Infection (2019) 47:793–803
https://doi.org/10.1007/s15010-019-01306-5

ORIGINAL PAPER

Analysis of hepatitis B virus-mixed genotype infection by ultra deep pyrosequencing in Sudanese patients, 2015–2016

Khalid Abdallah Enan[1] · Claudia Minosse[2] · Abdel Rahim Mohammed El Hussein[1] · Marina Selleri[2] · Emanuela Giombini[2] · Maria Rosaria Capobianchi[2] · Isam Mohamed Elkhidir[3] · Mohamed Omer Mustafa[1] · Osama Mohamed Khair[3] · Dina Ahamed Hassan[1] · Anna Rosa Garbuglia[2]

Received: 21 September 2018 / Accepted: 3 April 2019 / Published online: 8 April 2019

Abstract

Purpose The frequency of detection of HBV co-infection with multiple HBV genotypes is influenced by the detection method; usually co-infections are detected by multiplex PCR or hybridization assays, and are rarely confirmed by sequencing and conventional cloning. The objective of this study was to confirm by ultra-deep pyrosequencing (UDPS) mixed HBV infections, previously detected by multiplex-nested PCR.

Methods Sixteen samples from HBV co-infected Sudanese patients detected by multiplex-nested PCR, were amplified targeting the P/S region and sequenced by UDPS.

Results The only genotypes identified using UDPS were D and E, while A, B, C and F genotypes, previously detected by multiplex-nested PCR, were not detected. Specifically, 10 samples were shown to be mono-infected (D or E); in 3 out of 10 mono-infected D patients, a subtype combination was observed: D1 + D7 in 2 cases and D2 + D6 in 1 case. The remaining 6 subjects were D + E co-infected (harboring different mixtures of D subtypes).

Conclusions Overall, UDPS is more effective than multiplex-nested PCR for identifying multiple HBV genotypes and subtypes infections.

Keywords Hepatitis B virus · High-throughput nucleotide sequencing · Sudan · Co-infection

Introduction

Hepatitis B virus (HBV) is a DNA virus belonging to the *Hepadnaviridae* family showing high genetic diversity. At least 9 genotypes (A–I) have been identified, showing more than 8% diversity at nucleotide (nt) level, and a putative 10th genotype (J) has also been described [1]. Several subtypes have been identified for all but E, F and G genotypes [2]. Different HBV genotypes are distributed in different geographic areas: A and D are diffused in Europe, Africa, and United States of America (USA). Genotypes B and C are diffused in the Far East, whereas genotype E has been observed in West and Central Africa, but also in East Africa (Sudan and Madagascar). Genotype F is prevalent in Mexico and South America [3], genotype G has been described in France and USA [4], and genotype H in America and Japan [5]. HBV co-infection with two or more genotypes has previously been detected with serological [6, 7] and genetic typing [8, 9]. In 0.9% of infected intravenous drug users even triple infection with genotypes A, B, and C has been

Khalid Abdallah Enan and Claudia Minosse contributed equally to this work.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s15010-019-01306-5) contains supplementary material, which is available to authorized users.

✉ Anna Rosa Garbuglia
argarbuglia@iol.it

[1] Department of Virology, Central Laboratory, The Ministry of Higher Education and Scientific Research, 7099 Khartoum, Sudan

[2] Laboratory of Virology, National Institute for Infectious Diseases "Lazzaro Spallanzani"-IRCCS, Via Portuense, 292, 00149 Rome, Italy

[3] Department of Microbiology and Parasitology, Faculty of Medicine, University of Khartoum, 7099 Khartoum, Sudan

detected [10]. Co-infection frequencies are variable [11], and do not appear to be influenced by geographic area; rather, the method of detection seems to influence the reported co-infection frequency. Most of the reported co-infections with different genotypes are identified using multiplex PCR or hybridization assay, and they are rarely confirmed by sequencing and conventional cloning [11]. Moreover, the PCR-based genotyping assay does not take into account the genetic variability of local strains and some misleading results may occur. Toan and coworkers, using RFLP-PCR for HBV genotype detection, found that the HBV E genotype was more prevalent in Asian samples than in African samples, contradicting previous findings on HBV genotype distribution [12].

Ultra-deep Pyrosequencing (UDPS) is very useful to resolve mixed genotype infections in HCV-infected patients [13]. Moreover, such knowledge is also necessary for applying proper clinical management practices. In fact, correct HBV genotyping is important in clinical practice to predict the patients' response to different therapies [14, 15], and to determine which patients are more likely to develop resistance to nucleoside analogues [16–18]. Furthermore, a recent study demonstrated HBV genotypes and subtypes change during antiviral treatment, suggesting a different sensitivity to the drug among different subtypes during tenofovir therapy [18], due to a constant antiviral drug pressure. Therefore, an exact determination of HBV subtypes could predict the unfavourable response to an antiviral treatment.

In the present study, Sudanese HBV-positive samples, previously classified as mixed infection with multiplex-nested PCR, have been analyzed by UDPS to confirm the co-infections, and phylogenetic analysis has been performed. Furthermore, the P/S region has been sequenced using the Sanger method, to detect possible mutations in the antigenic sites.

Materials and methods

Sample collection and serological testing

A total of 454 plasma samples were selected from discarded HBsAg-positive blood donations (negative for HIV and HCV serological markers) from various States in Sudan (Khartoum, Elgazira, Kassala, River Nile, Northern Kordofan, Port-Sudan, El Fawa, and Northern Sudan).

Commercial ELISA kits (Prechek Bio, Inc, USA) were used to detect HBsAg according to manufacturer's instructions.

The study was designed and performed according to the Helsinki Declaration, and we received ethical approval from Ethical Committee of University of Khartoum", Khartoum, Sudan (meeting n.8/15-1-2015). All patients provided written informed consent before participating in this study.

DNA extraction and amplification

DNA was extracted using a commercial DNA extraction Kit (Analytikjena, Germany) according to manufacturer's instructions, and stored at − 20 °C until use. A genotyping system based on multiplex-nested PCR was employed to assign HBV genotypes, based on pre-S1 region, as previously described [19]. The PCR primers sequences used are shown in Table 1. The samples which were shown to be co-infected, were analyzed twice, in separate runs, to confirm the results.

In the 16 samples where DNA was available, UDPS analysis was carried out. To obtain a P/S amplicon for direct sequencing (DS) and UDPS, a nested PCR, previously described by Caballero et al. [20], was used, which targets the reverse transcriptase domain and the overlapping S gene (505 nt). To determine the sensitivity of primer set used in

Table 1 Primer sequences used for HBV genotyping by nested PCR

Primers	Sequences	Position	Specificity
Primer sequences used for the first PCR			
P1	5′-TCA CCA TAT TCT TGG GAA CAA GA-3′	nt 2823-2845	Universal
S1-2	5′-CGA ACC ACT GAA CAA ATG GC-3′	nt 685-704	Universal
Primer sequences used for the nested PCR (mix A)			
B2	5′-GGC TCM AGT TCM GGA ACA GT-3′	nt 67-86	Types A to E specific
BA1R	5′-CTC GCG GAG ATT GAC GAG ATG T-3′	nt 113-134	Type A specific
BB1R	5′-GGT CCT AGG AAT CCT GAT GTT G-3′	nt 165-186	Type B specific
BC1R	5′-CAG GTT GGT GAG TGA CTG GAG A-3′	nt 2979-2996	Type C specific
Primer sequences used for the nested PCR (mix B)			
B2R	5′-GGA GGC GGA TYT GCT GGC AA-3′	nt 3078-3097	Types D- to F-specific
BD1	5′-GCC AAC AAG GTA GGA GCT-3′	nt 2979-2996	Type D-specific
BE1	5′-CAC CAG AAA TCC AGA TTG GGA CCA-3′	nt 2955-2978	Type E-specific
BF1	5′-GYT ACG GTC CAG GGT TAC CA-3′	nt 3032-3051	Type F-specific

UDPS mixed HBV infection detection, we mixed different HBV genotypes (A + B + E) and (C + D + F) in varying ratio from 100 to 10,000 IU/ml. We used as template samples from diagnostic routine, for which we knew both the HBV viral load (determined by Roche COBAS® TaqMan® HBV, Roche Diagnostics GmbH, Mannheim, Germany and genotype) and HBVgenotype, which was identified by pol region sequencing [21]. The detection sensitivity for each HBV genotype was 100 IU/ml.

The HBsAg coding region was amplified according to Zhang et al. [22] and Sanger sequenced to detect possible mutations in the antigenic sites, particularly in the MHR and in the "a" determinant.

DS and UDPS

DS was performed on the automated ABI Prism 3100 instrument, using BigDye Terminator cycle sequencing kit (Applied Biosystems, Warrington, UK). Sanger sequences were used as reference genomes to map reads in UDPS analysis and to check the assembly accuracy of UDPS reads.

UDPS was performed with GS junior 454 (Roche); in this case multiplex identifiers (MIDs) for sample barcoding were added to the second-round PCR primers. The barcodes are then used to identify reads originated from the same sample, allowing to access multiple sequencing data of the same patient.

A median of 3160 reads (range 1682–4715) for each sample was obtained. All reads shorter than 400 bp, with more than one mismatch on the MID, two on the specific primer, three on the universal primer M13 and with indels or showing more than three gaps, were discarded. All identical sequences were clustered using CD-HIT software [23]. Haplotype sequences were identified as the most frequent representative reads and their frequencies computed as the number of observed reads identical to 96%. A threshold was identified as a minimum of five reads and 1% of the reads abundance in the single sample.

For each haplotype, the HBV genotype was assigned on the basis of similarity with reference sequences downloaded from NCBI database (genotype A, B, C, D, E and recombinant D/E: 46, 16, 41, 90, 69 and 6 sequences, respectively). A sample was classified as co-infected when more than one genotype was present.

For each haplotype [24], the intra-patient nucleotide diversity was calculated by Phylip DNADIST software package and expressed as mean substitutions/site $\times 10^{-4}$.

A phylogenetic analysis was performed by aligning the consensus haplotype sequences of each sample, using Muscle program [25] with 134 sequences genotype D (subtype D1–7) or E. One sequence (AY934771_SO_A; genotype A) has been used as the outgroup. All references are summarized in supplementary data (S1 Table).

The best fit model was identified and a phylogenetic analysis was performed using the maximum-likelihood method with Kimura-2-parameter model + G, implemented in MEGA6 software [26]. Bootstrap values (estimated with 500 replications) > 70% were considered significant.

For each haplotype, the inter-patient nucleotide distance and the mean intra-patient diversity were calculated by MEGA6 software and expressed as number of base substitutions per site between sequences.

The Sanger sequences obtained in this study have been submitted to GenBank: (MG267061-MG267076).

Results

Sanger sequencing and UDPS analysis of mixed infections

All samples classified as mixed infection by multiplex-nested PCR ($n = 16$) gave positive results with UDPS analysis. After the pipeline correction of UDPS raw data, a median of 2046 (range 1051–3514) reads for each sample was obtained.

In Table 2 the comparison of the genotyping/subtyping obtained by UDPS and multiplex-nested PCR is reported. The percentage of reads clustered in each haplotype, as well as the intra-patient nucleotide diversity of each haplotype is also shown.

The only genotypes identified using UDPS were genotypes D (15 samples) and E (7 samples), while genotypes A, B, C and F, detected by multiplex PCR, were not confirmed by UDPS.

According to the UDPS results, a mono-infection was identified in 10 cases (62.5%): 9 samples harbored genotype D, and 1 sample genotype E. In more detail, a unique genotype (either D or E) was detected in 7 patient samples, while 3 samples harbored two subtypes of genotype D (D1 + D7 in 2 cases; D2 + D6 in one case). The other 6 patients (37.5%) harbored combined D and E genotypes; of these, in three samples a D subtype combination was observed: D1 + D7 in 2 cases; D1 + D2 in 1 case.

Overall, the median diversity in mono-infected was significantly lower than in co-infected patients (median, range 0.62, 0.32–2.34 vs 1.54, 1.01–2.34, respectively; $p = 0.03$) (Table 2).

No specific geographic distribution of the HBV D subtypes seems to exist (Fig. 1), in agreement with previous studies [27–30]. Analysis of the D subtypes data revealed that D1 was present in 3 of 4 regions included in our study, while D7 was present in 2 of 4 regions. D2 and D6 were detected only in the Khartoum region. This is in agreement with previous studies [27–29].

Table 2 Comparison of results obtained with ultra deep pyrosequencing (UDPS) and multiplex-nested PCR analysis

ID patient[a]	Multiplex-nested PCR results	UDPS results											
	Genotype		Genotype	HAPL1	%[b]	Diversity[c]	HAPL2	%	Diversity	HAPL3	%	Diversity	Total corrected reads
SDG4	B_D	Mono-infected	D	**D7**	100	0.26							1955
SDF5	A_B_D_E		D	**D1**	66.0	1.74	**D7**	34.0	0.45				1464
SDF10	A_B_D_E_F		D	**D1**	90.7	0.39	**D7**	7.1	0.093				1204
SDKH15	A_B_C_D_E		D	**D1**	100	0.19							2795
SDKH16	A_B_D_E_F		D	**D2**	72.3	2.34	**D6**	27.7	1.20				2644
SDKH17	A_B_D_E_F		D	**D2**	100	1.86							2136
SDKH18	A_B_D		D	**D1**	100	0.69							1051
SDKH20	A_B_D_E		D	**D1**	100	0.39							3514
SDK22	A_B_E		E	**E**	100	0.97							2784
SDG1	A_B_D		D	**D1**	100	0.15							3364
Median (range)													2390 (1051–3514)
SDG2	A_B_D_E	Co-infected	D_E	**D1**	81.3	1.01	**E**	12.7	1.91	**D7**	6.0	1.52	1304
SDG3	C_D		D_E	**D7**	66.1	0.81	**E**	18.8	2.24	**D1**	15.1	2.34	1521
SDKH12	A_B_D_E		D_E	**E**	72.1	1.42	**D1**	27.9	0.87				1751
SDKH13	A_B_D_E		D_E	**E**	80.1	1.06	**D1**	19.9	1.73				1918
SDKH14	A_B_C_D_E		D_E	**D2**	98.6	0.40	**E**	1.4	1.73				2442
SDKH19	A_B_D_E		D_E	**D1**	62.9	1.71	**E**	35.4	1.58	**D2**	1.7	1.54	2397
Median (range)													1835 (1304-2442)
P													0.03

The HBV subtypes are indicated in bold

[a]SD, Sudan; G, El Gezira; F, Al Fasher; KH, Khartoum; K, Kassala

[b]Percentage of reads clustered in each haplotype

[c]Diversity was calculated using DNA distance (Phylip package) and is expressed as mean substitutions/site $\times 10^{-4}$

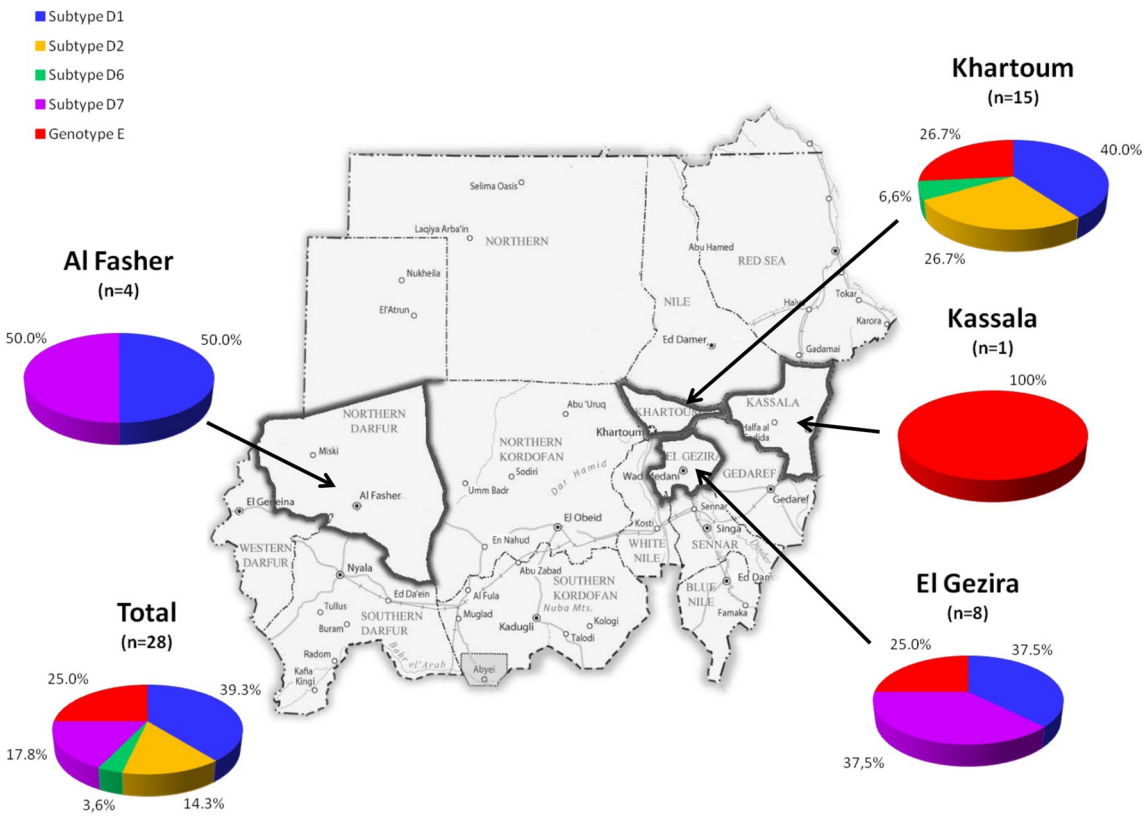


Fig. 1 Geographic distribution of HBV subtypes in four regions of Sudan. Total number of haplotypes, which were identified by UDPS analysis, per region are indicated in brackets and percentages of corresponding subtype or genotype are indicated near to the pie chart

Phylogenetic analysis

The phylogenetic analysis of P/S sequences is shown in Fig. 2a, b, each haplotype sequence obtained by UDPS clustered according to the expected genotype/subtype assignment and to geographic origin.

Regarding D1, haplotypes from every geographical region were found in different clusters suggesting that several strains are circulating in the same geographical area: SDKH20_HAPL1 and SDK15_HAPL1 clustered with the KF170771 Sudanese sequence, sharing 100% identity. SDG2_HAPL1 and SDKH12_HAPL2 shared 99.14% and 99.43% similarity with the KF170771 Sudanese sequence [27], whereas SDF5_HAPL1 had 100% identity with KF70758 Sudanese sequences. SDK13_HAPL2 and SG3_HAPL3 fell in the same cluster with the KF170766 Sudanese sequence, showing a similarity of 99.43% and 98.85%, respectively. SF10_HAPL1 fell in a clade with the KF170776 Sudanese strain (94.48% identity) [28], whereas SDK18_HAPL1 was very closely related to the GU456670 sequence from Iran (98.85% identity), while SDH19_HAPL1 was closely related to GU456664 from Iran (97.98% similarity).

Four haplotypes were classified as subtype D2, among which SDKH16_HAPL1 and SDKH17_HAPL1 fell in the same sub-cluster as the KF170761 Sudanese strain (98.27% and 97.97% similarity, respectively).

The only D6 haplotype detected (SKDH16_HAPL2) was most closely related to the KF170767 Sudanese sequence (97.09% identity) [28].

Five haplotypes (SDG4_HAPL1; SDG3_HAPL1, SDG2_HAPL3, SDF5_HAPL2, SDF10_HAPL2) formed a distinct cluster supported by a 82% bootstrap value with a mean identity of 96.9% with a D7 sequence from Niger (FN594771). D4 and D5 subtypes were not observed among the analyzed samples.

The 7 E haplotypes were found intermixed with strains from Sudan and West Africa in the phylogenetic tree. They were interspersed in three clades (Fig. 2b). In the major clade, four sequences clustered with sequences from Sudan, whereas in the second clade, one sequence clustered with sequences from Ghana, Guinea and African countries (not specified in GenBank). In the third clade two sequences were intermixed with sequences from Sudan, Niger, Cameroon, Ghana, Guinea or West Africa.

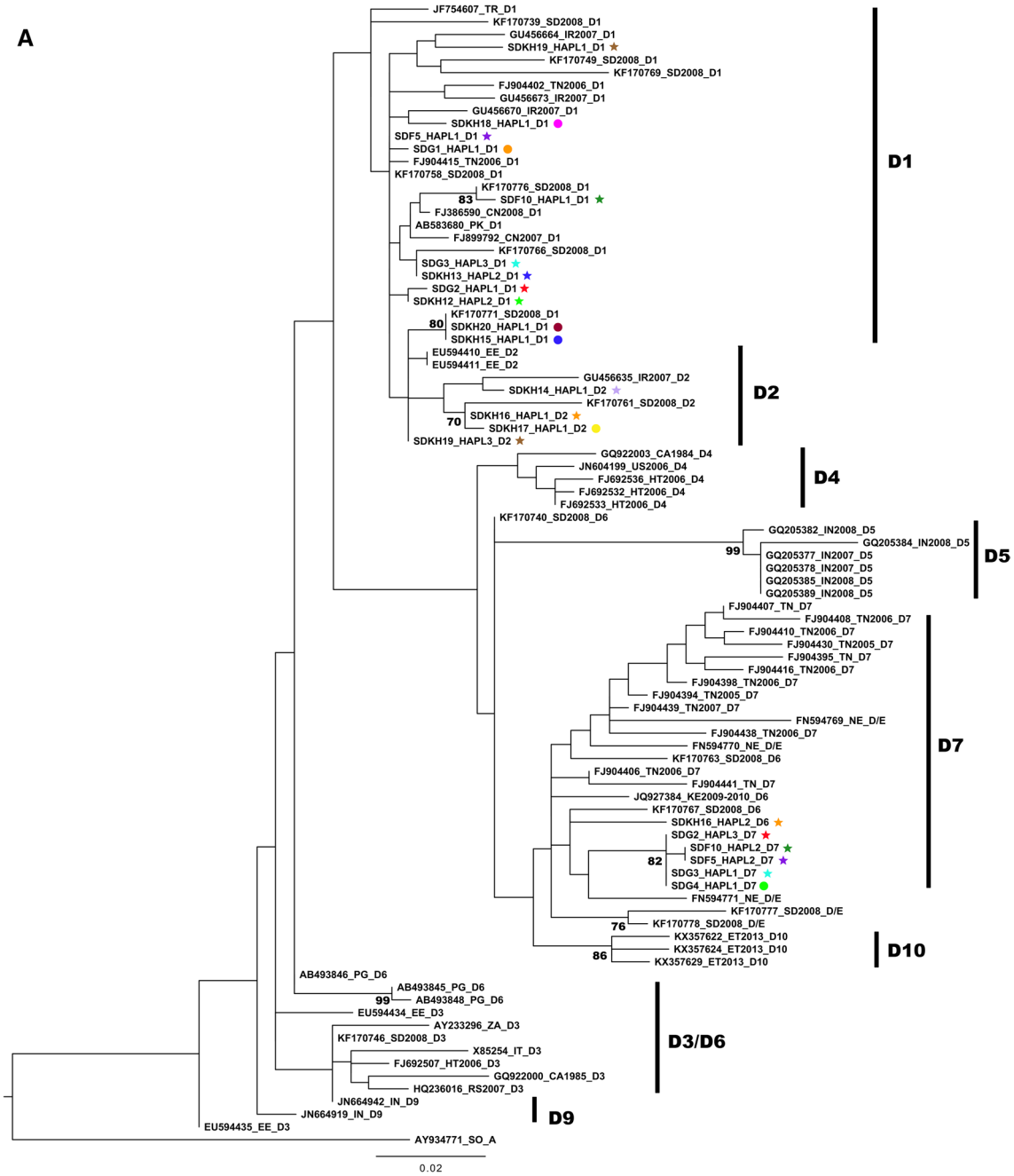


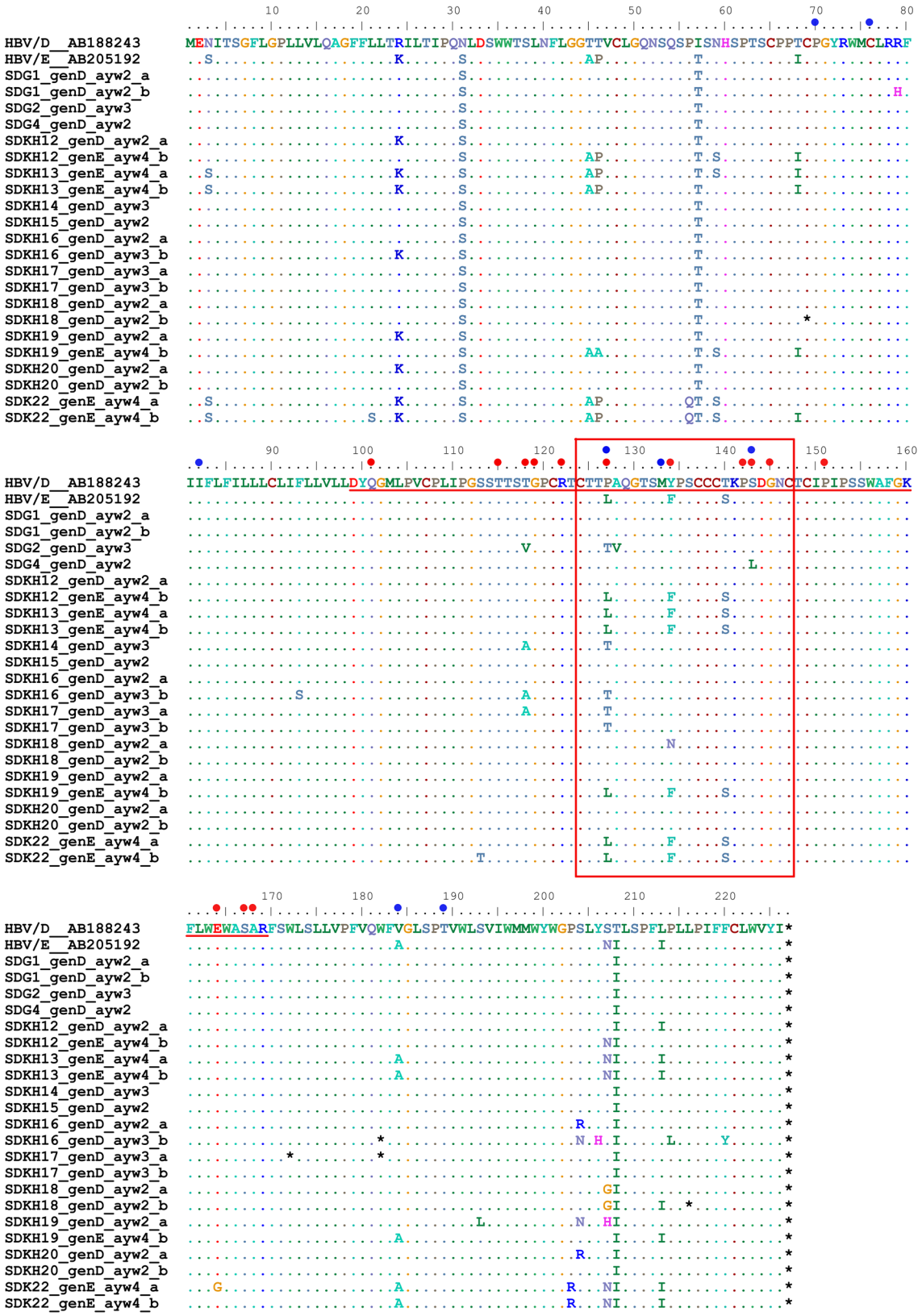
Fig. 2 Phylogenetic tree. Maximum-likelihood phylogenetic tree of the S fragment (amplified using the protocol described by Caballero et al. [20]). Nodes supported with a bootstrap value > 70% (500 replicates) are indicated. Sequences from mono-infected patients (i.e., with a single haplotype) are represented by a circle; sequences from co-infected patients (i.e., with two or more haplotypes) are represented by a star. For each group, a color identifies a single patient. The patients' ID define the Sudanese region (SD) from which they come (G, El Gezira; F, Al Fasher; KH, Khartoum; K, Kassala). Reference sequences are labeled by their accession numbers and their country of origin: AO, ANGOLA; AR, ARGENTINA; BE, BELGIUM; BJ, BENIN; BR, BRAZIL; CA, CANADA; CF, CENTRAL AFRICAN REPUBLIC; CG, CONGO (Republic of); CI, COTE D'IVOIRE; CM, CAMEROON; CN, CHINA; EE, ESTONIA; ; ET, ETHIOPIA; GA, GABON; GH, GHANA; GN, GUINEA; HT, HAITI; IE, IRELAND; IN, INDIA; IR, IRAN (Islamic Republic of); IT, ITALY; MG, MADAGASCAR; NA, NAMIBIA; NE, NIGER; PK, PAKISTAN; RS, SERBIA; SD, SUDAN; SN, SENEGAL; TN, TUNISIA; TR, TURKEY; US, UNITED STATES; WA, WESTERN AFRICA. The year of isolation (when available) and the genotype or subtype are also indicated. **a** Genotype D sequences. **b** Genotype E sequences

Fig. 2 (continued)

Among the mono-infected samples, SDK22 was the only specimen infected with genotype E, the remaining E strains were observed in mixed infected samples. The mean genetic diversity of genotype E was 1.29 ± 0.21%, which was significantly lower than genotype D strains (3.09 ± 0.15%, $p > 0.05$). When comparing the Sudanese to the African (excluding Sudan) E references diversities [1.77% (± SE: 0.31) and 1.68 (± SE: 0.35) respectively], our data indicate that there is no Sudanese E-specific strain. Identical sequences were found in patients from distant regions of Sudan (SDG3_HAPL2 and SDKH12_HAPL1), which were also identical to the Sudanese sequence KF170786; and SDKH19_HAPL2 was shown to be closely related to the KF170742 Sudanese strain. SDK13_HAPL1 was closely related to two sequences described in Guinea (98.85% similarity) and SDK14_HAPL2 showed high similarity (98.56%) to sequences from Guinea (GQ161789) and from Ghana (GQ161776).

HBV mutation in HBV S region

Thirteen samples were successfully amplified and sequenced in the S region. The amino acid sequence alignment of these samples is shown in Fig. 3. The major hydrophilic region (MHR) (aa 99-169) is underlined in the reference sequence and the "a determinant" region (aa 124-147) is boxed. If a double signal was observed in the electropherogram, the


HBV/D__AB188243
HBV/E__AB205192
SDG1_genD_ayw2_a
SDG1_genD_ayw2_b
SDG2_genD_ayw3
SDG4_genD_ayw2
SDKH12_genD_ayw2_a
SDKH12_genE_ayw4_b
SDKH13_genE_ayw4_a
SDKH13_genE_ayw4_b
SDKH14_genD_ayw3
SDKH15_genD_ayw2
SDKH16_genD_ayw2_a
SDKH16_genD_ayw3_b
SDKH17_genD_ayw3_a
SDKH17_genD_ayw3_b
SDKH18_genD_ayw2_a
SDKH18_genD_ayw2_b
SDKH19_genD_ayw2_a
SDKH19_genE_ayw4_b
SDKH20_genD_ayw2_a
SDKH20_genD_ayw2_b
SDK22_genE_ayw4_a
SDK22_genE_ayw4_b

◂**Fig. 3** Distribution of amino acid mutations detected in the S region among HBV isolates. The major hydrophilic region (MHR) (aa 99-169) is underlined in the reference sequence (AB188243, sequence for HBV genotype D; AB2051192 sequence for HBV genotype E). The "a determinant" region (aa 124-147) is boxed. In the name of the sequences is indicated: the patient ID and the genotype (by geno2pheno to the Sanger sequence of the long S region). If a double signal is present in the Sanger sequence, the polymorphic positions are reported separately in two different sequences indicated as "a" and "b". Amino acid substitutions Q101H, T115N, T118A, G119T, T123V, P127L/S, Y134H, P142L, S143L, G145R, P151L, E164G, S167L and A168T, which previously were reported to be associated with immune escape in HBV genotype D, are highlighted with red circle. Amino acid substitutions P70T, C76F, I82T, L127P, M133T, S143T, A184V, and T189I, which previously were reported to be associated with immune escape in HBV genotype E, are highlighted with blue circle

polymorphic positions are reported separately in two different sequences indicated as "a" and "b".

In two sequences (SDKH12 and SDKH19) from five patients where mixed infections were assessed by UDPS, both genotype D and E were recognized in the Sanger sequence of the S region.

All sequences of genotype D strains had the signature amino-acids 31S, 57T, and 208I.

In some patients, the sequences showed mutations in the MHR or in the "a determinant", or outside of these regions with respect to the reference sequences. Some of these substitutions have been associated with immune-escape phenomenon [30], including T118A in SDKH14, SDKH16_b and SDKH17_a, and S143L in SDG4. SDK22_a (genotype E mono-infected sample) showed the mutation E164G, which has already been associated with immune escape in genotype D samples.

In addition, other substitutions were observed in positions already described to be associated with immune escape (T118V, P/L127T and Y134N) [30].

Additional substitutions were also observed outside the MHR, including: L21S, R24K, T/P46A, P56Q, N59S, T68I, R79H, F93S, S193L, P203R, S204R/NK, Y206H, S207G/H, L213I, and F220Y.

Discussion

Different methods used in HBV genotyping show different sensitivities when compared to each other. In particular, discrepancies have been noted in RFLP analysis and DS of DNA, followed by phylogenetic analysis of these sequences [31–34]. Sanger sequencing is considered the gold standard in HBV genotyping, and is able to effectively detect mixed infections with an average rate of 20% of the samples sequenced. However, Bekondi et al. [35] demonstrated that when serum samples that presented mixed nucleotides in the electropherogram were screened for mixed infections by cloning the PreS, S region and Core PCR fragment, the resulting clones showed no evidence of mixed infection except for 1 patient. Multiplex PCR needs many primer sets and its efficacy is reduced by the variability of target sequences [31, 36]. MALDI-TOFF, capable of detecting wild-type and mutant alleles, may identify minority types if present at > 10% frequency [37]. The presence of mixed infections has frequently been confirmed by cloning and sequencing of PCR products [11], but this procedure is very time-consuming and the number of genotypes could be underestimated, depending on the analyzed clone number.

UDPS is able to detect genotypes present as a minority population (1%), so its sensitivity is higher than DS and MALDI-TOFF, as well as cloning/sequencing. In our study, we used UDPS to confirm mixed infections. The presence of co-infecting genotypes was confirmed in 6/16 patients (37.5%), whereas in the remaining samples 1 genotype, or 2 subtypes belonging to same genotype were observed. Co-infected samples were also observed to harbor different haplotypes. In 3 cases, 3 haplotypes were represented with a frequency of < 10%: Notably, genotypes A, B and C were not detected by UDPS, probably due to false positive signals in preliminary genotyping methods.

The phylogenetic analysis showed that the strains observed both in mono-infections and mixed infections did not fall in separate clusters, suggesting that no specific variants were involved in mixed infections, also confirmed by the absence of viral escape mutants in these cases (Fig. 3).

Regarding the geographic distribution, the genetic diversity observed among genotype E (1.3%) was similar to that observed in other countries such as the Central African Republic (1.8%) [38] and Rwanda (1.7%) [39]. Overall, D1, D2, D6 and D7 subtypes were closely related to strains circulating in Sudan or in Sub-Saharan area [40–42] as well as in Iran (Fig. 2a). It is possible that the introduction of these strains and their continued circulation in these countries could be related to the commercial relationship between Sudan and countries in the Middle East. The D1 subtype was the most frequently detected subtype showing high identity (94.68–100%) with sequences from Sudan (KF170758, KF170766, KF170776). Two haplotypes clustered with D1 subtypes from Iran (GU456664 and GU456670). In our samples, subtype D7 is the second most prevalent. It represents the main genotype in Morocco and Tunisia, nevertheless our strains clustered with isolates from Niger, and separately from D7 sequences from Tunisia, suggesting higher migratory flows between Niger and Sudan in comparison with Mediterranean Countries. Subtype D2 strains clustered with a Sudanese strain (KF170761) and an Iranian strain (GU456635). Only one D6 subtype was typed by UDPS and it clustered with a D6 strain from Sudan (KF170767).

It is important to point out that our phylogenetic analysis was carried out in a partial P/S region, and not with the

complete viral genome. However, we reported all subtypes declared by the authors and the subtypes indicated in GenBank (see Table S1 and Fig. 2) and all S regions that were analyzed by Sanger method confirmed the UDPS results.

This study has limitations, however: the UDPS was carried out in only four Sudanese States, while a nationwide survey could help delineate the actual frequency of HBV genotype mixed infection in Sudan. In conclusion, UDPS is more effective than multiplex-nested PCR analysis for identifying HBV genotypes and subtypes in mixed infections. This approach may be relevant for monitoring viral dynamics during antiviral therapy, when multiple strains coexist, possibly showing different sensitivities to anti-HBV drugs.

Acknowledgements This work was supported by Italian Ministry of Health (funds Ricerca Corrente, linea 3 hepatitis to National Institute for Infectious Diseases INMI IRCCS L. Spallanzani); the European Union's Horizon 2020 research and innovation program European Virus Archive goes Global under Grant agreement no. 653316.

Compliance with ethical standards

Conflict of interest All authors declare that they have no conflict of interest.

References

1. Kramvis A. Genotypes and genetic variability of hepatitis B virus. Intervirology. 2014;57:141–50. https://doi.org/10.1159/000360947.
2. Norder H, Couroucé A-M, Coursaget P, Echevarria JM, Lee S-D, Mushahwar IK, et al. Genetic diversity of hepatitis B virus strains derived worldwide: genotypes, subgenotypes, and HBsAg subtypes. Intervirology. 2004;47:289–309. https://doi.org/10.1159/000080872.
3. Santos AO, Alvarado-Mora MV, Botelho L, Vieira DS, Pinho JRR, Carrilho FJ, et al. Characterization of hepatitis B virus (HBV) genotypes in patients from Rondônia, Brazil. Virol J. 2010;7:315. https://doi.org/10.1186/1743-422X-7-315.
4. Stuyver L, De Gendt S, Van Geyt C, Zoulim F, Fried M, Schinazi RF, et al. A new genotype of hepatitis B virus: complete genome and phylogenetic relatedness. J Gen Virol. 2000;81:67–74. https://doi.org/10.1099/0022-1317-81-1-67.
5. Arauz-Ruiz P, Norder H, Robertson BH, Magnius LO. Genotype H: a new Amerindian genotype of hepatitis B virus revealed in Central America. J Gen Virol. 2002;83:2059–73. https://doi.org/10.1099/0022-1317-83-8-2059.
6. Hess G, Arnold W, Koesters W, Biswas R, Hütteroth TH, zum Büschenfelde KH. Simultaneous presence of HBsAg and anti-HBs in the serum of different subtypes (serological and immunofluorescent studies). Z Immunitatsforsch Immunobiol. 1977;153:143–51.
7. Tabor E, Gerety RJ, Smallwood LA, Barker LF. Coincident hepatitis B surface antigen and antibodies of different subtypes in human serum. J Immunol Baltim Md. 1950;1977:369–70.
8. Ding X, Gu H, Zhong Z-H, Zilong X, Tran HT-T, Iwaki Y, et al. Molecular epidemiology of hepatitis viruses and genotypic distribution of hepatitis B and C viruses in Harbin, China. Jpn J Infect Dis. 2003;56:19–22.
9. Kato H, Orito E, Sugauchi F, Ueda R, Koshizaka T, Yanaka S, et al. Frequent coinfection with hepatitis B virus strains of distinct genotypes detected by hybridization with type-specific probes immobilized on a solid-phase support. J Virol Methods. 2003;110:29–35.
10. Chen B-F, Kao J-H, Liu C-J, Chen D-S, Chen P-J. Genotypic dominance and novel recombinations in HBV genotype B and C co-infected intravenous drug users. J Med Virol. 2004;73:13–22. https://doi.org/10.1002/jmv.20051.
11. Candotti D, Diarra B, Bisseye C, Tao I, Pham Quang K, Sanou M, et al. Molecular characterization of hepatitis B virus in blood donors from Burkina Faso: prevalence of quasi-subgenotype A3, genotype E, and mixed infections. J Med Virol. 2016;88:2145–56. https://doi.org/10.1002/jmv.24589.
12. Toan NL, Song LH, Kremsner PG, Duy DN, Binh VQ, Duechting A, et al. Co-infection of human parvovirus B19 in Vietnamese patients with hepatitis B virus infection. J Hepatol. 2006;45:361–9. https://doi.org/10.1016/j.jhep.2006.03.013.
13. Minosse C, Giombini E, Bartolini B, Capobianchi MR, Garbuglia AR. Ultra-deep sequencing characterization of HCV samples with equivocal typing results determined with a commercial assay. Int J Mol Sci. 2016;17:1679. https://doi.org/10.3390/ijms17101679.
14. Chien R-N, Yeh C-T, Tsai S-L, Chu C-M, Liaw Y-F. Determinants for sustained HBeAg response to lamivudine therapy. Hepatol Baltim Md. 2003;38:1267–73. https://doi.org/10.1053/jhep.2003.50458.
15. Kobayashi M, Suzuki F, Akuta N, Suzuki Y, Arase Y, Ikeda K, et al. Response to long-term lamivudine treatment in patients infected with hepatitis B virus genotypes A, B, and C. J Med Virol. 2006;78:1276–83. https://doi.org/10.1002/jmv.20701.
16. Osiowy C, Villeneuve J-P, Heathcote EJ, Giles E, Borlang J. Detection of rtN236T and rtA181 V/T mutations associated with resistance to adefovir dipivoxil in samples from patients with chronic hepatitis B virus infection by the INNO-LiPA HBV DR line probe assay (version 2). J Clin Microbiol. 2006;44:1994–7. https://doi.org/10.1128/JCM.02477-05.
17. Bonino F, Marcellin P, Lau GKK, Hadziyannis S, Jin R, Piratvisuth T, et al. Predicting response to peginterferon alpha-2a, lamivudine and the two combined for HBeAg-negative chronic hepatitis B. Gut. 2007;56:699–705. https://doi.org/10.1136/gut.2005.089722.
18. Chauhan R, Singh AK, Rooge S, Varshney A, Kumar M, Sarin SK. Analysis of hepatitis B virus genotype changes in patients with chronic hepatitis B infection on tenofovir therapy. J Med Virol. 2016;88:1364–75. https://doi.org/10.1002/jmv.24489.
19. Naito H, Hayashi S, Abe K. Rapid and specific genotyping system for hepatitis B virus corresponding to six major genotypes by PCR using type-specific primers. J Clin Microbiol. 2001;39:362–4. https://doi.org/10.1128/JCM.39.1.362-364.2001.
20. Caballero A, Gregori J, Homs M, Tabernero D, Gonzalez C, Quer J, et al. Complex genotype mixtures analyzed by deep sequencing in two different regions of hepatitis B virus. PLoS One. 2015;10:e0144816. https://doi.org/10.1371/journal.pone.0144816.
21. Vincenti D, Piselli P, Solmone M, D'Offizi G, Capobianchi MR, Menzo S. Evolutionary trends of resistance mutational patterns of HBV reverse transcriptase over years (2002–2012) of different treatment regimens: the legacy of lamivudine/adefovir combination treatment. Antiviral Res. 2017;143:62–8. https://doi.org/10.1016/j.antiviral.2017.03.008.
22. Zhang Q, Wu G, Richards E, Jia S, Zeng C. Universal primers for HBV genome DNA amplification across subtypes: a case study for designing more effective viral primers. Virol J. 2007;4:92. https://doi.org/10.1186/1743-422X-4-92.
23. Li W, Jaroszewski L, Godzik A. Clustering of highly homologous sequences to reduce the size of large protein databases. Bioinforma Oxf Engl. 2001;17:282–3.
24. Ramírez C, Gregori J, Buti M, Tabernero D, Camós S, Casillas R, et al. A comparative study of ultra-deep pyrosequencing and

cloning to quantitatively analyze the viral quasispecies using hepatitis B virus infection as a model. Antiviral Res. 2013;98:273–83. https://doi.org/10.1016/j.antiviral.2013.03.007.
25. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32:1792–7. https://doi.org/10.1093/nar/gkh340.
26. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0. Mol Biol Evol. 2013;30:2725–9. https://doi.org/10.1093/molbev/mst197.
27. Yousif M, Mudawi H, Bakhiet S, Glebe D, Kramvis A. Molecular characterization of hepatitis B virus in liver disease patients and asymptomatic carriers of the virus in Sudan. BMC Infect Dis. 2013;13:328. https://doi.org/10.1186/1471-2334-13-328.
28. Yousif M, Mudawi H, Hussein W, Mukhtar M, Nemeri O, Glebe D, et al. Genotyping and virological characteristics of hepatitis B virus in HIV-infected individuals in Sudan. Int J Infect Dis IJID Off Publ Int Soc Infect Dis. 2014;29:125–32. https://doi.org/10.1016/j.ijid.2014.07.002.
29. Mahgoub S, Candotti D, El Ekiaby M, Allain J-P. Hepatitis B virus (HBV) infection and recombination between HBV genotypes D and E in asymptomatic blood donors from Khartoum, Sudan. J Clin Microbiol. 2011;49:298–306. https://doi.org/10.1128/JCM.00867-10.
30. Coppola N, Onorato L, Minichini C, Di Caprio G, Starace M, Sagnelli C, et al. Clinical significance of hepatitis B surface antigen mutants. World J Hepatol. 2015;7:2729–39. https://doi.org/10.4254/wjh.v7.i27.2729.
31. Lim CK, Tan JTM, Ravichandran A, Chan YC, Ton SH. Comparison of PCR-based genotyping methods for hepatitis B virus. Malays J Pathol. 2007;29:79–90.
32. Sertoz RY, Erensoy S, Pas S, Akarca US, Ozgenc F, Yamazhan T, et al. Comparison of sequence analysis and INNO-LiPA HBV DR line probe assay in patients with chronic hepatitis B. J Chemother Florence Italy. 2005;17:514–20. https://doi.org/10.1179/joc.2005.17.5.514.
33. Takahashi K, Akahane Y, Hino K, Ohta Y, Mishiro S. Hepatitis B virus genomic sequence in the circulation of hepatocellular carcinoma patients: comparative analysis of 40 full-length isolates. Arch Virol. 1998;143:2313–26.
34. Venegas M, Muñoz G, Hurtado C, Alvarez L, Velasco M, Villanueva RA, et al. Prevalence of hepatitis B virus genotypes in chronic carriers in Santiago, Chile. Arch Virol. 2008;153:2129–32. https://doi.org/10.1007/s00705-008-0231-6.
35. Bekondi C, Olinger CM, Boua N, Talarmin A, Muller CP, Le Faou A, et al. Central African Republic is part of the West-African hepatitis B virus genotype E crescent. J Clin Virol Off Publ Pan Am Soc Clin Virol. 2007;40:31–7. https://doi.org/10.1016/j.jcv.2007.05.009.
36. Repp R, Rhiel S, Heermann KH, Schaefer S, Keller C, Ndumbe P, et al. Genotyping by multiplex polymerase chain reaction for detection of endemic hepatitis B virus transmission. J Clin Microbiol. 1993;31:1095–102.
37. Ehrich M, Böcker S, van den Boom D. Multiplexed discovery of sequence polymorphisms using base-specific cleavage and MALDI-TOF MS. Nucleic Acids Res. 2005;33:e38. https://doi.org/10.1093/nar/gni038.
38. Bekondi C, Olinger CM, Boua N, Talarmin A, Venard V, Muller CP, et al. Characterization of hepatitis B virus strains from the Central African Republic: preliminary results. Pathol Biol (Paris). 2008;56:310–3. https://doi.org/10.1016/j.patbio.2007.12.007.
39. Hübschen JM, Andernach IE, Muller CP. Hepatitis B virus genotype E variability in Africa. J Clin Virol Off Publ Pan Am Soc Clin Virol. 2008;43:376–80. https://doi.org/10.1016/j.jcv.2008.08.018.
40. Yousif M, Kramvis A. Genotype D of hepatitis B virus and its subgenotypes: an update. Hepatol Res Off J Jpn Soc Hepatol. 2013;43:355–64. https://doi.org/10.1111/j.1872-034X.2012.01090.x.
41. Ochwoto M, Chauhan R, Gopalakrishnan D, Chen C-Y, Nganga Z, Okoth F, et al. Genotyping and molecular characterization of hepatitis B virus in liver disease patients in Kenya. Infect Genet Evol J Mol Epidemiol Evol Genet Infect Dis. 2013;20:103–10. https://doi.org/10.1016/j.meegid.2013.08.013.
42. Perrillo R. Benefits and risks of interferon therapy for hepatitis B. Hepatol Baltim Md. 2009;49:S103–11. https://doi.org/10.1002/hep.22956.