Received: 17 February 2022 | Revised: 25 April 2022 | Accepted: 27 April 2022

DOI: 10.1111/jvh.13702

JOURNAL OF VIRAL HEPATITIS JVH WILEY

ORIGINAL ARTICLE

Secreted hepatitis B virus splice variants differ by HBV genotype and across phases of chronic hepatitis B infection

Olivia Maslac[1,2] | Josef Wagner[1] | Vitina Sozzi[1] | Hugh Mason[1] | Jenny Svarovskaia[3] | Susanna Tan[3] | Anuj Gaggar[3] | Stephen Locarnini[1] | Lilly Yuen[1] | Margaret Littlejohn[1,4] | Peter A. Revill[1,2,5]

[1]Division of Molecular Research and Development, Victorian Infectious Diseases Reference Laboratory, Peter Doherty Institute for Infection and Immunity, Royal Melbourne Hospital, Melbourne, Victoria, Australia

[2]Department of Microbiology, Monash University, Clayton, Victoria, Australia

[3]Gilead Sciences, Foster City, California, USA

[4]Department of Infectious Diseases, University of Melbourne, Parkville, Victoria, Australia

[5]Department of Microbiology and Immunology, University of Melbourne, Parkville, Victoria, Australia

Correspondence
Peter Revill, Division of Molecular Research and Development, Victorian Infectious Disease Reference Laboratory, The Peter Doherty Institute for Infection and Immunity, Melbourne Health, 792 Elizabeth Street, Melbourne, Vic. 3000, Australia.
Email: peter.revill@mh.org.au

Funding information
The study was supported by NHMRC Project Grant APP1145977 and Gilead Sciences.

Abstract

Chronic hepatitis B (CHB) is characterized by progression through different phases of hepatitis B virus (HBV) infection and disease. Although not necessary for HBV replication, there is increasing evidence that HBV splice variants are associated with liver disease progression and pathogenesis. However, there have been no studies till date on the frequency or diversity of splice variants for different HBV genotypes across the phases of CHB. Next generation sequencing data from 404 patient samples of HBV genotype A, B, C or D in Phase I, Phase II or Phase IV of CHB was analysed for HBV splice variants using an in house bioinformatics pipeline. HBV splice variants differed in frequency and type by genotype and phase of natural history. Splice variant Sp1 was the most frequently detected (206/404, 51% of patients), followed by Sp13 (151/404 37% of patients). The frequency of variants was generally highest in Phase II (123/165, 75% of patients), a phase typically associated with enhanced immune activation, followed by Phase I (69/99, 70% of patients). Splice variants were associated with reduced hepatitis B e antigen (HBeAg) levels and statistically reduced likelihood of achieving HBsAg loss (functional cure) in Phase II patients for Sp1 and Sp13 (p = .0014 and .0156, respectively). The frequency of HBV splice variants in patient serum differed markedly by HBV genotype and phase of CHB natural history. The increased levels of HBV splice variants detected in CHB phase II patients compared with the higher replicative Phase I in particular warrants further investigation.

KEYWORDS
chronic hepatitis B, hepatitis B surface antigen, splicing

Abbreviations: CHB, Chronic hepatitis B; dlDNA, Duplex linear DNA; HBSP, Hepatitis B splice protein; HBV, Hepatitis B virus; HCC, hepatocellular carcinoma; pgRNA, pre-genomic RNA; Pol, Polymerase; TDF, tenofovir disoproxil; WT, wild-type.

Maslac and Wagner equal contributors.

Yuen, Littlejohn and Revill equal contributors.

1 | INTRODUCTION

More than 296 million people worldwide are living with chronic hepatitis B (CHB), for which there is currently no cure.[1] HBV-related mortality rates exceed 0.8 million individuals annually,[1] due to chronic HBV infection significantly elevating the likelihood of progression to cirrhosis and hepatocellular carcinoma (HCC).[2] This progression is gradual, as the course of chronic HBV infection is a dynamic process that illustrates the complex interactions between viral and host immune responses. The natural history of CHB is classified into five phases: (i) HBeAg-positive chronic infection (Phase I), (ii) HBeAg-positive chronic hepatitis (Phase II), (iii) HBeAg-negative chronic infection (Phase III), (iv) HBeAg-negative chronic hepatitis (Phase IV) and (v) HBsAg-negative phase (Phase V).[3] We have recently shown that HBV full-length haplotype sequence diversity increases across phases I, II and IV, for HBV genotypes (A–D)[4] and that increased diversity in baseline HBV sequences is associated with reduced likelihood of HBsAg loss on direct acting antiviral therapy in Phase II patients.[5]

Although the majority of HBV genomes are 3.2 kb in length, HBV variants generated by splicing of HBV pre-genomic RNA (pgRNA) are also frequently produced.[6,7] Alternative splicing of pgRNA generates variants that contain large genomic deletions predominantly within the polymerase (pol) and surface antigen (S) ORFs, with 17 splice variants (Sp1–17) characterized to date in serum and liver tissue from CHB patients or HBV cell culture studies[6,8] (Figure S1). Additional putative variants have also been identified,[7–10] suggesting the milieu of HBV splice variants is far more complex than previously appreciated. Sp1 is the most common splice variant detected in cell culture models and patient samples,[8,11] constituting up to 30% of HBV RNA in transfected cells.[7,12] Splicing removes sequences required for coding of functional polymerase and genome circularization, therefore splice variants are reverse transcribed by HBV pol provided *in trans* from wild-type (WT) pgRNA, and generally secreted as virus particles containing duplex-linear DNA (dlDNA) genomes.[13] RNA splicing is dispensable for the HBV replication cycle,[12] although splice variants may impact the replication of wild-type (WT) HBV,[14–17] due in part to the expression of novel proteins resulting from intron removal and subsequent formation of a new exon sequence. For example, splice variant Sp1 produces a novel precore protein that reduces RNA expression and decreases WT HBV replication,[17] as well as the hepatitis B splice protein (HBSP) associated with disease progression[18–20] and suppression of hepatocyte apoptosis.[21] In contrast, a protein encoded by Sp7 has been implicated in increased HBV replication.[15] HBV splice variants are also linked with liver disease severity, with Sp1 associated with more advanced liver disease[20,22,23] including liver cancer.[24] Splice variants have been found to reduce the likelihood of responding to IFN-α therapy[8]; however, their association with responses to direct acting antiviral therapy is unknown. To date, there have been no studies on the frequency and type of HBV splice variants present across different phases of CHB natural history.

In this study, we utilized deep sequencing data from baseline samples of patients subsequently treated with the direct acting antiviral tenofovir disoproxil (TDF) to investigate the type and frequency of HBV splice variants present across phases of HBV natural history and their association with treatment outcome, particularly HBsAg loss (functional cure).

2 | METHODS

2.1 | Patient cohorts

Baseline serum samples were available from HBV infected patients enrolled in clinical trials (Gilead Sciences) from Phases I, II and IV of CHB natural history.[4] Samples were not available from Phase III patients from clinical trials as these patients are not currently recommended for therapy.[25] Phase I patients (HBeAg-positive chronic infection) were from trial GS-US-203-0101, NTC00507507[26] ($n = 99$, predominantly HBV genotypes B and C),[26] Phase II patients (HBeAg-positive chronic hepatitis B) were from trial GS-US-174-0103, NTC00116805[27] ($n = 165$, HBV genotypes A–D) and Phase IV patients (HBeAg-negative chronic hepatitis B) were from trial GS-US-174-0102 trial, NTC00117676[27] ($n = 140$, HBV genotypes A to D) (Table S1).[27]

2.2 | Extraction of viral DNA, genome-length PCR and deep sequencing.

Whole HBV genome amplification and deep sequencing had previously been performed on DNA extracted from 200 μl of patient serum.[5,28] Briefly, HBV DNA was extracted from 200 μl of serum using the QIAGEN DNeasy Blood and Tissue Kit, or QIAamp DNA Mini Kit (QIAGEN) according to the manufacturer's instructions. A genome-length PCR product (fragment A) and an overlapping PCR product covering the primer binding region (fragment B) were generated using Hi-Fidelity DNA polymerase (Roche Diagnostics) and pooled in a 1:1 molar ration. PCR products from patient samples in the Phase I, Phase II and Phase IV studies were separately pooled for high throughput sequencing. Library preparation was performed using the Nextera XT Kit (Illumina), and NGS performed using the MiSeq platform (Illumina) according to the manufacturer's protocol at Micromon Genomics (Melbourne, Australia).[4,5,28] Viral genotype was determined during the parent study.[26,27]

2.3 | NGS data processing pipeline

The NGS FASTA.gz files were analysed using the in-house HBV-Quasiminer software package developed at VIDRL (Melbourne, Australia).[28] Briefly, the pipeline was established to process short NGS reads for HBV deep-sequencing analysis, facilitate batch processing of multiple NGS datasets, determine HBV genotype, map short NGS reads to a reference sequence of an equivalent HBV genotype, and to identify single/multiple nucleotide variants for each NGS dataset. For each dataset, the mapped reads represented in BAM format were screened using the Pindel program[29] to identify those that harbour large deletion break points which are potential splice sites.

2.4 | Identification of splice variants

Data files generated by Pindel were processed using custom scripts to map the large deletion breakpoints (>200 nucleotides in length) to a reference HBV genome sequence of equivalent genotype, and then, the coordinates were converted to equivalent positions on a reference HBV genotype D genome, numbered from the EcoR1 site (standard numbering system for HBV splice variants). Although the pipeline recognized the sequence at the breakpoint coordinates (splice donor and acceptor site locations) of known HBV splice variants[6] (Sp1–Sp14), it only called splice variants with single intron deletions because the large deletions detected from different short-length reads of a sample could not be guaranteed to have come from the same genome sequence, a limitation of using short-read NGS datasets. This means that reads classed as Sp1 could also include minor variants Sp2 and Sp4 (referred to as Sp1*), Sp9 may have included variants Sp10 and Sp12 (referred to as Sp9*) and Sp13 may have included variants Sp7 and Sp8 (referred to as Sp13*). As Sp3, Sp5, Sp6, Sp11 and Sp14 do not use the same combinations of donor and acceptor sites as any other known splice variants, they were detected independently (Figure S1). It was therefore not possible to compare frequency of single spliced variants to double spliced variants. As the software pipeline only identifies splice variants from the conserved donor and acceptor splice site sequences, this method will not detect replication derived deletions as these would have different sequences at the breakpoints.

The frequency of splice variants was determined by dividing the number of these variants by the total number of reads mapped to the same genome locations (specifically the splice donor site for each splice event detected). To investigate differences in the type and frequency of splice variants in different HBV genotypes and across multiple phases of CHB natural history, two levels of analysis were performed. Firstly, analysis data were plotted by CHB phase separated by genotype and splice group. Secondly, analysis data were plotted by splice groups and CHB phase separated by genotype.

2.5 | Clinical parameters

Hepatitis B virus markers (HBV viral load, HBeAg level, HBsAg level and alanine aminotransferase [ALT]) at baseline were obtained from the clinical trial data.[26,27] Haplotype number for each sample was determined as previously described.[4] HBsAg loss in the Phase II patients was defined as a negative HBsAg test, within 244 weeks on treatment.[30]

2.6 | Statistical analyses

The proportion of patients with splice variants and associations with HBV variants or clinical markers was determined using chi-square analysis and Fishers exact test. Graphical presentation and statistical analyses were performed using GraphPad Prism 9.0.1 for MacOS and V9.1.1 for Windows (GraphPad Software, www.graphpad.com/scientific-software/prism/). Normality of data was tested with both D'Agostino and Pearson and Shapiro–Wilk normality tests, and non-parametric data between the two groups was analysed using the Mann–Whitney test. A two-sided p value of .05 was considered statistically significant. To evaluate non-parametric data from three groups, a one-way ANOVA was performed using the Kruskal–Wallis test with Dunn's multiple comparisons test. Correlation was performed using Spearman non-parametric test. The p value was corrected using the false discovery rate method.

2.7 | Statistical and multiple logistic regression analysis to correlate splice number with HBsAg status

The frequency of Sp1* and Sp13* in the HBsAg negative and positive groups, respectively, were plotted using GraphPad Prism (v 9.0.1), and the differences were statistically assessed using Mann–Whitney test. MLR was used to analyse whether there was a statistical association between splice frequency and the patient losing HBsAg following treatment compared to those who remained HBsAg positive. MLR was done in GraphPad Prism for macOS (v9.0.2).

3 | RESULTS

3.1 | The number of patients with splice variants detected differed across phases of chronic hepatitis B disease and HBV genotype

NGS data encompassing the complete HBV genome sequence previously obtained from baseline patient samples was examined for the presence of HBV splice variants. The number of patients with splice variants detected differed significantly across the different disease phases. The highest proportion of patients harbouring splice variants was detected in Phase II (123/165 patients, 75%) followed by Phase I (69/99 patients, 70%) ($p = .39$). The proportion of patients positive for any splice variant was lowest in Phase IV (60/140 patients, 43%, $p < .001$ for both Phase I and Phase II) (Figure S2A). In Phase II, there was significantly higher number of patients with splice variants across all genotypes A–D compare with the number of patients without splice variants (t test $p = .0004$) (Figure S2A).

This pattern was similar when individual genotypes were examined except for genotype C. For both genotypes A and D, there was a significantly higher proportion of patients with splice variants in Phase II compared with Phase IV ($p < .001$ for both). For genotype B, there was also a significantly higher proportion of patients with splice variants in Phase II compared with both Phase I and Phase IV ($p = .046$ and $p = .01$, respectively). The number of patients with genotypes A or D in Phase I were not sufficient to allow comparison to the other phases. For genotype C, there was no significant difference in the proportion of patients with splice variants across any of the disease phases.

The proportion of patients positive for each individual splice variant also varied across disease phases and genotypes (Figure S2B). The most commonly detected splice variants were Sp1* (51% of patients [206/404]) and Sp13* (37% of patients [151/404]). These splice variants were detected in all HBV genotypes, in all three CHB phases (Figure S2B) in contrast to Sp5 and Sp9*, which were detected only in genotypes B and C. Phase I genotype A could not be included as there were no patients with genotype A in the Phase I group). The diversity of splice variants was greatest amongst genotypes B and C, with all splice variants identified, and lowest in genotype D with only 2 splice variants identified (Sp1* and Sp13*). Variants Sp6, 11 and 14 were detected in very few patients and were not included in the analysis.

3.2 | The frequency of HBV splice variants differed across phases of hepatitis B disease and HBV genotype

The frequency and type of splice variants within patients varied across phases of CHB natural history and HBV genotype. For disease phases, the frequency of splice variants was generally highest in Phase II (HBeAg positive hepatitis B) and lowest in Phase I (HBeAg positive chronic infection) (Figure 1). This was statistically significant for Sp1* in genotype B ($p = .0001$) and C ($p = .0005$), for SP9* in genotype B ($p = .0022$) and for Sp13* in genotype C ($p = .0022$). Within each disease phase, the frequency of splice variants varied by HBV genotype, but Sp1* was detected at a higher frequency than all other splice variants in all genotypes (Figure 2). This was statistically significant for Phase II in genotype A, Phase I and II for genotype B and C and Phase II and IV for genotype D with P ranging between <0.0001 and 0.0478 (Figure 2).

3.3 | Splice variants were detected less frequently at baseline in patients with genotypes A and D HBV who achieved HBsAg loss on treatment

Univariate analysis was employed to examine if there was any association between the level of splice variants at baseline and the likelihood of achieving HBsAg loss (functional cure) on TDF therapy. As HBsAg loss was almost exclusively limited to patients in the Phase II study[27] with genotypes A and D, analysis of associations of splice variants at baseline with subsequent HBsAg loss was restricted to these genotypes and this cohort. The frequency of splice variants Sp1* and Sp13* at baseline in patients was significantly lower in patients with genotypes A and D who achieved HBsAg loss on TDF treatment (mean 0.42% and 0.08%, respectively), compared with patients who remained HBsAg positive on therapy (mean 3.25% and 0.21%, respectively, Table S2 and Figure 3A for Sp1*, $p = .0014$; Figure 3B for Sp13*, $p = .0156$).

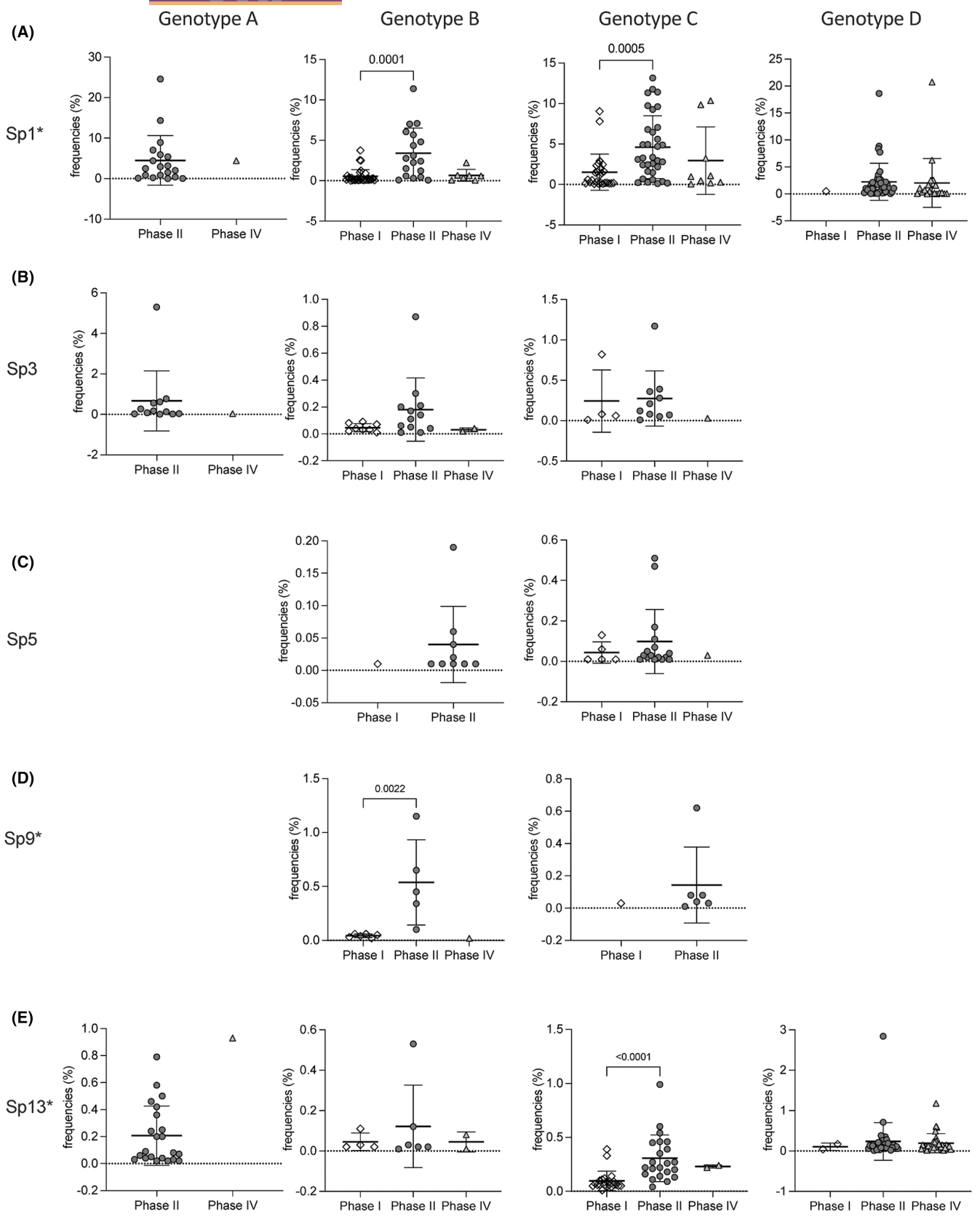
The association between splice frequency and HBsAg loss was stronger for Sp1* than Sp13*, indicated by the larger area under the ROC curve for Sp1*, Figure 3C,E (Area 0.8571, 95% CI 0.75–0.96, $p = .0025$) compared with Sp13*, Figure 3D,F (Area 0.7452, 95% CI 0.59–0.90, $p = .0168$).

Univariate logistic regression analysis showed that splice frequencies ($p = .037$), HBeAg titre ($p = .028$) and HBsAg titre ($p = .0243$) were independently associated with increased likelihood of HBsAg loss for splice variant Sp1* (Table 1). Multivariate logistic regression analysis confirmed the association between HBsAg loss and Sp1* splice frequencies after adjustment for HBsAg titre and viral load ($p = .031$ and $p = .048$ respectively). However, the association with HBeAg titre did not reach statistical significance in the multivariate analysis, indicating that HBeAg had a confounding effect on splice frequency (Table 1). For Sp13*, none of the clinical markers were independently associated with increased likelihood of HBsAg loss either in univariate or multivariate analysis (Table 1). However, Sp13* splice frequencies were statistically associated with HBsAg loss at 90% confidence interval ($p = .065$).

3.4 | The frequency of splice variants were positively associated with haplotype number and negatively associated with HBeAg levels at baseline in Phase II

The frequency of HBV splice variants at baseline in Phase II was compared to HBV markers, namely HBV haplotype number, hepatitis B e antigen (HBeAg level), hepatitis B surface (HBsAg) level, viral load, alanine aminotransferase (ALT) and age. The frequency of splice variants Sp1* and Sp13* in individual patients with HBV genotype A or D (combined) was strongly positively associated with increased haplotype number (Spearman $r = .48$, $p = .0003$, linear regression $R^2 = 0.11$ $p = .0133$ for Sp1* (Figure 4A) and Spearman $r = .55$, $p = <.0001$, linear regression $R^2 = 0.42$ $p = <.0001$ for Sp13* (Figure 4B). Further, the same splice variants were negatively associated with baseline HBeAg levels (Spearman $r = -.34$, $p = .015$ for Sp1* (Figure 4C) and Spearman $r = -.31$, $p < .0001$, linear regression

FIGURE 1 Frequency of splice variant reads per patient within HBV genotypes stratified by CHB natural history. Frequency of splice variant reads per patient across CHB natural history Phases I, II and IV within genotypes A–D. Each point on the graph represents the percentage of splice variant reads on the *y*-axis for one individual. Data are presented in a scatter dot plot for Sp1* (A), for Sp3 (B), for Sp5 (C), for SP9* (D) and for SP13* (E) with the line indicating the mean and the error bars indicating the standard deviation. Data were tested for normality with four different normality and log normality tests available in Prism 9 for macOS. In all cases, not all datasets within a figure panel were normally distributed, and therefore, the non-parametric two-tailed Mann–Whitney test was used for two group comparison and the non-parametric Kruskal–Wallis test was used for three group comparisons. For three group comparison, the Dunn's multiple comparisons test was used and a false discovery adjusted *p* value <.05 was reported. Non-significant observations are not shown in the graph panels



R^2 = 0.42, $p < .0001$, for Sp13* (Figure 4D). HBsAg titre and viral load were also negatively associated with Sp13* splice frequencies (Spearman $r = -.31$, $p = .032$, linear regression $R^2 = 0.13$, $p = .011$ for HBsAg (Figure 4F) and linear regression $R^2 = 0.2652$, $p = .0001$ for viral load (Figure 4H). Both splice variants were also negatively associated with age by Spearman correlation (Figure 4K,L).

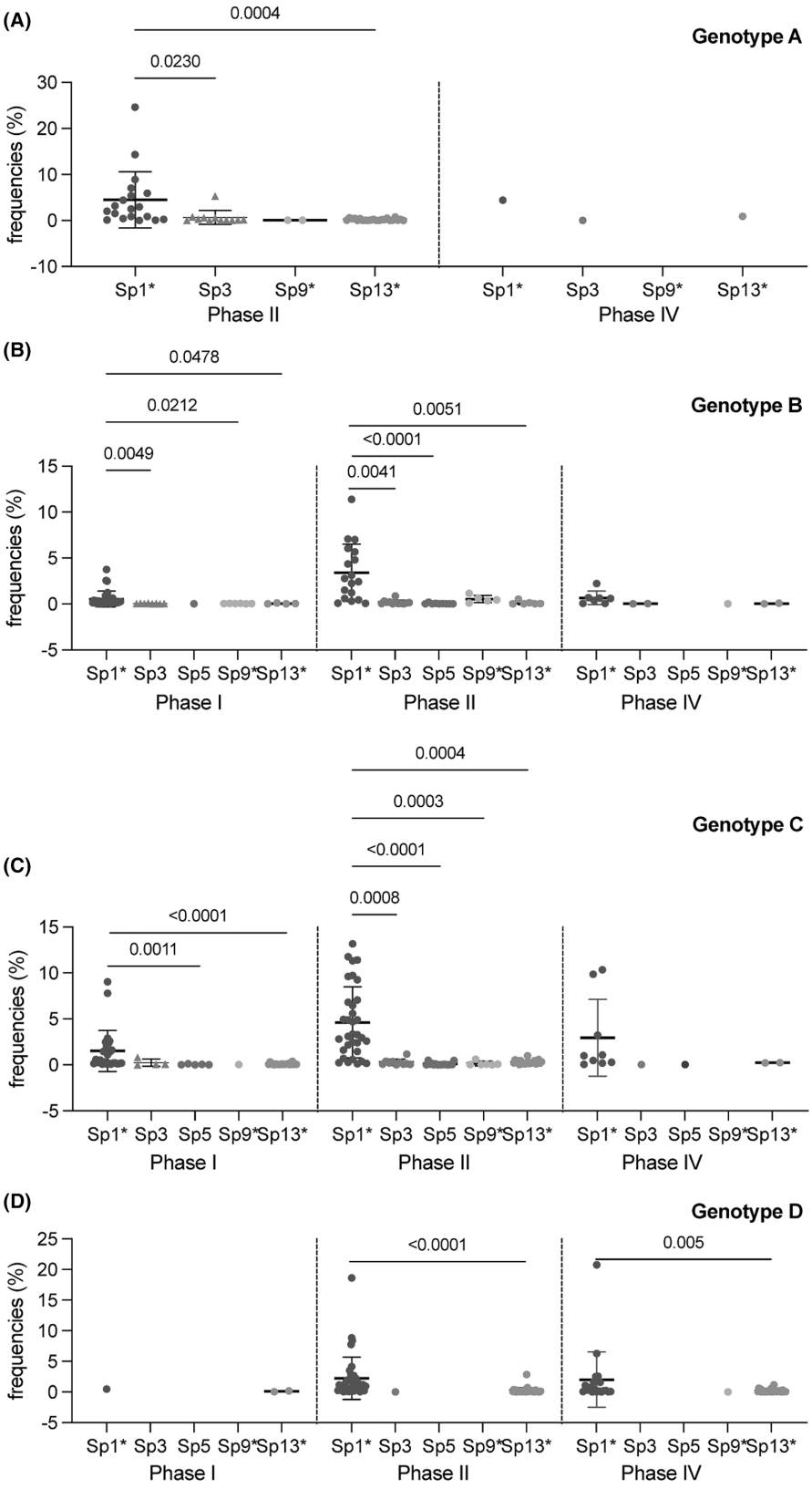


FIGURE 2 Frequency of splice variant reads per patient within HBV genotypes and CHB natural history stratified by splice variants. The number of splice frequencies for Phases I, II and IV for HBV genotypes A (A), B (B), C (C) and D (D) are shown. Each point on the graph represents the percentage of splice variant reads for one individual. Data are presented in a scatter dot plot with the line indicating the mean and the error bars indicating the standard deviation. Data were tested for normality with four different normality and log normality tests available in Prism 9 for macOS. In all cases, not all datasets within a figure panel were normally distributed, and therefore, the non-parametric Kruskal–Wallis test with Dunn's multiple comparisons test was used and a false discovery adjusted p value <.05 was reported. Non-significant observations are not shown in the graph panels

Separate analysis for genotypes A and D between Sp1* (Figure S3A) and Sp13* (Figure S3B) and clinical markers revealed fewer significant associations compared to the combined genotype A and D analysis. For patients with genotype B and C CHB, only viral load was positively associated with Sp1* frequency, and was only significant for genotype C (Figure S4A). No association was identified for Sp13* (Figure S4B). Interestingly, viral load was positively associated with Sp1* for genotype C, and negatively associated with Sp13* for genotypes A and D.

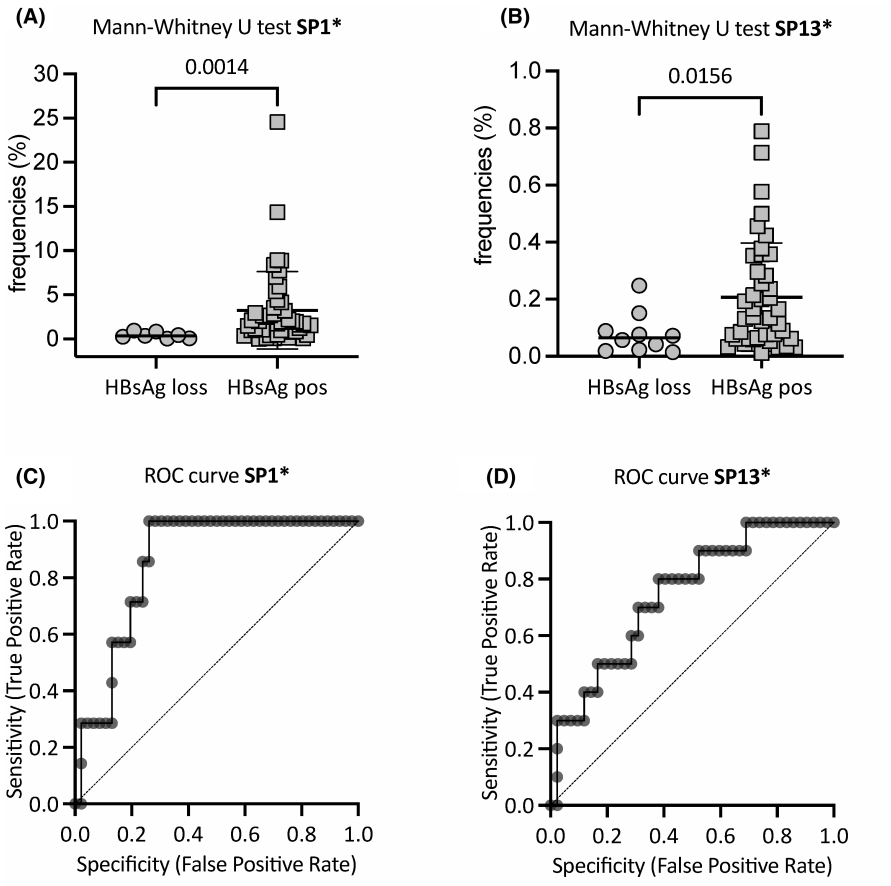


FIGURE 3 Splice variants Sp1* and SP13* were significantly lower in HBsAg negative patients. Frequency of splice variant reads per patient comparing patients who lost HBsAg on therapy and those who remained HBsAg positive for Sp1* (A) and Sp13* (B) variants. Each point on the graph represents the percentage of splice variant reads for one individual. Significant differences are indicated on the graphs. Error bars represent mean with standard deviation between percentages of splice reads per patient in each cohort. The Mann–Whitney test (Prism version 9) was used for A and B. The receiver operator characteristic (ROC) curve provides a more accurate test for Sp1* (C) compare with Sp13* (D) because the area under the ROC curve for Sp1* was 0.8571 ($p = .0025$) (E) which was higher compare with Sp13* (Area = 0.7452, $p = .0168$) (F). The ROC test was done using the multiple logistic regression analysis (Prism version 9). The ROC curve shows the true positive rate (sensitivity on *Y* axis) against the false positive rate (specificity on the *X* axis)

(E)

Area under the ROC curve	
Area	0.8571
Std. Error	0.05335
95% confidence interval	0.7526 to 0.9617
P value	0.0025

(F)

Area under the ROC curve	
Area	0.7452
Std. Error	0.08078
95% confidence interval	0.5869 to 0.9036
P value	0.0168

3.5 | The presence of splice variants were associated with HBV BCP/PC variants

Having identified an association between the frequency of splice variants and reduced HBeAg levels in Phase II patients, we next investigated if there were associations between HBV splice variants and HBV precore (PC) or basal core promoter (BCP) variants known to impact HBeAg levels (PC_G1896A, BCP_A1762T, BCP_G1764A). In HBeAg negative (Phase IV) patients, BCP variants were detected less frequently in patients who expressed splice variants ($p = .0147$ in Figure S5A). PC HBV variants were more prevalent in Phase II patients who had HBV splice variants than patients that did not harbour splice variants ($p = .0056$ in Figure S5D). The presence of splice variants was also associated with the *frequency* of BCP and PC variants within individual patients (Figure S5B, C, E and F). The median frequency of PC variants within individual patients was significantly higher in the presence of most splice variants for patients in Phase II (p values between .001 and .043 in Figure S5E); however, both BCP and PC median frequency was lower in the presence of splice variants in Phase IV although this only reached statistical significance for Sp1* and BCP variants ($p = .037$ in Figure S5C). We also compared the presence of other splice variants with HBeAg levels for patients in the HBeAg positive phases (Phase I and II), however this only reached significance for Sp3* in samples from Phase I patients (Figure S6).

4 | DISCUSSION

This is the first detailed analysis of secreted HBV splice variants Sp1–Sp14 detected from patient serum across disease phases of CHB natural history and their association with treatment response in the setting of direct acting antiviral (TDF) therapy. We observed significant differences in the number of patients with splice variants, and the type and frequency of HBV splice variants found within patients at different disease phases of CHB. The finding that splice variant Sp1* was the most common variant supports previous studies.[7] HBV splice variants were detected in a greater proportion of patients as well as at a higher frequency in the 'immune clearance' Phase II of HBV natural history (HBeAg-positive hepatitis) than in the 'immune tolerant' Phase I (HBeAg-positive infection) particularly in HBV genotypes B and C. This finding

TABLE 1 Multiple logistic regression analysis of clinical variables with HBsAg loss

	Univariate analysis			
	ß estimated (SE, 95% CI)		***p* Value**	**Odds ratio (95% CI)**
Parameter SP1*				
Splice frequencies	−1.764 (0.85, −3.91 to −0.48)		**.037**	0.171 (0.2–0.62)
HBeAg titre	3.853 (1.75, 1.21 to 8.3)		**.028**	47.13 (3.35–4023)
HBsAg titre	3.058 (1.36, 0.71 to 6.2)		**.0243**	21.28 (2.03–494.5)
Viral load	0.445 (0.63, −0.69 to 1.8)		.477	1.56 (0.5–6.08)
Parameter SP13*				
Splice frequencies	−9.355 (5.06, −21.96 to −1.68)		.065	0.0001 (0.29–0.19)
HBeAg titre	0.741 (0.6, −0.29 to 2.1)		.215	2.097 (0.75–8.13)
HBsAg titre	1.175 (1.06, −0.81 to 3.42)		.266	3.237 (0.5–30.46)
Viral load	0.477 (0.56, −0.56 to 1.68)		.394	1.612 (0.57–5.34)
	Multivariate analysis			
	Estimate	**Standard error**	**95% CI**	***p* Value**
Parameter SP1*				
Splice frequencies + HBeAg titre	−1.337	0.8407	−3.463 to −0.1326	.111
Splice frequencies + HBsAg titre	−2.724	1.265	−5.857 to −0.7633	**.031**
Splice frequencies + viral load	−2.111	1.066	−4.739 to −0.5561	**.048**
Splice frequencies + HBeAg + HBsAg + viral load	−1.953	1.190	−4.980 to −0.1147	.1
Parameter SP13*				
Splice frequencies + HBeAg titre	−9.269	5.152	−22.20 to −1.301	.072
Splice frequencies + HBsAg titre	−22.34	12.21	−53.91 to −5.296	.067
Splice frequencies + viral load	−9.151	5.046	−22.10 to −1.377	.07
Splice frequencies + HBeAg + HBsAg + viral load	−21.57	11.22	−52.28 to −5.963	.055

Note: In the multiple logistic regression model, the dependent (or outcome) variable Y was the HBsAg status with the positive outcome (HBsAg loss) was coded with 1 and the negative outcome (HBsAg positive) was coded with 0. MLR was done in Prism 9 for macOS. In the 'Univariate Analysis', the independent variables X (splice frequency and three clinical variables HBeAg titre, HBsAg titre and viral load) were analysed individually with HBsAg status (outcome variable Y). From the univariate analysis, the ß estimate (including standard and 95% confidence interval), the *p* value and the odds ratio (including 95% confidence interval) are reported. In the 'Multivariate Analysis', each of the three clinical factors (HBeAg, HBsAg and viral load) were added individually to the splice frequency variable and together. From the multivariate analysis, the ß estimate (including standard and 95% confidence interval) and the *p* value are reported. A significant *p* value <.05 is highlighted in bold font.

was unexpected, as HBV viral loads are generally higher in Phase I than Phase II, with a number of previous studies having shown that higher frequency of secreted splice variants is associated with higher HBV viral load.[24,31] While it is also possible that splice variants occurring with low frequency could be underestimated in patients with lower amounts of HBV DNA in the serum, especially in Phase IV patients with lower viral loads, this does not explain the genotype differences observed.

The proportion of patients with splice variants was lower in Phase IV, compared with both Phase I and II for genotypes A, B and D, but this was not observed for HBV genotype C. Indeed, genotype C showed no significant difference in proportion of patients with splice variants across the disease phases. We have previously shown that HBV genetic diversity increases across CHB phases for genotypes A, B and D, but not HBV genotype C.[4] Genotype C in particular exhibits striking differences in viral diversity, including the type and frequency of splice variants compared with other HBV genotypes. Taken together with our findings in the current study, this provides further evidence that chronic HBV is not a 'one size fits all' disease. Since HBV genotype C is generally thought to be the most pathogenic and oncogenic of the HBV genotypes,[32] the role of HBV haplotypes and splice variants in disease progression for HBV genotype C in particular warrants further investigation.

The frequency of splice variants was also associated with reduced HBeAg levels in Phase II patients and concommitant increases in the abundance and frequency of the HBV PC variant. It is interesting that in genotype C, there was a positive association between the frequency of splice variant Sp1* and HBeAg level, whilst for genotypes A and D, this association was negative adding further evidence to genotype differences in HBV splicing. Although removal of the major HBV intron to generate Sp1* or Sp13* also removes the terminal amino acid of the precore and overlapping core protein,[17] this will not impact the coding sequence of the HBeAg produced following post-translation modification of the precore protein, as this region of the precore protein is removed from the HBeAg.[33]

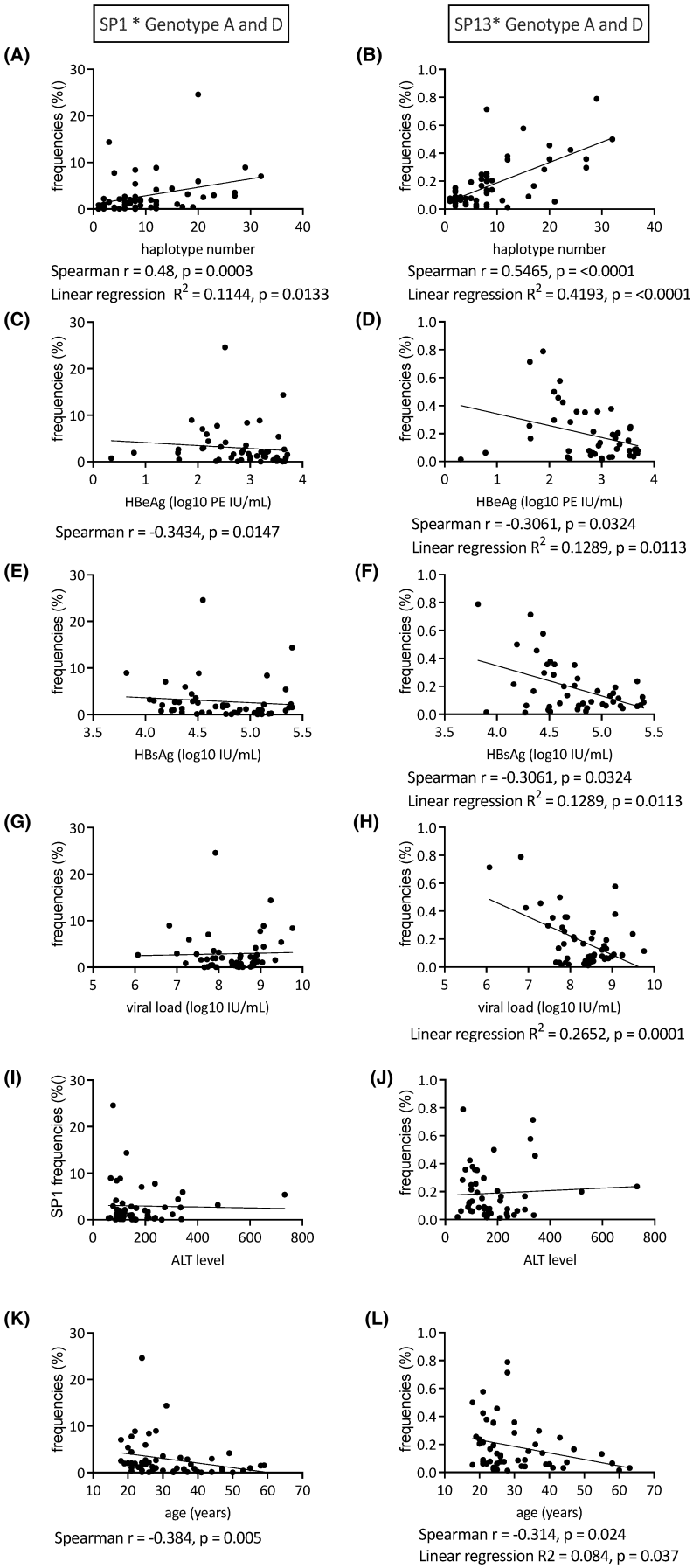


FIGURE 4 Association of splice variants Sp1* and Sp13* with clinical markers for genotype A and D. The association of frequency of splice variants Sp1* and Sp13* at baseline in Phase II with haplotype number (A and B) and HBV clinical markers, for HBeAg (C and D), HBsAg (E and F), viral load (G and H), ALT level (I and J) and patient age (K and L) are shown for A and D genotype combined. The correlation analysis was done using Spearman rho (*r*) correlation and simple linear regression analysis (Prism version 9). From the Spearman *r* analysis, the correlation coefficient value '*r*' and the *p* value is shown. From the linear regression analysis, the 'R^{2}' value and the *p* value is shown. Any significant association with a $p < .05$ is shown

Lower HBsAg antigen levels were observed in genotype A and D patients with higher levels of splice variants Sp1* and Sp13*; however, this association only reached statistical significance for Sp13*. Since removal of the major intron also removes the HBsAg coding sequence, the association with reduced HBsAg levels were not unexpected, however, the reason for the association of spliced variants with reduced HBeAg levels in Phase II patients is unclear.

Phase II is commonly referred to as the 'immune clearance' phase of HBV natural history and is characterized by increased activation of host immune responses, and the emergence of HBV specific antibodies that may ultimately lead to seroclearance of hepatitis B e antigen (HBeAg), or (rarely) hepatitis B surface antigen (HBsAg). In response to host immune pressure, HBV variants encoding mutations in the basal core promoter (BCP) or precore protein may be selected, which reduce or eliminate HBeAg production, but permit viral replication.[5,33] Indeed, we have recently shown that these variants are selected during Phase I (immune tolerant) CHB.[28] Interestingly, all splice variants identified in patient serum till date encode large deletions in the HBsAg gene rendering them incapable of expressing HBsAg, and the impact of this on host immune responses warrants further investigation.

Differences were also observed in the type and frequency of HBV splice variants across HBV genotypes (A–D). Greatest diversity was observed in the genotypes B and C, with all variant types identified. Sp1* and Sp13* were detected in the serum of patients across the major HBV genotypes (A–D), Sp3 was restricted to genotypes A to C, and Sp5 and Sp9* were restricted to genotypes B and C. The restriction of Sp5 to HBV genotypes B and C has been previously observed in cell culture studies of HBV spliced RNA[7] and in patient derived sequences.[7] This is most likely due to sequence variation at the Sp5 donor splice site (nt 2087) with the consensus donor splice sequence only present in genotype B and C sequences.[6,7] The restriction of Sp9* to genotypes B and C does not appear to be due to similar sequence differences, as Sp9 was originally described in a genotype A sequence,[6] and Sp9 has also previously been detected in genotype D patient samples and genotype A and D in cell culture samples.[7]

We also showed that the frequency of splice variants Sp1* or Sp13* at baseline was associated with reduced likelihood of HBsAg loss on subsequent TDF therapy for HBV genotypes A and D. The association between splice frequency and HBsAg loss was stronger for Sp1* than Sp13* (Figure 3E,F) and was confirmed by multivariate analysis adjusting for both HBsAg level and viral load for Sp1* (Table 1, multivariate analysis section). The association of lower levels of splice variants in the baseline samples of patients who subsequently lose HBsAg on therapy was concordant with our previous finding that lower HBV diversity is associated with increased likelihood of HBsAg loss in these patients[5] and suggests that HBV genome homogeneity at baseline is a major contributor to subsequent HBsAg loss on therapy. This suggests that the more heterogeneous the viral quasispecies pool, including the presence of splice variants, the lower the likelihood of achieving functional cure on TDF treatment. The mechanism for this association needs to be defined.

Sp1 is the most frequently detected splice variant in cell culture and chronic HBV patients[7,34,35]; however, ours is the first report showing that Sp13 is also prevalent in patient serum across phases of chronic HBV natural history. Initially identified in hepatoma cell lines transfected with replication competent HBV DNA,[36] Sp13 has also been identified integrated within the genome of the PLC/PRF/5 cell line derived from primary liver carcinoma,[6,14,35,37,38] with expression of the Sp13 Polymerase-Surface fusion transcript and protein associated with reduced HBV replication.[14] Sp1 has also been shown to modulate HBV replication,[17] and the novel hepatitis B splice protein encoded by Sp1 has been shown to inhibit apoptosis[21] and is associated with disease progression,[19,23,39] promotes hepatoma cell migration and invasion[40] and induces T-cell responses in CHB patients and HLA-transgenic mice.[41,42] Since both Sp1 and Sp13 are known to impact HBV replication in cell culture studies,[14,17] it remains to be determined if their increased frequency in Phase II of HBV natural history impacts HBV replication in patients.

Our study has some limitations. The first is that patient samples from different phases of CHB were not longitudinal samples from the same patient. This was unavoidable due to the retrospective nature of the studies, but future prospective studies on longitudinal samples from the same patient as they move through the different phases of CHB for different HBV genotypes would be informative. Second, it was not possible to determine the frequency of individual splice variants that had multiple intron deletions, due to the short reads generated by the Illumina Miseq NGS protocol. Nor was it possible to analyse minor variants Sp15–17 using our pipeline. Future targeted studies using single molecule (PacBio) sequencing are warranted to determine the contribution of minor splice species to HBV natural history. Unfortunately, samples were unavailable from phase III (HBeAg negative, HBV infection), as patients in this phase are not currently recommended for antiviral therapy and so were not included in the clinical trials. We speculate that as HBV viral loads are low in this phase, splice variants would also be present at low frequency, but this requires experimental confirmation. RNAseq studies of intrahepatic HBV splice variants in Phase III patients would be illuminating in this regard; however, liver tissue was not available for the current study. Finally, as our study was restricted to serum-derived HBV DNA, we were unable to assess the contribution of splice-derived RNA, or DNAs that may be integrated into the host genome across phases of natural history, particularly Sp13, previously identified integrated in the genome of PLC/PRF/5 (Alexander) cells.[13,14,37] Liver biopsy studies across phases of HBV natural history may provide further insights regarding the contribution of integrated splice variants to HBV natural history and pathogenesis.

SIGNIFICANCE

In conclusion, this study is significant because it provides new insights into the diversity of secreted HBV splice variants across phases of CHB natural history and genotype. The finding that secreted splice variants were more frequent in CHB phase II compared

with phases I or IV was unexpected, and futher studies on the contribution of HBV splice variants to HBV natural history, disease progression and treatment response are warranted.

ACKNOWLEDGEMENTS

We thank all of the patients and clinicians who contributed to the clinical trials for this study. The authors thank Laura McCoullough and Tania Candy for their contribution to the discussion and assistance with the preparation of this manuscript, respectively. Open access publishing facilitated by The University of Melbourne, as part of the Wiley – The University of Melbourne agreement via the Council of Australian University Librarians.

CONFLICT OF INTEREST

PAR had received research funding from Gilead Sciences. JS, ST and AG were employees of Gilead Sciences during this study.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from Gilead Sciences. Restrictions apply to the availability of these data, which were used under license for this study. Data are available from the author(s) with the permission of Gilead Sciences.

ORCID

Peter A. Revill https://orcid.org/0000-0003-2715-2541

REFERENCES

1. WHO. *Hepatitis B*. WHO; 2020. https://www.who.int/en/news-room/fact-sheets/detail/hepatitis-b. Accessed January 10, 2022.
2. Lavanchy D. Hepatitis B virus epidemiology, disease burden, treatment, and current and emerging prevention and control measures. *J Viral Hepat*. 2004;11(2):97-107.
3. European Association for the Study of the Liver. EASL 2017 clinical practice guidelines on the management of hepatitis B virus infection. *J Hepatol*. 2017;67(2):370-398.
4. Wagner J, Yuen L, Littlejohn M, et al. Analysis of hepatitis B virus haplotype diversity detects striking sequence conservation across genotypes and chronic disease phase. *Hepatology*. 2020;73:1652-1670.
5. Bayliss J, Yuen L, Rosenberg G, et al. Deep sequencing shows that HBV basal core promoter and precore variants reduce the likelihood of HBsAg loss following tenofovir disoproxil fumarate therapy in HBeAg-positive chronic hepatitis B. *Gut*. 2017;66(11):2013-2023.
6. Gunther S, Sommer G, Iwanska A, Will H. Heterogeneity and common features of defective hepatitis B virus genomes derived from spliced pregenomic RNA. *Virology*. 1997;238(2):363-371.
7. Lim CS, Sozzi V, Littlejohn M, et al. Quantitative analysis of the splice variants expressed by the major hepatitis B virus genotypes. *Microb Genom*. 2021;7(1):mgen000492.
8. Chen J, Wu M, Wang F, et al. Hepatitis B virus spliced variants are associated with an impaired response to interferon therapy. *Sci Rep*. 2015;5:16459.
9. Lam AM, Ren S, Espiritu C, et al. Hepatitis B virus capsid assembly modulators, but not nucleoside analogs, inhibit the production of extracellular pregenomic RNA and spliced RNA variants. *Antimicrob Agents Chemother*. 2017;61(8):e01734-18.
10. Betz-Stablein BD, Topfer A, Littlejohn M, et al. Single-molecule sequencing reveals complex genome variation of hepatitis B virus during 15 years of chronic infection following liver transplantation. *J Virol*. 2016;90(16):7171-7183.
11. Sommer G, van Bommel F, Will H. Genotype-specific synthesis and secretion of spliced hepatitis B virus genomes in hepatoma cells. *Virology*. 2000;271(2):371-381.
12. Wu HL, Chen PJ, Tu SJ, Lin MH, Lai MY, Chen DS. Characterization and genetic analysis of alternatively spliced transcripts of hepatitis B virus in infected human liver tissues and transfected HepG2 cells. *J Virol*. 1991;65(4):1680-1686.
13. Abraham TM, Lewellyn EB, Haines KM, Loeb DD. Characterization of the contribution of spliced RNAs of hepatitis B virus to DNA synthesis in transfected cultures of Huh7 and HepG2 cells. *Virology*. 2008;379(1):30-37.
14. Park GS, Kim HY, Shin HS, Park S, Shin HJ, Kim K. Modulation of hepatitis B virus replication by expression of polymerase-surface fusion protein through splicing: implications for viral persistence. *Virus Res*. 2008;136(1-2):166-174.
15. Ma ZM, Lin X, Wang YX, Tian XC, Xie YH, Wen YM. A double-spliced defective hepatitis B virus genome derived from hepatocellular carcinoma tissue enhanced replication of full-length virus. *J Med Virol*. 2009;81(2):230-237.
16. Tsai KN, Chong CL, Chou YC, et al. Doubly spliced RNA of hepatitis B virus suppresses viral transcription via TATA-binding protein and induces stress granule assembly. *J Virol*. 2015;89(22):11406-11419.
17. Wang YL, Liou GG, Lin CH, et al. The inhibitory effect of the hepatitis B virus singly-spliced RNA-encoded p21.5 protein on HBV nucleocapsid formation. *PLoS One*. 2015;10(3):e0119625.
18. Soussan P, Garreau F, Arnufl B, Pol S, Kremsdorf D. HBSP protein and related defective particles is associated to liver fibrosis. *Proceedings of the Molecular Biology of Hepatitis B Viruses*. Woodshole, MA of conference; 2004.
19. Soussan P, Garreau F, Zylberberg H, Ferray C, Brechot C, Kremsdorf D. In vivo expression of a new hepatitis B virus protein encoded by a spliced RNA. *J Clin Invest*. 2000;105(1):55-60.
20. Soussan P, Tuveri R, Nalpas B, et al. The expression of hepatitis B spliced protein (HBSP) encoded by a spliced hepatitis B virus RNA is associated with viral replication and liver fibrosis. *J Hepatol*. 2003;38(3):343-348.
21. Wu SX, Chen WN, Jing ZT, Liu W, Lin XJ, Lin X. Hepatitis B spliced protein (HBSP) suppresses fas-mediated hepatocyte apoptosis via activation of PI3K/Akt signaling. *J Virol*. 2018;92(23):e01273-18.
22. Redelsperger F, Lekbaby B, Mandouri Y, et al. Production of hepatitis B defective particles is dependent on liver status. *Virology*. 2012;431(1–2):21-28.
23. Soussan P, Pol J, Garreau F, et al. Expression of defective hepatitis B virus particles derived from singly spliced RNA is related to liver disease. *J Infect Dis*. 2008;198(2):218-225.
24. Bayliss J, Lim L, Thompson AJ, et al. Hepatitis B virus splicing is enhanced prior to development of hepatocellular carcinoma. *J Hepatol*. 2013;59(5):1022-1028.
25. Terrault NA, Bzowej NH, Chang KM, et al. AASLD guidelines for treatment of chronic hepatitis B. *Hepatology*. 2016;63(1):261-283.
26. Chan HL, Chan CK, Hui AJ, et al. Effects of tenofovir disoproxil fumarate in hepatitis B e antigen-positive patients with normal levels of alanine aminotransferase and high levels of hepatitis B virus DNA. *Gastroenterology*. 2014;146(5):1240-1248.
27. Marcellin P, Heathcote EJ, Buti M, et al. Tenofovir disoproxil fumarate versus adefovir dipivoxil for chronic hepatitis B. *N Engl J Med*. 2008;359(23):2442-2455.
28. Yuen L, Revill PA, Rosenberg G, et al. HBV variants are common in the 'immune-tolerant' phase of chronic hepatitis B. *J Viral Hepat*. 2020;27(10):1061-1070.
29. Ye K, Schulz MH, Long Q, Apweiler R, Ning Z. Pindel: a pattern growth approach to detect break points of large deletions and

medium sized insertions from paired-end short reads. *Bioinformatics*. 2009;25(21):2865-2871.
30. Marcellin P, Buti M, Krastev Z, et al. Kinetics of hepatitis B surface antigen loss in patients with HBeAg-positive chronic hepatitis B treated with tenofovir disoproxil fumarate. *J Hepatol*. 2014;61(6):1228-1237.
31. Preiss S, Littlejohn M, Angus P, et al. Defective hepatitis B virus DNA is not associated with disease status but is reduced by polymerase mutations associated with drug resistance. *Hepatology*. 2008;48(3):741-749.
32. Kramvis A. Genotypes and genetic variability of hepatitis B virus. *Intervirology*. 2014;57(3–4):141-150.
33. Messageot F, Salhi S, Eon P, Rossignol JM. Proteolytic processing of the hepatitis B virus e antigen precursor. Cleavage at two furin consensus sequences. *J Biol Chem*. 2003;278(2):891-895.
34. Kremsdorf D, Lekbaby B, Bablon P, et al. Alternative splicing of viral transcripts: the dark side of HBV. *Gut*. 2021;70(12):2373-2382.
35. Sommer G, Heise T. Posttranscriptional control of HBV gene expression. *Front Biosci*. 2008;13:5533-5547.
36. Huang HL, Jeng KS, Hu CP, Tsai CH, Lo SJ, Chang C. Identification and characterization of a structural protein of hepatitis B virus: a polymerase and surface fusion protein encoded by a spliced RNA. *Virology*. 2000;275(2):398-410.
37. Ramirez R, van Buuren N, Gamelin L, et al. Targeted long-read sequencing reveals comprehensive architecture, burden, and transcriptional signatures from hepatitis B virus-associated integrations and translocations in hepatocellular carcinoma cell lines. *J Virol*. 2021;95(19):e0029921.
38. Ziemer M, Garcia P, Shaul Y, Rutter WJ. Sequence of hepatitis B virus DNA incorporated into the genome of a human hepatoma cell line. *J Virol*. 1985;53(3):885-892.
39. Soussan P, Garreau F, Arnufl B, Pol S, Kremsdorf D. HBSP protein and related defective particles is associated to liver fibrosis. Woodshole, MA; 2004.
40. Chen WN, Chen JY, Jiao BY, et al. Interaction of the hepatitis B spliced protein with cathepsin B promotes hepatoma cell migration and invasion. *J Virol*. 2012;86(24):13533-13541.
41. Bayard F, Godon O, Nalpas B, et al. T-cell responses to hepatitis B splice-generated protein of hepatitis B virus and inflammatory cytokines/chemokines in chronic hepatitis B patients. ANRS study: HB EP 02 HBSP-FIBRO. *J Viral Hepat*. 2012;19(12):872-880.
42. Mancini-Bourgine M, Bayard F, Soussan P, et al. Hepatitis B virus splice-generated protein induces T-cell responses in HLA-transgenic mice and hepatitis B virus-infected patients. *J Virol*. 2007;81(10):4963-4972.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Maslac O, Wagner J, Sozzi V, et al. Secreted hepatitis B virus splice variants differ by HBV genotype and across phases of chronic hepatitis B infection. *J Viral Hepat*. 2022;29:604-615. doi: 10.1111/jvh.13702