High-resolution characterization of 12 classical and non-classical *HLA* loci in Southern Brazilians

Mariana de Sousa Castro[1], Hellen Caroline Issler[1], Geórgia Fernanda Gelmini[2], Bruna Laís Mauricio de Miranda[2], Verónica Calonga Solís[1], Alexander H. Schmidt[3, 4], Anke Stein[3], Maria da Graça Bicalho[2], Maria Luiza Petzl-Erler[1], Danillo G. Augusto[1 *]

[1]Laboratório de Genética Molecular Humana, Departamento de Genética, Universidade Federal do Paraná, Curitiba, Brazil. [2]Laboratório de Imunogenética e Histocompatibilidade, Departamento de Genética, Universidade Federal do Paraná, Curitiba, Brazil. [3]DKMS Life Science Lab, Dresden, Germany. [4]DKMS, Tübingen, Germany.

Correspondence to:
Prof. Danillo G. Augusto, PhD
Departamento de Genética
Universidade Federal do Paraná, Caixa Postal 19071,
81531-980 Curitiba, Brazil.
Tel. (+55 41) 33611724. e-mail: danillo@augusto.bio.br

*Current address:
Department of Neurology, University of California San Francisco,
San Francisco, CA 94158, USA.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/tan.13488

Abstract

The *human leukocyte antigen (HLA*) are the most polymorphic genes in the human genome. Due to their importance for antigen recognition, HLA molecules play a central role in host defense and graft rejection upon transplantation. The aim of this study was to characterize allelic diversity of the classical *HLA* genes *HLA-A, -B, -C, -DRA, -DRB1, -DQA1, -DQB1, -DPA, -DPB1*, and the non-classical class I genes *HLA-E, -F* and *-G* at high-resolution for a population of predominantly European ancestry from Curitiba, Brazil. Genotyping of 108 individuals was performed by next generation sequencing on the MiSeq platform and also by Sanger sequencing. The genotype distributions of all loci were in accordance with Hardy-Weinberg equilibrium ($p > 0.05$) and a total of 202 *HLA* variants at second field resolution were observed for the twelve loci. The strongest linkage disequilibrium ($r^2 = 1.0$, $p < 10^{-5}$) was observed for the following pair of alleles: *HLA-B*42:01:01 ~ HLA-DRB1*03:02:01; HLA-B*14:02:01 ~ HLA-C*08:02:01; B*42:01:01 ~ HLA-C*17:01:01; HLA-DRB1*03:01:01 ~ HLA-DQB1*02:01:01 DRB1*03:01:01 ~ HLA-DQB1*02:01:01; DRB1*13:01:01~ HLA-DQB1*06:03:01 and HLA-DRB1*09:01:02 ~ HLA-DQA1*03:02*. This is the first study to characterize all twelve *HLA* genes at high resolution in a single population. On the basis of the allelic frequencies of worldwide populations and principal component analysis, we confirmed the similarity of the study population to European and other Euro-descendant populations.

Key words: NGS, human leukocyte antigen, population genetics, Curitiba

Introduction

The *human leukocyte antigen* (*HLA)* are the most polymorphic genes in the human genome and are located within the major histocompatibility complex (MHC) on chromosome 6. The HLA molecules are essential for host defense against infection; classical class I and class II molecules present peptide antigens to CD8+ and CD4+ T cells, respectively, and play a pivotal role in self/non-self discrimination. Additionally, HLA class I molecules are involved in natural killer cell regulation.[1] HLA class I and class II molecules differ according to their structure and the region where their genes are located within the MHC complex. HLA class I molecules are formed by an alpha chain that is encoded by an *HLA* gene attached to an invariant beta chain, β2-microglobulin, which is encoded by a non-*HLA* gene. In contrast, HLA class II molecules have an alpha chain attached to a beta chain, which are encoded by two different *HLA* genes.[2]

The classical *HLA* class I genes are *HLA-A, HLA-B* and *HLA-C*, while the non-classical are *HLA-E, HLA-F* and *HLA-G*. The classical *HLA* class II genes are *HLA-DRA, HLA-DRB1, HLA-DQA1, HLA-DQB1, HLA-DPA1 and HLA-DPB1* and the non-classical are *HLA-DMA, HLA-DMB, HLA-DOA and HLA-DOB*, among others.[3] *HLA* allelic variation plays a central role in susceptibility to infection, autoimmune diseases, in graft-versus-host disease after hematopoietic stem cell transplantation and in graft rejection upon transplantation of cells or solid organs.[4]

Most of the classical *HLA* loci have thousands of known alleles, whose frequencies vary remarkably across populations.[5–8] Some alleles occur in most worldwide populations, while others are unique to certain populations. Therefore, characterizing the *HLA* allelic diversity across populations may contribute to the understanding of relationships among different populations.[9] *HLA* may provide valuable information about the evolutionary history of populations, especially if the cis combinations of alleles at different loci (haplotypes) are taken into consideration.[10–12]

The Brazilian population exhibits great diversity due to the long history of migrations and admixture since its multi-ethnic origin. Historically, Brazil underwent two main migration waves after its first colonization by the Portuguese: Africans were brought to work as slaves in sugarcane plantations, mainly in the Northeast region; and Europeans, especially Portuguese, Italian, Spanish and German, came after the abolition of slavery to work in coffee plantations in the South and Southeast regions.[13] Intermarriage between those ethnic groups and with the Native Americans contributed to the ethnic admixture of the contemporary Brazilian population.[14]

Paraná State is located in Southern Brazil and, similar to the other Southern states, its population is of predominantly European ancestry. The State of Paraná had, at the beginning, a greater proportion of immigrants from Italy, but also from Poland in the 1940's and from Japan from 1950 onwards.[15] Curitiba is the capital city of Paraná State and one of the largest Brazilian cities. According to the public data from the Brazilian Institute of Geography and Statistics, 79% of the Curitiba population self-declared themselves as white, 18% as mixed and only 3% as black (https://www.ibge.gov.br). Previous studies from our group showed that the ancestry of the predominantly Euro-descendants individuals from Curitiba was 80.6% European, 12.5% African and 7.0% Amerindian. The predominantly Afro-descendants individuals, on the other hand, exhibited 49.5% African and 41.8% European ancestry.[16,17] This large European component in Afro-descendant individuals has also been observed in other studies[18,19] and evidences the large contribution of the European ancestry in Southern Brazilians.

The *HLA* polymorphism has been deeply studied across populations, however, most studies focused on *HLA-A, HLA-B* and *HLA-DRB1* as their implication for transplantation outcome has been known for longer. The diversity of non-classical *HLA* genes and of genes that encode the alpha chain of the classical class II molecules is poorly known across populations, despite the growing evidence of their impact in disease and transplantation.[20–28] Even scarcer are the studies that describe the classical and non-classical genes in a single population sample.

Here, we characterized the allelic diversity of the classical *HLA* class I *HLA-A, HLA-B* and *HLA-C*, the non-classical class I *HLA-E, HLA-F* and *HLA-G* and the alpha and beta classical *HLA* class II genes *HLA-DRA, HLA-DRB1, HLA-DQA1, HLA-DQB1, HLA-DPA1 and HLA-DPB1* in a sample from Curitiba by the combination of next generation and Sanger sequencing techniques. This is, to the best of our knowledge, the first study describing alleles of all these 12 loci at high-resolution in a single population.

Methods

Population characterization

Peripheral blood samples (8 mL) were collected from 108 unrelated individuals (73.2% female, mean age 29 years) of predominantly European ancestry from Curitiba (25°25′ S, 49°17′ W, Paraná State, Brazil, Figure 1) and its metropolitan region. Individuals who reported known non-European background were not included in this study. We also excluded those who reported acute or chronic health alterations or who were under the use of prescribed medications. All individuals voluntarily agreed to participate and provided written informed consent. The Human Research Ethics Committee of the Federal University of Paraná, in accordance with the Brazilian Federal laws, approved this study under the protocol number CAAE 68282717.0.0000.0102.

DNA extraction

The blood samples were obtained in tubes containing the anticoagulant ethylenediaminetetraacetic acid tripotassium (K3EDTA). The DNA was extracted using the Wizard® Genomic DNA kit (Promega) following the instructions provided by the manufacturer.

HLA class I and class II sequencing

High-resolution genotyping of *HLA-A, -B, -C, -DRB1, -DQB1* and *-DPB1* genes was performed by next generation sequencing, MiSeq platform, with a previously validated protocol.[29]

The class II *HLA-DPA1, -DQA1, -DRA* and the non-classical class I *HLA-E*, *HLA-F*, *HLA-G* genes were genotyped by Sanger sequencing. The exons amplified for each locus were: 2 and 3 for *HLA-DPA1;* 1, 3 and 4 for *HLA-DQA1;* 4 for *HLA-DRA;* 2 and 3 for *HLA-E*; 1, 2, 4 and 5 plus introns 1, 2, 4, and 5 for *HLA-F*; 2, 3, and 4 for *HLA-G*. Primer sequences, annealing temperatures and thermal cycling

conditions are given in the Supplementary Tables 1-4. Sequencing reactions were performed with Big Dye® Terminator Cycle Sequencing Standard v3.1 kits (Life Technologies, Foster City, USA) according to the manufacturer instructions.

Data analysis

Allelic frequencies were obtained by direct counting and haplotypes were estimated by the ELB algorithm using the ARLEQUIN software version 3.5.2.[30] Hardy-Weinberg equilibrium was tested with Genalex v6.502 program (PEAKALL; SMOUSE, 2006). Haplotype frequencies were used to estimate linkage disequilibrium (LD) between pairs of alleles of different *HLA* loci using the exact test with Arlequin v3.5.2 software.

Allelic frequencies were used to perform principal component analysis (PCA) with Minitab 17 Statistics Software. The PCA analysis reduces a large number of variables into a small set that still contains most of the information, called principal components. The components accumulate the maximum variance of the data in a decreasing order, which makes it possible to visualize the variance in dimensions. The first two components comprise most of the variance and populations that have similar *HLA* frequencies will group on the XY graph. For this analysis, we used *HLA-A, HLA-B and HLA-DRB1* allelic frequencies at two-field resolution and included populations from Europe, East Asia and Africa, as well as other Brazilian populations (Table 1). Populations were also compared using the exact test of population differentiation with Arlequin.[31] Frequencies of each haplotype were compared between all populations two by two using the χ^2 test in 2×2 contingency tables.

Results

A total of 202 different *HLA* variants (corresponding to *HLA* alleles in each locus at the resolution level of the second field) were identified: 26 *HLA-A*, 47 *HLA-B*, 22 *HLA-C*, 37 *HLA-DRB1*, 15 *HLA-DQB1*, 23 *HLA-DPB1*, 12 *HLA-DQA1*, 8 *HLA-DPA1*, 2 *HLA-DRA*, 2 *HLA-E*, 3 *HLA-F* and 5 *HLA-G*. Tables 2 and 3 show the allele frequencies at second field resolution for *HLA* class I and class II, respectively. Frequencies of alleles at higher resolution can be found in Supplementary Table 5.

The genotype distributions for all loci were in accordance with Hardy-Weinberg equilibrium ($p > 0.05$).

Frequencies of the most common alleles and haplotypes were compared to those of Maringá, Northern Paraná, the geographically closest population for which high-resolution genotyping for *HLA-DRB1, -DQA1* and *-DQB1* is available.[32] Similar frequencies were observed for *HLA-DRB1*07:01:01* (f = 0.139; p = 0.765), *HLA-DQA1*01:02* (f = 0.178; p = 0.732), *HLA-DQB1*03:01:01* (f = 0.17; p = 0.438) and for the common haplotypes *HLA-DRB1*07:01~DQA1*02:01~DQB1*02:02* (f = 0.103; p = 0.553) and *HLA-DRB1*11:01~DQA1*05:05~DQB1*03:01* haplotypes (f = 0.591; p = 0.385). Since *HLA-A* and *HLA-B* data were not available, this population data was not used in further analyzes.

The *HLA-A~HLA-B~HLA-DRB1, HLA-A~-B~-C~-E~-F~-G and HLA-DRB1~-DQB1~*-DPB1 inferred haplotypes with frequency ≥ 0.015 are shown in Tables 4-6. The list of all inferred haplotypes and their frequencies is available in Supplementary Tables 6-9.

We analyzed the LD between pairs of loci; only haplotypes observed at frequencies ≥ 0.01 and presenting strong LD (D'> 0.8, r^2 >0.6 and $p < 10^{-5}$) are shown (Table 7). Applying these criteria, we observed significant LD between alleles from the following pairs of loci: *HLA-B~HLA-DRB1; HLA-B~HLA-C; HLA-C~HLA-DRB1; HLA-DQB1~HLA-DQA1; HLA-DRB1~HLA-DQB1 and HLA-DRB1~HLA-DQA1.* No significant LD was observed for pairs that included the non-classical loci *HLA-E*, *HLA-F* and *HLA-G*. The pair with the greatest number of alleles in LD was *HLA-DQB1~HLA-DQA1*.

The exact test of population differentiation was based on the *HLA-A*, *HLA-B* and *HLA-DRB1* frequencies (table 8). Frequencies observed in the study population differed significantly from those observed in all the populations, except the Portuguese, Italian and other Brazilian populations. The first two principal components in our analysis represented 39% of the total variance. The grouping was consistent with ancestry and geography (Figure 2). The population from Curitiba grouped with European populations from Germany, Portugal, England and Italy, but

was distant from Asian and African populations. Curitiba also grouped with other Brazilian populations, except Amerindians.

Discussion

The Brazilian population is admixed due to its history of colonization and migrations, and the ethnic background differs between regions of the country.[15,33–36] In our study, we show a great diversity of *HLA* alleles in Curitiba and a large number of inferred haplotypes. The similarity of the allele frequencies found between Curitiba and population of Maringá and other Brazilian populations used for comparisons in our analysis is expected due to the fact that all populations are of predominantly European ancestry.[32,37–39] The novelty of our study is the characterization of 12 classical and non-classical loci at high resolution in a single population. Most studies that describe *HLA* in populations focus on a few loci. Therefore, LD among all those genes as well as haplotypes that include classical and non-classical *HLA* are poorly described.

Brazil was originally occupied by Native American populations and throughout its history received numerous peoples from all over the world, especially from Africa, Europe and Japan. Despite intermarriage being common, the population in the South is of predominantly European ancestry.[13] In fact, the South was colonized after the Northeast and the Southeast regions and did not receive a significant contingent of Africans during the times of slavery; migrants from Eastern Asia were the last to arrive. Accordingly, the most common haplotypes observed in Curitiba are similar to those frequent in Europeans and their frequencies differ significantly from those frequent in Asian or African populations.[6] Despite this, we observed in Curitiba a few *HLA-DRB1~HLA-DQA1~HLA-DQB1* haplotypes that are frequent among Native American populations.[11]

The two most frequent *HLA-A~HLA-B~HLA-DRB1* haplotypes **29:02:01~*44:03:01~*07:01:01* (*f* = 0.025) and **01:01:01~*08:01:01~ *03:01:01* (*f* =

0.02) were also reported as the most common in Italian Americans (*f* = 0.015; *p* = 0.357 and f = 0.049%; *p* = 0.068, respectively) and in Spanish Americans (*f* = 0.023; *p* = 0.922 and *f* = 0.021; *p* = 0.871, respectively).[40] The most frequent classical class I and classical class II haplotypes corresponded to the most common haplotypes in other European populations such as English and Spanish (data not shown).[41,42]

For *HLA* class II, the haplotype *HLA-DRB1*07:01~HLA-DQA1*02:01~HLA-DPA1*01:03~HLA-DPB1*04:01* was observed at high frequency in Spain.[42] Based on our results, this haplotype includes the allele *HLA-DQB1*02:02*, which is the most common class II haplotype in Curitiba (*f* = 0.07) (Supplementary Table 9). Interestingly, *HLA-DQB1*02:02~HLA-DRB1*07:01* was seen in strong LD (D' = 0.953; r^2 = 0.768) as well as the pair *HLA-DQB1*02:02~HLA-DQA1*02:01 (*D' = 0.862; r^2 = 0.680). Our results contribute to the knowledge of previous haplotypes described in European populations.

HLA allelic frequencies were compared across worldwide populations. As expected, Curitiba grouped with Brazilians of predominantly European ancestry as well as with European populations. Because *HLA* frequencies drastically differ among different ethnic groups, Asians and Africans did not group with European and Euro-descendants. It is important to highlight that population grouping in PCA as well as homogeneity of allele and haplotype frequencies do not necessary reflect common ancestry and may be due to stochastic factors, such as gene flow and genetic drift. Therefore, the PCA results should be analyzed carefully as they simply reflect the distribution of *HLA* alleles among these populations, and grouping cannot be interpreted as being necessarily caused by common ancestry.

Polymorphism of *HLA-A*, *-B* and *-DRB1* in Brazil was deeply analyzed based on serologic or low-resolution genotyping volunteer donors enrolled in REDOME (Brazilian Registry of Bone Marrow Donor).[43–46] However, few studies describe the diversity of many *HLA* loci at high resolution in the same population as the present study. Population studies that describe *HLA* frequencies of multiple loci, as well as their LD and haplotype frequencies are important for evolutionary and anthropological studies. Additionally, they impact donor match, as more and more is

shown that *HLA* genes other than the most studied *HLA-A*, *-B*, *-C*, *-DQB1* and *-DRB1* also impact transplantation and graft survival.[20,47] Our results reinforce allelic and haplotypic similarities between the population from Curitiba with European populations, which corroborates other studies that analyzed Euro-descendants from Southern Brazil.[37,48,49]

Acknowledgements

This project received financial support from the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Fundação Araucária, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). We thank Kirsten M. Anderson for kindly reviewing this manuscript.

References

1. Augusto DG, Petzl-Erler ML. KIR and HLA under pressure: evidences of coevolution across worldwide populations. *Hum Genet*. 2015;134(9):929-940. doi:10.1007/s00439-015-1579-9
2. Williams TM. Human leukocyte antigen gene polymorphism and the histocompatibility laboratory. *J Mol Diagn*. 2001;3(3):98-104. doi:10.1016/S1525-1578(10)60658-7
3. Horton R, Wilming L, Rand V, et al. Gene map of the extended human MHC. *Nat Rev Genet*. 2004;5(12):889-899. doi:10.1038/nrg1489
4. Colvin RB, Smith RN. Antibody-mediated organ-allograft rejection. *Nat Rev Immunol*. 2005;5(10):807-817. doi:10.1038/nri1702
5. González-Galarza FF, Takeshita LYC, Santos EJM, et al. Allele frequency net 2015 update: New features for HLA epitopes, KIR and disease and HLA adverse drug reaction associations. *Nucleic Acids Res*. 2015;43(D1):D784-D788. doi:10.1093/nar/gku1166
6. ROBINSON J et al. The IPD and IMGT/HLA database: allele variant databases. *Nucleic Acids Res*. 2015;43:D423-43.
7. Bugawan TL, Klitz W, Blair A, Erlich HA. High-resolution HLA class I typing in the CEPH families: analysis of linkage disequilibrium among HLA loci. *Tissue Antigens*. 2000;56(5):392-404. http://www.ncbi.nlm.nih.gov/pubmed/11144287. Accessed November 29, 2018.
8. Gourraud P-A, Khankhanian P, Cereb N, et al. HLA Diversity in the 1000 Genomes Dataset. Colombo GI, ed. *PLoS One*. 2014;9(7):e97282. doi:10.1371/journal.pone.0097282
9. Sanchez-Mazas A, Meyer D. The relevance of HLA sequencing in population genetics studies. *J Immunol Res*. 2014;2014. doi:10.1155/2014/971818
10. Vina MAF, Hollenbach JA, Lyke KE, et al. Tracking human migrations by the analysis of the distribution of HLA alleles, lineages and haplotypes in closed and open populations. *Philos Trans R Soc B Biol Sci*. 2012;367(1590):820-829. doi:10.1098/rstb.2011.0320
11. Tsuneto LT, Probst CM, Hutz MH, et al. HLA class II diversity in seven Amerindian populations. Clues about the origins of the Ach?? *Tissue Antigens*. 2003;62(6):512-526. doi:10.1046/j.1399-0039.2003.00139.x
12. Parham P, Arnett KL, Adams EJ, et al. Episodic evolution and turnover of HLA-B in the indigenous human populations of the Americas. *Tissue Antigens*. 1997;50(3):219-232.
13. Salzano FM, Sans M. Interethnic admixture and the evolution of Latin American populations. *Genet Mol Biol*. 2014. doi:10.1590/S1415-47572014000200003
14. Giolo SR, Soler JMP, Greenway SC, et al. Brazilian urban population genetic structure reveals a high degree of admixture. *Eur J Hum Genet*. 2012;20(1):111-116. doi:10.1038/ejhg.2011.144

15. Levy MS. O papel da migração internacional na evolução da população brasileira (1872 a 1972). *Rev Saude Publica*. 1974;Suppl(0034-8910):49-90. doi:10.1590/S0034-89101974000500003
16. Probst CM, Bompeixe EP, Pereira NF, et al. HLA polymorphism and evaluation of European, African, and Amerindian contribution to the white and mulatto populations from Parana, Brazil. *Hum Biol*. 2000;72(4):597-617.
17. Braun-Prado K, Vieira Mion AL, Farah Pereira N, Culpi L, Petzl-Erler ML. HLA class I polymorphism, as characterised by PCR-SSOP, in a Brazilian exogamic population. *Tissue Antigens*. 2000;56(5):417-427. doi:10.1034/j.1399-0039.2000.560504.x
18. Pena SDJ, di Pietro G, Fuchshuber-Moraes M, et al. The genomic ancestry of individuals from different geographical regions of Brazil is more uniform than expected. *PLoS One*. 2011;6(2). doi:10.1371/journal.pone.0017063
19. Culpi L, Salzanot FM. MIGRATION , GENETIC MARKERS AND RACE ADMIXTURE IN CURITIBA . BRAZIL. 1984:127-135.
20. Pidala J, Lee SJ, Ahn KW, et al. Nonpermissive HLA-DPB1 mismatch increases mortality after myeloablative unrelated allogeneic hematopoietic cell transplantation. *Blood*. 2014;124(16):2596-2606. doi:10.1182/blood-2014-05-576041
21. Thomas R, Thio CL, Apps R, et al. A Novel Variant Marking HLA-DP Expression Levels Predicts Recovery from Hepatitis B Virus Infection. *J Virol*. 2012;86(12):6979-6985. doi:10.1128/JVI.00406-12
22. Gough SCL, Simmonds MJ. The HLA Region and Autoimmune Disease: Associations and Mechanisms of Action. *Curr Genomics*. 2007;8(7):453-465. doi:10.2174/138920207783591690
23. Jiang Y-G, Wang Y-M, Liu T-H, Liu J. Association between HLA class II gene and susceptibility or resistance to chronic hepatitis B. *World J Gastroenterol*. 2003;9(10):2221-2225.
24. Spínola H, Lemos A, Couto AR, et al. HLA-DQA1 and HLA-DQB1 allele diversity and its extended haplotypes in Madeira Island (Portugal). *Int J Immunogenet*. 2017;44(1):27-31. doi:10.1111/iji.12300
25. van der Ven K, Skrablin S, Engels G, Krebs D. HLA-G polymorphisms and allele frequencies in Caucasians. *Hum Immunol*. 1998;59(5):302-312. http://www.ncbi.nlm.nih.gov/pubmed/9619769. Accessed November 29, 2018.
26. Arnaiz-Villena A, Palacio-Gruber J, Enriquez de Salamanca M, et al. HLA-G, -A haplotypes in Amerindians (Ecuador): HLA-G*01:05N World distribution. *Hum Immunol*. 2018;79(2):89-90. doi:10.1016/j.humimm.2017.12.002
27. Felício LP, Porto IOP, Mendes-Junior CT, et al. Worldwide *HLA-E* nucleotide and haplotype variability reveals a conserved gene for coding and 3′ untranslated regions. *Tissue Antigens*. 2014;83(2):82-93. doi:10.1111/tan.12283
28. Castelli EC, Ramalho J, Porto IOP, et al. Insights into HLA-G genetics provided by worldwide haplotype diversity. *Front Immunol*. 2014;5(OCT). doi:10.3389/fimmu.2014.00476

29. Lange V, Böhme I, Hofmann J, et al. Cost-efficient high-throughput HLA typing by MiSeq amplicon sequencing. *BMC Genomics*. 2014;15(1):63. doi:10.1186/1471-2164-15-63
30. Excoffier, L. and Lischer HEL. An Integrated Software Package for Population Genetics Data Analysis. *Mol Ecol Resour*. 2010;10:564-567. doi:10.1111/j.1755-0998.2010.02847.x
31. Raymond M, Rousset F. An exact test for population differenciation. *Evolution (N Y)*. 1995;49(6):1280-1283. doi:10.2307/2410454
32. Reis PG, Sell AM, Sakita KM, Moliterno RA, Visentainer J EL. HLA-DRB1, DQA1 and DQB1 diversity in a mixed population of paraná, Southern Brazil. *Hum Immunol*. 2015;76(2015):153. doi:10.1016/j.humimm.2015.07.213
33. Carvalho-Silva DR, Santos FR, Rocha J, Pena SDJ. The Phylogeography of Brazilian Y-Chromosome Lineages. *Am J Hum Genet*. 2001;68:281-286.
34. Trachtenberg A, Jobim LFJ, Kraemer E, et al. The HLA polymorphism in five Brazilian populations. *Ann Hum Biol*. 1988. doi:10.1080/03014468800009651
35. Guerreiro-Junior V, Bisso-Machado R, Marrero A, Hünemeier T, Salzano FM, Bortolini MC. Genetic signatures of parental contribution in black and white populations in Brazil. *Genet Mol Biol*. 2009. doi:10.1590/S1415-47572009005000001
36. Kehdy FSG, Gouveia MH, Machado M, et al. Origin and dynamics of admixture in Brazilians and its effect on the pattern of deleterious mutations. *Proc Natl Acad Sci*. 2015;112(28):8696-8701. doi:10.1073/pnas.1504447112
37. Reis PG, Ambrosio-Albuquerque EP, Fabreti-Oliveira RA, et al. HLA-A, -B, -DRB1, -DQA1, AND -DQB1 PROFILE IN A POPULATION FROM SOUTHERN BRAZIL. *HLA*. 2018. doi:10.1111/tan.13368
38. Bortolotto AS, Petry MG, da Silveira JG, et al. HLA-A, -B, and -DRB1 allelic and haplotypic diversity in a sample of bone marrow volunteer donors from Rio Grande do Sul State, Brazil. *Hum Immunol*. 2012;73(2):180-185. doi:10.1016/j.humimm.2011.11.009
39. Meyer D, Single RM. 13th International Histocompatibility Workshop Anthropology / Human Genetic Diversity Joint Report – Chapter 1 : Introduction and Overview. *Genetics*. (1):1-4.
40. Mack SJ, Tu B, Yang R, Masaberg C, Ng J, Hurley K. between Italian and Spanish Americans. 2011;72(2):144-149. doi:10.1016/j.humimm.2010.10.017.Human
41. Doherty DG, Vaughan RW, Donaldson PT, Mowat AP. HLA DQA, DQB, and DRB genotyping by oligonucleotide analysis: distribution of alleles and haplotypes in British caucasoids. *Hum Immunol*. 1992;34(1):53-63. doi:10.1016/0198-8859(92)90085-2
42. Comas D, Mateu E, Calafell F, et al. HLA class I and class II DNA typing and the origin of Basques. *Tissue Antigens*. 1998;51(1):30-40. doi:10.1111/j.1399-0039.1998.tb02944.x
43. Bardi MS, Jarduli LR, Jorge AJ, et al. HLA-A, B and DRB1 allele and haplotype frequencies in volunteer bone marrow donors from the north of Parana State.

Rev Bras Hematol Hemoter. 2012;34(1):25-30. doi:10.5581/1516-8484.20120010

44. Ayo CM, da Silveira Camargo A V., Xavier DH, et al. Frequencies of allele groups HLA-A, HLA-B and HLA-DRB1 in a population from the northwestern region of S??o Paulo State, Brazil. *Int J Immunogenet*. 2015;42(1):19-25. doi:10.1111/iji.12159
45. Rodrigues C, Macedo LC, Bruder A V., et al. Allele and haplotype frequencies of HLA-A, B, C, DRB1 and DQB1 genes in polytransfused patients in ethnically diverse populations from Brazil. *Int J Immunogenet*. 2015;42(5):322-328. doi:10.1111/iji.12206
46. Carvalho MG, Tsuneto LT, Moita Neto JM, et al. HLA-A, HLA-B and HLA-DRB1 haplotype frequencies in Piauí's volunteer bone marrow donors enrolled at the Brazilian registry. *Hum Immunol*. 2013;74(12):1598-1602. doi:10.1016/j.humimm.2013.08.283
47. Pabón MA, Navarro CE, Osorio JC, et al. Impact of human leukocyte antigen molecules E, F, and G on the outcome of transplantation. *Transplant Proc*. 2014;46(9):2957-2965. doi:10.1016/j.transproceed.2014.07.010
48. Boquett JA, Nunes JM, Buhler S, et al. The HLA-A, -B and -DRB1 polymorphism in a large dataset of South Brazil bone marrow donors from Rio Grande do Sul. *HLA*. 2017;89(1):29-38. doi:10.1111/tan.12933
49. Augusto DG, Zehnder-Alves L, Pincerati MR, Martin MP, Carrington M, Petzl-Erler ML. Diversity of the KIR gene cluster in an urban Brazilian population. *Immunogenetics*. 2012;64(2):143-152. doi:10.1007/s00251-011-0565-1
50. Reis PG, Ambrosio-Albuquerque EP, Fabreti-Oliveira RA, et al. Hla-a, -B, -Drb1, -Dqa1, and -Dqb1 Profile in a Population From Southern Brazil. *Hla*. 2018;(July):298-303. doi:10.1111/tan.13368
51. Lázaro AM, Moraes ME, Marcos CY, Moraes JR, Fernández-Viña MA, Stastny P. Evolution of HLA-class I compared to HLA-class II polymorphism in Terena, a South-American Indian tribe. *Hum Immunol*. 1999;60(11):1138-1149. http://www.ncbi.nlm.nih.gov/pubmed/10600013. Accessed December 6, 2018.
52. Shen C, Zhu B, Ye S, et al. Allelic diversity and haplotype structure of HLA loci in the Chinese Han population living in the Guanzhong region of the Shaanxi province. *Hum Immunol*. 2010;71(6):627-633. doi:10.1016/j.humimm.2010.02.012
53. Alfirevic A, Gonzalez-Galarza F, Bell C, et al. In silico analysis of HLA associations with drug-induced liver injury: Use of a HLA-genotyped DNA archive from healthy volunteers. *Genome Med*. 2012;4(6):1-14. doi:10.1186/gm350
54. Schmidt AH, Baier D, Solloch U V., et al. Estimation of high-resolution HLA-A, -B, -C, -DRB1 allele and haplotype frequencies based on 8862 German stem cell donors and implications for strategic donor registry planning. *Hum Immunol*. 2009;70(11):895-902. doi:10.1016/j.humimm.2009.08.006
55. Rendine S, Ferrero NM, Sacchi N, Costa C, Pollichieni S, Amoroso A. Estimation of human leukocyte antigen class I and class II high-resolution

allele and haplotype frequencies in the Italian population and comparison with other European populations. *Hum Immunol*. 2012;73(4):399-404. doi:10.1016/j.humimm.2012.01.005
56. Saito S, Ota S, Yamada E, Inoko H, Ota M. Allele frequencies and haplotypic associations defined by allelic DNA typing at HLA class I and class II loci in the Japanese population. *Tissue Antigens*. 2000;56(6):522-529. doi:10.1034/j.1399-0039.2000.560606.x
57. Weiskopf D, Grifoni A, Arlehamn CSL, et al. Sequence-based HLA-A, B, C, DP, DQ, and DR typing of 339 adults from Managua, Nicaragua. *Hum Immunol*. 2018;79(1):1-2. doi:10.1016/j.humimm.2017.11.002
58. Assane AAA, Fabricio-Silva GM, Cardoso-Oliveira J, et al. Human leukocyte antigen-A, -B, and -DRB1 allele and haplotype frequencies in the Mozambican population: a blood donor-based population study. *Hum Immunol*. 2010;71(10):1027-1032. doi:10.1016/j.humimm.2010.06.017
59. Bettencourt BF, Santos MR, Pereira J, et al. HLA-A, -B, -C, -DQA1, -DQB1, -DRB1, -E, -F and -G genotyping of 130 individuals from Terceira Island, Azores, Portugal. *Hum Immunol*. 2016;77(6):445-446. doi:10.1016/j.humimm.2016.03.008
60. Grifoni A, Sidney J, Carpenter C, et al. Sequence-based HLA-A, B, C, DP, DQ, and DR typing of 159 individuals from the Worcester region of the Western Cape province of South Africa. *Hum Immunol*. 2018;79(3):143-144. doi:10.1016/j.humimm.2018.01.004

Table 1. Populations whose *HLA* frequency data was used for the comparisons.

Population	Reference	Sample size	Ancestry
Brazil Curitiba	Present study	108	Euro-descendant
Brazil Maringá	[50]	641	Euro-descendant
Brazil REDOME Piauí	*	46140	Admixed
Brazil Ribeirão Preto	[39]	108	Admixed
Brazil Terena	[51]	60	Native American
Brazil Rio Grande do Sul	[38]	4428	Euro-descendant
China	[52]	10.000	East Asian
England North west	[53]	161	European
Germany	[54]	8862	European
Italy	[55]	975	European
Japan	[56]	371	East Asian
Kenya	[57]	100	African
Mozambique	[58]	202	African
Portugal	[59]	130	European
South Africa	[60]	159	African

*Data published in Allele Frequency Net Database (http://www.allelefrequencies.net)[5]

Table 2. Allele frequencies of six *HLA* class I genes in 108 individuals from Curitiba, Brazil.

HLA-A		*HLA-B*		*HLA-C*		*HLA-E, -F, -G*	
Alleles	*f*	*Alleles*	*f*	*Alleles*	*f*	*Alleles*	*f*
*A*01:01*	0.075	*B*07:02*	0.082	*C*01:02*	0.024	*E*01:01*	0.538
*A*02:01*	0.226	*B*07:05*	0.005	*C*02:02*	0.052	*E*01:03*	0.462
*A*02:02*	0.009	*B*08:01*	0.034	*C*02:10*	0.010		
*A*02:04*	0.005	*B*13:02*	0.019	*C*03:03*	0.057	*F*01:01*	0.827
*A*02:05*	0.014	*B*14:01*	0.005	*C*03:04*	0.057	*F*01:03*	0.168
*A*02:11*	0.009	*B*14:02*	0.029	*C*03:04*	0.005	*F*01:04*	0.005
*A*02:17*	0.009	*B*15:01*	0.067	*C*04:01*	0.167		
*A*03:01*	0.108	*B*15:03*	0.010	*C*04:29*	0.005	*G*01:01*	0.794
*A*11:01*	0.047	*B*15:09*	0.010	*C*05:01*	0.038	*G*01:03*	0.049
*A*23:01*	0.038	*B*15:10*	0.005	*C*06:02*	0.052	*G*01:04*	0.118
*A*23:05*	0.005	*B*15:15*	0.005	*C*07:01*	0.148	*G*01:05*	0.01
*A*24:02*	0.142	*B*15:17*	0.019	*C*07:02*	0.124	*G*01:06*	0.029
*A*25:01*	0.033	*B*18:01*	0.053	*C*07:04*	0.014		
*A*26:01*	0.033	*B*27:02*	0.005	*C*08:02*	0.029		
*A*29:01*	0.005	*B*27:05*	0.024	*C*12:03*	0.09		
*A*29:02*	0.042	*B*35:01*	0.053	*C*14:02*	0.005		
*A*30:01*	0.028	*B*35:02*	0.024	*C*14:03*	0.005		
*A*30:02*	0.028	*B*35:03*	0.029	*C*15:02*	0.048		
*A*31:01*	0.038	*B*35:08*	0.014	*C*15:05*	0.010		
*A*32:01*	0.024	*B*38:01*	0.038	*C*16:01*	0.043		
*A*33:01*	0.019	*B*39:01*	0.029	*C*16:02*	0.010		
*A*33:03*	0.005	*B*39:02*	0.005	*C*17:01*	0.010		
*A*68:01*	0.028	*B*39:05*	0.010				
*A*68:02*	0.009	*B*39:06*	0.005				
*A*69:01*	0.009	*B*39:06*	0.005				
*A*74:01*	0.009	*B*39:09*	0.010				
		*B*39:13*	0.005				
		*B*40:01*	0.024				
		*B*40:02*	0.014				
		*B*40:04*	0.005				
		*B*41:01*	0.010				
		*B*42:01*	0.010				
		*B*44:02*	0.029				
		*B*44:03*	0.077				
		*B*44:05*	0.014				
		*B*45:01*	0.005				

		*B*47:01*	0.005		
		*B*49:01*	0.034		
		*B*50:01*	0.014		
		*B*51:01*	0.058		
		*B*52:01*	0.010		
		*B*53:01*	0.024		
		*B*55:01*	0.019		
		*B*56:01*	0.005		
		*B*57:01*	0.024		
		*B*58:01*	0.019		
		*B*81:01*	0.005		
HW (*p):*	0.13		0.57	0.77	0.65 0.60 0.70

f = frequency; HW = Hardy-Weinberg; *p* = p-value

Table 3. Allele frequencies of six *HLA* class II genes in a sample of 108 volunteer individuals from Curitiba, Brazil.

HLA-DRB1		*HLA-DQB1*		*HLA-DPB1*		*HLA-DRA, -DQA1, -DPA1*	
Alleles	*f*	*Alleles*	*f*	*Alleles*	*f*	*Alleles*	*f*
*DRB1*01:01*	0.072	*DQB1*02:01*	0.057	*DPB1*01:01*	0.052	*DRA*01:01*	0.656
*DRB1*01:02*	0.034	*DQB1*02:02*	0.118	*DPB1*01:05*	0.009	*DRA*01:02*	0.344
*DRB1*03:01*	0.058	*DQB1*03:01*	0.170	*DPB1*01:24*	0.005		
*DRB1*03:02*	0.010	*DQB1*03:02*	0.104	*DPB1*02:01*	0.132	*DQA1*01:01*	0.091
*DRB1*04:01*	0.019	*DQB1*03:03*	0.033	*DPB1*02:02*	0.014	*DQA1*01:02*	0.178
*DRB1*04:02*	0.029	*DQB1*03:19*	0.019	*DPB1*02:96*	0.005	*DQA1*01:03*	0.063
*DRB1*04:03*	0.010	*DQB1*04:02*	0.104	*DPB1*03:01*	0.047	*DQA1*01:04*	0.038
*DRB1*04:04*	0.019	*DQB1*05:01*	0.132	*DPB1*04:01*	0.316	*DQA1*01:05*	0.029
*DRB1*04:05*	0.014	*DQB1*05:02*	0.033	*DPB1*04:02*	0.151	*DQA1*02:01*	0.130
*DRB1*04:06*	0.005	*DQB1*05:03*	0.033	*DPB1*05:01*	0.028	*DQA1*03:01*	0.106
*DRB1*04:07*	0.01	*DQB1*06:02*	0.066	*DPB1*06:01*	0.024	*DQA1*03:02*	0.010
*DRB1*04:11*	0.019	*DQB1*06:03*	0.066	*DPB1*09:01*	0.009	*DQA1*03:03*	0.034
*DRB1*07:01*	0.139	*DQB1*06:04*	0.052	*DPB1*10:01*	0.038	*DQA1*04:01*	0.101
*DRB1*08:01*	0.048	*DQB1*06:09*	0.014	*DPB1*11:01*	0.028	*DQA1*05:01*	0.063
*DRB1*08:02*	0.010			*DPB1*13:01*	0.019	*DQA1*05:05*	0.159
*DRB1*08:04*	0.014			*DPB1*14:01*	0.033		
*DRB1*08:07*	0.019			*DPB1*15:01*	0.009	*DPA1*01:03*	0.701
*DRB1*09:01*	0.010			*DPB1*16:01*	0.009	*DPA1*01:05*	0.005
*DRB1*10:01*	0.005			*DPB1*17:01*	0.033	*DPA1*02:01*	0.210
*DRB1*11:01*	0.082			*DPB1*18:01*	0.009	*DPA1*02:02*	0.042
*DRB1*11:02*	0.014			*DPB1*19:01*	0.009	*DPA1*02:03*	0.005
*DRB1*11:03*	0.014			*DPB1*23:01*	0.005	*DPA1*02:07*	0.009
*DRB1*11:04*	0.038			*DPB1*27:01*	0.009	*DPA1*03:01*	0.023
*DRB1*12:01*	0.014			*DPB1*45:01*	0.005	*DPA1*04:01*	0.005
*DRB1*12:02*	0.005						
*DRB1*13:01*	0.067						
*DRB1*13:02*	0.077						
*DRB1*13:03*	0.005						
*DRB1*13:05*	0.005						
*DRB1*14:04*	0.005						
*DRB1*14:06*	0.005						
*DRB1*14:54*	0.024						
*DRB1*15:01*	0.043						
*DRB1*15:03*	0.024						
*DRB1*15:11*	0.005						
*DRB1*16:01*	0.029						

HW (*p):*	0.214	0.498	0.240	0.420 0.963 0.125

f = frequency; HW = Hardy-Weinberg; *p* = p-value

Table 4. Most common *HLA-A~HLA-B~HLA-DRB1* haplotypes (frequency ≥ 0.015).

HLA-A~HLA-B~HLA-DRB1	*f*
**29:02:01~*44:03:01~*07:01:01*	0.025
**01:01:01~*08:01:01~*03:01:01*	0.020
**23:01:01~*44:03:01~*07:01:01*	0.015
**02:01:01~*15:01:01~*11:01:01*	0.015

All the haplotypes exhibited significant LD ($p < 0.05$) between all pairs of neighboring loci. A total of 170 haplotypes was inferred by the ELB method (Supplementary table 6). *f* = frequency

Table 5. Most common *HLA* class I haplotypes (frequency ≥ 0.015).

HLA-A~HLA-B~HLA-C~HLA-E~HLA-F~HLA-G	*f*
**01:01:01~*08:01:01~*07:01:01~*01:01~*01:01:01~*01:01:02*	0.024
**02:01:01~*44:02:01~*05:01:01~*01:01~*01:01:01~*01:01:01*	0.024
**29:02:01~*44:03:01~*16:01:01~*01:03~*01:01:01~*01:01:01*	0.024
**03:01:01~*35:01:01~*04:01:01~*01:03~*01:03:01~*01:01:01*	0.018
**11:01:01~*35:01:01~*04:01:01~*01:01~*01:01:02~*01:01:03*	0.018
**02:01:01~*40:01:01~*03:04:01~*01:03~*01:01:01~*01:01:01*	0.018
**03:01:01~*07:02:01~*07:02:01~*01:03~*01:01:01~*01:01:01*	0.018
**02:01:01~*51:01:01~*15:02:01~*01:03~*01:01:01~*01:01:01*	0.018
**02:01:01~*15:01:01~*03:04:01~*01:03~*01:01:01~*01:01:01*	0.018

None of the most frequent inferred haplotypes exhibited significant LD ($p < 0.05$) between all pairs of neighboring loci. A total of 139 haplotypes was inferred by the ELB method (Supplementary table 7). *f* = frequency

Table 6. Most common *HLA-DRB1~HLA-DQB1~HLA-DPB1* haplotypes (frequency ≥ 0.015).

HLA-DRB1~HLA-DQB1~HLA-DPB1	*f*
07:01:01~*02:02:01~*04:01:01	**0.067**
**11:01:01~*03:01:01~*04:02:01*	0.034
01:01:01~*05:01:01~*04:01:01	**0.034**
13:02:01~*06:04:01~*02:01	**0.034**
**15:01:01~*06:02:01~*04:01:01*	0.024
01:02:01~*05:01:01~*04:01:01	**0.024**
**13:01:01~*06:03:01~*04:01:01*	0.019
13:01:01~*06:03:01~*04:02:01	**0.019**
07:01:01~*02:02:01~*11:01:01	**0.019**
**11:04:01~*03:01:01~*04:02:01*	0.019
07:01:01~*02:02:01~*04:02:01	**0.019**
**11:01:01~*03:01:01~*04:01:01*	0.019

In bold, the haplotypes with significant LD ($p < 0.05$) for all pairs of neighboring loci. A total of 112 haplotypes was inferred by the ELB method (Supplementary table 8). *f* = frequency

Table 7. Pairs of alleles at different *HLA* loci that present strong and significant linkage disequilibrium.

Loci	*f*	D'	r^2
HLA-B~HLA-DRB1			
**15:03:01~*08:04:01*	0.01	1.000	0.663
**42:01:01~*03:02:01*	0.01	1.000	1.000
HLA-B~HLA-C			
**07:02:01~*07:02:01*	0.060	1.000	0.623
**14:02:01~*08:02:01*	0.025	1.000	1.000
**42:01:01~*17:01:01*	0.010	1.000	1.000
**44:02:01~*05:01:01*	0.035	1.000	0.743
HLA-DQB1~HLA-DQA1			
**02:01:01~*05:01:01*	0.055	0.903	0.744
**02:02:01~*02:01:01*	0.095	0.862	0.680
**03:01:01~*05:05:01*	0.080	0.849	0.648
**04:02:01~*04:01*	0.055	0.883	0.698
**05:01:01~*01:01*	0.085	0.939	0.594
**05:03:01~*01:04*	0.013	1.000	0.871
**06:03:01~*01:03:01*	0.055	1.000	0.924
HLA-DRB1~HLA-DQB1			
**03:01:01~*02:01:01*	0.055	1.000	1.000
**07:01:01~*02:02:01*	0.120	0.953	0.767
**11:02:01~*03:19:01*	0.015	1.000	0.747
**13:01:01~*06:03:01*	0.060	1.000	1.000
**13:02:01~*06:04:01*	0.050	1.000	0.671
HLA-DRB1~HLA-DQA1			
**03:01:01~*05:01:01*	0.555	0.903	0.744
**07:01:01~*02:01:01*	0.120	0.913	0.769
**09:01:02~*03:02*	0.010	1.000	1.000
**13:01:01~*01:03:01*	0.065	1.000	0.924
**14:54:01~*01:04*	0.015	1.000	0.707

Only alleles with frequencies greater than 0.01. D' > 0.8 and r^2 >0.6 are shown. The significance is $p < 10^{-5}$. *f* = frequency

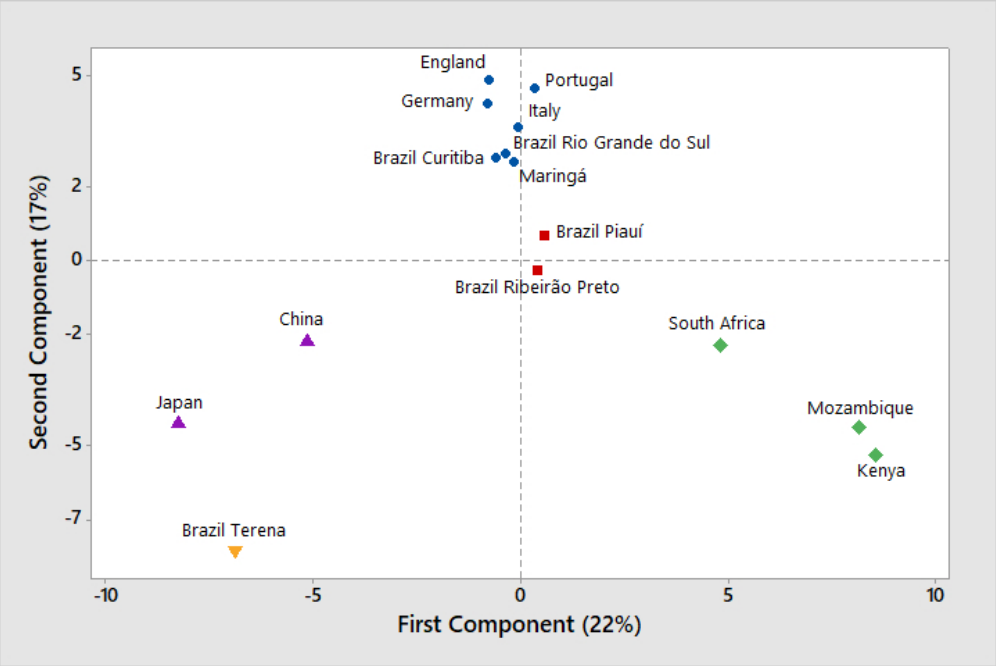
Table 8. Genetic differentiation between Curitiba and other populations based on *HLA-A. HLA-B* and *HLA-DRB1* at second field resolution.

	Curitiba (n = 108) p-value
Brazil Maringá (641)	0.52
Brazil Piauí (46140)	0.83
Brazil Ribeirão Preto (108)	0.95
Brazil Rio Grande do Sul (4428)	0.06
Brazil Terena (60)	$< 10^{-5}$
China (10000)	$< 10^{-5}$
England (161)	$< 10^{-5}$
Germany (8862)	$< 10^{-5}$
Italy (975)	0.09
Japan (371)	$< 10^{-5}$
Kenya (100)	$< 10^{-5}$
Mozambique (202)	$< 10^{-5}$
Portugal (130)	0.99
South Africa (59)	$< 10^{-5}$

Information about the populations and references are in table 1. The *p*-values shown are for test exact of population differentiation.

Figure 1 - Location of the study population. Language spoken: Portuguese; Linguistic family: Indo-European. Geographic location: South America, Brazil, South region; 25°25’ S, 49°17’ W

188x154mm (150 x 150 DPI)



Principal component analysis groups Curitiba with European populations. Each dot represents a population. Circle = European or Euro-descendant; square = Brazilian admixed; diamond = African; triangle = South East Asian or Amerindian. The list of populations and references can be found in Table 1.

110x73mm (165 x 166 DPI)