Human Heredity, no prelo

Color and genomic ancestry in Brazilians: a study with forensic microsatellites

Juliana R. Pimenta[1], Luciana W. Zuccherato[2], Adriana A. Debes[3], Luciana Maselli[3], Rosângela P. Soares[3], Rodrigo S. Moura-Neto[4,] Jorge Rocha[5], Sergio P. Bydlowski[3,6] and Sérgio D. J. Pena[1,2]

(1) GENE – Núcleo de Genética Médica, 30130-909 Belo Horizonte, Brazil, (2) Departamento de Bioquímica e Imunologia, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil, (3) Divisão de Pesquisa e Biologia Molecular, Fundação Pró-Sangue Hemocentro de São Paulo, São Paulo, SP, Brazil, (4) Instituto de Biologia, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, Brazil, (5) Instituto de Patologia e Imunologia Molecular da Universidade do Porto, Porto, Portugal, (6) Departamento de Hematologia, LIM-31, Faculdade de Medicina da Universidade de São Paulo, São Paulo, SP, Brazil.

Running title: Color and genomic ancestry in Brazilians

Corresponding author: Sérgio D. J. Pena - Departamento de Bioquímica e Imunologia, ICB, UFMG Av. Antônio Carlos, 6627, Caixa Postal 486, Belo Horizonte, MG Brazil. Tel: +55-31-32848000; fax: +55-31-32273792, E-mail address: spena@dcc.ufmg.br.

Keywords: Color, admixture, microsatellites, Brazilians, genetic markers, genome

Abstract

The population of Brazil, formed by extensive admixture between Amerindians, Europeans and Africans, is one of the most variable in the world. We have recently published a study that used ancestry-informative markers to conclude that in Brazil, at an individual level, color, as determined by physical evaluation, was a poor predictor of genomic ancestry, estimated by molecular markers. To corroborate these findings we undertook the present investigation based on data from 12 commercially available forensic microsatellites that were utilized to estimate the personal genomic origin for each of 752 individuals from the city of São Paulo, belonging to different Brazilian color categories (275 Whites, 192 Intermediates and 285 Blacks). The genotypes permitted the calculation of a personal likelihood-ratio estimator of African or European ancestry. Although the 12 markers set proved capable of discriminating between European and African individuals, we observed very significant overlaps among the three color categories of Brazilians. This was confirmed quantitatively using a Bayesian analysis of population structure that did not demonstrate significant genetic differentiation between the three color groups. These results corroborate and validate our previous conclusions using ancestry-informative markers that in Brazil at the individual level there is significant dissociation of color and genomic ancestry.

Introduction

Brazilians are one of the most heterogeneous populations in the world, the result of 5 centuries of interethnic crosses between peoples from three continents: Europeans, Africans and Amerindians. When the Portuguese arrived in 1500, there were approximately 2.5 million indigenous people living in the area of what is now Brazil [1]. The Portuguese-Amerindian admixture started soon after the arrival of the first colonizers and later became commonplace, being after 1755 even encouraged as a strategy for population growth and colonial occupation of the country [2]. From the middle of the 16th century, Africans were brought to Brazil to work on sugarcane farms and, later, in the gold and diamond mines and on coffee plantations. Historical records suggest that between 1551 and 1850 (when the slave trade was abolished), around 3.5 million Africans were brought to Brazil [3]. As to the European immigration, it is estimated that about 500,000 Portuguese arrived in the country between 1500 and 1808 [1]. From then on, after the Brazilian ports were legally opened to all friendly nations, Brazil received approximately 4 million other immigrants from several parts of the world. Portugal remained by far the most important source, followed by Italy, Spain, and Germany.

In 2003 we published a study in which the objective was to ascertain to what extent the physical appearance of a Brazilian individual was predictive of the degree of genomic African or European ancestry [4]. We used a panel of 10 ancestry-informative markers (AIMs), *i.e.* genetic polymorphisms that display large differences in allelic frequencies (> 0.40) between Europeans and Africans [5] to estimate, on an individual level, the ancestry of 173 Brazilians from a Southeastern rural community. When we compared these molecular results with the individual color classification of the same subjects, we observed that the correlation was very poor. In other words, at the individual level, color, as determined by physical evaluation, was a poor predictor of genomic European or African ancestry, estimated by molecular markers. We confirmed these results with 200 unrelated Brazilian white males who originated from cosmopolitan centers of the four major geographic regions of the country. There were, however, questions raised about the size and nature of the populations studied. We thus felt that there was the need to confirm the findings of the previous study.

Bydlowski et al. [7] published the allelic frequencies for 12 microsatellites commercially available for forensic studies (F13A01, F13B, FESFPS, LPL, CSF1PO, TPOX, TH01, vWA, D16S539, D7S820, D13S317 and D5S818) in a sample of 916 unrelated Brazilian subjects classified by joint phenotypic and genealogical criteria into four groups that they called European-derived (Whites), African-derived (Blacks), Brazilian Mulattos (Intermediates) and Asian-derived (Orientals). Knowing that others had already been successful with the use of such forensic microsatellites in the inference of human biogeographical ancestry [8-11], we felt that the data of Bydlowski et al [7] might provide us with the opportunity to verify, in a much larger sample, our results described above. We here report that the genotyping of the 12 forensic microsatellite loci did not show statistically significant degrees of genetic differentiation between White, Intermediate and Black Brazilian individuals, thus corroborating and validating our previous results [4].

Material and Methods

Populations Studied

Seven hundred and fifty two unrelated healthy volunteer blood donors of the city of São Paulo were studied after a written informed consent, as previously described by Bydlowski et al [7]. All individuals were classified by joint phenotypic and genealogical criteria as follows: subjects were asked about their color group and those of their parents and grand-parents, according to their own definition. Phenotype analysis (facial characteristic and skin pigmentation in the axilla, a body region not exposed to the sun) was performed by the interviewer. Subjects were then classified accordingly into three groups: Whites, Intermediates and Blacks. For instance, subjects were classified as Black when the characteristic phenotype was present and subjects described themselves, their parents and all their grand-parents as Blacks. We feel that with this methodology, color classification became as free as possible from subjective biases. The same procedure was followed for the classification of individuals into the other groups. In this study we only analyzed data of 752 individuals classified as Whites (275), Intermediates (192) and Blacks (285), which are the three main Brazilian color groups.

Additionally, we typed with the same 12 microsatellites two samples of populations considered representative of major ancestral groups of Brazil. The first included 20 individuals randomly drawn from a previously described [12] sample of 93 unrelated Portuguese men from the Porto District in Northern Portugal (41.11 N; 8.36 W). The second, representing Africa, comprised 20 individuals from the village of Santana in São Tomé Island in the West coast of Africa (1.00 N; 7.00 E). The population-based estimate of European genetic contribution to the Santana population was only 0.074 ± 0.015 [6].

DNA analysis

PCR amplification and STR genotyping were performed using multiplex systems kindly provided by Promega Corporation (Madison, WI, USA) and Dialab Diagnósticos (Belo Horizonte, Brazil): CTTv Multiplex (CSF1PO, TPOX, TH01 and vWA), Gamma STR Multiplex (D16S539, D7S820, D13S317 and D5S818) and FFFL Multiplex (F13A01, F13B, FESFPS and LPL). Amplified products were detected in silver nitrate stained 7 M urea–polyacrylamide denaturing gels.

Based on these results we assigned to each subject an individual "African Ancestry Index" (IAA) that was calculated as the logarithm of the ratio of the likelihood of a given multilocus genotype occurring in the African population to the likelihood of it occurring in the European population. Thus, the IAA represents a personal geographical ancestry estimate [4]. For the likelihood ratio calculations we pooled the African and European allele frequencies of each locus available for each at the DNA PCR Database of the Institut für Rechtsmedizin of the University of Düsseldorf (www.uni-duesseldorf.de/WWW/MedFak/Serology/database.html). The table is available in Supplemental data.

Data Analysis

Allele frequencies were estimated by the gene counting method. Fst values were calculated as described by Weir [13]. To estimate the distance between probability distributions we used the Kullback-Leibler divergence, which was calculated according to the original formulation [14]. Bayesian analysis of genetic differentiation between the color groups was performed as described by Corander et al. [15] using their *BAPS* software. To calculate admixture at an individual level we have also used the *Structure* program version 2.1 [16].

Results

Discrimination between Europeans and Africans

We first examined the discrimination power of the 12 microsatellite loci using samples of 20 males from Northern Portugal and 20 males from São Tomé Island in the Gulf of Guinea, West coast of Africa. These population sources were chosen because they are geographically related to the European and African population groups that participated in the peopling of Brazil. The box plot obtained from the data is shown in Fig. 1. For the 12 microsatellites there was good discrimination - five logs separated the two medians (2.35 and -2.65, respectively), but a small overlap occurred. We then compared these results with those utilizing a 10-allele set of ancestry-informative markers that had been previously typed in the same individuals [4]. With the latter, discrimination was much better, with no overlap between the two groups and more than 20 logs separating the two medians (9.75 and -11.93, respectively). To quantify the difference between the two analytical systems, we applied the Kullback-Leibler divergence (K-L distance), which is a measure of the distance between two probability distributions [14]. The K-L distance between the individuals from Portugal and São Tomé was 11.88 when measured with the AIMs, but only 1.29 for the microsatellites. We conclude that the set of 12 forensic microsatellite loci is useful in discriminating between Europeans and Africans, although with a much smaller efficiency (~10%) than AIMs.

White, Intermediate and Black Brazilians

Using the genotypes of the 12 forensic microsatellites we calculated the Index of African Ancestry (IAA) for 752 individuals from São Paulo initially described by Bydlowski et al. [7]. Using a methodology that minimized subjectivity by combining self-assessment, family history and phenotypical observation, the subjects were classified by color in three groups: 275 Whites, 192 Intermediates and 285 Blacks. When we compared the IAA values for these individuals, we observed that the groups had much wider ranges than those of Europeans and Africans (Fig. 2) and that there was very significant overlap between them. The medians were -1.52 for Whites, -0.54 for intermediates and -0.07 for Blacks. As previously [4] we additionally used the software *Structure* [16] to calculate the inferred proportion of African ancestry of each individual in our sample. There was a highly significant correlation between the two admixture estimates (data nor shown).

Level of genetic structure among the different color groups

All loci of the 12-STR set studied in Brazilian Whites, Intermediates and Blacks had been previously shown by Bydlowski et al [7] to be in Hardy-Weinberg equilibrium. Thus, we could calculate the degree of genetic differentiation between the three groups using the Fst statistic. Because microsatellites evolve by stepwise mutations rather than according to an infinite allele model [17], special measures of genetic distance have been proposed for them (e.g., Rst; 18). However, other authors [19, 20] have shown that the Fst statistic, which does not consider different mutational relationships among alleles and has a known relationship to differentiation by drift, actually reflects reality better than a mutation-based distance such as Rst. This occurs presumably because genetic drift has played the main role in generating the present distributions of microsatellite alleles and their variation among human populations; the role of mutation must have been less important owing to the time constraint imposed by the small timescale in which most human differentiation has occurred [19]. Our results showed a very small amount of genetic differentiation as measured by the Fst values of -0.008 for White-Black comparison, 0.009 for White-Intermediate and -0.004 for Black-Intermediate, none of them different from zero at the 0.05 level of significance.

This absence of significant genetic differentiation was confirmed by analysis of our data with the program Bayesian Analysis of Population Structure [15]. A major advantage of this software is that the number of distinct populations is treated as an unknown parameter. That means that if the program perceives that two populations, because of high degree of gene flow or recent foundation, can be considered a single panmitic population it will then combine them [15]. We entered in BAPS the allele distributions of Whites, Intermediates and Blacks from São Paulo at al 12 loci and also the African and European allele frequencies of each locus available at the DNA PCR Database of the Institut für Rechtsmedizin of the University of Düsseldorf (www.uni-duesseldorf.de/WWW/MedFak/Serology/ database.html). The program identified with a probability of 1 the presence of three clusters, as follows, cluster 1 – Europeans, cluster 2 = Africans and cluster 3 = Whites, Intermediates and Blacks from São Paulo together. The Fst values for differentiation of the clusters were 0.042 for Europeans versus Africans, 0.009 for Europeans versus Brazilians and 0.021 for Africans versus Brazilians. When we ran the Bayesian analysis (BAPS program) on the three Brazilian groups only, a single cluster was identified with probability 1.

Discussion

Microsatellites are highly polymorphic due to variation in the number of repeating units between alleles and this is believed to be primarily the result of frequent strand slippage during replication [17]. The high mutation rate of microsatellites [21] and the stepwise nature of their mutation process lead to frequent homoplasy in population studies. Indeed, if one takes into account the possibility of size constraints for their growth, different populations would tend to approach a common allelic distribution for these markers [22]. However, microsatellites can be useful in studies of human evolution and the genetic structure of human populations [23] especially if large numbers of loci are examined [24, 25].

Because of their high informativity, microsatellites have become markers of choice for forensic studies. Due to a series of factors that include considerations of intellectual property, commercial interests, convenience and the need for validation and uniformity, a relatively small number of microsatellites have been elected for use by the forensic community. An example of this is the CODIS (Combined DNA Index System) criminal database in the United States, for which 13 commercially available microsatellite loci were selected by the FBI, as reviewed by Budowle and Moretti [26]. The choice of few “forensic microsatellites” by the crime and paternity testing community has led to a wealth of information on their allele frequencies in populations all over the globe as available in different databases (for instance, the DNA PCR Database, www.uni-duesseldorf.de/WWW/MedFak/Serology/database.html and the Short Tandem Repeat DNA Internet DataBase, http://www.cstl.nist.gov/biotech/strbase/). These data have been used to show that forensic microsatellites can useful in population genetics and in the estimation of the probable biogeographical origin of a given DNA profile [e.g. 8-11, 27]. The realization that allele frequencies of forensic microsatellites are sensitive to geographical origin have led to the common practice of having different databases for "Caucasians", "Blacks", "Orientals" and "Hispanics" in the United States and Europe [28].

In the present study we used 12 commercially available forensic microsatellites to estimate the personal genomic origin for each of 752 individuals belonging to different Brazilian color categories, encompassing 275 Whites, 192 Intermediates and 285 Blacks (N=752). Our work differs from past admixture studies of Brazilians because our focus was the individual, and not the population as a whole. Since color is the main criterion used for racial categorization and prejudice in society, we wanted to ascertain to what degree it was correlated with genomic ancestry, and this could only be accomplished at a personal level. Each individual was first classified as White, Black or Intermediate, on the basis of both self-classification, family data and a multivariate phenotypic evaluation. The later is a fundamental component of the evaluation, since, as mentioned above, racial discrimination is mostly exerted on the basis of phenotype. The second component of the study was the calculation of a personal likelihood-ratio estimator of African or European ancestry [summarized in the individual Index of African Ancestry (IAA)] based on genotypes at 12 forensic microsatellite loci. We observed large variances and very significant overlaps among the three color categories (Fig. 2). In other words, at the individual level, it was not possible to obtain a reliable color classification on the basis of the genomic analysis of the 12 microsatellite loci. This was confirmed quantitatively using a Bayesian analysis of population structure [15] which did not demonstrate genetic differentiation between the three color groups. In other words, according to the analysis of the BAPS program the three color groups were considered a single panmitic population. These results corroborate and validate our previous study using ancestry-informative markers [4] that concluded that in Brazil, at an individual level, color, as determined by physical evaluation, was a poor predictor of genomic ancestry, estimated by molecular markers. An important corollary to this conclusion is that in forensic practice in Brazil one can dispense with the common American and European practice of having different microsatellite databases for "Caucasians" and "Blacks".

In conclusion, our two studies are concordant and illuminate the hazards of trying to equate color or "race" with geographical ancestry and of using interchangeably terms such as White, Caucasian and European in one hand, and Black, Negro or African in the other.

Acknowledgements

This work was supported by grants from the Brazilian agency CNPq and the Portuguese agency FCT (POCTI/42510/ANT/2001). Neuza A. Rodrigues and Kátia Barroso provided expert technical assistance. We are grateful to Promega Corporation and Dialab Diagnósticos for providing the kits used for the microsatellite analysis.

References

[1] Ribeiro D: O Povo Brasileiro: Formação e Sentido do Brasil. São Paulo, Companhia das Letras, 1995

[2] Mörner, M.: Race Mixture in the History of Latin America. Boston, Little, Brown and Company, 1967.

[3] Klein, HS: As origens africanas dos escravos brasileiros; in Pena SDJ (ed): Homo Brasilis, Ribeirão Preto, FUNPECRP, pp 93-112, 2002.

[4] 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.

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

[6] Tomas G, Seco L, Seixas S, Faustino P, Lavinha J, Rocha J: The peopling of Sao Tome: Gulf of Guinea: origins of slave settlers and admixture with the Portuguese. Hum Biol 2002;74:397-411.

[7] Bydlowski SP, de Moura-Neto RS, Soares RP, Silva R, Debes-Bravo AA, Morganti L: Genetic data on 12 STRs: F13A01, F13B, FESFPS, LPL, CSF1PO, TPOX, TH01, vWA, D16S539, D7S820, D13S317, D5S818 from four ethnic groups of Sao Paulo, Brazil. Forensic Sci Int 2003;135:67-71.

[8] Lowe AL, Urquhart A, Foreman LA, Evett IW: Inferring ethnic origin by means of an STR profile. Forensic Sci Int 2001;119:17-22.

[9] Sun G, McGarvey ST, Bayoumi R, Mulligan CJ, Barrantes R, Raskin S, Zhong Y, Akey J, Chakraborty R, Deka R.: Global genetic variation at nine short tandem repeat loci and implications on forensic genetics. Eur J Hum Genet 2003;11:39-49.

[10] Rowold DJ, Herrera RJ.: Inferring recent human phylogenies using forensic STR technology. Forensic Sci Int 2003;133:260-265.

[11] Agrawal S, Khan F: Reconstructing recent human phylogenies with forensic STR loci: a statistical approach. 2005;BMC Genet. 28:47.

[12] Carvalho-Silva DR, Santos FR, Rocha J, Pena SD: The phylogeography of Brazilian Y-chromosome lineages. Am J Hum Genet 2001;68:281-286.

[13] Weir BS: Genetic Data Analysis. Sunderland, Sinauer Associates, 1990.

[14] Kullback S, Leibler RA: On information and sufficiency. Ann Math Stat 1951;22:79-86.

[15] Corander J, Waldmann P, Sillanpaa MJ: Bayesian analysis of genetic differentiation between populations. Genetics 2003;163:367-374.

[16] Pritchard JK, Stephens M, Donnelly P: Inference of population structure using multilocus genotype data. Genetics 2000;155:945–959.

[17] Levison G, Gutman GA: Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol Biol Evol 1987;4:202–221

[18] Slatkin M.: A measure of population subdivision based on microsatellite allele frequencies. Genetics 1995;139:457-462.

[19] Perez-Lezaun A, Calafell F, Mateu E, Comas D, Ruiz-Pacheco R, Bertranpetit J: Microsatellite variation and the differentiation of modern humans. Hum Genet. 1997;99:1-7.

[20] Caglia A, Tofanelli S, Coia V, Boschi I, Pescarmona M, Spedini G, Pascali V, Paoli G, Destro-Bisol G: A study of Y-chromosome microsatellite variation in sub-Saharan Africa: a comparison between F_{ST} and R_{ST} genetic distances. Hum Biol 2003;75:313-330.

[21] Leopoldino AM, Pena SDJ: The mutational spectrum of human autosomal tetranucleotide microsatellites. Hum Mutat 2003;21: 71-79.

[22] Romualdi C, Balding D, Nasidze IS, Risch G, Robichaux M, Sherry ST, Stoneking M, Batzer MA, Barbujani G: Patterns of human diversity, within and among continents, inferred from biallelic DNA polymorphisms. Genome Res. 2002; 12:602-612.

[23] Bowcock AM, Ruiz-Linares A, Tomfohrde J, Minch E, Kidd JR, Cavalli-Sforza LL. : High resolution of human evolutionary trees with polymorphic microsatellites. 1994;Nature 31:455-457.

[24] Jorde LB, Rogers AR, Bamshad M, Watkins WS, Krakowiak P, Sung S, Kere J, Harpending HC: Microsatellite diversity and the demographic history of modern humans. Proc Natl Acad Sci U S A 1997;94:3100-3103.

[25] Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, Feldman MW: Genetic structure of human populations. Science 2002;298:2381–2385.

[26] Budowle, B., Moretti TR: Forensic analysis of short tandem repeat loci by multiplex PCR and real-time fluorescent detection during capillary

electrophoresis; in Epplen JT, Lubjuhn T (eds): DNA Profiling and DNA Fingerprinting. Basel, Birhäuser, pp 101-115, 1999.

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

[28] Butler JM. Forensic DNA Typing, ed 2. Amsterdam, Elsevier, 2005.

[29] Cavalli-Sforza L., Menozzi P, Piazza A: The History and Geography of Human Genes. Princeton, Princeton University Press, 1994.

[30] Sokal RR, Rohlf FJ: Biometry, ed 3, New York, W.H. Freeman, 1995.

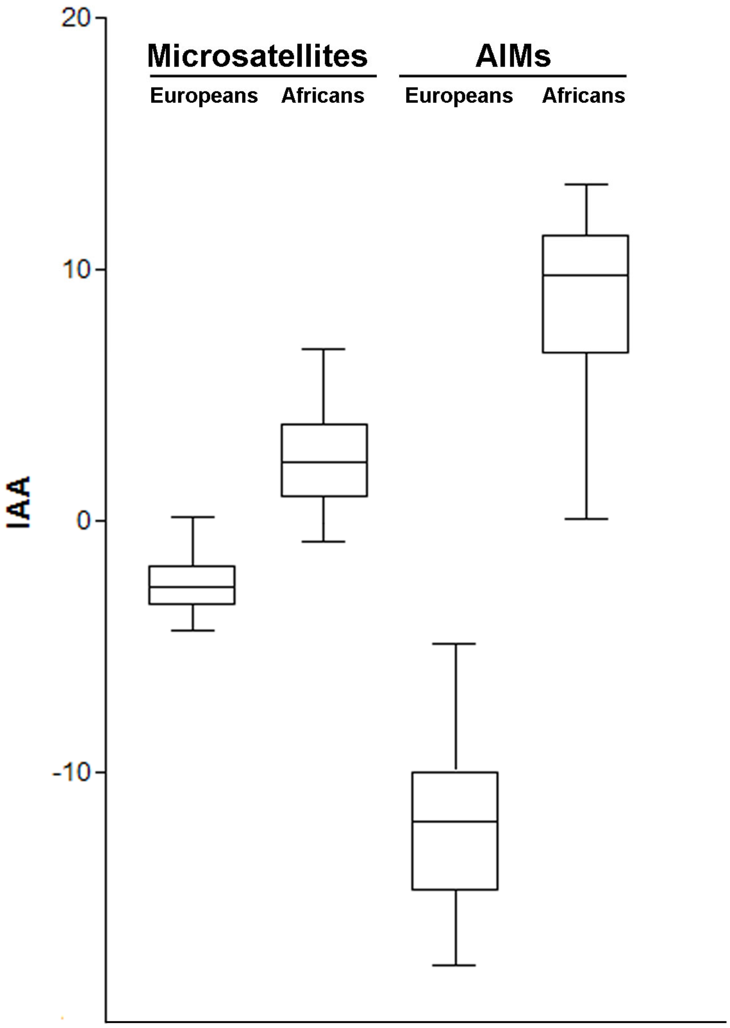

Figure 1 – Box plot showing the distribution of values of the Index of African Ancestry (IAA) calculated using forensic microsatellites or ancestry-informative markers (AIMs) in samples of the Northern Portuguese population (Europeans; n = 20) and Africans from São Tomé Island (Africans; n = 20). Each group is represented as a box whose top and bottom are drawn at the lower and upper quartiles, with a horizontal line at the median. Thus, the box contains the middle half of the scores in the distribution. Vertical lines outside the box extend to the largest and the smallest observations within 1.5 interquartile ranges from the box[30].

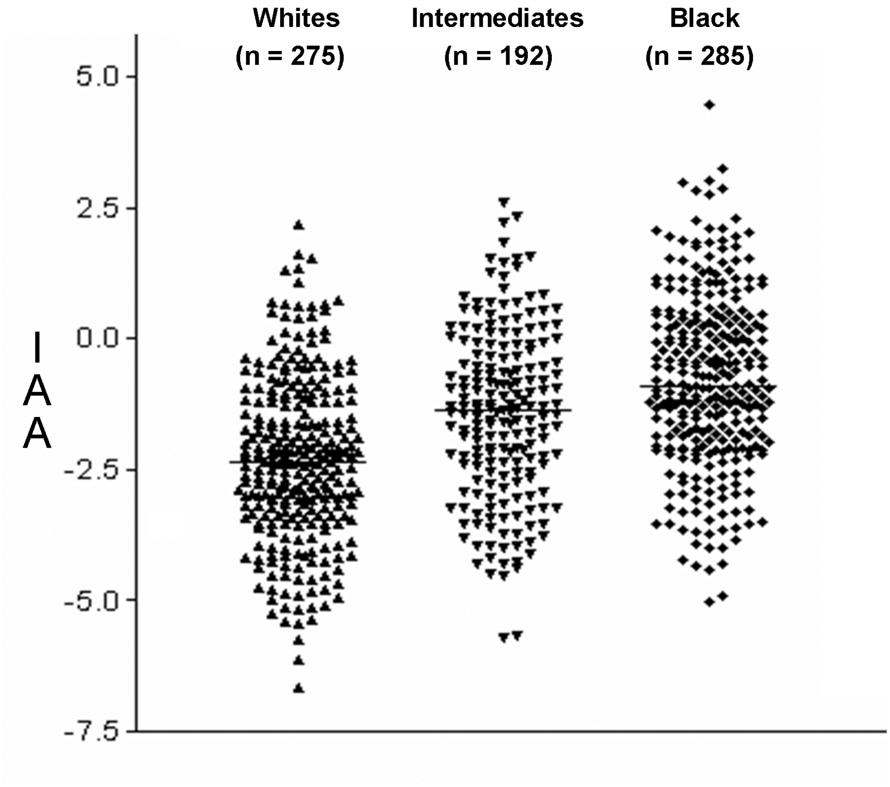

Figure 2 - Slot plot of the Index of African Ancestry (IAA) from 752 individuals from São Paulo, Brazil, separated according to their color groups (White, Intermediate and Black). Each symbol indicates the IAA value from one individual. The horizontal line indicates the median.