ABSTRACT
Ancestry informative SNPs can be useful to estimate individual and population biogeographical ancestry. Brazilian population is characterized by a genetic background of three parental populations (European, African, and Brazilian Native Amerindians) with a wide degree and diverse patterns of admixture. In this work we analyzed the information content of 28 ancestry-informative SNPs into multiplexed panels using three parental population sources (African, Amerindian, and European) to infer the genetic admixture in an urban sample of the five Brazilian geopolitical regions. The SNPs assigned apart the parental populations from each other and thus can be applied for ancestry estimation in a three hybrid admixed population. Data was used to infer genetic ancestry in Brazilians with an admixture model. Pairwise estimates of $F_{st}$ among the five Brazilian geopolitical regions suggested little genetic differentiation only between the South and the remaining regions. Estimates of ancestry results are consistent with the heterogeneous genetic profile of Brazilian population, with a major contribution of European ancestry (0.771) followed by African (0.143) and Amerindian contributions (0.085). The described multiplexed SNP panels can be useful tool for bioanthropological studies but it can be mainly valuable to control for spurious results in genetic association studies in admixed populations. Am. J. Hum. Biol. 22:187–192, 2010.

The recent and recurrent admixture process has been one of the key factors in the historical formation and current genetic structure of the Brazilian population. It has involved three major continental groups (Native Americans, Europeans, and Africans), and followed variable patterns of multidirectional introgression according to the social and historical conditions in each geopolitical region along the last five centuries up to the present day (Salzano, 2004; Sans, 2000). According to the 2006 National Household Survey of IBGE (Brazilian Institute of Geography and Statistics), the distribution of ethnic groups based on self-reported skin color is heterogeneous among the Brazilian geopolitical regions (IBGE, 2006). The self-reported classification of “Pardo,” a group including those who declared any sort of admixture or multiethnic origin, represents over 42% of the total Brazilian population, ranging from 15% in Southern region to 71.5% in the Northern region.

This legacy contributed to a unique heterogeneity of the allelic and genotypic frequencies that shapes the current admixed population. The genetic proportions of ancestry in several Brazilian population samples have been broadly investigated using uniparental inherited markers such as mitochondrial DNA (Alves-Silva et al., 2000; Bortolini et al., 1999; Hunemeier et al., 2007; Marrero et al., 2005; Silva et al., 2006) and Y-chromosome polymorphisms (Abe-Sandes et al., 2004; Bortolini et al., 1999; Carvalho-Silva et al., 2001; Hunemeier et al., 2007; Marrero et al., 2005), as well as autosomal STR markers (Callegari-Jacques et al., 2003; Ferreira et al., 2006; Pimenta et al., 2006). Some studies have attempted to use ancestry informative markers (AIM), such as SNPs and insertion/deletion polymorphisms, to estimate genetic ancestry in Brazilians (Blanton et al., 2008; Estrela et al., 2008; Luizón et al., 2008; Parra et al., 2003; Suarez-Kurtz et al., 2007).

The general unanimity presented in previous studies is that the genetic ancestry detected in Brazilian population is widely heterogeneous, follows a clear trend connected to historical facts and is an important issue to be taken into account in genetic association studies.

The use of genetic ancestry-informative markers to help infer historical and anthropological demography processes is in the spotlight of current human admixed population research (Benn-Torres et al., 2008; Martínez-Marignac et al., 2007; Seldin et al., 2007), applications for medical genomics to map genes that influence susceptibility to complex diseases underlying ethnicity factors (Bonilla et al., 2004b; Shaffer et al., 2007) as well as its usefulness for controlling for confounding effects of genetic associations in stratified populations (Gentil et al., 2007, 2009; Matsuzaki et al., 2004; Moreno Lima et al., 2007; Parra et al., 2004; Suarez-Kurtz et al., 2007). Thus the aim of the current study was to evaluate the usefulness of 28 ancestry informative SNPs to estimate the most likely ancestral composition of individuals in population samples from each one of the five Brazilian geopolitical regions and use them as stratification control in association studies.

Additional Supporting Information may be found in the online version of this article.

*Correspondence to: Rinaldo Wellerson Pereira, SGAN 916 Módulo B, Bloco C, 2 Andar Sala 220—Brasília—DF, Brazil, CEP70790-160. E-mail: rinaldo@pos.uchb.br

Received 20 April 2009; Revision received 1 June 2009; Accepted 22 June 2009

©2009 Wiley-Liss, Inc.
SUBJECTS AND METHODS

Brazilian population sample

The Brazilian population sample (BRZ) was composed by 200 healthy, unrelated individuals (103 males and 97 females), involved in cases of paternity investigation between 2003 and 2005. Sampled individuals were randomly chosen among those analyzed at no cost and signed an informed consent allowing the use of their DNA for paternity testing and further anonymous population genetics research. No attempt was made to classify individuals according to income, morphological traits, or to any self-declared classification. A stratified sampling approach was adopted so as to generate five groups of 40 unrelated individuals each derived by birth from one of the five Brazilian geopolitical regions (Center-West—CW, Northeast—NE, North—N, Southeast—SE, and South—S). The research protocol was approved by the Ethical Committee of Universidade Catolica de Brasilia.

Parental population genotypes

Genotypes from three parental populations were kindly provided by Dr. David Reich from a high-density admixture map study in a multiethnic panel of unrelated samples (Smith et al., 2004). The population selected for this work comprised African samples from Botswana, Cameroonian, Ghana, and Senegal (APR n = 120), European–American samples from Baltimore and Chicago (EUR n = 78) and Amerindian Native Mexican Zapotec (AMR n = 29) samples.

Ancestry informative markers

Twenty-eight ancestry informative markers were selected from previous publication (Smith et al., 2004). The ancestry informative markers (AIMs) reported large allelic frequency differences among European, African, and Native American populations and were therefore deemed as a potential set for assigning individual samples to their corresponding ethnic groups. PCR and single base extension primers to genotype the Brazilian samples were designed according to recommendations of the SNaPshot™ Multiplex Kit protocol (Applied Biosystems) using Primer3 web interface (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi). The AIMs were assembled into three multiplex panels and checked for primer dimer and hairpin formation with the Autodimer algorithm (Vallone and Butler, 2004) aiming to maximize the multiplexing success.

Genotyping assay

The PCR was optimized to coamplify the fragments in one single reaction for each AIM panel. The protocol was carried out in 12.5 μl reaction as follows: 1X Platinum® Taq polymerase buffer (Invitrogen, Brazil), 2.5 mM MgCl₂, 250 mM dNTPs, 1.6 mg ml⁻¹ BSA, 0.20 μM of each primer, 10–40 ng DNA, 1 U Platinum Taq DNA polymerase (Invitrogen, Brazil). PCR amplification was performed with an ABI9700 thermocycler using a specific cycling condition named as two-step touchdown PCR, here first described. This method combines the traditional touchdown PCR (Don et al., 1991) with a two-step PCR method (Lopez and Prezioso, 2001) consisting of two temperatures steps per cycle (denaturation and merged annealing/extension) in a way to minimize the thermocycling time. The two-step touchdown PCR cycling conditions were: 5 min at 94°C, 25 cycles of 30 s at 94°C, and 45 s at 65°C decreasing 0.5°C per cycle, and 10 additional cycles at 94°C for 30 s and 52°C for 30 s and a final extension step at 72°C during 5 min. The temperature of 65°C was chosen as an initial annealing/extension condition as its touchdown range was considered to suit most of the primers annealing temperatures. We do not recommend this approach for direct genotyped amplicons, such as STR or indels, since it can generate minus A products. It is suitable for single-base extension reactions only.

The single-base extension (SBE) multiplex reaction was performed using the SNaPshot™ Multiplex kit and electrophoresed on an ABI prism 3100 Genetic Analyzer (Applied Biosystems, USA), under an optimized methodology (Lins et al., 2007). Primer design and concentrations of the single-base extension reactions are detailed in Supporting Information.

Statistical analyses

Estimates of allelic frequency, deviations from Hardy–Weinberg equilibrium (HWE), genetic distance based on Wright’s statistics and a Pairwise Fst test were carried out for the 28 markers in Brazilian population sample using Arlequin v. 3.01 (Excoffier et al., 2005).

The Structure program version 2.1 (Falush et al., 2003; Pritchard et al., 2000) was used to evaluate the population structure via a multilocus genotype clustering of the parental population samples and to estimate the ancestral proportion of each individual in the Brazilian population (BRZ). Genotype data of the 28 markers from the reference population samples were analyzed together with the Brazilian sample to help infer the most likely ancestral proportions in Brazilians. The program parameters were set to use the putative population origin for each individual, a length of burn-in period of 80,000 and 50,000 MCMC extra repetitions, with the number of populations (K) varying from 1 to 10, with 10 replications to check consistency of results. In all calculations the model of admixture with correlated allele frequencies among populations was used. The number of populations (K) was defined by the highest posterior probabilities of log likelihood, log P(X|K), given by the software output summary (Falush et al., 2003; Pritchard et al., 2000). The graphical display of individual ancestry coefficient results was generated using the Distruct software (Rosenberg, 2004).

RESULTS

The set of 28 AIMs SNP provided a detailed picture of the genetic structure of the Brazilian urban population samples from the five geopolitical regions. Allelic frequencies were estimated for the overall Brazilian population and parental populations (Table 1).

A molecular analysis of variance (AMOVA) indicated a low degree of genetic differentiation among the Brazilian geopolitical regions, with 96% of the variance in the sample attributable to variation among individuals within groups. Estimates of Fst revealed no significant genetic differentiation among most of the pairwise comparisons (Table 2). The only exception involved samples from Southern region (S) that were significantly different from all other samples (Table 2). Deviations from HWE
TABLE 1. Ancestry informative markers alleles their genetic position and allelic frequency in African (AFR), Amerindian (AMR), European–American (EUR) as parental populations and Brazilian (BRZ) population samples

| Loci     | Allele | Genetic Position | AFR | AMR | EUR | BRZ | CW | NE | N   | SE | S   |
|----------|--------|------------------|-----|-----|-----|-----|----|----|-----|----|-----|
| rs1420654 | C      | 15q21            | 0.967 | 0.931 | 0.013 | 0.245 | 0.325 | 0.284 | 0.264 | 0.205 | 0.145 |
| rs1408642 | C      | 6q24             | 0.921 | 1.000 | 0.250 | 0.458 | 0.487 | 0.434 | 0.597 | 0.436 | 0.346 |
| rs727563  | C      | 22q13            | 0.811 | 0.948 | 0.253 | 0.424 | 0.423 | 0.527 | 0.347 | 0.461 | 0.353 |
| rs734780  | C      | 15q26            | 0.710 | 0.854 | 0.062 | 0.321 | 0.384 | 0.334 | 0.347 | 0.269 | 0.278 |
| rs730070  | A      | 14q32            | 0.197 | 0.054 | 0.896 | 0.610 | 0.600 | 0.605 | 0.625 | 0.564 | 0.658 |
| rs1129038 | C      | 15q13            | 0.996 | 0.983 | 0.224 | 0.685 | 0.687 | 0.737 | 0.681 | 0.744 | 0.577 |
| rs1240709 | A      | 1p36.3           | 0.500 | 0.103 | 0.766 | 0.570 | 0.462 | 0.592 | 0.500 | 0.615 | 0.679 |
| rs3769384 | C      | 3p14             | 0.783 | 0.875 | 0.154 | 0.364 | 0.400 | 0.382 | 0.400 | 0.436 | 0.205 |
| rs2278354 | T      | 5p15.2           | 0.703 | 0.839 | 0.120 | 0.271 | 0.287 | 0.355 | 0.306 | 0.269 | 0.141 |
| rs6034866 | A      | 20q13.33         | 0.954 | 0.448 | 0.578 | 0.363 | 0.196 | 0.391 | 0.431 | 0.523 | 0.271 |
| rs73418   | T      | 10p11.2          | 0.969 | 0.000 | 0.067 | 0.146 | 0.278 | 0.090 | 0.187 | 0.090 | 0.095 |
| rs803733  | C      | 8q24.3           | 0.013 | 0.410 | 0.880 | 0.701 | 0.681 | 0.718 | 0.650 | 0.755 | 0.757 |
| rs1871534 | C      | 8q24.3           | 0.071 | 1.000 | 0.981 | 0.767 | 0.757 | 0.795 | 0.625 | 0.687 | 0.973 |
| rs222541  | G      | 6q16             | 0.971 | 0.018 | 0.257 | 0.004 | 0.028 | 0.000 | 0.000 | 0.000 | 0.000 |
| rs267071  | C      | 5q22             | 0.088 | 1.000 | 0.654 | 0.545 | 0.520 | 0.646 | 0.548 | 0.500 | 0.520 |
| rs310612  | A      | 20q13.33         | 0.012 | 0.965 | 0.763 | 0.370 | 0.359 | 0.367 | 0.339 | 0.419 | 0.352 |
| rs7678641 | C      | 12p13.2          | 0.012 | 1.000 | 0.923 | 0.173 | 0.210 | 0.192 | 0.145 | 0.115 | 0.237 |
| rs8224803 | C      | 10p11.2          | 0.079 | 0.977 | 0.026 | 0.266 | 0.604 | 0.527 | 0.628 | 0.583 | 0.564 | 0.722 |
| rs7819886 | C      | 2q35             | 0.000 | 0.914 | 0.766 | 0.707 | 0.691 | 0.714 | 0.656 | 0.676 | 0.835 |
| rs1871534 | T      | 15q13            | 0.941 | 0.034 | 0.357 | 0.448 | 0.457 | 0.431 | 0.529 | 0.474 | 0.339 |
| rs7666807 | A      | 12q24.2          | 0.025 | 0.948 | 0.622 | 0.561 | 0.568 | 0.538 | 0.667 | 0.462 | 0.581 |
| rs730086  | C      | 17q21            | 0.063 | 1.000 | 0.667 | 0.581 | 0.500 | 0.577 | 0.597 | 0.564 | 0.722 |
| rs736556  | T      | 7p15             | 0.853 | 0.018 | 0.244 | 0.252 | 0.305 | 0.152 | 0.266 | 0.227 | 0.236 |

BRZ = total sample; CW = Central-west; NE = Northeast; N = North; SE = Southeast; S = South.

The Brazilian sample using Structure software was carried out using a variable K from 2 to K = 10, tested using the admixture model with correlated allele frequencies, hence assuming that each individual has a potential to represent fractions of its genome from each of the parental population. The test assigned individuals with a higher probability to three population cluster (log P(X|K) = −8405.57 for K = 3; log P(X|K) = −8997.2 for K = 2 and log P(X|K) = −8571.48 for K = 4) distinguishing the European, African, and Amerindian population and assigning individual admixture estimates of Brazilians to each of parental population. Individual admixture estimates of all populations are depicted in the bar plot (see Fig. 1) and Brazilian population ancestry proportions estimates are presented in Table 3.

The distribution of individual admixture estimates in the Brazilian population samples comprises a wide range of ancestry proportions (see Fig. 1). Most of individuals were on the extreme of European ancestry and several individuals exhibit a widely distributed three-hybrid pattern of variation. No individuals were near the extremes of the other two groups (African and Native American ancestry), with exception of only one individual from the southern region with 81.2% of Amerindian contribution.

Considering the full Brazilian sample, European ancestry between 90 and 99% was estimated for 104 individuals (52%). Among those, 29.8% are derived from the Southern region (representing 77.5% of the Southern sample), 23.1% from the Southeast (60% of the Southeastern sample), 19.2% from the Northeast (50% of the Northeastern sample), 15.4% from North (40% of the Northern sample), and 12.5% from the Center-west region (32.5% of the Center-western sample).

DISCUSSION

The discovery and validation of very large numbers of SNPs in humans developed panels of ancestry informative markers frequently involving several hundreds to thousands of SNPs and/or indels. However recent reports have shown that for most applications reduced panels involving a few tens of well-selected markers yields essentially the same discrimination power as much larger sets (Kosoy et al., 2008). In this work we evaluated the ancestry information content of a panel of 28 SNPs and applied it to...
investigate the genetic composition of five Brazilian population samples.

The information of population structure of this set of SNPs revealed that the panel displays an ability to correctly assign individual ancestry among African, European, and Amerindian populations. When population stratification is real, deviations from Hardy–Weinberg proportions at some or several loci are detected (Falush et al., 2003; Pritchard et al., 2000). Deviations from HWE were observed only in BRZ, suggesting recent admixture in this population. These results disclose the actual distinction power of this marker set for a population with a three hybrid background of admixture. Consistent with results, the highest log likelihood estimate provided by the Structure analysis indicated the existence of $K = 3$ populations, as expected. It is important to note that inference of $K$ using admixture model with correlated allele frequencies may be challenging if data of parental populations is not available, leading to roughly symmetric proportions ($1/K$) for each assigned population, which means that no structure can be found within the admixed population (Falush et al., 2003; Pritchard et al., 2000). Thus, it is extremely important to include genotype data of all parental population, or the nearest representation to the real parental populations, when testing population structure in admixture model (Lins, 2007).

Hence, regarding the representation of parental population used in this study, we believe that the only sample used in this study that reliably represents the history of migration in Brazil, are those from Cameroon, Ghana, and Senegal. Albeit Botswana is not significantly represented in Brazilian history, individuals from Botswana did not segregate from others of the same continental origin, underlying a more generalist power of informativeness of the SNPs. Furthermore, the sample used to represent Amerindian and European is far from the factual Brazilian native Amerindians or Europeans that colonized Brazil (mainly Portuguese, Spanish, Italian, and German). However, considering the current structure of world populations and its substratifications, and comparing the allelic frequencies of some markers used in this study and others among North-American natives and Amerindians from Brazilian Amazon (Luizon et al., 2008) and among European-Americans and Italian and Spanish samples (Seldin et al., 2006), we reinforce that, since the difference in allelic frequencies among aforementioned situations are small (Luizon et al., 2008; Seldin et al., 2006), the power of informativeness of the present set of SNPs defines ancestries in a continental level of stratification.

The choice of a set of ancestry informative markers depends mainly on two issues: (1) the model and complexity of the population admixture combined with (2) the informativeness of the set of markers used to differentiate individuals to the assigned parental population. Other issues, such as statistical methods of inference, demographic and genetic events and the use of parental population genotype data in the analysis are also relevant and should be taken into account (Tsai et al., 2005). Since Latin American people are derived from unique and specific interethnic admixture events that vary with local history (Salzano, 2004; Sans, 2000), our results reinforce the suggestion that ancestry informative markers should be used with caution regarding the ancestral populations and the markers used to distinguish them (Hoggart et al., 2003; Rosenberg et al., 2003). As more complex the model of admixture becomes, larger marker panels should be necessary, in a way that the power of discrimination must be as homogeneous as possible between all possible pairs of parental groups to avoid any bias in the inferences from one population to another (Kosoy et al., 2008).

Brazilian population samples used in the present study were derived in a stratified way from urban areas of the five geopolitical regions. These regions involve differences due to their historical background of colonization and the presence of isolated populations, such Native tribes in the Northern region and European colonies in Southern region. However, the urban centers from where the samples were derived have experienced significant immigration from inland areas and migrations between other regions, and, consequently, complex and recurrent patterns of ethnic introgression. Additionally, no attempt was made to classify individuals based on physical trait or self-reported ancestry. Any such attempt would likely result in a biased sampling. Since the paternity tests were free of charge, the population samples involved people of variable socioeconomic strata, although likely to be leaning slightly towards the “pardo” group. Interestingly, however, the molecular data showed that more than half of the 200 sampled Brazilians had a predominant European ancestry (>0.900).

**TABLE 3. Relative proportions of genetic ancestry in Brazilian populations samples estimated with the 28 SNP panel and comparison with proportions estimated with STR markers**

| 28 Ancestry SNP set | STR data* |
|---------------------|----------|
| EUR | AFR | AMR | EUR | AFR | AMR |
| North | 0.711 | 0.182 | 0.107 | 0.680 | 0.140 | 0.180 |
| Northeast | 0.774 | 0.136 | 0.089 | 0.750 | 0.150 | 0.100 |
| Center-West | 0.695 | 0.187 | 0.118 | 0.710 | 0.180 | 0.110 |
| Southeast | 0.799 | 0.141 | 0.061 | 0.750 | 0.180 | 0.070 |
| South | 0.877 | 0.070 | 0.052 | 0.810 | 0.110 | 0.080 |
| Brazil | 0.771 | 0.143 | 0.085 | 0.740* | 0.152* | 0.108* |

*Data from Callegari-Jacques et al., 2003.

*Obtained by average values among regional groups. Reference populations for admixture contribution are: EUR = European, AFR = African, AMR = Amerindian.
As expected from social demographic Survey (IBGE, 2006), the Southern region population displayed the highest proportion of European ancestry (Table 3) causing a significant genetic differentiation from the other four geopolitical regions (Table 2) and, consequently, with a lower degree of variance of the estimates of individual admixture (see Fig. 1). This corroborates the information of historical and genetic data of the European colonization and demographic process in the Southern region (Callegari-Jacques et al., 2003; Marrero et al., 2005).

Parra et al. (2003) investigated the ancestry in 200 samples from North, Northeast, Southeastern, and South of Brazil using 10 ancestry informative markers. Our ancestry estimates based on 28 AIMS chosen to maximize allelic frequencies among African, European, and Amerindian showed African population admixture estimate quite different from those presented by Parra et al. (2003), though an agreement is observed regarding the lower African admixture in Southern sample. Tsai et al. (2005) investigated the accuracy of different individual ancestry methods estimators. They showed that with large number of SNPs, both maximum likelihood estimate (MLE) methods, as those applied by Parra et al. (2003), and Bayesian methods (Structure software) give similar accuracy level. However, using smaller sets of SNPs (25–50 SNPs), Bayesian methods shows improved accuracy. We estimate population ancestry using a MLE approach (data not shown) implemented in the IAE3CI software (Bonilla et al., 2004a; Tsai et al., 2005) and the African ancestry estimates were overestimated for all regions, comparable to the values reported by Parra et al. (2003). In spite of the complexity to estimate ancestry, we are confident that our inference generated by Structure is more likely to represent the true population ancestry.

In a previous study the genetic ancestry of Brazilian urban population sample was ascertained based on multilocus profiles of 12 microsatellites commonly used in forensic analyses (Callegari-Jacques et al., 2003). Comparing our results to those reported by Callegari-Jacques et al. (2003), which regards the same source of individuals, the ancestral proportions is in agreement with the north–south gradient towards increased European ancestry, although the relative proportions of ancestries vary slightly (Table 3). The proportions of African ancestry are relatively closer but the European and Amerindian estimates diverge among the two types of DNA markers. In that case, only the population ancestry proportions were compared between very close Brazilian population samples, whereas it is not possible at the time to compare ancestry estimates, neither individual nor population, using the AIM set and forensic microsatellites markers within a particular sample. However, it has been shown that forensic STR loci are valid for admixture estimates, but did not provide precise individual estimate when compared to ancestry informative markers (Barnholtz-Sloan et al., 2005) in a US metropolitan population. Additionally, commercial forensic microsatellites did not demonstrate significant genetic differentiation in groups of self-reported skin color admixed Brazilians (Pimenta et al., 2006). Both accuracy and precision are desirable when such estimation is sought in population genetics and further correlation among forensic STR markers with ancestry informative SNPs regarding the individual ancestry is required. Hence, our results are comparable to those from Callegari-Jacques et al. (2003) in population level, whereas STR markers are not appropriate at individual level and thus the AIM set provides a wider use in biological research, either with bioanthropological aims as in genetic association studies to correct stratification.

Individual ancestry has an intrinsic role in control for population stratification in genetic association studies of markers with the investigated trait or disease, preventing spurious associations or identifying associations regardless of ethnicity or admixed proportions (Blanton et al., 2008; Bonilla et al., 2004b; Gentil et al., 2007, 2009; Martinez-Marignac et al., 2007; Moreno Lima et al., 2007). Regarding that biomedical research sustain the correction of population structure due to particularities of the concept of ethnicity (Pena, 2005), the genetic variability in admixed population (Estrela et al., 2008; Moreno Lima et al., 2007; Suarez-Kurtz et al., 2007)—including haplotype variation among particular populations used for admixture mapping (Piff et al., 2001; Smith et al., 2004)—as well as the lack of control of confounding effects in genetic association studies, our evaluation of a set of 28 SNP would be feasible for such intend. The variation in ancestry proportion is important to understand population stratification that confounds genetic association studies and the estimates can be used as quantitative trait or covariates in any statistical approach (Hoggart et al., 2008).

In conclusion, this study showed that the 28 AIM panel could provide useful genetic structure data for three hybrid admixed population samples. Studies imply that small subsets are shown to be useful tools for estimate genetic ancestry in admixed population (Kosoy et al., 2008; Tsai et al., 2005). Such AIM-panels and individual/ population ancestry estimates are useful not only for bioanthropological research but mainly for controlling the confounding effect of stratification in genetic association studies with candidate genes and phenotypes that underlies ethnicity factors.

### LITERATURE CITED

Abe-Standes K, Silva WA Jr, Zago MA. 2004. Heterogeneity of the Y chromosome in Afro-Brazilian populations. Hum Biol 76:77–86.

Alves-Silva J, da Silva Santos M, Guimarães PE, Ferreira AC, Bandelt HJ, Prado VP. 2004. The ancestry of Brazilian mtDNA lineages. Am J Hum Genet 67:444–461.

Barnholtz-Sloan JS, Pfaff CL, Chakraborty R, Long JC. 2005. Informativeness of the CODIS STR loci for admixture analysis. J Forensic Sci 50:1322–1326.

Benn-Torres J, Bonilla C, Robbins CM, Waterman L, Moses TY, Hernandez W, Santos ER, Bennett F, Aiken W, Tulloch T, Coard K, Hennais A, Wu S, Nemessure B, Leske MC, Freeman V, Carprien J, Kittles RA. 2009. Admixture and population stratification in African Caribbean populations. Ann Hum Genet 72 (Part 1):90–98.

Blanton RE, Silva LK, Morato VG, Parrado AR, Dias JP, Melo PR, Reis EA, Goddar RA, Nunes MR, Rodrigues SG, Vasconcelos PF, Castro JM, Reis MG, Barreto ML, Teixeira MG. 2008. Genetic ancestry and income are associated with dengue hemorrhagic fever in a highly admixed population. Eur J Hum Genet 16:762–765.

Bonilla C, Parra Ed, Pfaff CL, Dias S, Marshall JA, Hamman RF, Ferrell RE, Hoggart CL, McKiege PM, Shriver MD. 2004a. Admixture in the Hispanics of the San Luis Valley, Colorado, and its implications for complex trait gene mapping. Am J Hum Genet 73:139–153.

Bonilla C, Shriver MD, Parra Ed, Jones A, Fernandez JR. 2004b. Ancestral proportions and their association with skin pigmentation and bone mineral density in Puerto Rican women from New York City. Hum Genet 115:57–68.

Bortolini MC, Da Silva WAJW, De Guerra DC, Remonatto G, Miranda 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.
Callegari-Jacques SM, Grattapaglia D, Salzano FM, Salamoni SP, Cossotti SG, Ferreira ME, Hutz MH. 2003. Historical genetics: spatiotemporal analysis of the formation of the Brazilian population. Am J Hum Biol 15:824–834.

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

Don RH, Cox PT, Wainwright BJ, Baker K, Mattick JS. 1991. “Touchdown” PCR to circumvent spurious priming during gene amplification. Nucleic Acids Res 19:4003–4008.

Estrela RC, Ribeiro FS, Carvalho RS, Gregorio SP, Dias-Neto E, Struchiner CJ, Suarez-Kurtz G. 2008. Distribution of ABCB1 polymorphisms among Brazilian: impact of population admixture. Pharmacogenomics 9:267–276.

Exonfilter L, Laval G, Schneider S. 2005. Arelequin version 3.0: an integrated software package for population genetics data analysis. Évol Bioinfor Online 1:47–50.

Falush D, Stephens M, Pritchard JK. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164:1567–1587.

Ferreira LB, Mendes-Junior CT, Wielze CE, Luizon MR, Simeoa AL. 2006. Genomic ancestry of a sample population from the state of Sao Paulo, Brazil. Am J Hum Biol 18:702–705.

Gentil P, Lima RM, Lins TC, Abreu BS, Pereira RW, Oliveira RJ. 2007. Physical activity. Cdx-2 genotype, and BMI. Int J Sports Med 28:1065–1069.

Gentil P, Lins TC, Moreno Lima R, Abreu BS, Grattapaglia D, Bottaro M, Oliveira RJ, Pereira RW. 2009. Vitamin-D-receptor genotypes and bone-mineral density in postmenopausal women: interaction with physical activity. J Aging Phys Activity 17:31–45.

Hoggart CJ, Parra EJ, Shriver MD, Bonilla C, Kittles RA, Clayton DG, Gentil P, Lima RM, Lins TC, Abreu BS, Pereira RW, Oliveira RJ. 2007. A better way to optimize: two-step gradient dimer and hairpin structures. Biotechniques 37:226–231.

Hogsett CJ, Parra EJ, Shriver MD, Bonilla C, Kittles RA, Clayton DG, Bortolini MC. 2005. Heterogeneity of the genome ancestry of individuals classified as White in the state of Rio Grande do Sul, Brazil. Am J Hum Biol 17:496–506.

Martinez-Martignac VL, Valladares A, Cameron E, Chan A, Perera A, Bortolini MC. 2005. Heterogeneity of the genome ancestry of individuals classified as White in the state of Rio Grande do Sul, Brazil. Am J Hum Biol 17:496–506.

Matsuzaki H, Loi H, Dong S, Tsai YY, Fang J, Law J, Di X, Liu WM, Yang R, [et al. 2008. Ancestry profiles of a sample population from the state of Sao Paulo, Brazil. Am J Hum Biol 18:702–705.

McKeigue PM. 2003. Control of confounding of genetic associations in stratified populations. Am J Hum Genet 72:1492–1504.

Munemeyer T, Carvalho C, Marrero AR, Salzano FM, Junho Pena SD, Bortolin MC. 2007. Niger-Congo speaking populations and the formation of the Brazilian people. Phys Anthropol 133:384–387.

Nogueira LR, Lima RM, Gentil P, Oliveira RJ, Pereira RW. 2007. Impact of admixture on the performance of HapMap data in inference genetic structure: large variance in Amerindian contribution. Am J Phys Anthropol 132:455–462.

Parra EJ, Bonilla C, Kittles RA, Shriver MD. 2004. Implications of correlations between skin color and genomic ancestry for biomedical research. Nat Genet 36(11 Suppl):S84–S86.

Parra RC, Amado RC, Lambertucci JR, Rocha J, Antunes CM, Pena SD. 2003. Color and genomic ancestry in Brazilians. Proc Natl Acad Sci USA 100:179–182.

Pena SD. 2005. Reasons for banishing the concept of race from Brazilian medicine. Hist Cienc Saude Manguinhos 12:321–346.

Pfaff CL, Parra EJ, Bonilla C, Hester K, McKeigue PM, Kamboh MI, Hutchinson RG, Perrell RE, Boerwinkle E, Shriver MD. 2001. Population structure in admixed populations: effect of admixture dynamics on the pattern of linkage disequilibrium. Am J Hum Genet 68:198–207.

Pimenta JA, Zuccherato LW, Debes AA, Maselli L, Soares RP, Moura-Neto RS, Rocha J, Rydlewski SP, Pena SD. 2006. Color and genomic ancestry in Brazilians: a study with forensic microsatellites. Hum Hered 62:190–195.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945–955.

Rosenberg NA. 2004. Distruct: a program for the graphical display of population structure. Mol Ecol Notes 4:137–138.

Rosenberg NA, Li LM, Ward R, Pritchard JK. 2003. Informativeness of genetic markers for inference of ancestry. Am J Hum Biol 73:1402–1422.

Salzano FM. 2004. Interethnic variability and admixture in Latin American: social implications. Rev Bio Trop 52:405–415.

Sans M. 2000. Admixture studies in Latin America: from the 20th to the 21st century. Hum Biol 72:155–177.

Seldin MF, Shigeta R, Villoslada P, Selmi C, Tuomilehto J, Silva G, Belmont J, Klarskog L, Gregersen PK. 2006. European population substructure: clustering of northern and southern populations. PLoS Genet 2:e143.

Seldin MF, Tian C, Shigeta R, Scherbarth HR, Silva G, Belmont JW, Kittles R, Gammon S, Allevi A, Palatnik SA, et al. 2007. Argentinian population genetic structure: large variance in Amerindian contribution. Am J Phys Anthropol 132:455–462.

Shaffer Jr, Kammerer CM, Reich D, McDonald G, Patterson N, Goodpaster B, Bauer DC, Li J, Newman AB, Cauley JA, et al. 2007. Genetic markers for ancestry are correlated with body composition traits in older African Americans. Osteoporos Int 18:733–741.

Silva WA, Bortolini MC, Schneider MP, Marrero A, Elion J, Krishnanmorthy R, Zago MA. 2006. MiDNA haplogroup analysis of black Brazilian and sub-Saharan populations: implications for the Atlantic slave trade. Hum Biol 78:29–41.

Smith MW, Patterson N, Lautenberger JA, Truelove AL, McDonald GJ, Waliszewska A, Kessing BD, Malansky MJ, Scafe C, Le E, Dager PL, Migault AA, Yi Z, De The G, Exess M, Sankale JL, Moore JH, Poku K, Phair JP, Goedert JL, Vlahov D, Williams SM, Tishkoff SA, Winkler CA, De La Vega FM, Woodage T, Sivin JS, Zago MA, Pena SD. 2005. Heterogeneity of the genome ancestry of individuals classified as White in the state of Rio Grande do Sul, Brazil. Am J Hum Biol 17:496–506.

Tsiatis M, Houdry B, Barbosa F, Rezende L, Andrade F, Schork NJ, et al. 2007. Admixture in Mexico City: implications for admixture mapping of type 2 diabetes genetic risk factors. Hum Genet 120:807–819.

Tsai HJ, Choudhry S, Naqui M, Rodriguez-Cintron W, Burchard EG, Ziv E. 2005. Comparison of three methods to estimate genetic ancestry and control for stratification in genetic association studies among admixed populations. Hum Genet 118:424–435.

Vallone PM, Butler JM. 2004. AutoDimer: a screening tool for primer-dimer and hairpin structures. Biotechniques 37:226–231.