

BIOINFORMATICS ANALYSES OF POLYMERASE (*pol*) GENE OF HBV GENOME OCCURRED IN BLOOD SAMPLES FROM DIFFERENT REGIONS OF KSA

Al Harthi J.H.^a, Al-Yami M.R.^b and Sadik A.S.^{a,c*}

^aDepartment of Biology, Faculty of Science, Taif University, Taif, Al-Haweiah, P.O. Box 888, Zip code 21974, Taif, KSA. ^bChild Hospital in Taif, Shehar, El-Shafa Road, Zip code 26514, Ministry of Health, KSA.

^cDepartment of Agricultural Microbiology, Faculty of Agriculture, Ain Shams University, P.O. Box 68, Hadayek Shobra 11241, Cairo, Egypt. E. mail: *atef_sadik@yahoo.com

Article received 24.7.2018, Revised 15.9.2018, Accepted 23.9.2018

ABSTRACT

This study aimed to isolate of polymerase (*pol*) gene belonging to some hepatitis B virus (HBV) from some blood samples collected from different regions of KSA representing south, north, east, middle and west KSA. This was followed by determining the nucleotide sequences of this gene and carrying out some bioinformatics analyses on such genes. The presence of virus in some HBV-infected blood samples collected from different regions of KSA was serologically and molecularly confirmed by ELISA and Rt-PCR techniques. The nucleotide sequence of *pol* gene was determined and bioinformatically analyzed. Results showed that the nucleotide sequences of the partial sequences of the five HBV strains (LC152751.1, LC152752.1, LC152753.1, LC152757.1 and LC152759.1) were ranged from 600 to 899 nts and proved to be belonging to *pol* gene of HBV. Bioinformatics comparisons between the five HBV strains showed percent identities 92-98% and 80-95% at the levels of DNA and deduced amino acids, respectively, when compared to some overseas HBV strains recorded in GenBank. Types of domains (DNA_pol_viral_C and DNA pol viral-N) its accession (pfam00336 and pfam00242), super families (RT-like superfamily and cl02825) of the sequenced *pol* gene(s) were also addressed. Differences at the levels of DNA, encoding amino acids, *i.e.*, open reading frames and domains super families between the strains under investigation and those similar in GenBank were determined.

Keywords: Hepatitis, HBV, Polymerase (*pol*) gene, Bioinformatics.

INTRODUCTION

The relevance of infection caused by hepatitis viruses was related mainly to their wide geographic distribution and the large number of infected individuals in all parts of the world (Almeida et al., 2012; Lin and Kao, 2016 and Samal et al., 2017).

Hepatitis B virus (HBV) and Hepatitis C virus (HCV) are among the principal causes of severe liver disease (Akcam et al., 2009; Prasad et al., 2015 and Dai et al., 2015). Tadokoro et al. (2006) reported that HBV is a worldwide public health problem. HBV infection is transmitted sexually, parentally and prenatally from mother to child (Sharavanan et al., 2014). Therefore, Prasad et al. (2015) showed that health care workers are at high risk for transmission of hepatitis B.

There is limited data of epidemiology of Hepatitis B in community, more so in rural population (Prasad et al., 2015). Infection from Hepatitis B primarily results from peripartum vertical transmission and the risk increases in the presence of hepatitis Be antigen (Banks et al., 2016). Diagnostic assays allowing the quantification of HBV-DNA over a wide range of concentrations are important for monitoring patients during antiviral therapy (Zanella et al., 2002).

The epidemiological significance of HBV genotypes has been well established and becoming an essential concern day by day however, much little is known about the mixed infection with more than one HBV genotype and their clinical relevance (Alam et al., 2007). Using the emission transmission electron microscopy viral particles was measured at 42 nm in diameter (Howard 1986). It remains unclear whether HBV replicates in extra-hepatic tissues, and particularly in peripheral blood mononuclear cells, which may serve as a reservoir for the maintenance of infection (Mazet-Wagner et al., 2006). A serial invasive signal amplification react-ion assay (PCR-Invader assay) was developed for distinguishing the known eight genotypes (A to H) and four subgenotypes (Aa, Ae, Ba, Bj) of HBV. HBV, the hepadnavirus infecting humans, can be divided into the 7 genotypes A to G. By definition, genotypes differ by more than 8% at the nucleotide level. However, some genotypes differ by more than 14% from others (Schaefer et al., 2003).

The most prevalent genotype in the Mediterranean region is genotype D. Genotype E is observed in parts of East, Central, and West

Africa. Genotype F is found in South and Central America, genotype G in the USA and France, and genotype H is observed in Central America (Echevarría et al., 2005). This study was designed to analyze of the nucleotide sequence of polymerase (*pol*) gene of HBV amplified from the DNA-blood samples of different regions of Saudi Arabia, and this could be led to detecting the mutations at the level of DNA or its protein.

MATERIALS AND METHODS

Source of serum blood samples: Some HBV-infected serum blood samples were obtained from Central Laboratory [AJMS-01-2016 (Female-52y) and AJMS-02-2016 (Male-35y)] and King Abdul Aziz Hospital [AJMS-03-2016 (Male-41y)] at Taif; Asir Central Hospital [AJMS-04-2016 (Male-27y)] and General Abha Hospital [AJMS-05-2016 (Male-23y)] at Abha; General King Fahd Hospital [AJMS-06-2016 (Female-56y) and AJMS-07-2016 (Female -48y)] in Jeddah; Hospital in Hafr Al Batin [AJMS-08-2016 (Male-29y)] and Riyadh El Shamasy Hospital [AJMS-09-2016 (Female-46y) and AJMS-10-2016 (Male-34y)].

HBV serological testing: For detection of the surf-ace antigen of the hepatitis B in human serum or plasma by the enzyme immunoassay technique the Monolisa™ HBsAg ULTRA kit [BIO ELISA -BIO-RAD, la-Coquette, France] was used to assess the presence of HBsAg in collected blood samples.

PCR confirmation of HBV: To confirm the presence of HBV-DNA by PCR in the blood samples Cob-as *Taq*Screen MPX test was used among the COB-AS® *Taq*Man® Analyzer. Real time detection of PCR products was conducted using the following primer pair (Sense: 5' AGA CTC GTG GTG GAC TTC TCT 3' (5'252 position) and Antisense: 5' CAA AAG AAA ATT GGT AAC AGC GGT A 3' (5'794 position).

PCR program and amplification reaction: In a total volume of 25 μ L total which consists of Go*Taq* Green Master Mix, 2X (12.5 μ L); Sense Primer, 20 μ M (1.25 μ L); Antisense Primer, 20 μ M (1.25- μ L); HBV-DNA Template 2.50 μ L; Nuclease-Free water 7.50 μ L the PCR was conducted. The PCR program was started with one cycle of 95°C for five minutes followed by 35 cycles each of 95°C, 50°C and 72°C for minute for each. The final cycle (72°C) was extended for 10 minutes.

Sequencing of PCR products: Using three primers: Sense: 5' GGA TGT GTC TGC GGC GTT T3'; Sense: 5' AGA CTC GTG GTG GAC TTC TCT3' and Antisense: 5' CAA AAG AAA ATT GGT AAC AGC GGTA3' the PCR products

of round 2 were gel purified and sequenced using automated DNA sequencing system (ABI 3100) and BigDye® Terminator v3.1 cycle sequencing Kit (Applied Bio-systems, Foster City, CA, USA).

Bioinformatics analysis: By using BLASTN 2.2.23 + software (<http://www.ncbi.nlm.nih.gov/blast/>) DNA sequence was analyzed against compared to related HBV strains recorded in Genbank. The deduced amino acids *i.e.*, open reading frame (ORF), types of domains and their super families were also analyzed using the same program. The sequence that showed the lowest E-value and maximum identity was taken as the genotype of the sample analyzed. On sequencing, the general primers of major surface antigen region (F1, F2, S2, S3 P3 and TPR1), and primers of each of PCR (PCR products 1, 2 and 3), sequencing (Sense 5'252, Sense 5'377 and Antisense 5'794) and mutagenic (M741VAY WF and L669 MAYWF) were detected in the obtained partial sequences of HBV strains.

RESULTS AND DISCUSSION

The HBV is a major health problem worldwide with more than 350 million people being chronic carriers (Aljarbou, 2013). HBV is one of the main etiological agents of acute and chronic liver disease that is still a major public health problem in the world (Amini-Bavil-Olyaei et al., 2005).

In this study, a number of ten blood samples infected with HBV were collected from Taif, Jeddah, Riaydh, Hafr Al Batin and Abha regions of KSA for bioinformatics studies on HBV genome. HBV is an enveloped DNA virus that belongs to the Hepadnaviridae family (NCBI taxonomy, ICTV, Viral Zone). It contains a small, partially ds, relaxed-circular (RC) DNA genome that replicates by reverse transcription to an RNA intermediate, the pgR-NA (Datta et al., 2006). Its length is comprised between 3182 and 3248 bp depending on genotypes (Jones and Jianming 2013).

The experimental results showed that the nucleotide sequences of the partial sequences of the five HBV strains belonging to the polymerase (*pol*) gene of HBV were recorded as 600 (AJMS-02-2016, LC152751.1), 898 (AJMS-03-2016, LC-1527-52.1), 855 (AJMS-04-2016, LC1527-53.1), 838 (AJ-MS-08-2016, LC152757.1) and 899 (AJMS-10-2016, LC152759.1) nts (Table 1). By comparison these nucleotides of the five strains with overseas HBV strains in Genbank, percent identities 98-99; 94-95; 92; 96 and 95-96 % were recorded, respectively.

Table 1: Characters of nucleotide sequences, ORFs and putative domains of the *pol* gene of HBV strains: LC152751.1, LC152752.1, LC152753.1, LC152757.1 and LC152759.1.

Levels	Characters	HBV Strains belonging to <i>pol</i> gene				
		LC152751.1	LC152752.1	LC152753.1	LC152757.1	LC152759.1
DNA	Length (nts)	600	898	855	838	899
	Identities (%)	99	94-95	92	94	97-98
Protein	ORF#-Frame	1-1	1-2	1-1	1-1	1-2
	ORF Length (nts)	600	798	855	837	897
	Eaa	200	266	284	279	299
	Stop codon	-	-	+	-	-
Domains	Identities (%)	87-95	90-91	77-96	84-90	94
	Name	DNA_pol_viral_C		DNA_pol_viral_N		
	Accession	pfam00336		pfam00242		
	Description	DNA_pol_viral_C terminal		DNA_pol_viral_N		
Superfamily		RT-like superfamily				cl02825

Eaa: Encoding amino acids.

+ : Present.

- : Absent.

It is well known that HBV-polymerase region overlaps pre-S/S genes with high epitope density and plays an essential role in viral replication (Huang et al., 2013). The nucleotide sequences of open reading frames (ORFs) of the five HBV strains (LC152751.1, LC152752.1, LC152753.1, LC1527-57.1 and LC152759.1) were 600, 798, 855, 837 and 897 nts, respectively, and belonging to the *pol* gene of HBV (Table 2). These ORFs were deduced into 200, 266 (with stop codon), 285, 279, 299 amino acids, respectively. Identities of 90-91; 77-96; 84-90; 86 and 89-91%, respectively, were recorded between the ORFs of HBV strains compared to those similar HBV overseas strains recorded in GenBank. The intragenotypic divergence of the complete genome sequence of Iranian strains was 1.8% and the intergenotypic in genotype D was 3.8% and with the other genotypes was 7.9-15.4%. The largest ORF in the HBV genome encodes for the hepatitis B polymerase protein (HBp). The protein is 90 kDa in size and has RNA and DNA dependent polymerase activity (Toh et al., 1983). HBp plays a key role in HBV genome generation as well as pgRNA encapsidation. HBp is packaged together

with pgRNA with in HBV nucleocapsids) (Mack et al., 1988).

Differences of 9 & 36; 57 & 27; 70 & 75; 51 & 44 and 30 & 24 were recorded between the partial sequences of HBV strains: (LC152751.1, LC1527-52.1, LC152753.1, LC152757.1 and LC-152759.1) *pol* gene & its ORFs, i.e., deduced amino acids, respectively and the compared overseas strains of HBV in GenBank. Phylogenetic tree of the HBV *P* gene, partial sequence, strains of this study compared the most related genes confirmed that the gene was belonging to the *pol* gene (Fig. 1). HBp has been divided into four characterized domains. Based on sequence homologies and studies on the mechanism of viral genome replication, most parts of HBp are indispensable (Lanford et al., 1999). The primase domain acts in priming (-) DNA strand synthesis and ends up covalently linked to the 5' end of the (-) DNA strand. The subsequent domain does not appear to have any enzymatic function but acts as a spacer between the first and third domains. The third domain gives HBp its name. It occupies approximately 40% of the protein and encodes for the RNA and DNA dependent polymerase activity.

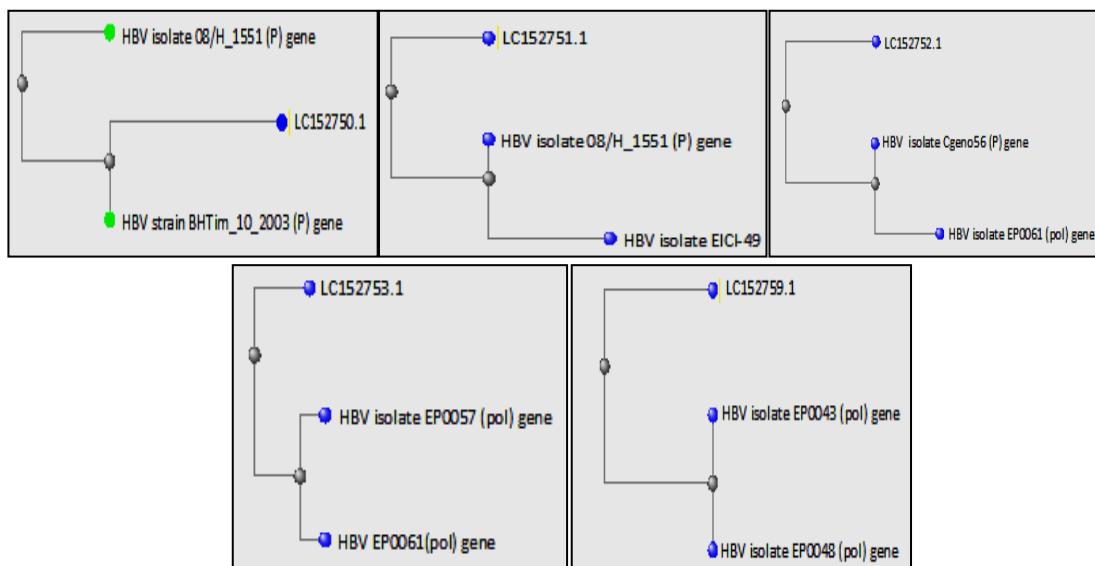


Figure 1. Phylogenetic tree of surface protein region genes of HBV strains and most related overseas strains in GenBank.

In the PCR-amplified sequences of HBV strains LC152751.1, LC152752.1, LC152753.1, LC152757.1 and LC152759.1, the general primers (Table 2) of surface antigen region genes were detected. Data show that the highest number of primers were recorded in the nucleotide sequences of LC152751.1 (Five primers) followed by LC152752.1 (Four primers). Three primers were

detected in LC152757.1 and LC152759.1, while two primers were found in LC152753.1. The S2 primer was the most repeated one (four strains) followed by F1 & P3 primers (three strains), S3 & TPR1 primers (one strain). Both of M741VA-YWF and L669MAYWF primers were only recorded in LC152751.1 (Fig. 2).

Table 2: Presences of general primers of the polymerase (*pol*) gene, and primers of each of PCR, sequencing and mutagenic recorded in the PCR-amplified sequences of HBV strains.

Primers names	LC152751.1	LC152752.1	LC152753.1	LC152757.1	LC152759.1
F1	-	+	-	+	+
F2	-	-	-	-	-
S2	-	+	+	+	+
S3	+	-	-	-	-
P3	-	+	-	+	+
TPR1	-	-	+	-	-
F58	+	+	-	-	-
R58	+	-	-	-	-
F252	-	-	-	-	-
M741VAYWF	+	-	-	-	-
L669MAYWF	+	-	-	-	-
Total	5	4	2	3	3

+: Present. -: Absent.

Only four partial sequences of HBV strains (LC152752.1, LC152753.1, LC152757.1 and LC152759.1) showed putative conserved domains belonging to DNA_pol_viral_N (pfam002-42) domain which described as DNA polymerase (viral) N-terminal domain of the superfamily cl02825 with Query E-value of 3.12e-89, 1.02e-96, 1.65e-89 and 2.89e-127, respectively. On the other hand, the putative conserved domains of the ORF number 1 of partial sequence of HBV strain LC152751.1 *pol* gene was DNA_pol_viral_C with

accession number of pfam00336 described as DNA polymerase (viral) C-terminal domain, with Query interval and Query E-value of 175-200 and 1.92e-08, respectively, and belonging to the RT-like superfamily (Fig. 3).

When the amino acids between the putative conserved domains (Queries) of the four partial sequences of HBV strains: LC152752.1, LC152753.1, LC152757.1 and LC152759.1 which belong to the DNA_pol_viral_N (pfam00242) domain and the superfamily cl02825 were compared

with those in the same family differences of 2, 8, 12, and 12, were recorded, respectively (Figures 4 and 5).

The viral polymerase is composed of four domains, bearing three enzymatic activities: primase activity, reverse transcriptase & DNA-dependent DNA polymerase and RNase H: ribonuclease H activity. The *pol* gene contains 842 or 843 amino acids (aa) in most of genotypes, but it has 832 aa in genotype D and 845 aa in genotype A. The enzyme exhibits both a DNA polymerase and a reverse transcriptase activity and replicates the HBV genome from an encapsidated pregenomic RNA template (Nassal, 2008).

However, HBp also requires the presence of metal ions and the presence of the stem loop for polymerase/reverse transcriptase activity to occur (Bartenschlager and Schaller, 1992; Urban et al., 1998; Tavis et al., 1998). The fourth domain of HBp possesses its RNase H activity (Chang et al., 1990; Radziwill et al., 1990). This domain also plays a key role in HBV genome replication. In case of the putative conserved domains of the partial sequences of LC152751.1, multi domains: reverse transcriptases (RTs) from RtvS (cd01645); DNA polymerase (vir-al) C-terminal domain (pfam00336) and reverse transcriptase (RNA-dependent DNA polymerase) (pfam00078) were obtained (Fig. 6).

TCGCT **GGATGTGCTGCGCGTTT** TATC ATCTT CCTT CAT CCT GCT GCT ATGCC TCA TCT TGT GGT CCT TGT GGATT ATCAAGG TAT GTT
GCC CTT TG CTC TCA ATT CCAGG AT CCT CA ACC ACC AGC AC GGG ACC AT GC AGA AC CT GC AC GACT CCT GCT CA AGGA AC CT TAT GT ATCC CTC
CT GT GCT GT ATCAA AC CCT CGG AC GGA ATT GC AGC GT ATT CCC AT CC AT CCT GGG CTT CGG AAA AT CCT AT GGG AGT GGG CCT **CAGC**
CCGTTCTCTGGCTCAGTTACTAGTGCC ATT GT CAGT GGT CG TAGG GTT TCCC ACT GTT **GGCTTCAGTTATATGGATGATGTGGTA**
ITGGGGGCCAACGCTGTAA CAGC AT CGT GAG TCC CT GATT **ACCGCTGTACCAATTTCCTTTG** TCT TGG GT ATAC AT TA AAC CCT AA CAAA AC
AAGAGATGGGGTACTCTGAATT TATGGGATATGT CATTGGAT GTT ATGGG TATTGGG CATTGCC ACA AGA AC AT CAT AC AAAAA ATCAA AGA AT GT
TTT AGAAA AT CCT GTT AAGC AG
Forward: 58-**GGATGTGCTGCGCGTTT**-38

L669MAYWF: (58-**CAGCCCGTTCTCATGGCTCAGTTACTAGTGCC**-38)
M741VAYWF: (58-**GGCTTCAGTTATGTGGATGATGTGGTATTGGG**-38)
S3: (Forward, nt 752-771): **TATTGGGGGCCAACGCTGTAA**

LC152751.1

Reverse: 58-**CAAAAGAAAATTGGTAACAGCGCTA**-38
GTTAATGAAAAAAGGAGATAAATTAATTATGCTGCTAGGGTCTATCTAACCTAACCAATTGGCCCTAGACAAAGGCATTAAACCGTAT
TATCTGAACATCGAGTAAATCATTAACCTTAAACATGGCATTATTACATACGCTGTGAATGCTGGCATTATATAAGAGAGAAACTACACGCA
CGCCTCATCTGTGG **GTCACCATATTCTGGGAA** AGAGCTACAGCATGGAGGTTGGCTTCAAACCTCGATAAGGAATTGGACGAGTCTT
CTGTCCTAACATCGTGTGGGATTATTCCAGATCACCGATGGCTGGACGCGTCCGGAGCCA ACT CAAACAATTCAAGACAGGGACTCTCAAC
AGGATCACCAGAACACAGGCAAATCAGGTAGGAGCGGGAGCTTGGGGCAGGGTTCACCCACACAGGGCGGTCTGTGGGGTGGAGCCATTAG
CTCAGGGCGTATTGACAACAGAGCCAGCCAGCATCTCGTGCCTCCGCAATCGGAGTGAGGAAGACAGCCTACTCCCATCTCCACCTCTAA
GAGACAGT **CATCCCTCAGGCCATGCCACTGG** ACTCCACAACTTCCACAAAGCTATGCTAGATCCCAGTGAAGGGCCTATATT **GCTGCTGGTG**
GCTCCAGTTC CGGAACAGTAAACCTGTTCCGACTACTGCCCTACCCATATCGTCATCTCTCGAGGACTGGGGACCTGCAACGAACATGGAGA
ACACAAACATCAGGATTCCCTAGGACCCCTGCTGTACAGGCCGGTCTTCTGTGACAAGAATCCTCACAAATACCACAGAGTCT **AGACTCGT**
GGTGGACTTCTCTCA ATT TTCTAGGGGAGCACC

S2: (Forward, nt 2816-2835) **GTCACCATATTCTGGGAAAC**
P3: (Forward, nt 3193-3213)

LC152752.1

F1: (sense, nt 56-76, **GTCGCTGGCTCAGTC**)
Forward: 58-**AGACTCGTGTGGACTTCTCT**-38
CGCCCCGTCGCAAGAGATCTCAATCTGGGAATCTCAATGTTAGTATCCCTGGACTCAAAGGTTGGAAACTTTACTGGGCTTTATTCTCTACT
GTACTCTGTCTTAATCCCGAGTGGCAAACCTCTCTTCTCACATTCACTTACAGGAGGACATTAAATAGATGTCACAAATATGTGGCCCT
CTTACAGTAAATGAAAATAGGAGTAAATTAAATTATGCTGCTAGGTCTATCTAACCTAACCAATTGGCCCTTAGACAAGGCATTAAA
CCGTATTATCTGAACATGCAGTAAATCATAACTTCAAACACTAGGCATTATTACATACTCTGTGAGGCTGGCATTCTATATAAGAGAGAGACT
ACACCGAGCGCTCATTTGTGG **GTCACCATATTCTGGGAA** CAAGAGCTACAGCATGGAGGAGGTTGGCTTCCAAACCTCGACAAGGCATGGGG
ACGAATTTCTGTTC **CCAATCTCTGGGATTCTCCCGAT** CACCACTGTTGGACCCCTGCGTCCGGAGCCA ACT CAAACAATCAGGATGGGACTTC
AACCCCAAAAGGATCACTGGGAGAGGCAAATCAGGTAGGAGCGGGAGCATCCACACCAAGGTTCCACCCCTCTACAAGGAGTCTTGGAGGT
GGAGCCCCGTCGGAGCTCCGGGACATTGCAACACAGCTGCCAGGGCAGACCCCCCAGGGCGTCCACAAATCTCTGTGGTACAGTCACCTACTCGA
ATATCTCCACCTGTTCCAACACTGATCCTCACGCCACGCCATGCCAACAGTCAGCTGCCACCTGCTGTAGAGCACAAGTGA

LC152753.1

S2: (Forward, nt 2816-2835) **GTCACCATATTCTGGGAAAC**
TPR1: Reverse, 50-**TCGGGAAAAGATCCCAAGAGGATTGG**-30 at nt 2933-2909

CCTTCGATAATGGCATTAACCTATTATCTGAACATGCAGTAACTATTAACTTAGGCATTATTAACATACTCTGTGAATGCTGGCA
TTATATAAGAGAGAAACTACACGCGCCTCATCTGTGG **GTCACCATATTCTGGGAA** AGAGCTACAGCATGGAGGTTGGCTTCCAAAC
TCGATAAGGAAATGGACGAGTCTTCTGTCTCAATCGTGTGGGATTATTCCAGATCACCAGTGGAGCCTGCGTCCGGAGCCA ACT CAAACA
TTCAAGACAGGGCTTCAACTCCAACAAGGATCACCGACCAAGGCCAAATCAGGTAGGAGGAGGCTTGGGAGGGTTCACCCACACAGGG
CGGCTCTGTGGGGTGGAGGCAATTAGGCTCAGGGCGTATTGACAACAGAGCCAGCCAGGAGCATCTCGTGCCTCCCAATCGGAGTGGAGGAGACA
GCCTACTCCCATCTCCACCTCTAAAGAGACAGT **CATCCCTCAGGCCATGCCACTGG** ACTCCACAACTTCCACCAAGCTATGCTAGATCCCAGAGT
GAGGGGCCTATATT **GCTGCTGGCTCAGTC** CGGAACAGTAAACCTGTTCCGACTACTGCCCTACCCATATCGTCATCTCTCGAGGAC
TGGGGACCTGTCACCGAACATGGAGAACACAAACATCAGGATTCTCTCAATTAACTAGGGGAGGACCCACGTGTCCT
CACAAATACCAAGAGTCTAGACTCGTGTGGACTTCTCTCAATTAACTAGGGGAGGACCCACGTGTCCT

LC152757.1

S2: (Forward, nt 2816-2835) **GTCACCATATTCTGGGAAAC**
P3: (Forward, nt 3193-3213)

CATCCCTCAGGCCATGCCACTGG

F1: (sense, nt 56-76, **GTCGCTGGCTCAGTC**)

CTATCCTAACCTTACTAAATATTGCCCTTAGACAAAGGCATTAACCGTATTATCTGAACATGCAGTAACTCATTACTTCAAACACTAGGCATTA
TTTACATACTCTGTGGAAGGCTGGCATCCTATATAAGAGAGAAACTACACGCGCCTCATTTGTGG **GTCACCATATTCTGGGAAAC** AAGAGCT

ACAGCATGGGAGGTGGCTTCCAAACCTCGAGAAGGAATCCGGACGAGTCTTCTGGGATTGTTCCCGATCACCAAGTGG
 ACCCTCGCTTCGGAGCCAACCTAACACAATCCAGATTGGACTTCACACTCCAAAGGATCACCGACCAGAGCCAATCAGGTAGGAGCGGGAGCAT
 TCGGGCCAGGGTTCACCCCACACAGGGCGTCTTGTGGGGTGGAGCCTTAGGCTCAGGGCTATTGACAACAGTGCAGCCGACCTCCCTG
 CCTCCGCCAACCGCAGTCAGGAGACAGCCTACTCCATCTCCACCTCTAACAGAGACAGT**CATCCTCAGGGCATGCAGTG**AACTCCACAAACAT
 TCCACCAAGCTATGCTAGATCCAGAGTGGGGCTATATTTC**CCTGCTGGTGGCTCCAGTTC**CGGAACAGTAAACCTGTTCCGACTACTGCCT
 CACCCATATCGCAATCTTCTCGAGGACTGGGACCTGCACCGAACATGGAGAACACACATCAGGATTCTCTAGGAGCCCTGCTGTACAGG
 CGGGGTTTACTTGTGACAAGAACCTCACAATACCACAGAGTCTAGACTCGTGGTGGACTTCTCAATTCTAGGGGAGCACCGTGT
 CTGGCCAATTCGCAGTCCCCAACCTCCAATCAC

S2: (Forward, nt 2816–2835) **GTCACCATATTCTGGGAAC** LC152759.1

P3: (Forward, nt 3193–3213) **CATCCTCAGGGCATGCAGTG**

F1: (sense, nt 56–76, **CCTGCTGGTGGCTCCAGTTC**)

Figure 2. Primers of surface protein genes recorded in the partial sequences of HBV strains belonging to *pol* gene of HBV.

Mutations in the catalytic domain of the polymerase gene can affect the amino-acid sequence of the envelope protein (HBsAg) (Bartholomeusz et al., 1998). In particular, the genetic sequence for the neutralization domain of HBV known as a determinant, which is found within the HBsAg and located between amino acids 99 and 169, 5 actually overlaps the major catalytic regions of the viral polymerase protein from amino acid 454 to 524 and known as domains A and B. The amino acid sequence deduced from the overlapping polymerase gene showed a substitution of serine to threonine at position 413 in reverse trans-

criptase domain, due to the above T-A substitution. One can conclude that the presence of virus in some HBV-infected blood samples collected from different regions of KSA was serologically and molecularly confirmed by ELISA and Rt-PCR techniques. The nucleotide sequence of *pol* gene was determined and bioinformatically analyzed, and therefore, differences at the levels of DNA and encoding amino acids between the strains under investigation and those similar in GenBank and types of domains, its accession, and superfamily of the sequenced *pol* gene were reported.

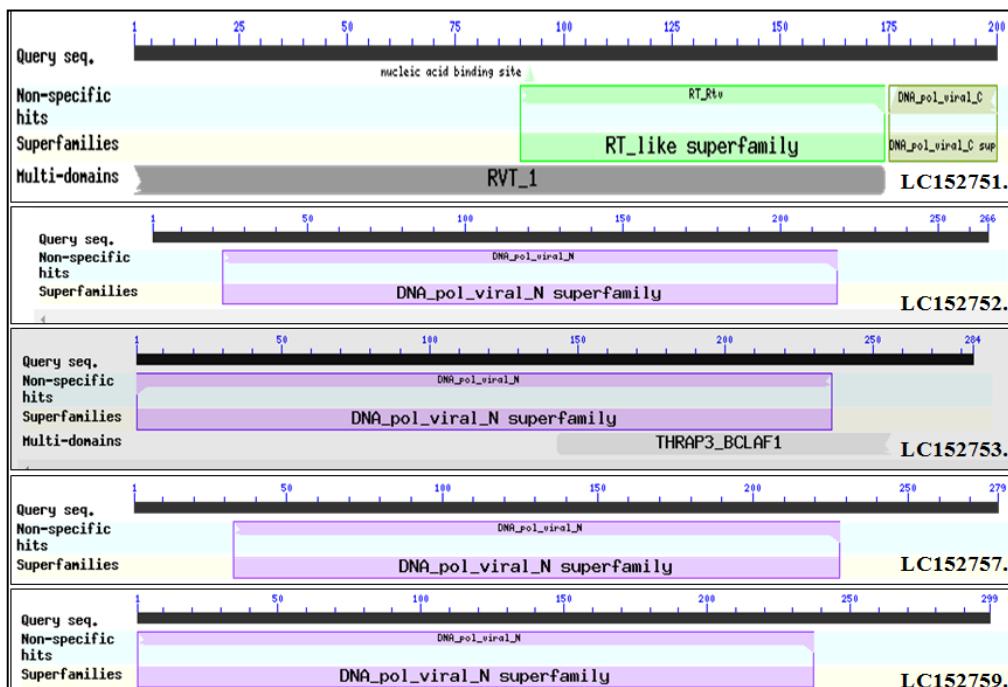


Figure 3. Putative conserved domains of the ORF number 1 of the partial sequences of HBV strains belonging to DNA_{pol}_viral_C and DNA_{pol}_viral_N domains.

gi_118896	1 . [161] . LYLRKNQTTLFKGKPYSWEHR. [2] . VQHNG. [12] . SSMVACSGHLLHNH. [25] . RTGLCSYKQIQTDR	256			
LC152752.1	1 . [22] . LYKRETRTRSASSCGSPYSWEQE	LQHGR. [11] . ESFCSQSCGIIRS	PVGPCVRSQLKQFRQ	89	
gi_82008114	1 . [156] . LYKRETRTRSASFCGSPYSWEQE	LQHGA	EPVCQQSQLGILPRA	SVGSPVQSQLQKSRL	212
gi_81967260	1 . [156] . LYKRETRTRSASFCGSPYSWEQE	LQHGS. [11] . ESLCAQSTGILSRP	SAGSSFQSKFQQSRL	223	
gi_81934952	1 . [160] . LYLRKNQTTLFKGKPYSWGHR. [2] . EQHNG. [12] . SSMVASSGHILHKQ. [25] . RTGGSVREKIQTNR	255			
gi_82005168	1 . [156] . LYKRETRTRSASFEGSPYSWEQE	LHHGA. [12] . ESFHQQSTRIFSR	PVGPCIQSKHQQSRL	224	
gi_123844506	1 . [156] . LYKRETRTRSASFEGSPYTWEQD	LQHGA. [10] . EPFHQQSSRIPRS	PVGPSIQSKYQQSRL	222	
gi_82001533	1 . [156] . LYKRETRTRSASFEGSPYSWEQE	LQHGA	ESFHQQSSGILSRP	PVGSSLQSKHRKSRL	212
gi_81946769	1 . [158] . LYKRETRTRSASFEGSPYSWEQE	LHHGR. [11] . EPFCSQPSGILSR	SVGPCIRSFQKSRL	225	
gi_75554236	1 . [156] . LYKRETRTRSASFEGSPYSWEQE	LQHGR. [11] . ESFCSQSSGILSR	SVGPRDRSQQHKQSRL	223	

<u>gi_118896</u>	257	EHLAR. [4]. SKITIGQQGSSPKTLYKSISSNFRNQTWAYNSSRNSGHTT. [1]. FSSASNNSNKSRSREKAYSSNSTSKR	331			
<u>LC152752.1</u>	90	GLQLQ. [1]. GSPTTGKSGRSGSFWARVHPTTRGSCGVEPLGSGRIDNRA	SRASSCLRQSARVKTAYSHLSTSKR 60			
<u>gi_82008114</u>	213	GLSQ. [1]. GQLARSHPGRSGSVRARVHSTTRRSFVELSGSGSNHIA	SSSSFCRHQSARVREAANSHLSTVER 283			
<u>gi_81967260</u>	224	GLQQK. [1]. GHLANGKQGRSGRLRSRVHHTTRWPVGMEPSGTRCSNNLA	SRSASCFHQSAVREEANPSLSTSKR 294			
<u>gi_81934952</u>	256	GFPKGK	SKITIGQQGSSQVSSPRSKSSSNFRNQTQANHSSWNQRHPT. [1]. YSTTSNTTQSRQREETYSSDAFKR 326			
<u>gi_82005168</u>	225	GLQPQ. [1]. GQLANSQRGRSWSVRSAHSSTRGSGFVEPSGTGQTNNA	SKPSCLCQQAAVRETAYPSLSTSER 295			
<u>gi_123844506</u>	223	GLSQ. [1]. GPLARGQQGRSWSLWTRVHPSPARRPFGVEPSVSGHTNNFA	SRSASCLHQSSVREAAYSHLSTTER 293			
<u>gi_82001533</u>	213	GLSQ. [1]. GHLARRQQGRSWSIRAGFHPTARRPFGVEPSGSGHTTYRA	SKFSACLYQSPVRKAAYSHLSTSKR 283			
<u>gi_81946769</u>	226	GLQPH. [1]. GHLARRQQGRSWSIRAGIHPTARRPFGVEPSGSGHTTNLA	SKFSACLYQSPVRKAAYSHLSTSKR 296			
<u>gi_75554236</u>	224	GLQPQ. [1]. GSMAGGKPGRRGSIRARVHPTTRRSFVGVEPSDSGHTDNA	SSASSCHHQSAADTKTAHDHLSNSR 294			
<u>gi_118896</u>	332	YSP. [4]. KSDFS. [1]. PGVRRI. [11]. CLWRSFYNTKPCGSCYIHHIVSSLDDWGPC	393			
<u>LC152752.1</u>	161	QSSS	GHAVE. [1]. HNIPPSY. [11]. CWWLQFRNSKPCSDYCLTHIVNLLEDWGPC 218			
<u>gi_82008114</u>	284	HSSS	GHEVE. [1]. YSIPPSY. [11]. CWWLQFRNSEPSCSDYCLSHLVNLLEDWGPC 341			
<u>gi_81967260</u>	295	HTST	GNAVE. [1]. NPVPGP. [11]. CWWLQFRDTEPCSDYCLSHIVNLLEDWGPC 352			
<u>gi_81934952</u>	327	HSPS. [4]. KSEPS. [1]. SGLCGGT. [11]. CLWRSFYNTEPCCGAYCLHHIVSSLEDWGPC	388			
<u>gi_82005168</u>	296	NSSS	GHALE. [1]. HDISPGS	CWWLQFRNSKPCSEYCLSHLVNLLEDWGPC 342		
<u>gi_123844506</u>	294	QSSS	GHAVE. [1]. YSIPFSS. [10]. CWWLQFRNSEPSCSDYCLSHLVNLLEDWGPC 350			
<u>gi_82001533</u>	284	HSSS	GHAVE. [1]. HNLPFNS. [11]. CWWLQFRNSKPCSDYCLSHLVNLLEDWGPC 341			
<u>gi_81946769</u>	297	QSSS	GHAVD	STVSTKL. [11]. LLVAPVQDTQPCSNYCLSHLVNLLEDWGPC 353		
<u>gi_75554236</u>	295	QSSS	GHAVE. [1]. HNFPFSS. [11]. CWWLQFRNSKPCSDYCLSHIVNLLEDWGPC 352			
<u>gi_118896</u>	1	[39]. HRVA. [1]. ALNLHLP TADLQWVHKTNAITGLYLSQNQAAQFNPHWIQKLPQYFGPLTINEK	109			
<u>LC152753.1</u>	1	RPAV. [1]. DLNLQLNLNVSI PWTQVKVGNFTGLYSSTPVPEWQTPSPHFIQKLPQYFGPLTINEK	70			
<u>gi_82008114</u>	1	[34]. RRVA. [1]. DLNLQLPNVS1 PWTQVKVGNFTGLYSSTPVPEWQTPSPHFIQKLPQYFGPLTINEK	104			
<u>gi_81967260</u>	1	[34]. RRVA. [1]. DLNLQLPNVS1 PWTQVKVGNFTGLYSSTPVPEWQTPSPHFIQKLPQYFGPLTINEK	104			
<u>gi_81934952</u>	1	[39]. HRVA	GLNLQLPTADL DWVHQTNAITGLYSTQTAKFNPEWKQPDFPKIHLSEDLFNLNYNNFCGPLTVNEK 108			
<u>gi_82005168</u>	1	[34]. QLVA. [1]. DLNLQLPNVS1 PWTQVKVGNFTGLYSSNIPVFNPDWQTPSFNPIHLQHNI	HRCEQFVGPLTVNEK 104			
<u>gi_123844506</u>	1	[34]. RRVA. [1]. DLNLQLPNVS1 PWTQVKVGNFTGLYSSNIPVFNPDWQTPSFNPIHLQHDI	ITKCEQFVGPLTVNEK 104			
<u>gi_82001533</u>	1	[34]. RRVA. [1]. DLNLGNLNVLNS1 PWTQVKVGNFTGLYSSTPVPEWQTPAFPKIHLHEDIA	NKCQQFVGPLTVNEK 106			
<u>gi_81946769</u>	1	[36]. RRVA. [1]. DLNLGNLNVLNS1 PWTQVKVGNFTGLYSSTPVPEWQTPAFPKIHLHEDIA	NKCQQFVGPLTVNEK 104			
<u>gi_75554236</u>	1	[34]. RRVA. [1]. DLNLGNLNVLNS1 PWTQVKVGNFTGLYSSTPVPEWQTPAFPKIHLQEDII	NRCQQYVGPLTVNEK 104			
<u>gi_118896</u>	110	RKLQLNPFPARFFPKATKYFPLIKGIKINNNYPNFALEHHFFATANLWTLWEAGILYLRNQTTTFKGPYWSWEHR.	[2]. V 186			
<u>LC152753.1</u>	71	RKLKLIMPARFPVNLTQYLPLDKGIKPYYPEHAVHNFKTRHLYLHQLKAGILYKRETRTRSASFCGSPYWSWGR.	[2]. A 147			
<u>gi_82008114</u>	105	RRLKLIMPARFPVNLTQYLPLDKGIKPYYPEHAVHNHYFQTRHLYLHQLKAGILYKRETRTRSASFCGSPYWSWEQE	L 179			
<u>gi_81967260</u>	105	RRLKLIMPARFPVNLTQYLPLDKGIKPYYPEHAVHNHYFQTRHLYLHQLKAGILYKRETRTRSASFCGSPYWSWEQE	L 179			
<u>gi_81934952</u>	109	RKLKLNPFPARFFPKATKYFPLSKGIKINNNYPDFSIEHHFAATYLWTLWESGILYLRNQTTTFKGPYWSWGR.	[2]. E 185			
<u>gi_82005168</u>	105	RRLNLNMPARFPVNSTKYLSELEKGIPYYPDNVNVNHFQTRHLYLHQLKAGILYKRETRTRSASFCGSPYWSWEQE	L 179			
<u>gi_123844506</u>	105	RRLKLVPMPARFPVNSTKYLPLDKGIKPYYPEAVNHVFQTRHLYLHQLKAGILYKRETRTRSASFCGSPYWSWEQE	L 179			
<u>gi_82001533</u>	105	RRLQLIMPARFPVNLTQYLPLDKGIKPYYPEHLVNVHYFQTRHLYLHQLKAGILYKRETHSASFCGSPYWSWEQE	L 179			
<u>gi_81946769</u>	107	RRLKLIMPARFPVNLTQYLPLDKGIKPYYPEPDHVNVNVHYFQTRHLYLHQLKAGILYKRETRTRSASFCGSPYWSWEQE	L 181			
<u>gi_75554236</u>	105	RRLKLIMPARFPVNLTQYLPLDKGIKPYYPEHLVNVHYFQTRHLYLHQLKAGILYKRETRTRSASFCGSPYWSWEQE	L 179			
<u>gi_118896</u>	187	QHN	G. [12]. SSMVACSG	HLLHN. [25]. RTGLCSYKIQTDRLEHLAR. [4]. SKITIGQQ. [120]. 393		
<u>LC152753.1</u>	148	WRG. [9]. G. [11]. FDHQLDP. [2]. GANSNN. [20]. GASTPRFTPLQGSLLGGAR. [1]. SSGHIAPQ. [48].	284			
<u>gi_82008114</u>	180	QHG	A	EPVCQSQL	SVGSPVSQLKQSRGLQSQ. [1]. GQLARSHP. [115]. 341	
<u>gi_81967260</u>	180	QHG	S. [11]. ESLCAQST	GILSRP	SAGSSFQSKFQQSRLGLQK. [1]. GHLANGKQ. [115]. 352	
<u>gi_81934952</u>	186	QHN	G. [12]. SSMVASSG	HILHQK. [25]. RTGGSREKIQTNRGLFFGK	SKITIGQQ. [120]. 388	
<u>gi_82005168</u>	180	HHG	A. [12]. ESFHQQST	RIFRSA	PVGPCIQSKHQQLQSQ. [1]. GQLANSQR. [104]. 342	
<u>gi_123844506</u>	180	QHG	A. [10]. EFPHQQSS	RIPRSR	PVGPSIQSKEYQSQSRGLQSQ. [1]. GPLARGQQ. [114]. 350	
<u>gi_82001533</u>	180	QHG	A	ESFHQQSS	GILSRP	PVGSSLQSKHRKSRSLGLQSQ. [1]. GHLARRQQ. [115]. 341
<u>gi_81946769</u>	182	HHG	R. [11]. EFPCSQPS	GILSR	SVGPCIRSQFKQSRGLQPH. [1]. GHLARRQQ. [114]. 353	
<u>gi_75554236</u>	180	QHG	R. [11]. EFPCSQSS	GILSR	SVGPRDRSQRHKQSRGLQPH. [1]. GSMAGGKP. [115]. 352	

Figure 4. The superfamily (cl02825) of DNA_pol_viral_N domain of ORF number 1 of partial sequence of HBV strains LC152752.1 and LC152753.1 *pol* gene which belongs to the DNA polymerase (viral) N-terminal domain (pfam00242).

<u>gi_118896</u>	1	[161]. LYLRKNQTTLTFKGPYWSWEHR.	[2]. VQHNG. [12]. SSMVACSGHLLHNH. [25]. RTGLCSYKIQTDRLEHLAR. [4]. SKITIGQQ.	[120]. 393	
<u>LC152758.1</u>	1	[32]. LYKRETRTRSASSCGSPYWSWEQE	LQHGR. [11]. ESFCSQSCGIISRS	PVGPCVRSQLQFQRQ 99	
<u>gi_82008114</u>	1	[156]. LYKRETRTRSASSCGSPYWSWEQE	LQHGA	EPVCQSQLGILPRA	SVGSPVSQLQFQRQ 212
<u>gi_81967260</u>	1	[156]. LYKRETRTRSASSCGSPYWSWEQE	LQHGS. [11]. ESLCAQSTGILSRP	SAGSSFQSKFQQSRL 223	
<u>gi_81934952</u>	1	[160]. LYLRKNQTTLTFKGPYWSWGR.	[2]. EQHNG. [12]. SSMVASSGHLHQK. [25]. RTGGSVREKIQTNRGLFFGK	255	
<u>gi_82005168</u>	1	[156]. LYKRETRTRSASSFCGSPYWSWEQE	LHHGA. [12]. ESFHQQSTRIFSRA	PVGPCIQSKHQQSRL 224	
<u>gi_123844506</u>	1	[156]. LYKRETRTRSASSFCGSPYWSWEQE	LQHGA. [10]. EPFHQQSSRIPSR	PVGPSIQSKEYQSQSR 222	
<u>gi_82001533</u>	1	[156]. LYKRETHSASFCGSPYWSWEQE	LQHGA	ESFHQQSSGILSRP	PVGSSLQSKHRKSRSL 212
<u>gi_81946769</u>	1	[158]. LYKRETRTRSASSFCGSPYWSWEQE	LHHGR. [11]. EPFCQSPGSLRSR	SVGPCIRSQFKQSR 225	
<u>gi_75554236</u>	1	[156]. LYKRETRTRSASSFCGSPYWSWEQE	LQHGR. [11]. ESFCQSSGILSR	SVGPRDRSQRHKQSR 223	
<u>gi_118896</u>	257	EHLAR. [4]. SKITIGQQSSPKTLYKSISSNFRNQTWAYNSSRNSGHTT.	[1]. FSSASNNSNKSRSREKAYSSNSTSKR	331	
<u>LC152758.1</u>	100	GLQLQ. [1]. GSPTTGKSGRSGSFWARVHPTTRGSCGVEPLGSGRIDNRA	SRASSCLRQSARVKTAYSHLSTSKR	170	
<u>gi_82008114</u>	213	GLQS. [1]. GQLARSHPGRSGSVRARVHSTTRRSFVELSGSGSNHIA	SSSSFCRHQSARVREAANSHLSTVER	283	
<u>gi_81967260</u>	224	GLQQK. [1]. GHLANGKQGRSGRLRSRVHHTTRWPVGMEPSGTRCSNNLA	SRSASCFHQSAVREEANPSLSTSKR	294	
<u>gi_81934952</u>	256	GFPKGK	SKITIGQQSSQVSSPRSKSSSNFRNQTQANHSSWNQRHPT.	[1]. YSTTSNTTQSRQREETYSSDAFKR 326	
<u>gi_82005168</u>	225	GLQPQ. [1]. GQLANSQRGRSWSVRSAHSSTRGSGFVEPSGTGQTNNA	SKPSCLCQQAAVRETAYPSLSTSER	295	
<u>gi_123844506</u>	223	GLQS. [1]. GPLARGQQGRSWSLWTRVHPSPARRPFGVEPSVSGHTNNFA	SRSASCLHQSSVREAAYSHLSTTER	293	
<u>gi_82001533</u>	213	GLQS. [1]. GHLARRQQGRSWSIRAGFHPTARRPFGVEPSGSGHTTYRA	SKSACLYQSPVRKAAYPSVSTFEK	283	
<u>gi_81946769</u>	226	GLQPH. [1]. GHLARRQQGRSWSIRAGIHPTARRPFGVEPSGSGHTTNLA	SKFASCLYQSPVRKAAYPSVSTFEK	296	
<u>gi_75554236</u>	224	GLQPQ. [1]. GSMAGGKPGRRGSIRARVHPTTRRSFVGVEPSDSGHTDNA	SSASSCHHQSAADTKTAHDHLSNSR	294	
<u>gi_118896</u>	332	YSP. [4]. KSDFS. [1]. PGVRRI. [11]. CLWRSFYNTKPCGSCYIHHIVSSLDDWGPC	393		
<u>LC152758.1</u>	171	QSSS	GHAVE. [1]. HNIPPSY. [11]. CWWLQFRNSKPCSDYCLTHIVNLLEDWGPC	228	
<u>gi_82008114</u>	284	HSSS	GHEVE. [1]. YSIPPN. [11]. CWWLQFRNSEPSCDYCLSHLVNLLEDWGPC	341	
<u>gi_81967260</u>	295	HTST	GNAVE. [1]. NPVPGP. [11]. CWWLQFRDTEPCSDYCLSHIVNLLEDWGPC	352	
<u>gi_81934952</u>	327	HSPS. [4]. KSEPS. [1]. SGLCGGT. [11]. CLWRSFYNTEPCCGAYCLHHIVSSLEDWGPC	388		
<u>gi_82005168</u>	296	NSSS	GHALE. [1]. HDISPGS	CWWLQFRNSKPCSEYCLSHLVNLLEDWGPC 342	
<u>gi_123844506</u>	294	QSSS	GHAVE. [1]. YSIPFSS. [10]. CWWLQFRNSEPSCDYCLSHLVNLLEDWGPC	350	
<u>gi_82001533</u>	284	HSSS	GHAVE. [1]. HNLPFNS. [11]. CWWLQFRNSKPCSDYCLSHLVNLLEDWGPC	341	
<u>gi_81946769</u>	297	QSSS	GHAVD	STVSTKL. [11]. LLVAPVQDTQPCSNYCLSHLVNLLEDWGPC 353	

gi_75554236	295	QSSS	GHAVE.[1].HNFPSS.[11].CWWLQFRNSKPCSDYCLSHIVNLLEDWGPC	352
gi_118896	1	[120]	FPKATKYFPLIKGIKNNYNPFALEHFATANYLWTLWEAGILYLRLNKNTTFTKGKPYPSWEHR.	[2].VQHNG 190
LC152759.1	1		YPNLTKYPLDKGIKPKYYPEHAVNHYFKTRHYLHTLWKAGILYKRETRTRSASF CGSPYSWEQE	LQHGR 68
gi_82008114	1	[115]	YPNVTKYPLDKGIKPKYYPEHAVNHYFKTRHYLHTLWKAGILYKRETRTRSASF CGSPYSWEQE	LQHGA 183
gi_81967260	1	[115]	FPKLTKYFPLEKGIKPKYYPEHAVNHYFKTRHYLHTLWKAGILYKRETRTRSASF CGSPYSWEQE	LQHGS 183
gi_81934952	1	[119]	FPKATKYFPLSKGIKNNYNPDFSIEHF AAATYLWTLWESGILYLRNKNTTFTKGKPYPSWGHR.	[2].EQHNG 189
gi_82005168	1	[115]	YPNSTKYLSLEKGIKPKYYPDNVNHYFQTRHYLHTLWKAGILYKRETRTRSASF CGSPYSWEQE	LHHGA 183
gi_123844506	1	[115]	FPNSTKYPLDKGIKPKYYPENVNHYFQTRHYLHTLWKAGILYKRETRTRSASF CGSPYSWTWQD	LQHGA 183
gi_82001533	1	[115]	YPNVTKYPLDKGIKPKYYPEHLVNHYFQTRHYLHTLWKAGILYKRETRTHASF CGSPYSWEQE	LQHGA 183
gi_81946769	1	[117]	YPNSTKYPLDKGIKPKYYPDHVNVNHYFQTRHYLHTLWKAGILYKRETRTRSASF CGSPYSWEQE	LHHGR 185
gi_75554236	1	[115]	YPNLTKYPLDKGIKPKYYPEYAVNHYFKTRHYLHTLWKAGILYKRETRTRSASF CGSPYSWEQE	LQHGR 183
gi_118896	191	[12]	SSMVCASGHLLHNH.	[25].RTGLCSYKIQJQTDRLEHLAR.[4].SKITIGQQGSSPKTLYKSISSNFRNQTWA 294
LC152759.1	69	[11]	ESFCSQSSGIVSRS	PVGPCVRSQLQKSRLGLQ. [1].GSPTRGKSGRSGSIRARVHPTT RSCGVE 143
gi_82008114	184		EPVCGQSQLGILPRA	SVGSPVQSQLQKSRLGLQ. [1].GOLARSHPGRSGSVRARVHSTT RSRFVE 247
gi_81967260	184	[11]	ESLCAQSTGILSRP	SAGSSFQSKFQQSRLGLQK. [1].GH LANGKQGRSGRLRSRVH TTRWPVGME 258
gi_81934952	190	[12]	SSMVASSGHILHQ.	[25].RTGGSVREKIQTNRGLGFPK SKITIGQQGSSQVS PRSKSSNFRNQTQA 289
gi_82005168	184	[12]	ESFHQGSTRIFSR	PVGPCIQSKHQQSRLGLQ. [1].GQ LANSQRGRSWRSRAHSSTRGSGFVE 259
gi_123844506	184	[10]	EPFHQGSSRIPSRS	PVGPSI QSKYQQSRLGLQ. [1].GPLARQGQGRSWSL WTRVHP SARRPFGVE 257
gi_82001533	184		ESFHQGSSGILSRP	PVGSSLQSKHRSRSLGLQ. [1].GH LARRQQGRSWSI RAGIHTARRPFGVE 247
gi_81946769	186	[11]	EPFCSQPSSGILSR	SVGPCIRSQFKQSRLGLQPH. [1].GH LARRQQGRSWSI RAGIHTARRPFGVE 260
gi_75554236	184	[11]	ESFCQSSGILSR	SVGPRDRSQQFKQSRLGLQ. [1].GSMAGGKPGRGGSIRARVHPTT RRSFGVE 258
gi_118896	295		YNSSRNSGHTT.	[1].FSSA NSNKSRSREKAYSSNSTS KRYSP. [4].KSDFS. [1].PGVRRRI. [11].CLWRSFY 370
LC152759.1	144	PLGSGRIDNSA	SRTSSCLRQSAVRKTAYSHLSTS KRSQSS	GHAVE. [1].HNIPPSY. [11].CWNLQFR 214
gi_82008114	248	LSGSGSNHNIA	SSSSFCRHQSAVREAANSHLSTVERHSS	GHEVE. [1].YSIPPNS. [11].CWNLQFR 318
gi_81967260	259	PSGTRCSNNLA	SRSASC FHQSAVREEANPSLSTS KRHTST	GNAVE. [1].NPVPPGP. [11].CWNLQFR 329
gi_81934952	290	NHSSWNQRHPT.	[1].YSTTSNTT QSRQREETYSSD S AFKRHSP. [4].KSEPS. [1].SGLCCGT. [11].CLWRSFY 365	
gi_82005168	260	PSGTQTNNA	SKPSCL QQAAVRETAYPSLSTS KRNSSS	GHALE. [1].HDISPGS CWNLQFR 319
gi_123844506	258	PSVSGHTNNFA	SRSASCL HQSSVREAAYSHLSTTERQSSS	GHAVE. [1].YSIPPSS. [10].CWNLQFR 327
gi_82001533	248	PSGSGHTTYRA	SKSASCL YQSPVRAA YPSVSTFEKHSSS	GHAVE. [1].HNLPNNS. [11].CWNLQFR 318
gi_81946769	261	PSGSGHTTNLA	SKFASCL YQSPVRAA YPSVSTFEKHSSS	GHAVD STVSTKL. [11].LLVAPVQ 330
gi_75554236	259	PSDSGHTDNA	SSASCHHQ SADTKTAHDHLS TSRNQSSS	GHAVE. [1].HNFPSS. [11].CWNLQFR 329
gi_118896	371	NTKPCGSYCTHIVSSLDDWGPC	393	
LC152759.1	215	NSKPCSDYCLTHIVNLLEDWGPC	237	
gi_82008114	319	NSEPCSDYCLSHLVNLLEDWGPC	341	
gi_81967260	330	DTEPCSDYCLSHINLLEDWGPC	352	
gi_81934952	366	NTEPCGAYCLHHIVSSLEDWGPC	388	
gi_82005168	320	NSKPCSEYCLSHLVNLLEDWGPC	342	
gi_123844506	328	NSEPCSDYCLSHLVNLQDWGPC	350	
gi_82001533	319	NSKPCSDYCLSLHVNLREDWGPC	341	
gi_81946769	331	DTQPCSNYCLSHLVNLLEDWGPC	353	
gi_75554236	330	NSKPCSDYCLSHIVNLLEDWGPC	352	

Figure 5. The superfamily (cl02825) of DNA_pol_viral_N domain of ORF number 1 of partial sequence of HBV strains LC152757.1 and LC152759.1 pol gene which belongs to the DNA polymerase (viral) N-terminal domain (pfam00242).

Name	Accession	Description	Interval	E-value
RT_Rtv	cd01645	RT_Rtv: Reverse transcriptases (RTs) from retroviruses (Rtvs). RTs catalyze the conversion of ...	90-174	1.77e-09

Pssm-ID: 238823 [Multi-domain] **Cd Length:** 213 **Bit Score:** 53.83 **E-value:** 1.77e-09

10	20	30	40	50	60	70	80
....********

LC152751.1 90 PMGVGLSPFLLAQFTSACISVVRRAFPHCLAFSYMDDVVLGAKSVQHRESLITAVTNFLLLSLGIHLNPKNKRWGYSLNF 169
Cdd:cd01645 130 PQGMKNSPTICQSFVAQALEPFRKQYPDIVIYH YMDDILIASLEGQLREIYELRQTLLRWGLTIPPEKVQK-EPPFQY 208
.....*

LC152751.1 170 MGYVI 174
Cdd:cd01645 209 LGYEL 213

Name	Accession	Description	Interval	E-value
DNA_pol_viral_C	pfam00336	DNA polymerase (viral) C-terminal domain;	175-200	1.92e-08

Pssm-ID: 144068 **Cd Length:** 245 **Bit Score:** 51.46 **E-value:** 1.92e-08

10	20
....**

.....*

LC152751.1 175 GCYGSLPQEHI IQKIKECFRKLPVNR 200
Cdd:pfam00336 1 GSYGSLPQDHIVKKISR CFRKLPVNR 26

Name	Accession	Description	Interval	E-value
RTV_1	pfam00078	Reverse transcriptase (RNA-dependent DNA polymerase); A reverse transcriptase gene is usually ...	1-174	4.15e-20

Pssm-ID: 249567 [Multi-domain] **Cd Length:** 196 **Bit Score:** 83.92 **E-value:** 4.15e-20

10	20	30	40	50	60	70	80
....********

.....*

LC152751.1 1 SLDVSAAFYHLPLHPAAMPHLLVGSSLSRYVARLssns RILNHQHGtmqn lhdscsrnly vsllly qt fgrklhv ysh 80
Cdd:pfam00078 61 KLDLKAFDSIPLDPLDRPLTAFGFPGRFIRTFSV --- RVNGNPGG ----- 103
90 100 110 120 130 140 150 160



Figure 6. The superfamilies of DNA_{pol}_viral_N domains of ORF number 1 of partial sequence of HBV strain LC152751.1 *pol* gene which belongs to the DNA polymerase (viral) C-terminal domain.

Acknowledgments

The authors would like to thank King Abdulaziz City for Science and Technology (KA-CST) represented by the General Directorate for Research Grants through Research Proposal # AT1518-37 for their financial support to accomplish this study. Thanks also due to Prof. Sonya Mohamed, ARC, Giza, Egypt for her sincere help to accomplish the bioinformatics analyses.

REFERENCES

- Akcam, F.Z., Uskun, E., Avsar, K., Songur, Y., Hepatitis B virus and hepatitis C virus seroprevalence in rural areas of the southwestern region of Turkey. International Journal of Infectious Diseases 13: 274-284 (2009).
- Alam, M.M., Zaidi, S.Z., Shaukat, S., Sharif, S., Angez, M., Naeem, A., Saleha, S., Butt, J.A., Malik, S.A., Common genotypes of Hepatitis B virus prevalent in injecting drug abusers (addicts) of North West Frontier province of Pakistan. Virology J. 4: 1-6 (2007).
- Aljarbou, A.N., The emergent concern of hepatitis of hepatitis B globally with special attention to Kingdom of Saudi Arabia. Int. J. Health Sci. (Qassim) 7: 333–340 (2013).
- Almeida, M.K.C., de Santos, K.N., dos Fecury, A. A., Oliveira, C.S.F., de Freitas, A.S., Quaresma, J.A.S., Fuzii, H.T., Martins, L.C., Prevalence of viral hepatitis B and C in riverside communities of the Tucurui Dam, Para, Brazil. Journal of Medical Virology 84, 1907-1912 (2012).
- Amini-Bavil-Olyaee S., Sarrami-Forooshani, R., Mahboudi, F., Sabahi, F., Adeli, A., Noorinayer, B., Azizi, M., Zali, M.R., Genotype characterization and phylogenetic analysis of hepatitis B virus isolates from Iranian patients. J. Med. Virol. 75: 227-234 (2005).
- Banks, T., Khang, J., Watts, I., Tyrosvoutis, M.E. G., Min A.M., Tun N.W., Keereecharoen, L., Simmawong, W., Wanyatip, S., Hanboonkunupakarn, B., Nosten, F., McGready, R., High hepatitis B seroprevalence and risk factors for infection in pregnant women on the Thailand-Myanmar Border. Journal of Infection in Developing Countries 10: 384-388 (2016).
- Bartenschlager, R., Schaller, H., Hepadnaviral assembly is initiated by polymerase binding to the encapsidation signal in the viral RNA genome. EMBO. 7: 4185-4192 (1992).
- Bartholomeusz, A., Schinazi, R., Locarnini, S., Significance of mutations in the hepatitis B virus polymerase selected by nucleoside analogues and implications for controlling chronic disease. Viral Hepatitis Rev. 4:67-187 (1998).
- Chang, L.J., et al., Mechanism of translation of the hepadnaviral polymerase (P) gene. Proc. Natl. Acad. Sci. USA. 87: 5158-5162 (1990).
- Dai, J.J., HuaMing, T., QingXi, M., QuanHong, Z., Anti-hepatitis B virus activities of friedelolactones from *Viola diffusa* Ging. Phytomedicine 22: 724-729 (2015).
- Datta, S., Banerjee, A., Chandra, P.K., Chowdhury, A., Chakravarty, R., Genotype, phylogenetic analysis and transmission pattern of occult hepatitis B virus (HBV) infection in families of asymptomatic HBsAg carriers. J. Med. Virol. 78: 53-59 (2006).
- Echevarría, J.M., Avellón A., Magnus L.O., Molecular epidemiology of hepatitis B virus in Spa-in: identification of viral genotypes and prediction of antigenic subtypes by limited sequencing. J. Med. Virol. 76: 176–184 (2005).
- Howard, C.R., The biology of hepadnaviruses. J. Gen. Virol. 67: 1215–1235 (1986).
- Huang, C.J., Wu, C.F., Lan, C.Y., Sung, F.Y., Lin C.L., et al., Impact of genetic heterogeneity in polymerase of hepatitis B virus on dynamics of viral load and hepatitis B progression. PLoS ONE 8: e70169 (2013).
- Jones, S.A., Jianming, H., Protein-primed terminal transferase activity of hepatitis VB virus polymerase. J. Virol. 87: 2563–2576 (2013).
- Lanford, R.E., Kim, Y.H., Lee, H., Notvall, L., Beames, B., Mapping of the hepatitis B virus reverse transcriptase TP and RT domains by trans complementation for nucleotide priming and by protein-protein interaction. J. Virol. 73:1885-1893 (1999).

- Lin, C.L., Kao, J.H., Review article: novel therapies for hepatitis B virus cure-advances and perspectives. *Aliment Pharmacol. Ther.* 44: 213–22 (2016).
- Mack, D.H., Bloch, W., Nath, N., Sninsky, J.J., Hepatitis B virus particles contain a poly-peptide encoded by the largest open reading frame: A putative reverse transcriptase. *J. Virol.* 62: 4786 -4790 (1998).
- Mazet-Wagner, A.A., Baclet, M.C., Loustaud-Ratti, V., Denis, F., Alain, S., Real-time PCR quantitation of hepatitis B virus total DNA and covalently closed circular DNA in peripheral blood mononuclear cells from hepatitis B virus-infected patients. *Journal of Virological Methods* 138: 70-79 (2006).
- Nassal, M., Hepatitis B viruses: reverse transcription a different way. *Virus Res.* 13: 235-249 (2008).
- Prasad, B., Saraf, N., Parikh, P., Ingle, M., Phadke, A., Sawant P., Cross sectional study of prevalence and risk factors of hepatitis B and heaptitis C infection in a rural village of India. *Arqu-ivos de Gastroenterologia.* 52: 321-324 (2015).
- Radziwill, G., et al., Mutational analysis of the heaptitis B virus P gene product: domain structure and Rnase H activity. *J. Virol.* 64: 613-620 (1990).
- Samal, J., Kandpal, M., Vivekanandan, P., HBeAg-induced miR-106b promotes cell growth by targeting the retinoblastoma gene. *Scientific Report* 7: 1-10 (2017).
- Schaefer, S., Glebe, D., Wend, U.C., Oyunbileg, J., Gerlich, W.H., Universal primers for real-time amplification of DNA from all known Orthohepadnavirus species. *Journal of Clinical Virology* 27: 30-37 (2003).
- Sharavanan, T.K.V., Premalatha, E., Dinakaran, N., Seroprevalence of hepatitis B surface antigen among rural pregnant women attending a tertiary care hospital. *Scholars Journal of Applied Medical Sciences* 2: 1351-1354 (2014).
- Tadokoro, K., Kobayashi, M., Yamaguchi, T., Suzuki, F., Miyauchi, S., Egashira, T., Kumada, H., Classification of hepatitis B virus genotypes by the PCR-Invader method with genotype-specific probes. *Journal of Virological Methods* 138: 30-39 (2006).
- Tavis, J.E., Massey, B., Gong, Y., The duck hepatitis B virus polymerase is activated by its RNA packaging signal, epsilon. *J. Virol.* 72: 5789-5796 (1998).
- Toh, H., Hayashida, H., Miyata, T., Sequence homology between retroviral reverse transcriptase and putative polymerases of hepatitis B virus and cauliflower mosaic virus. *Nature* 305: 827-829 (1983).
- Urban, M., McMillan, D.J., Canning, G., Newell, A., Brown, E., Mills, J.S., Jupp, R., In vitro activity of hepatitis B virus polymerase: requirement for distinct metal ions and the viral epsilon stem-loop. *J. Gen. Virol.* 79: 1121-1131 (1998).
- Zanella, I., Rossini, A., Domenighini, D., Albertini, A., Cariani, E., Quantitative analysis of hepatitis B virus DNA by real-time amplification. *European Journal of Clinical Microbiology & Infectious Diseases* 21: 22-26 (2002).