J Mol Evol (1996) 42:97–102

JOURNAL OF MOLECULAR EVOLUTION

Recombination Between Sequences of Hepatitis B Virus from Different Genotypes

Paul L. Bollyky, Andrew Rambaut, Paul H. Harvey, Edward C. Holmes

Wellcome Centre for the Epidemiology of Infectious Disease, Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK

Received: 11 July 1995 / Accepted: 16 October 1995

Abstract. A comparison of 25 hepatitis B virus (HBV) isolates for which complete genome sequences are available revealed two that occupied different positions in phylogenetic trees reconstructed from different open reading frames. Further analysis indicated that this incongruence was the result of recombination between viruses of different genomic and antigenic types. Both putative recombinants originated from geographic regions where multiple genotypes are known to cocirculate. A search of the sequence databases showed evidence of similar intergenotypic recombinants. These observations indicate that recombination between divergent strains may represent an important source of genetic variation in HBV.

Key words: Hepatitis B virus — Recombination — Phylogeny — Coinfection — Genotypes — Antigenic subtypes

Introduction

Hepatitis B virus (HBV) is a partially double-stranded DNA virus of the hepadnavirus family with a genome of only 3.2 kb, although there is translation in more than one reading frame of its circular genome. The virus causes considerable mortality and morbidity in many areas of the world, particularly tropical Africa and the Far East, where it has been estimated that between 10 and 15% of the population are chronic HBV carriers (Sherlock 1993).

Phylogenetic analyses of HBV sequences has led to the classification of the virus into six "genomic groups" or genotypes (denoted A–F), often with distinct geographic associations (Okamoto et al. 1988; Norder et al. 1992, 1993, 1994). In contrast, nine different antigenic subtypes have been previously distinguished using serological criteria (Bancroft et al. 1972; Le Bouvier 1971), so some antigenic subtypes span several genomic groups, although there is generally a good association between the genetic and antigenic classifications (Norder et al. 1993). Recombination between diverse viruses, in addition to point mutations, may represent one means by which the association between an antigenic subtype and genotype could change. Although recombination has already been documented in HBV (Georgi-Geisberger et al. 1992; Hino et al. 1991), exchange of sequences between viruses belonging to different subtypes or genotypes has not been reported to date and evidence for coinfection with divergent HBV strains is rare (Carman et al. 1993a). Furthermore, questions remain regarding the precise mechanism by which recombination takes place as well as its frequency and potential implications. Here we investigate, through phylogenetic analysis, whether recombination has been a factor in the generation of genetic diversity in 25 complete HBV genomes available on the molecular sequence databases.

Correspondence to: E.C. Holmes

Table 1. HBV sequences used in this study

GenBank ID	GenBank accession	Genotype	Antigenic subtype[a]	Reference
HBVADW	V00866, J02201	A	adw2	Ono et al. (1983)
HPBADWZCG	M57663	A	adw2	Estacio et al. (1988)
HBVADW2	X02763	A	adw2	Valenzuela et al. (1980)
HVHEPB	X51970	A	adw	Koechel (1990)
HPBADW1	D00329	B	adw	Okamoto et al. (1988)
HPBADW2	D00330	B	adw	Okamoto et al. (1988)
HPBADW3	D00331	B	NA	Okamoto et al. (1988)
HPBADWZ	M54923	B	adw	Satrosoewignjo et al. (1987)
HBVADR4	X01587	C	adr	Fujiyama et al. (1983)
HBVADR1CG	M38454	C	adr	Renbao et al. (1987)
HBVADRM	X14193	C	adr	Rho et al. (1989)
HEHBVAYR	X04615	C	ayr	Okamoto et al. (1986)
HPBADRA	M12906	C	adr	Kobayashi et al. (1984)
HPBADRC	D00630	C	adr	Ono et al. (1983)
HPBCG	D12980	C	adr	Mukaide et al. (1992)
HPBCGADR	M38636	C	adr	Kim et al. (1988)
HPBETNC	L08805, L08806	C	NA	Ogata et al. (1993)
HBVAYWMCG	X59795	D	ayw	Lai et al. (1991)
HBVDNA	X68292	D	ayw	Lai et al. (1992)
HPBHBVAA	M32138	D	NA	Tong et al. (1990)
HPBMUT	L27106	D	NA	Hasegawa et al. (1994)
XXHEPA	V01460, J02203	D	ayw	Galibert et al. (1979)
XXHEPAV	X02496	D	ayw	Bichko et al. (1985)
HBVADW4A	X69798	F	adw4	Naumann et al. (1993)

[a] NA: antigenic subtype not available

Methods

One simple instance in which recombination can be detected is when different genes, or different intergenic regions from a single genome, generate incongruent phylogenetic trees (Robertson et al. 1995a, 1995b). Because HBV has a circular genome it is sometimes difficult to determine the direction of the recombination event. In the analysis presented here putative recombinants were defined as those sequences falling into different viral genotypes in phylogenetic trees reconstructed for the four open reading frames (ORFs) from 25 complete genomes of HBV (Table 1).

All sequences were obtained from the GenBank/EMBL/DDBJ databases and aligned with the ClustalV program (Higgins et al. 1992). Phylogenetic trees were constructed using the maximum likelihood method implemented in the PHYLIP package (program DNAML; Felsenstein 1993). In order to assess the robustness of the groupings obtained a bootstrap neighbor-joining analysis with a 1,000 replications was also performed (programs SEQBOOT, DNADIST, NEIGHBOR, and CONSENSE). To localize the crossover points of the recombination events a method was used which identifies the sequence position which maximizes the chi-square of the differences in variable sites for which the recombined sequence shares a state with either of two proposed "parental" sequences (Maynard Smith 1992). A randomization test then estimates the probability that this could have occurred by chance. To find both ends of a recombination event the circular genome was cut into two sections through the center of the hypothesized recombined section (identified roughly by sight). The test was then performed on these two sections independently. All sequences are numbered according to scheme of Galibert et al. (1979).

Results and Discussion

Phylogenetic trees were constructed on each of the four ORFs of HBV separately. There were two viruses, HBVDNA and HPBADW1, which clustered with different genotypes in different ORFs, indicative of recombination. HBVDNA is a subtype ayw virus taken from an HBsAg (surface antigen) negative chronic liver disease patient in Italy (Lai et al. 1992). Phylogenetic analysis of the 25 complete genomic sequences shows that this virus is a member of genotype D (Fig. 1A), whereas its recombinant region mapped with the sequences of genomic group A (Fig. 1B). Of all the genotypes, A and D have the widest geographical distributions and are found on most continents (Norder et al. 1993). HPBADW1 is a virus of subtype adw and was taken from an asymptomatic carrier of HBsAg in Japan (Okamoto et al. 1988). A phylogeny reconstructed for the entire genome places this virus within genotype B (Fig. 1A), a genotype that is apparently restricted to Southeast Asia (Norder et al. 1993). The putative recombinant region clusters within genotype A, although the trees reconstructed in this case show less resolution because of the short length of sequence analyzed (Fig. 1C).

In the case of HBVDNA the region of recombination covered approximately half the genome (positions 735–2370) with the crossover points occurring in an overlapping area of the P and S genes and in the C gene (Fig. 2, Table 2). The P gene encodes a primase, reverse transcriptase, and a RNAase H which can cleave RNA in RNA/DNA hybrids; the S gene produces the viral surface antigen (HBsAg) while the C gene encodes the core protein. The region of recombination in HPBADW1 is much smaller, encompassing a 189-bp segment within

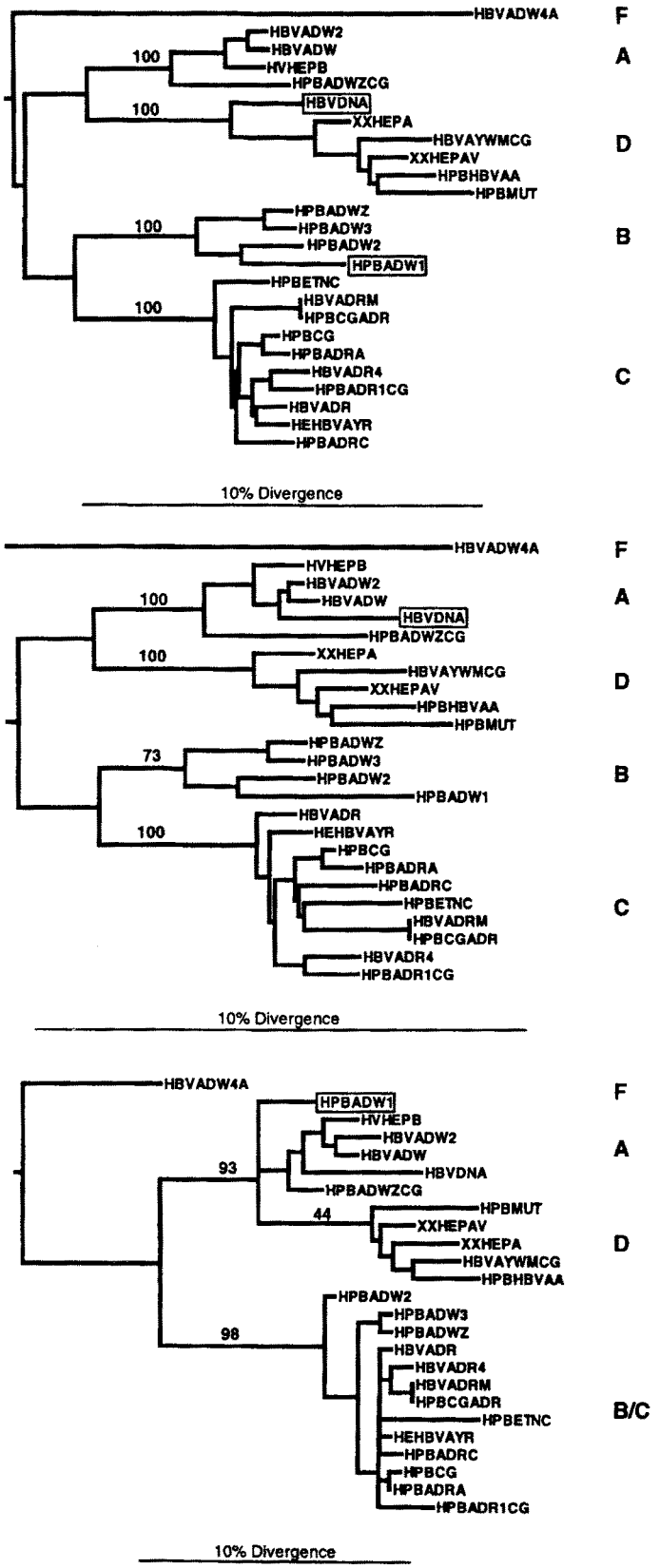

Fig. 1. **A** Maximum likelihood tree of the 25 complete genomes of HBV. **B** Maximum likelihood tree for the recombinant region in sequence HBVDNA (positions 735–2365). **C** Maximum likelihood tree for the recombinant region in HPBADW1 (positions 2014–2203). In all cases the putative recombinants are indicated with *boxes*. The *numbers* of bootstrap replications supporting each genotype are indicated. For ease of interpretation, trees are rooted on the longest branch which separates the single genotype F sequence, HBVADW4A, from the remaining sequences. Horizontal branch lengths are drawn to scale.

the C gene without overlapping reading frames (positions 2014–2203; Fig. 2, Table 2). A BLAST search and subsequent phylogenetic analysis detected additional sequences which had mosaic patterns (parental genotypes and recombination sites) and phylogenetic relationships identical to those in HPBADW1 (GenBank identifiers HPBHBCAGA, HPBHBCAGB, HPBHBCAGC; data not shown). These sequences are also from Japanese patients, which indicates that superinfection and recombination are not uncommon in localities where a number of

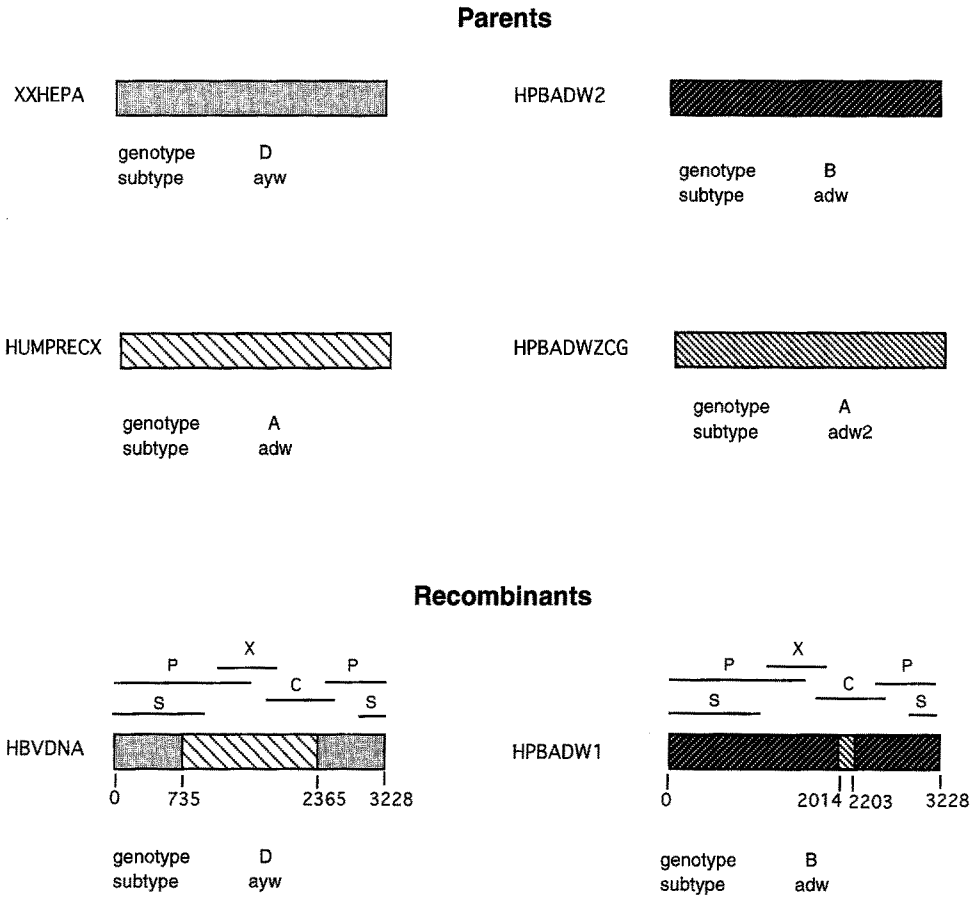

Fig. 2. Genotypic relationships of different regions of the putative recombinant HBV sequences. The circular genomes were split for ease of representation. The locations of the crossover points of the recombination events are indicated as are the positions of P, C, X, and S open reading frames.

Table 2. Localization of recombination events in HBV sequences[a]

Isolate	Parental lineage virus	Genotype	Crossover points	*P*	Region	Informative sites 1	Informative sites 2
HBVDNA	XXHEPA	D	728	<0.001			
			2365	<0.001			
					2365–735	27	208
					735–2365	96	41
	HUMPRECX[b]	A	735	<0.001			
			2365	<0.001			
HPBADW1	HPBADW2	B	1988	0.005			
			2203	<0.001			
					2203–2014	98	284
					2014–2203	28	10
	HPBADWZCG	A	2014	0.004			
			2203	<0.001			

[a] Localization of crossover points was determined using the suspected recombinant and the two parental sequences which came closest in a BLAST search of GenBank. Discrepancies in the crossover points given could arise by mutations subsequent to recombination or by the exact point of recombination occurring within an area conserved between the viruses in question. Column 1 denotes the number of informative sites grouping the recombinant and the sequence most closely matching the recombinant region, while column 2 denotes the number of informative sites grouping the recombinant and the sequence most closely related to its ancestrally derived region. Position 2014 was designated as a crossover point for HPBADW1 due to its greater statistical support while position 735 was chosen arbitrarily for isolate HBVDNA. Probability *P* was found using a randomization test with 1,000 trials.

[b] Meisel et al. (1993); GenBank accession number L13994

different genomic groups cocirculate and that these recombinants are viable and capable of generating new lineages, even though most cases of recombination would be expected to be fatal in a virus with multiple reading frames and such a small genome.

Recombination in 2 of the 25 viruses in our sample indicates that superinfection of individuals must occur with a noticeable frequency. Although there is little previous evidence for superinfection of HBV it is noteworthy that viruses with rare antigenic specificities, adwr and adyr, do appear, and these are thought to be caused by mixed infections (Carman et al. 1993b). Furthermore, incongruence between trees reconstructed from different ORFs of HBV had been documented elsewhere (Norder et al. 1994), although the possibility that this was caused by recombination was not discussed. Mixed infections could occur through either simultaneous transmission or sequential infection. The probability of the former must be low but there is a clear situation in which the latter can occur. An individual who has resolved an HBV infection will have antibody to the viral surface antigen and this will induce cross-protection to all antigenic subtypes (although some neutralizing antibodies may also be subtype specific). In contrast, in chronically infected patients, antibody is complexed to the viral surface antigen so that protection from reinfection is reduced. The mechanisms by which mosaic viruses are generated within a multiply infected host are also unclear, as there appears to be little opportunity for recombination in HBV. All replication steps are particle associated and a single RNA genome equivalent is taken up within the particle to be converted to the partially double-stranded DNA HBV virus, thereby reducing the probability of template switching (Georgi-Geisberger et al. 1992). Furthermore, it is thought that integrated HBV cannot serve as a template for virus replication as there is no evidence of integration of complete genomes, even with multiple integration events (Buendia et al. 1993).

The recombination events described here are significant because they represent a previously unrecognized source of variation in HBV, potentially both within individuals and populations at large. It is particularly important to determine whether the crossover points of recombination occur throughout the genome and whether recombinant viruses have altered biological properties, perhaps generating vaccine escape mutations (Harrison et al. 1991; Wallace et al. 1994; Yamamoto et al. 1994) or fulminant hepatitis (Carman et al. 1991; Kosaka et al. 1991; Omata et al. 1991), although to date these phenotypes have only been associated with point mutations and insertion/deletion events. Another possible outcome of recombination, the mixing of antigenic determinants and epitope types, might also lead to immune evasion through the generation of new and unfamiliar antigenic configurations.

Acknowledgments. This work was funded by the Marshall Scholarship Commission, The Royal Society, and The Wellcome Trust. We thank an anonymous referee for valuable comments.

References

Bancroft WH, Mundon FK, Russel PK (1972) Detection of additional antigenic determinants of hepatitis B antigen. J Immunol 109:842–848

Bichko V, Pushko P, Dreilina D, Pumpen P, Gren E (1985) Subtype ayw variant of hepatitis B virus. DNA primary structure analysis. FEBS Lett 185:208–212

Buendia MA, Paterlini P, Tiollas P, Brechot C (1993) Hepatitis B virus: liver cancer. In: Zuckerman AJ, Thomas HC (eds) Viral hepatitis. Churchill Livingstone, Edinburgh, pp 137–164

Carman WF, Hadziyannis S, Karayiannis P (1991) Association of the precore variant of HBV with acute and fulminant hepatitis B infection. In: Hollinger FB, Lemon SM, Mangolis HS (eds) Viral hepatitis and liver disease. Williams and Wilkins, Baltimore, pp 216–219

Carman WF, Thomas HK, Domingo E (1993a) Viral genetic variation: hepatitis B virus as a clinical example. Lancet 341:349–353

Carman WF, Thomas HC, Zuckerman AJ, Harrison T (1993b) Hepatitis B virus: molecular variants. In: Zuckerman AJ, Thomas HC (eds) Viral hepatitis. Churchill Livingstone, Edinburgh, pp 115–136

Estacio RC, Chavez CC, Okamoto H, Lingao AL, Reyes MT, Domingo E, Mayumi M (1988) Nucleotide sequence of a hepatitis B virus genome of subtype adw isolated from a Philippino: comparison with the reported three genomes of the same subtype. J Gastroenterol Hepatol 3:215–222

Felsenstein J (1993) PHYLIP (phylogeny inference package) version 3.5c. Distributed by the author. Department of Genetics, University of Washington, Seattle, USA

Fujiyama A, Miyanohara A, Nozaki C, Yoneyama T, Ohtomo N, Matsubara K (1983) Cloning and structural analyses of hepatitis B virus DNAs, subtype adr. Nucleic Acids Res 11:4601–4610

Galibert F, Mandart E, Fitoussi F, Tiollais P, Charnay P (1979) Nucleotide sequence of the hepatitis B virus genome (subtype ayw) cloned in E. coli. Nature 281:646–650

Georgi-Geisberger P, Berns H, Loncarevic IF, Yu ZY, Tang ZY, Zentgraf H, Scroder CH (1992) Mutations on free and integrated hepatitis B Virus DNA in a hepatocellular carcinoma: footprints of homologous recombination. Oncology 49:386–395

Harrison TJ, Hopes EA, Oon CJ, Zanetta AR, Zuckerman AJ (1991) Independent emergence of a vaccine induced escape mutant of hepatitis B virus. J Hepatol 13:105–107

Hasegawa K, Huang JK, Rogers SA, Blum HE, Liang TJ (1994) Enhanced replication of a hepatitis B virus mutant associated with an epidemic of fulminant hepatitis. J Virol 68:1651–1659

Higgins DG, Bleasby AJ, Fuchs R (1992) CLUSTALV: improved software for multiple sequence alignment. Comput Appl Biosci 8:189–191

Hino O, Tabata S, Hotta Y (1991) Evidence for increased in vitro recombination with insertion of human hepatitis B virus DNA. Proc Natl Acad Sci USA 88:9248–9252

Kim KT, Hyun SW, Kim YS, Rho HM (1988) Complete nucleotide sequence of hepatitis B virus (subtype adr). Korean J Biochem 21:319–331

Kobayashi M, Koike K (1984) Complete nucleotide sequence of hep-

atitis B virus DNA of subtype adr and its conserved gene organization. Gene 30:227–232

Koechel H (1990) Unpublished. Direct submission to the EMBL/GenBank/DDBJ databases

Kosaka Y, Takase K, Kojima M (1991) Fulminant hepatitis B: induction by hepatitis B virus mutants defective in the precore region and incapable of encoding e antigen. Gastroenterology 324:1087–1094

Lai ME, Mazzoleni AP, Melis A, Balestrieri A (1992) Unpublished. Direct submission to the EMBL/GenBank/DDBJ databases

Lai ME, Melis A, Mazzoleni AP, Farci P, Balestrieri A (1991) Sequence analysis of hepatitis B virus genome of a new mutant of ayw subtype isolated in Sardinia. Nucleic Acids Res 19:5078

Le Bouvier GL (1971) The heterogeneity of the Australia antigen. J Infect Dis 123:671–675

Maynard Smith J (1992) Analyzing the mosaic structure of genes. J Mol Evol 34:126–129

Meisel H, Jantschak J, Tars K, Prosch S, Pushko P, Pumpen P (1993) Unpublished. Direct submission to the EMBL/GenBank/DDBJ databases

Mukaide M, Kumazawa T, Hoshi A, Kawaguchi R, Hikiji K (1992) The complete nucleotide sequence of Hepatitis B virus, subtype adr (SRADR) and phylogenetic analysis. Nucleic Acids Res 20:6105

Naumann H, Schaefer S, Yoshida CFT, Gaspar AML, Repp R, Gerlich WH (1993) Identification of a new hepatitis B virus (HBV) genotype from Brazil that expresses HBV surface antigen subtype adw4. J Gen Virol 74:1627–1632

Norder H, Hammas B, Lofdahl S, Courouce A, Magnius L (1992) Comparison of the amino acid sequences of nine different serotypes of hepatitis B surface antigen and genomic classification of the corresponding hepatitis B virus strains. J Gen Virol 73:1201–1208

Norder H, Courouce AM, Magnius LO (1994) Complete genomes, phylogenetic relatedness, and structural proteins of six strains of the hepatitis B virus, four of which represented two new genotypes. Virology 198:489–503

Norder H, Hammas B, Lee SD, Bile K, Courouce A, Mushahwar I, Magnius LO (1993) Genetic relatedness of hepatitis B viral strains of diverse geographical origin and natural variations in the primary structure of the surface antigen. J Gen Virol 74:1341–1348

Ogata N, Miller RH, Ishak KG, Purcell RH (1993) The complete nucleotide sequence of a pre-core mutant of hepatitis B virus implicated in fulminant hepatitis and its biological characterization in chimpanzees. Virology 194:263–276

Okamoto H, Imai M, Shimozaki M, Hoshi Y, Iizuka H, Gotanda T, Tsuda F, Miyakawa Y, Mayumi M (1986) Nucleotide sequence of a cloned hepatitis B virus genome, subtype ayr: comparison with genomes of the other three subtypes. J Gen Virol 67:2305–2314

Okamoto H, Tsuda F, Sakugawa H, Sastrosoewignjo RI, Imai M, Miyakawa Y, Mayumi M (1988) Typing hepatitis B virus by homology in nucleotide sequence: comparison of surface antigen subtypes. J Gen Virol 69:2575–2583

Omata M, Ehata T, Yokosuka O, Hosoda K, Ohto M (1991) Mutations in the precore region of hepatitis B virus DNA in patients with fulminant and severe hepatitis. New Engl J Med 324:1699–1704

Ono Y, Onda H, Sasada R, Igarashi K, Sugino Y, Nishioka K (1983) The complete nucleotide sequences of the cloned hepatitis B virus DNA; subtype adr and adw. Nucleic Acids Res 11:1747–1757

Renbao G, Meijin C, Lueping S, Suwen Q, Zaiping L (1987) The complete nucleotide sequence of the cloned DNA of hepatitis B virus subtype adr in pADR-1. Sci Sin 30:507–521

Rho HM, Kim KT, Hyun SW, Kim YS (1989) The nucleotide sequence and reading frames of a mutant hepatitis B virus subtype adr. Nucleic Acids Res 17:2124

Robertson DK, Hahn BH, Sharp PM (1995a) Recombination in AIDS viruses. J Mol Evol 40:249–259

Robertson DL, Sharp PM, McCuthan FE, Hahn BH (1995b) Recombination in HIV-1. Nature 374:124–126

Satrosoewignjo RI, Omi S, Okamoto H, Mayumi M, Rustam M, Sujudi X (1987) The complete nucleotide sequence of HBV DNA clone of subtype adw (pMND122) from Menado in Sulawesi Island, Indonesia. ICMR Ann 7:51–60

Sherlock S (1993) Clinical features of hepatitis. In: Zuckerman AJ, Thomas HC (eds) Viral hepatitis. Churchill Livingstone, Edinburgh, pp 1–17

Tong SP, Li JS, Vitvitski L, Trepo C (1990) Active hepatitis B virus replication in the presence of anti-HBe is associated with viral variants containing an inactive pre-C region. Virology 176:596–603

Valenzuela P, Quiroga M, Zalvidar J, Gray P, Rutter WJ (1980) The nucleotide sequence of the hepatitis B viral genome and the identification of the major viral genes. In: Fields BN, Jaenisch R, Fox CF (eds) Animal virus genetics. Academic Press, New York, pp 57–70

Wallace LA, Carman WF (1994) Clinical implications of hepatitis B virus envelope protein variation. Int J Clin Lab Res 24:80–85

Yamamoto K, Hosikita M, Tsuda F, Itoh K, Akahane Y, Yotsumoto S, Okamoto H, Miyakawa Y, Mayumi M (1994) Naturally occurring escape mutants of hepatitis B virus with various mutations in the S gene in carriers seropositive for antibody to hepatitis B surface antigen. J Virol 68:2671–2676