HLA Haplotypes and Differential Regional Incidence of COVID-19 in Brazil: A Population Study Based in a Large Bone Marrow Donors Bank Dataset

Juliano André Boquett
Universidade Federal do Rio Grande do Sul
Fernanda Sales Luiz Vianna
Universidade Federal do Rio Grande do Sul
Nelson Jurandi Rosa Fagundes
Universidade Federal do Rio Grande do Sul
Lucas Schroeder
Unisinos: Universidade do Vale do Rio dos Sinos
Marcia Barbian
Universidade Federal do Rio Grande do Sul
Marcelo Zagonel-Oliveira
Unisinos: Universidade do Vale do Rio dos Sinos
Tiago Finger Andreis
Universidade Federal do Rio Grande do Sul
Luis Cristóvão de Moraes Sobrino Pôrto
UERJ: Universidade do Estado do Rio de Janeiro
José Artur Bogo Chies
Universidade Federal do Rio Grande do Sul
Lavinia Schuler-Faccini
Universidade Federal do Rio Grande do Sul
Patricia Ashton-Prolla
Hospital de Clinicas de Porto Alegre
CLÉVIA ROSSET (crosset@hcpa.edu.br)
Hospital de Clinicas de Porto Alegre https://orcid.org/0000-0002-1728-4770

Research Article

Keywords: COVID-19 variability, COVID-19 mortality, Genetic susceptibility, SARS-CoV-2

Posted Date: March 22nd, 2021

DOI: https://doi.org/10.21203/rs.3.rs-325153/v1

Abstract

Background

Coronavirus disease 2019 (COVID-19) rapidly spread all over the world causing high morbidity and mortality. Brazil is currently the third country in the world in the number of COVID-19 cases. Even though all Brazilian regions and states have reported a high number of cases, mortality rates varies among them. Environmental and genetic factors may influence the immune response towards SARS-CoV-2. The Brazilian population is highly heterogeneous, with different colonization and immigration histories in each region resulting in different genetic backgrounds. Here, we test if specific HLA haplotypes are associated with COVID-19 incidence and mortality in Brazil.

Methods

HLA data was obtained from The Brazilian Voluntary Bone Marrow Donors Registry (REDOME) which harbors data from more than four million individual donors, and COVID-19 data was retrieved from epidemiological bulletins issued by State Health Secretariats via the Ministry of Health of Brazil. We tested the association between the most frequent HLA haplotypes in Brazil and COVID-19 incidence and mortality using Spearman's correlation analysis.

Results

No correlation between HLA haplotypes and COVID-19 rates was found when we analyzed data from the 26 states and Federal District, as well as when we analyzed data from the 90 cities with at least 50 deaths registered in the São Paulo state. Significant negative correlation (suggestive of protection) between COVID-19 mortality and haplotypes HLA-A*01~B*08~DRB1*03, HLA-A*29~B*44~DRB1*07 and HLA-A*02~B*44~DRB1*04 was found when analyzing data from cities with at least 50 deaths registered in the entire country.

Conclusions

Our results do not support an association of specific HLA haplotypes with an increased risk of contracting SARS-CoV2 or dying from COVID-19 in Brazil. Nevertheless, using bone marrow donor registries for testing for associations between HLA variation and COVID-19 outcomes may represent an additional tool for health policymakers in the fight against COVID-19.

Background

Coronavirus disease 2019 (COVID-19), caused by SARS-CoV-2, rapidly spread to various countries causing high morbidity and mortality, and the World Health Organization (WHO) declared it a pandemic on March 11, 2020 [1]. Brazil is the third country in the world in the number of COVID-19 cases, with 8,933,356 confirmed cases and 218,878 deaths as of January 26th, 2021, accounting for 1,026 deaths per million inhabitants [2]. The first COVID-19 case in Brazil occurred in February 2020 in São Paulo state

(Southeast region). A few days later, suspected cases were reported from all remaining 26 Brazilian states. By the end of April 2020, the North region had high community transmission and the highest mortality rates in the country. In June and July, the number of daily confirmed cases continued to increase in all regions, reaching peaks in July 29th, 2020 and, again, in January 7th, 2021. The actual number of cases could be underestimated due to the low number of tests performed [3]. Even though all Brazilian regions and states have reported a high number of cases, mortality rates varied between them. Indeed, the symptomatology and mortality rates vary among different geographical regions and depend on the patient clinical profile, presence of comorbidities, and access to the health system [4–6]. Individual environmental and genetic factors may also influence susceptibility and immune response to the SARS-CoV-2.

Immune response variability may involve variants of the innate immune system and variants in the specific adaptive immune system [7–11]. Antigen presentation to T lymphocytes is one of the most important steps in determining immune response. The genetic complex of the classic human leukocyte antigen (HLA) encodes the proteins of the main histocompatibility complex (MHC), which mediates intracellular antigen presentation (HLA-A,-B,-C - MHC Class-I) and extracellular antigen presentation (HLA-DP, -DM, -DO, -DQ, and -DR - MHC class II) to the T cell receptors (TCR) on the surface of the CD8 + and CD4 + T lymphocytes, respectively [12]. Thus, by affecting the ability of T cells to respond to a specific antigen, HLA is one of the most important molecules of the immune system. The HLA alleles are highly polymorphic and their frequency varies widely in human populations, including among Brazilian regions [13]. In fact, Brazilian population is highly heterogeneous, with different colonization and immigration histories in each region resulting in different genetic backgrounds [14].

Previous case-control and bioinformatics studies have associated a few HLA genotypes with higher susceptibility or protection against the development of severe disease in those infected with SARS-CoV [15–18]. SARS-CoV-2 is very similar to SARS-CoV regarding aminoacid sequences, and, therefore, it is likely that HLA alleles associated with susceptibility to SARS-CoV may also play a role in COVID-19. Recently, HLA-A*02 and HLA-B*15 alleles were found to be the main presenters of SARS-CoV-2 antigens and HLA-A*25 and HLA-B*46 alleles were shown to have fewer predicted binding SARS-CoV-2 peptides in an *in-silico* analysis [19]. Another study analyzed the regional frequencies for the most common Italian haplotypes from the Italian Bone Marrow Donor Registry and found that the two most frequent HLA haplotypes in the Italian population were correlated with COVID-19 incidence and mortality [20]. The Brazilian Voluntary Bone Marrow Donors Registry (REDOME, in Portuguese *Registro de Doadores de Medula Óssea*) is the third largest bank of bone marrow donors in the world, with more than 5.2 million individuals registered to date. REDOME keeps information such as HLA-A, -B, and -DRB1 genotypes in low and/or medium resolution and city of residence of the donors.

In this study we evaluate if the most frequent HLA haplotypes in Brazil correlated to COVID-19 incidence and mortality. We also describe the frequencies of HLA genotypes previously associated with susceptibility or protection to SARS-CoV or SARS-CoV-2 infection in Brazil, which may help in establishing public control policies against COVID-19.

Methods

Data sources

For HLA data, we used a dataset composed by 4,148,713 individuals who volunteered as potential hematopoietic stem cell donors registered at REDOME until September, 2017. This registry includes information such as city of residence and HLA-A, -B, and -DRB1 genotypes. Donors come from recruitment centers distributed throughout the country and their DNA was genotyped in Health Ministry accredited Brazilian laboratories. Only data genotyped by molecular methods were included in the analyses. The volunteers are genotyped for the allelic group at the time of registration, and high-resolution genotyping is performed only on those potential donors selected after an initial screening. Thus, only low resolution (allelic group) genotypes were used in this study. The dataset was subdivided according to both state and city of residence. The Brazilian territory is divided into 26 states and one Federal District and counts with 5,570 municipalities.

Data about rates of COVID-19 cases and deaths were obtained from epidemiological bulletins issued by State Health Secretariats via the Ministry of Health of Brazil, available at https://bigdata-covid19a.icict.fiocruz.br/ [21]. COVID-19 rates for states were obtained on November 16th, 2020 and the rates for cities were obtained on October 24th, 2020.

This study was approved by the Ethics Committee in Research of Hospital de Clínicas de Porto Alegre, Brazil (CAAE 34701720300005327, GPPG 2020 – 0361), and all methods were carried out in accordance with local guidelines and regulations.

Statistical analysis

The incidence and mortality rates of COVID-19 were obtained based on the confirmed cases and deaths issued by each of the State Health Secretariats. To calculate the coefficient of incidence and mortality of each municipality, the number of confirmed cases or deaths, respectively, was divided by the resident population and multiplied by the population base of 100 thousand inhabitants.

Allele and haplotype frequency estimations and Hardy-Weinberg equilibrium (HWE) test were performed using the GENE[RATE] tools as described elsewhere [22–24]. Boxplots of allele and haplotype frequency were generated at RStudio Version 1.3.1093. Maps of rates of COVID-19 cases and deaths were performed in ArcGis v10.3. Spearman's correlation test between HLA alleles and the five most frequent haplotypes in Brazil [25] *versus* rates of COVID-19 cases and deaths was performed using IBM SPSS software, Version 20.0 (IBM Corp., Armonk, NY). We evaluated the HLA x COVID-19 cases and deaths correlations in the following scenarios: (1) the 26 states and Federal District of Brazil (data obtained until November 16th); (2) only cities with at least 50 deaths due to COVID-19 registered in the entire Brazilian territory (data obtained until October 24th, 2020) and; (3) only cities with at least 50 deaths registered in São Paulo state (data obtained until October 24th, 2020). In scenario #1 we replicated the approach previously performed by Pisanti *et al.* (2000) [20]; in scenario #2 we used municipalities with a defined

minimum number of deaths to increase the number of observations and, therefore, gain statistical power; finally, in scenario #3 we used data from a single state (São Paulo) in order to control population heterogeneity. São Paulo state was chosen as a study model since it can be considered as similar to a European country in terms of population and territorial dimensions. P-values below 0.05 were considered statistically significant. We applied the FDR (false discovery rate) correction for multiple tests to avoid having too many false-positives in haplotype and allele correlations.

Results

Geographical distribution of COVID-19 epidemic in Brazil

The states with the highest incidence of COVID-19 cases registered until November 16h, 2020 were Roraima, the Federal District and Amazonas, respectively; while the states with the lowest incidence were Pernambuco, Minas Gerais and Rio de Janeiro. Regarding COVID-19 deaths, the Federal District, Rio de Janeiro and Mato Grosso had the highest rates, respectively, while the lowest death rates were observed in Minas Gerais, Santa Catarina and Paraná, respectively (Table 1 and Fig. 1). The cities with the highest incidence of COVID-19 cases are Parauapebas (Pará), Boa Vista (Roraima) and Araguaína (Tocantins), all of them located in the North region of the country, while Santa Helena de Goiás (Goiás), Guajará-Mirim (Rodônia) and Rio de Janeiro (Rio de Janeiro), located in Central-West, North and Southeast regions of Brazil, respectively, had the highest rates of COVID-19 deaths.

Table 1
Regional (state) data relative to the impact of COVID-19 on the Brazilian population until November 16th, 2020.

State	Region	Total cases	Cases/100,000 inhabitants	Total deaths	Deaths/100,000 inhabitants
Distrito Federal	CW	220,403	7220.30	3,825	125.31
Goiás	CW	268,997	3780.10	6,060	85.16
Mato Grosso	CW	152,847	4334.58	4,053	114.94
Mato Grosso do Sul	CW	88,111	3136.30	1,693	60.26
Alagoas	NE	92,642	2764.53	2,298	68.57
Bahia	NE	374,721	2509.78	7,967	53.36
Ceará	NE	286,674	3120.13	9,443	102.78
Maranhão	NE	189,705	2666.42	4,187	58.85
Paraíba	NE	139,168	3445.37	3,211	79.49
Pernambuco	NE	171,216	1780.33	8,838	91.9
Piauí	NE	120,466	3671.96	2,541	77.45
Rio Grande do Norte	NE	84,021	2377.39	2,652	75.04
Sergipe	NE	86,146	3714.74	2,260	97.45
Acre	NO	33,295	3722.32	708	79.15
Amapá	NO	169,552	4029.55	4,699	111.68
Amazonas	NO	55,156	6400.29	780	90.51
Pará	NO	262,345	3018.67	6,838	78.68
Rondônia	NO	75,468	4200.93	1,503	83.66
Roraima	NO	60,579	9597.72	706	111.85
Tocantins	NO	78,169	4915.52	1,137	71.5
Espírito Santo	SE	171,237	4213.45	4,030	99.16
Minas Gerais	SE	383,473	1800.96	9,517	44.7
Rio de Janeiro	SE	327,455	1885.59	21,301	122.66
São Paulo	SE	1,169,377	2526.23	40,576	87.66
Paraná	SO	240,783	2090.70	5,728	49.74

State	Region	Total cases	Cases/100,000 inhabitants	Total deaths	Deaths/100,000 inhabitants
Rio Grande do Sul	SO	279,872	2450.08	6,243	54.65
Santa Catarina	SO	297,400	4100.65	3,330	45.92
CW: Central-West; NE: Northeast; NO: North; SE: Southeast; SO: South.					

Regional Distribution Of Most Frequent Hla Haplotypes And Alleles

The five most frequent HLA-A ~ B ~ DRB1 haplotypes in Brazil are presented in Table 2. The most frequent haplotype was HLA-A*01 ~ B*08 ~ DRB1*03, with a frequency ranging from 1.2% in Pará, Maranhão and Amapá, states located in the North (Pará, Amapá) and Northeast (Maranhão) of the country, to 3.1% in Rio Grande do Sul and Santa Catarina (both states located in the South region). Rio Grande do Sul and Santa Catarina had a high outlier frequency for haplotypes #1 (HLA-A*01 ~ B*08 ~ DRB1*03) and #3 (HLA-A*03 ~ B*07 ~ DRB1*15), while Santa Catarina and Rio Grande do Norte were high outliers for haplotype #4 (HLA-A*02 ~ B*44 ~ DRB1*04), and Amapá had a low outlier frequency for haplotype #5 (HLA-A*33 ~ B*14 ~ DRB1*01) (Fig. 2).

Considering HLA alleles, frequencies of COVID-19 deaths per state and per cities with at least 50 deaths are presented in Supplementary file 1. Allele frequencies did not deviate from HWE expectations. The most frequent alleles in Brazil were HLA-A*02, HLA-B*35 and HLA-DRB1*13. Boxplots with allele frequencies per state are presented in Supplementary file 2. It is possible to observe high and low outliers, mostly in states from the South of Brazil, such as Santa Catarina and Rio Grande do Sul, or by states from the North of Brazil, such as Amapá and Amazonas.

Table 2
Frequencies (%) of the five most common HLA haplotypes observed in the Brazilian population.

State	Region	#1 HLA-A*01 ~ B*08 ~DRB1*03	#2 HLA-A*29 ~ B*44 ~DRB1*07	#3 HLA-A*03 ~ B*07 ~DRB1*15	#4 HLA-A*02 ~ B*44 ~DRB1*04	#5 HLA-A*33 ~ B*14 ~DRB1*01
Distrito Federal	CW	1,97	1,53	0,97	0,83	0,92
Goiás	CW	2,16	1,69	0,81	0,90	0,76
Mato Grosso	CW	2,13	1,38	0,95	0,87	0,80
Mato Grosso do Sul	CW	2,03	1,39	0,97	0,86	0,73
Alagoas	NE	1,71	1,29	0,98	0,95	0,93
Bahia	NE	1,50	1,19	0,77	0,87	0,82
Ceará	NE	1,73	1,55	0,86	0,75	0,85
Maranhão	NE	1,27	1,25	0,75	0,79	0,74
Paraíba	NE	1,53	1,49	1,21	1,01	0,80
Pernambuco	NE	1,39	1,28	0,96	0,82	0,91
Piauí	NE	1,52	1,38	0,90	0,77	0,85
Rio Grande do Norte	NE	1,71	1,42	0,90	1,17	0,62
Sergipe	NE	1,65	1,22	0,61	0,87	0,59
Acre	NO	1,67	1,41	0,81	0,76	0,72
Amazonas	NO	1,30	1,26	0,61	0,65	0,79
Amapá	NO	1,29	1,31	0,63	0,77	0,46
Pará	NO	1,26	1,29	0,68	0,76	0,60
Rondônia	NO	2,05	1,39	1,01	0,84	0,85
Rorâima	NO	1,52	1,25	0,77	0,67	0,80
Tocantins	NO	1,61	1,58	0,87	0,82	0,75
Espírito Santo	SE	2,12	1,39	1,02	0,79	0,80
Minas Gerais	SE	2,11	1,65	0,89	0,96	0,85
Rio de Janeiro	SE	1,91	1,54	1,03	0,93	0,83

State	Region	#1 HLA-A*01 ~ B*08 ~DRB1*03	#2 HLA-A*29 ~ B*44 ~DRB1*07	#3 HLA-A*03 ~ B*07 ~DRB1*15	#4 HLA-A*02 ~ B*44 ~DRB1*04	#5 HLA-A*33 ~ B*14 ~DRB1*01
São Paulo	SE	2,18	1,46	0,93	0,74	0,83
Paraná	SO	2,69	1,39	1,27	0,75	0,79
Rio Grande do Sul	SO	3,16	1,64	1,48	0,98	0,61
Santa Catarina	SO	3,13	1,72	1,66	1,18	0,70
CW: Central-West; NE: Northeast; NO: North; SE: Southeast; SO: South.						

Correlation between HLA haplotype frequency and COVID-19 incidence and mortality

Spearman correlation coefficient was calculated to test if the regional COVID-19 incidence and mortality correlated with any of the five most frequent haplotypes in Brazilian population. Table 3 present the results for the first scenario, where no significant correlation was found between COVID-19 rates and haplotype frequency among the 26 states and the Federal District. COVID-19 rates and correlations with HLA haplotype frequencies among cities with at least 50 deaths are presented in Table 4. In total, 377 cities were included in this analysis, where haplotypes #1 (HLA-A*01 ~ B*08 ~ DRB1*03), #2 (HLA-A*29 ~ B*44 ~ DRB1*07) and #4 (HLA-A*02 ~ B*44 ~ DRB1*04) presented negative correlation between haplotype frequency and rates of COVID-19 deaths (-0.322, -0.193, -0.134, respectively). Finally, in the third scenario, we analyzed cities with at least 50 COVID-19 deaths in the São Paulo state. No significant correlation was observed in the 90 cities included in the analysis (Table 5).

Table 3

Bivariate correlation analysis among regional haplotypes estimated frequency in the population and COVID-19 incidence and mortality in Brazilian states.

N° cases/100,000 inhabitants (until November 16th, 2020)		
Haplotypes	**Correlation**	**P**
A*01 ~ B*08 ~ DRB1*03	-0.064	0.751
A*29 ~ B*44 ~ DRB1*07	-0.06	0.765
A*03 ~ B*07 ~ DRB1*15	-0.203	0.311
A*02 ~ B*44 ~ DRB1*04	-0.274	0.167
A*33 ~ B*14 ~ DRB1*01	-0.103	0.611
N° deaths/100,000 inhabitants (until November 16th, 2020)		
Haplotypes	**Correlation**	**P**
A*01 ~ B*08 ~ DRB1*03	-0.153	0.445
A*29 ~ B*44 ~ DRB1*07	-0.153	0.447
A*03 ~ B*07 ~ DRB1*15	-0.188	0.348
A*02 ~ B*44 ~ DRB1*04	-0.345	0.078
A*33 ~ B*14 ~ DRB1*01	0.194	0.332
Correlation = Spearman Correlation Coefficient (*rho*).		

Table 4
Bivariate correlation analysis among regional haplotypes estimated frequency in the population and COVID-19 incidence and mortality in Brazilian cities with at least 50 deaths.

N° cases/100,000 inhabitants (until October 24th, 2020)		
Haplotypes	**Correlation**	**P**
A*01 ~ B*08 ~ DRB1*03	-0.027	0.603
A*29 ~ B*44 ~ DRB1*07	0.111	0.031
A*03 ~ B*07 ~ DRB1*15	0.009	0.859
A*02 ~ B*44 ~ DRB1*04	0.084	0.103
A*33 ~ B*14 ~ DRB1*01	-0.041	0.425
N° deaths/100,000 inhabitants (until October 24th, 2020)		
Haplotypes	**Correlation**	**P**
A*01 ~ B*08 ~ DRB1*03	-0.322	< 0.001*
A*29 ~ B*44 ~ DRB1*07	-0.193	< 0.001*
A*03 ~ B*07 ~ DRB1*15	-0.110	0.033
A*02 ~ B*44 ~ DRB1*04	-0.134	0.009*
A*33 ~ B*14 ~ DRB1*01	0.015	0.772
Correlation = Spearman Correlation Coefficient (*rho*). *Remained significant after FDR correction		

Table 5
Bivariate correlation analysis among regional haplotypes estimated frequency in the population and COVID-19 incidence and mortality in cities with at least 50 deaths in São Paulo.

N° cases/100,000 inhabitants (until October 24th, 2020)		
Haplotypes	**Correlation**	**P**
A*01 ~ B*08 ~ DRB1*03	0.011	0.917
A*29 ~ B*44 ~ DRB1*07	0.009	0.935
A*03 ~ B*07 ~ DRB1*15	-0.056	0.603
A*02 ~ B*44 ~ DRB1*04	0.044	0.68
A*33 ~ B*14 ~ DRB1*01	0.011	0.918
N° deaths/100,000 inhabitants (until October 24th, 2020)		
Haplotypes	**Correlation**	**P**
A*01 ~ B*08 ~ DRB1*03	-0.184	0.083
A*29 ~ B*44 ~ DRB1*07	-0.149	0.16
A*03 ~ B*07 ~ DRB1*15	0.118	0.268
A*02 ~ B*44 ~ DRB1*04	0.074	0.488
A*33 ~ B*14 ~ DRB1*01	-0.025	0.817
Correlation = Spearman Correlation Coefficient (*rho*).		

Correlation between HLA allele frequency and COVID-19 incidence and mortality

Considering HLA allele frequency (at low resolution level or allelic group), there was no significant correlation with COVID-19 rates considering the 26 states and the Federal District. When the analysis included only Brazilian cities with at least 50 COVID-19 deaths, considering both positive and negative correlations, we found six HLA-A, four HLA-B and two HLA-DRB1 alleles significantly correlated with COVID-19 incidence, and 15 HLA-A, 20 HLA-B and seven HLA-DRB1 alleles significantly correlated with COVID-19 mortality. Finally, considering cities from São Paulo state with at least 50 COVID-19 deaths, we found three HLA-A, seven HLA-B and four HLA-DRB1 alleles that were significantly associated with COVID-19 incidence, while five HLA-A alleles were associated with COVID-19 mortality. Correlation values for all alleles are presented in Supplementary file 3.

Discussion

In the present study, we investigated the potential correlation between HLA allele and haplotype frequencies and the different regional distribution of COVID-19 incidence and mortality in Brazil. We

included HLA data at low resolution of 4,148,713 donors from REDOME and COVID-19 data registered until November 16th, 2020. No correlation between haplotype frequencies and COVID-19 rates was found when we analyzed data from the 26 states and Federal District, or when we analyzed data from the 90 cities with at least 50 deaths registered in São Paulo state (Tables 3 and 5). When analyzing data from cities with at least 50 deaths registered in the entire country (377), a significant negative correlation (suggestive of protection) was observed between COVID-19 mortality and haplotypes #1 (HLA-A*01 ~ B*08 ~ DRB1*03), #2 (HLA-A*29 ~ B*44 ~ DRB1*07) and #4 (HLA-A*02 ~ B*44 ~ DRB1*04), with haplotype #1 showing a larger effect. Recently, Pisanti *et al*. (2020) found significant correlation between the two most frequent haplotypes in the Italian population with both COVID-19 incidence and mortality [20]. In their study, HLA data at high resolution of 104,135 donors from Italian Bone Marrow Donors Registry (IBMDR) and COVID-19 mortality and incidence registered until May 24th, 2020 were included. The haplotype HLA-A*01:01g ~ B*08:01g ~ C*07:01g ~ DRB1*03:01g showed a positive correlation (suggestive of susceptibility) and HLA-A*02.01g ~ B*18.01g ~ C*07.01g ~ DRB1*11.04g showed negative correlation (suggestive of protection) with both COVID-19 incidence and mortality [20]. At low resolution, HLA-A*01 ~ B*08 ~ DRB1*03 is the most frequent HLA haplotype in both Brazil and Italy. However, while this haplotype has been positively correlated with COVID-19 incidence and mortality in Italy, suggesting increased susceptibility, we found, depending on the analyzed dataset, no correlation or a negative correlation with mortality in Brazil, which might suggest protection against COVID-19. One possibility is that HLA *versus* environment interaction is different in Italy and Brazil resulting in opposite associations between HLA variation and COVID-19 risk, in a context-dependent manner. Another possibility is that the positive association in the Italian population reflects regional genetic structure, as both COVID-19 rates [20] and genetic population structure in Italy [26, 27] show a North-South gradient. Another important point to highlight is that none of the studies took into consideration the SARS-CoV-2 variants, and it becomes increasingly important to consider this point in future studies.

In general, we observed that the four most common HLA haplotypes had higher frequencies in the South region and lower frequencies in the North of Brazil (Table 2 and Fig. 2). For example, the most frequent haplotype (HLA-A*-01 ~ B*08 ~ DRB1*03) ranged from 1.26% in Pará (North) to 3.16% in Rio Grande do Sul (South) (Fig. 2). These regional differences reflect the different contributions of Native American, European, and African populations across the country after a long history of colonization and immigration. From a broad perspective, the South region has more European influence; the Northeast region is characterized by more African influence, while in the North region the Native American influence is comparatively more pronounced [28].

In addition to the study that evaluated HLA haplotypes in the Italian population to the differential regional incidence of COVID-19, two bioinformatic studies predicted the binding affinity between HLA alleles and SARS-CoV-2 antigenic peptides [19, 29]. For HLA-A, it has been predicted that -A*02 had the best binding affinity, resulting in a more favorable immune response, while -A*25 had the weakest binding affinity, resulting in a less favorable response. Recently, Shkurnikov *et al*. (2021) [30] performed HLA class 1 genotyping through next-generation sequencing in deceased patients with COVID-19. In their analysis, HLA-A*01:01 allele (especially in homozygosis) was associated with high risk to present SARS-CoV-2,

while HLA-A*02:01 and HLA-A*03:01 mainly contributed to low risk. In our study, considering the scenario #2, for which we have more statistical power, we found significant negative correlations for HLA-A*01 (-0.276), HLA-A*02 (-0.175), HLA-A*03 (-0.302) and HLA-A*25 (-0.164) (Supplementary file 3). As for HLA-B alleles, both bioinformatics studies suggested that alleles -B*46:01 and -B*15:03 behaved as low and high affinity ligands against COVID-19 peptides, and Barquera *et al*. (2000) further characterized alleles -B*44:06, -B*51:07, -B*08:03, and B*52:01 as weak binders, and -B*35:10, as well as other molecules of the -B*15 lineage, as strong binders [29]. We found a significant negative correlation for -B*35 (-0.160) with COVID-19 mortality, but, unexpectedly, a significant positive correlation between -B*15 and COVID-19 mortality (0.280) (Supplementary file 3). Finally, for HLA-DRB1, Barquera *et al*. (2020) suggested that the alleles -DRB1*01:01 and -DRB1*10:01 were the strongest binders, while - DRB1*03:02 and -DRB1*03:03 were the weakest binders [29]. In line with these results, we found a significant negative correlation for allele -DRB1*01 (-0.246) and mortality (Supplementary file 3). Overall, our analysis provided mixed support for the inferences made by bioinformatics models (Supplementary file 3). This may be related to the complexity of the immune response against SARS-CoV-2, which probably depends on many genetic and environmental factors. For example, variants in other genes related to the immune response, like genes that code for inflammatory factors and interleukin 6 (IL-6), and genes related to SARS-CoV-2 entry into the cell may also influence the disease course. Recent studies have already correlated and reviewed the role of variants in *IL6* [31], *ACE2* [32], and *TMPRSS2* [33] genes in COVID-19 disease.

Together with genetic and immune system variation, environmental and social disparities among Brazilian regions may contribute to the differential burden of COVID-19, affecting disproportionally individuals carrying genetic factors of susceptibility and/or the most vulnerable people regarding social assistance. For instance, at the time we were collecting our data, Rio de Janeiro and Minas Gerais, both in the Southeast region, had about 1800 cases/100,000 inhabitants. However, the mortality rate was almost 3 times higher in Rio de Janeiro (122.66/100,000 inhabitants) than in Minas Gerais (44.7/100,000 inhabitants) (Table 1). Besides Minas Gerais, only Paraná and Santa Catarina, both in the South region, maintained mortality rates below 50/100,000 inhabitants, while higher mortality rates were found in states located in the Central-West and North regions (Table 1). These numbers should be taken with caution. The number of reported cases depends on the number and type of diagnostic tests performed by each city and state, and with limited testing it is unlikely that asymptomatic subjects would have been diagnosed for SARS-CoV-2 infection. The number of COVID-19 related deaths could be more reliable since, during the period analyzed in this study, the Brazilian health system was able to provide medical assistance for most people showing COVID-19 symptoms. Our study included only confirmed deaths by COVID-19 in our analysis, however, when considering deaths due to SARS reported in DATASUS (Department of Informatics of the Unified Healthcare System of Brazil), it was expected an average underreporting of 40% [34]. Therefore, it is possible that the number of cases have been underestimated in our analysis. The mortality rate may also vary according to the coverage of local health systems, number of hospital beds available, and measures of social isolation, for example. It is important to highlight that in Brazil there was no unified policy regarding COVID-19 protocols, resulting in huge differences in how different state governors and city mayors managed COVID-19 locally. Distinguishing

among genetic, social and environmental variability is essential for building an efficient way to prevent, control and understand disease outbreaks [35, 36]. Our study focused on genetic variation of the HLA system and is an important initial step in the understanding of COVID-19 dispersion and behavior in Brazil.

An important limitation of this study is that we used HLA data from bone marrow donors instead of directly genotyping individuals affected by COVID-19. However, this approach has some advantages. Bone marrow donor registries usually include very large sample sizes and a wide geographic coverage, as illustrated by the REDOME registry. In addition, because several countries maintain such large banks, statistically significant associations could reveal regions or populations in higher genetic risk for COVID-19, thus representing an additional tool for health policymakers in the fight against COVID-19.

Conclusions

In summary, we investigated if there was a correlation between HLA allele and haplotype frequencies and the different regional distribution of COVID-19 incidence and mortality in the country. No correlation between HLA haplotypes and COVID-19 rates was found when we analyzed data from the 26 states and Federal District, or when we analyzed data from the 90 cities with at least 50 deaths registered in São Paulo (Tables 3 and 5). However, when analyzing data from cities with at least 50 deaths registered in the entire country (n = 377), we found a significant negative correlation (suggestive of protection) between COVID-19 mortality and haplotypes #1 (HLA-A*01 ~ B*08 ~ DRB1*03), #2 (HLA-A*29 ~ B*44 ~ DRB1*07) and #4 (HLA-A*02 ~ B*44 ~ DRB1*04), with haplotype #1 showing a larger effect.

Abbreviations

COVID-19
Coronavirus Disease 2019
DATASUS
Department of Informatics of the Unified Healthcare System of Brazil
HWE
Hardy-Weinberg Equilibrium
HLA
Human Leukocyte Antigen
IBMDR
Italian Bone Marrow Donors Registry
MHC
Main Histocompatibility Complex
REDOME
Registro de Doadores de Medula Óssea
TCR
T Cell Receptors

WHO
World Health Organization

Declarations

Ethics approval and consent to participate

This study was approved by the Ethics Committee in Research of Hospital de Clínicas de Porto Alegre, Brazil (CAAE 34701720300005327, GPPG 2020-0361), and all methods were carried out in accordance with local guidelines and regulations.

Consent for publication

Not applicable.

Availability of data and materials

COVID-19 data used in this study are freely available from the source cited. HLA data are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This study was funded by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and Fundo de Incentivo a Pesquisa do Hospital de Clínicas de Porto Alegre (FIPE-HCPA), number 2020-0361.

Authors' contributions

JAB participated on study design, obtained HLA data, analyzed and interpreted HLA and COVID-19 data and was a major contributor in writing the manuscript; FSLV participated on study design, analyzed and interpreted HLA and COVID-19 data and participated in writing and reviewing the manuscript; NJRF was a major contributor on statistical analysis and participated in manuscript writing and revision; LS was responsible for graph and map design and contributed to data interpretation; MB participated on study design and statistical analysis; MZO participated on COVID-19 data collection and map design; TFA participated on COVID-19 data collection and writing and reviewing the manuscript; LCMSP participated on study design and availability of HLA data; JABC participated on study design, writing and reviewing the manuscript; LS-F participated in writing and reviewing the manuscript; PA-P participated in writing and reviewing the manuscript, and CR was responsible for study design and approval, participated in analysis and interpretation of HLA and COVID-19 data, and was a major contributor in writing and reviewing the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We would like to thank FIOCRUZ and REDOME for maintenance and availability of COVID-19 and HLA data, respectively.

References

1. Statement on the second meeting of the International Health Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus (2019-nCoV). Available at: https://www.who.int/news-room/detail/30-01-2020-statement-on-the-second-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-(2019-ncov).
2. CORONAVIRUS BRASIL. Available at: https://covid.saude.gov.br/; 2020.
3. Roser M, Ritchie H, Ortiz-Ospina E, Hasell J. Coronavirus Pandemic (COVID-19) Available at: OurWorldInData.org. 2020. Retrieved from: https://ourworldindata.org/coronavirus [Online Resource]
4. Richardson S, Hirsch JS, Narasimhan M, Crawford JM, McGinn T, Davidson KW, Barnaby DP, Becker LB, Chelico JD, Cohen SL, et al. Presenting Characteristics, Comorbidities, and Outcomes Among 5700 Patients Hospitalized With COVID-19 in the New York City Area. JAMA 2020; 26;323(20):2052-2059. doi: 10.1001/jama.2020.6775.
5. Li Z, Liu T, Yang N, Han D, Mi X, Li Y, Liu K, Vuylsteke A, Xiang H, Guo X. Neurological manifestations of patients with COVID-19: potential routes of SARS-CoV-2 neuroinvasion from the periphery to the brain. Front Med 2020; 14(5):533-541. doi: 10.1007/s11684-020-0786-5.
6. Zhou Z, Zhao N, Shu Y, Han S, Chen B, Shu X: Effect of gastrointestinal symptoms on patients infected with COVID-19. Gastroenterology 2020;158(8):2294-2297. doi: 10.1053/j.gastro.2020.03.020.
7. Ovsyannikova IG, Dhiman N, Jacobson RM, Poland GA. Human leukocyte antigen polymorphisms: variable humoral immune responses to viral vaccines. Expert Rev Vaccines 2006, 5:33-43. doi: 10.1586/14760584.5.1.33.
8. Jiang M, Li Y, Han M, Wang Z, Zhang Y, Du X. Recurrent PCR positivity after hospital discharge of people with coronavirus disease 2019 (COVID-19). J Infect 2020;81(1):147-178. doi: 10.1016/j.jinf.2020.03.024.
9. Ovsyannikova IG, Dhiman N, Haralambieva IH, Vierkant RA, O'Byrne MM, Jacobson RM, Poland GA. Rubella vaccine-induced cellular immunity: evidence of associations with polymorphisms in the Toll-like, vitamin A and D receptors, and innate immune response genes. Hum Genet 2010 ;127(2):207-21. doi: 10.1007/s00439-009-0763-1.
10. Dendrou CA, Petersen J, Rossjohn J, Fugger L: HLA variation and disease. Nat Rev Immunol 2018;18:325-339. doi: 10.1038/nri.2017.143.
11. Orrù V, Steri M, Sole G, Sidore C, Virdis F, Dei M, Lai S, Zoledziewska M, Busonero F, Mulas A, et al. Genetic variants regulating immune cell levels in health and disease. Cell 2013;155:242-256. doi:

10.1016/j.cell.2013.08.041

12. Saghazadeh A, Rezaei N: Introductory Chapter: Immunogenetics. Available at: https://www.intechopen.com/books/immunogenetics/introductory-chapter-immunogenetics. doi: 10.5772/intechopen.85505
13. Gonzalez-Galarza FF, McCabe A, Santos EJ, Jones J, Takeshita LY, Ortega-Rivera ND, Del Cid-Pavon GM, Ramsbottom K, Ghattaoraya GS, Alfirevic A, Middleton D, Jones AR. Allele frequency net database (AFND) 2020 update: gold-standard data classification, open access genotype data and new query tools. Nucleic Acid Research 2020;48:D783-8. doi: 10.1093/nar/gkz1029.
14. Kehdy FS, Gouveia MH, Machado M, Magalhães WC, Horimoto AR, Horta BL, Moreira RG, Leal TP, Scliar MO, Soares-Souza GB, Rodrigues-Soares F, Araújo GS, Zamudio R, Sant Anna HP, Santos HC, Duarte NE, Fiaccone RL, Figueiredo CA, Silva TM, Costa GN, Beleza S, Berg DE, Cabrera L, Debortoli G, Duarte D, Ghirotto S, Gilman RH, Gonçalves VF, Marrero AR, Muniz YC, Weissensteiner H, Yeager M, Rodrigues LC, Barreto ML, Lima-Costa MF, Pereira AC, Rodrigues MR, Tarazona-Santos E; Brazilian EPIGEN Project Consortium. Origin and dynamics of admixture in Brazilians and its effect on the pattern of deleterious mutations. Proc Natl Acad Sci USA. 2015;112(28):8696-701. doi: 10.1073/pnas.1504447112.
15. Lin M, Tseng HK, Trejaut JA, Lee HL, Loo JH, Chu CC, Chen PJ, Su YW, Lim KH, Tsai ZU, et al. Association of HLA class I with severe acute respiratory syndrome coronavirus infection. BMC Med Genet 2003;12;4:9. doi: 10.1186/1471-2350-4-9.
16. Wang SF, Chen KH, Chen M, Li WY, Chen YJ, Tsao CH, Yen MY, Huang JC, Chen YM. Human-leukocyte antigen class I Cw 1502 and class II DR 0301 genotypes are associated with resistance to severe acute respiratory syndrome (SARS) infection. Viral Immunol 2011;24:421-426. doi: 10.1089/vim.2011.0024.
17. Ng MH, Lau KM, Li L, Cheng SH, Chan WY, Hui PK, Zee B, Leung CB, Sung JJ. Association of human-leukocyte-antigen class I (B*0703) and class II (DRB1*0301) genotypes with susceptibility and resistance to the development of severe acute respiratory syndrome. J Infect Dis 2004;190:515-518. doi: 10.1086/421523.
18. Ng MH, Cheng SH, Lau KM, Leung GM, Khoo US, Zee BC, Sung JJ. Immunogenetics in SARS: a case-control study. Hong Kong Med J 2010;16:29-33.
19. Nguyen A, David JK, Maden SK, Wood MA, Weeder BR, Nellore A, Thompson RF. Human Leukocyte Antigen Susceptibility Map for Severe Acute Respiratory Syndrome Coronavirus 2. J Virol. 2020;16;94(13):e00510-20. doi: 10.1128/JVI.00510-20.
20. Pisanti S, Deelen J, Gallina AM, Caputo M, Citro M, Abate M, Sacchi N, Vecchione C, Martinelli R. Correlation of the two most frequent HLA haplotypes in the Italian population to the differential regional incidence of Covid-19. J Transl Med 2020;18:352. doi: 10.1186/s12967-020-02515-5.
21. Instituto de Comunicação e Informação Científica e Tecnológica em Saúde (ICICT). MonitoraCovid-19. Rio de Janeiro, 2020. Available at: https://bigdata-covid19.icict.fiocruz.br/. Accessed at: November 16th, 2020.

22. Buhler S, Nunes JM, Nicoloso G, Tiercy JM, Sanchez-Mazas A. The heterogeneous HLA genetic makeup of the Swiss population. PLoS ONE. 2012;7:e41400. doi: 10.1371/journal.pone.0041400.
23. Nunes JM. Using UNIFORMAT and GENE[RATE] to analyze data with ambiguities in population genetics. Evol Bioinfor. 2015;2:19-26. doi: 10.4137/EBO.S32415.
24. Boquett JA, Nunes JM, Buhler S, de Oliveira MZ, Jobim LF, Jobim M, Fagundes NJ, Schüler-Faccini L, Sanchez-Mazas A. 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.
25. Torres L, da Silva Bouzas LF, Almada A, de Sobrino Porto LCM, Abdelhay E. Distribution of HLA-A, -B and -DRB1 antigenic groups and haplotypes from the Brazilian bone marrow donor registry (REDOME). Hum Immunol. 2017;78(10):602-609. doi: 10.1016/j.humimm.2017.08.002.
26. Di Gaetano C, Voglino F, Guarrera S, et al. An overview of the genetic structure within the Italian population from genome-wide data. PLoS One. 2012;7(9):e43759. doi:10.1371/journal.pone.0043759.
27. Raveane A, Aneli S, Montinaro F, et al. Population structure of modern-day Italians reveals patterns of ancient and archaic ancestries in Southern Europe. Sci Adv. 2019;5(9):eaaw3492. doi:10.1126/sciadv.aaw3492.
28. IBGE, Centro de Documentação e Disseminação de Informações. Brasil: 500 anos de povoamento. Rio de Janeiro: IBGE, 2007. Available at: https://biblioteca.ibge.gov.br/visualizacao/livros/liv6687.pdf.
29. Barquera R, Collen E, Di D, Buhler S, Teixeira J, Llamas B, Nunes JM, Sanchez-Mazas A. Binding affinities of 438 HLA proteins to complete proteomes of seven pandemic viruses and distributions of strongest and weakest HLA peptide binders in populations worldwide. HLA. 2020;96(3):277-298. doi: 10.1111/tan.13956.
30. Shkurnikov M, Nersisyan S, Jankevic T, Galatenko A, Gordeev I, Vechorko V and Tonevitsky A. Association of HLA Class I Genotypes With Severity of Coronavirus Disease-19. Front. Immunol. 2021;12:641900. doi: 10.3389/fimmu.2021.641900
31. Ovsyannikova IG, Haralambieva IH, Crooke SN, Poland GA, Kennedy RB. The role of host genetics in the immune response to SARS-CoV-2 and COVID-19 susceptibility and severity. Immunol Rev. 2020;296(1):205-219. doi: 10.1111/imr.12897.
32. Benetti E, Tita R, Spiga O, Ciolfi A, Birolo G, Bruselles A, et al. ACE2 gene variants may underlie interindividual variability and susceptibility to COVID-19 in the Italian population. Eur J Hum Genet. 2020;28(11):1602-1614. doi: 10.1038/s41431-020-0691-z.
33. Hou Y, Zhao J, Martin W, Kallianpur A, Chung MK, Jehi L, et al. New insights into genetic susceptibility of COVID-19: an ACE2 and TMPRSS2 polymorphism analysis. BMC Med. 2020;15;18(1):216. doi: 10.1186/s12916-020-01673-z.
34. Veiga E Silva L, de Andrade Abi Harb MDP, Teixeira Barbosa Dos Santos AM, de Mattos Teixeira CA, Macedo Gomes VH, Silva Cardoso EH, S da Silva M, Vijaykumar NL, Venâncio Carvalho S, Ponce de Leon Ferreira de Carvalho A, Lisboa Frances CR. COVID-19 Mortality Underreporting in Brazil:

Analysis of Data From Government Internet Portals. J Med Internet Res. 2020; Aug 18;22(8):e21413. doi: 10.2196/21413

35. Floss M, Barros E, Bressel M, Hacon S, Stein A, Sirena S, et al. Lancet Countdown 2018 Report: Briefing for Brazilian Policymakers. Lancet Countdown. 2018;1:19. Available at: http://www.lancetcountdown.org/media/1417/2018-lancet-countdown-policy-brief-brazil.pdf.
36. Zhao Q, Li S, Coelho MSZS, Saldiva PHN, Hu K, Huxley RR, Abramson MJ, Guo. The association between heatwaves and risk of hospitalization in Brazil: A nationwide time series study between 2000 and 2015. PLoS Med. 2019;16(2):e1002753. doi:10.1371/journal.pmed.1002753.

Figures

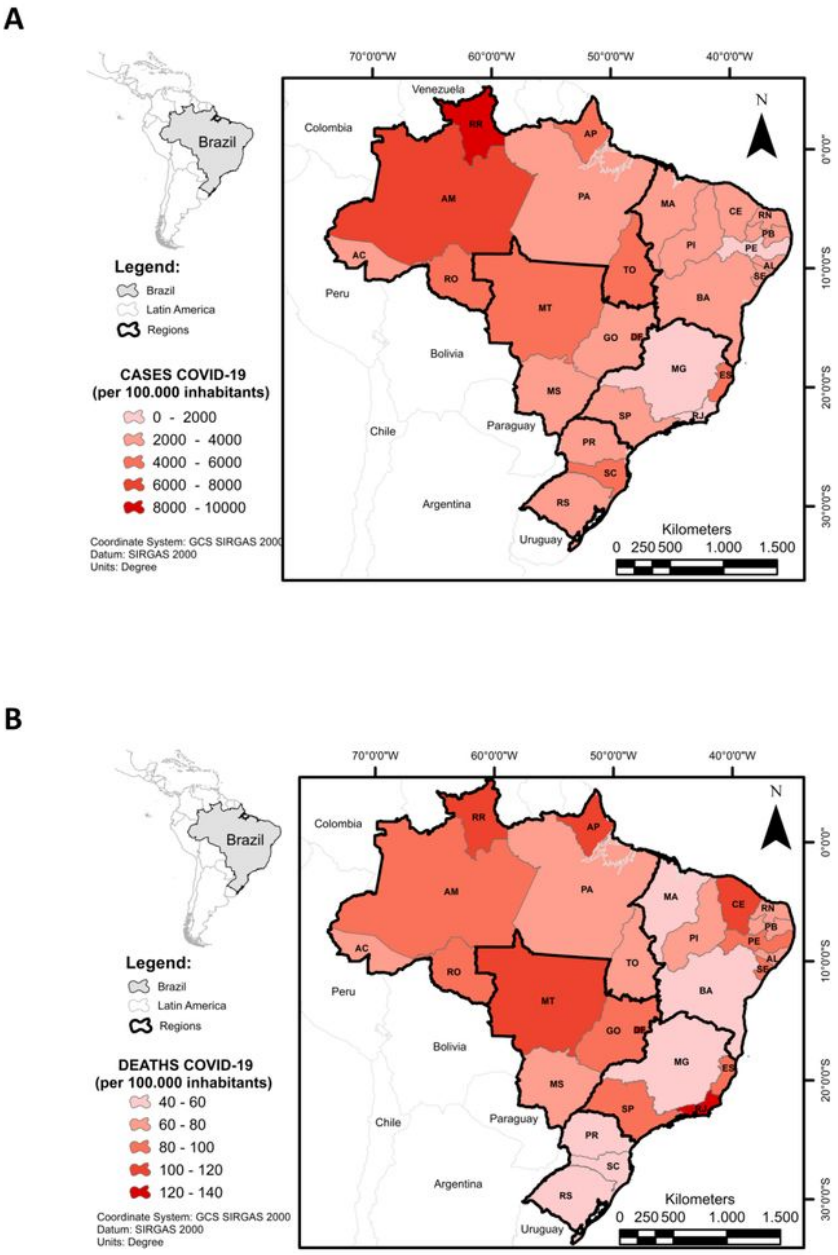

Figure 1

A) COVID-19 cases per 100,000 inhabitants in different Brazilian regions and states. B) COVID-19 deaths per 100,000 inhabitants in different Brazilian regions and states. Data were collected from February, 2020, until November 16th, 2020. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square

concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

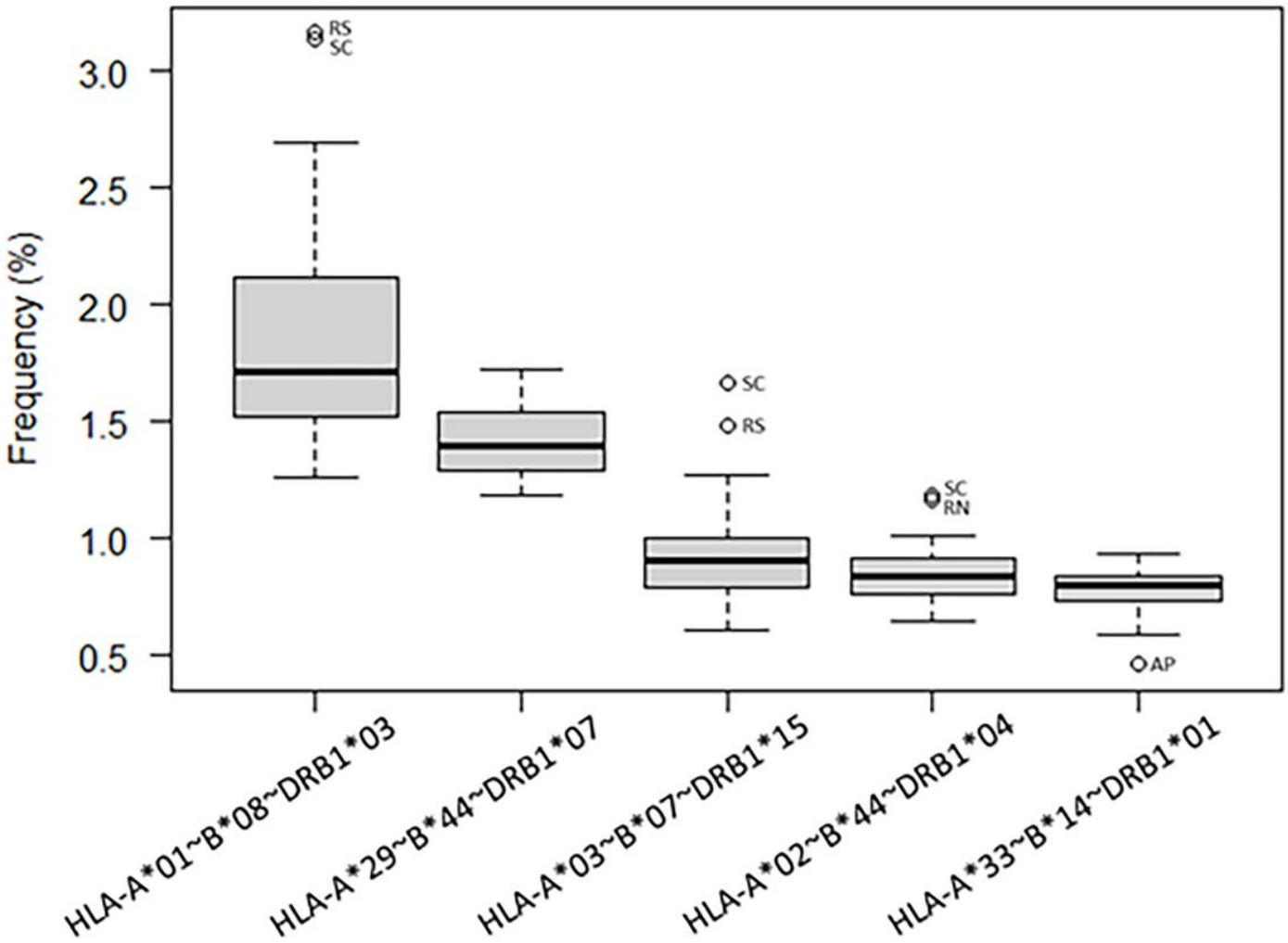

Figure 2

Boxplot representing the five most common haplotypes observed in the Brazilian population. Outliers are represented by the respective state's abbreviation. RS: Rio Grande do Sul; SC: Santa Catarina; RN: Rio Grande do Norte; AP: Amapá.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementaryfile1.xlsx
- Supplementaryfile2.docx
- Supplementaryfile3.docx
- letterapcwaiver.docx