viruses

MDPI

Article

In Silico Prediction of Human Leukocytes Antigen (HLA) Class II Binding Hepatitis B Virus (HBV) Peptides in Botswana

Wonderful Tatenda Choga [1,2], Motswedi Anderson [1], Edward Zumbika [3], Bonolo B. Phinius [1], Tshepiso Mbangiwa [1,2], Lynnette N. Bhebhe [1], Kabo Baruti [1,4], Peter Opiyo Kimathi [5], Kaelo K. Seatla [1,6], Rosemary M. Musonda [1,7], Trevor Graham Bell [8], Sikhulile Moyo [1,7], Jason T. Blackard [9] and Simani Gaseitsiwe [1,7,*]

1 Research Laboratory, Botswana Harvard AIDS Institute Partnership, Gaborone 0000, Botswana; wchoga@bhp.org.bw (W.T.C.); manderson@bhp.org.bw (M.A.); bphinius@bhp.org.bw (B.B.P.); tmbangiwa@bhp.org.bw (T.M.); lbhebhe@bhp.org.bw (L.N.B.); kabobaruti@gmail.com (K.B.); kseatla@bhp.org.bw (K.K.S.); rmusonda@bhp.org.bw (R.M.M.); smoyo@bhp.org.bw (S.M.)
2 Division of Human Genetics, Department of Pathology, Faculty of Health Sciences, University of Cape Town, Cape Town 7925, South Africa
3 Department of Applied Biology and Biochemistry, Faculty of Applied Sciences, National University of Science and Technology, Bulawayo 0000, Zimbabwe; edward.zumbika@nust.ac.zw
4 Department of Biological Sciences, Faculty of Science, University of Botswana, Gaborone 0000, Botswana
5 Centre for Proteomic and Genomic Research, Cape Town 7925, South Africa; popiyo@aims.ac.tz
6 Department of Medical Laboratory Sciences, Faculty of Health Sciences, University of Botswana, Gaborone 0000, Botswana
7 Department of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA
8 Independent Researcher, P.O. Box 497, Wits, Johannesburg 2050, South Africa; trevorgrahambell@gmail.com
9 Division of Digestive Diseases, University of Cincinnati College of Medicine, Cincinnati, OH 45267, USA; jason.blackard@uc.edu
* Correspondence: sgaseitsiwe@bhp.org.bw; Tel.: +267-390-2671; Fax: +267-3901284

Received: 12 June 2020; Accepted: 3 July 2020; Published: 6 July 2020

Abstract: Hepatitis B virus (HBV) is the primary cause of liver-related malignancies worldwide, and there is no effective cure for chronic HBV infection (CHB) currently. Strong immunological responses induced by T cells are associated with HBV clearance during acute infection; however, the repertoire of epitopes (*epi*) presented by major histocompatibility complexes (MHCs) to elicit these responses in various African populations is not well understood. In silico approaches were used to map and investigate 15-mers HBV peptides restricted to 9 HLA class II alleles with high population coverage in Botswana. Sequences from 44 HBV genotype A and 48 genotype D surface genes (*PreS/S*) from Botswana were used. Of the 1819 *epi* bindings predicted, 20.2% were strong binders (SB), and none of the putative *epi* bind to all the 9 alleles suggesting that multi-epitope, genotype-based, population-based vaccines will be more effective against HBV infections as opposed to previously proposed broad potency epitope-vaccines which were assumed to work for all alleles. In total, there were 297 unique *epi* predicted from the 3 proteins and amongst, S regions had the highest number of *epi* (n = 186). Epitope-densities (D_{epi}) between genotypes A and D were similar. A number of mutations that hindered HLA-peptide binding were observed. We also identified antigenic and genotype-specific peptides with characteristics that are well suited for the development of sensitive diagnostic kits. This study identified candidate peptides that can be used for developing multi-epitope vaccines and highly sensitive diagnostic kits against HBV infection in an African population. Our results suggest that viral variability may hinder HBV peptide-MHC binding, required to initiate a cascade of immunological responses against infection.

Keywords: hepatitis B virus (HBV); HLA class II alleles; T-cell epitopes; in silico; immunoinformatics; candidate multi-epitope vaccines (MEV); escape mutation; Botswana; Africa

1. Background

Hepatitis B virus (HBV), a member of the *Hepadnaviradae* family, is the major etiology of end stage liver diseases (ESLD), liver cirrhosis (LC) and hepatocellular carcinoma (HCC), and causes up to 887,000 deaths per year [1]. Although more than 90% of healthy adults resolve acute HBV infection within 6 months, there remain over 287 million people who test seropositive for hepatitis B surface antigen (HBsAg) [2] and have chronic HBV infection (CHB). Viral clearance is mediated by cytokines, lymphocytes, and the ability to mount a multi-specific polyclonal and vigorous T cell-mediated response against HBV antigens for a protective immunity [3–5]. The quality of these responses is influenced by host genetics, as well as the ability of certain viral variants to escape immune recognition [6–8].

The major histocompatibility complexes (MHCs)—known as human leukocytes antigens (HLAs) in humans—are integral components of host genes located at chromosome *6p21.* These highly polymorphic proteins serve as mediators of adaptive immune responses by presenting processed antigenic peptides to T cells. The two compatible types of MHCs—class I and class II—present exogenous and endogenous epitopes to $CD8^+$ cytolytic T cells and $CD4^+$ T helper (T_h) cells, respectively [9]. The MHC class II alleles (HLA-DR, -DQ and -DP) present epitopes to $CD4^+$ T cells [*epi*-HLA class II → $CD4^+$ T-cells] that in turn elicit adaptive immune responses against viral infections by facilitating the induction of $CD8^+$ cytotoxic T-lymphocytes (CTLs), production of cytokines crucial for survival, and maturation of B cells [10–13].

The link between HBV pathogenesis and host immunological profiles is still poorly understood [2]. HBV exhibits a high mutation rate, although only a small number of amino acid substitutions have been characterized functionally due to the costly and time-consuming nature of in vitro assays. Recent approaches have utilized in silico approaches such as machine learning techniques to prioritize candidate peptides for in vitro assays [14,15]. Thus far, the HBV mutations characterized have been associated with sensitivity of immunologic and molecular-based assays and viral escape leading to poor prognosis [16–19]. However, amino acid variations that influence HBV-MHC binding are poorly understood. In silico mapping of HLA class II binding peptides can be used to identify candidate peptides for in vitro assays to confirm $CD4^+$ T cell epitopes which are crucial for the design of epitope-based vaccines and highly sensitive diagnostic tools that can detect low HBV DNA levels which are frequently missed by diagnostic kits [20–23].

Sub-Sahara Africa (SSA) and the Western Pacific regions are highly endemic for CHB [24], where the circulating genotypes include A, D, and E for SSA, and B and C for the Western Pacific. HBV genotypes A (subgenotype A1) and D (subgenotype D3) have been reported in Botswana, with genotype E occurring rarely [25–29]. Not only do genotypes show unique geographic distribution, they also differ in treatment response, pathogenesis potential, and prognosis [30,31]. Studies conducted in China have mapped different T cell epitopes that may be eligible for epitope-based vaccines and some were evaluated in vitro [20,32]. However, these findings may be less applicable to African populations whose host genetic pool, circulating genotypes, and immune profiles for HBV (e.g., hepatitis B e antigen [HBeAg] and HBsAg positivity) differ considerably with those of the Chinese population. A prerequisite to determine the epitope(s) for inclusion in epitope-based vaccines include (1) identification of conserved regions of the genome and (2) characterization of those regions that elicit protective immune responses [33]. Although there are hepatitis B vaccines in use currently, vaccine escape does occur; thus, more optimized vaccine candidates may be needed to avoid vaccine failure. In this study, we utilized HLA class II alleles that occur at the highest frequency in Botswana and locally derived HBV strains to identify HLA class II binding peptides which are good candidates for confirmatory in vitro tests of

immunogenicity. The present study had three major aims: (1) to determine the repertoires of HLA class II epitopes within HBV envelope sequences of genotypes A and D isolated from different risk groups in Botswana (described in our earlier papers); (2) to compare if the predicted epitopes in genotypes A and D may vary across other HBV genotypes, suggesting that genotype-based multi-epitope vaccines would be more successful than the broad potency vaccines currently in use; (3) to investigate amino acid variations within these epitopes to determine if they may lead to immune escape (i.e., candidate escape mutations).

2. Materials and Methods

2.1. Mapping Peptides from HBV Surface Gene Restricted to HLA-class II Alleles

Three sequence datasets were included in the current analysis. The first database (N_1) was used to map epitopes (*epi*) that bind predominantly to HLA class II alleles in Botswana and consisted of 92 non-recombinant full-length S gene (*PreS/S*) sequences [25,26,28] retrieved from GenBank (accession numbers MF979142—MF979176, KR139743—KR139748, and MH464807—MH464854). The aligned sequences were sorted by genotype and included A (n = 44) and D (n = 48). The three domains of HBV surface proteins—PreS1, PreS2, and S—were manually extracted from an overlapping *Pol/S* fragment and were divided into genotypes whose amino acid (aa) sequence alignments were sorted according to column similarities. Nucleotide alignments and sorting were performed using AliView 1.21 software [34]. Each region was then used to create a consensus sequence with the threshold set at 90% for all positions. Variants that did not meet this threshold were investigated independently in post-analyses. To assess if the aa composition of consensus sequences was representative of existing HBV strains, BLAST searches were conducted using the NCBI database, and strains exhibiting 100% similarity and coverage were evaluated further (Supplementary Table S1).

The 15-mer HBV peptides overlapping by 14 aa were tested for binding to 9 HLA class II alleles—HLA-DRB1*0101, DRB1*0301, DRB1*0401, DRB1*0701, DRB1*0802, DRB1*1101, DRB1*1302, DRB1*1501, and DRB5*0101—that have high population coverage in Botswana [35]. The NetMHCIIPan version 3.2 online server (http://www.cbs.dtu.dk/services/NetMHCIIpan/) [36] was used to predict binding peptides, and their binding-affinity scores were categorized based on the Log-transformed binding affinity [1-Log50k (aff)]. The settings were adjusted starting with default [1-Log50k (aff)] of 0.426; 500nM and logAff = 0.638; 50nM for weak and strong binding respectively. Ten previously characterized HBV envelope proteins (PreS1, PreS2, S) epitopes—S: 18–37 epitope ID 51310; S: 70–84 (3966); S: 83–98 (46959); S: 200–211 (66307); S: 201–215 (59353); S: 211–244 (6574); S: 230–247 (47877); S: 363–378 (76458); S: 376–389 (17331); S: 378–386 (37664)—available at (www.iedb.org) [37] were included to calibrate the tool's settings, and the thresholds showing highest specificity were utilized in the present study. Using percentile rank of eluted ligand prediction score (%Rank_EL), the strong binders (SB) were determined between 2–10% Rank_EL and 10–50% Rank_EL for weak binders (WB). Peptides with binding affinity less than that of WB were deemed pseudo binders (NB). Figure 1 outlines the various analytical steps included in this study, and the NetMHCIIPan results are provided in the Supplementary Table S2.

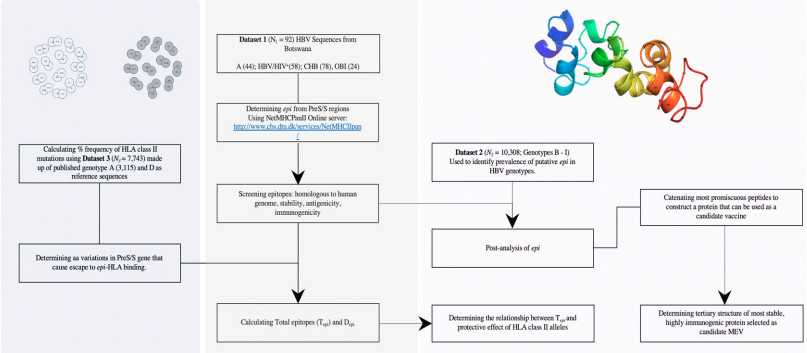


Figure 1. Schema illustrating the flow of data analysis used in this study. N = sample size; SB = strong binding peptides; WB = weak biding peptides; T_{epi} = total predicted epitopes; PreS/S = HBV surface gene; D_{epi} = epitope densities. Sequences were derived from patients with different clinical outcomes: −(HBV/HIV; CHB; OBI)—HIV = human immunodeficiency virus; OBI = occult hepatitis B infection, CHB = chronic hepatitis B infection, HBV/HIV = coinfection. The blue colored segment shows the pipeline used to evaluate the diversity of *epi*. The grey segment is the pipeline used to determine *epi* and measure of promiscuity and conservativeness. The pink segment is the pipeline used to determine the best candidate vaccine.

The relationship between the length of the protein and the frequency of binding-peptides were compared using epitope density score (D_{epi}) for all protein domains of S gene (PreS1, PreS2, S) and genotypes (A versus D). D_{epi} score was defined as the proportion of binding peptides $\sum_{i=1}^{n} I_{epi}$ (where; $\boldsymbol{I}_{epi}$ = WB + SB) to total predicted epitopes (T_{epi}) relative to protein size. T_{epi} represent the total count of all predicted proteins.

2.2. Determining Prevalence of Putative Epitopes in HBV Genotypes (A–I) Except A and D

Putative or promiscuous *epi* were defined as peptides that exhibit similar binding affinity to 2 or more alleles. The prevalence of predicted putative *epi* were determined in a second dataset (N_2) that included 10,308 PreS/S sequences (genotype B = 2905; C = 5575; E = 1118; F = 477; G = 86; H = 69; I = 78) retrieved from HBV database available at http://hvdr.bioinf.wits.ac.za/alignments/index.html [38]. N_2 was also used to determine the overall prevalence of the predicted escape mutations. The sequences used in this analysis are included in the supplementary file provided; Supplementary-Table S3.

2.3. Variations Causing Escape to HLA Class II Binding

Dataset N_3 consisted of 7743 HBV sequences (genotype A = 3115; genotype D = 4628) used to determine the frequency of aa variations which were termed *HLA escape mutations* in other HBV sequences. Sequences used were curated from http://hvdr.bioinf.wits.ac.za/alignments/index.html, partitioned by proteins (PreS1, PreS, S) [38]. Escape mutations were defined as those aa variations within the 15-mer core aa sequence that cause the binding affinity to change from either strong to pseudo binding (SB → NB) or from weak to pseudo binding (WB → NB). Several in-house customized Python pipelines were used to accurately investigate the frequency of escape mutations. The sequences used in this analysis are in the supplementary file provided; Supplementary Table S4.

2.4. Screening Putative Epi and Reconstruction of Tertiary Structure of the Modelled Vaccine

Since the predicted putative *epi* can be also homologous to human peptides that may (1) cause either autoimmune responses when used as a vaccine or (2) give false results when used as a diagnostic marker, a BLAST search was conducted with the NCBI protein database for all immunogenic *epi*. Afterwards, the predicted putative *epi* were catenated using a previously described method [39], and different combinations were used to construct candidate multi-epitope vaccines (MEVs). Physiological and

biochemistry proteins such as thermal stability, desirable shelf-life, and pH among other properties are prerequisites during development of an ideal vaccine. The biochemical properties of generated candidate proteins were evaluated using online ProtoParam tool [40]. The proteins exhibiting properties similarly to those of vaccines currently in use were deemed the best candidate. The properties of 3 current HepB vaccines—including VO_0011094, VO_0011095, and VO_0011093—are curated in the DNA vaccine database [41]. Properties predicted include; immunogenicity, antigenicity, instability index, estimated half-life in humans' molecular weight (mw), aliphatic index (AI), grand average of hydrophobicity (GRAVY), and theoretical (pH).

2.5. Determining the HLA-HBV Association Using T_{epi}

To test the hypothesis that T_{epi} could serve as a useful predictor of HLA-HBV associations, T_{epi} of S genes (A and D) were used to rank the 9 HLA class II alleles in the post *epi* prediction analyses. The available literature was used to corroborate the analyses.

3. Results

3.1. Predicting T-cell Epitopes Using Consensus from N_1 Dataset

We first generated 6 different consensus sequences from *PreS/S* sequences of genotypes A and D. Consensus sequences were validated by comparison to the NCBI database and identified 16 pre-existing sequences which exhibit 100% identity to: *PreS1* genotype A ($PreS1_A$) consensus sequence; 39 to $PreS1_D$, 6 to $PreS2_A$, 12 to $PreS2_D$, 19 to S_A, and 20 to S_D, respectively. The consensus sequences, identical sequences, and their country of origin are provided in supplementary file; Supplementary Table S1. *Epitope* is defined as a peptide that binds either weakly or strongly bind to HLA-DR alleles used, while *HLA class II escape mutations* were defined as aa variations within the 9-mer core aa sequence that changed the epitope-HLA-DR binding from SB/WB to pseudo binding. In total, there were 1819 total binding predicted (T_{epi}) from 297 unique epitopes restricted to 9 HLA class II including 20.2% SB. The number of signature aa differentiating HBV genotypes A and D were 31, 13, and 9 for the S, PreS1, and PreS1 regions, respectively (Table 1). S protein had the highest binding peptides constituting 79.9% of the sum of all T_{epi} ($\sum T_{epi}$), PreS1 constituted 11%, while PreS2 had the least (9.1%) $\sum T_{epi}$. The D_{epi} of the SB and WB among genotypes (A and D) and proteins (*PreS1*, *PreS2*, *S*) are summarized in Figure 2.

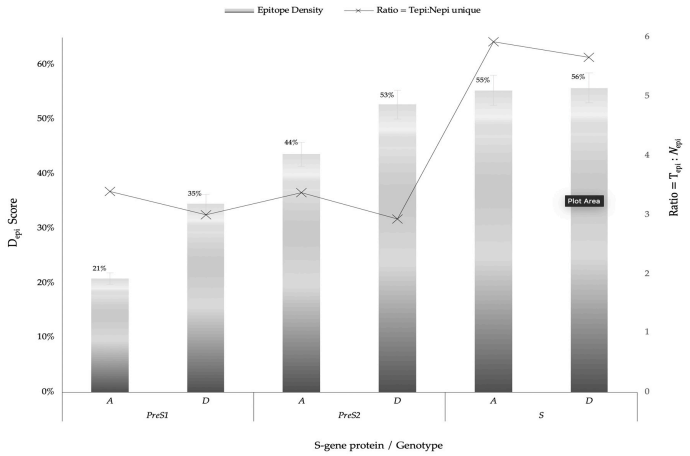


Figure 2. Epitope densities (D_{epi}) of different PreS/S proteins stratified by genotype (A or D) and protein (*PreS1*, *PreS2*, *S*). $D_{epi} = \frac{\sum_{i=1}^{n} Iepi}{T_{epi} \text{ X Protein length}}$ where *i* can be any protein ($PreS/S_A$ versus $PreS/S_D$). $PreS1_A$ represent genotype A large Hepatitis B surface antigen (HBsAg); $PreS1_D$ represent genotype D large HBsAg; $PreS2_A$ represent genotype A middle HBsAg; $PreS2_D$ represent genotype D middle HBsAg; S_A represent genotype A small HBsAg; S_D represent genotype D small HBsAg. T_{epi} = Total binding peptides (WB + SB). N_{epi} unique = count of unique binding peptides per each protein.

Table 1. Distribution of T cell epitopes restricted to 9 HLA class II alleles with high population coverage in Botswana.

	PreS1		**PreS2**		**S**	
	A (120 aa)	**D (108 aa)**	**A (55 aa)**	**D (55 aa)**	**A (226 aa)**	**D (226 aa)**
Total bindings (n = 1819) SB (367); WB (1452) (%)	6; 79 (4.7)	7; 107 (6.3)	4; 81 (4.7)	3; 78 (4.5)	122; 619 (40.7)	136; 577 (39.1%)
Unique *epi* (n = 297)	25 ¥	38 ¥	29 ¥	24 ¥	125 ¥	126 ¥
Ratio = T_{epi}: N_{epi}	3.4	3.0	3.4	2.9	5.9	5.7
Most active HLA: DRB * (SB; WB)	*0802; *0101	*0401; (*0401, *0101)	*0301; (*0401, *0101)	*0301; 5*0101	*0702; *0401	*0401; *1501
Genotype variation: $\frac{A}{D}$						
(A|D *epi*; p-value)	A|D > 0.05		A|D > 0.05		A|D > 0.05	

¥ indicates existence of *epi* that are common in both genotypes A and D. SB; strong binding peptides. WB; weak binding peptides. aa; amino acids. Web-logo diagrams represent signature aa between consensus sequences of genotype A and D set at a threshold of 100%. HBV; hepatitis B virus. * 0101 means HLA class II allele DRB1*0101 etc. 5*0101 means HLA class II allele DRB5*0101.

3.2. Prevalence of Putative Epitopes Across all HBV Genotypes

To assess if the predicted *epi* may be suitable for vaccine inclusion, the most common *epi* were compared to sequences other than genotypes A and D. Table 2 shows the number of aa sequences containing the indicted *epi*. Eight out of 53 *epi* were found in semi-conserved regions ranked as "++++" and had prevalence > 85% among sequences in N_2 dataset. These results suggest that multi-epitope genotype-based vaccines may be better to avoid vaccine escape.

3.3. Profiles of Strong Binding Epitopes

SB *epi* of the 3 proteins were sorted by the core aa sequence and analyzed based on genotype. Those found in both genotypes were considered putative when binding to alleles. There were no SB epitopes that were common between sequences of genotypes A and D for PreS1 and PreS2 regions. S protein had 89 out of 230 *epi* that satisfied the above criteria and were promiscuous for at least 5 alleles. There were 5 unique *epi* whose core aa were at S protein residues 41–49 (FLGGSPVCL), 14 *epi* with S protein residues 20–28 (FLLTRILTI), 10 *epi* with S residues 183–191 (FVGLSPTVW), 9 *epi* with S residues 22–30 (LTRILTIPQ), 6 *epi* with S residues 162–170 (LWEWASARF), 6 *epi* with S residues 184–192 (VGLSPTVWL), 12 *epi* with S residues 96–104 (VLLDYQGML), 2 *epi* with S residues 180–188 (VQWFVGLSP), 9 *epi* with S residues 163–171 (WEWASARFS), 4 *epi* with S residues 182–190 (WFVGLSPTV), and 12 *epi* with S residues 72–80 (YRWMCLRRF). Table 3 shows the full profiles of SB *epi* mapped for the S proteins for genotypes A and D.

3.4. Profiles of Most Promiscuous Epitopes

Since the majority of predicted *epi* were WB, a strict threshold was applied to select the most the promiscuous *epi*. Thus, *epi* were selected if they bind to least 6 alleles or more as shown in Table 4.

3.5. Designing and Predicting Structure of Candidate Multi-Epitope Vaccine

Table 4 shows 27 putative *epi* binding to at least 6 alleles and were selected to model the tertiary structure of candidate vaccines. The overlapping S region *epi* (S_A = 11, S_D = 10) were catenated to form proteins which were used to model a tertiary structure as shown in Figure 3. The overlapping S_A *epi* at S protein residues 6–20 with aa sequence SGFLGPLLVLQAGFF, aa sequence CIPIPSSWAFAKYLWEWASVRFSWLSLLVPFVQWF at S protein residues 155–183, and aa sequence WYWGPSLYNILSPFIPLLPIFFCLW at S protein residues 199–223 yields a 75 amino acid protein of mw = 8878.62. Of the 6 proteins predicted, the most stable S_A protein was vacci-S_A with amino acid sequence: 5′-SGFLGPLLVLQAGFFWYWGPSLYNILSPFIPLLPIFFCLWCIPIPSSWAFAKYLWEWASVRFSWLSLLV PFVQWF-3′. The overlapping *epi* in S_D occupied residues S: 6–20, S: 68–82, S: 197–223 and S: 155–183. The resulting protein was 85 aa long and had mw = 10,749.99. Of the 24-proteins predicted, the one selected for constructing tertiary structure was vacci-S_D with aa sequence: 5′-MMWYWGPSLYSILSPFLPLLPIFFCLWSGFLGPLLVLQAGFFSWAFGKFLWEWASARFSWLSLLVPF VQWFTCPGYRWMCLRRFIIFLF-3′. When a BLAST search was conducted with the NCBI protein database for the 2 protein sequences, results show that both sequences were similar to 2 domains of major surface antigen (vMSA) from hepadnavirus superfamily; accession number pfam00695. A similar approach was used to select proteins from the *epi* in PreS2 region to determine proteins that can be used to model the 3D tertiary structures of *epi* in the PreS2 region. In total, there were 6 *epi* selected 3 for each genotype. The overlapping *epi* were 20 aa long with mw of 2006.20 and occupying residues PreS2: 34–53 in genotype A sequences. vacci-$PreS_D$ was made from overlapping *epi* occupying residues PreS2: 37–53 and had a mw of 1719.91. All proteins generated using the different epi ordering (permutations and combinations) have been provided under the supplementary file Supplementary Table S5.

Table 2. Showing the prevalence of S protein *epi* in other HBV genotypes (B–I).

Epitope Sites in S Protein	AA Sequence	B (*n* = 2905)	C (*n* = 5575)	E (n = 1118)	F (*n* = 477)	G (*n* = 86)	H (*n* = 69)	I (*n* = 78)	Prevalence (%) = 1 − [Count of Seq / 10308] *%	Degree of Conservation ↓ (+: Variable)	*epi* Previously Discussed
17–31	AFGKFLWEWASARFS	E = 998	C = 32	F = 1					10,0	+	[42]
180–194	AGFFLLTRILTIPQS	B = 62	C = 4627	E = 5	F = 2	G = 75			46,3	++	[43]
90–104	CLIFLLVLLDYQGML	B = 2605	C = 4661	E = 1034	F = 442	G = 84	H = 65	I = 70	86,9	++++	[44]
69–83	CPGYRWMCLRRFIIF	B = 2384	C = 5064	E = 1027	F = 465	G = 84	H = 67	I = 65	88,8	++++	[42]
19–33	FFLLTRILTIPQSLD	B = 63	C = 4717	E = 5	F = 2	G = 76			47,2	++	[45–47]
158–172	FGKFLWEWASARFSW	E = 996	C = 31	F = 1					10,0	+	[48]
20–34	FLLTRILTIPQSLDS	B = 63	C = 4709	E = 5	F = 5	G = 78			47,1	++	[49]
93–107	FLLVLLDYQGMLPVC	B = 2616	C = 4720	E = 1031	F = 446	G = 84	H = 66	I = 72	87,7	++++	[50]
161–175	FLWEWASARFSWLSL	B = 1	E = 997	C = 81	F = 3	I = 16			10,7	+	[42]
179–193	FVQWFVGLSPTVWLS	B = 2630	C = 4698	E = 76	G = 85	I = 75			73,4	+++	[51]
18–32	GFFLLTRILTIPQSL	B = 63	C = 4618	E = 5	F = 2	G = 76			46,2	++	-
159–173	GKFLWEWASARFSWL	E = 1002	C = 31	F = 3					10,1	+	-
202–216	GPSLYSILSPFLPLL	C = 19	E = 2						0,2	+	[48]
71–85	GYRWMCLRRFIIFLF	B = 82	C = 5065	E = 1032	F = 464	G = 84	H = 67	I = 65	66,5	+++	-
92–106	IFLLVLLDYQGMLPV	B = 2606	C = 4659	E = 1030	F = 443	G = 84	H = 66	I = 70	86,9	++++	-
195–209	IWMMWYWGPSLYSIL	B = 3	C = 24	E = 2					0,3	+	-
160–174	KFLWEWASARFSWLS	B = 1	E = 1012	C = 34	F = 3	I = 16			10,3	+	-
91–105	LIFLLVLLDYQGMLP	B = 2611	C = 4656	E = 1031	F = 443	G = 84	H = 66	I = 70	86,9	++++	[52]
21–35	LLTRILTIPQSLDSW	B = 63	C = 4754	E = 5	F = 5	G = 77			47,6	++	[53]
94–108	LLVLLDYQGMLPVCP	B = 2653	C = 4729	E = 1032	F = 445	G = 84	H = 66	I = 72	88,1	++++	[54]
15–29	LQAGFFLLTRILTIP	B = 62	C = 4735	E = 5	F = 2	G = 76			47,3	++	[54]
192–206	LSVIWMMWYWGPSLY	B = 772	C = 4249	E = 889	G = 1	I = 50			57,8	++	[44]
95–109	LVLLDYQGMLPVCPL	B = 2619	C = 4708	E = 1022	F = 444	G = 84	H = 66	I = 73	87,5	++++	[55,56]
162–176	LWEWASARFSWLSLL	B = 7	E = 1001	C = 88	F = 445	G = 2	H = 65	I = 75	16,3	+	[54]
205–219	LYSILSPFLPLLPIF	C = 21	E = 2						0,2	+	-
178–192	PFVQWFVGLSPTVWL	B = 2649	C = 4723	E = 76	G = 85	I = 74			73,8	+++	-
70–84	PGYRWMCLRRFIIFL	B = 2375	C = 5117	E = 1033	F = 466	G = 85	H = 67	I = 65	89,3	++++	[49]
66–80	PPTCPGYRWMCLRRF	B = 60	C = 751	E = 14	F = 464	G = 77	H = 68		13,9	+	[44]
203–217	PSLYSILSPFLPLLP	C = 19	E = 2						0,2	+	-
67–81	PTCPGYRWMCLRRFI	B = 60	C = 750	E = 14	F = 464	G = 77	H = 68		13,9	+	[57]
16–30	QAGFFLLTRILTIPQ	B = 62	C = 4645	E = 5	F = 2	G = 76			46,5	++	[58]
181–195	QWFVGLSPTVWLSVI	B = 2588	C = 4310	E = 73	G = 6	I = 71			68,4	+++	[58]
38–52	SLNFLGGTTVCLGQN	B = 5	C = 2	E = 2					0,1	+	[55–57]
204–218	SLYSILSPFLPLLPI	C = 19	E = 2						0,2	+	[44]
193–207	SVIWMMWYWGPSLYS	B = 3	C = 24	E = 2					0,3	+	-

Table 2. *Cont.*

Epitope Sites in S Protein	AA Sequence	B (*n* = 2905)	C (*n* = 5575)	E (n = 1118)	F (*n* = 477)	G (*n* = 86)	H (*n* = 69)	I (*n* = 78)	Prevalence (%) = 1 − [Count of Seq / 10308] *%	Degree of Conservation ↓ (+: Variable)	*epi* Previously Discussed
155–169	SWAFGKFLWEWASAR	E = 998	C = 31	F = 1					10,0	+	-
68–82	TCPGYRWMCLRRFII	B = 63	C = 757	E = 17	F = 466	G = 77	H = 68		14,0	+	-
37–51	TSLNFLGGTTVCLGQ	B = 5	C = 2	E = 2					0,1	+	[59,60]
194–208	VIWMMWYWGPSLYSI	B = 3	C = 24	E = 2					0,3	+	-
14–28	VLQAGFFLLTRILTI	B = 61	C = 4676	E = 5	F = 2	G = 76			46,8	++	[61]*
180–194	VQWFVGLSPTVWLSV	B = 2636	C = 4584	E = 77	G = 6	I = 75			71,6	+++	[62]
156–170	WAFGKFLWEWASARF	E = 1000	C = 32	F = 1					10,0	+	[44]
163–177	WEWASARFSWLSLLV	B = 7	E = 1002	C = 87	F = 445	G = 2	H = 64	I = 75	16,3	+	-
182–196	WFVGLSPTVWLSVIW	B = 2506	C = 4125	E = 73	G = 6	I = 71			65,8	+++	[63]
201–215	WGPSLYSILSPFLPL	C = 19	E = 2						0,2	+	[52]
196–210	WMMWYWGPSLYSILS	B = 3	C = 25	E = 2					0,3	+	-
36–50	WTSLNFLGGTTVCLG	B = 5	C = 2	E = 2					0,1	+	-
35–49	WWTSLNFLGGTTVCL	B = 5	C = 2	E = 2					0,1	+	-
199–213	WYWGPSLYSILSPFL	C = 18	E = 2						0,2	+	-
206–220	YSILSPFLPLLPIFF	C = 19	E = 2						0,2	+	-
200–214	YWGPSLYSILSPFLP	C = 18	E = 2						0,2	+	-

+ The degree of conservation. The scale used: if score > = 100, then highly conserved and will be denoted by '+++++'. elif score > = 85: then semi conserved = '++++'. elif score > = 60: region of mutation and is denoted by '+++'. elif score > = 20, then highly variable region = '++'. else: high escape mutation = '+'. n represents the number of sequences used in the analysis. B represents full-length genotype B sequences, C represents full-length genotype C sequences, etc. 17–31 representing the beginning the position occupied by the 14-mer epitope predicts (e.g., 17 is the first amino acid of the epitope, while 31 is the last amino acid of the epitope).

Table 3. List of most promiscuous SB *epi* mapped from S protein.

S: Residues	AA Sequence	Geno	Core aa	HLA Class II Alleles																							
69–83	CPGYRWMCLRRFIIF	A	YRWMCLRRF																								
	CPGYRWMCLRRFIIF	D		*0101	*0301																						
68–82	ICPGYRWMCLRRFII	A																									
70–84	PGYRWMCLRRFIIFL	D				*0401	*0701	*0802	*1302																		
	PGYRWMCLRRFIIFL	A			*0301																						
71–85	GYRWMCLRRFIIFLF	A		*0101						*1501																	
	GYRWMCLRRFIIFLF	D																									
67–85	PICPGYRWMCLRRFI	A																									
66–84	PPICPGYRWMCLRRF	A		*0101						*1501	5*0101																
67–81	PTCPGYRWMCLRRFI	D								*1501																	
147–161	CTCIPIPSSWAFAKY	A	CIPIPSSWA																								
	CTCIPIPSSWAFGKF	D		*0101		*0401	*0701	*0802	*1501																		
145–159	GNCTCIPIPSSWAFA	A																									
146–160	NCTCIPIPSSWAFAK	A																									
	NCTCIPIPSSWAFGK	D									5*0101																
177–191	VPFVQWFVGLSPTVW	D	WFVGLSPTV	*0101		*0401	*0701	*0802	*1101		5*0101																
	VPFVQWFVGLSPTVW	A																									
168–182	ARFSWLSLLVPFVQW	D	WLSLLVPFV																								
169–183	RFSWLSLLVPFVQWF	D		*0101		*0401	*0701	*0802	*1101	*1501	5*0101																
	RFSWLSLLVPFVQWF	A																									
167–181	SARFSWLSLLVPFVQ	D																									
176–189	LVPFVQWFVGLSPTV	A	WFVGLSPTV	*0101		*0401	*0701	*0802			5*0101																
	LVPFVQWFVGLSPTV	D																									
166–180	ASARFSWLSLLVPFV	D									5*0101																
170–184	FSWLSLLVPFVQWFV	A		*0101		*0401	*0701	*0802	*1101																		
	FSWLSLLVPFVQWFV	D																									
167–181	SVRFSWLSLLVPFVQ	A									5*0101																

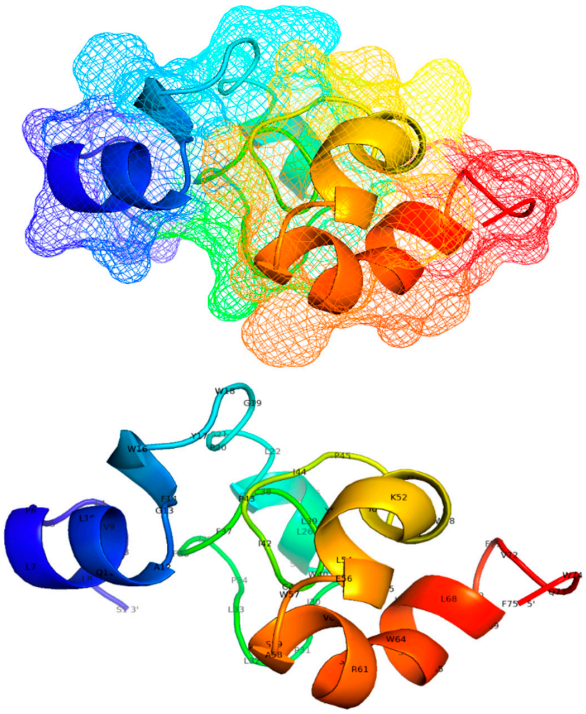
S: Residues	AA Sequence	Geno	Core aa	HLA Class II Alleles																							
68–82	TCPGYRWMCLRRFII	D	YRWMCLRRF	*0101	*0301	*0401	*0701	*0802		*1302	*1501																
168–182	VRFSWLSLLVPFVQW	A	WLSLLVPFV	*0101		*0401	*0701	*0802	*1101	*1302	*1501	5*0101															
14–28	VLQAGFFLLTRILTI	A	FLLTRILTI		*0301	*0401		*0802		*1302	*1501	5*0101															
	VLQAGFFLLTRILTI	D																									
165–179	WASARFSWLSLLVPF	D	FSWLSLLVP	*0101				*0802	*1101			5*0101															
174–188	SLLVPFVQWFVGLSP	A	FVQWFVGLS	*0101				*0802	*1101		*1501																
	SLLVPFVQWFVGLSP	D																									
6–20	SGFLGPLLVLQAGFF	A	LGPLLVLQA				*0401	*0701																			
	SGFLGPLLVLQAGFF	D		*0101				*0701				5*0101															
7–21	GFLGPLLVLQAGFFL	D							*0802	*1101																	
	GFLGPLLVLQAGFFL	A									*1501																
5–19	TSGFLGPLLVLQAGF	D																									
	TSGFLGPLLVLQAGF	A																									
93–107	FLLVLLDYQGMLPVC	D	LVLLDYQGM						*0802	*1101																	
	FLLVLLDYQGMLPVC	A		*0101		*0401			*0802	*1101	*1501																
92–106	IFLLVLLDYQGMLPV	A							*0802	*1101																	
	IFLLVLLDYQGMLPV	D							*0802	*1101																	
182–196	WFVGLSPTVWLSVIW	D	VGLSPTVWL		*0301	*0401			*0802	*1302	*1501																
	WFVGLSPTVWLSVIW	A			*0301	*0401			*0802	*1302																	
97–111	LLDYQGMLPVCPLIP	A	YQGMLPVCP			*0401	*0701		*1101		*1501	5*0101															
	LLDYQGMLPVCPLIP	D		*0101		*0401	*0701		*1101																		
96–110	VLLDYQGMLPVCPLI	A				*0401	*0701		*1101																		
	VLLDYQGMLPVCPLI	D				*0401	*0701		*1101																		
72–86	YRWMCLRRFIIFLFI	A	WMCLRRFII				*0701	*0802	*1101		*1501	5*0101															
	YRWMCLRRFIIFLFI	D	WMCLRRFII																								
194–208	VIWMMWYWGPSLYNI	A	WYWGPSLYN	*0101		*0401	*0701	*0802	*1101			5*0101															

$PreS1_A$ represent genotype A *epi* derived from sequences of large Hepatitis B surface antigen (HBsAg); $PreS1_D$ represent genotype D *epi* derived from sequences of large HBsAg; $PreS2_A$ represent genotype A *epi* derived from sequences of middle HBsAg; $PreS2_D$ represent genotype D *epi* derived from sequences of middle HBsAg; S_A represent genotype A *epi* derived from sequences of small HBsAg; S_D represent genotype D *epi* derived from sequences of small HBsAg. *0101 means HLA class II allele DRB1*0101 e.tc. 5*0101 means HLA class II allele DRB5*0101.

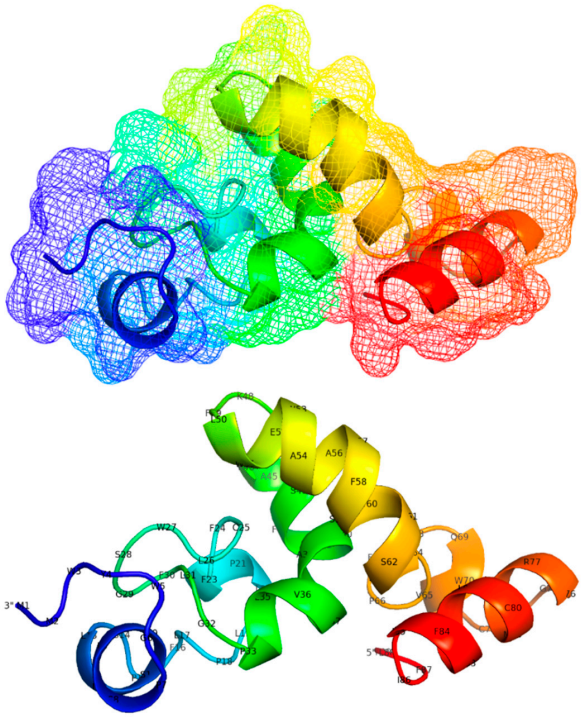
Table 4. Highlighting most promiscuous T cell epitopes restricted to 9 HLA class II alleles.

Epitope Site	AA Sequence	HBV Protein	Core AA	HLA Class II Alleles	Previously Discussed *epi*
34–48	PVPNIASHISSISSR	$PreS2_A$	IASHISSIS	1*0101, 1*0401, 1*0701, 1*0802, 1*1101, 1*1302, 1*1501, 5*0101	-
39–53	ASHISSISSRTGDPA	$PreS2_A$	ISSISSRTG	1*0101, 1*0401, 1*0701, 1*0802, 1*1101, 1*1501, 5*0101	-
38–52	IASHISSISSRTGDP	$PreS2_A$	ISSISSRTG	1*0101, 1*0401, 1*0701, 1*0802, 1*1101, 1*1501, 5*0101	-
37–51	TTASPLSSIFSRIGD	$PreS2_D$	LSSIFSRIG	1*0101, 1*0401, 1*0701, 1*0802, 1*1101, 1*1501, 5*0101	-
38–52	TASPLSSIFSRIGDP	$PreS2_D$	LSSIFSRIG	1*0101, 1*0401, 1*0701, 1*0802, 1*1101, 1*1501, 5*0101	-
39–53	ASPLSSIFSRIGDPA	$PreS2_D$	LSSIFSRIG	1*0101, 1*0401, 1*0701, 1*0802, 1*1101, 1*1501, 5*0101	-
208–222	ILSPFIPLLPIFFCL	S_A	FIPLLPIFF	1*0101, 1*0401, 1*0701, 1*0802, 1*1101, 1*1501, 5*0101	-
209–223	LSPFIPLLPIFFCLW	S_A	FIPLLPIFF	1*0101, 1*0401, 1*0701, 1*0802, 1*1101, 1*1501, 5*0101	-
207–221	NILSPFIPLLPIFFC	S_A	FIPLLPIFF	1*0101, 1*0401, 1*0701, 1*0802, 1*1101, 1*1302, 5*0101	-
149–163	CIPIPSSWAFAKYLW	S_A	IPSSWAFAK	1*0101, 1*0401, 1*0701, 1*0802, 1*1101, 1*1501, 5*0101	-
6–20	SGFLGPLLVLQAGFF	S_A	LGPLLVLQA	1*0101, 1*0401, 1*0701, 1*0802, 1*1101, 1*1501, 5*0101	-
200–214	YWGPSLYNILSPFIP	S_A	LYNILSPFI	1*0101, 1*0401, 1*0701, 1*0802, 1*1101, 1*1302, 1*1501	-
199–213	WYWGPSLYNILSPFI	S_A	LYNILSPFI	1*0101, 1*0401, 1*0701, 1*0802, 1*1101, 1*1501, 5*0101	-
164–178	EWASVRFSWLSLLVP	S_A	VRFSWLSLL	1*0101, 1*0401, 1*0802, 1*1101, 1*1302, 1*1501, 5*0101	[64]
168–182	VRFSWLSLLVPFVQW	S_A	WLSLLVPFV	1*0101, 1*0401, 1*0701, 1*0802, 1*1101, 1*1302, 1*1501, 5*0101	[44]
169–183	RFSWLSLLVPFVQWF	S_A	WLSLLVPFV	1*0101, 1*0401, 1*0701, 1*0802, 1*1101, 1*1501, 5*0101	[65]
156–170	WAFAKYLWEWASVRF	S_A	YLWEWASVR	1*0101, 1*0301, 1*0401, 1*0701, 1*0802, 1*1101, 1*1501, 5*0101	[49]
208–222	ILSPFLPLLPIFFCL	S_D	FLPLLPIFF	1*0101, 1*0401, 1*0701, 1*0802, 1*1101, 1*1501, 5*0101	-
209–223	LSPFLPLLPIFFCLW	S_D	FLPLLPIFF	1*0101, 1*0401, 1*0701, 1*0802, 1*1101, 1*1501, 5*0101	-
155–169	SWAFGKFLWEWASAR	S_D	FLWEWASAR	1*0101, 1*0301, 1*0401, 1*0701, 1*0802, 1*1101, 1*1501	-
6–20	SGFLGPLLVLQAGFF	S_D	LGPLLVLQA	1*0101, 1*0401, 1*0701, 1*0802, 1*1101, 1*1501, 5*0101	[44]
199–212	WYWGPSLYSILSPFL	S_D	LYSILSPFL	1*0101, 1*0401, 1*0802, 1*1101, 1*1302, 1*1501, 5*0101	-
167–181	SARFSWLSLLVPFVQ	S_D	WLSLLVPFV	1*0101, 1*0401, 1*0701, 1*0802, 1*1101, 1*1501, 5*0101	-
168–182	ARFSWLSLLVPFVQW	S_D	WLSLLVPFV	1*0101, 1*0401, 1*0701, 1*0802, 1*1101, 1*1501, 5*0101	[44]
169–183	RFSWLSLLVPFVQWF	S_D	WLSLLVPFV	1*0101, 1*0401, 1*0701, 1*0802, 1*1101, 1*1501, 5*0101	[42,49]
197–211	MMWYWGPSLYSILSP	S_D	WYWGPSLYS	1*0101, 1*0401, 1*0701, 1*0802, 1*1101, 1*1501, 5*0101	-
68–82	TCPGYRWMCLRRFII	S_D	YRWMCLRRF	1*0101, 1*0301, 1*0401, 1*0701, 1*0802, 1*1101, 1*1501, 5*0101	-

*0101 means HLA class II allele DRB1*0101 e.tc. 5*0101 means HLA class II allele DRB5*0101.



(**a**)



(**b**)

Figure 3. Showing a tertiary structure of candidate vaccines: (**a**) Tertiary structures of candidate epi modelled using 3Dpro webtool. The S_A protein in (a) has the aa composition: 5′-SGFLGPLLVLQAGFFWYWGPSLYNILSPFIPLLPIFFCLWCIPIPSSWAFAKYLWEWASVRFSWLSLLV PFVQWF-3′, and had following theoretical properties: antigenicity (0.04), instability index (II = 47.07), estimated half-life in vitro = 1.9 h, molecular weight (mw) = 8878.62, aliphatic index (AI = 114.40) and grand average of hydrophobicity (GRAVY = 0.995), and theoretical alkalinity (pI = 7,76). Using VaxiJen ver2.0 server set at threshold of 0.4, the overall prediction for the protective antigen was 0.53, displaying

it as a plausible antigen [66]. (**b**) (5′-MMWYWGPSLYSILSPFLPLLPIFFCLWSGFLGPLLVLQA GFFSWAFGKFLWEWASARFSWLSLL VPFVQWFTCPGYRWMCLRRFIIFLF-3′) protein modelled using from S_D epi. The protein had following theoretical properties: antigenicity (0.11), instability index (II = 53), estimated half-life in vitro 30 h, molecular weight (mw) = 10749.99, aliphatic index (AI = 101.91) and grand average of hydrophobicity (GRAVY = 0.965), and theoretical alkalinity (pI = 9.42),). The antigenicity score predicted in both candidate vaccine suggests that they are plausible antigens [66].

3.6. Mutations Associated with Escape from Class II HLA Binding

For amino acid variants present in more than 5% of sequences of N_1 database, we evaluated the in silico impact on immune recognition. Mutations were labelled relative to the HBV surface gene proteins—$PreS1_{1–115}$, $PreS2_{1–42}$, and $S_{1–226}$. Mutations that hindered HBV peptide-HLA binding were 86T, 90T, and 94P in $PreS1_A$; 54P, 79E, 84S, and 85Q in $PreS1_D$; 12I, 31I, and 54P in $PreS2_A$; 5F, 22H, 22L, 22P, 32H, 36L, and 42S in $PreS2_D$. The list of escape mutations and their prevalence in 7743 HBV genotype A and D sequences are shown in Table 5 and supplementary file Supplementary Table S5. There were coordinated variations among positions in aa sequence alignments that had an impact on HBV-HLA binding but showed no impact when analyzed individually. These were termed covariance mutations. The combinations of the covariance mutations and their impact on binding potential are shown in Tables 5–7 below.

3.7. Assessing the Distribution of HBV-HLA (epi) as a Predictor of HLA Protective Effective

We used existing information on immunological studies for meta-analysis to estimate T_{epi} as a predictor of the protective effect of MHC class II alleles. Figure 4 shows a Pareto analysis and the 20% threshold indicates the 3 HLA class II alleles—DRB1*0301, 1302, *1101—that are highly likely to be less protective against HBV. We therefore speculate, with caution, that there is a relationship between T_{epi} and protectiveness and this should be further investigated to establish correlations (p-value < 0.05) with high statistical confidence.

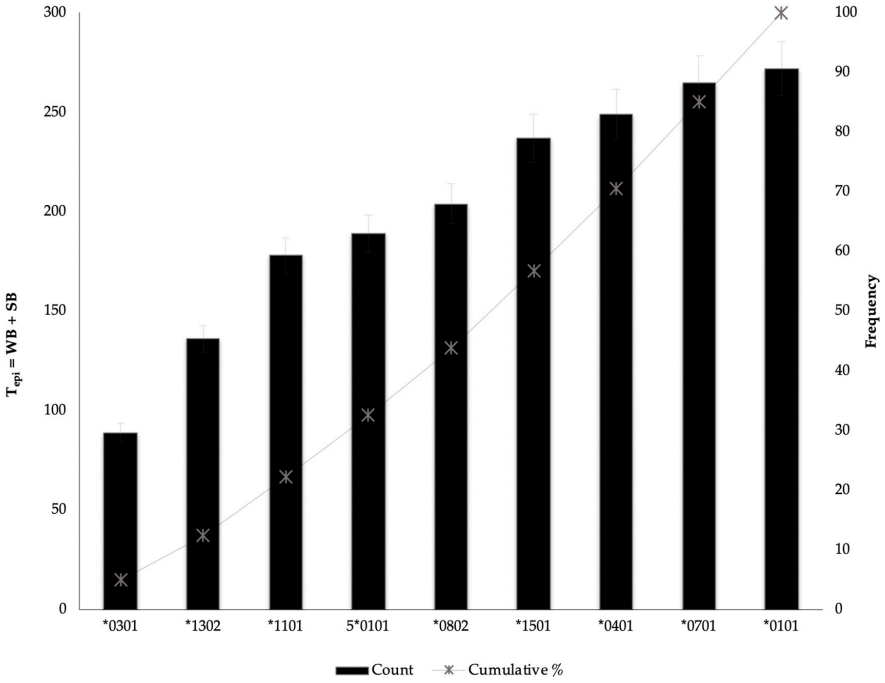


Figure 4. Pareto Analysis applied to rank the T_{epi} of alleles against their percentage frequency.

Table 5. The prevalence of S gene escape mutations.

Protein	AA Sequence	S Protein AA Residues (*epi*)	Escape Mutations	Count in A (*n* = 3115)	Count in D (*n* = 4628)	Count in Other Genotypes
S_A	MENITSGFLGPQLV	1–14	L12Q	13	1	3
		1–14	I4T	13	25	91^{Δ}
		1–14	I4Stop	10	-	-
S_A	GPLLVLQAGFFLLTR	10–24	L15Stop	2	-	17
S_A	LNFPGGSPVCLGQNS	39–53	L42P	13	6	53
S_A	TRILTIPQ *LDSWWT	23 -37	S31Stop	6	12	-
S_A	DLWWTSLNFLGDPPV	33–47	G44D	4	-	16
	DSWWTSLNFLGESPV		G44E	105	87	826
	IPIPSSWGFAKYLWE	150–164	A157G	10	5	3
S_A	IPIPSSWAFVKYLWE		A159V	20	3	234
S_A	PPICPGYRWMCQRRF	66–80	L77R	7	2	43
			L77Q	6	-	18
S_A	PPICPGYRWMCLR *F		R79Stop	9	10	20
			R79H	13	58	49
S_A	LIFLLVLLDYQDMLP	91–105	G102D	1	2	3
S_A	CLIFLLVLLDYQGML	90–104	D99Stop	15	5	8
			Y100C	196	12	17
			M103I	31	31	29
S_A	SLLVPFVQWFVGLTP LLVPFVQWFEGLSPT	174–188	S187T	1	-	-
			G185E	6	15	19
			V184E	1	2	2
			L186P	10	-	5
S_A	SPTVWLLAIWMMWYW	187–193	S193L	104	257	612
			A194V	494	Δ *	
S_A	GPSLYNISSPFIPLL	202–216	L209S	4	9	12
S_D	ENITSGCLGPLLVLQAGF	2–16	F8C	-	1	-
S_A	YLWEWASVRFSWPSL	161–175	L173P	1	3	11

Table 5. *Cont.*

Protein	AA Sequence	S Protein AA Residues (*epi*)	Escape Mutations	Count in A (*n* = 3115)	Count in D (*n* = 4628)	Count in Other Genotypes
S_A	SPFIPLLLIFFCLWV	210–224	P214L	11	41	23
S_D	SSWAFGKFLWEWASA	154–168 208–222	K160N	1	7	9
S_D	ILSPYILLLPIFFCI		F212Y + L213I + P214L	-; -; 11	40; 202; 41	3; $^{\Delta *}$;23
	NITSGFLGLLLVLQA	4 -18	P11L	2	4	3
S_D	MENITSGFLGPLLVL	1–15	T5P; N3S + I4T + T5A	2; -	4; 2	4; -29
S_D	SWWTSLNFLGETTVC	34–48	G44E	105	87	826
S_D	SWWTSLNFRGGTTVC		L42R	14	8	14
S_D	LSVIWMMWYWGPNLY	192–206	S204N	136	194	848
List of aa Variations within Genotype A and D *epi*	M1E, M1L, M1K, M1I, M1V, M1T, M1R, E2H, E2Q, E2*, E2V, E2K, E2A, E2G, E2D, N3H, N3E, N3P, N3A, N3C, N3D, N3I, N3R, N3Y, N3K, N3T, N3G, N3S, I4L, I4R, I4Y, I4Q, I4F, I4M, I4H, I4K, I4A, I4P, I4S, I4N, I4V, I4T, T5Y, T5Q, T5L, T5R, T5E, T5K, T5V, T5P, T5I, T5S, T5A, S6G, S6F, S6A, S6P, S6T, S6L, G7C, G7D, G7*, G7A, G7V, G7K, G7E, G7R, F8C, F8A, F8G, F8V, F8Y, F8I, F8H, F8P, F8S, F8L, L9R, L9K, L9V, L9I, L9H, L9Q, L9P, G10D, G10*, G10V, G10Q, G10T, G10A, G10K, G10E, G10R, P11A, P11T, P11H, P11L, L12V, L12E, L12M, L12R, L12P, L12Q, L13Q, L13V, L13I, L13F, L13R, L13H, L13P, V14R, V14L, V14E, V14I, V14M, V14G, V14A, L15E, L15K, L15T, L15*, L15I, L15F, L15V, L15S, Q16K, Q16E, Q16H, Q16L, Q16R, Q16P, A17R, A17T, A17S, A17P, A17V, A17G, A17E, G18W, G18K, G18E, G18A, G18R, G18V, F19I, F19V, F19L, F19S, F19Y, F19C, F20I, F20Y, F20L, F20S, L21V, L21M, L21F, L21*, L21W, L21S, L22Q, L22M, L22V, L22F, L22S, L22*, L22W, T23F, T23R, T23Q, T23S, T23P, T23A, T23I, R24N, R24T, R24Q, R24E, R24I, R24G, R24S, R24K, I25R, I25S, I25F, I25N, I25A, I25T, I25V, L26Y, L26I, L26F, L26Q, L26P, L26H, L26R, T27R, T27P, T27S, T27A, T27K, T27I, I28K, I28L, I28V, I28T, I28M, P29E, P29A, P29F, P29S, P29Q, P29T, P29L, Q30M, Q30S, Q30P, Q30A, Q30L, Q30H, Q30R, Q30K, S31K, S31I, S31D, S31T, S31C, S31G, S31R, S31N, L32V, L32G, L32Q, L32R, L32I, L32P, D33V, D33H, D33Y, D33E, D33N, D33G, S34*, S34A, S34T, S34W, S34P, S34L, W35P, W35C, W35*, W35G, W35R, W35L, W36E, W36G, W36R, W36*, W36L, T37D, T37P, T37L, T37I, T37S, T37N, T37A, S38E, S38A, S38Y, S38F, S38P, L39I, L39V, L39R, L39H, L39F, L39P, N40Q, N40H, N40K, N40I, N40D, N40S, F41Q, F41I, F41Y, F41C, F41L, F41S, L42*, L42S, L42I, L42V, L42Q, L42R, L42P, G43V, G43W, G43K, G43R, G43E, G44Q, G44K, G44R, G44A, G44D, G44V, G44E, S45M, S45D, S45G, S45E, S45H, S45Q, S45R, S45I, S45K, S45N, S45V, S45L, S45P, S45A, S45T, P46N, P46R, P46S, P46I, P46A, P46H, P46L, P46T, V47Q, V47N, V47L, V47P, V47M, V47E, V47R, V47K, V47A, V47G, V47T, C48W, C48L, C48F, C48R, C48G, C48S, C48Y, L49A, L49T, L49C, L49S, L49I, L49V, L49F, L49H, L49R, L49P, G50C, G50W, G50P, G50R, G50D, G50V, G50S, G50A, Q51E, Q51K, Q51H, Q51P, Q51R, Q51L, N52T, N52G, N52H, N52K, N52I, N52D, N52S, S53G, S53M, S53*, S53T, S53P, S53W, S53L, Q54K, Q54*, Q54H, Q54L, Q54P, Q54R, S55T, S55Y, S55A, S55P, S55F, S55C, P56A, P56*, P56S, P56R, P56H, P56L, P56Q, T57S, T57A, T57I, S58Y, S58A, S58L, S58T, S58P, S58F, S58C, N59L, N59I, N59R, N59H, N59D, N59K, N59S, H60D, H60S, H60N, H60Y, H60Q, H60R, H60P, S61T, S61P, S61*, S61L, P62S, P62Q, P62L, T63V, T63F, T63N, T63S, T63A, T63I, S64W, S64P, S64F, S64Y, S64C, C65G, C65F, C65S, C65Y, C65R, P66S, P66L, P66T, P66A, P66Q, P66H, P67T, P67A, P67R, P67S, P67L, P67Q, I68D, I68P, I68V, I68F, I68S, I68N, I68A, I68T, C69F, C69G, C69S, C69R, C69Y, C69W, C69*, P70S, P70H, P70R, P70A, P70T, P70L, G71S, G71D, G71W, G71V, Y72S, Y72D, Y72N, Y72H, Y72F, Y72C, R73L, R73C, R73P, R73H, W74C, W74G, W74R, W74*, W74S, W74L, M75K, M75L, M75R, M75S, M75V, M75T, M75I, C76G, C76*, C76R, C76W, C76S, C76F, C76Y, L77P, L77G, L77M, L77V, L77Q, L77R, R78I, R78G, R78W, R78P, R78L, R78Q, R79P, R79G, R79L, R79C, R79S, R79H, F80G, F80N, F80Y, F80I, F80L, F80S, I81S, I81Y, I81N, I81F, I81M, I81V, I81T, I82F, I82N, I82V, I82M, I82T, I82L, F83Y, F83I, F83L, F83C, F83S, L84F, L84H, L84P, F85Q, F85R, F85P, F85L, F85Y, F85S, F85C, I86M, I86K, I86L, I86F, I86V, I86T, L87M, L87V, L87R, L87Q, L87P, L88R, L88V, L88M, L88Q, L88P, L89V, L89Y, L89T, L89R, L89I, L89Q, L89P, C90G, C90W, C90N, C90I, C90*, C90Y, C90F, C90S, L91G, L91F, L91P, L91R, L91H, I92H, I92N, I92V, I92M, I92L, I92S, I92T, F93R, F93W, F93A, F93Y, F93I, F93L, F93C, F93S, L94G, L94*, L94V, L94F, L94M, L94W, L94S, L95P, L95F, L95M, L95V, L95S, L95*, L95W, V96I, V96E, V96N, V96F, V96L, V96C, V96D, V96A, V96G, L97C, L97R, L97V, L97H, L97F, L97P, V98M, V98W, V98A, V98Q, V98P, V98R, V98L, D99H, D99K, D99V, D99A, D99E, D99N, D99G, Y100P,					

Table 5. *Cont.*

Protein	AA Sequence	S Protein AA Residues (*epi*)	Escape Mutations	Count in A (*n* = 3115)	Count in D (*n* = 4628)	Count in Other Genotypes
	Y100*, Y100K, Y100D, Y100H, Y100N, Y100L, Y100W, Y100S, Y100F, Y100C, Q101P, Q101L, Q101*, Q101K, Q101H, Q101R, G102N, G102R, G102A, G102D, G102C, G102S, G102V, M103K, M103R, M103L, M103T, M103V, M103I, L104Q, L104M, L104G, L104V, L104W, L104S, L104F, P105T, P105S, P105H, P105A, P105L, P105R, V106L, V106R, V106Q, V106D, V106E, V106G, V106I, V106A, C107A, C107W, C107*, C107G, C107S, C107Y, C107R, P108A, P108T, P108S, P108L, P108H, L109R, L109V, L109I, L109Q, L109M, L109P, I110R, I110H, I110V, I110N, I110F, I110S, I110P, I110T, I110M, I110L, A111N, A111R, A111T, A111Q, A111S, A111L, A111P, G112V, G112K, G112N, G112E, G112A, G112R, S113C, S113R, S113F, S113L, S113Y, S113K, S113P, S113N, S113A, S113T, T114V, T114*, T114L, T114R, T114N, T114M, T114K, T114A, T114P, T114S, T115P, T115S, T115A, T115I, T115N, T116S, T116P, T116A, T116I, T116N, S117Y, S117A, S117V, S117R, S117I, S117G, S117N, S117T, T118W, T118N, T118G, T118I, T118E, T118P, T118S, T118K, T118R, T118M, T118A, T118V, G119S, G119W, G119T, G119*, G119C, G119V, G119E, G119R, P120H, P120N, P120R, P120L, P120A, P120Q, P120S, P120T, C121T, C121G, C121F, C121L, C121Y, C121S, C121W, C121R, K122G, K122V, K122E, K122N, K122T, K122I, K122S, K122Q, K122R, T123G, T123V, T123P, T123S, T123N, T123I, T123A, C124G, C124R, C124W, C124Y, C124S, C124F, T125P, T125I, T125A, T125M, T126P, T126Q, T126K, T126G, T126M, T126L, T126V, T126N, T126A, T126S, T126I, P127F, P127R, P127V, P127H, P127S, P127A, P127I, P127T, P127L, A128T, A128R, A128G, A128V, Q129E, Q129K, Q129L, Q129P, Q129N, Q129H, Q129R, G130V, G130*, G130D, G130K, G130A, G130E, G130S, G130R, G130N, N131Q, N131G, N131D, N131H, N131K, N131S, N131I, N131A, N131P, N131T, S132H, S132Y, S132P, S132C, S132F, M133R, M133G, M133K, M133V, M133Q, M133S, M133I, M133L, M133T, F134K, F134D, F134W, F134T, F134Q, F134C, F134R, F134V, F134H, F134S, F134L, F134N, F134I, F134Y, P135S, P135R, P135T, P135H, P135L, S136T, S136C, S136A, S136L, S136P, S136F, S136Y, C137S, C137R, C137*, C137Y, C137W, C138S, C138*, C138W, C138G, C138R, C138Y, C139S, C139G, C139*, C139Y, C139W, C139R, T140P, T140K, T140M, T140A, T140L, T140I, T140S, K141N, K141Q, K141T, K141I, K141R, K141E, P142T, P142A, P142H, P142R, P142I, P142S, P142L, T143W, T143F, T143A, T143P, T143M, T143L, T143S, D144Y, D144N, D144V, D144A, D144G, D144E, G145D, G145V, G145I, G145S, G145K, G145E, G145A, G145R, N146I, N146T, N146K, N146S, N146D, C147W, C147S, C147R, C147G, C147Y, T148S, T148I, T148A, C149S, C149G, C149W, C149Y, C149R, I150S, I150N, I150M, I150V, I150T, P151S, P151R, P151T, P151L, P151H, I152H, I152F, I152L, I152T, I152V, P153A, P153T, P153L, P153Q, P153S, S154Q, S154A, S154H, S154V, S154T, S154P, S154L, S155T, S155A, S155Y, S155F, S155P, W156S, W156C, W156G, W156R, W156*, W156L, A157S, A157N, A157D, A157P, A157V, A157T, A157G, F158I, F158Y, F158S, F158L, A159P, A159Q, A159L, A159S, A159T, A159R, A159E, A159V, A159G, K160Q, K160T, K160E, K160G, K160S, K160N, K160R, Y161K, Y161C, Y161V, Y161H, Y161I, Y161L, Y161S, Y161F, L162S, L162*, L162R, L162V, L162I, L162P, L162Q, W163G, W163C, W163R, E164*, E164K, E164A, E164V, E164D, E164G, W165Q, W165*, W165G, W165C, W165S, W165R, W165L, A166S, A166D, A166P, A166T, A166V, A166G, S167Y, S167E, S167T, S167G, S167P, S167L, V168L, V168I, V168F, V168G, V168D, V168S, V16887T, V168P, V168A, R169G, R169L, R169S, R169C, R169P, R169H, F170V, F170L, F170S, S171N, S171Y, S171C, S171L, S171P, S171F, W172M, W172F, W172G, W172R, W172S, W172L, W172C, W172*, P173I, P173Y, P173S, P173V, P173F, P173L, S174G, S174K, S174C, S174R, S174T, S174I, S174N, L175I, L175*, L175F, L175S, L176I, L176V, L176Q, L176P, V177P, V177I, V177S, V177G, V177E, V177M, V177L, V177A, P178G, P178R, P178H, P178A, P178L, P178S, P178Q, F179I, F179C, F179L, F179Y, F179S, V180P, V180L, V180E, V180F, V180I, V180A, Q181K, Q181E, Q181L, Q181H, Q181R, W182F, W182G, W182S, W182R, W182C, W182L, W182*, F183V, F183Y, F183I, F183L, F183S, F183C, V184L, V184P, V184D, V184F, V184I, V184E, V184G, V184A, G185K, G185L, G185W, G185R, G185A, G185E, L186M, L186A, L186R, L186C, L186S, L186I, L186V, L186F, L186P, L186H, S187A, S187T, S187C, S187P, S187F, S187L, P188A, P188S, P188R, P188T, P188H, P188L, T189L, T189N, T189A, T189S, T189P, T189I, V190S, V190E, V190P, V190D, V190G, V190I, V190F, V190A, W191K, W191S, W191*, W191G, W191L, W191R, W191C, L192R, L192H, L192V, L192F, L192P, S193*, S193T, S193P, S193F, S193L, A194G, A194D, A194S, A194T, A194F, A194L, A194I, A194V, I195K, I195V, I195L, I195T, I195M, W196G, W196V, W196K, W196R, W196*, W196S, W196L, M197L, M197K, M197R, M197V, M197I, M197T, M198L, M198V, M198R, M198K, M198T, M198I, W199P, W199C, W199R, W199G, W199S, W199*, W199L, Y200H, Y200*, Y200L, Y200N, Y200S, Y200C, Y200F, W201C, W201L, W201G, W201S, W201R, W201*, G202V, G202E, G202R, G202A, P203H, P203T, P203G, P203S, P203A, P203L, P203Q, P203R, S204Q, S204H, S204C, S204D, S204I, S204T, S204G, S204K, S204R, S204N, L205Q, L205R, L205V, L205P, L205M, Y206G, Y206P, Y206Q, Y206*, Y206I, Y206W, Y206V, Y206D, Y206R, Y206L, Y206N, Y206S, Y206F, Y206H, Y206C, N207G, N207K, N207C, N207D, N207P, N207H, N207I, N207T, N207R, N207S, I208L, I208C, I208F, I208S, I208V, I208N, I208T, L209K, L209A, L209F, L209*, L209G, L209M, L209S, L209W, L209V, S210M, S210G, S210I, S210T, S210K, S210R, S210N, P211A, P211T, P211S, P211R, P211L, P211H, F212C, F212S, F212L, F212Y, I213A, I213V, I213S, I213F, I213T, I213M, I213L, P214G, P214R, P214T, P214H, P214S, P214Q, P214L, L215D, L215I, L215V, L215T, L215M, L215Q, L215R, L215P, L216T, L216G, L216C, L216R, L216Y, L216V, L216I, L216S, L216F, L216*, P217A, P217T, P217Q, P217S, P217L, I218P, I218K, I218S, I218F, I218N, I218T, I218L, F219C, F219I, F219L, F219S, F220V, F220I, F220S, F220W, F220Y, F220L, F220C, C221V, C221R, C221W, C221L, C221S, C221G, C221F, C221Y, H222Q, H222V, H222S, H222R, H222F, H222I, H222P, H222L, W223F, W223G, W223C, W223R, W223*, W223L, V224L, V224I, V224E, V224G, V224A, Y225A, Y225*, Y225N, Y225D, Y225L, Y225I, Y225H, Y225C, Y225F, Y225S, I226R, I226H, I226L, I226F, I226N, I226V, I226T, I226M, I226S,					

1R represents the position of the mutation '1' and R is the change in amino acid "candidate escape mutations". $^{\Delta}$ Genotype C sequences were excluded in the analysis. $^{\Delta*}$ genotype B, C, D, E, F, G, H and I sequences were excluded in this analysis. * = Stop codon.

Table 6. Summary of *PreS1* epitopes binding to respective alleles, and mutations that exists in the Botswana HBV sequences.

Epitope Position	Core AA Sequence	HLA_ DRB* 1/5_ Genotype A	Epitopes Position	Core AA Sequence	HLA_ DRB *1/5_ Genotype D	Mutations Relative to Protein: (WT, aa#, Mut)	
						A	D
84–92	ILATVPAVP ¥	*0802	73–81	ILQTLPANP		I84V, I84T, A86V, A86T, V88M, V88L, A90P, A90V, P92L	I73M, L75H, L77M, L77V, A79E
85–93	LATVPAVPP ¥	*0802	74–82	LQTLPANPP		I84V, I84T, A86V, A86T, V88M, V88L, A90P, A90V, P92L	I73M, L75H, L77M, L77V, A79E
34–42	FGANSNNPD		23–30	FRANTANP ¥	*0401	-	R24K
16–24	LSVPNPLGF		5–13	LSTSNPLGF ¥	*0401, *1302	-	-
24–32	FFPDHQLDP		13–21	FFPDHQLDP	*0401,	F25L	-
63–71	FGPGFTPPH		52–60	FGLGFTPPH	*0101, *0401, *0701, 5*0101	F67L	L54M, L54P, F56L
34–42	FGANSNNPD		23–30	FRANTANPD	*0101, *0401, *0802, *1302	-	R24K
67–75	FTPPHGGVL		56–64	FTPPHGGLL	*0101, *0701,	F67L, G73R	-
28–36	HQLDPAFGA		17–25	HQLDPAFRA	*0301, *0401,	-	R24K
84–92	ILATVPAVP	*0101, *0301, *0401, *0701, *0802, *1101, *1302, *1501, 5*0101	73–81	ILQTLPANP	*0101, *0301, *0401, *0701, *0802,*1302,	I84V, I84T, A86V, A86T, V88M, V88L, A90P, A90V, P92L	I73M, L75H, L77M, L77V, A79E
85–93	LATVPAVPP	*0401, *0802, 5*0101	74–82	LQTLPANPP	*0802, 5*0101	I84V, I84T, A86V, A86T, V88M, V88L, A90P, A90V, P92L	I73M, L75H, L77M, L77V, A79E
74–82	LLGWSPQAQ	*1501	63–71		*0101, *0401, *0802, *1501	W77stop, S78R, P79S, A81S, A81P	S67N, P68S
16–24	LSTSNPLGF		5–13		*0301, *0701, *1302	-	-
12–21	MGTNLSVPN	*0401, *0802, *1302	1–10	MGQNLSTSNP		-	G2E
15–22	NLSVPNPLG	*0401,	4–12	NLSTSNPLGF		-	-
83–90	QGILATVPA	*0101, *0802	71–79	QGILQTLPA	*0101,	G83D, I84V, I84K, A86T, A86V, V88L, V88M, A90T, A90V	I73M, L75H, L77M, L77V, A79E
14–22	TNLSVPNPL	*0101, *0701, *1302	3–11	QNLSTSNPL	*0101, *0401,	-	-
4–12	WSAKPRKGM	*1101, 5*0101	-			S5P, A6S, A6T, K10N, K10I	-
77–85	WSPQAQGIL	*0101, *0701,	66–75	WSPQAQGIL	*0101, *0701,	W77stop, S78R, P79S, A81P, A81S, G83D, I84V, I84K	P68S,

*0101 means HLA class II allele DRB1*0101 e.tc. 5*0101 means HLA class II allele DRB5*0101. ¥ indicates SB.

Table 7. Summary of *PreS2* epitopes binding to respective alleles, and mutations that exist within the Botswana HBV sequences.

Genotype	AA Sequence Core	AA Sequence of *epi*	Pres2 *epi* Residues	Count of *epi* with Core AA Sequence	HLA Class II Alleles	Mutations in the Core Sequence
A	ALQDPRVRG	AFHQALQDPRVRGLYFPA	7–24	7	*0301$^{\Delta}$	R16K, V17I
A	ASHISSISS	VPNTASHISSISSRT	35–49		*0401	A39T, I45S
D	ASPLSSIFS	VPTTASPLSSIFSRIG	35–50	2	*0101, *0401	L42I, L42S
A	FHQALQDPR	NSTAFHQALQDPRVRG	4–19	2	*0301, 5*0101	A11T, R16K
D	FHQTLQDPR	NSTTFHQTLQDPRVR	4–18	1	*0301	-
D	FSRIGDPAL	SPLSSIFSRIGDPALN	40–55	1	*0101, *0401, *0701, 5*0101	R48T, P52H, A53V, L54P
A	HISSISSRT	VPNTASHISSISSRTG	35–50	4	*0101, *0401, *0701	I45S, R48T
D	IFSRIGDPA	SPLSSIFSRIGDPALN	40–55	2	*0802, *1302, *1501	L42I, L42S
A	ISSISSRTG	PNTASHISSISSRTGDPALN	36–55	6	*0101, *0401, *0701, *0802, *1101, *1501, 5*0101	I45S, R48T
A	LNPVPNTAS	SSSGTLNPVPNTASHISSI	27–45	5	*0401, *0802	L32H, L32R, P34L, N37H, N37T, I38T, A38T
D	LSSIFSRIG	PTTASPLSSIFSRIGDPALN	36–55	6	*0101, *0401, *0701, *0802, *1101, *1501, 5*0101	P52L, A53V
A	MQWNSTAFH	MQWNSTAFHQALQDP	1–15	1	*1302	Q2I, A7T
		MQWNSTTFHQTLQDP		-	-	S5F, S5Y, T6A, T7I, T7N
D	PLSSIFSRI	VPTTASPLSSIFSRI	35–49	1	*0701	L42I, L42S
A	PVPNTASHI	SSGTLNPVPNTASHISSI	27–45	6	*1302	P34L, N37H, N37T
D	PVPTTASPL	VNPVPTTASPLSSIF	32–46	2	*0701	P34H, P36L, L42I, L42S
A	QALQDPRVR	TAFHQALQDPRVRGLYF	6–22	3	5*0101	-
D	QTLQDPRVR	STTFHQTLQDPRVRGLYF	5–22	4	5*0101	R22K
D	TLQDPRVRG	STTFHQTLQDPRVRGLYFPA		6	*0301	R18K, G19D, G19A
A	VRGLYFPAG	DPRVRGLYFPAGGSSSG	14–30	3	*0802, *1101, *1501	V17I, Y21N, Y21S, F22L, F22P, F22T, P23A
D				3	*0802, *1101, *1501	R18K, G19D, G19A, F2222Q, F22H, F22P, F22L
A	WNSTAFHQA	MQWNSTAFHQALQDP	1–15	1	*0401, *0701	A7T, A11T
D	WNSTTFHQT	MQWNSTTFHQTLQDP		1	*0701	S5F, S5Y, T6A, T7I, T7N
A	YFPAGGSSS	PRVRGLYFPAGGSSSGTLNP	15–34	6	*0101, *0401, *1101, 5*0101	Y21N, Y21S, F22L, F22P, F22T, P23A, S29L
D	YFPAGGSSS	PRVRGLYFPAGGSSSGTVNP		6	*0101, *0401, 5*0101, *1101	F22Q, F22H, F22P, F22L

*0101 means HLA class II allele DRB1*0101 e.tc. 5*0101 means HLA class II allele DRB5*0101.

4. Discussion

This detailed HBV immunoinformatics approach outlines the candidate peptides that can be used to develop biologicals against CHB. However, mutations within the T cell epitopes may impair HBV-HLA complex formation, a crucial component responsible for initiating cascade of responses for viral clearance [67]. In this study, we observed that there were no *epi* that bind to all alleles, and there was a large difference between the *epi* profiles of genotype A compared to genotype D. This phenomenon may explain previous failures during preclinical trials of candidate vaccines against CHB generated thus far, which were designed for broad potency. [68–75]. This may suggest that a genotype- and population-based, multi-epitope vaccine would be the best candidate to combat HBV. Using the cut-off set in this analysis, 21 *epi* from S regions and 6 from PreS2 regions were identified and used to construct tertiary structure of candidate vaccine. The candidate vaccines showed high binding in all alleles except for DRB1*0301. Populations with high coverage of DRB1*0301 have been closely associated with high susceptibility to CHB infection and nonresponse vaccination with envelope proteins [53,76,77].

Among the 3 types of proteins (PreS/S) analyzed in this study, S had the highest binding peptides constituting 79.9% of the sum of all T_{epi} ($\sum T_{epi}$), and genotype D had the most epitopes. However, when D_{epi} were compared between genotypes (A and D) and proteins (PreS/S), it was clear that the length of proteins where independent of the *epi* but dependent on the aa compositions and the position of the *epi* within the proteins. While S region was the longest protein (213 aa), it had similar D_{epi} to the smallest protein $PreS2_D$ (PreS2 -54 aa). Previous studies that investigated all 7 proteins of HBV and showed that S protein had the most T cell antigenic epitopes [78,79]. However, further investigation should be conducted to determine any relationships between antigenicity and PreS1 protein with least *epi*.

We also compared the frequency of *epi* between 2 genotypes, A and D, and observed that genotype D had generally more immunogenic *epi* than A with exception for $PreS2_A$ whose *epi* were 12% more than those of $PreS2_D$, (Figure 1). Clinically, these trends correspond to data reported from studies that investigated the prognosis of patients infected with HBV genotype A1 strains compared to genotype D3. Others have reported a 10-fold increased progression to HCC among HBV patients infected with genotype A compared to genotype D and that patients infected with genotype A strains were likely to progress to CHB compared to genotype D [8,80]. Most countries in SSA, including Botswana, have low prevalence of HBV genotype D3 among different risk groups than genotype A1 [25–27,81].

Using existing information from studies that investigated the impact of alleles on different HBV outcomes to validate our statistical associations [53,82–85], we observed that out of 9 alleles, HLA-DRB*1301/2 and *0401 alleles—which had most *epi*—have been associated with spontaneous clearance of HBV infection [18,24,33,68,78,82,85–87], and HLA DRB1*0301 that had the least *epi* in our study has been previously associated with susceptibility to HBV infection, autoimmune hepatitis, chronicity, and non-responsiveness to HBV vaccination across different ethnic groups [9,88,89]. This strongly suggests that T_{epi} of *PreS/S* should be explored further as a predictor of the protective effect of HLA class II alleles.

A host immune system can recognize foreign antigens (*epi*) and clear the infection in some cases [90–92]; however, most pathogens including HBV can mutate within epitopes, and this may result in an escape from host immune surveillance leading to persistence of infection [93]. This characteristic is regarded as one of the major hindrances in developing high potency therapeutic drugs. This mechanism of aa changes within epitopes (escape mutations) interferes with both peptide processing reducing the intracellular antigen load and downregulation of MHC expression hence increased risk of developing liver malignancies (HCC, LC) among CHB patients [4,5,94–98]. Studies investigating the role of escape mutations within the T cell epitopes are relatively rare. In this study, we observed coordinated aa variations, which reveal genetic dependencies (i.e., *epi* that escaped HLA binding when there were two or more mutations); however, some single aa mutations altered the binding potential. These mutations were termed covariance mutations. For instance, in the proportion of *PreS1* binding peptides from both

genotype A and D against alleles shown in Figure 3, 15-mers with core aa sequence ILATVPAVP$_{84–92}$ in PreS1$_A$—corresponding to ILQTLPANP$_{73–81}$ in PreS1$_D$—weakly binds to alleles: *0101; 0401; 0701; 0802; *1101; 1302; *1501; 5*0101 → PreS1$_A$, and *0101; 0401; 0701; 0802; 1302→ PreS1$_D$ respectively. The aa *Ala* in genotype A is replaced by aa *Gln* for genotype D [A$_{(A)}$ → Q $_{(D)}$] and [Val$_{(A)}$ → Lue$_{(D)}$][91] causing *PreS1*$_D$ to pseudo bind to 3 alleles (*1101, 1302, and 5*0101). Additionally, the epitope strongly binds to allele *0802, but the changes in genotype D result in pseudo binding. Tables 3–7 summarize all the core aa of *PreS/S* epitopes that are restricted to 9 HLAs. We observed that the HBV epitope-HLA is greatly influenced by the position of core aa. For instance PreS1$_D$ *epi* AFGLGFTPPHGGLLG$_{51–65}$ is a WB to alleles: *0101, *0401 and 5*0101 when using core-aa: FGLGFTPPH$_{52–60}$ but it can only bind to *0701 when the core aa is FTPPHGGLL$_{57–64}$. Furthermore, post *epi* analyses show that there were mutations outside the core aa sequences but had impact on the HBV-HLA blinding. For instance, 2 *epi* with aa residues S: 41–55—FLGGPPVCLGQNSQS and FLGGSPVCLGQNSQS—and all with core aa sequences FLGGPPVCL show different binding affinities thus NB and WB respectively. The escape mutations defined in the present study were those found within the core aa sequence.

Overall, the present computational study facilitates the development of experimental epitope and escape mutation mapping studies.

5. Conclusions

Vaccines act by inducing strong immunity to counteract viral antigens presented by MHC-epitopes. However, their success is affected by virus evolution (e.g., escape mutations) within known protective epitopes; hence, multi-epitope, population-based vaccine constructs are preferred in order to generate a potent immunologic response against HBV. We demonstrate the quality of T cell epitopes among different HBV genotypes and reconstructed a candidate multi-epitope population-based vaccine. Our results suggest that among aa variations classified as polymorphisms do exit T-cell escape mutations and. In silico studies should be followed up with preclinical assays to validate the novelty of their findings.

Viral hepatitis (VH) is a global burden, and the WHO has put forth an ambitious goal to eliminate VH as a public threat by 2030. HBV contributes a vast majority (77%) of VH cases and there are no therapeutic cures for chronic hepatitis B infections (CHB). We hypothesized that epitope vaccines are a potential CHB treatment because they can induce strong immune systems with ability to achieve hepatitis B virus surface antigen (HBsAg) loss. While several trials have failed to produce effective vaccines against CHB from T cell epitopes, we aimed to investigate the repertoire of T cell epitopes from different HBV genotypes (A and D), MHC class II alleles with high population coverage in Botswana. In silico analyses were used to map promiscuous epitopes (15-mers) using alleles -9 MHC class II alleles-, and PreS/S sequences -genotype A and D- with high population coverage in Botswana. Some epitopes mapped within PreS/S conserved regions, and none were promiscuous to all alleles suggesting that multi-epitope, population-based vaccines (MEPBV) may be more effective candidate vaccines against CHB compared to previously reported broad potency epitope-based candidate vaccines. Highly promiscuous peptides may also be considered as candidate peptides for designing highly sensitive diagnostic chips since current serological kits may fail to detect other HBV clinical phenotypes. The mapped T epitopes exhibited high mean diversity among genotypes and others had coordinated amino acid variations that were genetically dependent on each other in order to escape epitope-HLA binding.

Supplementary Materials: The following are available online at http://www.mdpi.com/1999-4915/12/7/731/s1, Table S1: Supplementary_Main, Table S2: Supplementary_NetMHCIIPan_Results, Table S3 Supplementary-10,308-[B-I], Table S4 Supplementary-7743-[A&D], Table S5 Supplementary_vacc-gen-proteins.

Author Contributions: Conceptualization, W.T.C., M.A., B.B.P., T.M., R.M.M., S.M., J.T.B. and S.G.; formal analysis, W.T.C., P.O.K. and S.M.; funding acquisition, S.G.; investigation, W.T.C., K.K.S., T.G.B., J.T.B. and S.G.; methodology, W.T.C., M.A., B.B.P., T.M., R.M.M., S.M., J.T.B. and S.G.; supervision, M.A., E.Z., K.K.S., R.M.M., T.G.B., S.M., J.T.B. and S.G.; validation, P.O.K. and S.G.; writing—original draft, W.T.C.; writing—review and editing, M.A., E.Z.,

B.B.P., T.M., L.N.B., K.B., K.K.S., T.G.B., S.M., J.T.B. and S.G. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported through the Sub-Saharan African Network for TB/HIV Research Excellence (SANTHE), a DELTAS Africa Initiative [grant # DEL-15-006]. The DELTAS Africa Initiative is an independent funding scheme of the African Academy of Sciences (AAS)'s Alliance for Accelerating Excellence in Science in Africa (AESA) and supported by the New Partnership for Africa's Development Planning and Coordinating Agency (NEPAD Agency) with funding from the Welcome Trust [grant # 107752/Z/15/Z] and the UK government. The views expressed in this publication are those of the author(s) and not necessarily those of AAS, NEPAD Agency, Welcome Trust or the UK government.

Conflicts of Interest: All authors declare no conflict of interest.

Abbreviations

HBV	hepatitis B virus
HLA class II	human leukocyte antigen class II
CHB	chronic hepatitis B infection;
epi	epitopes
SB	strong binding
WB	weak binding
MHC	major histocompatibility complex
aa	amino acids

References

1. Ladep, N.G.; Lesi, O.A.; Mark, P.; Lemoine, M.; Onyekwere, C.; Afihene, M.; Crossey, M.M.; Taylor-Robinson, S.D. Problem of hepatocellular carcinoma in West Africa. *World J. Hepatol.* **2014**, *6*, 783–792. [CrossRef] [PubMed]
2. Spearman, C.W.; Afihene, M.; Ally, R.; Apica, B.; Awuku, Y.; Cunha, L.; Dusheiko, G.; Gogela, N.; Kassianides, C.; Kew, M.; et al. Hepatitis B in sub-Saharan Africa: Strategies to achieve the 2030 elimination targets. *Lancet Gastroenterol. Hepatol.* **2017**, *2*, 900–909. [CrossRef]
3. Kim, J.M.; Lee, K.W.; Sinn, D.H.; Choi, G.S.; Yi, N.J.; Kwon, C.H.D.; Suh, K.S.; Joh, J.W. Use of direct antiviral agents in liver transplant recipients with hepatitis C virus in Korea: 2-center experience. *Ann. Surg. Treat. Res.* **2018**, *95*, 147–151. [CrossRef] [PubMed]
4. Chisari, F.V.; Isogawa, M.; Wieland, S.F. Pathogenesis of hepatitis B virus infection. *Pathol. Biol. (Paris)* **2010**, *58*, 258–266. [CrossRef] [PubMed]
5. Rosenberg, W. Mechanisms of immune escape in viral hepatitis. *Gut* **1999**, *44*, 759–764. [CrossRef] [PubMed]
6. Karan, M.A.; Tascioglu, N.E.; Ozturk, A.O.; Palanduz, S.; Carin, M. The role of HLA antigens in chronic hepatitis B virus infection. *J. Pak. Med. Assoc.* **2002**, *52*, 253–256.
7. Zhang, Z.H.; Wu, C.C.; Chen, X.W.; Li, X.; Li, J.; Lu, M.J. Genetic variation of hepatitis B virus and its significance for pathogenesis. *World J. Gastroenterol.* **2016**, *22*, 126–144. [CrossRef]
8. Shi, Y.H. Correlation between hepatitis B virus genotypes and clinical outcomes. *Jpn. J. Infect. Dis.* **2012**, *65*, 476–482. [CrossRef]
9. Thio, C.L.; Thomas, D.L.; Karacki, P.; Gao, X.; Marti, D.; Kaslow, R.A.; Goedert, J.J.; Hilgartner, M.; Strathdee, S.A.; Duggal, P.; et al. Comprehensive analysis of class I and class II HLA antigens and chronic hepatitis B virus infection. *J. Virol.* **2003**, *77*, 12083–12087. [CrossRef]
10. Whitmire, J.K. Induction and function of virus-specific CD4+ T cell responses. *Virology* **2011**, *411*, 216–228. [CrossRef]
11. Juno, J.A.; van Bockel, D.; Kent, S.J.; Kelleher, A.D.; Zaunders, J.J.; Munier, C.M. Cytotoxic CD4 T Cells-Friend or Foe during Viral Infection? *Front. Immunol.* **2017**, *8*, 19. [CrossRef] [PubMed]
12. Luckheeram, R.V.; Zhou, R.; Verma, A.D.; Xia, B. CD4(+)T cells: Differentiation and functions. *Clin. Dev. Immunol.* **2012**, *2012*, 925135. [CrossRef] [PubMed]
13. Gaseitsiwe, S.; Maeurer, M.J. Identification of MHC class II binding peptides: Microarray and soluble MHC class II molecules. *Methods Mol. Biol.* **2009**, *524*, 417–426. [CrossRef] [PubMed]
14. Sette, A.; Fikes, J. Epitope-based vaccines: An update on epitope identification, vaccine design and delivery. *Curr. Opin. Immunol.* **2003**, *15*, 461–470. [CrossRef]

15. Dimitrov, I.; Flower, D.R.; Doytchinova, I.A. Improving in Silico Prediction of Epitope Vaccine Candidates by Union and Intersection of Single Predictors. *World J. Vaccines* **2011**, *1*, 8. [CrossRef]
16. Horvat, R.T.; El Atrouni, W.; Hammoud, K.; Hawkinson, D.; Cowden, S. Ribosomal RNA sequence analysis of Brucella infection misidentified as Ochrobactrum anthropi infection. *J. Clin. Microbiol.* **2011**, *49*, 1165–1168. [CrossRef]
17. Pollicino, T.; Cacciola, I.; Saffioti, F.; Raimondo, G. Hepatitis B virus PreS/S gene variants: Pathobiology and clinical implications. *J. Hepatol.* **2014**, *61*, 408–417. [CrossRef]
18. Simon, B.; Kundi, M.; Puchhammer, E. Analysis of mutations in the S gene of hepatitis B virus strains in patients with chronic infection by online bioinformatics tools. *J. Clin. Microbiol* **2013**, *51*, 163–168. [CrossRef]
19. Weber, B. Diagnostic impact of the genetic variability of the hepatitis B virus surface antigen gene. *J. Med. Virol.* **2006**, *78* (Suppl. 1), S59–S65. [CrossRef]
20. Zheng, J.; Lin, X.; Wang, X.; Zheng, L.; Lan, S.; Jin, S.; Ou, Z.; Wu, J. In Silico Analysis of Epitope-Based Vaccine Candidates against Hepatitis B Virus Polymerase Protein. *Viruses* **2017**, *9*, 112. [CrossRef]
21. Shi, T.D.; Wu, Y.Z.; Jia, Z.C.; Zou, L.Y.; Zhou, W. Therapeutic polypeptides based on HBV core 18-27 epitope can induce CD8+ CTL-mediated cytotoxicity in HLA-A2+ human PBMCs. *World J. Gastroenterol.* **2004**, *10*, 1902–1906. [CrossRef]
22. Dimitrov, I.; Garnev, P.; Flower, D.R.; Doytchinova, I. MHC Class II Binding Prediction-A Little Help from a Friend. *J. Biomed. Biotechnol.* **2010**, *2010*, 705821. [CrossRef]
23. Rosa, D.S.; Ribeiro, S.P.; Cunha-Neto, E. CD4+ T cell epitope discovery and rational vaccine design. *Arch. Immunol. Et Ther. Exp.* **2010**, *58*, 121–130. [CrossRef]
24. Schweitzer, A.; Horn, J.; Mikolajczyk, R.T.; Krause, G.; Ott, J.J. Estimations of worldwide prevalence of chronic hepatitis B virus infection: A systematic review of data published between 1965 and 2013. *Lancet* **2015**, *386*, 1546–1555. [CrossRef]
25. Choga, W.T.; Anderson, M.; Zumbika, E.; Moyo, S.; Mbangiwa, T.; Phinius, B.B.; Melamu, P.; Kayembe, M.K.; Kasvosve, I.; Sebunya, T.K.; et al. Molecular characterization of hepatitis B virus in blood donors in Botswana. *Virus Genes* **2019**, *55*, 33–42. [CrossRef]
26. Anderson, M.; Choga, W.T.; Moyo, S.; Bell, T.G.; Mbangiwa, T.; Phinius, B.B.; Bhebhe, L.; Sebunya, T.K.; Lockman, S.; Marlink, R.; et al. Molecular Characterization of Near Full-Length Genomes of Hepatitis B Virus Isolated from Predominantly HIV Infected Individuals in Botswana. *Genes (Basel)* **2018**, *9*, 453. [CrossRef]
27. Mbangiwa, T.; Kasvosve, I.; Anderson, M.; Thami, P.K.; Choga, W.T.; Needleman, A.; Phinius, B.B.; Moyo, S.; Leteane, M.; Leidner, J.; et al. Chronic and Occult Hepatitis B Virus Infection in Pregnant Women in Botswana. *Genes (Basel)* **2018**, *9*, 259. [CrossRef]
28. Anderson, M.; Gaseitsiwe, S.; Moyo, S.; Wessels, M.J.; Mohammed, T.; Sebunya, T.K.; Powell, E.A.; Makhema, J.; Blackard, J.T.; Marlink, R.; et al. Molecular characterisation of hepatitis B virus in HIV-1 subtype C infected patients in Botswana. *Bmc Infect. Dis.* **2015**, *15*, 335. [CrossRef]
29. Matthews, P.C.; Beloukas, A.; Malik, A.; Carlson, J.M.; Jooste, P.; Ogwu, A.; Shapiro, R.; Riddell, L.; Chen, F.; Luzzi, G.; et al. Prevalence and Characteristics of Hepatitis B Virus (HBV) Coinfection among HIV-Positive Women in South Africa and Botswana. *PLoS ONE* **2015**, *10*, e0134037. [CrossRef]
30. Wai, C.T.; Fontana, R.J. Clinical significance of hepatitis B virus genotypes, variants, and mutants. *Clin. Liver Dis.* **2004**, *8*, 321–352. [CrossRef]
31. Chu, C.J.; Lok, A.S. Clinical significance of hepatitis B virus genotypes. *Hepatology* **2002**, *35*, 1274–1276. [CrossRef]
32. Zhang, T.Y.; Guo, X.R.; Wu, Y.T.; Kang, X.Z.; Zheng, Q.B.; Qi, R.Y.; Chen, B.B.; Lan, Y.; Wei, M.; Wang, S.J.; et al. A unique B cell epitope-based particulate vaccine shows effective suppression of hepatitis B surface antigen in mice. *Gut* **2020**, *69*, 343–354. [CrossRef] [PubMed]
33. Patronov, A.; Doytchinova, I. T-cell epitope vaccine design by immunoinformatics. *Open Biol.* **2013**, *3*, 120139. [CrossRef] [PubMed]
34. Larsson, A. AliView: A fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* **2014**, *30*, 3276–3278. [CrossRef]
35. Ndung'u, T.; Gaseitsiwe, S.; Sepako, E.; Doualla-Bell, F.; Peter, T.; Kim, S.; Thior, I.; Novitsky, V.A.; Essex, M. Major histocompatibility complex class II (HLA-DRB and -DQB) allele frequencies in Botswana: Association with human immunodeficiency virus type 1 infection. *Clin. Diagn. Lab. Immunol.* **2005**, *12*, 1020–1028. [CrossRef] [PubMed]

36. Jensen, K.K.; Andreatta, M.; Marcatili, P.; Buus, S.; Greenbaum, J.A.; Yan, Z.; Sette, A.; Peters, B.; Nielsen, M. Improved methods for predicting peptide binding affinity to MHC class II molecules. *Immunology* **2018**, *154*, 394–406. [CrossRef]
37. Vita, R.; Mahajan, S.; Overton, J.A.; Dhanda, S.K.; Martini, S.; Cantrell, J.R.; Wheeler, D.K.; Sette, A.; Peters, B. The Immune Epitope Database (IEDB): 2018 update. *Nucleic Acids Res.* **2019**, *47*, D339–D343. [CrossRef]
38. Bell, T.G.; Yousif, M.; Kramvis, A. Bioinformatic curation and alignment of genotyped hepatitis B virus (HBV) sequence data from the GenBank public database. *Springerplus* **2016**, *5*, 1896. [CrossRef]
39. Miruka, C.O.; Matunda, N.C.; Ejekwumadu, N.J.; Mokembo, J.N. Design of a Recombinant Hepatitis B Vaccine Based on Stably Binding HLA-I Peptides. *J. Biomol. Res. Ther.* **2015**, *4*. [CrossRef]
40. Artimo, P.; Jonnalagedda, M.; Arnold, K.; Baratin, D.; Csardi, G.; de Castro, E.; Duvaud, S.; Flegel, V.; Fortier, A.; Gasteiger, E.; et al. ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Res.* **2012**, *40*, W597–W603. [CrossRef]
41. Racz, R.; Li, X.; Patel, M.; Xiang, Z.; He, Y. DNAVaxDB: The first web-based DNA vaccine database and its data analysis. *Bmc Bioinform.* **2014**, *15* (Suppl. 4), S2. [CrossRef]
42. Zhang, X.; Wang, J.; Lu, J.; Li, R.; Zhao, S. Correction to: Immunogenicity of adenovirus-vector vaccine targeting hepatitis B virus: Non-clinical safety assessment in non-human primates. *Virol. J.* **2018**, *15*, 137. [CrossRef]
43. Mizukoshi, E.; Sidney, J.; Livingston, B.; Ghany, M.; Hoofnagle, J.H.; Sette, A.; Rehermann, B. Cellular immune responses to the hepatitis B virus polymerase. *J. Immunol.* **2004**, *173*, 5863–5871. [CrossRef]
44. Schneider, J.; Chorlton, J.; Pearce, G.; Nicola, J.; Brown, D. Compositions for Inducing an Immune Response against Hepatitis B. U.S. Patent Application No. 11/986,294, 30 October 2008.
45. King, T.H.; Kemmler, C.B.; Guo, Z.; Mann, D.; Lu, Y.; Coeshott, C.; Gehring, A.J.; Bertoletti, A.; Ho, Z.Z.; Delaney, W.; et al. A whole recombinant yeast-based therapeutic vaccine elicits HBV X, S and Core specific T cells in mice and activates human T cells recognizing epitopes linked to viral clearance. *PLoS ONE* **2014**, *9*, e101904. [CrossRef]
46. Zeigler, D.F.; Gage, E.; Roque, R.; Clegg, C.H. Epitope targeting with self-assembled peptide vaccines. *Npj Vaccines* **2019**, *4*, 30. [CrossRef]
47. Falugi, F.; Petracca, R.; Mariani, M.; Luzzi, E.; Mancianti, S.; Carinci, V.; Melli, M.L.; Finco, O.; Wack, A.; Di Tommaso, A.; et al. Rationally designed strings of promiscuous CD4(+) T cell epitopes provide help to Haemophilus influenzae type b oligosaccharide: A model for new conjugate vaccines. *Eur. J. Immunol.* **2001**, *31*, 3816–3824. [CrossRef]
48. Publicover, J.; Gaggar, A.; Jespersen, J.M.; Halac, U.; Johnson, A.J.; Goodsell, A.; Avanesyan, L.; Nishimura, S.L.; Holdorf, M.; Mansfield, K.G.; et al. An OX40/OX40L interaction directs successful immunity to hepatitis B virus. *Sci. Transl. Med.* **2018**, *10*. [CrossRef]
49. Krebber, W.J.T.A.; Kessler, J.H.; Melief, C.J.M.; Kwappenberg, K.M.C. Synthetic Long Peptides (slp) for Therapeutic Vaccination Against Hepatitis B Virus Infection. U.S. Patent No. 10,376,576, 13 August 2019.
50. Chen, L.; Zhang, Y.; Zhang, S.; Chen, Y.; Shu, X.; Lai, J.; Cao, H.; Lian, Y.; Stamataki, Z.; Huang, Y. A novel T-cell epitope in the transmembrane region of the hepatitis B virus envelope protein responds upon dendritic cell expansion. *Arch. Virol.* **2019**, *164*, 483–495. [CrossRef]
51. Sette, A.; Sidney, J.; Southwood, S. Identification of Broadly Reactive DR Restricted Epitopes. U.S. Patent No. 6,413,517, 2 July 2002.
52. Kent, S. Immunomodulating Compositions, Uses Therefore and Processes for Their Production. U.S. Patent No. 2007/0248584 A1, 25 October 2007.
53. Godkin, A.; Davenport, M.; Hill, A.V. Molecular analysis of HLA class II associations with hepatitis B virus clearance and vaccine nonresponsiveness. *Hepatology* **2005**, *41*, 1383–1390. [CrossRef]
54. Shen, K.; Shen, L.; Wang, J.; Jiang, Z.; Shen, B. Understanding Amino Acid Mutations in Hepatitis B Virus Proteins for Rational Design of Vaccines and Drugs. *Adv. Protein. Chem. Struct. Biol.* **2015**, *99*, 131–153. [CrossRef]
55. Lerner, R.A.; Green, N.; Alexander, H.; Liu, F.T.; Sutcliffe, J.G.; Shinnick, T.M. Chemically synthesized peptides predicted from the nucleotide sequence of the hepatitis B virus genome elicit antibodies reactive with the native envelope protein of Dane particles. *Proc. Natl. Acad. Sci. USA* **1981**, *78*, 3403–3407. [CrossRef]

56. Milich, D.R.; Peterson, D.L.; Leroux-Roels, G.G.; Lerner, R.A.; Chisari, F.V. Genetic regulation of the immune response to hepatitis B surface antigen (HBsAg). VI. T cell fine specificity. *J. Immunol.* **1985**, *134*, 4203–4211. [PubMed]
57. Krchnak, V.; Mach, O.; Maly, A. Computer prediction of B-cell determinants from protein amino acid sequences based on incidence of beta turns. *Methods Enzym.* **1989**, *178*, 586–611. [CrossRef]
58. Wang, H.; Luo, H.; Wan, X.; Fu, X.; Mao, Q.; Xiang, X.; Zhou, Y.; He, W.; Zhang, J.; Guo, Y.; et al. TNF-alpha/IFN-gamma profile of HBV-specific CD4 T cells is associated with liver damage and viral clearance in chronic HBV infection. *J. Hepatol.* **2020**, *72*, 45–56. [CrossRef]
59. Pajot, A.; Michel, M.L.; Mancini-Bourgine, M.; Ungeheuer, M.N.; Ojcius, D.M.; Deng, Q.; Lemonnier, F.A.; Lone, Y.C. Identification of novel HLA-DR1-restricted epitopes from the hepatitis B virus envelope protein in mice expressing HLA-DR1 and vaccinated human subjects. *Microbes Infect.* **2006**, *8*, 2783–2790. [CrossRef]
60. Zhou, C.; Li, C.; Gong, G.Z.; Wang, S.; Zhang, J.M.; Xu, D.Z.; Guo, L.M.; Ren, H.; Xu, M.; Xie, Q.; et al. Analysis of immunological mechanisms exerted by HBsAg-HBIG therapeutic vaccine combined with Adefovir in chronic hepatitis B patients. *Hum. Vaccin. Immunother.* **2017**, *13*, 1989–1996. [CrossRef]
61. Bahrami, A.A.; Bandehpour, M.; Khalesi, B.; Kazemi, B. Computational Design and Analysis of a Poly-Epitope Fusion Protein: A New Vaccine Candidate for Hepatitis and Poliovirus. *Int. J. Pept. Res. Ther.* **2020**, *26*, 389–403. [CrossRef]
62. Sette, A.; Sidney, J.; Southwood, S.; Vitiello, M.A.; Livingston, B.D.; Celis, E.; Kubo, R.T.; Grey, H.M.; Robert, W. Inducing cellular immune responses to hepatitis B virus using peptide compositions. U.S. Patent No. 2005/0063983 A1, 24 March 2005.
63. Paoletti, E.; Gettig Russell, R.; Francis, J.-C.; Cox William, I.; Pincus Steven, E.; Johnson Gerard, P.; Limbach Keith, J.; Taisne Charles, D.E.; Riviere, M.; Norton Elizabeth, K.; et al. Genetically Engineered Vaccine Strain. AU 1997/010012 A, 3 January 1997.
64. Shin, Y.W.; Ryoo, K.H.; Hong, K.W.; Chang, K.H.; Choi, J.S.; So, M.; Kim, P.K.; Park, J.Y.; Bong, K.T.; Kim, S.H. Human monoclonal antibody against Hepatitis B virus surface antigen (HBsAg). *Antivir. Res.* **2007**, *75*, 113–120. [CrossRef]
65. Yuan, Q.; Zhang, T.; Luo, W.; Chen, Y.; Zhang, J.; Xia, N. Polypeptides and antibodies for treating HBV infection and related diseases. U.S Patent No. 9,751,914 B2, 5 September 2017.
66. Doytchinova, I.A.; Flower, D.R. VaxiJen: A server for prediction of protective antigens, tumour antigens and subunit vaccines. *Bmc Bioinform.* **2007**, *8*, 4. [CrossRef]
67. Klein, J.; Sato, A. The HLA system. First of two parts. *N. Engl. J. Med.* **2000**, *343*, 702–709. [CrossRef]
68. Xu, D.Z.; Wang, X.Y.; Shen, X.L.; Gong, G.Z.; Ren, H.; Guo, L.M.; Sun, A.M.; Xu, M.; Li, L.J.; Guo, X.H.; et al. Results of a phase III clinical trial with an HBsAg-HBIG immunogenic complex therapeutic vaccine for chronic hepatitis B patients: Experiences and findings. *J. Hepatol.* **2013**, *59*, 450–456. [CrossRef] [PubMed]
69. Kosinska, A.D.; Bauer, T.; Protzer, U. Therapeutic vaccination for chronic hepatitis B. *Curr. Opin. Virol.* **2017**, *23*, 75–81. [CrossRef] [PubMed]
70. Lobaina, Y.; Michel, M.L. Chronic hepatitis B: Immunological profile and current therapeutic vaccines in clinical trials. *Vaccine* **2017**, *35*, 2308–2314. [CrossRef]
71. Boni, C.; Janssen, H.L.A.; Rossi, M.; Yoon, S.K.; Vecchi, A.; Barili, V.; Yoshida, E.M.; Trinh, H.; Rodell, T.C.; Laccabue, D.; et al. Combined GS-4774 and Tenofovir Therapy Can Improve HBV-Specific T-Cell Responses in Patients With Chronic Hepatitis. *Gastroenterology* **2019**, *157*, 227–241 e227. [CrossRef]
72. Lok, A.S. Hepatitis B: 50 years after the discovery of Australia antigen. *J. Viral Hepat.* **2016**, *23*, 5–14. [CrossRef]
73. Lok, A.S.; McMahon, B.J.; Brown, R.S., Jr.; Wong, J.B.; Ahmed, A.T.; Farah, W.; Almasri, J.; Alahdab, F.; Benkhadra, K.; Mouchli, M.A.; et al. Antiviral therapy for chronic hepatitis B viral infection in adults: A systematic review and meta-analysis. *Hepatology* **2016**, *63*, 284–306. [CrossRef]
74. Lok, A.S.; Pan, C.Q.; Han, S.H.; Trinh, H.N.; Fessel, W.J.; Rodell, T.; Massetto, B.; Lin, L.; Gaggar, A.; Subramanian, G.M.; et al. Randomized phase II study of GS-4774 as a therapeutic vaccine in virally suppressed patients with chronic hepatitis B. *J. Hepatol.* **2016**, *65*, 509–516. [CrossRef]
75. Gaggar, A.; Coeshott, C.; Apelian, D.; Rodell, T.; Armstrong, B.R.; Shen, G.; Subramanian, G.M.; McHutchison, J.G. Safety, tolerability and immunogenicity of GS-4774, a hepatitis B virus-specific therapeutic vaccine, in healthy subjects: A randomized study. *Vaccine* **2014**, *32*, 4925–4931. [CrossRef]

76. 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*, 2221–2225. [CrossRef]
77. Thio, C.L.; Carrington, M.; Marti, D.; O'Brien, S.J.; Vlahov, D.; Nelson, K.E.; Astemborski, J.; Thomas, D.L. Class II HLA alleles and hepatitis B virus persistence in African Americans. *J. Infect. Dis.* **1999**, *179*, 1004–1006. [CrossRef]
78. Huang, C.J.; Wu, C.F.; Lan, C.Y.; Sung, F.Y.; Lin, C.L.; Liu, C.J.; Liu, H.F.; Yu, M.W. Impact of genetic heterogeneity in polymerase of hepatitis B virus on dynamics of viral load and hepatitis B progression. *PLoS ONE* **2013**, *8*, e70169. [CrossRef]
79. Ferrari, C.; Bertoletti, A.; Penna, A.; Cavalli, A.; Valli, A.; Missale, G.; Pilli, M.; Fowler, P.; Giuberti, T.; Chisari, F.V.; et al. Identification of immunodominant T cell epitopes of the hepatitis B virus nucleocapsid antigen. *J. Clin. Investig.* **1991**, *88*, 214–222. [CrossRef] [PubMed]
80. Tonetto, P.A.; Goncales, N.S.; Fais, V.C.; Vigani, A.G.; Goncales, E.S.; Feltrin, A.; Goncales, F.L., Jr. Hepatitis B virus: Molecular genotypes and HBeAg serological status among HBV-infected patients in the southeast of Brazil. *Bmc Infect. Dis.* **2009**, *9*, 149. [CrossRef] [PubMed]
81. Anderson, M.; Choga, W.T.; Moyo, S.; Bell, T.G.; Mbangiwa, T.; Phinius, B.B.; Bhebhe, L.; Sebunya, T.K.; Makhema, J.; Marlink, R.; et al. In Silico Analysis of Hepatitis B Virus Occult Associated Mutations in Botswana Using a Novel Algorithm. *Genes (Basel)* **2018**, *9*, 420. [CrossRef] [PubMed]
82. Tamori, A.; Kawada, N. HLA class II associated with outcomes of hepatitis B and C infections. *World J. Gastroenterol.* **2013**, *19*, 5395–5401. [CrossRef]
83. Thursz, M.R.; Kwiatkowski, D.; Allsopp, C.E.; Greenwood, B.M.; Thomas, H.C.; Hill, A.V. Association between an MHC class II allele and clearance of hepatitis B virus in the Gambia. *N. Engl. J. Med.* **1995**, *332*, 1065–1069. [CrossRef]
84. Wang, L.; Zou, Z.Q.; Wang, K. Clinical Relevance of HLA Gene Variants in HBV Infection. *J. Immunol. Res.* **2016**, *2016*, 9069375. [CrossRef]
85. Al-Qahtani, A.A.; Al-Anazi, M.R.; Abdo, A.A.; Sanai, F.M.; Al-Hamoudi, W.; Alswat, K.A.; Al-Ashgar, H.I.; Khalaf, N.Z.; Eldali, A.M.; Viswan, N.A.; et al. Association between HLA variations and chronic hepatitis B virus infection in Saudi Arabian patients. *PLoS ONE* **2014**, *9*, e80445. [CrossRef]
86. Yan, Z.H.; Fan, Y.; Wang, X.H.; Mao, Q.; Deng, G.H.; Wang, Y.M. Relationship between HLA-DR gene polymorphisms and outcomes of hepatitis B viral infections: A meta-analysis. *World J. Gastroenterol.* **2012**, *18*, 3119–3128. [CrossRef]
87. Li, Z.K.; Nie, J.J.; Li, J.; Zhuang, H. The effect of HLA on immunological response to hepatitis B vaccine in healthy people: A meta-analysis. *Vaccine* **2013**, *31*, 4355–4361. [CrossRef]
88. Yang, P.M.; Sung, J.L.; Chen, D.S. HLA-A, B, C and DR antigens in chronic hepatitis B viral infection. *Hepatogastroenterology* **1989**, *36*, 363–366.
89. Czaja, A.J.; Doherty, D.G.; Donaldson, P.T. Genetic bases of autoimmune hepatitis. *Dig. Dis. Sci.* **2002**, *47*, 2139–2150. [CrossRef]
90. Parkin, J.; Cohen, B. An overview of the immune system. *Lancet* **2001**, *357*, 1777–1789. [CrossRef]
91. Starzl, T.E.; Zinkernagel, R.M. Antigen localization and migration in immunity and tolerance. *N. Engl. J. Med.* **1998**, *339*, 1905–1913. [CrossRef]
92. Scully, C.; Georgakopoulou, E.A.; Hassona, Y. The Immune System: Basis of so much Health and Disease: 3. Adaptive Immunity. *Dent. Update* **2017**, *44*, 322–324, 327. [CrossRef]
93. Finlay, B.B.; McFadden, G. Anti-immunology: Evasion of the host immune system by bacterial and viral pathogens. *Cell* **2006**, *124*, 767–782. [CrossRef]
94. Thimme, R.; Wieland, S.; Steiger, C.; Ghrayeb, J.; Reimann, K.A.; Purcell, R.H.; Chisari, F.V. CD8(+) T cells mediate viral clearance and disease pathogenesis during acute hepatitis B virus infection. *J. Virol.* **2003**, *77*, 68–76. [CrossRef]
95. Rehermann, B. Intrahepatic T cells in hepatitis B: Viral control versus liver cell injury. *J. Exp. Med.* **2000**, *191*, 1263–1268. [CrossRef]
96. Hass, M.; Hannoun, C.; Kalinina, T.; Sommer, G.; Manegold, C.; Gunther, S. Functional analysis of hepatitis B virus reactivating in hepatitis B surface antigen-negative individuals. *Hepatology* **2005**, *42*, 93–103. [CrossRef]

97. Chang, J.J.; Lewin, S.R. Immunopathogenesis of hepatitis B virus infection. *Immunol. Cell Biol.* **2007**, *85*, 16–23. [CrossRef]
98. Albayrak, A.; Ertek, M.; Tasyaran, M.A.; Pirim, I. Role of HLA allele polymorphism in chronic hepatitis B virus infection and HBV vaccine sensitivity in patients from eastern Turkey. *Biochem Genet.* **2011**, *49*, 258–269. [CrossRef]