Self-reported skin color, genomic ancestry and the distribution of *GST* polymorphisms

Guilherme Suarez-Kurtz[a], Daniela D. Vargens[a], Claudio J. Struchiner[b], Luciana Bastos-Rodrigues[c] and Sergio D.J. Pena[c]

Background and objective **Skin color and self-reported ethnicity have systematically been used in the pharmacogenetic/-genomic literature as phenotypic proxies for geographical ancestry. Population admixture, however, challenges the appropriateness of this approach. We compared the effectiveness of color-based and marker-based biogeographical ancestry classifications in typing polymorphisms in *GSTM1*, *GSTM3* and *GSTT1* in the heterogeneous Brazilian population.**

Methods **Individual DNA from 335 healthy Brazilians was typed for a set of insertion/deletion polymorphisms, previously validated as ancestry informative markers. *GSTM1-null* and *GSTT1-null* polymorphisms were detected by multiplex PCR and the *GSTM3*B* polymorphism by restriction-fragment length polymorphism. Nonlinear logistic regression modeling was developed to describe the association between the *GST* polymorphisms and ancestry estimated by the ancestry informative markers.**

Results **Analysis of the ancestry informative markers data with the *Structure* software revealed the existence of only two significant clusters, one of which was inferred to be an estimate of the African component of ancestry. Nonlinear logistic regression showed that the odds of having the *GSTM1-null* genotype decreases ($P<0.0004$, Wald statistics), whereas the odds of having the *GSTM3*B* allele increases ($P<0.0001$) with the increase of the African component of ancestry, throughout the range (0.13–0.95) observed in the population sample. The African component of ancestry proportion was not associated with *GSTT1-null* frequency. Within the self-reported Black and Intermediate groups, there were significant differences in ancestry informative markers between *GSTM1-null* and *non-null* individuals, and between carriers and noncarriers of the *GSTM3*B* allele.**

Conclusions **Interethnic admixture is a source of cryptic population structure that may lead to spurious genotype–phenotype associations in pharmacogenetic/-genomic studies. Logistic regression modeling of *GST* polymorphisms shows that admixture must be dealt with as a continuous variable, rather than proportioned in arbitrary subcategories for the convenience of data quantification and analysis. *Pharmacogenetics and Genomics* 17:765–771 © 2007 Lippincott Williams & Wilkins.**

Pharmacogenetics and Genomics 2007, **17**:765–771

Keywords: ancestry informative markers, Brazil, glutathione-S-transferases, nonlinear logistic analysis, population admixture, population stratification

[a]Pharmacology Division, Instituto Nacional de Câncer, [b]Program of Scientific Computation, Fundação Oswaldo Cruz, Rio de Janeiro and [c]Department of Biochemistry and Immunology, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil

Correspondence to Dr Guilherme Suarez-Kurtz, MD, PhD, Divisão de Farmacologia, Instituto Nacional de Câncer, Rua André Cavalcanti 37, Rio de Janeiro 21230-050, Brazil
Tel: +5521 3233 1310; fax: +5521 3233 1340;
e-mail: kurtz@inca.gov.br

Received 10 January 2007 **Accepted** 3 April 2007

Introduction

Continental populations of the world vary considerably in their predisposition to diseases and in the allele frequencies of pharmacogenetically important loci, probably as a result of genetic drift, and also because of adaptation to local selective factors such as climate and available nutrients. In Brazil and elsewhere, skin color has systematically been used in the scientific literature as phenotypic proxies for geographical ancestry. The extensive admixture of the Brazilian population, as a result of five centuries of interethnic crosses between people from three continents – Europeans, Africans and Amerindians – however, casts a shadow of doubt on the appropriateness of this approach. Indeed, we have previously demonstrated that at the individual level in Brazil, color, as determined by physical evaluation is a poor predictor of genomic European or African ancestry, estimated by molecular markers [1,2]. The impact of the weakness of this correlation on the frequency distribution of polymorphism in genes of pharmacological relevance among different color groups in Brazil has not been previously examined. We thus decided to compare the effectiveness of color-based classification and marker-based biogeographical ancestry classification in typing polymorphisms in the *GSTM1*, *GSTM3* and *GSTT1* genes that encode the phase 2 drug metabolizing enzymes, glutathione-S-transferase mu 1 (GSTM1), mu 3 (GSTM3) and theta 1 (GSTT1). The choice of these targets was based on various factors. First, these GSTs play a key role in cellular detoxification, protecting macromolecules from attack by reactive electrophyles, such as products of oxidative stress, environmental

pollutants and carcinogens [3]. Not surprisingly, polymorphisms in *GSTM1*, *GSTM3* and *GSTT1* have been widely investigated as risk factors for cancer [4–6]. Second, there are large differences in the frequency distribution of some common *GST* polymorphisms between Europeans and sub-Saharan Africans, two of the major ancestral roots of the Brazilian population [7]. Distinct examples are the *GSTM1-null* (homologous deletion) and the *GSTM3*B* (a 3-pb deletion in intron 6, that has been postulated to increase the expression levels of GSTM3) polymorphisms. The *GSTM1-null* genotype is twice as common in Europeans (0.42–0.60) than in sub-Saharan populations (0.20–0.33), whereas the homozygous *GSTM3*B* genotype is ca. 20 times more frequent among African Bantus (0.64) than in Europeans (< 0.05) [8–10]. In contrast, the *GSTT1-null* genotype displays similar frequencies (ca. 0.25) among Europeans [8] and sub-Saharan African Vendas and Zimbabweans [11]. Third, there is evidence that commonly used skin color or racial labels are insufficient and inaccurate representations of ancestral clusters and do not capture population substructure differences in distribution of risk genotypes in drug metabolizing genes, including *GSTM1* [12,13]. Finally, *GSTM1*, *GSTT1* and *GSTM3*, have been extensively studied as risk factors for various pathologies in the Brazilian population, with conflicting results and substantial variation in the frequency distribution of variant alleles and genotypes (reviewed in [14,15]).

Methods

Study population

The experimental protocols were approved by the Ethics Committee of the Brazilian National Cancer Institute, and written informed consent was obtained from all participating individuals. A sample of 335 randomly chosen, unrelated, healthy individuals (218 men, 117 women) from Rio de Janeiro, was analyzed in this study. Each individual answered a questionnaire about his/her ancestry and demographics. The pronounced level of admixture in the trihybrid Brazilian population poses special challenges to ethnic/racial categorization [1,7]. In this study we adopted the classification scheme used in the 2000 Brazilian Census (*http://www.ibge.gov.br/home/estatistica/populacao/censo2000/*), which relies on self-perception of skin color. Accordingly, the enrolled individuals self-identified as 'Branco' (white, $n = 107$; 32%), 'Pardo' (meaning brown, here denoted as intermediate, $n = 119$; 35%) and 'Preto' (black, $n = 109$; 33%). A single blood sample (3 ml) was drawn from each individual and DNA was extracted using the GFX Genomic Blood DNA Purification Kit (Amersham Biosciences) following the manufacturer's instructions.

Population structure analysis

Individual DNA was independently typed for a set of 40 biallelic short insertion/deletion polymorphisms (indels). This set has been previously validated as useful ancestry informative markers or AIMs [16] using the CEPH-HGDP panel composed of 1064 individuals drawn from 55 populations worldwide [17]. As population clustering algorithm we have applied the *Structure* program version 2.1 [18], available at *http://pritch.bsd.uchicago.edu/software.html*. This software uses multilocal genotypes to allocate ancestry proportions of individuals to different clusters (populations). The software defines 'K' clusters (where K is provided by the user), each of them being characterized by a set of allelic frequencies for each locus. The individuals are grouped (probabilistically) on the basis of their genotypes. Chi-square and Fisher exact probability test for 2×2 and larger tables was performed online at *http://faculty.vassar.edu/lowry/VassarStats.html*

Glutathione-S-transferase genotyping

For *GSTM1* and *GSTT1* typing, a multiplex PCR protocol, which detects the presence or the absence (*GST-null* genotypes) of gene fragments was used, following previously described procedures [19]. Genotyping for the *GSTM3*B* polymorphism was performed by restriction-fragment length polymorphism as described by Inskip *et al.* [20]. The allelic and genotypic frequencies were derived by gene counting. The statistical tests applied for analysis of the *GST* polymorphisms are indicated in the results. Significance level was set at $P < 0.05$.

Population structure and *GST* polymorphisms

To describe the association between the *null*-polymorphisms in *GSTT1* and *GSTM1*, and ancestry estimated by the AIMs indel set, we fitted a nonlinear logistic regression model using maximum likelihood estimation. Variable *GSTM3* has more than two categories that is, genotypes *GSTM3*A/*A*, **A/*B* and **B*/*B* and the binary logistic regression model cannot be applied. In this case, we fitted the proportional odds nonlinear logistic regression. The covariate ancestry entered the model transformed as a linear tail-restricted cubic spline allowing for nonlinear contributions. The result of the model fitting exercise is presented as a graph relating the log odds [i.e. $\log(p/(1-p))$, where p is the proportion of individuals with the variant marker] against the proportion of ancestry in a cluster specified by the software *Structure*, and the respective 95% confidence interval. This method, described by Harrell [21] is implemented as function 'lrm' available in the R package 'Design' [22]. Analysis of variance tables describe the Wald statistics for testing the model components [23].

Results

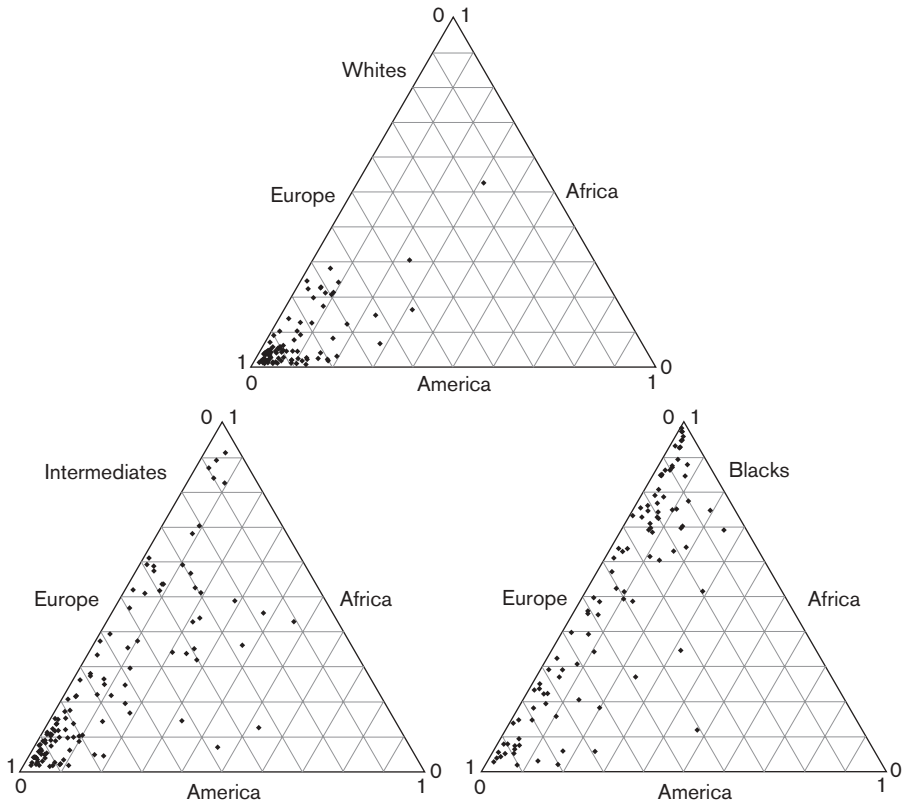
Population structure

The Brazilian population was formed from three ancestral roots: Amerindians, Europeans and Africans. Thus, in our initial analysis of the three groups of individuals recruited for this study (i.e. White, Intermediate and Black) we focused on this trihybrid nature. Each individual was

typed with the 40-indel set and the data were analyzed using the *Structure* program [18]. As ancestral population references we used our previous typing data with the CEPH-HGDP panel on 161 Europeans, 126 Africans and 103 Amerindians [16]. Using these data as a departure point, the software assigned to each of our individuals a proportion of Amerindian, European and African ancestry. The results are shown as triangular plots in Fig. 1.

The data in Table 1 show that the Amerindian contribution was relatively small and did not differ significantly among the three color groups ($\chi^2 = 0.29$, $P > 0.8$). To test whether this uniform Amerindian contribution was relevant for analysis of our data we ran the *Structure* software again, but with a different set of parameters, most specifically without using prior population information. Under these circumstances, we made *Structure* runs asking the program to separate our individuals for respectively $K = 1$, 2 or 3 clusters and then compared the data to reach an estimate of the posterior probability of each value of K. *Structure* produced a posterior probability of 1 for $K = 2$, thus indicating the existence of only two significant clusters differing in allele frequencies. The program then assigned to each individual a proportion of cluster 1 and cluster 2 ancestry. The average proportion of cluster 2 ancestry among Whites,

Fig. 1



Triangular plots of the genomic proportions of African European and Amerindian ancestry in three self-reported color groups of Brazilians. (a) Whites, (b) Intermediates and (c) Blacks. Each point represents a separate individual and the ancestral proportions can be determined by dropping a line parallel to the grid lines to each of the three axes. The graphs were drawn using the Triangular Plot program available in the R project in statistical computing, version 2.3.1, available at *http://www.r-project.org/*

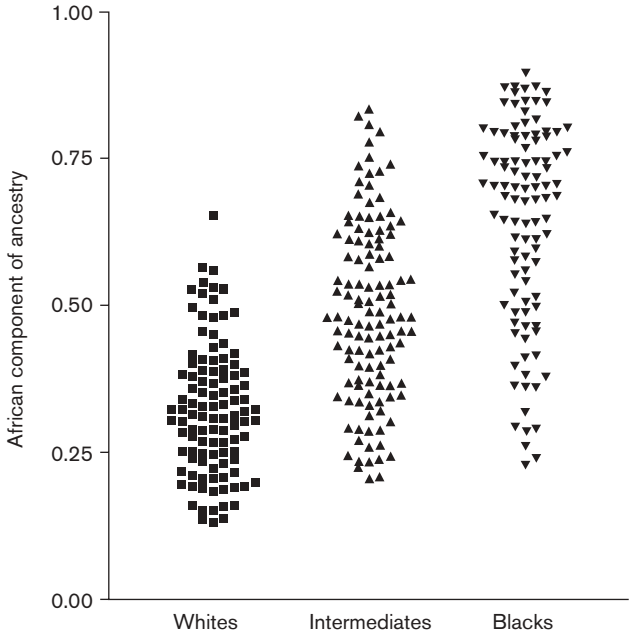
Table 1 Proportion of membership in predefined ancestral population for each color group[a]

Group	Individuals (*n*)	Proportion of membership in ancestral population		
		Amerindian	African	European
White	107	0.067 (0.054–0.069)	0.069 (0.053–0.085)	0.864 (0.840–0.888)
Intermediate	119	0.083 (0.064–0.102)	0.236 (0.192–0.279)	0.681 (0.631–0.731)
Black	109	0.073 (0.059–0.088)	0.509 (0.449–0.569)	0.418 (0.359–0.476)

[a]Data are expressed as mean (95% CI).

Intermediates and Blacks was respectively 0.326, 0.488 and 0.644. Thus, we infer that cluster 2 is an estimate of African contribution. This was confirmed when we compared the cluster 2 results with the proportion of African ancestry as measured previously using $K = 3$: there was a highly significant correlation between the two estimates ($r = 0.925$, $P < 0.0001$). We then decided to rename 'cluster 2' as the 'African component of ancestry' (ACA). It should, however, be clear that the value of ACA is a correlate of the proportion of African origin and not a direct measure of it. The individual results of ACA proportions for the self-identified White, Intermediate and Black individuals are shown in Fig. 2. It is clear that there is great overlap between the three groups, making it almost impossible to estimate color from marker data at an individual level. Yet, the pair-wise comparison of the three color groups using the Mann–Whitney test showed that they all differed significantly from each other with $P < 0.0001$.

Fig. 2



Slot plot of the proportion of ACA for 335 individuals from Rio de Janeiro, Brazil, sorted according to their self-reported color groups (White, Intermediate and Black). Each symbol indicates the proportion of ACA value for one individual, determined by the *Structure* software version 2.1 [17].

Allele and genotype distribution according to self-categorization

The distribution of the *GST* polymorphisms in the studied population is shown in Table 2. No difference existed in the *GSTT1-null* frequency across the three self-identified population groups ($P > 0.09$, χ^2 test). A significant difference was detected in relation to the *GSTM1-null* polymorphism ($P < 0.01$, χ^2 test), with a trend for decreasing frequency from White, to Intermediate to Black individuals ($P < 0.002$, χ^2 test for trend in proportions). Pairwise comparisons showed that the *GSTM1-null* frequency was lower in the Black than in either the White or the Intermediate groups ($P < 0.01$ and $P = 0.025$, respectively, χ^2 test), but did not differ between the latter two groups. Regarding the *GSTM3*B* polymorphism, the genotype frequencies in the overall population and in each group did not deviate from Hardy–Weinberg proportions ($P > 0.8$, goodness-of-fit χ^2 test). The frequency of allele **B* varied significantly across the three population groups ($P < 0.01$ χ^2 test), with a trend for increasing frequency from White to Intermediate to Black individuals ($P < 0.002$, χ^2 test for trend in proportions). In pairwise comparisons, the **B* allele frequency was higher in the Black than in either the White or the Intermediate groups ($P < 0.001$ and < 0.025, respectively, χ^2 test), but did not differ between the latter two groups. The *GSTM3* genotype distribution varied significantly across the three subgroups ($P < 0.001$, χ^2 test), and between Blacks and either Whites or Intermediate individuals ($P < 0.001$ and < 0.01, respectively, χ^2 test). No difference existed in the distribution on the *GSTM3* genotypes between the White and the Intermediate groups.

Distribution of *GST* polymorphisms according to ancestry

We developed a nonlinear logistic regression model to explore the association of polymorphisms in *GSTT1*, *GSTM1* and *GSTM3* with the estimated individual ancestry in the African component. The results are presented graphically in Fig. 3. ACA associates significantly

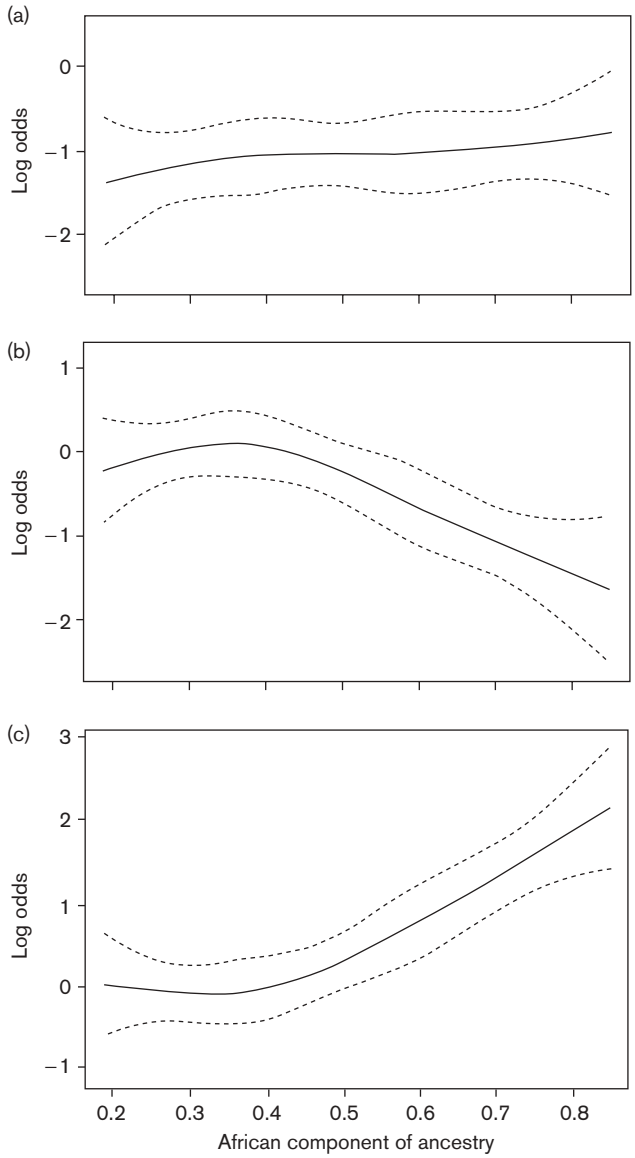
Table 2 Allele and genotype frequencies of *GSTM1*, *GSTT1* and *GSTM3* in healthy Brazilians[a]

Group (*n*)	*GSTM1-null*	*GSTT1-null*	*GSTM3 genotype*			*GSTM3 allele*	
			**A/*A*	**A/*B*	**B/*B*	**A*	**B*
White (107)	0.48 (0.38–0.57)	0.26 (0.17–0.35)	0.51 (0.41–0.61)	0.43 (0.33–0.52)	0.07 (0.03–0.13)	0.72 (0.62–0.80)	0.28 (0.20–0.38)
Intermediate (119)	0.44 (0.35–0.53)	0.24 (0.17–0.33)	0.40 (0.31–0.49)	0.48 (0.39–0.58)	0.12 (0.06–0.19)	0.64 (0.54–0.72)	0.36 (0.27–0.45)
Black (109)	0.28 (0.19–0.37)	0.27 (0.18–0.36)	0.25 (0.17–0.33)	0.47 (0.37–0.56)	0.28 (0.20–0.38)	0.48 (0.38–0.57)	0.52 (0.42–0.61)
P[b]	<0.01	0.15		<0.001		<0.01	

[a]Data are expressed as mean (95% CI).
[b]*P* represents the χ^2 test for comparisons across the three groups.

with the polymorphisms in *GSTM1* and *GSTM3*, but not in *GSTT1*, in our Brazilian population sample. Specifically, the odds of having the *GSTM1-null* genotype decreases ($P < 0.0004$, Wald statistics), whereas the odds of having the *GSTM3*B* allele increases ($P < 0.0001$ Wald statistics) with the increase of the ACA. Inspection of the respective graphs (Fig. 3b and c) discloses an apparent threshold of ACA, at approximately 0.4, above which a monotonic correlation is observed between polymorphism frequency and African ancestry. It is noteworthy that 61% of the overall population sample had an estimated ACA > 0.4, corresponding to 8.8% of the self-identified White, 70.3% of Intermediate and 87.5% of Black individuals.

Fig. 3



Fitted logistic models showing the logit proportions describing the association between ancestry and *GST* polymorphisms (a) *GSTT1-null*, (b) *GSTM1-null*, (c) *GSTM3*B*.

We next examined whether the distribution of polymorphisms in *GSTM1*, *GSTM3* and *GSTT1* within each group was affected by the individual proportion of ACA. The results are shown in Table 3. Within the Black and the Intermediate, but not the White group, *GSTM1-null* individuals have significantly lower ACA than *GSTM1 non-null* individuals in the same group. By contrast, Black and Intermediate, but not White carriers of the *GSTM3*B* allele (**B/*B* and **A/*B* genotypes) have significantly greater ACA, than homozygous wild-type *GSTM3*A/*A*. Regarding the *GSTM1* polymorphism, no significant ancestral differences were disclosed between individuals genotyped as *null* or *non-null* within each color group.

Discussion

We used the *Structure* program with prior population structure to determine European, African and Amerindian ancestry proportions in 335 Brazilians living in Rio de Janeiro. These individuals had self-classified according to color as White (107), Intermediate (119) and Black (109). When we compared the mean proportions (Table 1) it stood out that this group of White Brazilian individuals had a European ancestry (86.4%), higher than we observed in our previous studies [1,2]. We could not identify any bias in sampling that might have led to this tendency. One possibility is that this is owing to peculiarities of the population structure of Rio de Janeiro, a large metropolis. The proportion of European ancestry in the Black individuals, on the other hand was perfectly consistent with our previous studies [1].

We also observed that the proportion of Amerindian ancestry was fairly uniform across the three groups. We then repeated the *Structure* analysis without prior population structure and the program identified that the individuals could be best explained by admixture of

Table 3 Proportion of the African component of ancestry in each color group according to genotype[a]

Group	*GSTM1*			*GSTM3*			*GSTT1*		
	Non-null	Null	P^b	*A/*B + *B/*B	*A/*A	P^b	Non-null	Null	P^b
White	0.35 (0.12)	0.32 (0.17)	0.60	0.33 (0.12)	0.30 (0.10)	0.50	0.32 (0.11)	0.34 (0.13)	0.28
Intermediate	0.51 (0.16)	0.45 (0.14)	0.04	0.52 (0.16)	0.45 (0.13)	0.02	0.51 (0.15)	0.48 (0.16)	0.19
Black	0.67 (0.17)	0.58 (0.17)	0.02	0.67 (0.16)	0.55 (0.19)	0.002	0.64 (0.17)	0.65 (0.18)	0.73

[a]Data are expressed as mean (SD).
[b]*P* values for Student's *t*-test.

two clusters (1 and 2). Cluster 2 frequencies correlated very well ($r = 0.925$) with the proportion of African ancestry determined in the initial analysis. This observation has several important implications. First, it reduces the individual admixture proportions from three to two components and this considerably simplifies and increases the power of statistical analyses. Second, it demonstrates that the software *Structure* can be successfully used for ancestry analysis without any previous information about allele frequencies in parental populations. This is of great importance, as the true allele frequencies of historical ancestral groups can never be known with certainty. On the other hand, this approach has a problem. As the Bayesian *K*-means algorithm used by the program concentrates in ascertaining clusters that best explain differences between the samples, it will miss ancestral contributions that are more or less uniformly distributed, such as the Amerindian one in this case.

When we plotted the individual proportions of the ACA in the three color groups, we observed large variation and significant overlap. This is consistent with our previous observations that in Brazil, at the individual level, color, as determined by physical evaluation, is a poor predictor of genomic European or African ancestry, estimated by molecular markers [1,2]. Nevertheless, the proportion of ACA differed significantly across the three color groups of this study. Significant differences across these groups were also observed in relation to the distribution of the *GSTM1-null* and the *GSTM3*B*, but not the *GSTT1-null* polymorphisms. We detected significant trends for increasing frequency of the *GSTM3*B* allele and for decreasing frequency of the *GSTM1-null* genotype from White, to Intermediate and to Black Brazilians, whereas the frequency of the *GSTT1-null* genotype was unaffected by the population structure. These data are consistent with previous studies in the Brazilian population [15]. Our study, however, shows for the first time a significant within-group heterogeneity in the frequency distributions of *GSTM1-null* and the *GSTM3*B* polymorphisms in Brazilians: within the Black and the Intermediate color groups, the proportion of African ancestry was significantly higher in carriers than in noncarriers of the *GSTM3*B* allele, whereas this proportion was significantly lower in *GSTM1-null* as compared with *GSTM1-non-null* individuals. These results are compatible with the notion that the frequency of *GSTM1-null*, and *GSTM3*B* may vary by ancestral root and that self-reported race may be an insufficient and inaccurate representation of the ancestral clusters [12,13].

To represent and analyze the influence of ancestral cluster on the distribution of *GST* polymorphisms in the Brazilian population we introduced a novel approach, based on nonlinear logistic regression models. This approach disclosed significant associations of the *GSTM1-null* and the *GSTM3*B* polymorphisms with the proportion in African ancestry, throughout the range of ACAs observed in the study population (0.13–0.95). Importantly, a threshold proportion of the African component of ancestry was detected at approximately 0.4, above which the frequency of the polymorphisms shows a monotonic dependence on the ACA. This range encompasses 61% of the overall population sample, representing 87.5% of Blacks, 70.3% of Intermediates but only 8.8% of self-identified Whites. By contrast to the *GSTM1-null* and *GSTM3*B* polymorphisms, the nonlinear logistic regression model approach revealed no influence of ACA on the frequency distribution of the *GSTT1-null* polymorphism. This result is consistent with the relatively constant frequency of the *GSTT1-null* polymorphism among Europeans and sub-Saharan Africans (see Introduction), who represent two major ancestral roots of the Brazilian population.

Population admixture is a likely source of cryptic population structure that may not be recognized or acknowledged by investigators, and thus lead to spurious genotype–phenotype associations in pharmacogenetic/-genomic studies [23,24]. Our logistic regression analysis revealed that the variable frequency of *GSTM1-null* and *GSTM3*B* polymorphisms in the highly admixed Brazilian population is best fit by continuous functions of the proportion of African ancestry, across three self-reported color categories acknowledged by the Brazilian Census, namely White, Intermediate and Black individuals. Of noteworthy, linear relationships were observed for the range of estimated African ancestry (0.4–0.95) that encompasses the majority (> 70%) of non-White Brazilians. The relevance of these findings to the design and interpretation of genotype/phenotype association studies of *GSTM1-null* and *GSTM3*B* polymorphisms in Brazilians – and most likely other admixed populations is evident. Admixture must be dealt with as a continuous variable, rather than proportioned in arbitrary subcategories for the convenience of data quantification and analysis.

Acknowledgement

This research was supported by grants from Conselho Nacional de Desenvolvimento Científico e Tecnológico, Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro and Swiss Bridge Foundation.

References

1 Parra FC, Amado RC, Lambertucci JR, Rocha J, Antunes CM, Pena SDJ. Color and genomic ancestry in Brazilians. *Proc Natl Acad Sci U S A* 2003; **100**:177–182.

2 Pimenta JR, Zuccherato LW, Debes AA, Maselli L, Soares RP, Moura-Neto RS, *et al.* Color and ancestry in Brazilians: a study with forensic microsatellites. *Hum Hered* 2006; **62**:190–195.

3 Hayes JD, Flanagan JU, Jowsey IR. Glutathione transferases. *Annu Rev Pharmacol Toxicol* 2005; **45**:51–88.

4 Dalhoff K, Buus Jensen K, Enghusen Poulsen H. Cancer and molecular biomarkers of phase 2. *Methods Enzymol* 2005; **400**:618–627.

5 Reszka E, Wasowicz W, Gromadzinska J. Genetic polymorphism of xenobiotic metabolising enzymes, diet and cancer susceptibility. *Br J Nutr* 2006; **96**:609–619.
6 McIlwain CC, Townsend DM, Tew KD. Glutathione S-transferase polymorphisms: cancer incidence and therapy. *Oncogene* 2006; **25**:1639–1648.
7 Salzano FM, Bortolini MC. *The evolution and genetics of Latin American populations*. Cambridge, UK: Cambridge University Press; 2002.
8 Garte S, Gaspari L, Alexandrie AK, Ambrosone C, Autrup H, Autrup JL, *et al*. Metabolic gene polymorphism frequencies in control populations. *Cancer Epidemiol Biomarkers Prev* 2001; **10**:1239–1248.
9 Tetlow N, Robinson A, Mantle T, Board P. Polymorphism of human mu class glutathione transferases. *Pharmacogenetics* 2004; **14**:359–68.
10 Ye Z, Song H, Higgins JP, Pharoah P, Danesh J. Five glutathione s-transferase gene variants in 23 452 cases of lung cancer and 30 397 controls: meta-analysis of 130 studies. *PLoS Med* 2006; **3**:e91. Epub 2006 Mar 7.
11 Adams CH, Werely CJ, Victor TC, Hoal EG, Rossouw G, van Helden PD. Allele frequencies for glutathione S-transferase and N-acetyltransferase 2 differ in African population groups and may be associated with oesophageal cancer or tuberculosis incidence. *Clin Chem Lab Med* 2003; **41**:600–605.
12 Wilson JF, Weale ME, Smith AC, Gratrix F, Fletcher B, Thomas MG, *et al*. Population genetic structure of variable drug response. *Nat Genet* 2001; **29**:265–269.
13 Barnholtz-Sloan JS, Chakraborty R, Sellers TA, Schwartz AG. Examining population stratification via individual ancestry versus self-reported race. *Cancer Epidemiol Biomarkers Prev* 2005; **14**:1545–1551.
14 Filho MB, Albano RM, Rossini A, Pinto LFR. Pharmacogenomics of xenobiotic metabolizing enzymes in South American populations. *Curr Pharmacogenomics* 2006; **4**:9–18.
15 Suarez-Kurtz G, Pena SDJ. Pharmacogenetics in the Brazilian population. In: Suarez-Kurtz G, editor. Pharmacogenomics in admixed populations. Austin: Landes Bioscience, 2006. Available at: http://www.eurekah.com/chapter/3157 [Accessed 18 December 2006].
16 Bastos-Rodrigues L, Pimenta JR, Pena SDJ. The genetic structure of human populations studied through short insertion-deletion polymorphisms. *Ann Human Genet* 2006; **70**:658–665.
17 Cann HM, de Toma C, Cazes L, Legrand MF, Morel V, Piouffre L, *et al*. A human genome diversity cell line panel. *Science* 2002; **296**: 261–262.
18 Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics* 2000; **155**:945–959.
19 Gattas GJ, Kato M, Soares-Vieira JA, Siraque MS, Kohler P, Gomes L, *et al*. Ethnicity and glutathione S-transferase (GSTM1/GSTT1) polymorphisms in a Brazilian population. *Braz J Med Biol Res* 2004; **37**:451–458.
20 Inskip A, Elexperu-Camiruaga J, Buxton N, Dias PS, MacIntosh J, Campbell D, *et al*. Identification of polymorphism at the glutathione S-transferase, GSTM3 locus: evidence for linkage with GSTM1*A. *Biochem J* 1995; **312**:713–716.
21 Harrell FE. *Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis*. New York: Springer; 2001.
22 Development Core Team. R. *A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing; 2005.
23 Suarez-Kurtz G. Pharmacogenetics in admixed populations. *Trends Pharmacol Sci* 2005; **26**:196–201.
24 Suarez-Kurtz G, Pena SDJ. Pharmacogenomics in the Americas: Impact of genetic admixture. *Curr Drug Targets* 2006; **7**:1649–1658.