AMERICAN JOURNAL OF HUMAN BIOLOGY 15:824–834 (2003)

Historical Genetics: Spatiotemporal Analysis of the Formation of the Brazilian Population

SIDIA M. CALLEGARI-JACQUES,[1,2] DARIO GRATTAPAGLIA,[3,4] FRANCISCO M. SALZANO,[2] SABRINA P. SALAMONI,[2] SHAIANE G. CROSSETTI,[2] MÁRCIO E. FERREIRA[4] AND MARA H. HUTZ[2]
[1]*Departamento de Estatística, Instituto de Matemática, Universidade Federal do Rio Grande do Sul, Campus do Vale, 91540-000 Porto Alegre, RS, Brazil*
[2]*Departamento de Genética, Instituto de Biociências, Universidade Federal do Rio Grande do Sul, 91501-970 Porto Alegre, RS, Brazil*
[3]*Pós-Graduação em Ciências Genômicas e Biotecnologia, Universidade Católica de Brasília, SGAN 916, Mod. B, Asa Norte, 70790-160 Brasília, DF, Brazil*
[4]*Hereditas – Tecnologia em Análise de DNA, SHIN CA 02 lote 19, Lago Norte, 71505-000 Brasília, DF, Brazil*

ABSTRACT A total of 1,037 individuals living in five different sociogeographic regions of Brazil were studied in relation to 12 short tandem repeat polymorphisms. The objective was to assess the degree of European, African, and Amerindian contributions to their gene pools. Although most of the genetic variability was found within regions, significant differences were also observed between regions. The estimated relative proportions of the above-indicated continental contributions showed intermediate values between those obtained with uniparental (mtDNA, Y-chromosome) data, and a higher percentage of European heritage as compared to previous autosome results. A north–south trend of increasing European contribution was also found, in agreement with the history of the Brazilian population. Am. J. Hum. Biol. 15:824–834, 2003.

Brazil, a country of continental size, was first colonized by a wide array of Amerindian groups, which comprised about 2 million persons at the time of the European discovery, in 1500 AD. Afterwards, immigration involved around 9 million Africans who were forced to cross the Atlantic between the 15th and 19th century, a number only matched by the intentional overseas European migration that occurred in the 19th and 20th centuries. Land colonization occurred unevenly along the vast territory of 8.5 million km^2. The first target was the littoral area, but the process of inland expansion started as early as the 17th century. Presently, demographic density is uneven, with large tracts of unoccupied land, especially in the Amazon region (Wehling and Wehling, 1994; Johnson, 1997; Bueno, 1998; http://www.ibge.gov.br; review in Salzano and Bortolini, 2002).

These peoples met and mated among themselves in distinct ways, giving rise to a highly multiethnic admixed population. In the last 50 years many researchers have tried to quantify the relative contributions of these three main groups of people to the present-day Brazilian population. Early studies used blood group and protein genetic loci. A comprehensive review of quantitative estimates can be found in Salzano and Bortolini (2002). Most recent general evaluations involved mitochondrial DNA (mtDNA; Alves-Silva et al., 2000), Y-chromosome (Carvalho-Silva et al., 2001), and population-specific insertion/deletion or single nucleotide polymorphisms (Parra et al., 2003) systems.

Short tandem repeat (STR) polymorphisms have been widely employed in studies aimed at the understanding of the history, demography, and evolution of human populations; they are also frequently used for forensic (paternity and identification) investigations. Here we report data on 12 STR loci in representative samples of Brazil's five sociogeographic regions, which were colonized in diverse ways in Brazilian history. In principle, we would expect much more Amerindian and African influences in the northern and eastern regions than in the south. The questions asked were the following: 1) How do these estimates compare with those pre-

Contract grant sponsors: Programa de Apoio a Núcleos de Excelência (PRONEX), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS), and Pró-Reitoria de Pesquisa da Universidade Federal do Rio Grande do Sul (PROPESQ-UFRGS).

*Correspondence to: Dr. Sidia M. Callegari-Jacques, Departamento de Genética, UFRGS, Caixa Postal 15053, 91501-970 Porto Alegre, RS, Brazil. E-mail: sidia.jacques@ufrgs.br

Received 22 May 2003; Revision received 9 July 2003; Accepted 16 July 2003

Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/ajhb.10217

viously obtained concerning the relative contributions of Europeans, Africans, and Amerindians to the present-day Brazilian population? 2) Are these contributions significantly different among these diverse regions? and 3) If different periods of admixture are considered in the several regions, do the observed rates vary?

SUBJECTS AND METHODS

Samples and DNA extraction

Total genomic DNA was extracted from whole blood samples collected from healthy, unrelated individuals involved in disputed cases of paternity analyzed in the privately owned Hereditas Laboratory, located in Brasília, DF, between 2000 and 2001. All sampled individuals signed an informed consent allowing the use of their DNA samples for paternity testing and further research and statistical investigations. They were born in or close to cities with over 200,000 inhabitants in the five official Brazilian sociogeographic regions. The following cities (all of them with at least 700,000 inhabitants) were the main contributors (about 80%) to the regional samples: Porto Alegre (Southern region); Campinas (Southeastern region); Brasília and Goiânia (Central-West region); Natal, Fortaleza, and Recife (Northeastern region) and Manaus (Northern region) (Fig. 1). Since there is a high degree of interethnic admixture in Brazil, no attempt was made to classify the individuals according to morphological traits. Furthermore, because the laboratory has offered several hundred free paternity tests to subjects of lower socioeconomic strata, the unrelated individuals sampled for this study should not be too much biased toward European-derived subjects, who have a better economic situation and therefore can pay for the tests. The sample, therefore, can be regarded as being reasonably representative of the Brazilian population as a whole. DNA was extracted and purified using the GFX Genomic Blood Kit (Amersham-Pharmacia, Uppsala, Sweden) and quantitated by inspecting the fluorescence of ethidium bromide stains of aliquots in agarose minigels in comparison to standards of known quantity.

STR typing and data analysis

The AmpF*l*STR® Profiler Plus™ kit was used to coamplify the following loci: D3S1358, D5S818, D7S820, D8S1179, D13S317, D18S51, D21S11, FGA, and VWA. The AmpF*l*STR® Cofiler™ kit was used to coamplify the loci CSF1PO, TH01, TPOX, and the overlapping D3S1358 and D7S820 loci. All 12 STR loci are unlinked. Overlapping loci between multiplexes added a quality control check against sample mixing during the development of the profile database. PCR reactions were prepared according to the manufacturer's instructions using between 1–2 ng of template DNA per reaction. For both multiplexes, PCR was carried out in a Perkin Elmer (Norwalk, CT) GeneAmp® PCR system 9600 thermocycler using the following program: 95°C for 11 minutes, followed by 28 cycles at 94°C for 1 minute, 59°C for 1 minute, and 72°C for 1 minute. A final extension was conducted at 60°C for 45 minutes.

PCR products were subjected to electrophoresis in an ABI (Foster City, CA) Prism® 377-XL DNA sequencer. A formamide loading solution (FLS) was prepared by combining 100 μl of blue dextran loading buffer provided in the GeneScan-500 ROX internal Lane Size Standard Kit (PE Biosystems, Foster City, CA) and 500 μl of deionized formamide. After amplification, 4 μl of the PCR product were added to 4.5 μl of FLS, and 0.5 μl of GeneScan-500 ROX size standard, denatured at 95°C for 2 minutes, and immediately chilled on ice. The PCR products (1.5 μl) were loaded in a vertical denaturing, 5% LongRanger (FMC Bioproducts, Rockland, MD)/6.0 M Urea gel solution. The appropriate multilocus allelic ladders were loaded in the first and last lanes of the gel. Data were collected with the ABI Prism® 377 Collection software application using filter F and analyzed using the software GeneScan® Analysis 2.1. Automated genotyping was carried out using the Genotyper® 2.1 software.

Allele frequencies were estimated by direct counting. Admixture proportions were evaluated by means of two estimators. The first (Chakraborty, 1985) is based on the allele frequencies of the source and admixed populations and provides least-squares estimates of the admixture rates using gene identity probabilities. The calculations were performed by the Admix routine written by R. Chakraborty, modified and adapted for Windows by B. Bertoni (Fac. Medicina, Univ. de la República, Montevideo) and available at http://www.genetica.fmed.edu.uy/software.htm. The second estimator, derived by Bertorelle and Excoffier (1998) for two parental populations and

Fig. 1. Map of Brazil indicating its five sociogeographic regions (separated by heavy lines), as well as the states' contours (lighter lines). Also shown are the locations of the cities from which most (about 80%) of the sample was drawn.

extended by Dupanloup and Bertorelle (2001) for any number of such populations, is based on a coalescent approach that explicitly takes into account molecular information as well as gene frequencies. The time from the creation of the hybrid population until present days (t_A) is an important contributor to the admixture coefficients, and can be estimated by the minimum number of pairwise differences observed between a gene drawn from the admixed population and a gene drawn from a parental population. When the molecular information comes from microsatellite loci, and the single-step stepwise model of mutation is assumed, t_A can be estimated by the minimum number of squared differences in allele size observed between a gene drawn from the admixed population and a gene from a parental population. External (e.g., historical) information can also be used to estimate the age of the admixture

event. Using this possibility, we estimated the admixture coefficients for four putative lengths of time considered between the epoch of the hybrid population formation and present time. The admixture proportions were obtained with the ADMIX 2-0 program (http://www2.unife.it/genetica/Isabelle/Isabelle.html), as the average of 1,000 bootstrap replications, considering first the infinite-site and afterwards the stepwise mutation models. Heterogeneity among regions was tested using analysis of molecular variation (AMOVA; Arlequin software; Schneider et al., 2000). Whenever the stepwise mutation model was assumed for the calculations, imperfect repeat alleles were grouped to those nearest in size.

RESULTS

Allele frequencies for the 12 STR loci considering individuals from the five regions as well as for the whole sample are displayed in Table 1. The most variable loci were D18S51, D21S11, and FGA, with 11–18 alleles occurring, depending on the geographic region considered. In these and in other STRs, the number of alleles observed is somewhat less in the Southeast, but this may be due to the fact that this is the region with the smallest sample size. There are differences among the frequencies of the most common alleles of the several regions, but the overall pattern did not display striking differences.

Information about the populations considered as representative of the Brazilian parental populations is listed in Table 2. Unfortunately, sample sizes for the African and Amerindian populations are much smaller than those for Europeans (Spain and Portugal), which number as much 8.8 thousand for TH01, while for Africans the larger sample size is 663 (FGA), and for Amerindians 139 (several loci). Unfortunately, this is inevitable due to the number of investigators presently working in these groups. We assume that no significant deviations, due to this variable sampling, occur in the estimation of the putative parental frequencies.

Table 3 shows the accumulated proportion of ancestry deriving from different continental sources, estimated for people from the five Brazilian regions based on two methods of estimation that do not consider allele size variation. The differences obtained with the two methods are minimal, suggesting that they are estimating reasonably well the different proportions. While the variability among regions is not marked, the percentage of European contribution, as expected by the history of these populations, is higher in the South (81–82%) and lowest in the North (68–71%). The African component is lowest in the South (11%), while the highest values are found in the Center-West and Southeast (18–20%). Extreme values for the Amerindian fraction were found in the South and Southeast (7–8%) and North (17–18%). It is clear, therefore, that a north–south trend of increasing European contribution occurs, with complementary values being observed in relation to the two other sources of genes.

In the last two columns of Table 4 another estimate of accumulated admixture, using Dupanloup and Bertorelle's (2001) method, is given. Different from that provided by the same method and displayed in Table 3, now the allele sizes have been considered. No marked differences were found between the results of the two procedures, and in only one case (Southeast) were they above 10%. The latter is expected, however, if due consideration is given to the high SDs obtained for these figures.

One of the methods of analysis (Dupanloup and Bertorelle, 2001) allows evaluations considering different years of contact, and the corresponding figures are shown in Table 4. Brazil was discovered by the Portuguese in 1500, so the maximum number of years of population admixture can be set as 500 years, while in more remote areas of the country these encounters could have occurred much later, say, 200 years ago. Despite different among-region colonization histories the results are relatively uniform, suggesting about 1% accumulation of African-derived genes in the common gene pool per century. On the other hand, the figures related to the Amerindian heritage indicate that the major miscegenation occurred at the time of the European arrival, with negligible Amerindian contributions in the following centuries.

The AMOVA performed on these data involved two approaches. In the first, alleles were considered as units, independent of size variation, while in the second, size variability was taken into account. Both analyses yielded, however, essentially the same results. By far the highest percentage of total variance (99.9%) was found within regions, but the remaining (0.1%) interregion variability is still significant at the 0.01 (alleles as units) and 0.001 (considering allele sizes) levels, thus validating the differences discussed above.

TABLE 1. Allele frequencies for the 12 short tandem repeat polymorphisms studied in subjects from five Brazilian sociogeographic regions

	Sociogeographic regions					
Systems and alleles	North	Northeast	Center-West	Southeast	South	All samples
CSF1PO						
*6	0.000	0.000	0.002	0.000	0.002	0.001
*7	0.016	0.025	0.014	0.027	0.002	0.015
*8	0.014	0.022	0.010	0.027	0.009	0.014
*9	0.030	0.031	0.026	0.027	0.016	0.026
*10	0.274	0.234	0.260	0.245	0.257	0.257
*11	0.278	0.324	0.309	0.321	0.283	0.299
*12	0.317	0.308	0.314	0.261	0.350	0.317
*13	0.067	0.056	0.058	0.087	0.077	0.067
*14	0.004	0.000	0.007	0.005	0.004	0.004
No. of individuals	252	162	285	92	226	1017
D3S1358						
*12	0.002	0.000	0.002	0.005	0.002	0.002
*13	0.004	0.003	0.002	0.000	0.002	0.002
*14	0.079	0.073	0.089	0.083	0.093	0.084
*15	0.300	0.268	0.275	0.278	0.259	0.277
*16	0.263	0.281	0.300	0.268	0.264	0.277
*17	0.241	0.238	0.198	0.194	0.206	0.216
*18	0.107	0.122	0.123	0.167	0.159	0.131
*19	0.004	0.012	0.009	0.005	0.013	0.009
*20	0.000	0.003	0.002	0.000	0.002	0.002
No. of individuals	253	164	285	108	226	1036
D5S818						
*7	0.026	0.012	0.019	0.009	0.029	0.021
*8	0.010	0.009	0.014	0.018	0.004	0.011
*9	0.056	0.040	0.028	0.005	0.031	0.035
*10	0.056	0.055	0.061	0.078	0.051	0.058
*11	0.372	0.314	0.348	0.321	0.339	0.343
*12	0.315	0.366	0.344	0.385	0.376	0.352
*13	0.151	0.177	0.177	0.170	0.155	0.165
*14	0.014	0.027	0.009	0.014	0.011	0.014
*15	0.000	0.000	0.000	0.000	0.004	0.001
No. of individuals	252	164	286	109	226	1037
D7S820						
*7	0.008	0.012	0.014	0.000	0.020	0.012
*8	0.145	0.174	0.147	0.170	0.173	0.159
*9	0.116	0.118	0.119	0.124	0.104	0.116
*10	0.301	0.280	0.269	0.261	0.323	0.289
*11	0.231	0.224	0.252	0.211	0.200	0.227
*12	0.165	0.152	0.173	0.193	0.142	0.163
*13	0.032	0.031	0.026	0.037	0.038	0.032
*14	0.002	0.009	0.000	0.004	0.000	0.002
No. of individuals	251	161	286	109	225	1032
D8S1179						
*8	0.014	0.012	0.010	0.009	0.013	0.012
*9	0.008	0.009	0.005	0.009	0.009	0.008
*10	0.095	0.049	0.054	0.070	0.104	0.076
*11	0.077	0.101	0.077	0.079	0.053	0.076
*12	0.095	0.113	0.119	0.117	0.150	0.119
*13	0.308	0.308	0.278	0.332	0.272	0.294
*14	0.221	0.241	0.300	0.252	0.240	0.253
*15	0.146	0.149	0.122	0.108	0.128	0.132
*16	0.032	0.018	0.033	0.019	0.022	0.027
*17	0.004	0.000	0.002	0.000	0.009	0.003
*18	0.000	0.000	0.000	0.005	0.000	<0.001
No. of individuals	253	164	286	107	226	1036
D13S317						
*7	0.000	0.000	0.000	0.000	0.002	0.001
*8	0.087	0.094	0.100	0.124	0.129	0.105
*9	0.095	0.119	0.065	0.078	0.102	0.090
*10	0.075	0.040	0.056	0.018	0.060	0.055

(Cont.)

TABLE 1. (Cont.)

Systems and alleles	Sociogeographic regions					All samples
	North	Northeast	Center-West	Southeast	South	
*11	0.259	0.229	0.318	0.243	0.247	0.266
*12	0.259	0.311	0.287	0.344	0.285	0.289
*13	0.148	0.137	0.129	0.110	0.142	0.136
*14	0.071	0.067	0.040	0.078	0.033	0.054
*15	0.006	0.003	0.005	0.005	0.000	0.004
No. of individuals	253	164	286	109	225	1037
D18S51						
*10	0.006	0.006	0.012	0.019	0.007	0.009
*10.2	0.002	0.000	0.000	0.000	0.000	<0.001
*11	0.004	0.009	0.011	0.005	0.007	0.007
*12	0.125	0.117	0.112	0.133	0.111	0.118
*13	0.110	0.089	0.119	0.152	0.151	0.123
*13.2	0.000	0.000	0.000	0.000	0.005	0.001
*14	0.172	0.160	0.144	0.129	0.187	0.161
*14.2	0.000	0.000	0.002	0.000	0.000	0.001
*15	0.131	0.141	0.157	0.143	0.131	0.141
*15.2	0.000	0.003	0.007	0.000	0.000	0.002
*16	0.169	0.160	0.160	0.143	0.138	0.156
*16.2	0.000	0.000	0.004	0.000	0.000	0.001
*17	0.120	0.132	0.112	0.090	0.102	0.113
*17.2	0.000	0.000	0.000	0.000	0.002	<0.001
*18	0.086	0.086	0.071	0.076	0.075	0.078
*19	0.030	0.046	0.041	0.086	0.036	0.043
*20	0.018	0.037	0.030	0.024	0.027	0.027
*21	0.021	0.009	0.007	0.000	0.005	0.010
*21.2	0.000	0.003	0.000	0.000	0.000	<0.001
*22	0.004	0.003	0.007	0.000	0.007	0.005
*23	0.000	0.000	0.004	0.000	0.005	0.002
*24	0.002	0.000	0.000	0.000	0.000	<0.001
*25	0.000	0.000	0.000	0.000	0.000	0.000
*26	0.000	0.000	0.000	0.000	0.002	<0.001
*27	0.000	0.000	0.000	0.000	0.002	<0.001
No. of individuals	251	163	281	105	221	1021
D21S11						
*24.2	0.004	0.000	0.002	0.000	0.005	0.002
*25	0.002	0.003	0.000	0.005	0.000	0.002
*26	0.002	0.000	0.000	0.000	0.002	0.001
*27	0.014	0.027	0.028	0.032	0.038	0.027
*28	0.119	0.137	0.184	0.111	0.140	0.143
*29	0.261	0.213	0.233	0.231	0.233	0.237
*29.2	0.000	0.000	0.005	0.000	0.002	0.002
*30	0.227	0.263	0.203	0.218	0.230	0.225
*30.2	0.030	0.034	0.016	0.028	0.038	0.028
*31	0.048	0.058	0.071	0.079	0.075	0.065
*31.2	0.107	0.119	0.102	0.120	0.097	0.107
*32	0.016	0.015	0.011	0.009	0.014	0.013
*32.2	0.115	0.082	0.087	0.139	0.084	0.098
*33	0.002	0.006	0.003	0.000	0.002	0.003
*33.2	0.039	0.031	0.044	0.023	0.038	0.037
*34	0.004	0.000	0.002	0.000	0.000	0.002
*34.2	0.004	0.003	0.002	0.000	0.000	0.002
*35	0.006	0.003	0.005	0.005	0.002	0.004
*36	0.000	0.003	0.002	0.000	0.000	0.001
*37	0.000	0.003	0.000	0.000	0.000	0.001
No. of individuals	253	164	283	108	221	1029
FGA						
*17	0.002	0.003	0.000	0.000	0.000	0.001
*18	0.006	0.012	0.003	0.005	0.016	0.008
*18.2	0.002	0.003	0.005	0.000	0.002	0.003
*19	0.061	0.076	0.078	0.084	0.069	0.072
*20	0.105	0.113	0.129	0.102	0.077	0.107
*21	0.165	0.162	0.147	0.167	0.168	0.160
*21.2	0.002	0.003	0.000	0.000	0.000	0.001

(Cont.)

TABLE 1. (Cont.)

Systems and alleles	Sociogeographic regions					All samples
	North	Northeast	Center-West	Southeast	South	
*22	0.149	0.180	0.145	0.121	0.190	0.160
*22.2	0.000	0.000	0.005	0.005	0.007	0.003
*23	0.155	0.116	0.170	0.200	0.164	0.158
*23.2	0.002	0.000	0.002	0.000	0.000	0.001
*24	0.130	0.158	0.115	0.153	0.144	0.135
*24.2	0.000	0.003	0.000	0.000	0.000	<0.001
*25	0.146	0.110	0.115	0.098	0.104	0.118
*25.2	0.000	0.003	0.000	0.000	0.000	<0.001
*26	0.047	0.043	0.064	0.051	0.053	0.053
*27	0.010	0.009	0.012	0.014	0.004	0.010
*28	0.010	0.000	0.005	0.000	0.002	0.005
*29	0.002	0.000	0.002	0.000	0.000	0.001
*>30	0.006	0.006	0.003	0.000	0.000	0.004
No. of individuals	253	164	283	75	226	1001
TH01						
*5	0.000	0.000	0.003	0.000	0.000	0.001
*6	0.248	0.228	0.185	0.183	0.254	0.223
*7	0.235	0.235	0.252	0.292	0.204	0.238
*8	0.136	0.125	0.128	0.114	0.106	0.123
*9	0.156	0.171	0.157	0.188	0.153	0.161
*9.3	0.213	0.238	0.264	0.213	0.281	0.246
*10	0.012	0.003	0.011	0.010	0.002	0.008
No. of individuals	253	164	286	101	226	1030
TPOX						
*<6	0.000	0.000	0.000	0.000	0.002	<0.001
*6	0.014	0.015	0.009	0.011	0.004	0.010
*7	0.014	0.009	0.002	0.005	0.004	0.007
*8	0.415	0.491	0.456	0.452	0.436	0.447
*9	0.111	0.125	0.118	0.141	0.133	0.123
*10	0.069	0.055	0.063	0.065	0.066	0.064
*11	0.286	0.247	0.296	0.299	0.306	0.288
*12	0.091	0.055	0.054	0.016	0.049	0.059
*13	0.000	0.003	0.002	0.011	0.000	0.002
No. of individuals	253	164	286	92	226	1021
VWA						
*11	0.002	0.006	0.000	0.000	0.000	0.001
*12	0.000	0.000	0.000	0.000	0.000	0.000
*13	0.004	0.000	0.005	0.005	0.004	0.004
*14	0.075	0.070	0.084	0.069	0.077	0.077
*15	0.126	0.143	0.146	0.153	0.128	0.138
*16	0.299	0.266	0.275	0.250	0.238	0.268
*17	0.275	0.284	0.246	0.283	0.253	0.264
*18	0.148	0.125	0.154	0.194	0.210	0.164
*19	0.061	0.085	0.063	0.037	0.077	0.066
*20	0.008	0.018	0.023	0.009	0.011	0.015
*21	0.002	0.003	0.004	0.000	0.002	0.003
No. of individuals	253	164	285	108	226	1036

DISCUSSION

How do the present values compare with those previously obtained? A first contrast can be made between the estimates obtained by Salzano and Bortolini (2002), based on blood groups and proteins, with those obtained with the STR data presented here. In both cases admixture estimates were made using Chakraborty's (1985) method, therefore avoiding methodological differences. In all five regions the STR series seemed to systematically overestimate the European and underestimate the African contributions, the extreme discrepancy occurring in the Northeast (STR: 75 $\pm$ 0.2% European; 15 $\pm$ 0.2% African contributions; blood groups + proteins: 46 $\pm$ 2% and 44 $\pm$ 2%, respectively). For the African proportions we can also compare our data with those obtained by Parra et al. (2003) in 200 Brazilian self-defined "whites" using 10 autosomal population-specific alleles. The African contributions

TABLE 2. Parental populations employed in the admixture analysis and sources of the information used

	Parental populations		
Systems	European	African	Amerindian
CSF1PO	Portugal, Spain	Benin, Cameroon, Central African Rep., Nigeria	Gavião, Suruí, Wai Wai, Xavante, Zoró
Sample sizes	1498	241	136
D3S1358	Portugal, Spain	Angola, Cameroon, Central African Rep., Guinea-Bissau, Mozambique	Gavião, Suruí, Wai Wai, Xavante, Zoró
Sample sizes	3263	740	139
D5S818	Portugal, Spain	Angola, Guinea-Bissau, Mozambique	Gavião, Suruí, Wai Wai, Xavante, Zoró
Sample sizes	2534	286	139
D7S820	Portugal, Spain	Angola, Guinea Bissau, Mozambique	Gavião, Suruí, Wai Wai, Xavante, Zoró
Sample sizes	2762	286	139
D8S1179	Portugal, Spain	Angola, Guinea-Bissau, Mozambique	Gavião, Suruí, Wai Wai, Xavante, Zoró
Sample sizes	2534	286	139
D13S317	Portugal, Spain	Angola, Guinea Bissau, Mozambique	Gavião, Suruí, Wai Wai, Xavante, Zoró
Sample sizes	2776	286	139
D18S51	Portugal, Spain	Guinea-Bissau, Cameroon, Central African Rep., Angola, Mozambique	Gavião, Suruí, Wai Wai, Xavante, Zoró
Sample sizes	2687	564	139
D21S11	Portugal, Spain	Cameroon, Central African Rep., Guinea-Bissau, Mozambique, Namibia (Ovambo), Uganda (Bantu group)	Gavião, Suruí, Wai Wai, Xavante, Zoró
Sample sizes	2787	615	139
FGA	Portugal, Spain	Guinea Bissau, Mozambique, Namibia (Ovambo, Southwest Bantu group)	Gavião, Suruí, Wai Wai, Xavante, Zoró
Sample sizes	3078	673	139
TH01	Portugal, Spain	Benin, Cameroon, Central African Rep., Mozambique, Namibia (Ovambo), Nigeria, R. of South Africa (Cape Town, Xhosa), Uganda	Gavião, Suruí, Wai Wai, Xavante, Zoró
Sample sizes	8676	627	136
TPOX	Portugal, Spain	Cameroon, Central African Rep., Namibia (Ovambo)	Gavião, Suruí, Wai Wai, Xavante, Zoró
Sample sizes	2322	346	136
VWA	Portugal, Spain	Angola, Guinea-Bissau, Mozambique, Uganda (Bantu group)	Gavião, Suruí, Wai Wai, Xavante, Zoró
Sample sizes	6837	376	139

Source of the data: Europeans and Africans: http://www.uni-duesseldorf.de/WWW/MedFak/Serology/dna.html. Amerindians: Hutz et al. (2002).

TABLE 3. Accumulated proportion of ancestry deriving from different continental sources obtained by two methods of analysis based on the infinite-allele mutation model, compared with estimates derived from mtDNA and Y-chromosome markers

		STR					
		Method 1		Method 2		mtDNA[a]	Y-chrom.[b]
Region	Source of ancestry	Proportion	Standard error	Proportion	Standard error	Proportion	Proportion
North	European	0.68	0.014	0.71	0.032	0.31	0.98
	African	0.14	0.011	0.12	0.029	0.15	0.02
	Amerindian	0.18	0.006	0.17	0.020	0.54	0.00
Northeast	European	0.75	0.002	0.73	0.040	0.34	0.96
	African	0.15	0.002	0.17	0.036	0.44	0.04
	Amerindian	0.10	0.001	0.10	0.024	0.22	0.00
Center-West	European	0.71	0.002	0.72	0.032	—	—
	African	0.18	0.001	0.18	0.029	—	—
	Amerindian	0.11	0.001	0.10	0.018	—	—
Southeast	European	0.75	0.002	0.73	0.051	0.31	0.96
	African	0.18	0.001	0.20	0.045	0.34	0.04
	Amerindian	0.07	0.001	0.07	0.028	0.33	0.00
South	European	0.81	0.014	0.82	0.035	0.66	1.00
	African	0.11	0.012	0.11	0.032	0.12	0.00
	Amerindian	0.08	0.005	0.07	0.021	0.22	0.00

Method 1: Chakraborty (1985); **Method 2:** Dupanloup and Bertorelle (2001).
[a]According to Alves-Silva et al. (2000). Subjects mainly classified as "whites."
[b]According to Carvalho-Silva et al. (2000). Subjects self-defined as "whites."

TABLE 4. Proportion of ancestry deriving from diverse continental sources considering different assumptions about years of contact

	Years of contact									
	500		400		300		200		Unknown	
Region and source of ancestry	Proportion	SE	Proportion	SE	Proportion	SE	Proportion	SE	Proportion	SE
North										
European	0.80	0.068	0.79	0.069	0.78	0.067	0.76	0.062	0.74	0.059
African	0.05	0.073	0.06	0.073	0.07	0.071	0.09	0.065	0.11	0.062
Amerindian	0.15	0.027	0.15	0.026	0.15	0.026	0.15	0.025	0.15	0.025
Northeast										
European	0.83	0.079	0.82	0.075	0.81	0.073	0.80	0.073	0.78	0.068
African	0.15	0.088	0.16	0.083	0.17	0.080	0.18	0.083	0.19	0.076
Amerindian	0.02	0.031	0.02	0.029	0.02	0.030	0.02	0.029	0.03	0.029
Center-West										
European	0.85	0.065	0.84	0.063	0.83	0.061	0.81	0.059	0.79	0.055
African	0.08	0.071	0.09	0.070	0.10	0.068	0.12	0.065	0.14	0.060
Amerindian	0.07	0.025	0.07	0.025	0.07	0.023	0.07	0.023	0.07	0.023
Southeast										
European	0.98	0.086	0.96	0.085	0.95	0.084	0.93	0.082	0.91	0.075
African	−0.06	0.100	−0.04	0.096	−0.02	0.094	−0.01	0.092	0.01	0.083
Amerindian	0.08	0.038	0.08	0.038	0.07	0.037	0.08	0.037	0.08	0.028
South										
European	0.89	0.070	0.88	0.073	0.87	0.066	0.85	0.065	0.83	0.064
African	0.04	0.078	0.05	0.077	0.06	0.072	0.07	0.072	0.09	0.069
Amerindian	0.07	0.026	0.07	0.025	0.07	0.024	0.08	0.025	0.08	0.024

Method used: Dupanloup and Bertorelle (2001).

estimated with this approach were always higher (range of differences: 2–15%) than those obtained with the STR loci studied here. The method of estimation employed by Parra et al. (2003), however, has relatively large standard errors and postulates only two parental populations, preventing valid inferences about the significance of these differences between the estimates. On the other hand, the observed differences between the blood group + protein and STR datasets in the Amerindian proportional contributions did not show any apparent trend.

Uniparental markers (mtDNA, Y-chromosome) can distinguish how much of the observed mixed inheritance derives from female or male differential contributions. It is now a well-established fact that interethnic crosses were asymmetrical in relation to sex in the Brazilian past, European males and African and Amerindian females contributing disproportionally more to the process. A comparison of the present results with those of the mtDNA (Alves-Silva et al., 2000) and Y-chromosome (Carvalho-Silva et al., 2001) estimates, reproduced in the two last columns of Table 3, yielded, as expected, intermediate results. They consistently show more European and less African and Amerindian contributions than the mtDNA data, the opposite being true in relation to the Y-chromosome data. This is expected due to the asymmetrical pattern of interethnic crossings which occurred during Brazilian history: dominant European-derived males crossed predominantly with African-derived and/or Amerindian-derived females due, to a certain extent, to the shortage of females of their own continental derivation (details in Salzano and Bortolini, 2002).

In the interpretation of the above-indicated findings, it should be noted that overall differences in the STR allele frequencies between continental groups are not high. In relation to the 12 loci considered here the average and median differences are as follows: European/African: 3.6% and 2.0%; European/Indian: 5.7% and 4.5%; African/Indian: 6.7% and 4.8%. Some systems, however, show more marked distinctions, namely, TH01, European-African, 9% and 6%; D5S818, European/Indian, 12% for both descriptors; and D5S818, African/Indian, 16% and 13% (http://www.uni-duesseldorf.de/WWW/MedFak/Serology/dna.html; Hutz et al., 2002). Be that as it may, the low differences between Europeans and Africans may condition random deviations, which could influence the admixture estimations. But an alternative explanation is that since the bulk of our sample is composed of individuals who could pay for paternity determinations, it may reflect the marked socioeconomic differentials that exist among people of different ethnic extraction in Brazil. Results of the last (year 2000) census of the Brazilian population (http://www.ibge.gov.br) showed marked economic differences associated with ethnic/color classification. The average monthly income of people self-defined as "black or brown" is about 60.0% (South region) to 51.3% (Southeast) of the amount reported by self-classified "white" persons. Thus, people in a better economic condition are mostly of European extraction, thus at least partly explaining the different proportions observed here and in other samples.

Brazil's colonization started on the coast and only gradually was the interior peopled by Europeans, Africans, and their interethnic descendants. Therefore, the rates of gene flow that occurred especially in the North and Center-West may have been different from those prevailing in the other regions. This should be taken into consideration when the data of Table 4 are examined. On the other hand, the colonization process was also different in the diverse regions, as far as European, African, and Amerindian parentage is concerned. The result is that present (2001) evaluations according to the official National Research by Domicile Sample (http://www.ibge.gov.br) furnishes the following percentages for persons classified as "white": North, 27.9; Northeast, 29.5; Center-West, 43.8; Southeast, 63.5; South, 84.0, a trend in agreement with the north–south gradient obtained in the present research.

The questions posed in the introduction can now be addressed. 1) The present estimates suggest more European and less African contributions to the gene pools of the diverse Brazilian regions. This conclusion, however, should be tempered by the fact that STR frequencies do not differ markedly between these two continental groups. Different socioeconomic stratification may also explain the divergence between the previous and present results. 2) Yes, there are small but significant genetic differences among the inhabitants of the different regions. and 3) The African-derived gene influx was

apparently constant over the centuries, while the Amerindian contribution to the gene pool probably occurred mostly at the time of the European arrival.

ACKNOWLEDGMENTS

We thank the subjects of this investigation for allowing us to use the data originally collected for paternity determination purposes for the present analysis.

LITERATURE CITED

Alves-Silva J, Santos MS, Guimarães PEM, Ferreira ACS, Bandelt H-J, Pena SDJ, Prado VF. 2000. The ancestry of Brazilian mtDNA lineages. Am J Hum Genet 67:444–461.

Bertorelle G, Excoffier L. 1998. Inferring admixture proportions from molecular data. Mol Biol Evol 15: 1297–1311.

Bueno E. 1998. História do Brasil. Porto Alegre: Zero Hora/RBS Jornal.

Carvalho-Silva DR, Santos FR, Rocha J, Pena SDJ. 2001. The phylogeography of Brazilian Y-chromosome lineages. Am J Hum Genet 68:281–286.

Chakraborty R. 1985. Gene identity in racial hybrids and estimation of admixture rates. In: Neel JV, Ahuja Y, editors. Genetic microdifferentiation in man and other animals. New Delhi: Indian Anthropological Association. p 171–180.

Dupanloup I, Bertorelle G. 2001. Inferring admixture proportions from molecular data: extension to any number of parental populations. Mol Biol Evol 18:672–675.

Hutz MH, Callegari-Jacques SM, Almeida SEM, Armborst T, Salzano FM. 2002. Low levels of STRP variability are not universal in American Indians. Hum Biol 74:781–806.

Johnson HB. 1997. A colonização portuguesa do Brasil, 1500–1580. In: Bethell L, editor. História da América Latina, vol. 1. América Latina colonial. São Paulo: Editora da Universidade de São Paulo. p 241–281.

Parra FC, Amado RC, Lambertucci JR, Rocha J, Antunes CM, Pena SDJ. 2003. Color and genomic ancestry in Brazilians. Proc Nat Acad Sci USA 100:177–182.

Salzano FM, Bortolini MC. 2002. The evolution and genetics of Latin American populations. Cambridge, UK: Cambridge University Press.

Schneider S, Kueffer JM, Roessli D, Excoffier L. 2000. Arlequin version 2000: a software for population genetic data analysis. Geneva: Genetics and Biometry Laboratory, University of Geneva.

Wehling A, Wehling MJC. 1994. Formação do Brasil colonial. Rio de Janeiro: Editora Nova Fronteira.