table 2 and others previously reported using DISPAN, based on 1,000 replications. The genetic distance between populations is proportional to the length of the branches. The numbers on the nodes indicate the percentage of the bootstrap replicates that support those branches.

TABLE 4. Main characteristics of the principal components based on the allele frequencies of four polymorphic Alu *repeats (APO, ACE, TPA25, and FXIIIB)*

Component	Eigenvalue	Proportion explained (%)	Cumulative proportion explained (%)
1	2.337	58.437	58.437
2	0.914	22.861	81.298
3	0.496	12.391	93.689
4	0.252	6.311	100.000

Maria Rosa were placed closer to the Europeans than to the Africans.

We determined the amount of gene flow experienced by each population by plotting the expected heterozygosities of each population against the distance of that population from the centroid (Fig. 4). We also added to our sample 14 other world populations (Batzer et al., 1996), the Bantu Sotho population (Stoneking et al., 1997), and four Amerindian populations (Battilana et al., 2002).

When compared to other world populations (Fig. 4a), Pedro Cubas, Galvão and São Pedro, the two African-derived populations (African-American and Afro-Caribbean), and two African populations (Nigerians and Sotho) fell well above the theoretical line. When only African and African-derived populations were compared (Fig. 4b), the only population that fell well above the line was Pedro Cubas.

In the first attempt to estimate genetic admixture under a three-hybrid model (Africans, Europeans, and Amerindians) with the French as the European parental population, negative scores and scores greater than 100% were obtained in three out of the four loci analyzed. In a second estimate, using São Paulo as European parental population instead of the French, estimates were obtained in two of the four loci: ACE and FXIIIB. The estimates of the genetic contribution from the three groups,

TABLE 5. Population mean scores for the first two principal components based on the allele frequencies of four polymorphic Alu *repeats (APO, ACE, TPA25, and FXIIIB)*

Population	ID	PC1	PC2
São Paulo	1	−0.54083	1.34488
Abobral	2	−0.9103	−0.29144
Pedro Cubas	3	−0.87331	−0.19134
Galvão	4	−1.76423	−0.17517
São Pedro	5	−2.04588	−0.52134
Pilões	6	−0.83342	0.45703
Maria Rosa	7	−0.77706	0.55673
Guarani	8	1.21369	0.12426
Sotho-Bantu	9	−1.58745	0.15655
African-Americans	10	−1.6493	−0.53835
French	11	−0.12162	1.14179
Karitiana	12	1.63903	−0.33031
Surui	13	0.87178	−1.39089
Cainang	14	0.6964	0.98153
Xavante	15	0.77496	−0.50594
European-Americans	16	−0.10556	0.58758
British_Afro-Caribbeans	17	−1.64707	−0.90938
Indonese-Moluccas	18	0.15736	−0.18601
Indonese-Nusa_Tengarras	19	−0.07077	−0.78449
Javanese	20	0.42035	−1.52773
Philipine	21	0.44505	1.0014
Malayan	22	−0.04278	−0.2867
Australian	23	−0.05374	−2.3154
PNG-coast	24	−1.30215	−1.5295
PNG-interior	25	−1.13683	−1.74134
Greeks	26	−0.11571	1.0764
Turks	27	−0.35282	1.68372
Bretons	28	−0.33994	0.97997
French_Acadians	29	−0.32118	0.37183
Swiss	30	−0.45847	0.92778
Bass-CAV	31	−0.50555	1.85081
Nguni	32	−1.98692	−0.48302
!Kung	33	−1.54136	0.20386
Nigerian	34	−1.43338	−0.80625
RCA_Pygmies	35	−2.2757	0.756
Zaire+RCA_Pygmies	36	−1.96079	0.58202
Zaire_Pygmies	37	−1.32642	1.19999
Ache	38	1.62477	0.43527
Wayuu	39	1.06817	2.15489
Arhuaco	40	0.43875	−2.12443
Chimila	41	1.19358	0.14364
Ingano	42	1.07624	−0.4539
Guambiano	43	1.30956	0.11304
Guayabero	44	1.41413	0.07171
Kogui	45	0.71466	−0.32551
Paez	46	1.0893	−0.39118
Inca	47	−1.01976	−1.9013
Ngobe	48	0.59169	−1.25701
Waunana	49	0.91562	−1.018
Quechua	50	0.85257	0.53395
Toba	51	0.70805	−1.469
Navajo	52	1.32658	−0.51806
Moskoke	53	0.54316	−0.17957
Zuni	54	1.24159	0.14383
Sioux	55	0.49626	0.04901
Cree-Ojibwa	56	0.63906	−1.70365
Maya-Campeche	57	0.75265	−0.01473
Maya-Buctzoz	58	0.35199	0.43547
Maya	59	0.85271	0.38462
Alaska	60	0.15985	−0.75046
Alaska-Aleuth	61	0.09005	−0.72133
Greenland	62	0.03455	−0.35242
Greenland	63	0.64628	−1.65039
Chinese	64	0.41537	0.57519
Chinese-Taiwan	65	0.63924	0.83686
Ami	66	0.95144	1.05668
Atayal	67	0.47044	−0.22114
Bunun	68	0.48163	1.02858
Paiwan	69	0.63859	1.01158
Goiherri	70	−0.32686	1.0127
Arratians	71	−0.09409	1.02835
Arabans	72	−0.04689	1.1277
Bascs-Bilbao	73	−0.379	1.43948

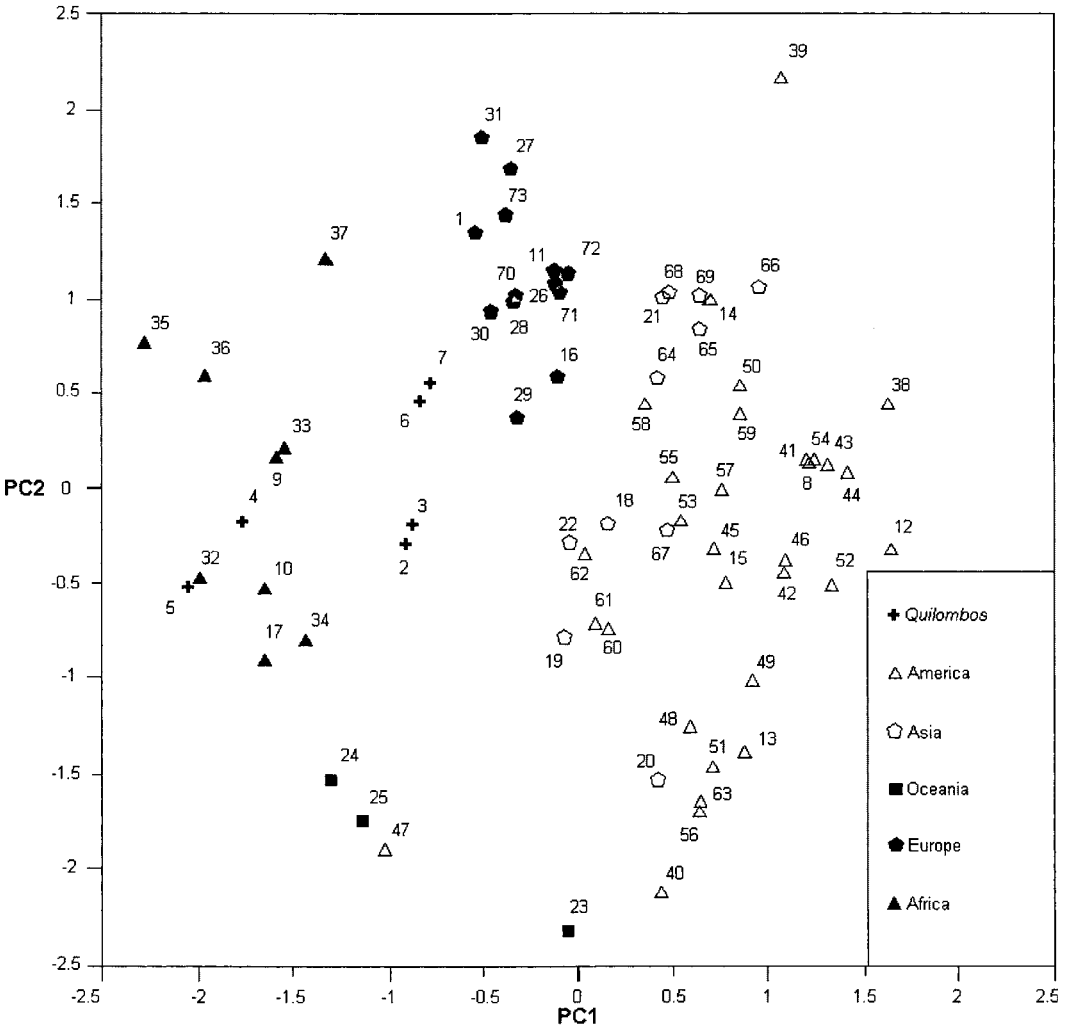

Fig. 3. Worldwide population affinities using the allele frequencies of four polymorphic *Alu* repeats (APO, ACE, TPA25, and FXIIIB): first two principal component scores. The ID numbers are the same as those in Table 5. São Paulo was considered a European-derived population.

based on the ACE and FXIIIB loci only, are presented in Table 6. We did not obtain results in any of the loci for Pedro Cubas.

DISCUSSION

The geographical isolation and the small size of the populations (between 50 and 350 individuals) would lead us to expect some degree of inbreeding or substructuration in the quilombo remnants. Nevertheless, we did not observe significant departures from the Hardy-Weinberg equilibrium in any of the populations, which probably is a consequence of the small sample sizes. On the other hand, it is possible that these populations have undergone a significant degree of gene flow. Indeed, it is not unusual for an individual born in one of these populations to have one of his parents born in a different quilombo remnant within the area. This was confirmed after the pedigrees of the populations were completed. This semi-isolation characteristic of quilombo remnants has been observed by Silva et al. (1999) in a study based on VNTRs and STRs allelic

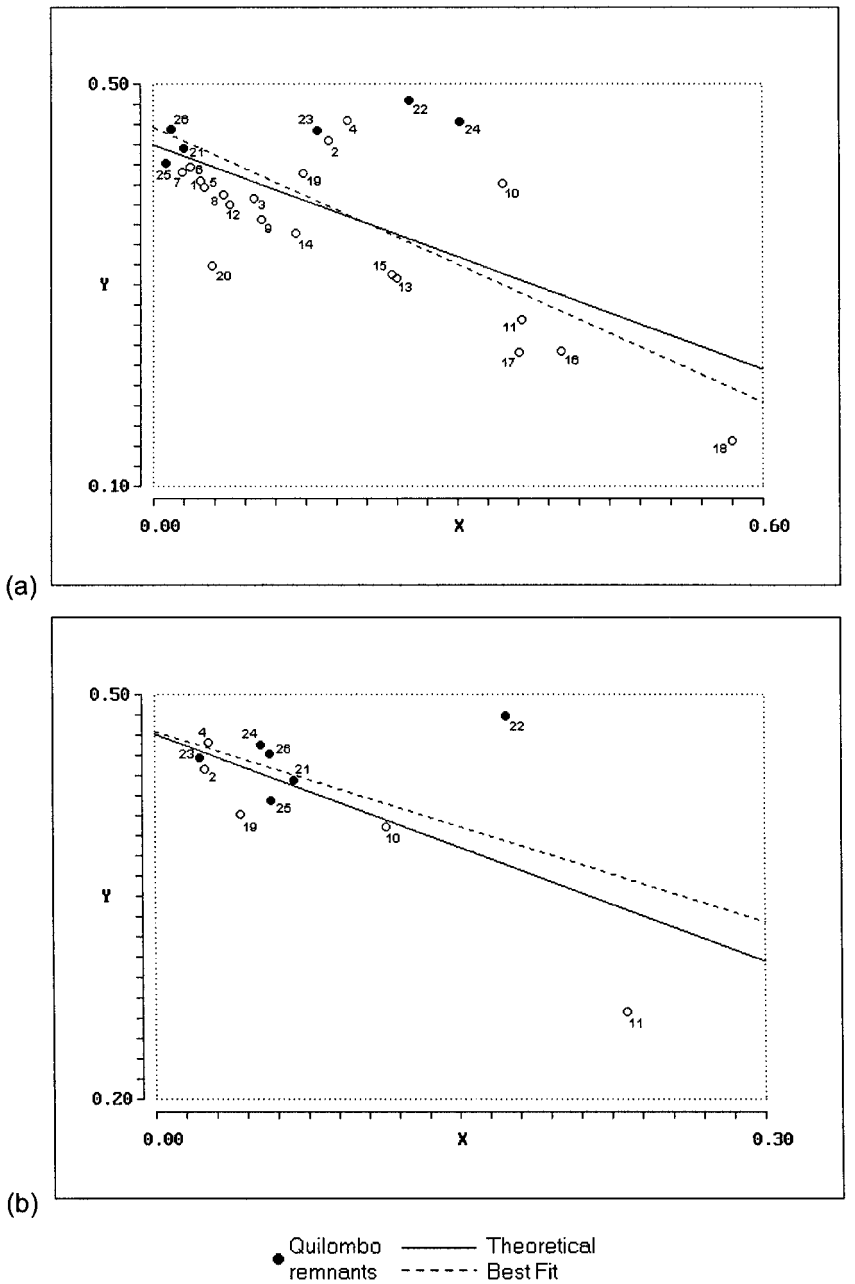

Fig. 4. Distance from the centroid (X) vs. expected heterozygosities (Y). **a:** Worldwide populations. **b:** African and African-derived populations. (1) European-Americans; (2) African-Americans; (3) Hispanics; (4) Afro-Caribbeans; (5) Swiss; (6) Bretons; (7) French Acadians; (8) Greek Cypriots; (9) Turkish Cypriots; (10) Nigerians; (11) Pygmies; (12) French; (13) Alaska Natives; (14) Greenland Natives; (15) Cainang; (16) Guarani; (17) Xavante; (18) Ache; (19) Sotho–Bantu; (20) São Paulo; (21) Abobral; (22) Pedro Cubas; (23) Galvão; (24) São Pedro; (25) Pilões; (26) Maria Rosa. References: populations 1–14: Batzer et al. (1996); populations 15–18: Battilana et al. (2002); population 19: Stoneking et al. (1997); populations 20–26: present study.

TABLE 6. Genetic admixture estimates, based on the allele frequencies of two polymorphic Alu *repeats (ACE and FXIIIB)*

Loci	Parental populations	*Quilombos*				
		Abobral	Galvão	São Pedro	Pilões	Maria Rosa
ACE/FXIIIB	Sotho	0.15	0.74	0.67	0.15	0.26
ACE/FXIIIB	São Paulo	0.63	0.26	0.28	0.74	0.55
ACE/FXIIIB	Guarani	0.23	0.01	0.06	0.12	0.19

frequencies. These authors suggest that the diversity eventually lost due to isolation may have been compensated by the admixture of different ethnic groups (Africans, Europeans, and Amerindians) by the time of the foundation of the quilombos. In our study, the expected heterozygosities were overall higher in quilombo remnants than in São Paulo. This can be explained not only by admixture of the three main ethnic groups, but also among the African groups, since individuals from many different African ethnic groups were shipped together as slaves. Another factor that may account for the results reported here is that African populations usually present higher genetic diversity indexes when compared to other world populations, as reported by many studies that used mitochondrial DNA, Y chromosome, and autosomal microsatellites. This larger genetic diversity has been explained as a consequence of the probable African origin of modern humans, of the larger effective population size of the Africans or of the likely earlier populational expansion that took place in Africa (Bowcock et al., 1994; Jorde et al., 1995, 1997; Shriver et al., 1997; Jorde et al., 2000; Ingman et al., 2000).

The differentiation indexes estimated for the quilombo remnants were similar to those observed for other African populations analyzed. The overall F_{st} in our quilombos was 0.073, higher than the 0.042 value estimated by Watkins et al. (2002). Stoneking et al. (1997) observed G_{st} values of 0.088 in Africans, similar to the 0.067 value estimated in our quilombo remnants. These were higher than those observed for other world populations, with the exception of Amerindians (Stoneking et al., 1997; Watkins et al., 2001, 2002). Within Amerindians, G_{st} values obtained with *Alu* insertions range from 0.102–0.452 (Mateus-Pereira et al., 2004; Oliveira, pers. commun.). Despite the higher F_{st} and G_{st} values, the total variability is clearly within populations rather than among them, a fact that has been observed for almost all human population groups. As a whole, our results suggest that the quilombo remnants have indeed experienced some isolation, capable of generating some degree of differentiation, but their isolation was certainly not so intense as the one experienced by Amerindians. Nevertheless, genetic drift, and founder effect in particular, could explain the larger degree of differentiation observed among these populations and other general groups.

The relationships among these populations are suggested by the topology of the neighbor-joining tree of Figure 2.

Galvão and São Pedro are close to each other and also to the African populations. Galvão and São Pedro are geographically close (3 km). According to their oral lore, Galvão and São Pedro have a common origin. A slave who arrived in the area around 1850 took at least two different wives, having with them at least 24 children. Some of his children probably founded some other populations in the area.

Pilões and Maria Rosa are also geographically close to each other (6 km) and exhibit many links in pedigrees and clustered together as expected in Figure 3.

Pilões, Maria Rosa, and Abobral were genetically closer to the European samples, a finding that might indicate a higher European contribution to these populations. Summing up, some quilombo populations are closer to the African and African-derived populations than to the European or Amerindian populations (São Pedro and Galvão). European contribution, however, is evidenced for Pilões, Maria Rosa, and Abobral, as they were placed closer to the Europeans.

The plotting of the principal component scores revealed approximately the same main clusters of Figure 2. The most interesting findings were the plotting of Pedro Cubas and Abobral between the Africans and the Asian/Amerindian populations and the plotting of Pilões and Maria Rosa

between the Africans and the Europeans, confirming the trend observed in Figure 2.

The admixed nature of quilombo populations is also apparent from the plot of heterozygosity versus distance from the centroid (Fig. 4). Quilombo populations also presented heterozygosity levels higher than the average value, which indicates a larger degree of gene flow experienced by these groups, similar to other admixed African-derived populations (African-Americans and Afro-Caribbeans). Pedro Cubas showed the highest degree of gene flow when compared to other African and African-derived populations (Fig. 4b). Because in Figure 2 Pedro Cubas is in a different cluster than all the other quilombos, we hypothesize that Pedro Cubas may have received a larger Amerindian contribution than other quilombo remnants. This is confirmed by the positioning of Pedro Cubas between the Africans and the Asian and Amerindian populations in Figure 3. In fact, Macedo-Souza (pers. commun.) found in Pedro Cubas some Y chromosome haplotypes that contained the T allele in the DYS199 loci, characteristic of Amerindians. This allele was not detected in the other quilombos.

Admixture estimates gave inconsistent results in two out of the four loci studied. Even when estimates were made for the ACE and FXIIB loci only, it was not possible to obtain results for Pedro Cubas. The partial results obtained for the other quilombo populations are in accordance with the results from PCA: higher levels of African contribution in São Pedro and Galvão, and higher levels of European contribution for Pilões, Maria Rosa, and Abobral. However, the inconsistent results obtained in these analyses indicate that a simple model of admixture is not enough to explain the allelic frequencies observed. Admixture estimates do not account for drastic population size changes or genetic drift. All of the quilombos are small populations. Some of them experienced founder effect (São Pedro and Galvão), while others have experienced severe bottlenecks. In 1998, Pilões had 250 and Maria Rosa 140 inhabitants, whereas in 2001 they had only 128 and 56 inhabitants. These facts may partially explain the difficulties in estimating genetic admixture in these populations.

CONCLUSIONS

The allelic frequencies estimated for the four *Alu* insertions in this study are similar to those observed for other African and African-derived populations. However, some populations exhibited important European genetic contribution. Their heterozygosities were high, a trend also observed in other African and African-derived populations. In conclusion, the quilombo remnants of the Ribeira River Valley behave, as expected, as semi-isolated populations. F_{st} and G_{st} values estimated in our quilombo remnants were high, similar to some African populations, but lower than those observed among Amerindians. A larger degree of differentiation was observed within the quilombo populations than among them. In the neighbor-joining tree some clustered with the African and African-derived populations (São Pedro and Galvão). Others presented an important degree of European contribution (Pilões, Abobral, and Maria Rosa). Pedro Cubas probably also received Amerindian admixture. Similar trends were also observed when the two first principal components scores were plotted against each other. Overall, the quilombos presented a higher degree of gene flow than average when compared to worldwide populations, but similar degree when compared to other African-derived populations.

ACKNOWLEDGMENTS

We thank the people living in the quilombo for their willingness to participate. We also thank L.M. Macedo-Souza, C.B. Angeli, and E. Pardono for helping with the genealogical inferences and with the collection of samples; Dr. Cláudio Leone (Departamento de Pediatria, Hospital das Clínicas, Faculdade de Medicina, Universidade de São Paulo, São Paulo, Brazil), for clinical assistance; and Dr. Walter A. Neves and J.P.V. Atui (Departamento de Biologia, Instituto de Biociências, Universidade de São Paulo) for assistance with the principal component analyses.

LITERATURE CITED

Arpini-Sampaio Z, Costa MC, Melo AA, Carvalho MF, Deus MS, Simoes AL. 1999. Genetic polymorphisms and ethnic admixture in African-derived black communities of northeastern Brazil. Hum Biol 71:69–85.

Bamshad M, Kivisild T, Watkins WS, Dixon ME, Ricker CE, Rao BB, Naidu JM, Prasad BV, Reddy PG, Rasanayagam A, Papiha SS, Villems R, Redd AJ, Hammer MF, Nguyen SV, Carroll ML, Batzer MA, Jorde LB. 2001. Genetic evidence on the origins of Indian caste populations. Genome Res 11:994–1004.

Battilana J, Bonatto SL, Freitas LB, Hutz MH, Weimer TA, Callegari-Jacques SM, Batzer MA, Hill K,

Hurtado AM, Tsuneto LT, 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.

Batzer MA, Kilroy GE, Richard PE, Shaikh TH, Desselle TD, Hoppens CL, Deininger PL. 1990. Structure and variability of recently inserted Alu family members. Nucleic Acids Res 18:6793–6798.

Batzer MA, Arcot SS, Phinney JW, Alegria-Hartman M, Kass DH, Milligan SM, Kimpton C, Gill P, Hochmeister M, Ioannou PA, Herrera RJ, Boudreau DA, Scheer WD, Keats BJB, Deininger PL, Stoneking M. 1996. Genetic variation of recent *Alu* insertions in human populations. J Mol Evol 42:22–29.

Bortolini MC, Zago MA, Salzano FM, Silva-Junior WA, Bonatto SL, da Silva MC, Weimer TA. 1997. Evolutionary and anthropological implications of mitochondrial DNA variation in African Brazilian populations. Hum Biol 69:141–159.

Bortolini MC, Da Silva WA Junior W, De Guerra DC, Remonatto G, Mirandola R, Hutz MH, Weimer TA, Silva MC, Zago MA, Salzano FM. 1999. African-derived South American populations: a history of symmetrical and asymmetrical matings according to sex revealed by bi- and uni-parental genetic markers. Am J Hum Biol 11:551–563.

Bowcock AM, Ruiz-Linares A, Tomfohrde J, Minch E, Kidd JR, Cavalli-Sforza LL. 1994. High resolution of human evolutionary trees with polymorphic microsatellites. Nature 368:455–457.

Comas D, Calafell F, Benchemsi N, Helal A, Lefranc G, Stoneking M, Batzer MA, Bertranpetit J, Sajantila A. 2000. Alu insertion polymorphisms in NW Africa and the Iberian Peninsula: evidence for a strong genetic boundary through the Gibraltar Straits. Hum Genet 107:312–319.

Comas D, Plaza S, Calafell F, Sajantila A, Bertranpetit J. 2001. Recent insertion of an Alu element within a polymorphic human-specific Alu insertion. Mol Biol Evol 18:85–88.

Cooper DN. 1999. Human gene evolution. London: Scientific Publishers. p 221–264, 329–388.

de Pancorbo MM, Lopez-Martinez M, Martinez-Bouzas C, Castro A, Fernandez-Fernandez I, de Mayolo GA, de Mayolo AA, de Mayolo PA, Rowold DJ, Herrera RJ. 2001. The Basques according to polymorphic Alu insertions. Hum Genet 109:224–233.

Deininger PL, Batzer MA, Hutchison CA III, Edgell MH. 1992. Master genes in mammalian repetitive DNA amplification. Trends Genet 8:307–311.

Figueiredo MS, Silva MC, Guerreiro JF, Souza GP, Pires AC, Zago MA. 1994. The heterogeneity of the beta S cluster haplotypes in Brazil. Gene Geogr 8:7–12.

Gonçalves MS, Nechtman JF, Figueiredo MS, Kerbauy J, Arruda VR, Sonati MF, Saad SO, Costa FF, Stoming TA. 1994. Sickle cell disease in a Brazilian population from Sao Paulo: a study of the beta S haplotypes. Hum Hered 44:322–327.

Guerreiro JF, Ribeiro-dos-Santos AKC, Santos EJM, Vallinoto ACR, Cayres-Vallinoto IMV, Aguiar GFS, Santos SEB. 1999. Genetical-demographic data from two Amazonian populations composed of descendants of African slaves: Pacoval and Curiau. Genet Mol Biol 22:163–167.

Harpending HC, Ward R. 1982. Chemical systematics and human populations. In: Nitecki M, editor. Biochemical aspects of evolutionary biology. Chicago: University of Chicago Press. p 213–256.

Hollies CR, Monckton DG, Jeffreys AJ. 2001. Attempts to detect retrotransposition and de novo deletion of Alus and other dispersed repeats at specific loci in the human genome. Eur J Hum Genet 9:143–146.

Ingman M, Kaessmann H, Paabo S, Gyllensten U. 2000. Mitochondrial genome variation and the origin of modern humans. Nature 408:708–713.

International Human Genome Sequencing Consortium. 2001. Initial sequencing and analysis of the human genome. Nature 409:860–921.

Jorde LB, Bamshad MJ, Watkins WS, Zenger R, Fraley AE, Krakowiak PA, Carpenter KD, Soodyall H, Jenkins T, Rogers AR. 1995. Origins and affinities of modern humans: a comparison of mitochondrial and nuclear genetic data. Am J Hum Genet 57:523–538.

Jorde LB, Rogers AR, Bamshad M, Watkins WS, Krakowiak P, Sung S, Keres J, Harpending HC. 1997. Microssatelite diversity and the demographic history of modern humans. Proc Natl Acad Sci USA 94:3100–3103.

Jorde LB, Watkins WS, Kere J, Nyman D, Eriksson AW. 2000. Gene mapping in isolated populations: new roles for old friends? Hum Hered 50:57–65.

Jorde LB, Watkins WS, Bamshad MJ. 2001. Population genomics: a bridge from evolutionary history to genetic medicine. Hum Mol Genet 10:2199–207.

Long JC. 1991. The genetic structure of admixed populations. Genetics 127:417–428.

Majumder PP, Roy B, Banerjee S, Chakraborty M, Dey B, Mukherjee N, Roy M, Thakurta PG, Sil SK. 1999. Human-specific insertion/deletion polymorphisms in Indian populations and their possible evolutionary implications. Eur J Hum Genet 7:435–446.

Mateus-Pereira LH, Socorro A, Fernandez I, Masleh M, Vidal D, Bianchi NO, Batzer MA, Bonatto SL, Salzano FM, Herrera RJ. 2004. Phylogenetic information in polymorphic L1 and Alu insertions from East Asians and Native American populations. Am J Phys Anthropol (in press).

Melton T, Clifford S, Martinson J, Batzer M, Stoneking M. 1998. Genetic evidence for the proto-Austronesian homeland in Asia: mtDNA and nuclear DNA variation in Taiwanese aboriginal tribes. Am J Hum Genet 63: 1807–1823.

Mingroni-Netto RC, Angeli CB, Auricchio MT, Leal-Mesquita ER, Ribeiro-dos-Santos AK, Ferrari I, Hutz MH, Salzano FM, Hill K, Hurtado AM, Vianna-Morgante AM. 2002. Distribution of CGG repeats and FRAXAC1/DXS548 alleles in South American populations. Am J Med Genet 111: 243–252.

Nasidze I, Risch GM, Robichaux M, Sherry ST, Batrzer MA, Stoneking M. 2001. Alu insertion polymorphisms and the genetic structure of human populations from the Caucasus. Eur J Hum Genet 9:267–272.

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

Novick GE, Batzer MA, Deininger PL, Herrera RJ. 1996. The mobile element *Alu* in the human genome. BioScience 46:32–41.

Novick GE, Novick CC, Yunis J, Yunis E, de Mayolo PA, Scheer WD, Deininger PL, Stoneking M, York DS, Batzer MA, Herrera R. 1998. Polymorphic *Alu* insertions and the Asian origin of Native American populations. Hum Biol 70:23–39.

Oliveira SF, Dos Santos EB, De Souza Mendonca PJ, Da Cruz Rocha DC, Dos Santos SE. 2001. Group-specific component (GC) in Curiau and Pacoval, two African-derived Brazilian populations. Am J Hum Biol 13: 718–720.

Ota T. 1983. DISPAN: genetic and phylogenetic analysis. University Park, PA: Institute of Molecular Evolutionary Genetics, Pennsylvania State University.

Pante-de-Sousa G, Mousinho-Ribeiro RC, Santos EJM, Zago MA, Guerreiro JF. 1998. Origin of the hemoglobin

S gene in a northern Brazilian population: the combined effects of slave trade and internal migrations. Genet Mol Biol 21:1415–4757.

Parra EJ, Marcini A, Akey J, Martinson J, Batzer MA, Cooper R, Forrester T, Allison DB, Deka R, Ferrel RE, Shriver MD. 1998. Estimating African American admixture proportions by use of population-specific alleles. Am J Hum Genet 63:1839–1851.

Ribeiro-dos-Santos AK, Pereira JM, Lobato MR, Carvalho BM, Guerreiro JF, Batista Dos Santos SE. 2002. Dissimilarities in the process of formation of Curiau, a semi-isolated Afro-Brazilian population of the Amazon region. Am J Hum Biol 14:440–447.

Roy AM, Carroll ML, Kass DH, Nguyen SV, Salem AH, Batzer MA, Deininger PL. 1999. Recently integrated human Alu repeats: finding needles in the haystack. Genetica 107:149–161.

Roy-Engel AM, Salem AH, Oyeniran OO, Deininger L, Hedges DJ, Kilroy GE, Batzer MA, Deininger PL. 2002. Active Alu element "A-tails": size does matter. Genome Res 12:1333–1344.

Shen MR, Batzer MA, Deininger PL. 1991. Evolution of the master *Alu* gene(s). J Mol Evol 33:311–320.

Sherry ST, Harpending HC, Batzer MA, Stoneking M. 1997. *Alu* evolution in human populations: using the coalescent to estimate effective population size. Genetics 147:1977–1982.

Shriver MD, Jin L, Ferrel RE, Deka R. 1997. Microsatelite data support an early population expansion in Africa. Genome Res 7:586–591.

Silva WA Jr, Bortolini MC, Meyer D, Salzano FM, Elion J, Krishnamoorthy R, Schneider MP, De Guerra DC, Layrisse Z, Castellano HM, Weimer TD, Zago MA. 1999. Genetic diversity of two African and sixteen South American populations determined on the basis of six hypervariable loci. Am J Phys Anthropol 109:425–437.

Stoneking M, Fontius JJ, Clifford SL, Soodyall H, Arcot SS, Saha N, Jenkins T, Tahir MA, Deininger PL, Batzer MA. 1997. Alu insertion polymorphisms and human evolution: evidence for a larger population size in Africa. Genome Res 7:1061–1071.

Wagner SC, Friedrish JR, Job F, Hutz MH. 1996. Caracterização molecular da anemia falciforme em pacientes de Porto Alegre. Rev Bras Genet (Suppl) 19:244.

Watkins WS, Ricker CE, Bamshad MJ, Carrol ML, Nguyen SV, Batzer MA, Harpending HC, Rogers AR, Jorde LB. 2001. Patterns of ancestral human diversity: an analysis of *Alu*-insertion and restriction-site polymorphisms. Am J Hum Genet 68:738–752.

Watkins WS, Ostler CT, Brassington A, Bamshad MJ, Batzer MA, Carrol ML, Nguyen SV, Jorde LB. 2002. Population structure and distribution of 100 polymorphic *Alu* insertion polymorphisms. Am J Hum Genet 71:354.

Weir BS, Cockerham CC. 1984. Analysis of disequilibrium coefficients. In: Hill WG, Mackay TFC, editors. Evolution and animal breeding. London: Commonwealth Agricultural Bureaux. p 45–51.

Zago MA, Figueiredo MS, Ogo SH. 1992. Bantu beta S cluster haplotype predominates among Brazilian blacks. Am J Phys Anthropol 88:295–298.

Zietkiewicz E, Richer C, Sinnet D, Labuda D. 1998. Monophyletic origins of *Alu* elements in primates. J Mol Evol 47:172–182.