Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Review

Risk HLA alleles in South America and potential new epitopes for SARS-CoV2

Samantha Sáenz Hinojosa, Vanessa Romero *

Universidad San Francisco de Quito, School of Medicine, Ecuador

ARTICLE INFO

ABSTRACT

HLA alleles are associated with the body's response to infection and the regulation of the immune system. HLA alleles have been reported to be involved in response to viral infections such as SARS-CoV2. Our study reviews the HLA alleles associated with protection or susceptibility to SARS-CoV2 and the prevalence of these HLA alleles in South America. Previous studies on HLA and SARS-CoV2 infection reported that HLA-A*02:02, HLA-B*15:03, and HLA-C*12:03 are protective; while HLA-A*25:01, HLA-B*46:01, and HLA-C*01:02 increase susceptibility. We identified that these alleles are not frequent in South America, confirmed that the spike protein is the most immunogenic protein of SARS-CoV2, and detected new immunogenic epitopes that bound to protective HLA alleles and to HLA alleles common in South America (binding score > 0.90). These could be used as vaccine targets.

© 2021 Published by Elsevier Inc. on behalf of American Society for Histocompatibility and Immunogenetics.

Contents

1. Introduction

Coronavirus has been related to many respiratory zoonotic infections and is responsible for three documented pandemics in humans: Severe Acute Respiratory Syndrome (SARS-CoV) in 2002, the Middle East Respiratory Syndrome (MERS-CoV) in 2012, and the current SARS-CoV2, which began at the end of 2019, has been responsible so far for over 58 million infected [1].

These three viruses are beta coronaviruses and it has been reported that 38% of HLA-class I epitopes are conserved among them. Some of the most immunogenic regions of these viruses are part of the structural proteins, including the spike glycoprotein, membrane protein, and nucleocapsid phosphoprotein. Additionally, coronaviruses use the spike glycoprotein for neutralizing antibodies and mediate membrane fusion and virus entry [2].

Nowadays, a variety of studies are being done to understand the interaction between SARS-CoV2 and the host immune system. This will help to develop preventive measures to confront this pandemic, such as vaccines. The constrain in which HLA binds to different regions of the virus influences the adequate immune response and clearance of the virus. It also affects the susceptibility to get infected and, in some cases, the development of complications.

We aimed to understand how the HLA influences susceptibility or protection to SARS-CoV2 infection; therefore, we performed a literature review to find the HLA alleles that provide susceptibility or protection, and to evaluate how these specific HLA alleles bind to the most immunogenic structural regions of SARS-CoV2. Additionally, as most of the studies are limited to Asian populations, we analyze the frequency of these HLA alleles in South America and perform a binding prediction of the most common HLA alleles in this continent to SARS-CoV2 immunogenic regions.

2. Methods

2.1. Literature search strategy

This literature review focuses on the available information on susceptibility or protectiveness of different HLA types against MERS-CoV, SARS-CoV and SARS-CoV2 on the PUBMED database in the English language from 2003 until May 2020. The terms used included 'COVID 19 HLA' (n = 10), 'SARS-CoV-2 HLA' (n = 8), 'Severe SARS HLA' (n = 50), 'MERS-CoV HLA' (n = 9), 'SARS-CoV HLA' (n = 52) and 'HLA SARS-CoV2' (n = 12). The exclusion criteria were: (I) No relation between specific HLA and SARS-CoV, SARS-CoV2, or MERS-CoV; (II) Vaccine development; (III) Not related to humans; (IV) Duplicated articles; (V) Complete article not available; and (VI) Unpublished data.

We performed an additional review to obtain the prevalence of those HLA found to be related to SARS-CoV2 in South America, as well as the most common HLA in this region, from PUBMED database as well as The Allele Frequencies Net Database [3].

2.2. Sequences

SARS-Cov2 Nucleocapsid Phosphoprotein (**YP_009724397.2**), Membrane Glycoprotein (**YP_009724393.1**) obtained from NCBI [4] were used. SARS-Cov2 Spike Protein (**PDB ID: 6VSB**) was obtained from RCSB Protein DataBank [5].

2.3. Epitope and HLA prediction

The Epitope Analysis Resource from the Immune Epitope Database and Analysis Resource (IEDB) [6] was used to predict the epitope and MHC I and MHC II binding, and high and low binders were

Table 1
HLA associations with respiratory diseases related to coronavirus.

Disease	Susceptible	Protective
SARS-CoV2	HLA-B*46:01 [9]	HLA-B*15:03 [9]
	HLA-A*25:01 [9]	HLA-A*02:02 [9]
	HLA-C*01:02 [9]	HLA-C*12:03 [9]
SARS-CoV	HLA-B*07:03 [10]	HLA-DR*03:01 [11]
	HLA-DRB1*1202 [12]	HLA-Cw*15:02 [11]
	HLA-Cw*0801 [13]	
	HLA-A*46:01 [14]	
MERS-CoV	HLA-DRB1*11:01 [15]	
	HLA-DQB1*02:02 [15]	

Next to each HLA allele is the corresponding bibliography.

inferred quantitatively. The length of the peptides is aleatory, and the ANN validation is part of the IEDB. The peptides were reviewed manually to exclude duplication.

First, we calculated the binding prediction using the Prediction Method from IEDB 2020.4 (NetMHCpan EL 4.0)[6] between the HLA alleles found in our systematic review (six), the most common HLA alleles worldwide (four) and the epitopes described by Grifoni et al. [7] which included: five highly immunogenic SARS-CoV regions, six dominant SARS-CoV2 B Cell epitopes and seventeen T Cell epitopes shown to have ≥ 90% identity with SARS-CoV region.

Second, Grifoni et al. [7] reported SARS-CoV2 epitopes that were strongly bound to the most common worldwide HLA. We consider that it was useful to identify the epitopes that were strongly bound to the most protective HLA allele since it can help identify new immunogenic regions. Therefore, we used the nucleotide sequence from SARS-Cov2 nucleocapsid, membrane glycoprotein, and spike protein to calculate the prediction binding towards the most susceptible or protective HLA alleles found on the literature review.

Finally, we calculated the binding prediction between SARS-CoV2 most immunogenic regions and the most common HLA alleles found in South America, from our literature review.

To confirm the consistency of the results, we analyzed the bindings with the additional algorithm: ANN 4.0 [6].

3. Results

3.1. HLA allele susceptibility and protectiveness literature review

SARS-CoV2 shares a similar S protein with over 70% of identity with SARS-CoV [2] and has a 59% nucleotide sequence homology to MERS-CoV [8]. It was reported that 38% of HLA-class I epitopes are conserved between all three viruses, located in different parts of the sequence, including the spike, membrane, and nucleocapsid proteins [8]. Therefore, we scoped, on the PUBMED database, for the association between MERS-CoV, SARS-CoV, and SARS-CoV2 infection and the most susceptible or protective HLA alleles worldwide. For SARS-CoV2, the most protective HLA alleles were HLA-B*15:03 [9], HLA-A*02:02 [9] and HLA-C*12:03 [9] (Table 1).

3.2. Epitopes and HLA binding prediction

Grifoni et al. [7] identified five highly immunogenic SARS-CoV regions, six dominant SARS-CoV2 B Cell epitopes and seventeen SARS-CoV2 T Cell epitopes, which have ≥ 90% similar identity to SARS-CoV regions. We calculated the binding scores between all epitopes individually to the six susceptible/protective HLA alleles (HLA-A*02:02, HLA-A*25:01, HLA-B*15:03, HLA-B*46:01, HLA-C*01:02, HLA-C*12:03) and also to the most common worldwide HLA types described by Grifoni et al., (HLA-A*02:01, HLA-B*58:01, HLA-B*40:01, HLA-A*24:02) [7] (Table 2). We did this to identify if the affinity was higher or lower on SARS-CoV or SARS-

Table 2

MHC I binding prediction > 0.90 score between HLA alleles found to be protective and susceptible to SARS-CoV2 and common worldwide HLA alleles, and five immunogenic SARS-CoV regions, seventeen dominant SARS-CoV 2 T Cell epitopes and six dominant SARS-Cov 2 B Cell epitopes reported by Grifoni et al.

HLA			VIRUS			
Allele[a]	Cell response	P/C	Virus	Peptide	Score	Protein[b]
HLA-A*02:02	Dominant T Cell	P	SARS-CoV2	VLNDILSRL	0.99	S
HLA-B*40:01	Dominant T Cell	C	SARS-CoV2	AEIRASANL	0.97	S
HLA-B*15:03	Dominant T Cell	P	SARS-CoV2	KAYNVTQAF	0.97	N
HLA-C*12:03	Dominant T Cell	P	SARS-CoV2	KAYNVTQAF	0.97	N
HLA-A*02:02	Dominant T Cell	P	SARS-CoV2	LLLDRLNQL	0.96	N
HLA-B*40:01	Immunogenic Epitope	C	SARS-CoV	SELVIGAVI	0.95	M
HLA-B*40:01	Dominant B Cell	C	SARS-CoV2	SELVIGAVI	0.95	M
HLA-B*58:01	Dominant T Cell	C	SARS-CoV2	KAYNVTQAF	0.95	N
HLA-A*02:01	Dominant T Cell	P	SARS-CoV2	VLNDILSRL	0.95	S
HLA-A*02:01	Dominant T Cell	P	SARS-CoV2	LLLDRLNQL	0.94	N
HLA-B*15:03	Immunogenic Epitope	P	SARS-CoV	LQLPQGTTL	0.93	N
HLA-B*15:03	Dominant T Cell	P	SARS-CoV2	LQLPQGTTL	0.93	N
HLA-B*15:03	Dominant B Cell	P	SARS-CoV2	LQLPQGTTL	0.93	N
HLA-B*15:03	Dominant B Cell	P	SARS-CoV2	LQIPFAMQM	0.92	S
HLA-B*40:01	Dominant B Cell	C	SARS-CoV2	SELVIGAVIL	0.91	M
HLA-A*02:02	Dominant T Cell	P	SARS-CoV2	RLNEVAKNL	0.91	S
HLA-A*02:01	Dominant B Cell	P	SARS-CoV2	KLLEQWNLV	0.90	M

Prediction method: IEDB recommended 2020.04 (NetMHCpan EL 4.0) | High Score = good binder.

[a] P, HLA allele found to be protective on the literature; C, HLA allele common worldwide.

[b] M, membrane protein; N, nucleocapsid phosphoprotein; S, spike glycoprotein.

Table 3

Best bindings (score > 0.90) between the most immunogenic regions SARS-CoV2, and the most common worldwide HLA alleles and the three HLA alleles that protect against SARS-CoV2.

Response	Allele	P/C[a]	Start-End	Peptide	Score	Protein[b]	Description
Dominant T Cells	HLA-A*02:02	P	976–984	VLNDILSRL	0.99	S	
	HLA-B*15:03	P	265–274	TKAYNVTQAF	0.97	N	
	HLA-C*12:03	P	265–274	TKAYNVTQAF	0.97	N	
	HLA-A*02:02	P	222–230	LLLDRLNQL	0.96	N	
	HLA-B*15:03	P	159–167	LQLPQGTTL	0.93	N	
	HLA-A*02:02	P	1185–1193	RLNEVAKNL	0.91	S	
	HLA-B*40:01	C	1011–1082	QLIRAAEIRASANLAATK	0.97	S	
	HLA-B*58:01	C	265–274	TKAYNVTQAF	0.95	N	
	HLA-A*02:01	C	976–984	VLNDILSRL	0.95	S	
	HLA-A*02:01	C	222–230	LLLDRLNQL	0.94	N	
Dominant B Cells	HLA-B*15:03	P	153–172	NNNAATVLQLPQGTTLPKGF	0.93	N	
	HLA-B*15:03	P	889–909	FGAGAALQIPFAMQMAYRFNGI	0.92	S	
	HLA-B*40:01	C	132–151	PLLESELVIGAVILRGHLRI	0.95	M	
	HLA-B*40:01	C	132–151	PLLESELVIGAVILRGHLRI	0.91	M	
	HLA-A*02:01	C	43854.00	MADSNGTITVEELKKLLEQWNLVI	0.90	M	
Most immunogenic epitopes	HLA-A*02:02	P	976–984	VLNDILSRL	0.99	S	
	HLA-A*02:02	P	269–277	YLQPRTFLL	0.98	S	
	HLA-B*15:03	P	266–274	KAYNVTQAF	0.97	N	
	HLA-B*15:03	P	305–314	AQFAPSASAF	0.97	N	
	HLA-A*02:02	P	417–425	KIADYNYKL	0.97	S	Novel
	HLA-C*12:03	P	266–274	KAYNVTQAF	0.97	N	
	HLA-A*02:02	P	222–230	LLLDRLNQL	0.96	N	
	HLA-C*12:03	P	37–45	FAYANRNRF	0.96	M	Novel
	HLA-B*15:03	P	557–565	KKFLPFQQF	0.96	S	Novel
	HLA-B*15:03	P	159–167	LQLPQGTTL	0.93	N	
	HLA-B*15:03	P	894–902	LQIPFAMQM	0.92	S	
	HLA-B*15:03	P	408–421	RQIAPGQTGKIADY	0.92	S	Novel
	HLA-B*15:03	P	443–451	SKVGGNYNY	0.91	S	
	HLA-A*02:02	P	1185–1193	RLNEVAKNL	0.91	S	

Prediction method: IEDB recommended 2020.04 (NetMHCpan EL 4.0) | High Score = good binder.

[a] P, HLA allele found to be protective on the literature; C, HLA allele common worldwide.

[b] M, membrane protein; N, nucleocapsid phosphoprotein; S, spike glycoprotein.

CoV2. The top three highest scores or best binders (score > 0.97) are HLA-A*02:02 (protective), HLA-B*40:01 (common) and HLA-B*15:03 (protective). As expected, the dominant SARS-CoV2 T and B cell regions have better bindings compared to SARS-CoV immunogenic regions. We used an additional algorithm to confirm the results: ANN 4.0 [6] (Supplementary Table 1).

SARS-CoV 2′s most immunogenic regions are the spike glyco-protein, membrane protein, and nucleoprotein [7]; and the most

protective HLA alleles are HLA-B*15:03, HLA-A*02:02, and HLA-C*12:03 [9]. We consider that it was useful to identify the epitopes that are strongly bound to the most protective HLA alleles since it can help identify new immunogenic regions. Thus, we calculated the binding prediction for spike glycoprotein, the nucleoprotein, and the membrane protein and detected good binding scores (score > 90), especially on the spike glycoprotein and the nucleoprotein (Table 3). Table 3 shows the best binding scores between

Table 4
Susceptible and Protective HLA allele average frequency in South America countries.

| | HLA allele frequency in South America (3,17) | | | | | | | | | | | |
| | Susceptible | | | | | | Protective | | | | | |
	HLA-B*46:01	HLA-B*46	HLA-A*25:01	HLA-A*25	HLA-C*01:02	HLA-C*01	HLA-B*15:03	HLA-B*15	HLA-A*02:02	HLA-A*02	HLA-C*12:03	HLA-C*12
ARGENTINA	0	0	0.04195	0.0175	0.03325	0.029	0.0048	0.0655	0.01519333	0.259	0.0654	0.095
BRAZIL	0.0002	0.0415712	0.01245	0.011311111	0.0071	0.0692	0.0195	0.0880125	0.00925	0.266055556	0.0342	0.04765
CHILE	0.0031	0	0.017525	0.0065	0.045966667	0	0.00565	0.0522	0.10365	0.2663	0.049	0
COLOMBIA	0	0	0.011266667	0.006	0.11605	0	0.0103	0.0503333	0.07806666	0.2565	0.0455	0
ECUADOR	0	0	0.0079	0.0066	0	0	0.011533333	0	0	0.372	0	0
PERU	0.0022	0.0032	0.0064	0	0	0.133	0.0086	0.0845333	0.2511	0.5745	0.031	0.02
URUGUAY	0	0.001	0	0.012	0	0.043	0	0.006	0	0.249	0	0
VENEZUELA	0	0	0.0058	0.0092	0.0238	0.04885	0	0.0764333	0	0.225533333	0	0.0545

the most common HLA alleles and the protective HLA alleles towards the dominant SARS-CoV2 B Cell epitopes, the dominant SARS-CoV2 T Cell epitopes and the most immunogenic SARS-CoV2 epitopes. We found four new SARS-CoV2 regions that represent a good binding with the most protective HLA alleles which could represent vaccine targets.

The prevalence of the HLA alleles related to SARS-CoV2 (Table 1) could be used to predict populations at risk; therefore, we searched for their frequency in South American populations (Table 4). The most frequent HLA alleles found were: HLA-C*01:02, HLA-B*15:03 and HLA-A*02 in Colombia; HLA-A*02 and HLA-C*12:03 in Argentina; and HLA-A*02 in Chile and in Peru (Table 4). In Ecuador, from a sample of 1010 Ecuadorians [16], the most frequent allele was HLA-A*02, followed by HLA-B*15:03 and HLA-A*25:01. None of the six HLA alleles that contribute to susceptibility or protection to SARS-CoV2 infection had a higher frequency than 0.58. By the time we performed our research, there was no available data from Paraguay, Bolivia, Guyana, Suriname, and French Guiana. Supplementary Table 2 demonstrates the complete data, including the number of populations for each country, showing no significant changes with the summary scores shown in Table 4.

Additionally, we found that the most common HLA alleles in South America are HLA class I: HLA-A*02, HLA-A*24, HLA-B*35, HLA-B*44, HLA-C*04; and HLA class II: HLA-DQB1*03, HLA-DQB1*02, HLA-DQB1*04, and HLA-DRB1*04. When looking for the most common HLA alleles in South America we only found alleles defined at the first-field. We performed a prediction binding between the mentioned alleles and the most immunogenic SARS-CoV2 regions. The software used to calculate the binding, IEDB, only allows to perform bindings with alleles defined at the second-field; therefore, we calculated the binding with the closest HLA allele to the first-field, which correspond to the first second-field HLA allele found in the software. The binding between the MHC I alleles and SARS-CoV2 regions is summarized in Table 5 -- only the highest scores > 0.90 are included. HLA-B*35 and HLA-B*44 have the top three best bindings (score > 0.98), directed towards the spike region. Moreover, we found new epitopes that have good binding, which could be considered as vaccine targets. We used an additional algorithm to confirm the results: ANN 4.0 [6] (Supplementary Table 3). HLA-DQB1*04 and HLA-DQB1*03 have the best bindings between the MHC II alleles and SARS-CoV2 regions (Supplementary Table 4).

4. Discussion

In December 2019, patients with respiratory symptoms of unknown origin were described in the city of Wuhan, China. These patients shared a common exposure to the same seafood market in the city [18]. The disease was recognized to be caused by a novel coronavirus, later named SARS-CoV2 [18]. The virus was rapidly transmitted and expanded worldwide. By the end of January 2020, the World Health Organization (WHO), declared it to be a "public health emergency of international concern" [19]. The infection briskly spread through all continents, reaching a total of 58,238,200 infected and 1,383,641 deaths by November 21, 2020 [1]. In South America, the total number of cases was 10,625,450 by the same date, and the countries with a higher number of affected individuals were: Brazil with 6,020,164 cases; Argentina with 1,359,042 cases; Colombia with 1,233,444 cases; Peru with 946,087 cases, and Chile with 539,143 cases [1].

The rapid spread of the virus may respond to either a characteristic of the virus or of the host. Thus, we investigated the association with one of the main immune system features: the HLA. The development of protection or susceptibility to a specific disease is influenced by the binding between parts of the pathogen and

Table 5
MHC I binding between the most common HLA alleles in South America and the most immunogenic SARS-CoV2 regions.

Allele	Start-End	Peptide	Score	Protein[a]	Description
HLA-B*35:01	84–92	LPFNDGVYF	0.98	S	Novel
HLA-B*35:01	896–904	IPFAMQMAY	0.98	S	
HLA-B*44:02	95–104	TEKSNIIRGW	0.98	S	Novel
HLA-B*35:01	325–333	TPSGTWLTY	0.98	N	
HLA-A*02:01	269–277	YLQPRTFLL	0.97	S	
HLA-B*44:02	322–330	MEVTPSGTW	0.97	N	
HLA-A*24:02	635–643	VYSTGSNVF	0.97	S	Novel
HLA-A*24:02	95–103	YFIASFRLF	0.97	M	
HLA-C*04:01	1137–1145	VYDPLQPEL	0.97	S	Novel
HLA-B*35:01	687–695	VASQSIIAY	0.96	S	
HLA-A*24:02	1208–1216	QYIKWPWYI	0.96	S	Novel
HLA-A*24:02	159–168	VYSSANNCTF	0.96	S	
HLA-B*35:01	79–87	SPDDQIGYY	0.95	N	Novel
HLA-A*02:01	976–984	VLNDILSRL	0.95	S	
HLA-A*24:02	1066–1075	TYVPAQEKNF	0.95	S	
HLA-B*35:01	321–329	QPTESIVRF	0.95	S	Novel
HLA-A*24:02	489–497	YFPLQSYGF	0.94	S	Novel
HLA-B*44:02	1201–1209	QELGKYEQY	0.94	S	Novel
HLA-A*02:01	222–230	LLLDRLNQL	0.94	N	
HLA-A*24:02	94–102	SYFIASFRL	0.94	M	Novel
HLA-A*24:02	448–456	NYNYLYRLF	0.93	S	
HLA-A*24:02	94–103	SYFIASFRLF	0.93	M	
HLA-A*24:02	788–797	IYKTPPIKDF	0.92	S	
HLA-B*44:02	11–20	EELKKLLEQW	0.92	M	Novel
HLA-A*24:02	368–377	LYNSASFSTF	0.91	S	
HLA-A*02:01	417–425	KIADYNYKL	0.91	S	Novel
HLA-B*44:02	989–997	AEVQIDRLI	0.90	S	Novel
HLA-A*02:01	15–23	KLLEQWNLV	0.90	M	Novel
HLA-A*24:02	265–275	YYVGYLQPRTF	0.90	S	

Prediction method: IEDB recommended 2020.04 (NetMHCpan EL 4.0) | High Score = good binder.

[a] M, membrane protein; N, nucleocapsid phosphoprotein; S, spike glycoprotein.

the HLA type [20]. For instance, in the case of the HIV infection, HLA*74:01 is associated with a lower viral load (protection) while HLA-A*36:01 with a higher viral load (susceptibility) [20]. Also, regarding the HIV-1 infection outcome, HLA-B*08, HLA-B*35, HLA-B*53, HLA-B*55, and HLA-B*56 have a lower binding affinity and a worse prognosis [20]. Likewise, for Hepatitis C virus infection, HLA-A*03 is associated with spontaneous clearance and HLA-A*25 with an unfavorable treatment outcome [20]. Another case is dengue fever, in which HLA-B*35, DRB1*04, *07, *11, and DQB1*03:02 [21] provide protection. In HPV-16 positive invasive cervical carcinoma, HLA DRB1*04:07-DQB1*03:02, and DRB1*15:01-DQB1*06:02 increase vulnerability [21].

Our literature review found that HLA-A*02:02, HLA-B*15:03, and HLA-C*12:03 have been reported in laboratory and silico as protective alleles and HLA-A*25:01, HLA-B*46:01 and HLA-C*01:02 as susceptible alleles for SARS-CoV2 (Table 1) [9]. Our data supports Barquera et al. findings that described HLA-A*02:02 and HLA-B*15:03 are the strongest binders of SARS-CoV2 peptides; whereas HLA-A*25:01, HLA*B46:01 and HLA*C:01:02 are the weakest binders [22]. HLA-B*46:01 provides susceptibility to SARS-CoV and SARS-CoV2.

SARS-CoV and SARS-CoV2 have over 90% genetic similarity in S, M and N proteins [23]. Ou et al. compared convalescent sera from SARS-CoV and SARS-CoV2 patients and concluded moderate cross-neutralization activity between viruses [2]. We compared the binding interaction between SARS-CoV2 T and B cell regions and SARS-CoV immunogenic regions described by Grifoni et al. [7], demonstrated limited cross-neutralization and a unique immunity for SARS-CoV2, and confirmed Ou et al. results [2].

We demonstrated that the protective HLA identified from our systemic review bound tighter to immunogenic epitopes described by Grifoni et al. [7]. Additionally, we identified four new epitopes with strong bindings to the most protective HLA (HLA-B*15:03, HLA-A*02:02 and HLA-C*12:03), which could be used as potential

vaccine targets and contribute to an enhanced immunity response [24] (Table 5). These epitopes are mainly found in the spike protein, the most virulent and immunogenic region [7,17].

We searched for regional HLA frequency databases in South America that could provide accurate data; however, the information available is limited. The published studies analyzed primarily organ donors such as renal and bone marrow donors in Colombia or solid organs donors in Brazil. Furthermore, the Allele Frequencies Net Database also did not provide information from Paraguay, Bolivia, Guyana, Suriname, and French Guiana. We found in our literature review that the six HLA allele frequencies resulting in protection or susceptibility to SARS-CoV2, are uncommon in the region. These data are comparable with Requena et al., results that demonstrates only HLA-C*01:02 (susceptible) and HLA-C*12:03 (protective) have a weighted allele frequency (WAF) $\geq$ 5% [17].

We suggest that different protective HLA alleles could interact with the virus in South American populations. Therefore, we searched for the most common HLA alleles in South America (HLA-A*02, HLA-A*24, HLA-B*25, HLA-B*44, HLA-C*04, HLA-DQB1*03, HLA-DQB1*02, HLA-DQB1*04, and HLA-DRB1*04), however, the information is limited, and we could only find HLA alleles defined at first-field. We mentioned HLA-A*02:02 is a protective allele in previous studies and we found HLA-A*02 is common in the region. HLA-A*02:01 and HLA-A*24:02 are also mentioned by Cuspoca et al., as being the most frequent alleles in Latin America [25].

For the most common HLA alleles in South America, we calculated the binding between SARS-CoV2 immunogenic regions. As mentioned, the IEDB software only allows to perform bindings with second-field HLA alleles, therefore, we calculated the binding with the closest HLA allele to the first-field, which correspond to the first second-field HLA allele found in the software. We are aware that the first-field HLA allele do not always correspond to the second-field HLA allele. However, our results gives a general

idea of the mentioned prediction. Our results demonstrated that HLA-B*35:01 and HLA-B*44:02 have the tighter bindings. Peru and Ecuador populations have the highest prevalence HLA-B*35 and may be protected. In contrast, Colombia has the lowest prevalence of HLA-B*35 and HLA-B*44. South America is composed of many ancient American ancestry populations, which share unique MHC class I loci within them. For instance, the Waorani, a native tribe from Ecuador and Brazil, have a high frequency of rare globally MHC class I alleles including HLA-A2 and HLA-A24 [26] which suggests that selection and genetic drift may explain the presence of new HLA alleles in South American tribes [26]. Our study reinforces the importance of researching HLA variants in different populations and encourages the creation of a South America HLA variant database.

Additionally, we detected the epitopes interacting with the most common HLA in South America. Requena et al., reported a list of the best HLA-I candidate epitopes and our data overlap with IPFAMQMAY, YLQPRTFLL, VASQSIIAY, and YYVGYLQPRTF (score > 90) [17]. Additionally, the epitopes IPFAMQMAY, YLQPRTFLL and SYFIASFRLF are also mentioned by Cuspoca et al., as potential vaccine peptides [25]. Moreover, we report new epitopes for HLA-I that could be useful in the development of a regional vaccine (Table 5).

In conclusion, we provide bioinformatic evidence that HLA-B*15:03, HLA-A*02:02, HLA-C*12:03 are protective, and HLA-A*25:01, HLA-B*46:01, HLA-C*01:02 increase susceptibility towards SARS-CoV2. We reported two lists of new epitopes (most of them located in the spike protein), one for the previously reported protective/susceptibility HLA alleles, and a second for the most common HLA alleles found in South America. The novel epitopes could aid in the current vaccine research and alert about possible adverse immunogenic outcomes. Finally, we showed that the protective/susceptible HLA alleles are uncommon in South America, which suggests either other HLA respond to the infection, or the influence of other environmental factors in the development of the infection.

CRediT authorship contribution statement

Samantha Sáenz Hinojosa: Investigation, Resources, Writing - original draft, Writing - review & editing, Visualization. **Vanessa Romero:** Conceptualization, Writing - review & editing, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.humimm.2021.04.005.

References

[1] Coronavirus Update (Live): 58,317,427 Cases and 1,384,174 Deaths from COVID-19 Virus Pandemic – Worldometer n.d. https://www.worldometers.info/coronavirus/?utm_campaign=homeAdvegas1?%20, (accessed November 21, 2020).

[2] X. Ou, Y. Liu, X. Lei, P. Li, D. Mi, L. Ren, et al., Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV, Nat. Commun. 11 (2020) 1620, https://doi.org/10.1038/s41467-020-15562-9.

[3] F.F. Gonzalez-Galarza, A. McCabe, E.J.M.D. Santos, J. Jones, L. Takeshita, N.D. Ortega-Rivera, et al., Allele frequency net database (AFND) 2020 update: gold-standard data classification, open access genotype data and new query tools, Nucl. Acids Res. 48 (2020) D783–D788, https://doi.org/10.1093/nar/gkz1029.

[4] NCBI Resource Coordinators, Database resources of the National Center for Biotechnology Information, Nucl. Acids Res. 46 (2018) D8–D13, https://doi.org/10.1093/nar/gkx1095.

[5] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, et al. The Protein Data Bank 2000. http://Nucleic Acids Research, 28: 235-242, (accessed May 14, 2020).

[6] R. Vita, S. Mahajan, J.A. Overton, S.K. Dhanda, S. Martini, J.R. Cantrell, et al., The immune epitope database (IEDB): 2018 update, Nucl. Acids Res. 47 (2019) D339–D343, https://doi.org/10.1093/nar/gky1006.

[7] A. Grifoni, J. Sidney, Y. Zhang, R.H. Scheuermann, B. Peters, A. Sette, A sequence homology and bioinformatic approach can predict candidate targets for immune responses to SARS-CoV-2, Cell Host Microb. 27 (671–680) (2020), https://doi.org/10.1016/j.chom.2020.03.002 e2.

[8] K. Kiyotani, Y. Toyoshima, K. Nemoto, Y. Nakamura, Bioinformatic prediction of potential T cell epitopes for SARS-Cov-2, J. Hum. Genet. 65 (2020) 569–575, https://doi.org/10.1038/s10038-020-0771-5.

[9] A. Nguyen, J.K. David, S.K. Maden, M.A. Wood, B.R. Weeder, A. Nellore, et al., Human leukocyte antigen susceptibility map for severe acute respiratory syndrome coronavirus 2, J. Virol. 94 (2020), https://doi.org/10.1128/JVI.00510-20.

[10] M.H.L. Ng, K.-M. Lau, L. Li, S.-H. Cheng, W.Y. Chan, P.K. Hui, et al., 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. 190 (2004) 515–518, https://doi.org/10.1086/421523.

[11] S.-F. Wang, K.-H. Chen, M. Chen, W.-Y. Li, Y.-J. Chen, C.-H. Tsao, et al., 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. 24 (2011) 421–426, https://doi.org/10.1089/vim.2011.0024.

[12] N. Keicho, S. Itoyama, K. Kashiwase, N.C. Phi, H.T. Long, L.D. Ha, et al., Association of human leukocyte antigen class II alleles with severe acute respiratory syndrome in the Vietnamese population, Hum. Immunol. 70 (2009) 527–531, https://doi.org/10.1016/j.humimm.2009.05.006.

[13] Y.-M.-A. Chen, S.-Y. Liang, Y.-P. Shih, C.-Y. Chen, Y.-M. Lee, L. Chang, et al., Epidemiological and genetic correlates of severe acute respiratory syndrome coronavirus infection in the hospital with the highest nosocomial infection rate in Taiwan in 2003, J. Clin. Microbiol. 44 (2006) 359–365, https://doi.org/10.1128/JCM.44.2.359-365.2006.

[14] M. Lin, H.-K. Tseng, J.A. Trejaut, H.-L. Lee, J.-H. Loo, C.-C. Chu, et al., Association of HLA class I with severe acute respiratory syndrome coronavirus infection, BMC Med. Genet. 4 (2003) 9, https://doi.org/10.1186/1471-2350-4-9.

[15] A.H. Hajeer, H. Balkhy, S. Johani, M.Z. Yousef, Y. Arabi, Association of human leukocyte antigen class II alleles with severe Middle East respiratory syndrome-coronavirus infection, Ann Thorac Med 11 (2016) 211–213, https://doi.org/10.4103/1817-1737.185756.

[16] J.M. Galarza, R. Barquera, A.M.T. Álvarez, D.I. Hernández Zaragoza, G.P. Sevilla, A. Tamayo, et al., Genetic diversity of the HLA system in human populations from the Sierra (Andean), Oriente (Amazonian) and Costa (Coastal) regions of Ecuador, Hum. Immunol. 79 (2018) 639–650, https://doi.org/10.1016/j.humimm.2018.06.004.

[17] D. Requena, A. Médico, R.D. Chacón, M. Ramírez, O. Marín-Sánchez, Identification of Novel Candidate Epitopes on SARS-CoV-2 Proteins for South America: A Review of HLA Frequencies by Country, Front. Immunol. 11 (2020) 2008, https://doi.org/10.3389/fimmu.2020.02008.

[18] D. Wu, T. Wu, Q. Liu, Z. Yang, The SARS-CoV-2 outbreak: What we know, Int. J. Infect. Dis. 94 (2020) 44–48, https://doi.org/10.1016/j.ijid.2020.03.004.

[19] X. Li, W. Wang, X. Zhao, J. Zai, Q. Zhao, Y. Li, et al., Transmission dynamics and evolutionary history of 2019-nCoV, J. Med. Virol. 92 (2020) 501–511, https://doi.org/10.1002/jmv.25701.

[20] N.B. Crux, S. Elahi, Human Leukocyte Antigen (HLA) and Immune Regulation: How Do Classical and Non-Classical HLA Alleles Modulate Immune Response to Human Immunodeficiency Virus and Hepatitis C Virus Infections?, Front. Immunol. 8 (2017) 832, https://doi.org/10.3389/fimmu.2017.00832.

[21] D. Maira, A. Vansan, A. Maria, J.E. Laguila Visentainer, C.A. de Souz, HLA and infectious diseases. HLA and associated important diseases, InTech (2014), https://doi.org/10.5772/57496.

[22] R. Barquera, E. Collen, D. Di, S. Buhler, J. Teixeira, B. Llamas, et al., 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 96 (2020) 277–298, https://doi.org/10.1111/tan.13956.

[23] S.F. Ahmed, A.A. Quadeer, M.R. McKay, Preliminary identification of potential vaccine targets for the COVID-19 coronavirus (SARS-CoV-2) based on SARS-CoV immunological studies, Viruses 12 (2020), https://doi.org/10.3390/v12030254.

[24] J. Trowsdale, J.C. Knight, Major histocompatibility complex genomics and human disease, Ann. Rev. Genomics Hum. Genet. 14 (2013) 301–323, https://doi.org/10.1146/annurev-genom-091212-153455.

[25] A.F. Cuspoca, L.L. Díaz, A.F. Acosta, M.K. Peñaloza, Y.R. Mendez, D.C. Clavijo, et al. A Multi-Epitope Vaccine Against Sars-Cov-2 Directed Towards the Latin American Population: An Immuno informatics Approach 2020. doi:10.21203/rs.3.rs-70414/v1.

[26] D.I. Watkins, S.N. McAdam, X. Liu, C.R. Strang, E.L. Milford, C.G. Levine, et al., New recombinant HLA-B alleles in a tribe of South American Amerindians indicate rapid evolution of MHC class I loci, Nature 357 (1992) 329–333, https://doi.org/10.1038/357329a0.