A DISSERTATION ON
STUDY OF PHYLOGENETIC ANALYSIS OF HEPATITIS B VIRUS SUB GENOTYPES IN INDIA

SUBMITTED TO THE
DEPARTMENT OF BIOSCIENCES
INTEGRAL UNIVERSITY, LUCKNOW

IN PARTIAL FULFILLEMENT
FOR THE
DEGREE OF MASTER OF SCIENCE
IN MICROBIOLOGY

BY
Urooj Fatima
MSc Microbiology (IV Semester)
DEPARTMENT OF BIOSCIENCES
INTEGRAL UNIVERSITY

UNDER THE SUPERVISION OF
Dr Sabihur Rahman Farooqui
Assistant Professor
Integral University, Lucknow

INTEGRAL UNIVERSITY

Established Under U.P. Act No 09
of 2004 by State Legislation
Approved by University
Grants Commission
Phone No.: +91 (0552) 2890812, 2890730, 3296117, 6451039,
Fax No.: 0522-2890809
Kursi Road, Lucknow-226026, Uttar Pradesh (INDIA)

TO WHOM IT MAY CONCERN

This is to certify that Mrs. Urooj Fatima a student of M.Sc. Microbiology (II Year, IV semester) Integral University, has completed her four months(Feb-June) dissertation work entitled "Phylogenetic Analysis of HBV sub genotypes in India" successfully.
She has completed this work from 11-02-2022-11-05-2022 under the supervision of Dr. Sabihur Rahman Farooqui. The dissertation was compulsory part of her M.Sc. degree. I wish her good luck and future

(Dr. Snober S. Mir)
Head,
Department of Biosciences,

INTEGRAL UNIVERSITY

Established Under U.P. Act No 09 of 2004 by
State Legislation Approved by
University Grants Commission
Phone No.: +91 (0552) 2890812, 2890730, 3296117, 6451039, Fax No.: 0522-2890809
Kursi Road, Lucknow-226026, Uttar Pradesh (INDIA)

CERTIFICATE OF ORIGINAL WORK

This is to certify that Mrs. Urooj Fatima a student of M.Sc. Microbiology (II Year, IV semester) Integral University, has completed her four months (Feb-June) dissertation work entitled "Phylogenetic Analysis of HBV sub genotypes in India" successfully, under my guidance and supervision. The results reported by her are genuine and script of the thesis has been written by the candidate himself. The thesis entitled "Phylogenetic analysis of hepatitis B virus subgenotypes in India" is therefore, being forwarded for the acceptance in fulfilment of the requirements for the award of the degree of Masters of Science in Microbiology, Department of biosciences, Integral university, Lucknow

(Dr. Sabihur Rahman Farooqui)
Assistant Professor
Department of Biosciences

ACKNOWLEDGEMENT

First of all I am grateful to God for the good health and wellbeing that were necessary to complete this work.

I wish to express my sincere thanks to Dr. Snober Mir (Head of department of Biosciences) for giving me an opportunity to join the departmental laboratory and providing all the necessary facilities for the project.

I place on record, my sincere thank you to Dr Sabihur Rahman Farooqui (Assistant professor, Department of biosciences) for the invaluable guidance throughout my project and academic session. It would not be possible for me to complete my work on time without sir's continuous guidance. He cooperated with me in every part of this project. Thank you sir for such guidance.
I am happy to work under such a sincere mentor.

I take this opportunity to express gratitude to all of the Department faculty members for their help and support.

I am also thankful to my parents and my partner Mr Noor Ahmad for motivating me and providing me with all the facilities to complete my work on time. Their blessings and support have helped me a lot to achieve my goal successfully.
I also place on record my sense of gratitude to one and all, who directly or indirectly, have lent their hand in this venture.

UROOJ FATIMA

TABLE OF CONTENTS

LIST OF ABBREVIATIONS:

C
cccDNA- Closed covalent circular DNA

D
DNA- Deoxyribonucleic acid

G
GTP- Guanosine triphosphate

H
HBcAg- Hepatitis B core antigen
HBeAg- Hepatitis E core antigen
HBsAg- Hepatitis B surface antigen
HCC- Hepatocellular carcinoma
HBV- hepatitis B virus

M
mRNA- messenger RNA

N
NCBI- National Center for Biotechnology Information
NPC- Nuclear pore complex
NTCP- Sodium taurocholate cotransporting polypeptide

O
ORF- Open Reading Frame

P
PgRNA- Pregenomic RNA

R
rcDNA- relaxed circular DNA
RNA-Ribonucleic Acid

LIST OF FIGURES:

INTRODUCTION

Hepatitis B virus (HBV) is partially double stranded DNA virus, a species of the genus Orthohepadanavirus and a member of the Hepadanaviridae family of viruses. HBV is a small, partially double-stranded DNA genome (3.2 kb) encoding four genes—HBsAg (surface envelope glycoprotein), HBcAg (viral capsid protein), HBV Pol/RT (polymerase reverse transcriptase), and X gene (transcriptional activator).

Hepatitis B virus (HBV) is a viral infection that attacks the liver and can cause both acute and chronic disease.Despite the availability of an effective vaccine, HBV infections continue to be a public health problem . In 2015, the World Health Organization (WHO) estimated that over 257 million people are chronically infected with HBV . Globally, HBV infections account for an estimated 887,000 deaths mostly from cirrhosis and hepatocellular carcinoma, with at least 250,000 of these recorded in Africa

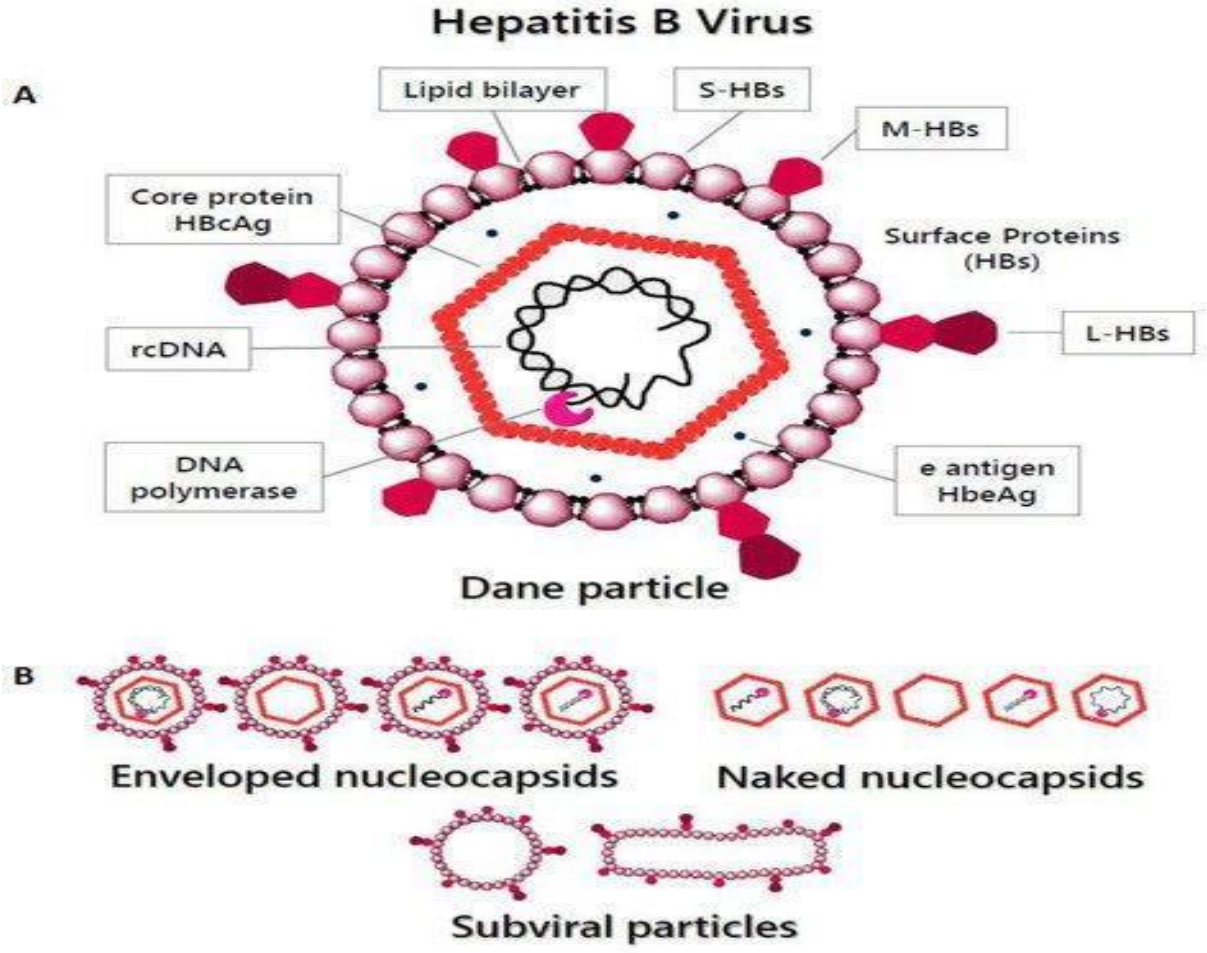


FIGURE 1 - Structure of HBV
FIGURE TAKEN FROM Venkatakrishnan, B. and Zlotnick, A. (2016). The Structural Biology of Hepatitis B Virus: Form and Function. Annual Review of Virology. https://dx.doi.org/10.1146%2Fannurev-virology-110615-042238

The unusual mechanism of HBV replication by reverse transcription and the lack of proof-reading ability of an RNA intermediate result in sequence heterogeneity. HBV is classified into at least 9 genotypes; A to I and a putative tenth genotype (J). Genotypes A–D, F, H and I are further classified into at least 35 subgenotypes
Most HBV genotypes and subgenotypes have a distinct geographical distribution. In sub-Saharan Africa, (comprising Eastern Africa, Central Africa, Southern Africa and Western Africa), HBV genotypes A, D and E circulate, with genotype A predominating in southern and eastern parts of the continent, while genotype D is found in the northern regions. West Africa is the only major region in the world where HBV is still hyperendemic—[> 8% of hepatitis B surface antigen (HBsAg) chronic carriers]. HBV/E, which was first described in 1992 ,is the predominant genotype prevailing in this region.

“Viral hepatitis," refers to infections that affect the liver and are caused by viruses. It is a major public health issue in the United States and worldwide. Not only does viral hepatitis carry a high morbidity, but it also stresses medical resources and can have severe economic consequences. The majority of all viral hepatitis cases are preventable. Viral hepatitis includes five distinct disease entities, which are caused by at least five different viruses. Hepatitis A and hepatitis B (infectious and serum hepatitis, respectively) are considered separate diseases and both can be diagnosed by a specific serologic test. Hepatitis C and E comprise a third category, each a distinct type, with Hepatitis C parenterally transmitted, and hepatitis E enterically transmitted. Hepatitis D, or delta hepatitis, is another distinct virus that is dependent upon hepatitis B infection.

There are four major serologic types of hepatitis B virus (adw, ayw, adr, and ayr), based on antigenic epitopes present on its envelope protiens.These serotypes are based on a common
determinant (a) and two mutually exclusive determinant pairs (d/y and w/r). The viral strains are also divided into ten genotypes (A-J) and forty subgenotypes according to the overall nucleotide sequence variation of the genome. The genotypes have a distinct geographical distribution and are used in tracing the evolution and transmission of viruses. The clinical significance of the four types remains unclear. Common to all of these subtypes, however, is an immuno-dominant epitope, the “a” determinant, that is the target of a neutralizing antibody in hepatitis B viral infection (anti-HBsAg). Recently, mutations in the “a” determinant have been reported to be associated with recurrence of hepatitis B viremia in serum despite the presence of protective antibodies (anti-HBsAg).

The virus is most commonly transmitted from mother to child during birth and delivery, as well as through contact with blood or other body fluids during sex with an infected partner, unsafe injections or exposures to sharp instruments, that means it is communicable disease.

REVIEW OF LITERATURE

STRUCTURE OF HBV

The infectious HBV virion (Dane particle) has a spherical, double-shelled structure 42 nm in diameter, consisting of a lipid envelope containing HBsAg that surrounds an inner nucleocapsid composed of hepatitis B core antigen (HBcAg) complexed with virally encoded polymerase and the viral DNA genome. The genome of HBV is a partially double-stranded circular DNA of about 3.2 kilobase (kb) pairs. The viral polymerase is covalently attached to the 5′ end of the minus strand. They have 3 different viral structures that are observed in serum of HBV infected persons by electron microscopy. The virus particle, called Dane particle (virion), consists of an outer lipid envelope and an icosahedral nucleocapsid core composed of protein. The nucleocapsid encloses the viral DNA and a DNA polymerase thats reverse transcriptase activity similar to retroviruses. The outer envelope contains embedded proteins which are involved in viral binding of, and entry into, susceptible cells. The virus is one of the smallest enveloped animal viruses with a virion diameter of 42 nm, but pleomorphic forms exist, including filamentous and spherical bodies lacking a core. These particles are not infectious and are composed of the lipid and protein that forms part of the surface of the virion, which is called the surface antigen (HBsAg), and is produced in excess during the life cycle of the virus.

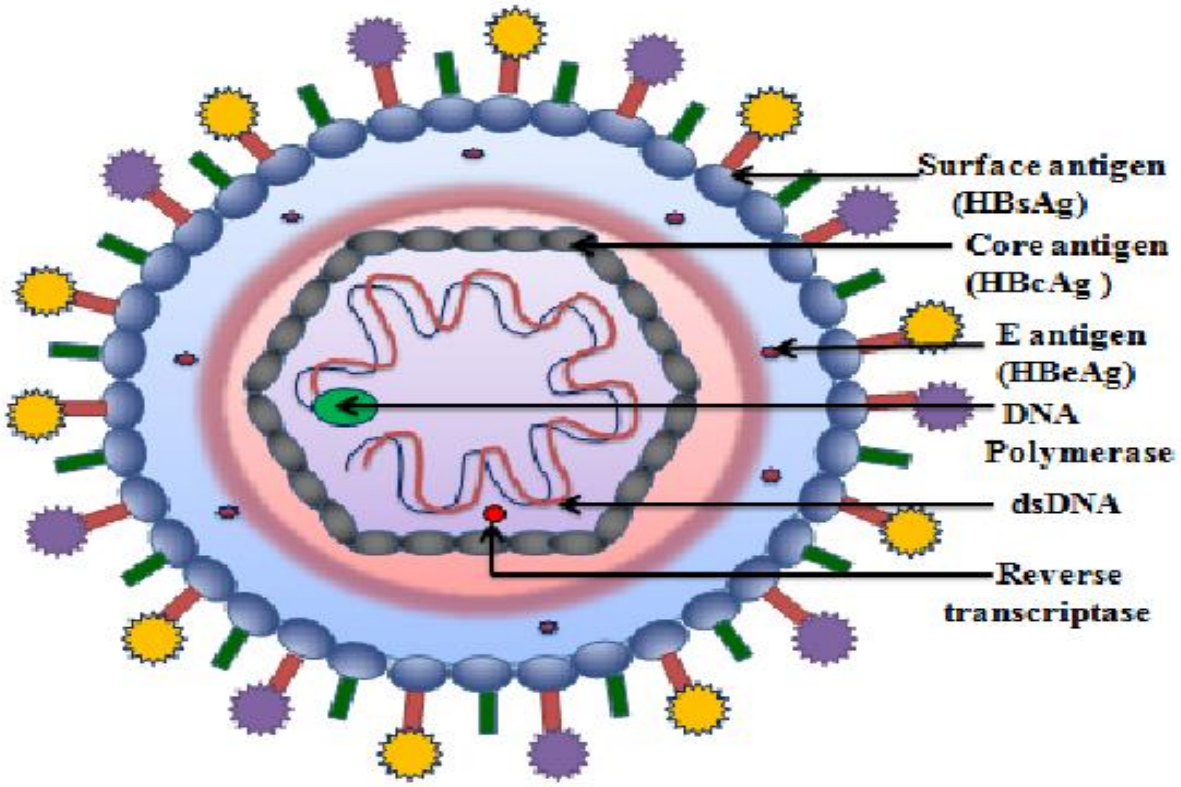


FIGURE 2- STRUCTURE OF HEPATITIS B VIRUS

FIGURE TAKEN FROM A brief history of hepatitis.
http://www.cevhap.org/index.php/en/ about-viral-hepatitis/a-brief-history-of-hepatitis.

The structural proteins of HBV

The envelope proteins

The HBV surface proteins are encoded by one open reading frame that is divided by three in-frame AUG start codons into the following domains: PreS1, PreS2 and S. The large HBV surface protein (LHBs) encompasses the PreS1 domain (108 or 119 aa depending on the genotype), the PreS2 domain (55aa) and the S domain (226 aa); the middle surface protein (MHBs) encompasses the PreS2 and S domain and the small (SHBs) consists of the S domain [1] .The S-domain, that is common to all three surface proteins, harbors at Asn-146 an N-glycosylation site, which is partially used in all three surface proteins [2]. Moreover, there is in the PreS2 domain at Asn-4 a glycosylation site that is used in MHBs, but not in LHBs

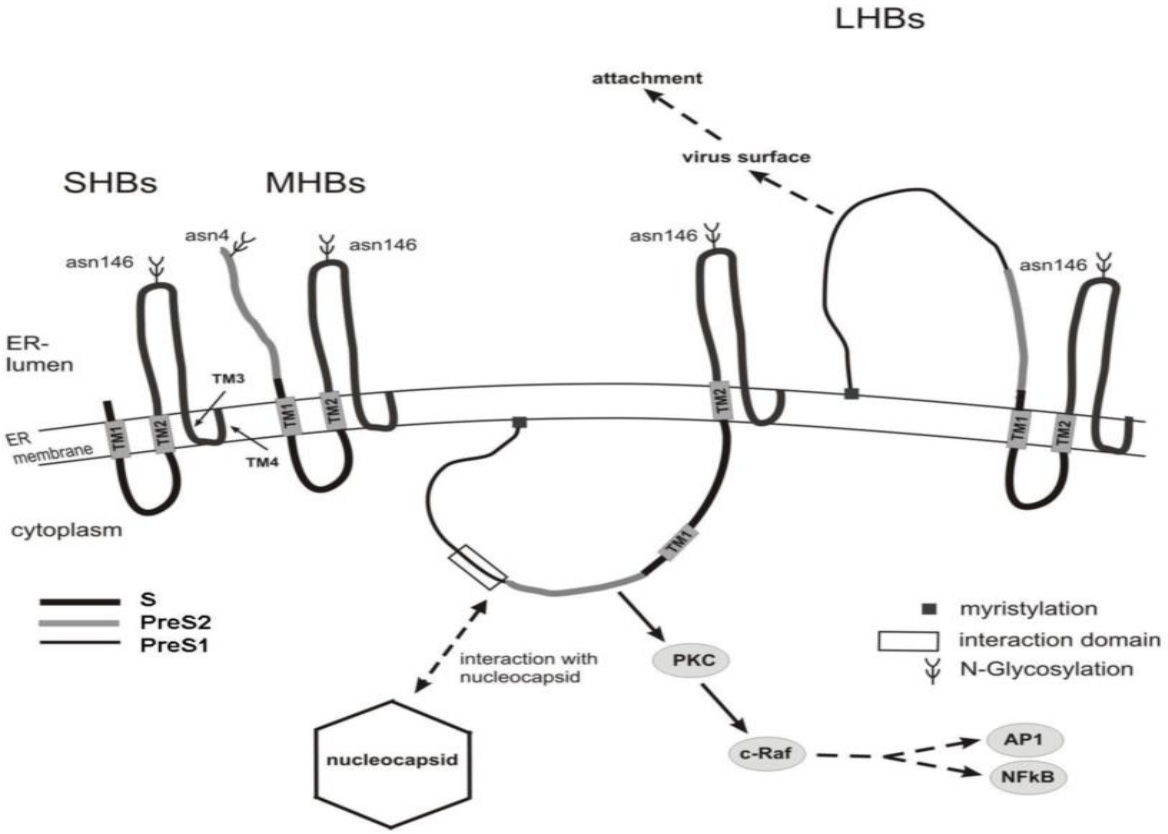


Figure 3. Schematic structure of the HBV surface proteins. The S-domain is common to all three HBV surface proteins. In the case of the large surface protein (LHBs), TM1 is not used as a start transfer signal, resulting in a cytoplasmic orientation of the PreS1PreS2 region. In a fraction of LHBs the PreSPreS2 domain is posttranslationally translocated across the ER membrane. In this case, the PreS1PreS2 domain faces the lumen of the ER. The two forms of LHBs fulfill different functions. This fraction that faces the ER lumen is exposed to the viral surface in the mature viral particle and is involved in the attachment process. The cytoplasmic form mediates the contact to the nucleocapsid and triggers intracellular signal transduction cascades by the interaction of the PreS2 domain with protein kinase C (PKC).FIGURE TAKEN FROM- Locarnini, S. Molecular virology of hepatitis B virus. Semin. Liver Dis. 2004,

HBV envelope proteins are integral membrane proteins that are anchored by the S-domain to the membrane. Membrane insertion of the S protein is initiated by an N-terminal signal sequence (aa 8-22) that is not cleaved and forms the first transmembrane region (TM1). In the mature protein, the aa 23-79 face the cytoplasm. At aa position 80-98 a second signal that forms TM2 directs the translocation of the growing chain though the ER membrane into the ER lumen, until at about aa 170 the hydrophobic C-terminus of the S-domain that is localized within the ER membrane starts. The detailed structure of the C-terminal region (aa 170-226) is not fully understood and is considered as one or two transmembrane regions (TM3/4). In the complete S protein, the N-terminus (aa 1-7) and the loop between aa 99 and aa 169 face the lumen of the ER, the domain between aa 23-79 face the cytoplasm. The luminal orientation of the loop between aa 99-169 enables the N-glycosylation of Asn-146 by the N-glycosyltransferases that are localized within the lumen of the ER. Moreover, this loop contains the major conformational epitope of the HBV surface antigen (HBsAg). After budding of the mature viral particle, these luminal domains are exposed on the external surface of the viral particle .

In the case of the MHBs protein there is no difference in the topology of its S-domain as compared to SHBs. The N-terminal PreS2 domain of MHBs is cotranslationally translocated into the ER-lumen resulting in the accessibility of Asn 4 to the N-glycosyltransferases. Therefore, MHBs is found in three forms: the unglycosylated (p30), the monoglycosylated (gp33) and the biglycosylated (gp36).

LHBs show an unusual biosynthesis. Its hydrophilic PreS1-PreS2 domain is not cotranslationally translocated, since TM1 of the S-domain is not used as a cotranslational signal sequence. As a consequence, the PreS1-PreS2 domain and a part of the S domain, up aa 79, remain on the cytosolic face of the ER. TM2 anchors the growing LHBs in the membrane and causes translocation of the downstream sequences into the ER lumen, enabling glycosylation at Asn-146 of the S-domain that still faces the ER-lumen. Due to this topology, the Asn-4 in the PreS2 domain is, in contrast to MHBs, not accessible to N-glycosylation. The glycine residue at aa 2 of the PreS1 domain in LHBs is myristylated. An interesting aspect is that in about 50% of the LHBs proteins a posttranslational translocation occurs, resulting in a dual membrane topology of the mature LHBs. As a consequence of this posttranslational translocation, the “unused " TM1 is integrated into the ER membrane and in this fraction the PreS1-PreS2 domain faces the ER lumen. There seems to be an equilibrium between the two topologies of LHBs, and recent reports suggested that chaperons are involved in the control of the posttranslational topological reorientation. With respect to the viral life cycle, the different topological forms of LHBs have different functions. In the mature secreted virion, the PreS1-PreS2 domain originally oriented to the ER lumen is exposed on the outer surface of the viral particle and is important for the virus-cell interaction. The form of the LHBs protein that faces the cytoplasm with its PreS1-PreS2 domain plays an essential role in virus morphogenesis by interacting with the nucleocapsid

Apart from these morphogenic functions, the cytoplasmic orientation of the PreS2 domain is associated with the additional function of PreS2 as a regulatory protein. Cytoplasmic PreS2 is found in LHBs and in C-terminally truncated MHBs proteins [3] that are encoded by 3´-end truncated pres/S sequences isolated from HBV-associated HCCs. In this conformation, the PreS2 domain binds to and activates PKC. The PreS2-dependent activation of PKC results in the activation of

the c-Raf/MEK signaling cascade that controls the expression of a variety of cellular and viral promoters [4]. PreS2-dependent activation and HBx-dependent activation share a variety of common features [5]. It has been shown that functionality of HBx- or PreS2-dependent activation is crucial for HBV replication. Regarding HBV expression, the selective knock-out of PreS2 or HBx can be compensated by the respective other unaffected HBV regulatory protein [6]. However, simultaneous knock-out of both PreS2 and HBx abolishes HBV replication [7].

The HBV surface proteins are not only part of the viral particle. They are suggested to bud from the post-ER pre-Golgi membranes [8] without envelopment of nucleocapsids into the lumen of vesicular structures and are finally released by secretion. These subviral particles have a diameter of 20 nm and an octahedral symmetry [9]. In addition, there are filaments of variable length. As compared to the virion, the subviral particles are highly overproduced. In the serum of HBV-infected patients, a 10,000-fold excess of subviral to viral particles can be found. The relevance of the subviral particles for the viral life cycle is not understood. These particles may interfere with the host immune system or support the infection process [10].

A detailed analysis revealed that the HBV subviral particles form by self-assembly of the S protein into branched filaments in the lumen of the ER. These long filaments are then folded and bridged for packing into crystal-like structures before they are transported by ER-derived vesicles to the ER-Golgi intermediate compartment (ERGIC). In the ERGIC, they are unpacked and relaxed. Due to their size, further progression through the secretory pathway might be limited. Therefore, their conversion into spherical particles is required. Small branched filaments can be formed by the L protein in the ER lumen, but these filaments are not packed into transport vesicles, accounting for the retention of the L protein within cells

The nucleocapsid

The HBV nucleocapsid is formed by the core protein (HBcAg) that is conserved between the different genotypes [9]. HBcAg encompasses 183 or 185 aa depending on the genotype, and the primary sequence of HBcAg can be divided into two parts: 1) The N-terminal 149 or 151 aa (depending on the genotype) are sufficient for the self-assembly of capsids. This part of the HBcAg is called the assembly domain. 2) The C-terminal 34 aa, designated the protamine domain, is rich in arginine residues that confer a positive charge to this domain. This domain is essential for the packaging of the pregenome / HBVPol complex [11].

HBcAg can be overproduced in pro- and eukaryotic expression systems and assembled in these systems to capsid particles [12]. Capsid assembly starts with the formation of HBcAg dimers that are crosslinked by a disulfide bridge between Cys-61 [13]. *In vitro* capsid assembly can proceed independent of cellular factors [14]. Crucial factors controlling assembly of purified HBcAg dimers are the concentration of HBcAg dimers and the buffer composition [15]. During the viral life cycle, capsid formation initiates with the binding of the Pol-complexed viral pregenome, but depends on cellular factors as chaperones and kinases [16]. Moreover, there are reports describing the incorporation of PKC in the HBV capsid [17]. A recent report established a correlation between PKC-dependent phosphorylation of the assembly

domain at Ser-106 and an increased level of assembled core. Moreover, in addition to the promotion of capsid assembly, the phosphorylation at Ser-106 seems to increase the stability of the assembled core, although the ratio of α-helical content was decreased in the capsid [18]. These factors might contribute to decreasing the threshold concentration that is required for the initiation of capsid formation.

In principle, HBcAg dimers can assemble into two different types of particles: on the one hand, particles with a diameter of 30 nm that consist of 90 HBcAg dimers and display a T= 3 symmetry, and on the other hand larger particles with a diameter of 34 nm, that are assembled by 120 dimers and display a T = 4 symmetry [19]. Although both particle forms can be found in HBcAg-producing systems, in infectious viral particles mainly the T = 4 capsids are found. However, in 10% of the virions T=3 capsids are found [20]. Based on cryo-electron microscopy [21] and crystallization of the capsid [22] details of the capsid structure were revealed. The most prominent features are surface spikes, flanked at either side by holes. Each spike can be arranged in one of two non-identical environments that are either part of two hexagons or of a hexagon and a pentagon. The spikes are formed by the core dimers. From each subunit, two anti-parallel α-helices, which are connected by a short loop in between aa 78-83, associate into a four-helix bundle [23]. The loop connecting the α-helices forms the spike tip and represents the major epitope of the capsid antigen. The assembled capsid is neither a tight shell nor a rigid inflexible particle. The conversion of the RNA pregenome into the DNA genome requires the import of nucleotides into the capsid lumen. The capsid shell contains pores with a diameter between 12 Å - 15 Å that enable the diffusion of small molecules into and out of the capsid lumen [24]. Based on the observation that foreign sequences can be inserted into the spike tip in between aa 78-81 without affecting the capacity to form capsids [25], it could be shown that the capsid structure is both highly stable and enormously flexible [26]. Induction of a conformational change from the outside is associated with a change in the internal capsid organization [27]. This conformational cross talk between capsid lumen and surface might be relevant for the control of capsid maturation during conversion of the RNA pregenome into the mature DNA genome [28].

Affecting nucleocapsid formation could be an alternative therapeutic approach to control HBV infection. The heteraryldihydropyrimidine (HAP) Bay 41-4109 recently was found to act in an HBcAg-specific manner and thereby inhibits virus production [29]. A more detailed analysis revealed that Bay 41-4109 can accelerate and misdirect capsid assembly [30]. Depending on the ratio of inhibitor molecule to HBcAg dimer, Bay 41-4109 can exert a stabilizing effect (up to a ratio of one inhibitor molecule per two dimers) or a destabilizing effect yielding large non-capsid HBcAg aggregates (Bay41-4109: dimer ratio of 1:1 or greater). From these data it is concluded that Bay41-4109 affects virus replication at low concentrations by induction of an inappropriate assembly of the capsid, and at higher concentrations by misdirecting the assembly from capsid formation to the formation of large HBcAg aggregates.

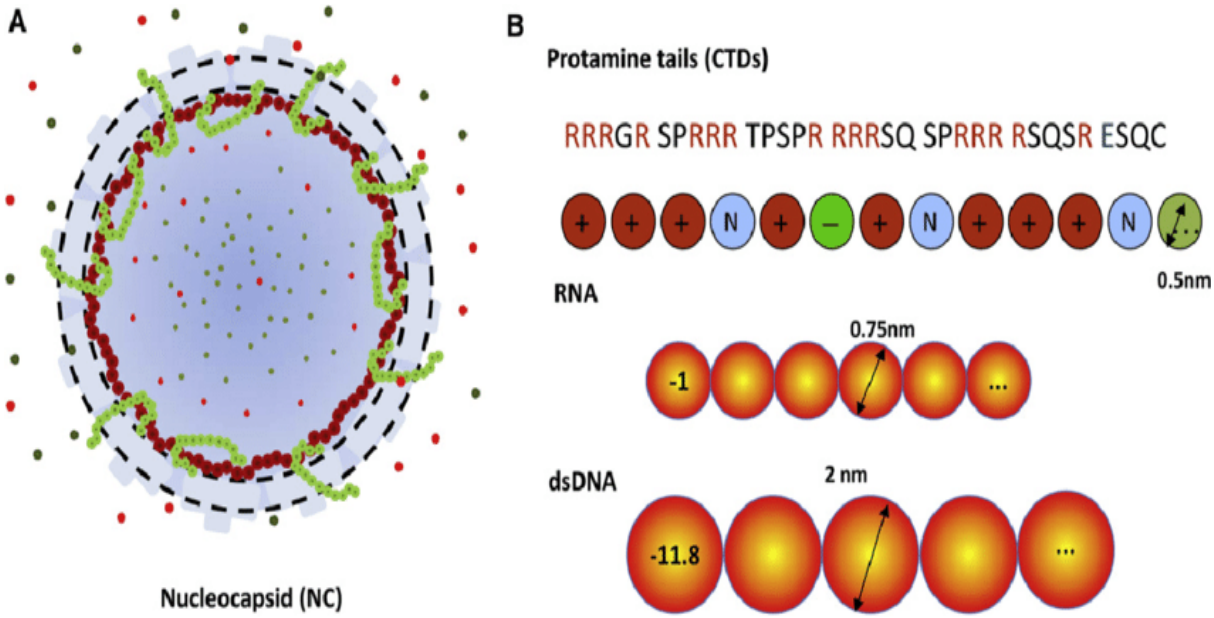


FIGURE 4-
(A) Schematic representation of the HBV nucleocapsid. The capsid shell is represented by a spherical shell (dashed lines) permeable to small ions (small dots) and the coarse-grained segments of the C-terminal domain but not to the segments of pgRNA or dsDNA. (B) Coarse-grained representation of pertinent biomolecules. Shown here are only portions of the coarse-grained chains of biomacromolecules.

TAKEN FROM- Bartenschlager, R., and H. Schaller. 1992. Hepadnaviral assembly isinitiated by polymerase binding to the encapsidation signal in the viralRNA genome. EMBO J

EPIDEMIOLOGY OF HEPATITIS B INFECTION

Hepatitis B is one of the most common infectious diseases in the world. Of the 2 billion individuals infected by hepatitis B virus (HBV), over 350 million are chronically infected, ([1]) with chronic infection manifested by persistence of the virus and HBV surface antigen (HBsAg) in serum and production of viral antigens and HBV DNA in the liver. Over a million individuals die annually of HBV-related chronic liver disease ([2]). Cirrhosis, liver failure, or hepatocellular carcinoma (HCC) develop in 15–40% of individuals with **chronic HBV infection ([3]).**

Figure 5 -. Global prevalence of hepatitis B virus infection. (From the Centers for Disease Control 2012.)
TAKEN FROM- Centers for Disease Control. The yellow book 2012. Chapter 3. Infectious diseases related to travel – hepatitis B. Available at www.nc.cdc.gov/ travel/yellowbook/2012/chapter-3-infectious-diseases-related-to-travel/ hepatitis-b [Accessed 28 July 2012]

The prevalence of chronic HBV infection varies widely in different parts of the world The majority of individuals in areas with a high prevalence of chronic HBV infection, where carrier rates are >5%, are infected at the time of birth or during early childhood ([2]). In areas of low prevalence such as Northern Europe, infection is typically acquired during adulthood via percutaneous or sexual transmission. In the United States, although prevalence of chronic HBV infection is low, around 1.2 million people are infected . Moreover, there are certain areas and populations in the United States, such as Alaskan natives, Pacific Islanders, and infants of first-generation immigrant mothers from highly endemic areas, where prevalence of HBV infection is high ([6]).

COMPONENTS OF HBV

It consists of:

- HBsAg (Hepatitis B surface antigen) was the first hepatitis B virus protein to be discovered. It consists of small (S), medium (M) and large (L) protein.
- HBcAg (Hepatitis B core antigen) is the main structural protein of HBV icosahedral nucleocapsid and it has function in replication of the virus. Capsid formation is the main factor for infection of the cell. HBcAg contributes to HBV clearance *in vivo*, but it is unknown whether HBcAg has to be in the capsid form to contribute to viral clearance.
- Hepatitis B virus DNA polymerase is incorporated into the nucleocapsid along with the pre-genomic RNA (pgRNA). Inside the capsid, the pgRNA

undergoes reverse transcription, making the (-) DNA strand. At the same time, most of the RNA template is degraded by the RNase activity of the polymerase. This is followed by (+) DNA strand synthesis, and the polymerase ends up covalently bound to the (-) DNA strand. The polymerase is discarded after the virion infects a new cell.

- HBeAg (Hepatitis B envelope antigen) can be found between the icosahedral nucleocapsid core and the lipid envelope, but is considered "nonparticulate" and is secreted and accumulates in serum. HBeAg and HBcAg are made from the same reading frame.
- HBx is small,[35] 154 amino acid long, nonstructural and has an important role in HBV-associated liver disease and in HBV replication in HepG2 cells. Many activities have been linked to the expression of HBx. However, the molecular mechanisms of many of these activities are unknown. This protein is multifunctional and it activates cellular signaling pathways and is essential for viral infection.

Hepatitis D virus requires HBV envelope particles to become virulent

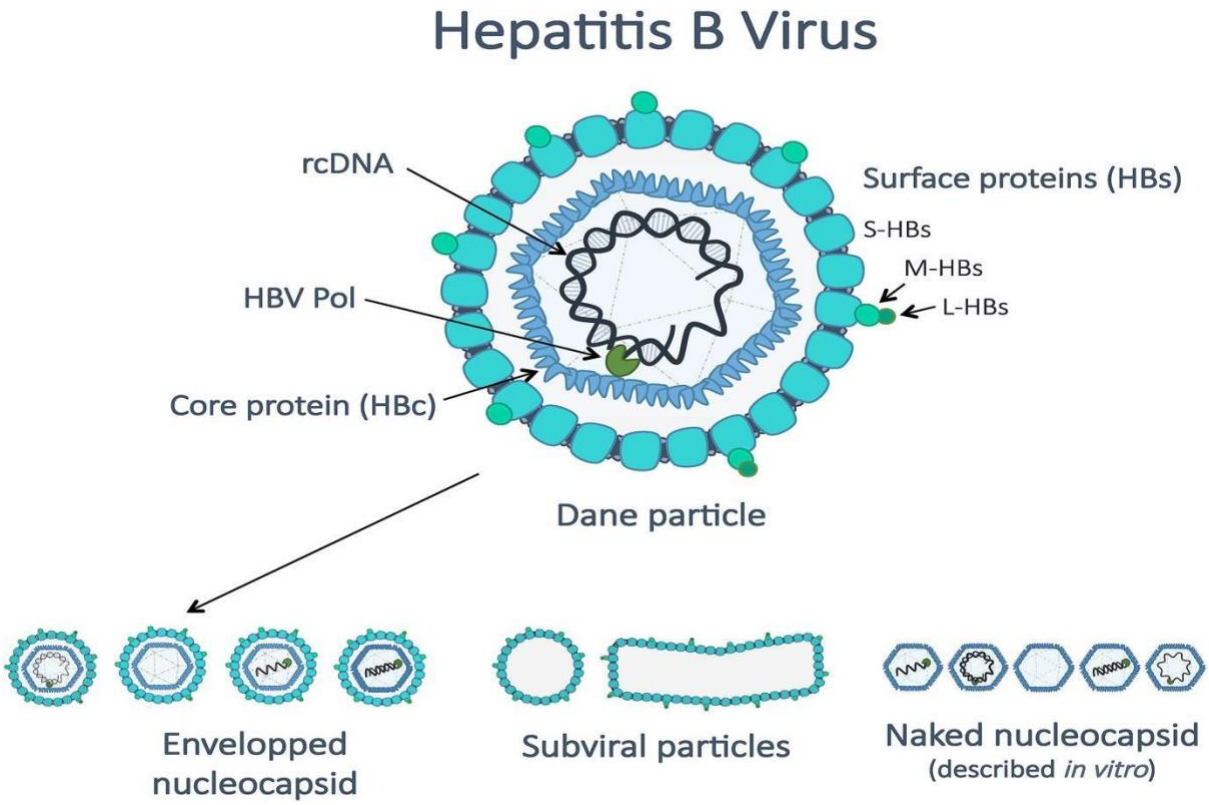


FIGURE 6- Schematic representation of HBV particles. Infectious HBV virion (Dane particle) (upper) and non-infectious HBV particles, including enveloped capsids containing immature DNA/RNA, subviral particles (sphere and filament), and naked nucleocapsids (lower).

TAKEN FROM-Bardens,A,Doring, Prange, R,2011.Alix regulates egress of hepatitis B virus naked capsid particles in ESCRT-independent manner

GENOME OF HBV

HBV belongs to the virus family *Hepadnaviridae* (infecting different avian and mammalian hosts), which includes several genera of partially double stranded DNA genome of approximately 3.2 kb length, generated through reverse transcription from a longer intermediate RNA (approximately 3.5 kb, generally referred to as pregenomic RNA or pgRNA) [31]. The HBV genome encodes four partially overlapped open reading frames (ORF): the *surface (preS1, preS2, S)*, *core (precore, core)*, *polymerase* and the *'x'* genes respectively. High genetic variability is a characteristic feature of the HBV as the viral polymerase lacks proofreading activity and uses an RNA intermediate during its replication [32]. On the other hand, the extreme overlapping of the open reading frames of the HBV genome limits the possibility of fixation of all these mutations [33]. These opposite aspects render the substitution rate of HBV to an intermediate level between RNA and DNA viruses.

Such a replication system makes random errors during genomic replication, which are the source of genetic variation, upon which natural selection can act, leading to evolution of the HBV genome [34]. The nucleotide substitution rate, for HBV has been estimated to be $1.4 – 5.0 \times 10^{-5}$ per site per year, being 10 fold superior than other DNA viruses, but the rate of synonymous (silent) substitutions is higher than the rate of non-synonymous substitutions, suggestive of a constrained evolution of the HBV genome [35]. In contrast, in a liver transplantation setting, the mutation rate has been found to be almost 100-fold higher [36]while mutation rate is negligible in silent or occult HBV infection, where there is minimal host response over many decades [37]. However, Hannoun et al., [38]calculated a mean frequency of fixation of nucleotide substitution of a wider range ($2.1–25 \times 10^{-5}$ nucleotide change per site per year) depending on the HBeAg/anti-HBeAg status of the host. Thus it appears that host-virus interaction and immune selective pressures, imposed by the host immune system, either naturally or medically, can affect the variability of the HBV genome.

Random errors/variations in the HBV genome, occurring due to long periods of persistence and immune selection pressures operating at the population level have led to the emergence of distinct genotypes and their subgenotypes in specific geo-ethnic populations, and being transmission competent these variants stably circulate within the given geo-ethnic population [39]. In addition, certain mutations may also emerge under medical pressures (vaccine, or antiviral therapy), which are selected at the individual level. During specific phases of chronic HBV infection, mutations (e.g. 587^{A}, 1896^{A}, $1762^{T}/1764^{A}$ etc.) emerge that are advantageous for escaping the natural or therapy induced antiviral immune pressure and thus favours viral persistence.

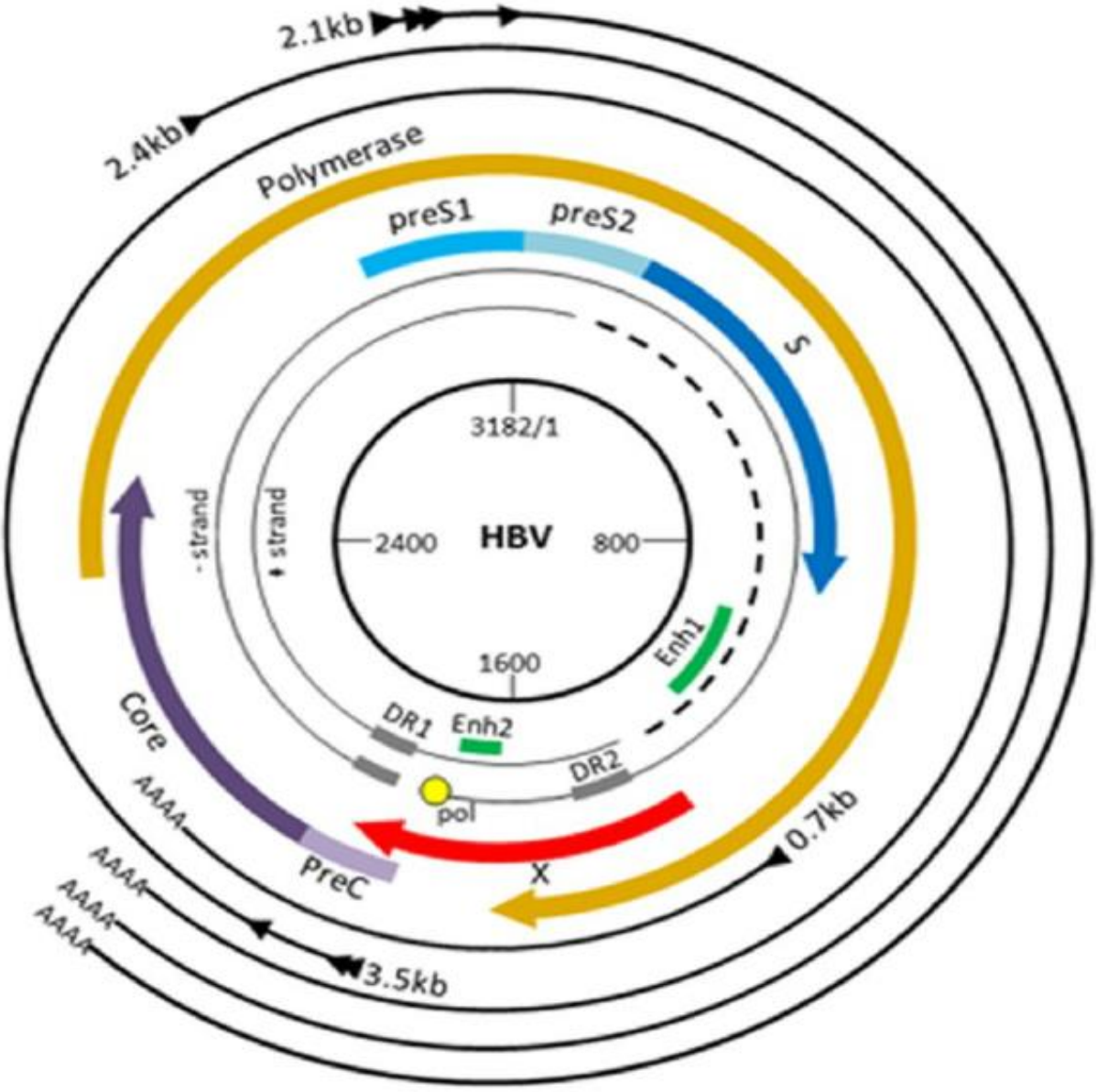


FIGURE 7-HBV genome organization. HBV contains a small, partially double-stranded (~dsDNA) genome (see inner black circles) that consists of a full-length negative strand and an incomplete (dashed lines) positive strand. The genome contains four promoters, two enhancer regions (Enh1, Enh2), and two direct repeats (DR1, DR2). The four ORFs are depicted by the colored arrows. During virus replication, the ~dsDNA genome is repaired into covalently closed circular (ccc) DNA, which serves as the template for viral transcription. The four major transcripts, shown as thin black outer arrows, are described in the text. Note that the X ORF (red) is present in all four HBV mRNAs

TAKEN FROM - Hepatitis B Virus HBx Protein Interactions with the Ubiquitin Proteasome System - Scientific Figure on ResearchGate. Available from: https://www.researchgate.net/figure/HBV-genome-organization-HBV-contains-a-small-partially-double-stranded-dsDNA-genome_fig2_268880855 [accessed 20 Jun, 2022]

HBV genetic diversity: genotypes & subgenotypes

Classically, HBV strains were distinguished by the presence of two pairs of mutually exclusive serotype determinants '*d*'/'*y*' and '*w*'/'*r*', in the HBsAg along with the main antigenic determinant 'a', which led to the description of 4 serotypes, namely *adw, adr, ayw* or *ayr*. Additional serotypes were subsequently characterized leading to the description of nine serotypes namely *ayw1, ayw2, ayw3, ayw4, ayr, adw2, adw4, adrq+* and *adrq-* and a distinct geographical pattern for the distribution of serotypes was also documented [40]. However, with the advent of molecular biological techniques and advanced computational methods for the phylogenetic analysis of complete viral genome sequences, HBV genotypes and subgenotypes have been described, which have largely replaced the classical serotype based classification of HBV strains.

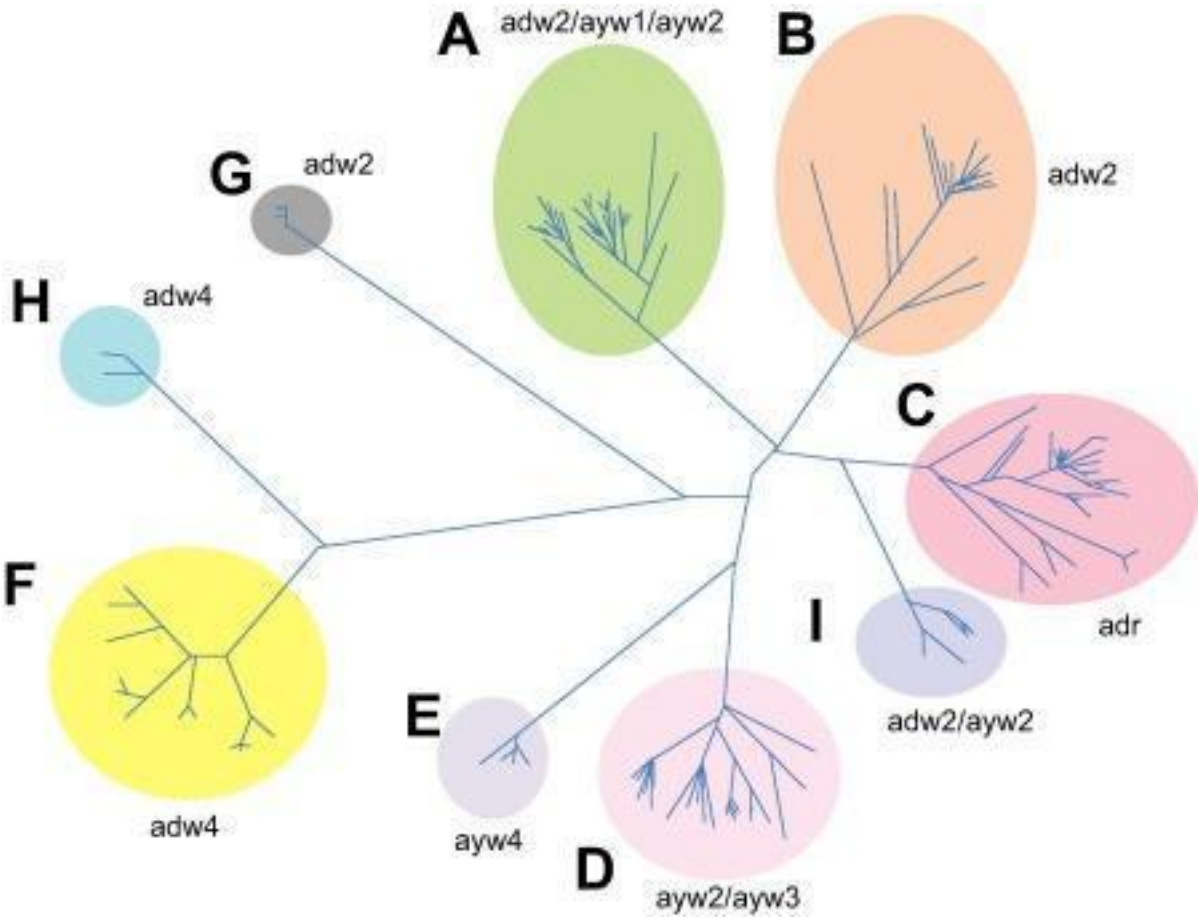


FIGURE 8- Hepatitis B virus genomic variability: an unrooted phylogenetic tree representative of the nine major genotypes (A–I). Based on HBsAg heterogeneity, ten serological subtypes have been identified: ayw1, ayw2, ayw3, ayw4, ayr, adw2, adw4, adwq, adrq+ and adrq-.

Based on more than 8% genetic variability among HBV strains found worldwide, eight HBV genotypes namely A, B, C, D, E, F, G, and H have been well established [41-43]. Further extensive phylogenetic analyses of the HBV genotypes have resulted in recognition of subgenotypes of genotypes A, B, C, D and F, based on more than 4% intra-genotypic divergence. Until now, the presence of 5 sub genotypes have been recognized for each of the HBV genotypes A, B, C and D, while 4 sub genotypes have been well reported for genotype F [44]. Having evolved distinctly in specific geo-ethnic populations, HBV genotypes/subgenotypes have a distinct geographical distribution pattern

TRANSMISSION OF HBV AND ITS PREVENTION

HBV is present in blood, saliva, semen, vaginal secretions, and menstrual blood of infected individuals. Because HBV is resistant to breakdown outside the body, it is easily transmitted through contact with infected bodily fluids ([2]). Perinatal vertical transmission is the most common mode of transmission worldwide. Presence of HBV e antigen (HBeAg) in the mother's serum is associated with greater infectivity: ([6, 7]) The risk of perinatal HBV infection among infants born to HBV-infected mothers ranges from 10–40% in HBeAg-negative mothers to 70–90% in HBeAg-positive mothers ([8]). Children of HBsAg-positive mothers who do not become infected perinatally remain at high risk of infection during early childhood ([8]). In households of a chronically infected individual, HBV infection can occur via person-to-person, nonsexual contact ([8]).

Immunization is the most effective means of preventing HBV infection; vaccination offers >95% protection against the development of chronic infection ([2, 6]). The failure of immunization programs targeted at individuals with risk factors has led to the recommendation that hepatitis B vaccination should be included in routine vaccination schedules for infants

LIFE CYCLE OF HBV

The life cycle of *Hepatitis B virus* is complex. Hepatitis B is one of a few known non-retroviral viruses which use reverse transcription as a part of its replication process.

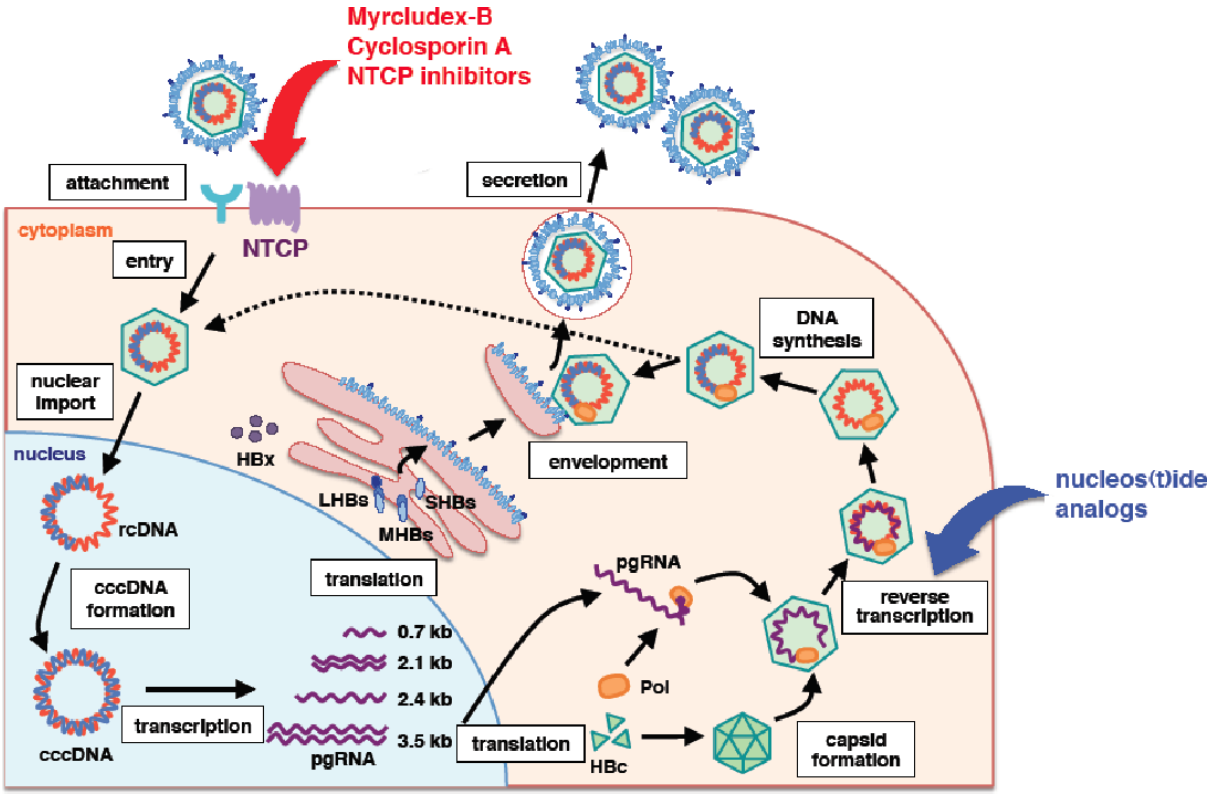


FIGURE 9- Schematic representation of the hepatitis B virus (HBV) lifecycle. Nucleos(t)ide analogs inhibit reverse transcription. Myrcludex-B, cyclosporin A and some NTCP inhibitors can inhibit the viral entry process by targeting NTCP.

1. Attachment

According to a general concept of viral infection, the first step is an energy-independent attachment of the viral particle to a structure at the host cell surface. After the primary attachment, which is characterized by low affinity and reversibility, the virus particle is transferred to a more specific receptor. In case of enveloped viruses the binding to the receptor normally is followed by a fusion step, either at the plasma membrane or in an endosomal compartment [45]. Many aspects of the initial steps in the HBV life cycle at the present are still enigmatic. However, based on the establishment of different *in vitro* infection symptoms significant progress has been achieved recently.

2. Experimental systems

Primary human hepatocytes (PHH) are a classic *in vitro* infection system to study HBV infection [46-48]. The major disadvantage of these cells is their limited

availability and their heterogeneous quality, varying from donor to donor. Moreover, there is a low infection efficiency leading to only a few percent of infected cells [49].

Primary hepatocytes from *Tupaia belangeri* (PTH) represent an alternative cell culture system [50]. PTHs are more readily available and there is less variability between different preparations. The infectability of PTHs is comparable to that of PHHs. However, in contrast to PHH, PTH can be infected by woolly monkey HBV in addition to HBV [51,52]. A drawback of using PTHs as an infection model, however, is that this cell culture system is less characterized and the cross-reactivity of many antisera specific for mouse or human targets with the corresponding *Tupaia* protein so far is not clear.

Based on this, the availability of a human hepatoma cell line that can be infected with HBV is desirable. For the human hepatoma cell line HepG2 there are many reports describing a specific binding and uptake of HBV [53], however, there are only two reports of infection of HepG2 that were cultivated in the presence of DMSO and 5-aza-2`deoxycytidine [54]. A new hepatoma cell line, HepaRG, established from a female HCV-positive patient with an HCC was reported to be susceptible to HBV infection after differentiation in DMSO and hydrocortisone [55]and to enable reproducible infection by HBV.

A general problem in the analysis of productive infection is to differentiate between input and *de novo* synthesized viral or subviral particles. An unequivocal marker for productive HBV infection is the formation of covalently closed circular (ccc) DNA (for a recent review see that can be detected by Southern blotting or by a real-time PCR approach for selective amplification of cccDNA, but not of the other viral DNAs.The cccDNA amplification can be detected for infected PTH and PHH, but in the case of HepaRG cells there seems to be no amplification of cccDNA. Detection of the viral mRNAs provides an additional approach to discriminate between input and *de novo* synthesis.

When using enriched or purified viral inocula, detection of HBeAg by ELISA represents a good marker, since after the enrichment procedure HBeAg is not present in the inoculum. HBsAg detection is more sensitive as compared to HBeAg, it requires, however, multiple cell washing steps.

3. Virus-cell interaction

HBV infection is thought to follow a multistep process. While for the DHBV system heparin or dextran sulphate have no effect on the infection process,in the case of HBV there are reports about the relevance of the initial attachment to the carbohydrate side chains of hepatocyte-associated heparan sulphate proteoglycans as attachment receptors for HBV infection.This interaction is suggested to initiate the multistep entry process of HBV and is followed by yet unknown high-affinity step(s) mediating HBV uptake. Identification of the “HBV receptor” or of HBV binding partners is one of the challenging open questions in the field of HBV biology. There is a constantly growing list of proteins that were found to bind to HBV, but for none of these potential binding factors is there convincing evidence of its essential relevance for the infection process (for a detailed list see.While the cellular structures that mediate viral binding and entry are less understood, more is known about the viral structures involved in binding and entry. A milestone in the characterization of viral prerequisites for the binding to hepatocytes was the observation of Neurath *et al.* in 1986. Neurath and colleagues reported that a short fragment of the surface protein encompassing aa 21-47 of the PreS1 domain (this corresponds to aa 10-36 in

genotype D, E and G) binds to HepG2 cells and completes the binding of HBV to these cells. Consistent with this finding, it has been observed that aa 3-77 of HBV PreS1 are crucial for infectivity. Paran *et al.* identified a QLDAPF sequence motif (corresponding to aa 18-25) as an essential domain for HBV binding. A further structural prerequisite for the infectivity of HBV is the myristoylation at glycine 2 of the PreS1 [56]. It was observed that a myristoylated peptide PreS1 domain showed significantly stronger binding to HepG2 cell-derived membranes as compared to the non-myristoylated PreS1-domain [57]. Detailed studies revealed that acylated peptides encompassing the N-terminal part of the PreS1 domain efficiently inhibit HBV and HDV infection [58-60]. One essential parameter is the hydrophobicity of the N-terminal acyl residue. The increase in the inhibitory potential is associated with an increase in the chain length of the fatty acid. A pentanoyl group is less efficient than a decyl group, which is less efficient than a myristoyl group (C14). Regarding the peptide sequence, it was revealed that residues 1-8 and 19-28 are dispensable for the inhibitory effect; residues aa 9-18, however, are crucial. In accordance with this, recombinant HBVs mutated between aa 9-18 of the PreS1 domain are not infectious.

The mechanism by which these peptides exert their inhibitory effect is presently unclear. The very low IC50 of 8 nM and the observation that these peptides are inhibitory even if they are added after the virus attachment has occurred argue against a simple competitive inhibition. Moreover, there is no clear correlation between the specificity of peptide binding, on the one hand, and susceptibility of the respective cell for HBV infection on the other hand. Analysis of the bio-distribution of these peptides in immunodeficient urokinase-type plasminogen activator (uPA) mice, repopulated with primary human or *Tupaia belangeri* hepatocytes, demonstrates an accumulation of the acetylated peptides in the liver, but no preferential binding to the implanted PTHs or PHHs [61]. Based on this, it can be speculated that these peptides inhibit viral infection by interfering with signal transduction cascades that regulate HBV infection or with early post-entry steps.

Although the PreS1 domain contains the major cell attachment epitope, there are reports about further epitopes outside the PreS1 domain that are involved in HBV-cell attachment. Paran *et al.* describe the existence of a secondary attachment site in the S domain. Moreover, antibodies recognizing epitopes within the PreS2 domain or the S domain were found to inhibit infection [62]. However, it is unclear whether these antibodies act by directly masking an essential sequence for HBV-cell attachment, or whether their binding acts as a spacer, preventing the close contact between the virus and the cell surface. Moreover, the interference with a post-entry step is possible.

Data from the DHBV system suggest that *hepadnaviridae* are internalized by an endocytotic step[63]. It was observed that DHBV particles colocalize with fluorophore-labelled transferrin in the endosomal compartment. Moreover, it has been demonstrated that bafilomycin A1, which inhibits vacuolar proton ATPases, impairs infection.

4. Entry and release of the nucleocapsid into the cytoplasm

Productive infection with HBV requires delivery of the genome into the nucleus [64]. Resulting questions concern the entry of the virus/nucleocapsid into the cell and the subsequent transport of the genome towards the nucleus. In contrast to viruses harboring type I fusion proteins on their surface, HBV does not possess a classic fusion peptide sequence. In a recent report, a fusogenic function was ascribed to the

PreS1 domain of HBV. Based on sequence analysis, it was suggested by Rodriguez-Crespo *et al.* that the N-terminus of the S-domain (aa 1-23), including the first transmembrane region (TM1), might act as a fusogenic sequence. This hypothesis is supported by the observation that a chimeric fusion protein of influenza virus, hemagglutinin, with the sequence aa 7-18 of the S domain, showed significant hemifusion activity. Moreover, it was demonstrated that DHBV subviral particles upon low pH treatment expose hydrophobic domains on their surface that could mediate membrane contact [65]. Further analysis revealed that a decrease in the hydrophobicity of the TM-1 domain in DHBV L protein but not in S-protein resulted in a loss of infectivity. Moreover, *in vitro* experiments with synthetic peptides corresponding to TM1 indicate that the hydrophobicity of TM1 is required for aggregation and lipid mixing of phospholipid vesicles [66]. Although these data suggest that TM1 could act as a fusogenic sequence, so far there is no direct experimental evidence for fusion to host cell membranes during HBV entry.

In the case of genotype ayw, the PreS2 domain of HBV harbors between aa 41-52 a membrane-permeable peptide designated TLM (translocation motif). The presence of this TLM is conserved in all *hepadnaviridae*. The TLM belongs to the family of membrane-permeable peptides. Fusion of the TLM to other peptides or proteins enables their energy and receptor-independent translocation across cellular membranes into the cytoplasm. The functionality of the TLM as a membrane-permeable peptide depends on a defined pattern of hydrophilic and hydrophobic amino acids that form a labile amphipathic alpha helix. Fusion of the TLM to HBcAg revealed that fully assembled nucleocapsids that are decorated on their surface with TLM-peptides are able to translocate across cellular membranes and deliver the packaged nucleic acid to the nucleus. These data demonstrate that even particles, if they bear TLM peptides on their surface, are able to translocate across cellular membranes. Based on this, it was analyzed whether the TLM peptide could play a role in the HBV entry process using the DHBV system. In contrast to HBV, DHBV harbors two TLMs in the PreS domain. Infection experiments revealed that destruction of the TLMs, or even of one TLM, abolished infectivity. More detailed analysis revealed that TLM-deficient DHBV particles still bind to the cell and are able to enter the endocytotic pathway, but the TLM deficient mutants accumulate in an endosomal compartment. Further experiments revealed that in the endosomal compartment a proteolytic processing of the internalized viral particle occurs, resulting in an unmasking of the TLM peptide. It was concluded that, due to the endosomal proteolytic processing, unmasked TLM enabled the translocation across the endosomal membrane into the cytoplasm, where the proteolytically processed envelope dissociates from the nucleocapsid. The capacity of viral particles that have been proteolytically processed by preincubation with endosomal lysate to translocate across cellular membranes has been shown for HBV and DHBV. Incubation of HepG2 cells with proteolytically processed HBV or LMH resulted in a productive infection of these cells, which are not permissive to infection with the unprocessed virus..

Based on these data, it was suggested that *hepadnaviridae* could deliver their nucleocapsid into the cytoplasm not by a fusion process, but by a novel mechanism that is based on membrane translocation. However, there are publications arguing against this model: HDV entry does not depend on the functionality of a TLM, and mutation or deletion of the TLM in HBV seems not to affect its infectivity. Recent data, however, raise the question whether HDV is a suitable model system to study HBV entry. While chimeric particles harboring woodchuck envelope proteins are

unable to infect PHHs, a recombinant HDV assembled with envelope proteins of WHV infects PHHs, indicating significant differences between the entry process of HDV and HBV. Detailed analysis of the TLM-mutated HBV reveals that due to the partial deletion of either the C-terminal or N-terminal part of the TLM a novel functional TLM was generated. However, destruction of the TLM by point mutations that convert the structure to a stable beta sheet resulted in a complete loss of infectivity using PHH (E. Hildt's unpublished results).

3.5. Import of the genome into the nucleus

Lipofection of mature nucleocapsid or transfection of primary human hepatocytes or hepatoma cells with membrane-permeable nucleocapsids indicates that the nucleocapsid moves by a directed transport towards the nucleus. This can be deduced from the kinetics of intracellular trafficking [66]. A perinuclear accumulation after delivery of the nucleocapsid into the cytoplasm can be observed within 15 min, while a diffusion-based process would take over 1h. A central role for the intracellular trafficking of the nucleocapsids is ascribed to the microtubule system. The controversial issue of the relevance of actin filaments for intracellular nucleocapsid transport has been discussed.

Productive viral infection requires the transport of the HBV genome into the nucleus, where the conversion into cccDNA occurs (recently reviewed in. It is questionable whether the import of the viral genome into the nucleus occurs in association with HBcAg or not. The limited efficiency of the available infection systems, as well as the very small amounts of nucleocapsids which are finally released into the cytoplasm, make it difficult to address this question. One approach to investigate this is based on digitonin-permeabilized cells, which are subsequently exposed to nucleocapsids. Based on this experimental system, a phosphorylation-dependent binding of the core particle to the nuclear pore complex was observed. According to their previous observation that PKC can be encapsidated into the core particle, the authors assume that the encapsidated PKC phosphorylates C-terminal Ser residues in the core protein giving rise to mature phosphorylated progeny core particles. However, this appears to be in contrast to more recent observations that correlate HBV capsid maturation with stepwise dephosphorylation.

Further work of this group suggested that immature capsids reached the basket of the nuclear pore complex, but neither released capsid proteins nor immature genomes into the nucleoplasm. In the case of mature capsids, intranuclear staining for HBcAg was observed. However, the digitonin permeabilization procedure affects the integrity of the cell. The permeabilization kills the cell and causes loss of cellular proteins. Traces of digitonin might affect the stability of the nucleocapsids. Moreover, the chosen antiserum (Dako HBcAg) detects HBcAg dimers as well as fully assembled particles, and, therefore, does not allow the conclusion that assembled particles have translocated into the nucleus. Electron microscopy data after microinjection of nucleocapsids into *Xenopus laevis* oocytes, however, demonstrate that in this system the capsid passed the nuclear pore and entered the nuclear basket. However, even immature nucleocapsids were found to enter the nuclear basket. It is assumed that only disaggregation of the mature nucleocapsid can occur, resulting in the release of the polymerase-linked viral genome into the nuclear basket.

In a recent report, an efficient system for gene transfer into hepatocytes based on cell-permeable nucleocapsids was described. The cell permeability of the nucleocapsids was achieved by fusion of the TLM peptide to the HBcAg. Dimers of TLM-HBcAg assemble into the icosahedral capsid. This peptide enables the

receptor-independent translocation of cargo (proteins or peptides that are fused to the TLM) across the plasma membrane without affecting the integrity of the cell. These TLM-nucleocapsids translocate as fully assembled particles across the plasma membrane without affecting the cellular integrity. Finally, the HBV genome, or its derivative, packaged into these TLM nucleocapsids, is efficiently expressed, indicating that a productive trafficking ending with expression of the packaged genome occurs [69]. Using this system and an antibody that selectively recognizes fully assembled nucleocapsids (mab 3120) no evidence for nuclear import of nucleocapsids was obtained, but a perinuclear accumulation could be observed. If the disassembly of the nucleocapsid does not occur within the nucleus, the viral genome that is linked to the polymerase, and therefore is too big for a free diffusion through the nuclear pore complex, must be transported actively into the nucleus. One possibility could be the association to HBcAg dimers that possess a nucleic acid binding domain and an NLS sequence. Another possibility could be that the polymerase mediates the final import of the viral genome. This is supported by the observation that the Pol-DNA complex is efficiently imported into the nucleus. Deproteinization of the viral genome, however, caused retention in the cytoplasm. Recently it was revealed that the TP-domain of HBV polymerase harbors a functional bipartite nuclear localization signal that is crucial for the HBV infectivity.

5. Replication

5.1. rcDNA to cccDNA conversion

At the end of the viral entry process, the viral genome is delivered into the nucleus. The viral genome exists at this stage as rcDNA (relaxed circular DNA). rcDNA consists of a complete (-)-DNA strand covalently linked to the viral polymerase P at its 5' end, and an incomplete (+)-DNA strand with an RNA oligonucleotide at its 5' end, which serves as primer for the (+)-strand synthesis. To establish a viral infection, the viral genome has to be present in a stable form within the infected cell. In the case of HBV, the viral rcDNA is converted into a nuclear, episomal covalently closed circular DNA (cccDNA), which represents the central intracellular intermediate in viral replication and also serves as an experimental marker for the successful establishment of an infection. For both genome amplification and cccDNA formation, the shorter (+)-DNA strand has to be completed, both strands need to be covalently ligated and the obstructive terminal modifications must be removed. The mechanism of how the viral polymerase and RNA primer is removed from the (-)-DNA strand and the (+)-DNA strand, respectively, is still not fully understood. Two independent studies identified a protein-free rcDNA containing the identical nucleotide sequence as the encapsidated rcDNA, but the polymerase is not bound anymore to the (-)-DNA strand and might be an intermediate during cccDNA formation. Infection experiments with primary *Tupaia* hepatocytes showed that blocking reverse transcriptase activity of the viral polymerase strongly reduces cccDNA formation. Furthermore, recent *in vitro* experiments revealed that DDX3 DEAD-Box RNA helicase is incorporated into nucleocapsids inhibiting reverse transcription, which further leads to a reduced level of double-linearized DNA (dlDNA) [69]. Taken together, these findings suggest a role of the P protein in completion of the (+)-DNA strand. However, the detailed process of cccDNA generation still remains unclear and needs to be further investigated.

5.2. pgRNA transcription from cccDNA

The nuclear cccDNA serves as template for the synthesis of the pregenomic RNA (pgRNA), an RNA intermediate for viral replication and other subgenomic RNAs. The bicistronic pgRNA has two major roles in the viral life cycle: it serves as translation and reverse transcription template. The pgRNA is an overlength transcript containing a second copy of the direct repeat 1 (DR1), the ε signal and a poly-A tail, serving as a transcript for the translation of the 90 kDa viral polymerase, the 21 kDa core protein, and a 24 kDa precursor early antigen. Second, it serves as a template for reverse transcription of the viral (-)-DNA strand and is, therefore, indispensable for viral replication. Besides the pgRNA, there are three additional subgenomic RNAs coding for the surface proteins (2.4 kb RNA and 2.1 kb RNA) and the HBx protein (0.7 kb RNA). Transcription of all hepadnaviral RNAs is processed by host cell polymerase II. Besides the wt RNA, there are splicing variants being translated into hepatitis B splice-generated proteins and encapsidated into defective viral particles.

5.3. Reverse Transcription

Hepatitis B viruses, as well as other members of the *hepadnaviridae*, use pgRNA as replication intermediate for reverse transcription. First, the pgRNA-polymerase complex is packed in the lumen of assembling capsids, whereas the viral polymerase binds to the encapsidation signal ε which is a cis-element on the pgRNA. How the interaction of ε and the polymerase takes place is currently not understood in detail, but it is supposed to play a role in the recruitment of core protein homodimers, which finally leads to capsid formation through self-assembly. Besides initiation of pgRNA-polymerase encapsidation, the ε-polymerase interaction also induces reverse transcription, whereby first the (-)-DNA strand is synthesized, followed by (+)-DNA strand generation finally leading to rcDNA. A protein-priming mechanism is crucial for the start of DNA synthesis. A short DNA oligonucleotide covalently linked to the polymerase binds to ε and initiates (-)-DNA strand synthesis. Besides these two factors, the core protein is also essential for reverse transcription, as concluded from several reports which could clearly demonstrate that deletions or modification of the C-terminal assembly region of the core protein lead to defective DNA synthesis. The mature secreted HBV virions have completed reverse transcription, which takes place in intact nucleocapsids and contain only DNA. After the DNA genome is synthesized, the nucleocapsid can either continue with the viral life cycle and interact with envelope proteins and get secreted as infectious virions or they can re-deliver their rcDNA to the nucleus and build up a cccDNA pool within the nucleus.

Morphogenesis

Capsid maturation

HBV nucleocapsid formation starts when the complex of the RNA pregenome, HBV polymerase and HBcAg dimers has formed. When nucleocapsid assembly is

completed, the conversion of the RNA into single-stranded and then into partially double-stranded DNA takes place (reviewed in detail in. In contrast to the nucleocapsids isolated from secreted viruses that contain only mature partially double-stranded DNA, intracellular nucleocapsids show all these different stages of the viral DNA synthesis. Based on these observations, it was concluded that the early RNA-containing capsids (the immature nucleocapsid) are not incorporated into viral particles. The resulting questions are whether there are changes in the capsid structure associated with capsid maturation that enable discrimination between the immature and the mature nucleocapsid. For DHBV and HBV, it was reported that mutations in Pol that destroyed the reverse transcriptase activity resulted in an accumulation of immature nucleocapsids that are not enveloped [70]. Further experiments based on the DHBV system demonstrated that the envelopment of the nucleocapsid occurs at a late stage of the replication cycle [71].

Detailed analyses revealed that capsid maturation is associated with a dephosphorylation of the nucleocapsid. Phosphorylation is required for efficient RNA packaging. In the case of HBV, it has been shown that three Ser-Pro-residues (Ser 155, 162, 170) that are located in the C-terminal domain can be phosphorylated [72]. Further analysis based on *in vitro* experiments in HepG2 cells revealed that Ser-162 in the HBV core protein is necessary and sufficient for the encapsidation of HBV RNA. However, both Ser-162 and Ser-170 are required for the production of HBV DNA replicative intermediates. The core Ser-155 is essential for the formation of relaxed circular DNA intermediates. Destruction of these phosphorylation sites by a conversion of Ser to Ala resulted in a nuclear accumulation of these nucleocapsids that do not contain significant amounts of DNA. HBx is proposed to support core phosphorylation at these residues to different extents, and thereby to exert a regulatory effect on HBV **replication.**

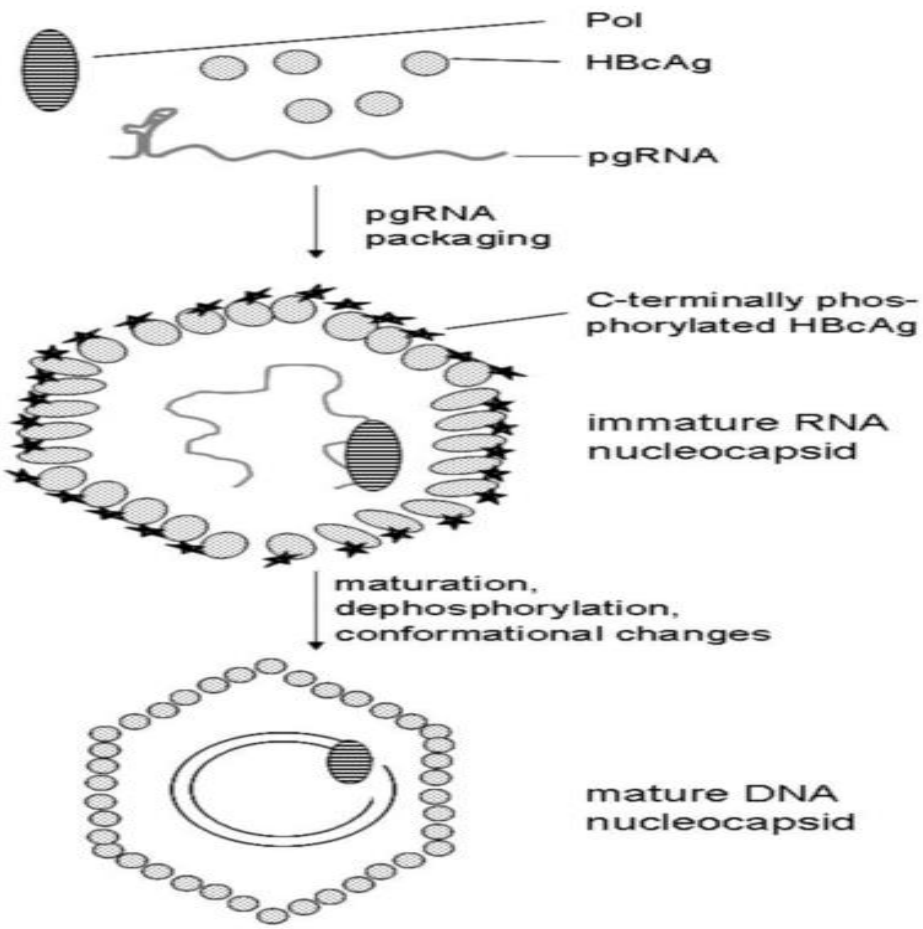


FIGURE 10- Genome packaging and nucleocapsid maturation. HBV nucleocapsid formation starts when the complex of the RNA pregenome , HBV polymerase and HBcAg dimers has formed. Efficient packaging of the RNA pregenome requires phosphorylation in the C-terminal part of the core protein. Conversion of the immature RNA-containing nucleocapsid to the mature DNA-containing nucleocapsid is associated with dephosphorylation and conformational changes. These significant differences in the structure between the RNA-containing immature nucleocapsid and the mature nucleocapsid trigger the envelopment of the mature nucleocapsid.

The identity of the kinase(s) that are involved in the phosphorylation of these residues is not fully understood. Based on *in vitro* experiments, protein kinase C or members of the SPRK kinase family are suggested to be involved. Analysis of DHBV capsid phosphorylation by detailed mass specrometric analysis revealed that the core protein from immature nucleocapsids was phosphorylated on at least six sites, whereas the mature nucleocapsid was completely dephosphorylated. In accordance with this, it was observed that mutation of the DHBV core phosphorylation sites to Ala completely blocked reverse transcription at a very early stage. Aspartate mutations, however, enabled complete first-strand DNA synthesis, but were defective in accumulating mature double-stranded DNA. This reflects, on the one

hand, the instability of the Asp-core mutants, and, on the other hand, the block in the mature second-strand DNA synthesis.

Based on the data from the HBV and DHBV systems, it has been concluded that nucleocapsid maturation can be described by a sequential phosphorylation (immature nucleocapsid) and dephosphorylation (mature nucleocapsid). This dephosphorylation during capsid maturation is associated with significant differences in the structure between the RNA- and the DNA-containing cores. In particular, there is a strong change affecting a hydrophobic pocket close to the spike that is required for the interaction of the preS1 domain with the nucleocapsid. This pocket is formed largely by residues that upon mutation have been shown to lead to abnormal viral secretion.

Envelopment and budding

In contrast to type C retroviruses or lentiviruses, where mutants with impaired envelope protein formation are still released coated with lipid bilayer, in the case of HBV the envelopment of the mature nucleocapsid strictly depends on the presence of the viral surface proteins. However, it was shown that MHBs are dispensable for virus production.. Formation of LHBs and SHBs are strictly required. Moreover, virion formation requires that in a fraction of LHBs the PreS1PreS2 domain faces the cytoplasm. Fusion of a secretion signal to the N-terminus of LHBs results in the exclusive formation of LHBs molecules that expose the PreS1PreS2 domain to the lumen of the ER. This enables the formation of subviral particles but prevents the secretion of viral particles. This observation is supported by findings from the DHBV system. Here it was found that the L protein is required for the envelopment of the nucleocapsid. Absence of L results in transport of the capsid to the nucleus, reimport and amplification of the viral genome. For DHBV, aa 116-137 of the PreS domain were considered to be essential for virus morphogenesis; in the case of HBV, aa 103-124 (aa 92-113 depending on the genotype).

It is assumed that this part of the PreS1 domain interacts with the capsid during envelopment. This hypothesis is supported by the observation that mutations within this part of PreS1 impair capsid envelopment. Moreover, *in vitro* binding assays with HBV PreS1-derived peptides and recombinant peptides support this (E. Hildt's unpublished results). In addition to the PreS1PreS2 domain that faces the cytoplasm, there is a short loop in the S domain between TM1 and TM2 that faces the cytoplasm and could interact with the nucleocapsid. Deletions within this domain inhibited virion formation, while the production of subviral particles was not affected.

To identify nucleocapsid residues that are crucial for envelopment, a variety of natural and engineered mutants were analyzed. Based on these experiments, it was concluded that the spike tip seems to have no impact on the capsid envelopment. In contrast, it was observed that a peptide that binds to the spike tip prevents secretion of mature viral particles. Cryo-electron microscopy of HBV particles purified by sucrose density gradient centrifugation supports the observation that the spike tip interacts via electrostatic interactions with HBsAg. Mutagenesis of HBcAg further demonstrated that the aa residues clustered around the base of the spike and in the grove between the spikes are essential for the interaction of the nucleocapsid with the envelope. Electron microscopy data from CsCl-purified HBV particles support this hypothesis, while electron microscopy of sucrose gradient purified viral particles [73] fails to demonstrate stable envelope contacts at these sites.

Mature hepadnaviral nucleocapsids form in the cytoplasm. For DHBV, it has been shown that mature nucleocapsids attach to intracellular membranes. This attachment does not require the presence of envelope proteins. Immature nucleocapsids do not bind. The exact mechanism that mediates the delivery of mature nucleocapsids to the post-ER, pre-Golgi-compartment, where envelopment occurs, is presently not understood. For retroviruses and some enveloped RNA viruses it was shown that budding from the plasma membranes depends on host functions involved in protein sorting into late endosomal multivesicular bodies (MVBs). Inhibition of different MVB proteins by coexpression of dominant-negative mutants of AIP1/ALIX, and VPS4B revealed that MVB functions are required for efficient budding and release of enveloped HBV virions. Moreover, HBV virions and subviral particles are all released by distinct pathways with separate host factor requirements

EVOLUTION

The early evolution of HBV, like that of all viruses, is difficult to establish. The identification of hepadnaviruses in a wide range of vertebrates suggests a long coevolution. The identification of endogenous hepadnaviridae elements shared by various bird species shows the presence of these viruses in birds for at least 70M years. Although similar evidence is missing for mammals, the phylogenetic position of ortho hepadnaviruses as a sister clade to avi hepadnaviruses suggests a presence of the virus in the amniote ancestor and a subsequent coevolution with both birds and mammals after their divergence (>300M years ago). It has also been proposed that a New World bat hepadnavirus may be the origin of the primate hepadnaviruses. Avi Hepadnaviruses lack the X protein but a vestigial X reading frame is present in the genome of duck hepadnavirus. The X protein may have evolved from a DNA glycosylase.

Recently, the reconstruction of HBV genomes from ancient human remains has allowed investigating the evolution of this virus in humans in more detail. In 2021, a study reconstructed 137 ancient HBV genomes and proved the presence of the virus in humans for at least 10,000 years. The most recent common ancestor of all known human HBV lineages was dated to between 20,000 and 12,000 years ago. However, it cannot be said whether the virus was present in humans long before that or acquired shortly before from another animal species. The evolution of HBV in humans was shown to reflect known events of human history such as the first peopling of the Americas during the late Pleistocene and the Neolithic transition in Europe. These studies also showed that some ancient HBV strains still infect humans, while other became extinct. HBV strains found in African and South-East Asian apes (chimpanzees, gorillas, orangutans and gibbons) appear related to human HBV strains, which could reflect past cross-species transmission events.

AIM / OBJECTIVES

The main aim of this study is:

- To retrieve the complete nucleotide sequences of the Whole genome of HBV from NCBI databases from various hospitals and research institutes in India.

- To retrieve various sub genotype sequences of HBV submitted to NCBI from around the world.

- To perform Phylogenetic analysis of selected HBV sequences with the help of MEGA 11 software.

METHODOLOGY

1. RETRIEVAL OF HBV SEQUENCES:

The NCBI nucleotide database was used for retrieval of complete HBV sequences.

NCBI

The National Center for Biotechnology Information (NCBI) was established in November 1988, at the National Library of Medicine (NLM) in the United States. The NLM was chosen because it had experience in creating and maintaining biomedical databases and as part of the National Institutes of Health (NIH), it could establish an intramural research program in computational molecular biology. The mission of the NCBI is to develop new information technologies to aid in the understanding of fundamental molecular and genetic processes that control health and disease.

The NCBI houses a series of databases relevant to biotechnology and biomedicine and is an important resource for bioinformatics tools and services. Major databases include

Genbank for DNA sequences and PubMed, a bibliographic database for biomedical literature. Other databases include the NCBI Epigenomics database. All these databases are available online through Entrez search engine. NCBI was directed by David Lipman, one of the original authors of the BLAST sequence alignment program and a widely respected figure in bioinformatics.

PHYLOGENETIC ANALYSIS

Phylogenetics is a powerful approach in finding evolution of current day species. By studying phylogenetic trees, scientists gain a better understanding of how species have evolved while explaining the similarities and differences among species. The phylogenetic study can help in analysing the evolution and the similarities among diseases and viruses, and further help in prescribing their vaccines against them

PHYLOGENETICS

Phylogenetics can be considered as one of the best tools for understanding the spread of contagious disease, for example, transmission of the human immunodeficiency virus (HIV) and the origin and subsequent evolution of the severe acute respiratory syndrome (SARS) associated coronavirus (SCoV) Earlier, morphological traits were used for assessing similarities between species and building phylogenetic trees. Presently, phylogenetics relies on information extracted from genetic material such as deoxyribonucleic acid (DNA), ribonucleic acid (RNA) or protein sequences

PHYLOGENETIC TREE

A phylogenetic tree, also known as a phylogeny, is a diagram that depicts the lines of evolutionary descent of different species, organisms, or genes from a common ancestor. Phylogenies are useful for organizing knowledge of biological diversity, for structuring classifications, and for providing insight into events that occurred during evolution. Furthermore, because these trees show descent from a common ancestor, and because much of the strongest evidence for evolution comes in the form of common ancestry, one must understand phylogenies in order to fully appreciate the overwhelming evidence supporting the theory of evolution.

Tree diagrams have been used in evolutionary biology since the time of Charles Darwin. Therefore, one might assume that, by now, most scientists would be exceedingly comfortable with "tree thinking"--reading and interpreting phylogenies. However, it turns out that the tree model of evolution is somewhat counterintuitive and easily misunderstood. This may be the reason why biologists have only in the last few decades come to develop a rigorous understanding of phylogenetic trees. This understanding allows present-day researchers to use phylogenies to visualize evolution, organize their knowledge of biodiversity, and structure and guide ongoing evolutionary research[74].

A phylogenetic tree is an estimate of the relationships among taxa (or sequences) and their hypothetical common ancestors (Nei and Kumar 2000; Felsenstein 2004; Hall 2011). Today most phylogenetic trees are built from molecular data: DNA or protein sequences. Originally, the purpose of most molecular phylogenetic trees was to estimate the relationships among the species represented by those sequences, but today the purposes have expanded to include understanding the relationships among the sequences themselves without regard to the host species, inferring the functions of genes that have not been studied experimentally (Hall et al. 2009), and elucidating mechanisms that lead to microbial outbreaks (Hall and Barlow

2006) among many others. Building a phylogenetic tree requires four distinct steps: (Step 1) identify and acquire a set of homologous DNA or protein sequences, (Step 2) align those sequences, (Step 3) estimate a tree from the aligned sequences, and (Step 4) present that tree in such a way as to clearly convey the relevant information to others.

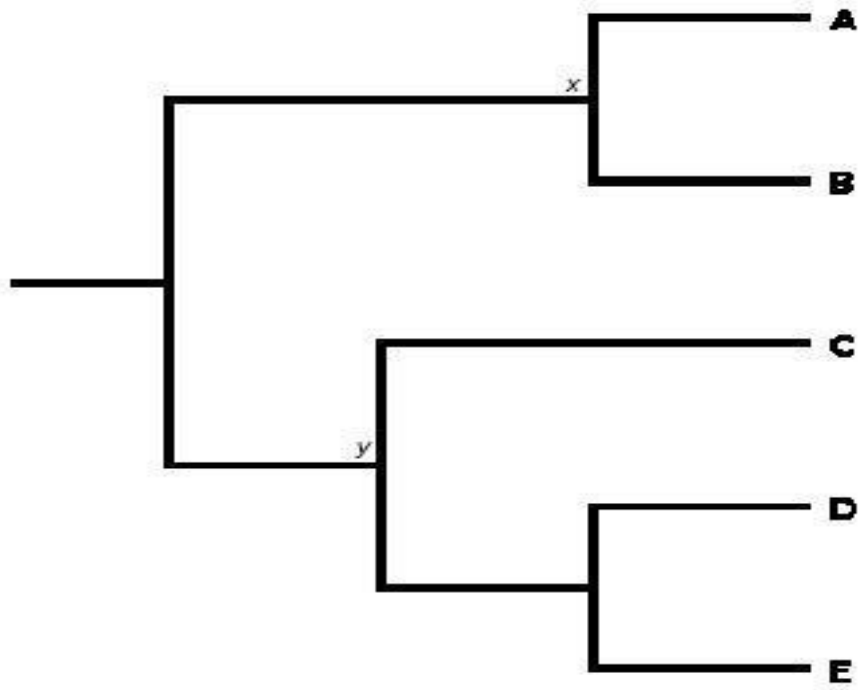


FIGURE11 - Trees contain information on the relative timing of nodes only when the nodes are on the same path from the root (i.e., when one node is a descendant of another).
In this tree, nodes x and y are not on the same path, so we cannot tell whether the ancestral organisms in node x lived before or after those in node y.
TAKEN FROM- Avise, J. C. *Evolutionary Pathways in Nature: A Phylogenetic Approach* (Cambridge University Press, Cambridge, UK, 2006)

Moreover, phylogenetics can be used to evaluate the reciprocal evolutionary interaction between microorganisms, as well as to identify mechanisms (horizontal gene transfer) responsible for the rapid adaptation of pathogens in an ever-changing host microenvironment.

MEGA 6 SOFTWARE

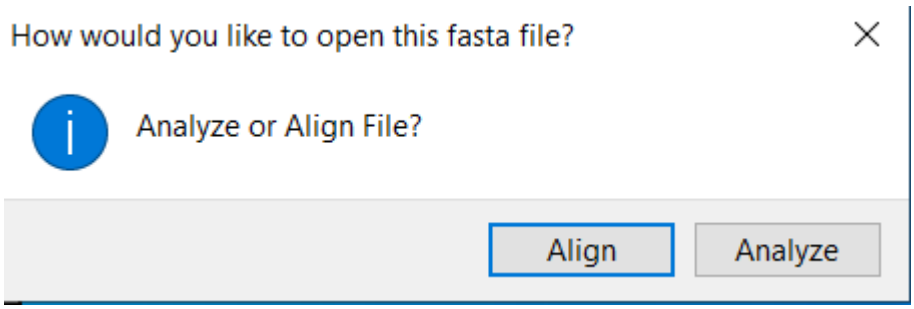
The Molecular Evolutionary Genetics Analysis (MEGA) software is developed for comparative analyses of DNA and protein sequences that are aimed at inferring the molecular evolutionary patterns of genes, genomes, and species over time (Kumar et al. 1994; Tamura et al. 2011). MEGA is currently distributed in two editions: a graphical user interface (GUI) edition with visual tools for exploration of data and analysis results (Tamura et al. 2011) and a command line edition (MEGA-CC), which is optimized for iterative and integrated pipeline analyses (Kumar et al. 2012).

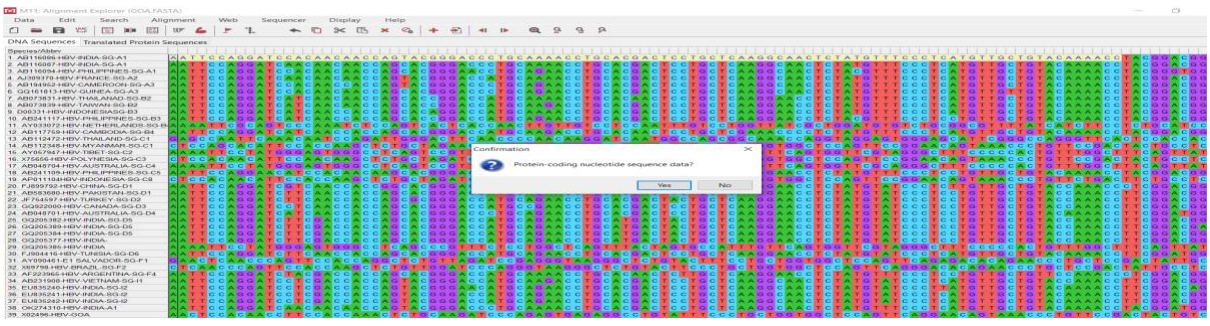
Step 1: Acquiring the Sequences

Ironically, the first step is the most intellectually demanding, but it often receives the least attention. If not done well, the tree will be invalid or impossible to interpret or both. If done wisely, the remaining steps are easy, essentially mechanical, operations that will result in a robust meaningful tree.

Step 2: Aligning the Sequences

If the Alignment Explorer window is not already open, in MEGA5's main window choose **Open a File/Session** from the **File menu.** Choose the MEGA5 alignment file (.mas) or the sequence file (.fasta) that you saved in Step 1. In the resulting dialog choose **Align**.





Step 3: Estimate the Tree

Step 4: Present the Tree

A drawing of a phylogenetic tree conveys a lot of information, both explicit and implicit. Some of that implicit information can be misleading, so it is up to the investigator to ensure that the information conveyed, both explicit and implicit, is correct.

A phylogenetic tree consists of external nodes (the tips) that represent the actual sequences that exist today, internal nodes that represent hypothetical ancestors, and branches that connect nodes to each other. The lengths of the branches represent the amount of change that is estimated to have occurred between a pair of nodes. That is the explicit information conveyed by a tree drawing.

In version 6.0, we have now added facilities for building molecular evolutionary trees scaled to time (timetrees), which are clearly needed by scientists as an increasing number of studies are reporting divergence times for species, strains, and duplicated genes (e.g., Kumar and Hedges 2011; Ward et al. 2013). For this purpose, we have implemented the RelTime method, which can be used for large numbers of sequences comprising contemporary data sets, is the fastest method among its peers, and is shown to perform well in computer simulations (Tamura et al. 2012). RelTime produces estimates of relative times of divergence for all branching points (nodes) in any phylogenetic tree without requiring knowledge of the distribution of the lineage rate variation and without using clock calibrations and associated distributions. Relative time estimates produced by MEGA will be useful for determining the ordering and spacing of sequence divergence events in species and gene family trees. The (relative) branch rates produced by RelTime will also enable users to determine the statistical distribution of evolutionary rates among lineages and detect rate differences between species and duplicated gene clades. In addition, relative times obtained using molecular data can be directly compared with the times from nonmolecular data (e.g., fossil record) to test independent biological hypotheses.

RESULT-

Table showing the various Subgenotype sequences of HBV retrieved from NCBI. These sequences were used as reference sequences to identify the subgenotype of HBV sequences submitted to NCBI from INDIA.

S NO.	SUBGENOTYPES	ACCESSION NO.	REGION
1	A1	AB116086	INDIA
		AB116087	INDIA
		AB116094	PHILIPPINES
	A2	AY233280	SOUTH AFRICA
		AJ309370	FRANCE
	A3	AB194952	CAMEROON
		GQ161813	GUINEA
2	B1	AB010290	JAPAN
		D23768	JAPAN
	B2	AB073831	THAILAND
		AB073839	TAIWAN
	B3	D00331	INDONESIA

		AB241117	PHILIPPINES
	B4	AY033072 AB117759	NETHERLANDS CAMBODIA
3	C1	AB112472 AB112348	THAILAND MYANMAR
	C2	AY057947 AB1014376	TIBET JAPAN
	C3	X75656	POLYNESIA
	C4	AB048704	AUSTRALIA
	C5	AB241109	PHILIPPINES
	C6	AB493842	INDONESIA
	C7	EU670263	PHILIPPINES
	C8	AP011104	INDONESIA
4	D1	FJ899792 AB583680	CHINA PAKISTAN
	D2	JF754597	TURKEY
	D3	GQ922000	CANADA
	D4	AB048701	AUSTRALIA
	D5	GQ205382 GQ205389 GQ205384	INDIA INDIA INDIA

	D6	FJ904416	TUNISIA
5	F1 F2 F4	AY090461 X69798 AF223965	E1 SALVADOR BRAZIL ARGENTINA
6	I1 I2	AB231908 EU835240 EU835241 EU835242	VIETNAM INDIA INDIA INDIA

The below table lists the HBV sequences that were submitted to NCBI from various institutes and hospitals of INDIA.

INDIA-LUCKNOW-KGMU

S NO.	GENOTYPE	ACCESSION NO.	REGION	NOMENCLATURE
1	A	KT151612	LUCKNOW	L1
2	A	KT151614	LUCKNOW	L2
3	A	KT151615	LUCKNOW	L3
4	A	KT151617	LUCKNOW	L4
5	D	KT151620	LUCKNOW	L5

> KT151612-HBV-INDIA-LUCKNOW-KGMC

AATTCCAGGATCCACAACAACCAGTACGGGACCCTGCAAAACCTGCACGACTC
CTGCTCAAGGCAACTCTATGTTTCCCTCATGTTGCTGTACAAAACCTACGGATG
GAAATTGCACCTGTATTCCCATCCCATCATCYTGGGCTTTCGCAAAATACCTATG
GGAGTGGGCCTCAGTCCGTTTCTCTTGGCTCAGTTTACTAGTGCCATTTGTTCA

GTGGTTCGTAGGGCTTTCCCCCACTGTTTGGCTTTCAGCTATATGGATGATGTG
GTCCTGGGGGCCAAGTCTGTACAACACCTTGAATCCCTTTTTACCGCTGTTACC
AATTTTCTTTTGTCTTTGGGTATACATTTAAACCCTAACAAAACAAAGAGATGGG
GTTATTCCCTAAATTTCATGGGGTATGTAATTGGAAGTTGGGGTACATTGCCAC
AGGATCATATTAGACAAAAAATCAAACACTGTTTTAGAAAACTTCCTGTCAATCG
GCCTATTGATTGGAAAGTATGTCAGAGAATTGTGGGTCTTTTGGGCTTTGCCGC
CCCATTTACACAATGTGGTTACCCTGCYTTAATGCCTTTGTATGCATGTATACAA
GCGAAACAGGCTTTTACTTTCTCGCCAACTTACAAGGCCTTTCTAAGTCAACAG
TATGTGAACCTTTACCCCGTTGCCCGGCAACGGCCTGGTCTGTGCCAAGTGTTT
GCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATCGGCCATCAGCGCATGC
GCGGAACCTTTGTGGCTCCTCTGCCGATCCATACTGCGGAACTCCTAGCCGCT
TGTTTTGCTCGCAGCAGGTCTGGAGCAAAACTCATCGGGACTGATAATTCTGTC
GTCCTTTCTCGGAAATATACATCATTTCCATGGCTGCTAGGTTGTACTGCCAACT
GGATTCTTCGCGGGACGTCCTTTGTTTACGTCCCGTCGGCGCTGAATCCCGCG
GACGACCCCTCGCGRGGCCGCTTGGGACTCTATCGTCCCCTTCTCCGTCTGCC
GTACCGTCCGACCACGGGGCGCACCTCTCTTTACGCGGTCTCCCCGTCTGTGC
CTTCTCACCTGCCGGTCCGTGTGCACTTCGCTTCACCTCTGCACGTTGCATGGA
GACCACCGTGAACGCCCATCAGAGCCTGCCCAAGGTCTTACATAAGAGGACTC
TTGGACTCCCAGCAATGTCAACGACCGACCTTGAGGCCTACTTCAAAGACTGTG
TGTTTAAAGACTGGGAGGAGCTGGGGGAGGAGATTAGGTTAATGATCTTTGTAT
TAGGAGGCTGTAGGCATAAAATGGTCTGCGCACCATCCTCATGCAACTTTTTCA
CCTCTGCCTAATCATCTCTTGTACATGTCCCACTTTTCAAGCCTCCAAGCTGTGC
CTTGGATGGCTTTGGGGCATGGACATTGACCCTTATAAAGAATTTGGAGCTACT
AGCGAGTTACTCTCTTTTTTGCCTTCTGACTTCTTTCCTTCCGTYCGGGATCTAC
TAGATACAGCCTCAGCTCTGTATCGGGAAGCCTTAGAGTCTCCTGAGCATTGCT
CACCTCACCATACTGCACTCAGGCAAGCCGTTCTCTGCTGGGGGGAATTAATG
ACTCTAGCTACCTGGGTGGGTAATAATTTGGAAGATCCAGCATCCAGGGATCTA
GTAGTCAATTATGTTAATACTAACATGGGCCTAAAGATCAGGCAACTATTGTGGT
TTCATATTTCTTGCCTTACTTTTGGAAGAGAAACTGTCCTTGAGTATTTGGTTTCT
TTCGGAGTGTGGATCCGCACTCCTCCACCCTATAGACCAGCAAATGCCCCTATC
TTATCAACACTTCCGGAAACTACTGTTATTAGAGGAAGAGGCCGAGGCAGGTCC
CCTAGAAGAAGAACTCCCTCGCCTCGCAGACGAAGATCTCAATCGCCGCGTCG
CAGAAGATCTCAATCTCGGGAATCTCAATGTTAGTATTCCTTGGACTCATAAGGT
GGGAAATTTTACTGGACTTTATTCTTCTACTGTTCCTATCTTTAATCCTGACTGG
CAAACTCCTTCYTTTCCTAAAATTCATTTACATGAAGACATTGCTAATAGGTGTC
AGCAATTTGTAGGCCCTCTCACTGTAAATGAAAAGAGAAGACTGAAATTAATTAT
GCCTGCTAGGTTTTATCCTAACAGCACTAAATATTTGCCTCTAGACAAAGGGATT
AAACCTTATTATCCTGATCAGGTAGTTAATCATTACTTTAAAACCCGACATTATTT
ACATACTCTTTGGAAGGCTGGGATTCTATATAAGAGGGAATCTACACGTAGCGC
CTCATTTTGCGGGTCACCATATTCTTGGGAACAAGAGCTACATCATGGGAGGTT
GGTCATCAAAACCTCGCAAAGGCATGGGGACGAACCTTTCTGTTCCCAATCCTC
TGGGATTCTTTCCCGATCATCAGTTGGACCCTGCATTCGGAGCCAATTCAAACA
ATCCAGATTGGGACTTCAACCCCGTCAAGGACCATTGGCCAMMAGCCAACCAG
GTAGGAGTGGGAGCATTCGGGCCAGGGTTCACTCCCCCACACGGAGGTGTTTT
GGGGTGGAGCCCTCAGGCTCAGGGCATATTGGCCACAGTGCCAGCAGTTCCT
CCTCCTGCCTCCACCAATCGGCAGGCAGGACGACAGCCTACTCCCATCTCTCC
ACCTCTAAGAGACAGTCATCCTCAGGCCATGCAGTGGAATTCCACTGCCTTCCA
CCAAGCTCTGCAGGATCCCAGAGTCAGGGGTCTGTATTTTCCTGCTGGTGGCT
CCAGTTCAGGAACAGTAAACCCTGCTCCGAATATTGCCTCTCACATCTCGTCAA
TCTCCGCGAGGACTGGGGACCCTGTGGCGAACATGGAGAACATCACATCAGGA

TTCCTAGGACCCCTGCTCGTGTTACAGGCGGGGTTTTTCTTGTTGACAAGAATC
CTCACAATACCGCAGAGTCTAGACTCGTGGTGGACTTCTCTCAATTTTCTAGGG
GGCTCACCCGTGTGTCTTGGCCAAAATTCGCAGTCCCCAACCTCCAATCACTCA
CCAACCTCCTGTCCTCCAATTTGTCCTGGTTATCGCTGGATGTGTCTGCGGCGT
TTTATCATATTCCTCTTCATCCTGCTGCTATGCCTCATCTTCTTATTGGTTCTTCT
GGATTATCAAGGTATGTTGCCCGTTTGTCCTCT

INDIA-MAHARASHTRA-NATIONAL INSTITUTE OF VIROLOGY

S NO.	GENOTYPE	ACCESSION NO.	REGION	NOMENCLATURE
1	D	GQ183486	PUNE	P1
2	D	GQ183485	PUNE	P2
3	D	GQ183484	PUNE	P3
4	D	GQ183483	PUNE	P4
5	D	GQ183482	PUNE	P5
6	A/G/C	EU835242	PUNE	P6
7	I	KF214680	PUNE	P7
8	A	KF214659	PUNE	P8
9	A	KF214663	PUNE	P9
10	A	KF214666	PUNE	P10
11	C	KF214668	PUNE	P11

> GQ183486-HBV-INDIA-MAHARASHTRA-PUNE-NATIONAL INSTITUTE OF VIROLOGY

AATTCCAGGATCCTCAACCACCAGCACGGGACCATGCAGAACCTGCACGACTC
CTGCTCAAGGAACCTCTATGTATCCCTCCTGTTGCTGTACCAAACCTTCGGACG
GAAATTGCACCTGTATTCCCATCCCATCATCCTGGGCTTTCGGAAAATTCCTAT
GGGAGTGGGCCTCAGCCCGTTTCTCCTGGCTCAGTTTACTAGTGCCATTTGTTC
AGTGGTTCGTAGGGCTTTCCCCCACTGTTTGGCTTTCAGTTATATGGATGATGT
GGTATTGGGGGCCAAGTCTGTACAGCATCTTGAGTCCCTTTTTACCGCTGTTAC
CAATTTTCTTTTGTCTTTGGGTATACATTTAAACCCTAACAAAACAAAAAGATGG
GGTTACTCTCTAAATTTTATGGGCTATGTCATTGGATGTTATGGATCTTTGCCAC
AAGAACACATCAAACAAAAAATCAAAGAATGTTTTCGCAAACTTCCTGTTAACCG
GCCTATTGATTGGAAAGTGTGTCAACGAATTGTGGGTCTTTTGGGTTTTGCTGC
TCCTTTTACACAATGTGGTTATCCTGCTTTAATGCCTTTGTATGCATGTATTCAAT
CCAAGCAGGCTTTTACCTTCTCGCCAACTTACAAGGCCTTTCTGTGTAAACAATA
CCTGAACCTTTACCCCGTTGCTCGGCAACGGCCAGGTCTGTGCCAAGTGTTTG
CTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCATGCG
TGGAACCTTTATGGCTCCTCTGCCGATCCATACTGCGGAACTCTTAGCAGCGTG
TTTTGCTCGCAGCAGGTCTGGAGCAAACCTTATCGGGACTGATAACTCTGTTGT
CCTCTCCCGCAAATATACATCGTTTCCATGGCTGCTCGGCTGTGCTGCCAACTG
GATCCTGCGCGGGACGTCCTTTGTTTACGTCCCGTCGGCGCTGAATCCCGCGG
ACGACCCTTCTCGGGGTCGCTTGGGACTCTCTCGTCCCCTTCTCCGTCTGCCG
TTCCGACCGACCACGGGGCGCACCTCTCTTTACGCGGACTCCCCGTCTGTGCC
TTCTCATCTGCCGGACCGTGTGCACTTCGCTTCACCTCTGCACGTCGCATGGA
GACCACCGTGAACGCCCACCAAATCTTGCCCAAGGTCTTACATAAGAGGACTCT
TGGACTCTCTGCAATGTCAACGACCGACCTTGAGGCATACTTCAAAGACTGTTT
GTTTAAAGACTGGGAGGAGTTGGGGGAGGAGATTAGATTAAAGGTCTTTGTACT
AGGAGGCTGTAGGCATAAATTGGTCTGCGCACCAGCACCATGCAACTTTTTCAC
CTCTGCCTAATCATCTCTTGTTCATGTCCTACTGTTCAAGCCTCCAAGCTGTGCC
TTGGGTGGCTTTGGGGCATGGACATTGACCCTTATAAAGAATTTGGAGCTACCG
TGGAGTTACTCTCGTTTTTGCCTTCTGACTTCTTTCCTTCAGTACGAGATCTTCT
AGATACCGCCTCAGCTCTGTATCGGGATGCCTTAGAGTCTCCTGAGCATTGCTC
ACCTCACCATACTGCACTCAGGCAAGCAATTCTGTGCTGGGGGGAACTAATGA
CTCTAGCTACCTGGGTGGGTGTTAATTTGGAAGATCCAGCATCTAGGGACCTAG
TAGTCAGTTATGTCAACACTAATATGGGCCTAAAGTTCAGGCAACTCTTGTGGTT
TCACATTTCTTGTCTTACTTTTGGAAGAGAAACAGTTATAGAATATTTGGTGTCTT
TCGGAGTGTGGATTCGCACTCCTCCAGCTTATAGACCACCAAATGCCCCTATCT
TATCAACACTTCCGGAGACTACTGTTGTTAGACGACGAGGCAGGTCCCCTAGAA
GAAGAACTCCCTCGCCTCGCAGACGAAGGTCTCAATCGCCGCGTCGCAGAAGA
TCTCAATCTCGGGAATCTCAATGTTAGTATTCCTTGGACTCATAAGGTGGGAAA
CTTTACGGGGCTTTATTCTACTACTGTACCTGTCTTTAATCCTCATTGGAAAACA
CCCTCTTTTCCTAATATACATTTACATCAAGACATTATCAAAAAATGTGAACAGTT
TGTAGGCCCACTCACAGTTAATGAAAAAAGAAGATTGCAATTGATTATGCCTGC
TAGGTTTTATCCTAATGTTACCAAATATTTGCCATTGGATAAGGGTATTAAACCTT
ATTATCCAGAACACCTAGTTAATCATTACTTCCAAACTAGACACTATTTACACACT
CTTTGGCAGGCGGGTATATTATATAAGAGAGAAACAACACATAGCGCCTCATTT
TGTGGGTCACCATATTCTTGGGAACAAGAGCTACAGCATGGGGCAGAATCTTTC
CACCAGCAATCCTCTGGGATTCTTTCCCGACCACCAGTTGGATCCAGCCTTCAG
AGCAAACACCGCAAATCCAGATTGGGACTTCAATCCCAACAAGGACACCTGGC
CAGACGCCAACAAGGTAGGAGCTGGAGCATTCGGGCTGGGTTTCACCCCACC
GCACGGAGGCCTTTTGGGGTGGAGCCCTCAGGCTCAGGGCATACTACAAACTT

TGCCAGCAAATCCGCCTCCTGCCTCCACCAATCGCCAGTCAGGAAGGCAGCCC
ACCCCGCTGTCTCCACCTTTGAGAAACACTCATCCTCAGGCCATGCAGTGGAA
CTCCACAACCTTCCACCAAACTCTGCAAGATCCCAGAGTGAGAGGCCTGTATTT
CCCTGCTGGTGGCTCCAGTTCAGGAACAGTAAACCCTGTTCCGACTACTGCCT
CTCACTTATCGTCAATCTTCTCGAGGATTGGGGACCCTGCGCTGAACATGGAGA
ACATCACATCAGGATTCCTAGGACCCCTTCTCGTGTTACAGGCGGGGTTTTTCT
TGTTGACAAGAATCCTCACAATACCGCAGAGTCTAGACTCGTGGTGGACTTCTC
TCAATTTTCTAGGGGGAACTACCGTGTGTCTTGGCCAAAATTCGCAGTCCCCAA
CCTCCAATCACTCACCAACCTCCTGTCCTCCAACTTGTCCTGGTTATCGCTGGA
TGTGTCTGCGGCGTTTTATCATCTTCCTCTTCATCCTGCTGCTATGCCTCATCTT
CTTGTTGGTTCTTCTGGACTATCAAGGTATGTTGCCCGTTTGTCCTCT

INDIA-KOLKATA-ICMR

S NO.	GENOTYPE	ACCESSION NO.	REGION	NOMENCLATURE
1	A	DQ315786	KOLKATA	K1
2	D	KT366499	KOLKATA	K2
3	D	KT366497	KOLKATA	K3
4	A	KR905431	KOLKATA	K4

> DQ315786-HBV-INDIA-KOLKATA-ICMR

AATTCCAGGATCCACAACAACCAGCACGGGACCCTGCAAAACCTGCACGACTC
CTGCTCAAGGCAACTCTATGTTTCCCTCATGTTGCTGTACAAAACCTACGGATG
GAAATTGCACCTGTATTCCCATCCCATCATCTTGGGCTTTCGCAAAATACCTATG
GGAGTGGGCCTCAGTCCGTTTCTCTTGGCTCAGTTTACTAGTGCCATTTGTTCA
GTGGTTCGTAGGGCTTTCCCCCACTGTTTGGCTTTCAGCTATATGGATGATGTG
GTACTGGGGGCCAAGTCTGTACAACATCTTGAGTCCCTTTATACCGCTGTTACC
AATTTTCTTTTGTCTTTGGGTATACATTTAAACCCTAACAAAACAAAGAGATGGG
GTTATTCCCTGAATTTCATGGGTTATGTAATTGGAAGTTGGGGTACATTGCCACA
GGATCATATTGTACAAAAAATCAAACACTGTTTTAGAAAACTTCCTGTCAATCGA
CCTATTGATTGGAAAGTATGTCAGAGAATTGTGGGTCTTTTGGGTTTTGCCGCT
CCATTTACACAATGTGGTTACCCTGCATTAATGCCTTTGTATGCATGTATACAAG
CGAAACAGGCTTTTACTTTCTCGCCAACTTACAAGGCCTTTCTAAGTAAACAGTA
TATGAACCTTTACCCCGTTGCCCGGCAACGGCCTGGTCTGTGCCAAGTGTTTG
CTGACGCAACCCCCACTGGCTGGGGCTTGGCCATCGGCCATCAGCGCATGCG

TGGAACCTTTGTGGCTCCTCTGCCGATCCATACTGCGGAACTCCTAGCCGCTT
GTTTTGCTCGCAGCCGGTCTGGAGCAAAACTCATCGGGACTGATAATTCTGTCG
TCCTCTCTCGGAAATATACATCATTTCCATGGCTGCTAGGTTGTACTGCCAACT
GGATACTACGAGGGACGTCCTTTGTTTACGTCCCGTCGGCGCTGAATCCCGCG
GACGACCCTTCGCGAGGCCGCTTGGGACTGTATCGTCCCCTTCTCCGTCTGCC
GTACCGTCCGACCACGGGGCGCACCTCTCTTTACGCGGTCTCCCCGTCTGTGC
CTTCTCATCTGCCGGTCCGTGTGCACTTCGCTTCACCTCTGCACGTTGCATGGC
GACCACCGTGAACGCCCATCAGATCCTGCCCAAGGTCTTACATAAGAGGACTC
TTGGACTCCCAGCAATGTCAACGACCGACCTTGAGGCCTACTTCAAAGACTGTG
TGTTTAAAGACTGGGAGGAGTTGGGGGAGGAGATTAGGTTAATGATCTTTGTAT
TAGGAGGCTGTAGGCATAAATTGGTCTGCGCACCATCATCATGCAACTTTTTCA
CCTCTGCCTAATCATCTCTTGTACATGTCCCACTGTTCAAGCCTCCAAGCTGTG
CCTTGGTTGGCTTTGGGGCATGGACATTGACCCTTATAAAGAATTTGGAGCTAC
TGTGGAGTTACTCTCATTTTTGCCTTCTGACTTCTTTCCTTCCGTCCGGGATCTA
CTAGACACAGCCTCAGCTCTGTATCGGGAAGCCTTAGAGTCTCCTGAGCATTG
CTCACCTCACCATACAGCACTCAGGCAAGCCATTCTCTGCTGGGGGGAATTAAT
GACTCTAGCTACCTGGGTGGGTAATAATTTGGAAGATCCAGCATCCAGGGATCT
AGTAGTCAATTATGTTAATACTAACATGGGCCTAAAGATCAGGCAACTATTGTGG
TTTCATATTTCTTGCCTTACTTTTGGAAGAGAAACTGTCCTTGAGTATTTGGTCTC
TTTCGGAGTGTGGATTCGCACTCCTCCAGCCTATAGACCACCAAATGCCCCTAT
CTTATCAACACTTCCGGAAACTACTGTTGTTAGACGACGAGACCGAGGTAGGTC
CCCTAGAAGAAGAACTCCCTCGCCTCGCAGACGAAGATCTCAATCGCCGCGTC
GCAGAAGATCTCAATCTCGGGAATCTCAATGTTAATATTCCTTGGACTCATAAG
GTGGGAAATTTTACTGGGCTTTATTCTTCTACTGTCCCTATCTTTAATCCTGAAT
GGCAAACCCCTTCTTTTCCAAAAATTCATTTACATGAAGACATTGCTAATCGGTG
TCAGCAATTTGTAGGCCCTCTCACTGTAAATGAAAAGAGAAGACTGAAATTAATT
ATGCCTGCTAGGTTTTACCCTAACAGCACAAAATATTTGCCTCTAGACAAAGGG
ATTAAACCTTATTATCCTGATCATGTAGTTAATCATTACTTTCAAACCCGACATTA
TTTACATACTCTTTGGAAGGCTGGGATTCTATATAAGAGGGAAACTACACGTAG
CGCCTCATTTTGCGGGTCACCATATTCTTGGGAACAAGAGCTACATCATGGGAG
GTTGGTCATCAAAACCTCGCAAAGGCATGGGGACGAACCTTTCTGTTCCCAACC
CTCTGGGATTCTTTCCCGATCATCAGTTGGACCCTGCATTCGGAGCCAATTCAA
ACAATCCAGATTGGGACTTCAACCCCATCAAGGACCACTGGCCACAAGCCAAC
CAGGTAGGAGTGGGAGCATTCGGGCCAGGGTTCACTCCCCCACACGGAGGTG
TCTTGGGGTGGAGCCCTCAGGCTCAGGGCATATTGGCCACAGTGCCAGCAGT
GCCTCCTCCTGCCTCCACCAATCGGAAGTCAGGAAGGCAGCCTACTCCCATCT
CTCCACCTCTAAGAGACAGTCATCCTCAGGCCATGCAGTGTTCCACAGCTTTCC
ACCAAGCTCTGCAAGATCCCAGAGTCAGGGGCCTGTATCTTCCTGCTGGTGGC
TCCAGTTCAGGGACTCTCAACCCTGTTCCCAATATTGCCTCTCACATATCGTCA
ATCTCCTCGAGGACTGGGGACCCTGCGCCGAACATGGAGAACATCACATCAGG
ATTCCTAGGACCCCTGCTCGTGTTACAGGCGGGGTTTTTCTTGTTGACAAGAAT
CCTCACAATACCGCAGAGTCTAGACTCGTGGTGGACTTCTCTCAATTTTCTAGG
GGGATCACCCGTGTGTCTTGGCCAAAATTCGCAGTCCCCAACCTCCAATCACTC
ACCAACCTCCTGTCCTCCAATTTGTCCTGGTTATCGCTCGATGTGTCTGCGGCG
TTTTATCATATTCCTCTTCATCCTGCTGCTATGCCTCATCTTCTTATTGGTTCTTC
TGGATTATCAAGGTATGTTGCCCGTTTGTCCTCT

INDIA-M.P.-JABALPUR-ICMR-NIRTH

S NO.	GENOTYPE	ACCESSION NO.	REGION	NOMENCLATURE
1	D	MK541689	JABALPUR	J1

> MK541689 -HBV-INDIA-M.P.-JABALPUR-ICMR-NIRTH

CTCCAAAACCTTCCACCAAGCTCTACAAGATCCCAGAGTGAGAGGCCTGTATTT
CCCTGCTGGTGGCTCCAGTTCAGGAACAGTAAACCCTGTTCCGACTACTGCCT
CTCCCATATTGTCTCTCTTCTCGAAGATTGGGGACCTTGCACCGAACATGGAGA
ACATCACATCAGGATTCCTAGGACCCCTGCTCGTGTTACAGGCGGGGTTTTTCT
TGTTGACAAGAATCCTCACAATACCGCAGAGTCTAGACTCGTGGTGGACTTCTC
TCAATTTTCTAGGGGGAACCACCGTGTGTCTTGGCCAAAATTCGCAGTCCCCAA
CCTCCAATCACTCACCAACCTCTTGTCCTCCAACTTGTCCTGGTTATCGCTGGA
TGTGTCTGCGGCGTTTTATCATCTTCCTCTTCATCCTGCTGCTATGCCTCATCTT
CTTGTTGGTTCTTCTGGACTATCAAGGTATGTTGCCCGTTTGTCCTCTGATTCCA
GGATCTTCGACCACCAGTACGGGACCATGCAGAACCTGCATGACTACTGCTCA
AGGAACCTCTATGTATCCCTCCTGTTGCTGTACAAAACCTTCGGACGGAAATTG
CACCTGTATTCCCATCCCATCATCCTGGGCTTTCGGAAAATTCCTATGGGAGTG
GGCCTCAGCCCGTTTCTCCTGGCTCAGTTTACTAGTGCCATTTGTTCAGTGGTT
CGTAGGGCTTTCCCCCACTGTTTGGCTTTCAGTTATATGGATGATGTGGTATTG
GGGGCCAAGTCTGTACAGCATCTTGAGTCCCTTTTTACCGCTGTTACCAATTTT
CTTTTGTCTTTGGGCATACATTTAAACCCTAACAAGACGAAAAGATGGGGTTATT
CTCTACACTTCATGGGGTATGTCATTGGTAGTTGGGGATCATTGCCACAAGAGC
ACATTCTACACAAGCTGAAAGAATGTTTTAGAAAACTCCCTGTTAACAGGCCTAT
TGATTGGAAAGTATGTCAAAGAATTGTGGGTCTTTTGGGTTTTGCTGCCCCTTTT
ACACAATGTGGTTATCCTGCTTTAATGCCCTTATATGCATGTATTCAATCTAAAC

AGGCTTTTACTTTCTCGCCAACTTACAAGGCCTTTCTATGTAAACAATACATGAA
CCTTTACCCCGTTGCCCGGCAACGGTCAGGTCTGTGCCAAGTGTTTGCTGACG
CAACCCCCACTGGCTGGGGCTTGGTCATGGGCCATCAGCGCATGCGTGGAAC
CTTTCAGGCTCCTCTGCCGATCCATACTGCGGAACTCCTAGCCGCTTGTTTTGC
TCGCAGCAGGTCTGGAGCAAACATTCTCGGGACTGATAACTCTGTTGTTCTCTC
CCGCAAATATACATCGTTTCCATGGCTGCTAGGCTGTGCTGCCAACTGGATCCT
GCGCGGGACGTCCTTTGTCTACGTCCCGTCGGCGCTGAATCCCGCGGACGAC
CCGTCTCGGGGCCGCTTGGGTCTCGGTCGTCCTCTCATCCGTCTGCCGTTCCG
ACCGACCACGGGGCGCACCTCTCTTTACGCGGACTCCCCGTCTGTGCCTTCTC
ATCTGCCGGACCGTGTGCACTTCGCTTCACCTCTGCACGTCGCATGGAGACCA
CCGTGAACGCCCACCAAATCTTGCCCAAGATCTTATATAAGAGGACTCTTGGAC
TCTCCACAATGTCAACGACCGATCTTGAGCAATACTTCAAAGACTGTTTGTTTAA
GGACTGGGAGGAGTTGGGGGAGGAGATTAGGTTAATGATCTTTGTATTAGGAG
GCTGTAGGCATAAATTGGTCTGCGCACCAGCACCATGCAACTTTTTCACCTCTG
CCTAATCATCTTTTGTTCATGTCCTACTGTTCAAGCCTCCAAGCTGTGCCTTGGG
TGGCTTTAGGGCATGGACATTGACCCTTATAAAGAATTTGGAGCGACTGTGGAG
TTACTCTCATTTTTGCCTTCTGATTTCTTTCCGTCTGTACGAGATCTTCTAGACAC
CGCCGCAGCTCTATATCGGGATGCGTTAGAGTCTCCTGAGCATTGCTCACCTC
ACCATACTGCACTCAGGCAAGCAATTCTTTGCTGGGGAGAATTAATGACTCTAG
CTACCTGGGTGGGTACTAATTTGGAAGATCCAGCATCTAGGGACCTAGTGGTC
AGTTATGTCAACACTAACATGGGCCTAAAGTTCAGGCAACTATTGTGGTTTCAC
ATTTCTTGTCTTACTTTTGGGAGAGAAACGGTTTTGGAGTATTTGGTGTCTTTCG
GAGTGTGGATTCGCACTCCTCCAGCTTATAGACCACCAAATGCCCCTATCTTAT
CAACACTTCCGGAGACTACTGTTGTTAGACGACGAGGCAGGACCCCTAGAAGA
AGAACTCCCTCGCCTCGCAGACGAAGGTCTCAATCGCCGCGTCGCAGAAGATC
TCAATCTCGGGAATCCCAATGTTAATATTCCCTGGACTCATAAGGTGGGAAACT
TTACGGGGCTTTATTCTTCTACTGTACCTGTCTTTAACCCTCATTGGAAAACACC
AACTTTTCCTAACATACATTTACACCAAGACATCATCAACAAATGTGAACAGTTT
GTAGGCCCTCTTACTGTCAATGAGAAAAGAAGATTGCAATTGATTATGCCTGCT
AGGTTTTACCCTAAAGTTACCAAATATTTACCCTTAGATAAGGGTATTAAACCTT
ACTATCCAGAACATCTAGTTAATCATTACTTCCAAACCAGACACTATTTACACAC
TCTATGGAAGGCGGGTATTTTATATAAGAGAGAAACAACACATAGCGCCTCATT
TTGTGGGTCACCATATTCTTGGGAACAAAAGCTACAGCATGGGGCAGAATCTTT
CCACCAGCAATCCTCTGGGATTCTTTCCCGACCACCAGTTGGATCCAGCCTTCA
GAGCAAACACCAGCAATCCAGATTGGGACTTCAATCCCAACAAGGACACCTGG
CCAGACGCCAACAAGGTAGGAGCTGGAGCATTCGGGCCAGGGTTCACCCCAC
CGCACGGAGGCCTTTTGGGGTGGAGCCCGCAGGCTCAGGGCATACTACAAAC
CTTGCCAGCAAATCCGCCTCCTGCAGCTACCAATCGCCAGTCAGGAAGGCAGC
CTACCCCACTGACTCCACCTCTGAGAGACACTCATCCTCAGATGCAGTGGAC

INDIA-GOA

S NO.	GENOTYPE	ACCESSION NO.	REGION	NOMENCLATURE
1		X02496	GOA	G1

> X02496-HBV-GOA

AACTCCACAACCTTCCACCAAACTCTGCAAGATCCCAGAGTGAGAGGCCTGTAT
TTCCCTGCTGGTGGCTCCAGTTCAGGAACAGTAAACCCTGTTCCGACTACTGTC
TCTCCCATATCGTCAATCTTCTCGAGGATTGGGGACCCTGCGCTGAACATGGA
GAACATCACATCAGGATTCCTAGGACCCCTGCTCGTGTTACAGGCGGGGTTTTT
CTTGTTGACAAGAATCCTCACAATACCGCAGAGTCTAGACTCGTGGTGGACTTC
TCTCAATTTTCTAGGGGGAACTACCGTGTGTCTTGGCCAAAATTCGCAGTCCCC
AACCTCCAATCACTCACCAACCTCCTGTCCTCCAACTTGTCCTGGTTATCGCTG
GATGTGTCTGCGGCGTTTTATCATCTTCCTCTTCATCCTGCTGCTATGCCTCATC
TTCTTGTTGGTTCTTCTGGACTATCAAGGTATGTTGCCCGTTTGTCCTCTAATTC
CAGGATCTTCAACTACCAGCACGGGACCATGCAGAACCTGCACGACTCCTGCT
CAAGGAACCTCTATGTATCCCTCCTGTTGCTGTACCAAACCTTCGGACGGAAAT
TGCACCTGTATTCCCATCCCATCATCCTGGGCTTTCGGAAAATTCCTATGGGAG
TGGGCCTCAGCCCGTTTCTCCTGGCTCAGTTTACTAGTGCCATTTGTTCAGTGG
TTCGTAGGGCTTTCCCCCACTGTTTGGCTTTCAGTTATATGGATGATGTGGTATT
GGGGGCCAAGTCTGTACAGCATCTTGAGTCCCTTTTTACCGCTGTTACCAATTT
TCTTCTGTCTTTGGGTATACATTTAAACCCTAACAAAACAAAAAGATGGGGTTAC
TCTTTACATTTCATGGGCTATGTCATTGGATGTTATGGGTCATTGCCACAAGATC
ACATCATACAGAAAATCAAAGAATGCTTTAGAAAACTTCCTGTTAACAGGCCTAT
TGATTGGAAAGTCTGTCAACGTATTGTGGGTCTTTTGGGTTTTGCTGCCCCTTTT
ACACAATGTGGTTATCCTGCTTTAATGCCTTTGTATGCATGTATTCAGTCGAAGC
AGGCTTTTACTTTCTCGCCAACTTACAAGGCCTTTCTGTGTAAACAATACCTGAA

CCTTTACCCCGTTGCCCGGCAACGGCCAGGTCTGTGCCAAGTGTTTGCTGACG
CAACCCCCACTGGCTGGGGCTTGGTCATGGGCCATCAGCGCATGCGTGGAAC
CTTTCTGGCTCGTCTGCCGATCCATACTGCGGAACTCCTAGCCGCTTGTTTTGC
TCGCAGCAGGTCTGGAGCAAACATTCTCGGGACGGATAACTCTGTTGTTCTCTC
CCGCAAATATACATCGTATCCATGGCTGCTAGGCTGTGCTGCCAACTGGATCCT
GCGCGGGACGTCCTTTGTTTACGTCCCGTCGGCGCTGAATCCCGCGGACGAC
CCTTCTCGGGGTCGCTTGGGACTCTCTCGTCCCCTTCTCCGTCTGCCGTTTCGA
CCGACCACGGGGCGCACCTCTCTTTACGCGGACTCCCCGTCTGTGCCTTCTCA
TCTGCCGGACCGTGTGCACTTCGCTTCACCTCTGCACGTCGCATGGAGACCAC
CGTGAAAGCCCAACCATTCTTGCCCAAGGTCTTACATAAGAGGACTCTTGGACT
CTCTGTAATGTCAACGACCGACCTTGAGGCATACTTCAAAGACTGTTTGTTTAAA
GACTGGGAGGAGTTGGGGGAGGAGATTAGATTAAAGGTCTTTGTATTAGGAGG
CTGTAGGCATAAATTGGTCTGCGCACCAGCACCATGCAACTTTTTCACCTCTGC
CTAATCATCTCTTGTTCATGTCCTACTGTTCAAGCCTCCAAGCTGTGCCTTGGGT
GGCTTTGGGGCATGGACATTGATCCTTATAAAGAATTTGGAGCTACTGTGGAGT
TACTCTCGTTTTTGCCTTCTGACTTCTTTCCTTCAGTACGAGATCTTCTAGATAAC
GCCTCAGCTCTGTATCGGGAAGCCTTAGAGTCTCCTGAGCATTGTTCACCTCAC
CATACTGCACTCAGGCAAGCAATACTGTGCTGGGGGGAACTAATGACTCTAGC
TACCTGGGTGGGTGGTAATTTGGAAGATCCAATATCCAGGGACCTAGTAGTCA
GTTATGTCAACACTAATATGGGCCTAAAATTCAGGCAACTATTGTGGTTTCACAT
TTCTTGTCTCACTTTTGGAAGAGAAACAGTTATAGAGTATTTGGTGTCTTTTGGA
GTGTGGATTCGCACTCCTCCAGCTTATAGACCACCAAATGCCCCTATCTTATCA
ACACTTCCGGAGACTACTGTTGTTAGACGACGAGGCAGGTCCCCTAGAAGAAG
AACTCCCTCGCCTCGCAGACGAAGGTCTCAATCGCCGCGTCGCAGAAGATCTC
AATCTCGGGAATCTCAATGTTAGTATTCCTTGGACTCATAAGGTGGGAAACTTTA
CGGGGCTTTATTCTTCTACTGTACCTGTCTTTAACCCTCATTGGAAAACACCCTC
TTTTCCTAATATACATTTACACCAAGACATTATCAAAAAATGTGAACAATTTGTAG
GCCCACTCACAGTCAATGAGAAAAGAAGACTGCAATTGATTATGCCAGCTAGGT
TTTATCCAAATGTTACCAAATATTTGCCATTGGATAAGGGTATTAAACCTTATTAT
CCAGAATATTTAGTTAATCATTACTTCCAAACTAGACATTATTTACACACTCTATG
GAAGGCGGGTATATTATACAAGAGAGAAACTACACATAGCGCCTCATTTTGTGG
GTCACCATATTCTTGGGAACAAGAGCTACAGCATGGGGCAGAATCTTTCCACCA
GCAATCCTCTGGGATTCTTTCCCGACCACCAGTTGGATCCAGCCTTCAGAGCAA
ACACCGCAAATCCAGATTGGGACTTCAATCCCAACAAGGACACCTGGCCAGAC
GCCAACAAGGTAGGAGCTGGAGCATTCGGGCTGGGATTCACCCCACCACACG
GAGGCCTTTTGGGGTGGAGCCCTCAGGCTCAGGGCATACTAGAAACGTTGCCA
GCAAATCCGCCTCCTGCCTCTACCAATCGCCAGTCAGGAAGGCAGCCTACCCC
GCTGTCTCCACCTTTGAGAAACACTCATCCTCAGGCCATGCAGTGG

INDIA- DELHI- AIIMS

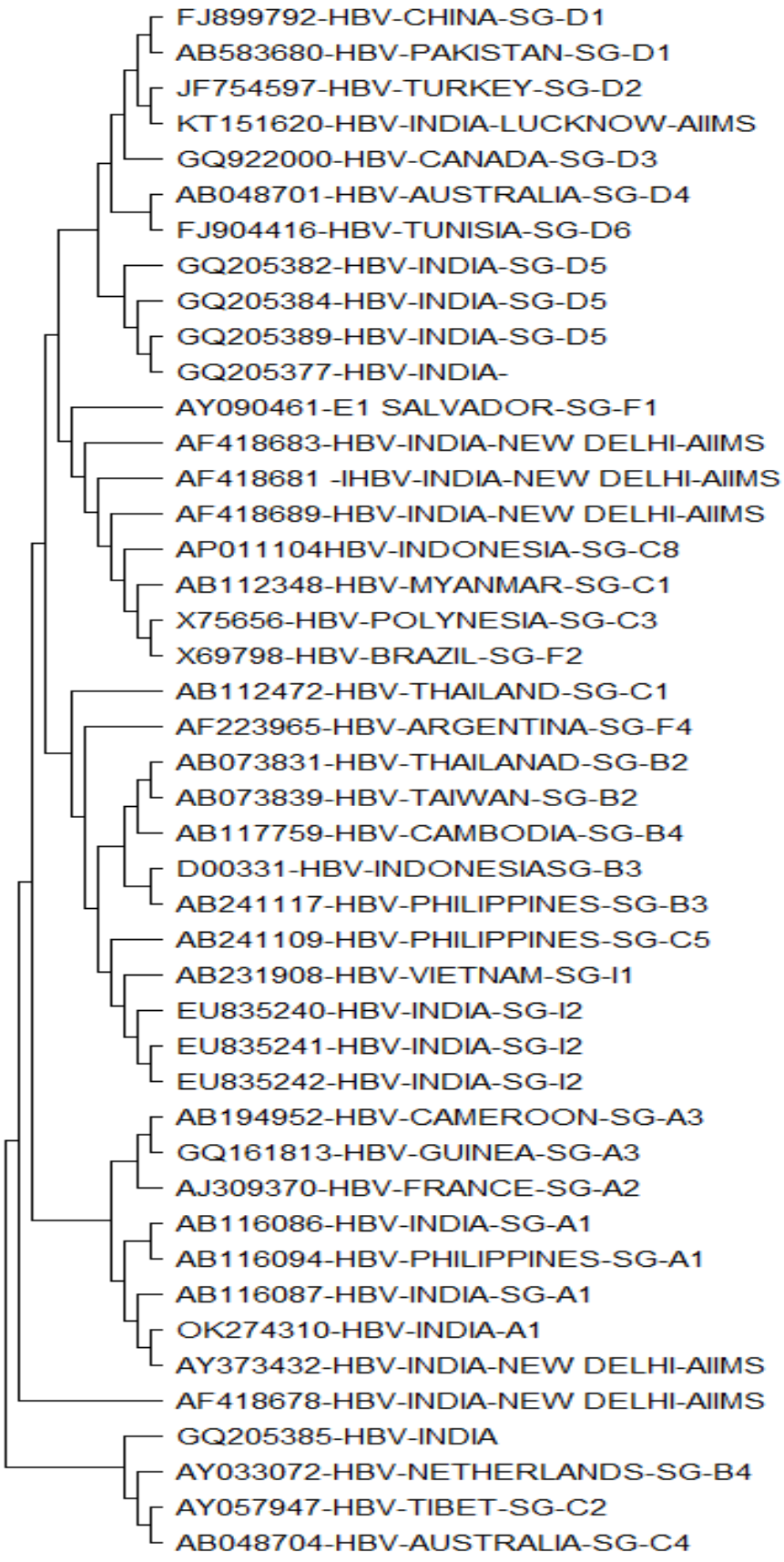
S NO.	GENOTYPE	ACCESSION NO.	REGION	NOMENCLATURE
1		AY373432	AIIMS	A1
2		AF418678	AIIMS	A2
3		AF418681	AIIMS	A3
4		AF418683	AIIMS	A4
5		AF418689	AIIMS	A5
6	D	KT151620	AIIMS	A6

> AY373432-HBV-INDIA-NEW DELHI-AIIMS

AATTCCAGGATCAACAACAACCAGCACGGGACCCTGCAAAACCTGCACGACTC
CTGCTCAAGGCAACTCTATGTTTCCCTCATGTTGCTGTACAAAACCTACGGATG
GAAATTGCACCTGTATTCCCATCCCATCATCTTGGGCTTTCGCAAAATACCTATG
GGAGTGGGCCTCAGTCCGTTTCTCTTGGCTCAGTTTACTAGTGCCATTTGTTCA
GTGGTTCGTAGGGCTTTCCCCCACTGTTTGGCTTTCAGCTATATGGATGATGTG
GTATTGGGGGCCAAGTCTGTACAACATCTTGAGTCCCTTTATACCGCTGTTACC
AATTTTCTTTTGTCTTTGGGCATACATTTAAACCCTAACAAAACAAAGAGATGGG
GTTACTCCCTAAATTTTATGGGTTATGTAATTGGAAGTTGGGGTACATTGCCACA
GGATCATATTGTACAAAAAATCAAACACTGTTTTAGAAAACTTCCTGTCAATCGA
CCTATTGATTGGAAAGTATGTCARAGAATTGTGGGTCTTTTGGGCTTTGCCGCT
CCATTTACACAATGTGGGTACCCTGCATTAATGCCTTTGTATGCATGTATACAAG

CGAAACAGGCTTTTACTTTCTCGCCAACTTACAAGGCCTTTCTAAGTAAACAGTA
TATGAACCTTTACCCCGTTGCACGGCAACGGCCTGGTCTGTGCCAAGTGTTTG
CTGACGCAACCCCCACTGGCTGGGGCTTGGCCATCGGCCATCAGCGCATGCG
TGGAACCTTTGTGGCTCCTCTGCCGATCCATACTGCGGAACTCCTAGCCGCTT
GTTTTGCTCGCAGCCGGTCTGGAGAAAAACTCATCGGGACTGATAATTCTGTCG
TCCTTTCTCGGAAATATACATCATTTCCATGGCTGCTAGGTTGTACTGCCAACTG
GATTCTTCGCGGGACGTCCTTTGTTTACGTCCCGTCGGCGCTGAATCCCGCGG
ACGACCCCTCGCGAGGCCGCTTGGGGCTGTATCGTCCCCTTCTCCGTCTGCCG
TACCGTCCGACCACGGGGCGCACCTCTCTTTACGCGGTCTCCCCGTCTGTGCC
TTCTCATCTGCCGGTCCGTGTGCACTTCGCTTCACCTCTGCACGTTGCATGGAG
ACCACCGTGAACGCCCATCAGGTCCTGCCCAAGGTCTTACATAAGAGGACTCT
TGGACTCTCAGCAATGTCAACGACCGACCTTGAGGCCTACTTCAAAGACTGTGT
GTTTAAAGACTGGGAGGAGTTGGGGGAGGAGATTAGGTTAATGATCTTTGTATT
AGGAGGCTGTAGGCATAAATTGGTCTGCGCACCATCATCATGCAACTTTTTCAC
CTCTGCCTAATCATCTCTTGTACATGTCCCACTTTTCAAGCCTCCAAGCTGTGCC
TTGGATGGCTTTGGGGCATGGACATTGACCCTTATAAAGAATTTGGAGCTACTG
TGGAGTTACTCTCATTTTTGCCTTCTGACTTCTTTCCTTCCGTCCGGGATCTACT
AGATACAGCCTCAGCTCTATATCGGGAAGCCTTAGAGTCTCCTGAGCATTGCTC
ACCTCACCATACAGCACTCAGGCAAGCCATTCTCTGCTGGGGGGAATTAATGA
CTCTAGCTACCTGGGTGGGTAATAATTTGGAAGATCCAGCATCCAGGGATCTAG
TAGTCAATTATGTTAATACTAACATGGGCCTAAAGATCAGGCAATTATTGTGGTT
TCATATTTCTTGCCTTACTTTTGGAAGAGAAACTGTCCTTGAGTATTTGGTCTCTT
TCGGAGTGTGGATTCGCACTCCTCCAGCCTATAGACCACCAAATGCCCCTATCT
TATCAACACTTCCGGAAACTACTGTTGTTAGACGACGAGACCGAGGCAGGTCC
CCTAGAAGAAGAACTCCCTCGCCTCGCAGACGAAGATCTCAATCGCCGCGTCG
CAGAAGATCTCAATCTCGGGAATCTCAATGTTAGTATTCCTTGGACTCATAAGGT
GGGAAATTTTACTGGACTTTATTCTTCTACTGTCCCTATCTTTAATCCTGAATGG
CAAACTCCTTCTTTTCCTAAAATTCATTTACATGAAGACATTATTAATAGGTGTCA
GCAATTTGTAGGCCCTCTCACTGTAAATGAAAAAAAAAGATTGAACTTAATTATG
CCTGCTAGGTTGTATCCTATCAGCACGAGATATTTGCCTTTACACAAAGGGATT
AAACCTTATTATCCTGATCTTGTAGTTAATCATTACTTTCAAACCCGACATTATTT
ACATACTCTTTGGAAGGCTGGGATTCTATATAAGAGGGAAACTACACGTAGCGC
CTCATTTTGCGGGTCACCATATTCTTGGGAACAAGAGCTACATCATGGGAGGTT
GGTCATCAAAACCTCGCAAAGGCATGGGGACGAATCTTTCTGTTCCCAACCCTC
TGGGATTCTTTCCCGATCATCAGTTGGACCCTGCATTCGGAGCCAATTCAAACA
ATCCAGATTGGGACTTCAACCCCATCAAGGACCACTGGCCACAAGCCAACCAG
GTAGGAGTGGGAGCATTCTGGCCAGGGCTCACTCCCCCACACGGAGGTGTTTT
GGGGTGGAGCCCTCACGCTCAGGGCATATTGGCCACAGCGCCAGCAGTGCCT
CCTCCTGCCTCCACCAATCGGCAGTCAGGAAGGCAGCCTACTCCCATATCTCC
ACCTCTAAGAGACAGTCATCCTCAGGCCATGCAGTGGAATTCCACTGCCTTCCA
CCAAGCTCTGCAGAATCCCAAAGACAGGGGTCTGTATCTTCCTGCTGGTGGCT
CCAGTTCAGGAACACTCAACCCTGTTCCAACTATTGCCTCTCACATATCGTCAAT
CTCCTCGAGGATTGGGGACCCTGCGACGAACATGGAGAACATCACATCAGGAT
TCCTAGGACCCCTGCTCGTGTTACAGGCGGGGTTTTTCTTGTTGACAAGAATCC
TCACAATACCGCAGAGTCTAGACTCGTGGTGGACTTCTCTCAATTTTCTAGGGG
GATCACCCGTGTGTCTTGGCCAAAATTCGCAGTCCCCAACCTCCAATCACTCAC
CAACCTCCTGTCCTCCAATTTGTCCTGGTTATCGCTGGATGTGTCTGCGGCGTT
TTATCATATTCCTCTTCATCCTGCTGCTATGCCTCATCTTCTTATTGGTTCTTCTG
GATTATCAAGGTATGTTGCCCGTTTGTCCTCT

PHYLOGENETIC ANALYSIS OF WHOLE GENOME OF HBV

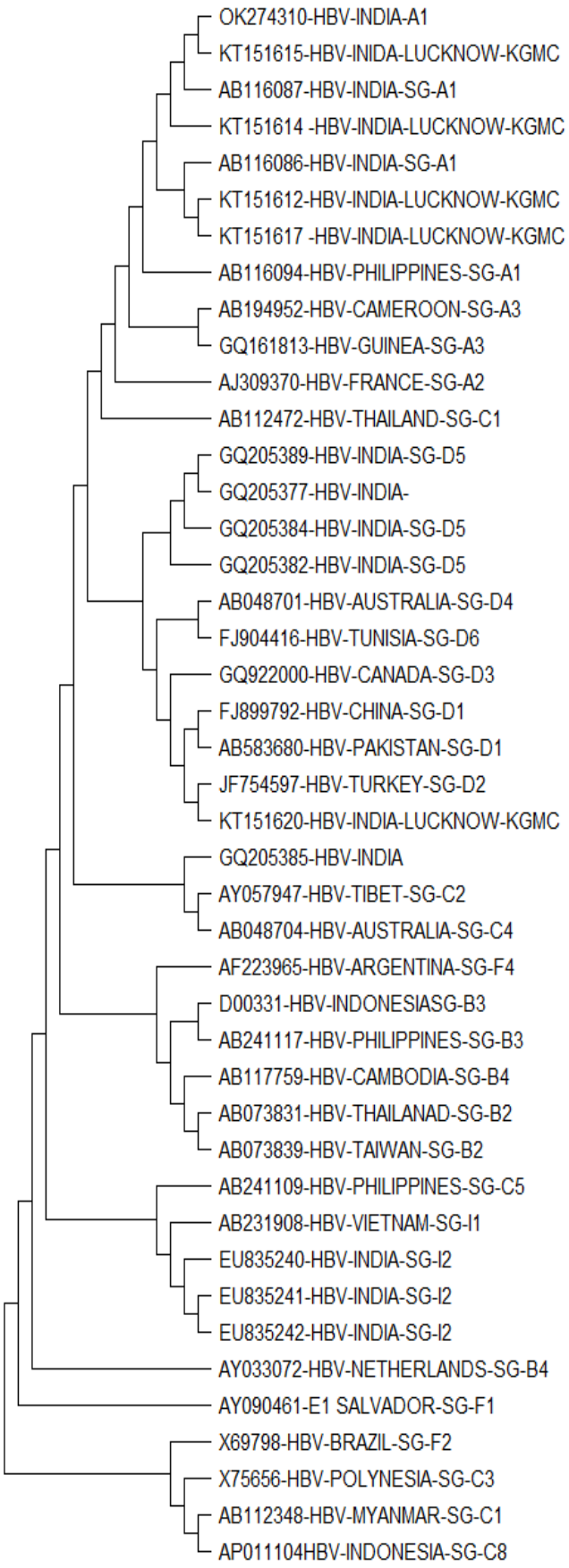

FJ899792-HBV-CHINA-SG-D1
AB583680-HBV-PAKISTAN-SG-D1
JF754597-HBV-TURKEY-SG-D2
KT151620-HBV-INDIA-LUCKNOW-AIIMS
GQ922000-HBV-CANADA-SG-D3
AB048701-HBV-AUSTRALIA-SG-D4
FJ904416-HBV-TUNISIA-SG-D6
GQ205382-HBV-INDIA-SG-D5
GQ205384-HBV-INDIA-SG-D5
GQ205389-HBV-INDIA-SG-D5
GQ205377-HBV-INDIA-
AY090461-E1 SALVADOR-SG-F1
AF418683-HBV-INDIA-NEW DELHI-AIIMS
AF418681 -IHBV-INDIA-NEW DELHI-AIIMS
AF418689-HBV-INDIA-NEW DELHI-AIIMS
AP011104HBV-INDONESIA-SG-C8
AB112348-HBV-MYANMAR-SG-C1
X75656-HBV-POLYNESIA-SG-C3
X69798-HBV-BRAZIL-SG-F2
AB112472-HBV-THAILAND-SG-C1
AF223965-HBV-ARGENTINA-SG-F4
AB073831-HBV-THAILANAD-SG-B2
AB073839-HBV-TAIWAN-SG-B2
AB117759-HBV-CAMBODIA-SG-B4
D00331-HBV-INDONESIASG-B3
AB241117-HBV-PHILIPPINES-SG-B3
AB241109-HBV-PHILIPPINES-SG-C5
AB231908-HBV-VIETNAM-SG-I1
EU835240-HBV-INDIA-SG-I2
EU835241-HBV-INDIA-SG-I2
EU835242-HBV-INDIA-SG-I2
AB194952-HBV-CAMEROON-SG-A3
GQ161813-HBV-GUINEA-SG-A3
AJ309370-HBV-FRANCE-SG-A2
AB116086-HBV-INDIA-SG-A1
AB116094-HBV-PHILIPPINES-SG-A1
AB116087-HBV-INDIA-SG-A1
OK274310-HBV-INDIA-A1
AY373432-HBV-INDIA-NEW DELHI-AIIMS
AF418678-HBV-INDIA-NEW DELHI-AIIMS
GQ205385-HBV-INDIA
AY033072-HBV-NETHERLANDS-SG-B4
AY057947-HBV-TIBET-SG-C2
AB048704-HBV-AUSTRALIA-SG-C4

AIIMS

The phylogenetic analysis of AIIMS give following information

Accession no. KT151620 of unknown genotype and accession no. JF754597 of known genotypes follow the same path and they originate from the same node and are closely related to each other.
Hence genotype of KT151620 is D and subgenotype is D2

Accession no. AY373432 of unknown genotype and accession no. OK274310 of known genotype shows similarity and are homologous so unknown genotype is A and subgenotype is A1


OK274310-HBV-INDIA-A1
KT151615-HBV-INIDA-LUCKNOW-KGMC
AB116087-HBV-INDIA-SG-A1
KT151614 -HBV-INDIA-LUCKNOW-KGMC
AB116086-HBV-INDIA-SG-A1
KT151612-HBV-INDIA-LUCKNOW-KGMC
KT151617 -HBV-INDIA-LUCKNOW-KGMC
AB116094-HBV-PHILIPPINES-SG-A1
AB194952-HBV-CAMEROON-SG-A3
GQ161813-HBV-GUINEA-SG-A3
AJ309370-HBV-FRANCE-SG-A2
AB112472-HBV-THAILAND-SG-C1
GQ205389-HBV-INDIA-SG-D5
GQ205377-HBV-INDIA-
GQ205384-HBV-INDIA-SG-D5
GQ205382-HBV-INDIA-SG-D5
AB048701-HBV-AUSTRALIA-SG-D4
FJ904416-HBV-TUNISIA-SG-D6
GQ922000-HBV-CANADA-SG-D3
FJ899792-HBV-CHINA-SG-D1
AB583680-HBV-PAKISTAN-SG-D1
JF754597-HBV-TURKEY-SG-D2
KT151620-HBV-INDIA-LUCKNOW-KGMC
GQ205385-HBV-INDIA
AY057947-HBV-TIBET-SG-C2
AB048704-HBV-AUSTRALIA-SG-C4
AF223965-HBV-ARGENTINA-SG-F4
D00331-HBV-INDONESIASG-B3
AB241117-HBV-PHILIPPINES-SG-B3
AB117759-HBV-CAMBODIA-SG-B4
AB073831-HBV-THAILANAD-SG-B2
AB073839-HBV-TAIWAN-SG-B2
AB241109-HBV-PHILIPPINES-SG-C5
AB231908-HBV-VIETNAM-SG-I1
EU835240-HBV-INDIA-SG-I2
EU835241-HBV-INDIA-SG-I2
EU835242-HBV-INDIA-SG-I2
AY033072-HBV-NETHERLANDS-SG-B4
AY090461-E1 SALVADOR-SG-F1
X69798-HBV-BRAZIL-SG-F2
X75656-HBV-POLYNESIA-SG-C3
AB112348-HBV-MYANMAR-SG-C1
AP011104HBV-INDONESIA-SG-C8

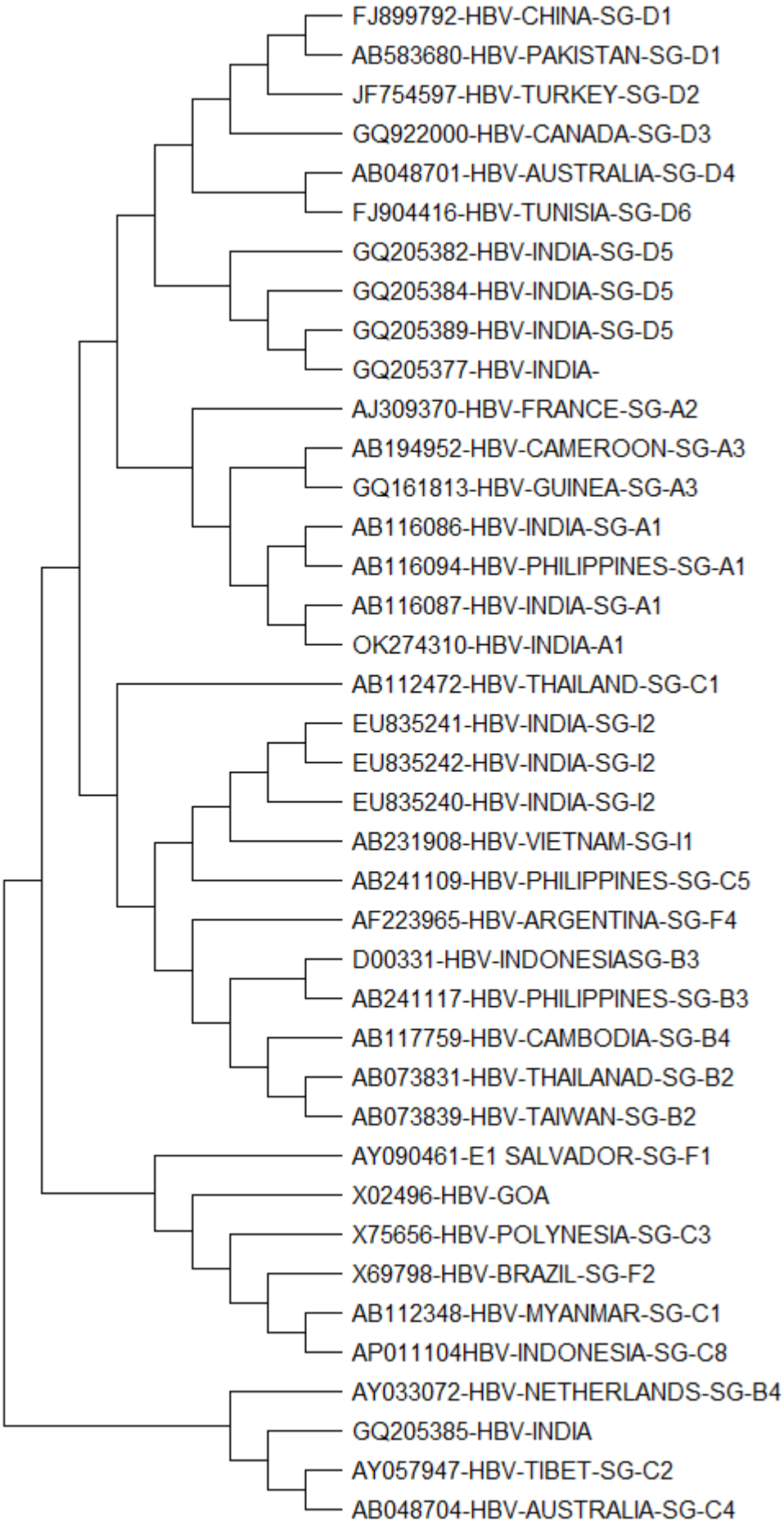
KGMU-

Unknown genotype and subgenotype (KT151615) of KGMU is A and A1 respectively on seeing the subgenotype of OK274310 from the phylogenetic tree, as their nodes are common and their tips are closely related and show similarity between them.

Accession no. KT151614 have genotype A, and subgenotypes of the same may be A1,A2, and A3.

Accession no. KT151612 & KT151617 are similar from the phylogenetic analysis and they have genotype A and subgenotype may be A1 as they both show similarity to Accession no. AB116086.

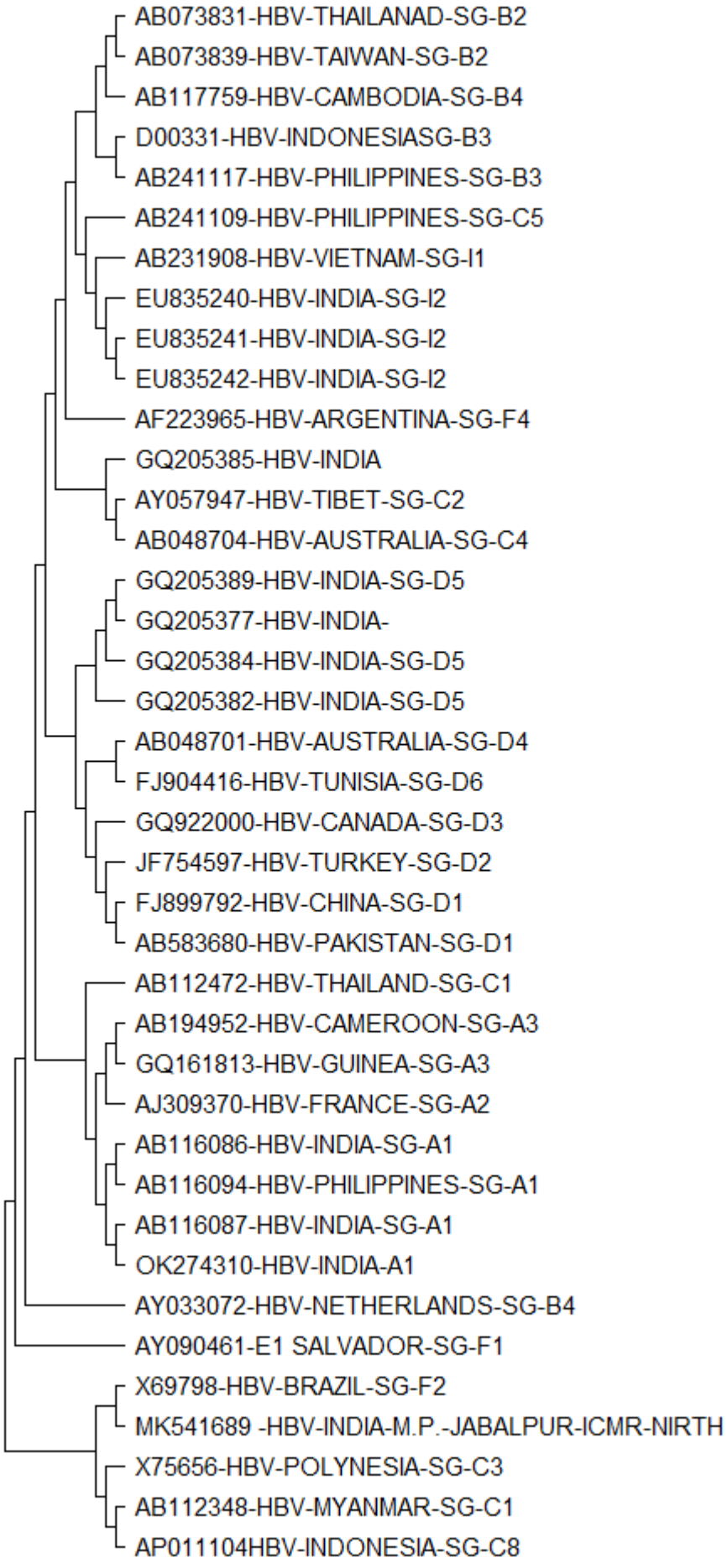
Accession no. KT151620 shows similarity with accession no. JF754597 is from the phylogenetic tree so its genotype is D and subgenotype is D2.


FJ899792-HBV-CHINA-SG-D1
AB583680-HBV-PAKISTAN-SG-D1
JF754597-HBV-TURKEY-SG-D2
GQ922000-HBV-CANADA-SG-D3
AB048701-HBV-AUSTRALIA-SG-D4
FJ904416-HBV-TUNISIA-SG-D6
GQ205382-HBV-INDIA-SG-D5
GQ205384-HBV-INDIA-SG-D5
GQ205389-HBV-INDIA-SG-D5
GQ205377-HBV-INDIA-
AJ309370-HBV-FRANCE-SG-A2
AB194952-HBV-CAMEROON-SG-A3
GQ161813-HBV-GUINEA-SG-A3
AB116086-HBV-INDIA-SG-A1
AB116094-HBV-PHILIPPINES-SG-A1
AB116087-HBV-INDIA-SG-A1
OK274310-HBV-INDIA-A1
AB112472-HBV-THAILAND-SG-C1
EU835241-HBV-INDIA-SG-I2
EU835242-HBV-INDIA-SG-I2
EU835240-HBV-INDIA-SG-I2
AB231908-HBV-VIETNAM-SG-I1
AB241109-HBV-PHILIPPINES-SG-C5
AF223965-HBV-ARGENTINA-SG-F4
D00331-HBV-INDONESIASG-B3
AB241117-HBV-PHILIPPINES-SG-B3
AB117759-HBV-CAMBODIA-SG-B4
AB073831-HBV-THAILANAD-SG-B2
AB073839-HBV-TAIWAN-SG-B2
AY090461-E1 SALVADOR-SG-F1
X02496-HBV-GOA
X75656-HBV-POLYNESIA-SG-C3
X69798-HBV-BRAZIL-SG-F2
AB112348-HBV-MYANMAR-SG-C1
AP011104HBV-INDONESIA-SG-C8
AY033072-HBV-NETHERLANDS-SG-B4
GQ205385-HBV-INDIA
AY057947-HBV-TIBET-SG-C2
AB048704-HBV-AUSTRALIA-SG-C4

GOA
As we can observe in the phylogenetic tree

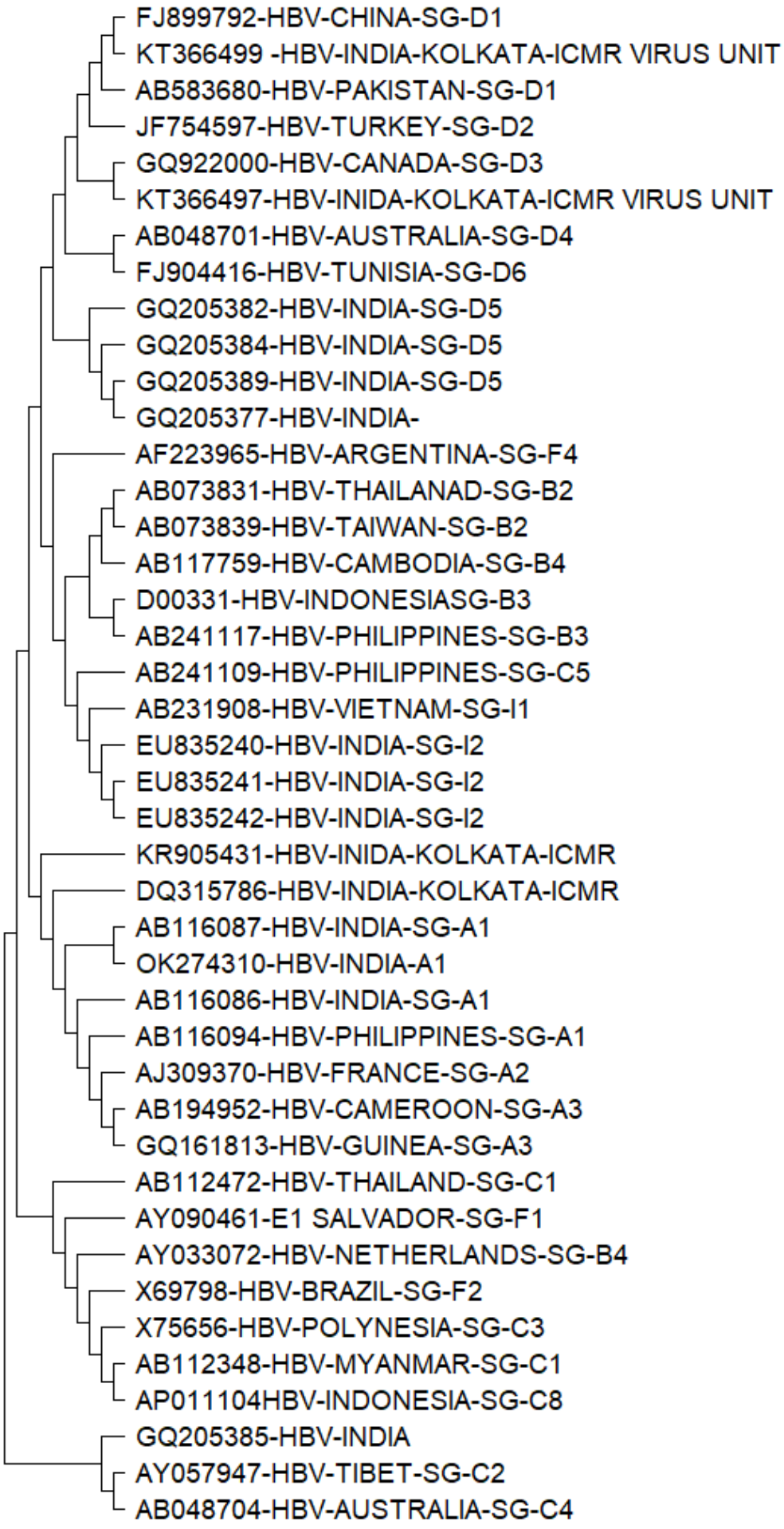
Certain genotypes are also unknown.so from the concept of homology we can say that the genotype of accession no. GQ205377 is D5, on comparing with genotype of known accession no.GQ205389
And genotype of accession no. X02496 may be C or F,
from the phylogenetic tree .

An unknown genotype of Goa of accession no.GQ205385 may be B or C.


AB073831-HBV-THAILANAD-SG-B2
AB073839-HBV-TAIWAN-SG-B2
AB117759-HBV-CAMBODIA-SG-B4
D00331-HBV-INDONESIASG-B3
AB241117-HBV-PHILIPPINES-SG-B3
AB241109-HBV-PHILIPPINES-SG-C5
AB231908-HBV-VIETNAM-SG-I1
EU835240-HBV-INDIA-SG-I2
EU835241-HBV-INDIA-SG-I2
EU835242-HBV-INDIA-SG-I2
AF223965-HBV-ARGENTINA-SG-F4
GQ205385-HBV-INDIA
AY057947-HBV-TIBET-SG-C2
AB048704-HBV-AUSTRALIA-SG-C4
GQ205389-HBV-INDIA-SG-D5
GQ205377-HBV-INDIA-
GQ205384-HBV-INDIA-SG-D5
GQ205382-HBV-INDIA-SG-D5
AB048701-HBV-AUSTRALIA-SG-D4
FJ904416-HBV-TUNISIA-SG-D6
GQ922000-HBV-CANADA-SG-D3
JF754597-HBV-TURKEY-SG-D2
FJ899792-HBV-CHINA-SG-D1
AB583680-HBV-PAKISTAN-SG-D1
AB112472-HBV-THAILAND-SG-C1
AB194952-HBV-CAMEROON-SG-A3
GQ161813-HBV-GUINEA-SG-A3
AJ309370-HBV-FRANCE-SG-A2
AB116086-HBV-INDIA-SG-A1
AB116094-HBV-PHILIPPINES-SG-A1
AB116087-HBV-INDIA-SG-A1
OK274310-HBV-INDIA-A1
AY033072-HBV-NETHERLANDS-SG-B4
AY090461-E1 SALVADOR-SG-F1
X69798-HBV-BRAZIL-SG-F2
MK541689 -HBV-INDIA-M.P.-JABALPUR-ICMR-NIRTH
X75656-HBV-POLYNESIA-SG-C3
AB112348-HBV-MYANMAR-SG-C1
AP011104HBV-INDONESIA-SG-C8

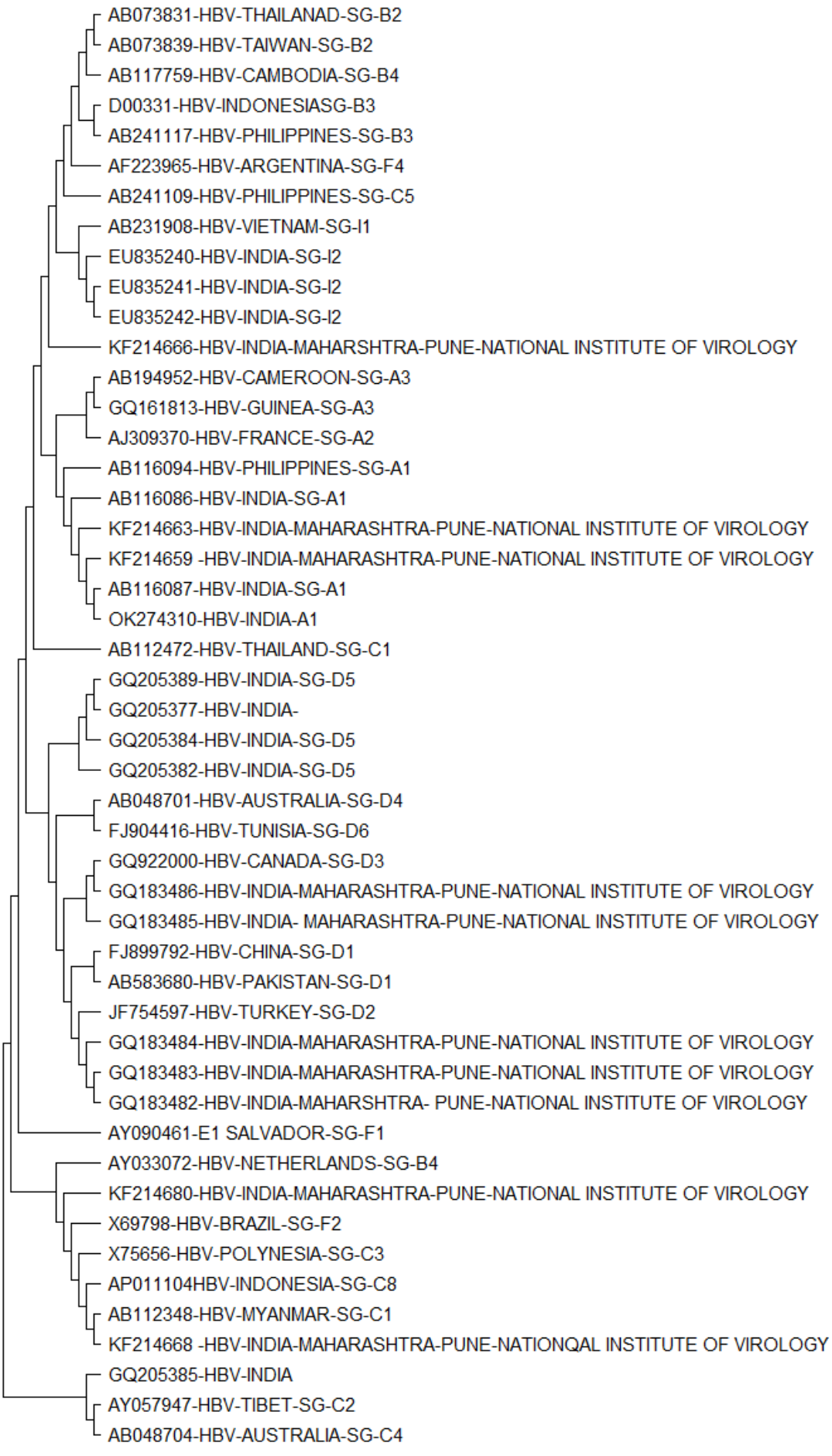
JABALPUR

As we can observe from the tree that the tips of both known and unknown genotypes are same, so they show similarity, so unknown genotype of accession no. MK541689 is F2


FJ899792-HBV-CHINA-SG-D1
KT366499 -HBV-INDIA-KOLKATA-ICMR VIRUS UNIT
AB583680-HBV-PAKISTAN-SG-D1
JF754597-HBV-TURKEY-SG-D2
GQ922000-HBV-CANADA-SG-D3
KT366497-HBV-INIDA-KOLKATA-ICMR VIRUS UNIT
AB048701-HBV-AUSTRALIA-SG-D4
FJ904416-HBV-TUNISIA-SG-D6
GQ205382-HBV-INDIA-SG-D5
GQ205384-HBV-INDIA-SG-D5
GQ205389-HBV-INDIA-SG-D5
GQ205377-HBV-INDIA-
AF223965-HBV-ARGENTINA-SG-F4
AB073831-HBV-THAILANAD-SG-B2
AB073839-HBV-TAIWAN-SG-B2
AB117759-HBV-CAMBODIA-SG-B4
D00331-HBV-INDONESIASG-B3
AB241117-HBV-PHILIPPINES-SG-B3
AB241109-HBV-PHILIPPINES-SG-C5
AB231908-HBV-VIETNAM-SG-I1
EU835240-HBV-INDIA-SG-I2
EU835241-HBV-INDIA-SG-I2
EU835242-HBV-INDIA-SG-I2
KR905431-HBV-INIDA-KOLKATA-ICMR
DQ315786-HBV-INDIA-KOLKATA-ICMR
AB116087-HBV-INDIA-SG-A1
OK274310-HBV-INDIA-A1
AB116086-HBV-INDIA-SG-A1
AB116094-HBV-PHILIPPINES-SG-A1
AJ309370-HBV-FRANCE-SG-A2
AB194952-HBV-CAMEROON-SG-A3
GQ161813-HBV-GUINEA-SG-A3
AB112472-HBV-THAILAND-SG-C1
AY090461-E1 SALVADOR-SG-F1
AY033072-HBV-NETHERLANDS-SG-B4
X69798-HBV-BRAZIL-SG-F2
X75656-HBV-POLYNESIA-SG-C3
AB112348-HBV-MYANMAR-SG-C1
AP011104HBV-INDONESIA-SG-C8
GQ205385-HBV-INDIA
AY057947-HBV-TIBET-SG-C2
AB048704-HBV-AUSTRALIA-SG-C4

KOLKATA
As observed from the phylogenetic tree that the tip of both the known and unknown genotype is same,they are homologous and show close similarities,
so unknown genotype of accession no.KT366499 is D1
Accession no. KT366497 is D3

On observing tree some subgenotypes are still not confirmed but still it can be said that Unknown Genotypes of these are
Accession no. KR905431 -A
Accession no. DQ315786 -A


AB073831-HBV-THAILANAD-SG-B2
AB073839-HBV-TAIWAN-SG-B2
AB117759-HBV-CAMBODIA-SG-B4
D00331-HBV-INDONESIASG-B3
AB241117-HBV-PHILIPPINES-SG-B3
AF223965-HBV-ARGENTINA-SG-F4
AB241109-HBV-PHILIPPINES-SG-C5
AB231908-HBV-VIETNAM-SG-I1
EU835240-HBV-INDIA-SG-I2
EU835241-HBV-INDIA-SG-I2
EU835242-HBV-INDIA-SG-I2
KF214666-HBV-INDIA-MAHARSHTRA-PUNE-NATIONAL INSTITUTE OF VIROLOGY
AB194952-HBV-CAMEROON-SG-A3
GQ161813-HBV-GUINEA-SG-A3
AJ309370-HBV-FRANCE-SG-A2
AB116094-HBV-PHILIPPINES-SG-A1
AB116086-HBV-INDIA-SG-A1
KF214663-HBV-INDIA-MAHARASHTRA-PUNE-NATIONAL INSTITUTE OF VIROLOGY
KF214659 -HBV-INDIA-MAHARASHTRA-PUNE-NATIONAL INSTITUTE OF VIROLOGY
AB116087-HBV-INDIA-SG-A1
OK274310-HBV-INDIA-A1
AB112472-HBV-THAILAND-SG-C1
GQ205389-HBV-INDIA-SG-D5
GQ205377-HBV-INDIA-
GQ205384-HBV-INDIA-SG-D5
GQ205382-HBV-INDIA-SG-D5
AB048701-HBV-AUSTRALIA-SG-D4
FJ904416-HBV-TUNISIA-SG-D6
GQ922000-HBV-CANADA-SG-D3
GQ183486-HBV-INDIA-MAHARASHTRA-PUNE-NATIONAL INSTITUTE OF VIROLOGY
GQ183485-HBV-INDIA- MAHARASHTRA-PUNE-NATIONAL INSTITUTE OF VIROLOGY
FJ899792-HBV-CHINA-SG-D1
AB583680-HBV-PAKISTAN-SG-D1
JF754597-HBV-TURKEY-SG-D2
GQ183484-HBV-INDIA-MAHARASHTRA-PUNE-NATIONAL INSTITUTE OF VIROLOGY
GQ183483-HBV-INDIA-MAHARASHTRA-PUNE-NATIONAL INSTITUTE OF VIROLOGY
GQ183482-HBV-INDIA-MAHARSHTRA- PUNE-NATIONAL INSTITUTE OF VIROLOGY
AY090461-E1 SALVADOR-SG-F1
AY033072-HBV-NETHERLANDS-SG-B4
KF214680-HBV-INDIA-MAHARASHTRA-PUNE-NATIONAL INSTITUTE OF VIROLOGY
X69798-HBV-BRAZIL-SG-F2
X75656-HBV-POLYNESIA-SG-C3
AP011104HBV-INDONESIA-SG-C8
AB112348-HBV-MYANMAR-SG-C1
KF214668 -HBV-INDIA-MAHARASHTRA-PUNE-NATIONQAL INSTITUTE OF VIROLOGY
GQ205385-HBV-INDIA
AY057947-HBV-TIBET-SG-C2
AB048704-HBV-AUSTRALIA-SG-C4

PUNE

As observed from the phylogenetic tree that the tip of both the known and unknown genotypes is same, they are homologous and show close similarities,

So unknown subgenotype of

Accession no. GQ183486 is D3

Accession no. KF214668 is C1

But subgenotype is still not clear for certain unknown sequences

But their genotype may be

Accession no. GQ183482 -D

CONCLUSION

In this whole work we performed Phylogenetic analysis of sequences obtained from NCBI with the help of MEGA 11 software, which shows that following information can only be taken out when the sequences show close similarity among the known sequences. If they are homologous that means they derive from the same ancestor so most probably their genotype will also be similar.

After obtaining the phylogenetic tree for different regions of INDIA we conclude that there are various strains of the same type of HBV and their symptoms and treatment may also be the same even if they are in circulation in different regions of INDIA. Knowing the genotypes further, pharmaceuticals and researchers can work on the treatment and prevention of the following disease , because Chronic HBV is a significant public health concern worldwide. The natural history and outcome of the infection depend on several viral, host, and external factors. Individuals with chronic HBV are at increased risk of developing cirrhosis or HCC; effective treatment interventions need to be considered and if possible implemented *before* the development of decompensated cirrhosis or HCC. Prevention via vaccination is a central strategy to reduce the future impact of the disease**.**

This work demonstrated that the natural course of HBV infection may vary by geographic area and that these differences are associated predominantly with HBV genotypes as well as other viral and host factors. However the relationship between HBV genotypes and clinical course of disease has thus far been determined largely from clinic-based studies, rather than population based studies. This may potentially lead to selection bias, since patients with more advanced diseases are more likely to visit hospitals.

To conclude we ascertain that accurate determination of HBV genotypes and subgenotypes will become increasingly important in our understanding of HBV epidemiology and its clinical implications.

REFERENCES

1.Ganem, D.; Schneider, R.J. Hepadnaviridae: the viruses and their replication. In Fields Virology, 4th ed; Knipe, D.M., Howley, P.M., Eds.; 2001; Lippincott Williams: Wilkins, Philadelphia, USA. [Google Scholar]

2. Lupberger, J.; Hildt, E. Hepatitis B virus-induced oncogenesis. World J. Gastroenterol. 2007, 13, 74–81. [Google Scholar] [PubMed]

3.Hepatitis B: Fact sheets WHO.[http://www.who.int/mediacentre/factsheets/fs204/en]

4.Locarnini, S. Molecular virology of hepatitis B virus. Semin. Liver Dis. 2004, 24, 3–10. [Google Scholar] [CrossRef] [PubMed]

5.Lupberger, J.; Hildt, E. Hepatitis B virus-induced oncogenesis. World J. Gastroenterol. 2007, 13, 74–81. [Google Scholar] [PubMed]

6.Cougot, D.; Neuveut, C.; Buendia, M.A. HBV induced carcinogenesis. J. Clin. Virol. 2005, 34, S75–S78. [Google Scholar] [CrossRef] [PubMed]

7.Ganem, D.; Schneider, R.J. Hepadnaviridae: the viruses and their replication. In Fields Virology, 4th ed; Knipe, D.M., Howley, P.M., Eds.; 2001; Lippincott Williams: Wilkins, Philadelphia, USA. [Google Scholar]

8.Gerlich, W.H.; Heermann, K.H.; Lu, X. Functions of hepatitis B surface proteins. Arch. Virol. Suppl. 1992, 4, 129–132. [Google Scholar] [PubMed]

9.Lu, X.; Mehta, A.; Dwek, R.; Butters, T.; Block, T. Evidence that N-linked glycosylation is necessary for hepatitis B virus secretion. Virology 1995, 213, 660–665. [Google Scholar] [CrossRef] [PubMed]

10.Schmitt, S.; Glebe, D.; Tolle, T.K.; Lochnit, G.; Linder, D.; Geyer, R.; Gerlich, W.H. Structure of pre-S2 N- and O-linked glycans in surface proteins from different genotypes of hepatitis B virus. J. Gen. Virol. 2004, 85, 2045–2053. [Google Scholar] [CrossRef] [PubMed]

11. Eble, B.E.; Lingappa, V.R.; Ganem, D. The N-terminal (pre-S2) domain of a hepatitis B virus surface glycoprotein is translocated across membranes by downstream signal sequences. J. Virol. 1990, 64, 1414–1419. [Google Scholar] [PubMed]

12.Hedges SB, Kumar S. Precision of molecular time estimates. Trends Genet. 2004;20:242–247. [PubMed] [Google Scholar]

13.Kumar S, Hedges SB. Timetree2: Species divergence times on the iPhone. Bioinformatics. 2011;27:2023–2024. [PMC free article] [PubMed] [Google Scholar]

14.Kumar S, Stecher G, Peterson D, Tamura K. MEGA-CC: computing core of molecular evolutionary genetics analysis program for automated and iterative data analysis. Bioinformatics. 2012;28:2685–2686. [PMC free article] [PubMed] [Google Scholar]
15.Kumar S, Tamura K, Nei M. MEGA: molecular evolutionary genetics analysis software for microcomputers. Comput Appl Biosci. 1994;10:189–191. [PubMed] [Google Scholar]

16. Gerlich WH. Medical virology of hepatitis B: how it began and where we are now. Virol J. 2013; 10: 239.

17. Schädlers, Hildt E. HBV life cycle: entry and morphogenesis. Viruses. 2009; 1:185-209.

18. Zhang X, Hou J, Lu M. Regulation of hepatitis B virus replication by epigenetic mechanisms and microRNAs. Front Genet. 2013; 14:202.

19. Robinson BJ, Pierson SLL. Clinical Virology. In: Mahon CR, Lehman DC, Manuselis G (eds.), Textbook of Diagnostic Microbiology. 4th ed. China: Saunders; 2011. p.732.

20. Ay P, Torunoglu MA, Com S, Çipil Z, Mollahaliloğlu S, Erkoc Y, Dilman U. Trends of hepatitis B notification rates in Turkey, 1990 to 2012. Euro Surveill. 2013; 18 pii:20636.

21. Uhl P, Fricker G, Haberkorn U, Mier W. Current status in the therapy of liver diseases. Int J Mol Sci. 2014; 15:7500-7512.

22. Bertoletti A. Gehring AJ. Immune therapeutic strategies in chronic hepatitis B virus infection: virus or inflammation control? PloS Pathog. 2013; 9: e1003784.

23. Watashi K, Urban S, Li W, Wakita T. NTCP and beyond: opening the door to unveil hepatitis B virus entry. Int J Mol Sci. 2014; 15:2892-2905.

24. Liang TJ. Hepatitis B: the virus and disease. Hepatology. 2009; 49(Suppl 5):13-21.

25. Zhang Q, Cao G. Genotypes, mutations, and viral load of hepatitis B virus and the risk of hepatocellular carcinoma: HBV properties and hepatocarcinogenesis. Hepat Mon. 2011; 11:86-91.

26. Grimm D, Thimme R, Blum HE. HBV life cycle and novel drug targets. Hepatol Int. 2011; 5:644-653.

27. Tanwar S, Dusheiko G. Is there any value to hepatitis B virus genotype analysis? Curr Gastroenterol Rep. 2012; 14:37-46.

28. Lupberger, J.; Hildt, E. Hepatitis B virus-induced oncogenesis. World J. Gastroenterol. 2007, 13, 74–81. [Google Scholar] [PubMed]

29. Stockl, L.; Berting, A.; Malkowski, B.; Foerste, R.; Hofschneider, P.H.; Hildt, E. Integrity of c-Raf-1/MEK signal transduction cascade is essential for hepatitis B virus gene expression. Oncogene 2003, 22, 2604–2610. [Google Scholar] [CrossRef] [PubMed]

30. Hafner, A.; Brandenburg, B.; Hildt, E. Reconstitution of gene expression from a regulatory-protein-deficient hepatitis B virus genome by cell-permeable HBx protein. EMBO Rep. 2003, 4, 767–773. [Google Scholar] [CrossRef] [PubMed]

31. Hafner, A.; Brandenburg, B.; Hildt, E. Reconstitution of gene expression from a regulatory-protein-deficient hepatitis B virus genome by cell-permeable HBx protein. EMBO Rep. 2003, 4, 767–773. [Google Scholar] [CrossRef] [PubMed]

32. Kann M. Structural and Molecular Virology. In: Lai CL, Locarnini S(eds), Hepatitis B Virus. 2nd ed. London: International Medical Figure 7. Press; 2008. P.2-2. Secretion of HBV virions (modified from ref 21) 7 İnan et al. Hepatitis B Virus: Biology and Life Cycle

33. Bruns, M.; Miska, S.; Chassot, S.; Will, H. Enhancement of hepatitis B virus infection by noninfectious subviral particles. J. Virol. 1998, 72, 1462–1468. [Google Scholar] [PubMed]

34.Patient, R.; Hourioux, C.; Sizaret, P.Y.; Trassard, S.; Sureau, C.; Roingeard, P. Hepatitis B virus subviral envelope particle morphogenesis and intracellular trafficking. J. Virol. 2007, 81, 3842–3851. [Google Scholar] [CrossRef] [PubMed]

35.Chain, B.M.; Myers, R. Variability and conservation in hepatitis B virus core protein. BMC Microbiol. 2005, 5, 33. [Google Scholar] [CrossRef] [PubMed][Green Version]

36. Gallina, A.; Bonelli, F.; Zentilin, L.; Rindi, G.; Muttini, M.; Milanesi, G. A recombinant hepatitis B core antigen polypeptide with the protamine-like domain deleted self-assembles into capsid particles but fails to bind nucleic acids. J. Virol. 1989, 63, 4645–4652. [Google Scholar] [PubMed]

37. Hong HJ, Ryu CJ, Hur H, Kim S, Oh HK, Oh MS, Park SY. In vivo neutralization of hepatitis B virus infection by an anti-preS1 humanized antibody in chimpanzees. Virology. 2004; 318:134- 141.

38. Glebe D. Recent advances in hepatitis B virus research: A German point of view. World J Gastroenterol. 2007; 13:8-13.

39. Beck J, Nassal M. Hepatitis B virus infection. World J Gastroenterol. 2007; 13:48-64.

40. Prange R. Host factors involved in hepatitis virus maturation, assembly, and egress. Med Microbiol Immunol. 2012; 201:449-461.

41. Locarnini S. Molecular virology of hepatitis B virus. Semin Liver Dis. 2004; 24(Suppl 1):3-10

42. Ceres, P.; Zlotnick, A. Weak protein-protein interactions are sufficient to drive assembly of hepatitis B virus capsids. Biochemistry 2002, 41, 11525–11531. [Google Scholar] [CrossRef] [PubMed]

43.Zlotnick, A.; Johnson, J.M.; Wingfield, P.W.; Stahl, S.J.; Endres, D. A theoretical model successfully identifies features of hepatitis B virus capsid assembly. Biochemistry 1999, 38, 14644–14652. [Google Scholar] [CrossRef] [PubMed]

44.Bartenschlager, R.; Schaller, H. Hepadnaviral assembly is initiated by polymerase binding to the encapsidation signal in the viral RNA genome. EMBO J. 1992, 11, 3413–3420. [Google Scholar] [PubMed]
45.Beck, J.; Nassal, M. Efficient Hsp90-independent in vitro activation by Hsc70 and Hsp40 of duck hepatitis B virus reverse transcriptase, an assumed Hsp90 client protein. J. Biol. Chem. 2003, 278, 36128–36138. [Google Scholar] [CrossRef] [PubMed]

46. Hirsch, R.C.; Lavine, J.E.; Chang, L.J.; Varmus, H.E.; Ganem, D. Polymerase gene products of hepatitis B viruses are required for genomic RNA packaging as well as for reverse transcription. Nature 1990, 344, 552–555. [Google Scholar] [CrossRef] [PubMed]

47.Junker-Niepmann, M.; Bartenschlager, R.; Schaller, H. A short cis-acting sequence is required for hepatitis B virus pregenome encapsidation and sufficient for packaging of foreign RNA. EMBO J. 1990, 9, 3389–3396. [Google Scholar] [PubMed]

48.Lott, L.; Beames, B.; Notvall, L.; Lanford, R.E. Interaction between hepatitis B virus core protein and reverse transcriptase. J. Virol. 2000, 74, 11479–11489. [Google Scholar] [CrossRef] [PubMed]

49.HC, Huang EY, Su PY, Wu SY, Yang CC, Lin YS, Chang WC, Shih C. Nuclear export and import of human hepatitis B virus capsid protein and particles. PLoS pathogens. 2010;6(10):e1001162. PubMed PMID: 21060813; PubMed Central PMCID: PMC2965763. [PMC free article] [PubMed] [Google Scholar]

50. Perlman DH, Berg EA, O'Connor PB, Costello CE, Hu J. Reverse transcription-associated dephosphorylation of hepadnavirus nucleocapsids. Proceedings of the National Academy of Sciences of the United States of America. 2005;102(25):9020–9025. PubMed PMID: 15951426; PubMed Central PMCID: PMC1157036. [PMC free article] [PubMed] [Google Scholar]

51. Ludgate L, Ning X, Nguyen DH, Adams C, Mentzer L, Hu J. Cyclin-dependent kinase 2 phosphorylates s/t-p sites in the hepadnavirus core protein C-terminal domain and is incorporated into viral capsids. Journal of virology. 2012;86(22):12237–12250. PubMed PMID: 22951823; PubMed Central PMCID: PMC3486511. [PMC free article] [PubMed] [Google Scholar]

52. Melegari M, Wolf SK, Schneider RJ. Hepatitis B virus DNA replication is coordinated by core protein serine phosphorylation and HBx expression. Journal of virology. 2005;79(15):9810–9820. PubMed PMID: 16014942; PubMed Central PMCID: PMC1181610. [PMC free article] [PubMed] [Google Scholar]

53. Bock CT, Schwinn S, Locarnini S, Fyfe J, Manns MP, Trautwein C, Zentgraf H. Structural organization of the hepatitis B virus minichromosome. Journal of molecular biology. 2001;307(1):183–196. PubMed PMID: 11243813. [PubMed] [Google Scholar]

54. Nassal M. The arginine-rich domain of the hepatitis B virus core protein is required for pregenome encapsidation and productive viral positive-strand DNA synthesis but not for virus assembly. Journal of virology. 1992;66(7):4107–4116. PubMed PMID: 1602535; PubMed Central PMCID: PMC241213. [PMC free article] [PubMed] [Google Scholar]

55. Le Pogam S, Chua PK, Newman M, Shih C. Exposure of RNA templates and encapsidation of spliced viral RNA are influenced by the arginine-rich domain of human hepatitis B virus core antigen (HBcAg 165-173) Journal of virology. 2005;79(3):1871–1887. PubMed PMID: 15650211; PubMed Central PMCID: PMC544126. [PMC free article] [PubMed] [Google Scholar]

56. Chu TH, Liou AT, Su PY, Wu HN, Shih C. Nucleic acid chaperone activity associated with the arginine-rich domain of human hepatitis B virus core protein. Journal of virology. 2014;88(5):2530–2543. PubMed PMID: 24352445; PubMed Central PMCID: PMC3958103. [PMC free article] [PubMed] [Google Scholar]

57. Llewellyn EB, Loeb DD. The arginine clusters of the carboxy-terminal domain of the core protein of hepatitis B virus make pleiotropic contributions to genome replication. Journal of virology. 2011;85(3):1298–1309. PubMed PMID: 21084467; PubMed Central PMCID: PMC3020508. [PMC free article] [PubMed] [Google Scholar]

58. Le Pogam S, Shih C. Influence of a putative intermolecular interaction between core and the pre-S1 domain of the large envelope protein on hepatitis B virus secretion. Journal of virology. 2002;76(13):6510–6517. PubMed PMID: 12050364; PubMed Central PMCID: PMC136289. [PMC free article] [PubMed] [Google Scholar]

59. Zlotnick A, Venkatakrishnan B, Tan Z, Lewellyn E, Turner W, Francis S. Core protein: A pleiotropic keystone in the HBV lifecycle. Antiviral research. 2015 PubMed PMID: 26129969. [PMC free article] [PubMed] [Google Scholar]

60. Mason WS, Seal G, Summers J. Virus of Pekin ducks with structural and biological relatedness to human hepatitis B virus. Journal of virology. 1980;36(3):829–836. PubMed PMID: 7463557; PubMed Central PMCID: PMC353710. [PMC free article] [PubMed] [Google Scholar]

61. Summers J, Mason WS. Replication of the genome of a hepatitis B--like virus by reverse transcription of an RNA intermediate. Cell. 1982;29(2):403–415. PubMed PMID: 6180831. [PubMed] [Google Scholar]

62. Bartenschlager R, Schaller H. The amino-terminal domain of the hepadnaviral P-gene encodes the terminal protein (genome-linked protein) believed to prime reverse transcription. The EMBO journal. 1988;7(13):4185–4192. PubMed PMID: 2854056; PubMed Central PMCID: PMC455131. [PMC free article] [PubMed] [Google Scholar]

63. Zoulim F, Seeger C. Reverse transcription in hepatitis B viruses is primed by a tyrosine residue of the polymerase. Journal of virology. 1994;68(1):6–13. PubMed PMID: 7504742; PubMed Central PMCID: PMC236258. [PMC free article] [PubMed] [Google Scholar]

64. Weber M, Bronsema V, Bartos H, Bosserhoff A, Bartenschlager R, Schaller H. Hepadnavirus P protein utilizes a tyrosine residue in the TP domain to prime reverse transcription. Journal of virology. 1994;68(5):2994–2999. PubMed PMID: 7512155; PubMed Central PMCID: PMC236789. [PMC free article] [PubMed] [Google Scholar]

65. Radziwill G, Tucker W, Schaller H. Mutational analysis of the hepatitis B virus P gene product: domain structure and RNase H activity. Journal of virology. 1990;64(2):613–

620. PubMed PMID: 2153228; PubMed Central PMCID: PMC249151. [PMC free article] [PubMed] [Google Scholar]

66. Kim S, Lee J, Ryu WS. Four conserved cysteine residues of the hepatitis B virus polymerase are critical for RNA pregenome encapsidation. Journal of virology. 2009;83(16):8032–8040. PubMed PMID: 19515776; PubMed Central PMCID: PMC2715788. [PMC free article] [PubMed] [Google Scholar]

67. Xiong Y, Eickbush TH. Origin and evolution of retroelements based upon their reverse transcriptase sequences. The EMBO journal. 1990;9(10):3353–3362. PubMed PMID: 1698615; PubMed Central PMCID: PMC552073. [PMC free article] [PubMed] [Google Scholar]

68. Nassal M. Hepatitis B viruses: reverse transcription a different way. Virus research. 2008;134(1–2):235–249. PubMed PMID: 18339439. [PubMed] [Google Scholar]

69. Wang YX, Luo C, Zhao D, Beck J, Nassal M. Extensive mutagenesis of the conserved box E motif in duck hepatitis B virus P protein reveals multiple functions in replication and a common structure with the primer grip in HIV-1 reverse transcriptase. Journal of virology. 2012;86(12):6394–6407. PubMed PMID: 22514339; PubMed Central PMCID: PMC3393573. [PMC free article] [PubMed] [Google Scholar]

70. Tavis JE, Cheng X, Hu Y, Totten M, Cao F, Michailidis E, Aurora R, Meyers MJ, Jacobsen EJ, Parniak MA, Sarafianos SG. The hepatitis B virus ribonuclease H is sensitive to inhibitors of the human immunodeficiency virus ribonuclease H and integrase enzymes. PLoS pathogens. 2013;9(1):e1003125. PubMed PMID: 23349632; PubMed Central PMCID: PMC3551811. [PMC free article] [PubMed] [Google Scholar]

71. Bartenschlager R, Junker-Niepmann M, Schaller H. The P gene product of hepatitis B virus is required as a structural component for genomic RNA encapsidation. Journal of virology. 1990;64(11):5324–5332. PubMed PMID: 2214019; PubMed Central PMCID: PMC248581. [PMC free article] [PubMed] [Google Scholar]

72. Jones SA, Hu J. Hepatitis B virus reverse transcriptase: diverse functions as classical and emerging targets for antiviral intervention. Emerg Microbes Infect. 2013;2(9):e56. PubMed PMID: 26038488; PubMed Central PMCID: PMC3820986. [PMC free article] [PubMed] [Google Scholar]

73 . Hirsch, R.C.; Lavine, J.E.; Chang, L.J.; Varmus, H.E.; Ganem, D. Polymerase gene products of hepatitis B viruses are required for genomic RNA packaging as well as for reverse transcription. Nature 1990, 344, 552–555. [Google Scholar] [CrossRef] [PubMed]