2012

Brazilian urban population genetic structure reveals a high degree of admixture

EUROPEAN JOURNAL OF HUMAN GENETICS, LONDON, v. 20, pp. 111-116, JAN, 2012
http://www.producao.usp.br/handle/BDPI/42318

European Journal of Human Genetics (2012) 20, 111–116

www.nature.com/ejhg

ARTICLE

Brazilian urban population genetic structure reveals a high degree of admixture

Suely R Giolo*[,1,2], Júlia MP Soler[3], Steven C Greenway[4], Marcio AA Almeida[1], Mariza de Andrade[5], JG Seidman[4], Christine E Seidman[4], José E Krieger[1] and Alexandre C Pereira*[,1]

Advances in genotyping technologies have contributed to a better understanding of human population genetic structure and improved the analysis of association studies. To analyze patterns of human genetic variation in Brazil, we used SNP data from 1129 individuals – 138 from the urban population of Sao Paulo, Brazil, and 991 from 11 populations of the HapMap Project. Principal components analysis was performed on the SNPs common to these populations, to identify the composition and the number of SNPs needed to capture the genetic variation of them. Both admixture and local ancestry inference were performed in individuals of the Brazilian sample. Individuals from the Brazilian sample fell between Europeans, Mexicans, and Africans. Brazilians are suggested to have the highest internal genetic variation of sampled populations. Our results indicate, as expected, that the Brazilian sample analyzed descend from Amerindians, African, and/or European ancestors, but intermarriage between individuals of different ethnic origin had an important role in generating the broad genetic variation observed in the present-day population. The data support the notion that the Brazilian population, due to its high degree of admixture, can provide a valuable resource for strategies aiming at using admixture as a tool for mapping complex traits in humans.
European Journal of Human Genetics (2012) **20,** 111–116; doi:10.1038/ejhg.2011.144; published online 24 August 2011

Keywords: genetic structure; Brazilian; admixture mapping; admixture

INTRODUCTION

The advances in genotyping technologies have provided important and considerable insights regarding our views of human population structure. The knowledge of patterns of genetic variation within and among human populations have contributed to a better understanding of the relationship between genetics and ethnicity, as well as improved the design and analysis of case–control association studies. Although there are several studies that have investigated the genetic structure of non-Caucasian populations, including individuals of African, African Americans, Asian, and Native American ancestry, most studies have primarily focused on individuals of European ancestry.[1–12] Therefore, coverage of the global human population remains incomplete with populations from South America being underrepresented in the databases of human genetic variation. Included in these understudied populations are individuals from Brazil, a country of almost 200 million people, which represents approximately 52% of the South American population and 3% of the world's population.

Historically, the Brazilian population always experienced large degrees of intermarriage between ethnic groups, and Brazilians are known to be heavily admixed with Amerindian, European, and African ancestries. In general, Brazilians trace their origins to the original Amerindians and two main sources of immigration: Africans and Europeans.[13,14] In the five geographical regions of Brazil (North, Northeast, Center–West, Southeast, and South), Northern Brazilians are mostly of Amerindian ancestry, with some African ancestry. Current inhabitants of Northeast and Center–West are mostly of African origin, although some individuals whose ancestors migrated from Southern Brazil can trace their roots to Europe. Southern and Southeastern Brazilians are mostly of European origin. However, individuals of African and Asian descent are also found in several localities of the Southeast. For decades, new immigrants, as well as migrants from other parts of Brazil, have flocked to Southeast Brazil where intermarriage between individuals of different ancestry is very common. The goals of the present work are to: (i) identify patterns of population structure among the Southeast Brazilian population enabling individuals from this region to be included in future studies of genetic variation, (ii) to identify marker panels that can effectively capture the variation revealed by dense genotyping from samples of the Southeast Brazilian population and samples from the 11 populations of the HapMap Project, Phase III, which include individuals of Asian, African, European, and Mexican ancestry, and (iii) assess global and local ancestry inferences of the Southeast Brazilian population.

MATERIALS AND METHODS

Datasets and preprocessing steps

Analysis was performed considering samples of the Southeast Brazilian population (BRZ), as well as samples from the following 11 populations of the HapMap database, Phase III: African ancestry in Southwest (ASW), Utah residents with Northern and Western European ancestry from the CEPH collection (CEU), Han Chinese in Beijing, China (CHB), Chinese in Metropolitan Denver, Colorado (CHD), Gujarati Indians in Houston, Texas from the western state of Gujarat in India (South Asia) (GIH), Japanese in Tokyo, Japan

[1]Laboratory of Genetics and Molecular Cardiology, Heart Institute, Medical School of University of Sao Paulo, Sao Paulo, Brazil; [2]Department of Statistics, Federal University of Parana, Curitiba, Brazil; [3]Department of Statistics, University of Sao Paulo, Sao Paulo, Brazil; [4]Department of Genetics, Harvard Medical School, Boston, MA, USA; [5]Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA
*Correspondence: Dr AC Pereira, Laboratory of Genetics and Molecular Cardiology, Heart Institute, University of Sao Paulo Medical School, Sao Paulo, Brazil. Tel: +55 11 3069 5929; Fax: +55 11 3069 5929; E-mail: alexandre.pereira@incor.usp.br or
Dr SR Giolo, Department of Statistics, Federal University of Parana, PO Box 19081, ZIP 81531-990, Curitiba, Brazil. Tel: +55 41 3361 3141; E-mail: giolo@ufpr.br
Received 28 September 2010; revised 27 April 2011; accepted 24 May 2011; published online 24 August 2011

(JPT), Luhya in Webuye, Kenya (LWK), Mexican ancestry in Los Angeles, California (MEX), Masai in Kinyawa, Kenya (MKK), Tuscans in Italy (TSI), and Yoruba in Ibadan, Nigeria (YRI). International HapMap Project, Phase III is available at http://www.sanger.ac.uk/humgen/hapmap3.

The Southeast Brazilian population samples are from a study conducted with trios of individuals (mother, father, and son or daughter), whose children have a congenital heart disease and parents do not. All individuals are from the general urban population of Sao Paulo, the largest metropolitan area of the country. In the present analysis, we have only used data from those unrelated individuals (mothers and fathers). These individuals were enrolled in the current study at the Heart Institute of the University of Sao Paulo. Genotyping for these samples was performed using the Affymetrix SNP array 6.0 platform (Affymetrix, Santa Clara, CA, USA). All subjects gave verbal and written consent. The present protocol was approved by the University of Sao Paulo Medical School IRB (CAPPesq). Samples from the HapMap were genotyped using two platforms, Affymetrix SNP 6.0 and Illumina Human 1M arrays (Illumina, San Diego, CA, USA). More details from the HapMap populations are available from the HapMap Project webpage. Only unrelated individuals were considered in the present analysis. Only SNPs located on the autosomal chromosomes and successfully genotyped in all populations were used for this analysis.

SNPs that were not accurately assessed on the Affymetrix 6.0 array were excluded from the final analysis. That is, we removed, separately for each of the 12 populations, SNPs with more than 5% missing genotype, SNPs that were not in Hardy–Weinberg equilibrium ($P \leq 10^{-4}$), and also those with a minor allele frequency less than or equal to 0.01. At the end of these steps, 365 116 autosomal SNPs, shared by all 12 population data sets and 1129 unrelated individuals representing the 11 HapMap populations (n=991) and the Brazilian population (n=138), remained.

Statistical analysis

We used Principal Components Analysis (PCA), a dimensionality reduction technique,[1,2] to analyze the data. For each population k, the data set consists of n_k unrelated subjects, where each subject has m biallelic SNPs common for all populations. Data for all 12 populations were then displayed in a matrix G of dimension m by n with $n = \sum_{k=1}^{12} n_k$. The values 0, 1, 2, or empty, correspond to the genotypic information assigned to each SNP.[2] After mean-centering and normalizing each row i of the matrix G, n eigenvalues and n corresponding eigenvectors (axes of variation) were calculated, using the covariance matrix of individuals $\psi = G'G$. Plots of the eigenvectors associated with the largest eigenvalues were then used to investigate the structure of the populations under analysis. PCA was run without the removal of outliers and without eliminating SNPs in linkage disequilibrium.

To investigate whether a smaller number of SNPs could effectively capture the variation revealed by the 365 116 common SNPs, we built three panels of markers. The first panel has 250 SNPs, consisting of the top 50 SNPs retained from each of the top five axes of variation. SNPs were ranked on the basis of their loading scores (in absolute value) obtained from the axes of variation. The second and third panels were obtained by retaining the top-ranked 100 and 150 SNPs from each of the same top five axes, respectively. As there were no common SNPs among those retained, the total number of SNPs left in each panel was 250, 500, and 750, respectively. The relationship between the different populations was also investigated by calculating the *F*st statistic, a metric representation of the effect of population subdivision[15,16] for each pair of populations, using the SNPs in the three panels, and also the 365 116 common SNPs. *F*st statistic is often expressed as the proportion of genetic diversity due to allele frequency differences among populations. A zero value implies that the two populations are interbreeding freely and a value of one that the two populations are completely separate.

For global ancestry analysis, we applied the model-based STRUCTURE program[17] to estimate the admixture proportion for the BRZ samples. This was done by applying the STRUCTURE program to two different pooled data sets consisting of four reference populations each (CEU, YRI, MEX, and BRZ, for model 1) and (TSI, ASW, MEX, and BRZ, for model 2), without informing the program which samples were the reference samples. The reason for selecting model 2 was based on the smallest *F*st values obtained between the BRZ and HapMap, Phase III samples of Caucasian and African origin. As seen in Figure 3, the performance of the first two PCs in each of the two different pooled data sets is similar. We allowed the program in such an unsupervised mode to infer the underlying ancestral populations, as well as the ancestral proportion for each subject. The number of ancestral populations K was fixed at 3, 4, 5, and 7. For a given K, we ran STRUCTURE 10 times with different random seeds (10 000 iterations for burn-in phase, and 10 000 iterations for Markov chain optimization and recorded $L(K)$, the log likelihood of the data given K, from each run. We used the metric ΔK to find the optimal K, which is selected to have the largest ΔK value.[18] The inferred number of ancestral populations for the pooled data was 3.

Analyses described above were carried out using the publicly available STRUCTURE,[17] and EIGENSTRAT[2,7] software packages.

RESULTS

Principal components analysis

PCA using the 12 populations showed pronounced patterns of genetic variation within and amongst the populations. To visualize these patterns graphically, we shall consider the top three axes of variation chosen on the basis of their eigenvalues (Figure 1).

The two and three most informative axes of variation, PC1 and PC2 (Figure 2a), and PC1, PC2, and PC3 (Figure 2b), can resolve the 11 populations available in the HapMap study. That is, despite some overlap, we observed that the individuals from the 11 HapMap populations were clearly separated by their different ancestries of origin (African, Asian, European, and Mexican). Asian populations were tightly clustered and distinct from the African and European populations. The Southeast Brazilian population formed a continuum between Europeans and Africans, with some overlap of the Mexican population. The continuum of genotypes observed in the Brazilian population is consistent with the high degree of intermarriage between individuals of the European and African descent.

F_{st} statistic results

The F_{st} statistic was calculated for all population pairs using the 365 116 common SNPs (Table 1). Small *F*st values (0.001 to 0.008) were found for each pair of Asian populations (CHB, CHD, and JPT), indicating less pronounced genetic differences between these populations. Similarly, each pair of African populations (ASW, LWK, MKK, and YRI) is separated by low *F*st scores. Greater *F*st distances (0.128 to 0.168) were observed between Asian and African populations. Populations with European ancestry (CEU and TSI) are also separated by small *F*st values (0.003). Three distinct clusters of ancestral populations (Asian, African, and European) are distinguished by *F*st scores. MEX and GIH populations are closer to the European cluster than to the African cluster as measured by *F*st distance. *F*st scores confirm that the Southeast Brazilian population is close to both the European, African, and Mexican populations.

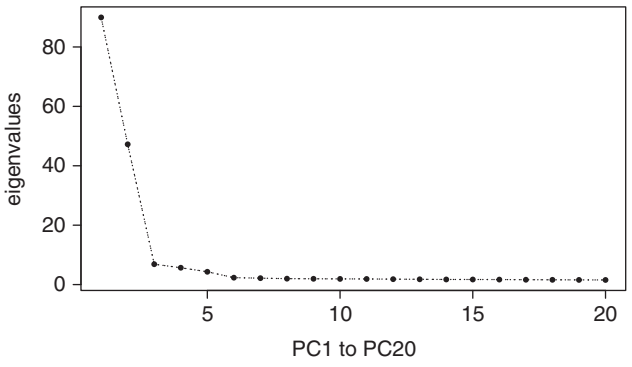

Figure 1 Eigenvalues associated with the 20 first PCs (axes of variation) obtained from the PCA, in which all common SNPs were used.

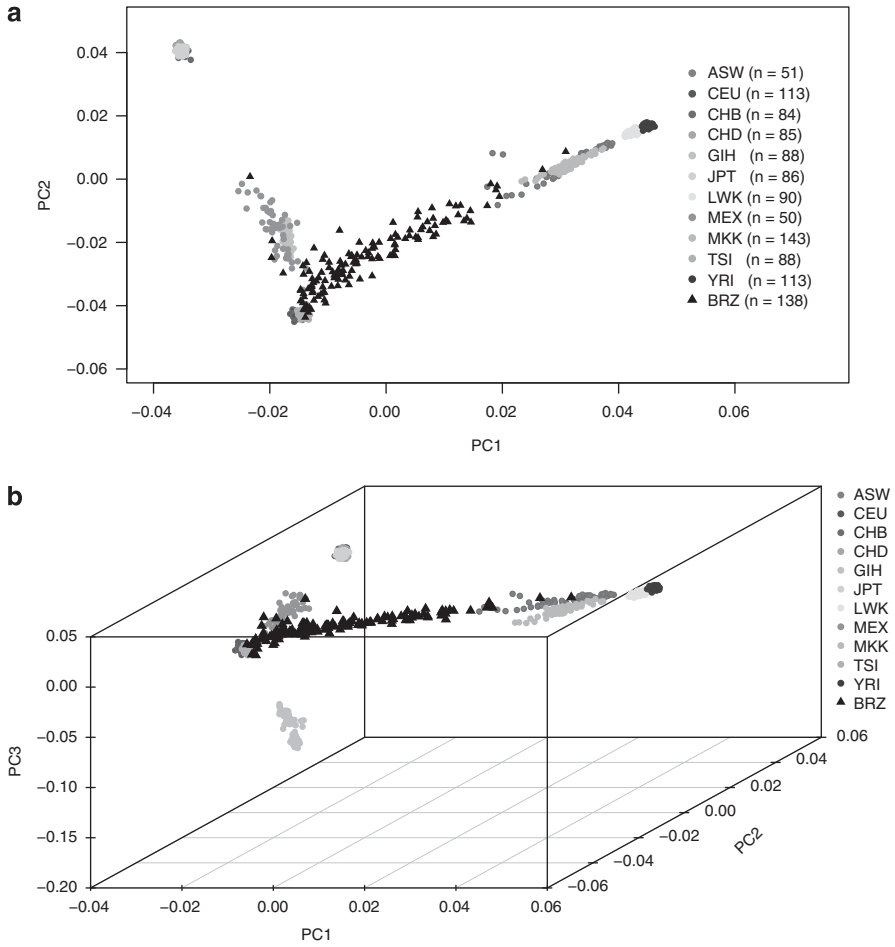

Figure 2 Projection of 1129 individuals from 11 populations of the HapMap Project, Phase III, and the Brazilian population on their (**a**) first and second, and (**b**) first, second, and third axes of variation obtained from PCA, which used 365 116 SNPs. ASW, African ancestry in Southwest; CEU, Utah residents with Northern and Western European ancestry from the CEPH collection; CHB, Han Chinese in Beijing; China, CHD, Chinese in Metropolitan Denver, Colorado; GIH, Gujarati Indians in Houston, Texas; JPT, Japanese in Tokyo, Japan; LWK, Luhya in Webuye, Kenya; MEX, Mexican ancestry in Los Angeles, California; MKK, Masai in Kinyawa, Kenya; TSI, Tuscans in Italy; YRI, Yoruba in Ibadan, Nigeria; and BRZ, Brazilians in São Paulo, Brazil.

Table 1 F_{ST} **statistics calculated between each pair of populations using all 365 116 common SNPs**

	ASW	*CEU*	*CHB*	*CHD*	*GIH*	*JPT*	*LWK*	*MEX*	*MKK*	*TSI*	*YRI*
CEU	0.090										
CHB	0.128	0.103									
CHD	0.129	0.105	0.001								
GIH	0.085	0.034	0.073	0.073							
JPT	0.129	0.105	0.007	0.008	0.074						
LWK	0.010	0.131	0.158	0.159	0.119	0.160					
MEX	0.084	0.029	0.066	0.067	0.035	0.066	0.121				
MKK	0.013	0.092	0.130	0.131	0.086	0.131	0.015	0.087			
TSI	0.089	0.003	0.104	0.105	0.034	0.106	0.128	0.030	0.089		
YRI	0.009	0.141	0.167	0.168	0.129	0.168	0.008	0.130	0.024	0.139	
BRZ	0.047	0.011	0.083	0.084	0.028	0.084	0.079	0.018	0.050	0.010	0.087

Abbreviations: ASW, African ancestry in Southwest; CEU, Utah residents with Northern and Western European ancestry from the CEPH collection; CHB, Han Chinese in Beijing, China; CHD, Chinese in Metropolitan Denver, Colorado; GIH, Gujarati Indians in Houston, Texas; JPT, Japanese in Tokyo, Japan; LWK, Luhya in Webuye, Kenya; MEX, Mexican ancestry in Los Angeles, California; MKK, Masai in Kinyawa, Kenya; TSI, Tuscans in Italy; YRI, Yoruba in Ibadan, Nigeria.

Ancestry informative markers

Small sets of ancestry informative markers (AIMs) that can provide substantial substructure information have been the focus of several studies.[19–21] AIM sets consisting of 200 markers or less can map ancestral origin to Africa, Europe, or Asia. We considered three panels of markers. SNPs on each panel were selected on the basis of their

Table 2 F_{ST} statistics calculated between each pair of populations using Panel 1 (A), Panel 2 (B), and Panel 3 (C)

	ASW	*CEU*	*CHB*	*CHD*	*GIH*	*JPT*	*LWK*	*MEX*	*MKK*	*TSI*	*YRI*
(A)											
CEU	0.3586										
CHB	0.3761	0.3834									
CHD	0.3757	0.3794	0.0004								
GIH	0.2890	0.1031	0.2810	0.2766							
JPT	0.3692	0.3752	0.0059	0.0056	0.2736						
LWK	0.0324	0.4792	0.4596	0.4601	0.4021	0.4519					
MEX	0.2643	0.0625	0.2379	0.2331	0.0778	0.2284	0.3975				
MKK	0.0107	0.3515	0.3561	0.3567	0.2935	0.3504	0.0317	0.2690			
TSI	0.3393	0.0046	0.3828	0.3791	0.0959	0.3744	0.4643	0.0615	0.3336		
YRI	0.0392	0.4987	0.4755	0.4757	0.4286	0.4679	0.0065	0.4223	0.0472	0.4859	
BRZ	0.1793	0.0502	0.2450	0.2429	0.0794	0.2384	0.2988	0.0259	0.1905	0.0418	0.3204
(B)											
CEU	0.3263										
CHB	0.3520	0.3504									
CHD	0.3492	0.3484	0.0002								
GIH	0.2656	0.1004	0.2560	0.2537							
JPT	0.3460	0.3445	0.0051	0.0063	0.2488						
LWK	0.0295	0.4439	0.4327	0.4298	0.3738	0.4272					
MEX	0.2457	0.0566	0.2158	0.2145	0.0780	0.2093	0.3734				
MKK	0.0110	0.3235	0.3359	0.3320	0.2716	0.3222	0.0306	0.2521			
TSI	0.3158	0.0036	0.3562	0.3545	0.0948	0.3494	0.4370	0.0594	0.3134		
YRI	0.0371	0.4662	0.4509	0.4480	0.4030	0.4457	0.0060	0.4005	0.0469	0.4607	
BRZ	0.1666	0.0428	0.2323	0.2310	0.0786	0.2277	0.2805	0.0255	0.1784	0.0383	0.3029
(C)											
CEU	0.3145										
CHB	0.3351	0.3383									
CHD	0.3329	0.3387	3.8e–5								
GIH	0.2559	0.1011	0.2436	0.2422							
JPT	0.3299	0.3352	0.0058	0.0079	0.2397						
LWK	0.0289	0.4298	0.4146	0.4119	0.3623	0.4099					
MEX	0.2307	0.0543	0.2062	0.2067	0.0776	0.2020	0.3551				
MKK	0.0115	0.3119	0.3204	0.3180	0.2623	0.3175	0.0298	0.2385			
TSI	0.3024	0.0030	0.3425	0.3430	0.0959	0.3388	0.4209	0.0569	0.3003		
YRI	0.0363	0.4525	0.4328	0.4303	0.3908	0.4279	0.0061	0.3817	0.0456	0.4452	
BRZ	0.1572	0.0416	0.2267	0.2271	0.0795	0.2237	0.2683	0.0241	0.1698	0.0373	0.2907

Abbreviations: ASW, African ancestry in Southwest; CEU, Utah residents with Northern and Western European ancestry from the CEPH collection; CHB, Han Chinese in Beijing, China; CHD, Chinese in Metropolitan Denver, Colorado; GIH, Gujarati Indians in Houston, Texas; JPT, Japanese in Tokyo, Japan; LWH, Luhya in Webuye, Kenya; MEX, Mexican ancestry in Los Angeles, California; MKK, Masai in Kinyawa, Kenya; TSI, Tuscans in Italy; YRI, Yoruba in Ibadan, Nigeria.

loading scores obtained from a PCA performed on the covariance matrix of the SNPs. The first panel has 250 SNPs consisting of 50 SNPs with highest loading scores (in absolute value) on the top five axes of variation. The second and third panels retained 100 and 150 SNPs, respectively, of the top five axes of variation, and have 500 and 750 markers, respectively. Plots of the two first axes of variation (PC1 and PC2) were obtained by performing PCA for each of the three panels of SNPs (data not shown). The 250 SNP set reproduced the stratification observed with the entire 365 116 SNP set (Figure 2). The 500 and 750 SNP set produced results that were indistinguishable from the 250 SNP set. The chromosomal distribution of the 500 SNP set was uniform. Although the magnitude of the *F*st values varied, the same pattern could be observed for all three panels of markers (Table 2). All three SNP marker panels captured the variation revealed by the entire $>300\,000$ SNP set. Indeed, calculation of the pairwise Spearman correlation coefficient between the four *F*st matrices yielded results always higher than 0.964.

Global ancestry inference of the Brazilian population

Global ancestry inference of the studied samples was able to determine mean ancestries for Amerindian, African, and European. For such, we have first recalculated Eigenstrat principal components, using two different subsets of HapMap samples as 'ancestral' populations. In the first model, we have used the CEU, YRI, and MEX samples to represent, respectively, a Caucasian, African, and Amerindian ancestral population. In the second model, we used the TSI, ASW, and MEX samples to represent such populations. The reason for using the first model was because of the common use of these as ancestral populations in most of the earlier reports. In the second model, we have used the populations with smallest *F*st pairwise differences with the BRZ sample. No significant differences between these two models were observed (Figure 3). Structural analysis, using the 100 most important SNPs from PC1 and PC2, from these two models is presented in Figure 4. In our sampled individuals from the Brazilian Southeast region, mean values were 0.15, 0.24, and 0.61, respectively, for

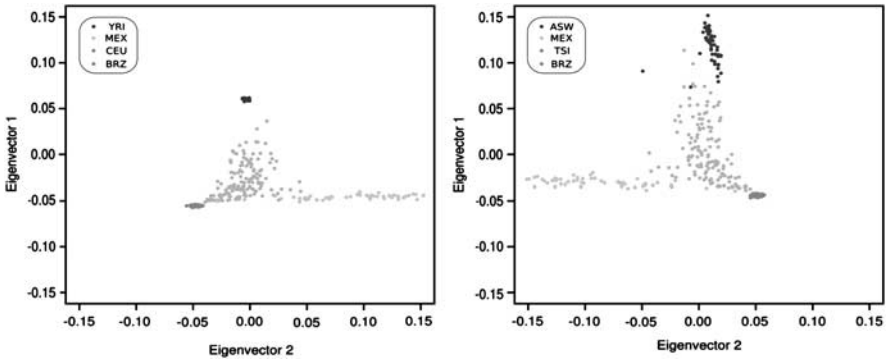

Figure 3 Projection of individuals from three potentially ancestral populations of the HapMap Project, Phase III, and the Brazilian population on their first and second axes of variation (PCs) using Model 1=YRI, CEU, MEX, and BRZ, and Model 2=ASW, TSI, MEX, and BRZ.

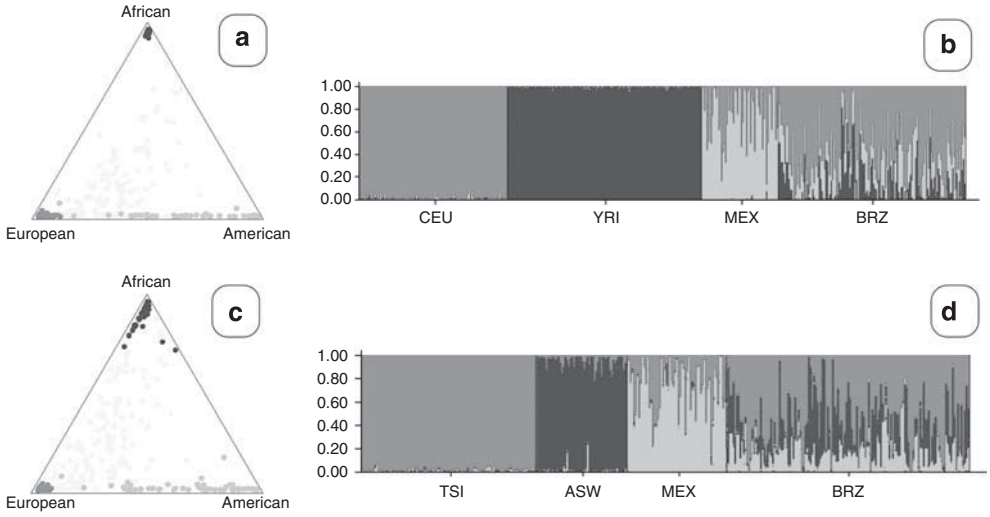

Figure 4 Proportion of membership of each pre-defined population in each of the three clusters. (**a**) Triangular plot of the genomic proportions of African, European, and American ancestry, of the sampled populations from Model I (CEU, YRI, and MEX). (**b**) Barplot structure analyses with admixture model for sampled populations from Model I. (**c**) Triangular plot of the genomic proportions of African, European, and American ancestry, of the sampled populations from Model II (TSI, ASW, and MEX). (**d**) Barplot Structure analyses with admixture model for sampled populations from Model II. Red, blue, and green, represent the proportions of inferred ancestry from European, African, and American ancestral populations. (MEX, Mexican ancestry in Los Angeles, California; CEU, Utah residents with Northern and Western European ancestry from the CEPH collection; YRI, Yoruba in Ibadan, Nigeria; TSI, Tuscans in Italy; ASW, African ancestry in Southwest; and BRZ, Brazilians in São Paulo, Brazil).

Amerindian, African, and European ancestries for Model I markers, and 0.17, 0.27, and 0.56, respectively, for Amerindian, African, and European ancestries for Model II markers (Figure 4).

DISCUSSION

We have compared the genotypic variation of 365 116 SNPs among 1129 unrelated individuals of five continents (Asia, Europe, Africa, and North and South America) to individuals from Southeast Brazil. We demonstrate that this population is a highly admixed population and quite distinct from other HapMap populations. Principle component analyses demonstrate extensive of intermarriage between individuals of African and European descent. This intermarriage occurred between 1500 and the present day reflecting about 20 generations of intermarriage. Thus, the genomes of Brazilian individuals consist of chromosomal segments of distinct ancestry with substantial European and African-related admixture. These findings will have important implications for the correct design and analytical planning of studies exploring complex traits in this population. We expect that the large degree of admixture observed in the Southeast Brazilian population can be exploited for the gene mapping of important disease loci.

The study cohort was collected in Southeast Brazil, in Sao Paulo state. Individuals of African, Amerindian, and perhaps Asian ancestries, may be underrepresented in this study, as individuals with European ancestry comprise a majority in this region. Thus, additional analyses using larger and random samples that can cover all five Brazilian regions might perhaps show an even more pronounced degree of genetic variation than the one suggested by our analysis. Whether the same degree of intermarriage will be observed in other parts of Brazil or other parts of Latin America will be addressed in future studies.

New dense genotyping data from other forthcoming Brazilian studies will determine whether the same pattern of extensive genetic admixture exists in other parts of Brazil.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We thank the CNPq (Brazil, Grant 150653/2008–5) for partial financial support (SRG). This work was supported by FAPESP (Grant 2007/58150-7), and Hospital Samaritano, Sao Paulo.

1 Patterson N, Price AL, Reich D: Population structure and eigenanalysis. *PLoS Genet* 2006; **2**: e190.
2 Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D: Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006; **38**: 904–909.
3 Seldin MF, Shigeta R, Villoslada P *et al*: European population substructure: clustering of northern and southern populations. *PLoS Genet* 2006; **2**: e143.
4 Paschou P, Ziv E, Burchard EG *et al*: PCA-correlated SNPs for structure identification in worldwide human populations. *PLoS Genet* 2007; **3**: 1672–1686.
5 Heath SC, Gut IG, Brennan P *et al*: Investigation of the fine structure of European populations with applications to disease association studies. *Eur J Hum Genet* 2008; **16**: 1413–1429.
6 Paschou P, Drineas P, Lewis J *et al*: Tracing sub-structure in the European American population with PCA-informative markers. *PLoS Genet* 2008; **4**: e1000114.
7 Price AL, Butler J, Patterson N *et al*: Discerning the ancestry of European Americans in genetic association studies. *PLoS Genet* 2008; **4**: e236.
8 Biswas S, Scheinfeldt LB, Akey JM: Genome-wide insights into the patterns and determinants of fine-scale population structure in humans. *Am J Hum Genet* 2009; **84**: 641–650.
9 Xing J, Watkins WS, Witherspoon DJ *et al*: Fine-scaled human genetic structure revealed by SNP microarrays. *Genome Res* 2009; **19**: 815–825.
10 McEvoy BP, Montgomery GW, McRae AF *et al*: Geographical structure and differential natural selection among North European populations. *Genome Res* 2009; **19**: 804–814.
11 Auton A, Bryc K, Boyko AR *et al*: Global distribution of genomic diversity underscores rich complex history of continental human populations. *Genome Res* 2009; **19**: 795–803.
12 Adeyemo A, Gerry N, Chen G *et al*: A genome-wide association study of hypertension and blood pressure in African Americans. *PLoS Genet* 2009; **5**: e1000564.
13 Goncalves VF, Carvalho CM, Bortolini MC, Bydlowski SP, Pena SD: The phylogeography of African Brazilians. *Hum Hered* 2008; **65**: 23–32.
14 Suarez-Kurtz G: *Pharmacogenomics in Admixed Populations*. Landes Bioscience: Austin, 2007.
15 Wright S: Genetical structure of populations. *Nature* 1950; **166**: 247–249.
16 Duan S, Zhang W, Cox NJ, Dolan ME: FstSNP-HapMap3: a database of SNPs with high population differentiation for HapMap3. *Bioinformation* 2008; **3**: 139–141.
17 Falush D, Stephens M, Pritchard JK: Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 2003; **164**: 1567–1587.
18 Wang Z, Hildesheim A, Wang SS *et al*: Genetic admixture and population substructure in Guanacaste Costa Rica. *PLoS One* 2010; **5**: e13336.
19 Yang N, Li H, Criswell LA *et al*: Examination of ancestry and ethnic affiliation using highly informative diallelic DNA markers: application to diverse and admixed populations and implications for clinical epidemiology and forensic medicine. *Hum Genet* 2005; **118**: 382–392.
20 Kosoy R, Nassir R, Tian C *et al*: Ancestry informative marker sets for determining continental origin and admixture proportions in common populations in America. *Hum Mutat* 2009; **30**: 69–78.
21 Enoch MA, Shen PH, Xu K, Hodgkinson C, Goldman D: Using ancestry-informative markers to define populations and detect population stratification. *J Psychopharmacol* 2006; **20**: 19–26.