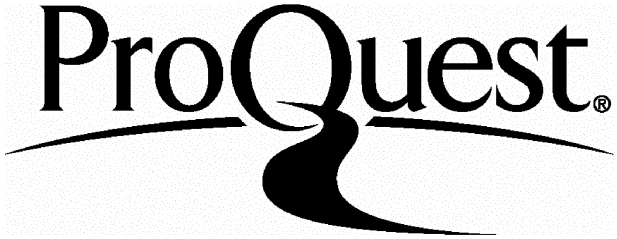
Molecular epidemiological studies of defined variants of hepatitis B and TT viruses

Rachel Louise Hallett

Submitted for the degree of Doctor of Philosophy

Department of Virology
University College London



Abstract

During surveillance of viral hepatitis in England, three particular viral variants, two of hepatitis B virus (HBV) and the other of TT virus (TTV), were identified. Molecular biological and phylogenetic analysis tools, largely developed to study eukaryotic genes, were adapted to define the genetic characteristics of the three variants, to estimate their relative prevalences and to study their relationships with other similar viral strains. The hypothesis investigated in this thesis is that such studies permit the epidemiological and evolutionary significance of particular viral variants to be elucidated.

One of the HBV variants, HBV^{PV}, was discovered while analysing clusters of acute hepatitis B infection in northern English prisons. A high throughput assay using denaturing gradient gel electrophoresis (DGGE) was developed to screen HBV PCR products for sequence variation. This approach revealed that while the prevalence of HBV^{PV} was similar in communities in the north and south east of England in 1990, over the next 6 years it dramatically increased in the north, but not the south east. The spread of HBV^{PV} into the community thus appeared to be linked to its prior dissemination among prison inmates.

The other HBV variant, HBV^{AV}, was shown by DGGE to infect a large number of patients attending an alternative therapy clinic. Unlike HBV^{PV}, it was not circulating commonly in the community, further implicating it as the cause of the point source outbreak.

DNA from the TTV-like variant, PM virus (PMV), was isolated while conducting a search for possible viral causes of acute non-A to E viral hepatitis. The entire PMV genome was characterised and several phylogenetic analysis tools were applied to study the evolutionary and taxonomic relationships between PMV and other TTVs. These approaches permitted clarification of the status of TTVs as viruses that belong to a new genus of the *Circoviridae* family rather than variants of a single species, and PMV itself as the prototype of a unique species within the genus.

Acknowledgements

Many thanks to my supervisors Dr Chong Gee Teo and Dr Peter Balfe, and also to Dr Jonathan Clewley, for their help in the preparation of this thesis.

I am indebted to my family, friends and colleagues in VRD for their support.

Table of contents

List of figures

List of tables

Abbreviations

ALT	alanine aminotransferase
BFDV	beak and feather disease virus
BLAST	basic local alignment search tool
bp	base pair
CAV	chicken anaemia virus
cccDNA	covalently closed circle DNA
CDSC	Communicable Disease Surveillance Centre
CPHL	Central Public Health Laboratory
CTL	cytotoxic T lymphocyte
DGGE	denaturing gradient gel electrophoresis
DHBV	duck hepatitis B virus
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
EDTA	disodium ethylenediaminetetra acetate
EMBL	European Molecular Biology Laboratory
ε	encapsidation signal
HBV	hepatitis B virus
HBV^{AV}	hepatitis B virus autohaemotherapy variant
HBV^{PV}	hepatitis B virus prisoner variant
HBcAg	hepatitis B core antigen
HBeAg	hepatitis B e antigen
HBsAg	hepatitis B surface antigen
HBIG	hepatitis B immunoglobulin
HCC	hepatocellular carcinoma
HCV	hepatitis C virus
HIV	human immunodeficiency virus
HLA	human leukocyte antigen
IDU	injecting drug user
IFN-α	alpha interferon
kb	kilobase
kDa	kiloDalton

MHR	major hydrophilic region
mRNA	messenger RNA
NCR	non-coding region
nt	nucleotide
ORF	open reading frame
P	polymerase
PCR	polymerase chain reaction
PCV	porcine circovirus
PHLS	Public Health Laboratory Service
PHYLIP	phylogeny inference package
PMV	PM virus
RCR	rolling circle replication
RDA	representational difference analysis
RIP	recombinant identification program
RNA	ribonucleic acid
SSCP	single strand conformation polymorphism
S proteins	surface proteins
TAE	tris-acetate-EDTA
TBE	tris-borate-EDTA
TLMV	TTV-like mini virus
T_m	melting temperature
TTV	TT virus
UPGMA	unweighted pair group method with arithmetic means

Chapter 1

Introduction

1.1 Viral causes of hepatitis

Knowledge of the existence of hepatitis can be traced back to the fifth century BC, when Hippocrates first described epidemic jaundice. Throughout history, outbreaks of hepatitis have been described in association with wars and natural disasters (Mahoney, 1999; Paver and Mortimer, 1996). In 1885, Lurman recorded 191 cases of jaundice in workers at a shipyard in Bremen who had all received a smallpox vaccine preparation containing glycerinated human lymph. This report provided the first evidence that hepatitis could be transmitted from blood-containing products. Similar outbreaks have been documented in diabetics, in patients treated for syphilis, and following the administration of yellow fever vaccines. However, it was only after World War II that a distinction could be made between hepatitis A (infectious hepatitis) which was transmitted via the faecal-oral route, and hepatitis B (serum hepatitis) which resulted from percutaneous exposure. Descriptions of their different epidemiologies resulted from human volunteer studies (Havens, 1946; MacCallum and Bauer, 1947) and Krugman and colleagues later verified that homologous immunity occurred after infection with each type (Krugman *et al.*, 1967).

At about this time, Blumberg et al. (1967) described an immunoprecipitin that was present in the blood of a leukaemic Australian Aborigine, which they named Australia antigen. Subsequent studies found that this antigen was highly prevalent in multiply transfused patients and a link with hepatitis B was hypothesised. Its viral origin was established when particles that reacted with antisera to Australia antigen were visualised using immune electron microscopy (Dane *et al.*, 1970). The various components that make up the hepatitis B "Dane particle" were rapidly dissected. Hepatitis B virus (HBV) is now classified as a hepadnavirus and several related mammalian and avian members have since been characterised. HBV will be further described in Section 1.2.

The discovery of the hepatitis A virus, also by the use of immune electron microscopy, was reported later (Feinstone *et al.*, 1973). It could be experimentally transmitted to primates, and was finally isolated in cell culture in 1979 (Provost and Hilleman, 1979).

Diagnostic assays for both viruses quickly became available. In the early 1970s, many developed countries introduced blood donor screening for the Australia antigen, now renamed hepatitis B surface antigen (HBsAg), in an effort to reduce post-transfusion jaundice. However, this condition did not disappear, and hepatitis was still observed in approximately 10% of blood transfusion recipients. It thus became apparent that at least one further bloodborne hepatitis agent remained to be found.

Hepatitis D virus (delta virus) was first described in 1977, and was initially thought to be a new antigen of hepatitis B (Rizzetto *et al.*, 1977). However, it was later shown to be a separate transmissible, but defective, virus. Its 1.7 kb RNA genome encodes only one antigen and it requires HBV to replicate and to coat its genome with an envelope consisting of hepatitis B surface proteins.

A further enterically transmitted hepatitis virus, hepatitis E virus, was detected when clinical samples, collected during water-borne epidemics of hepatitis in India, were serologically tested for HAV and HBV. It was found that most patients had evidence of anti-HAV IgG and so were immune to infection by this virus, which had initially been postulated as the causative agent (Wong *et al.*, 1980). HEV was visualised in 1983 and later experimental transmission studies demonstrated its aetiologic role in hepatitis (Balayan *et al.*, 1983; Krawczynsky and Bradley, 1989).

Meanwhile, the principal cause of post-transfusion non-A non-B hepatitis cases remained undiscovered until 1989, when the hepatitis C virus (HCV) genome was characterised using molecular cloning techniques (Choo *et al.*, 1989). Around 170 million people worldwide are thought to be chronically infected with this flavivirus.

It was found, however, that some cases of post-transfusion non-A non-B hepatitis still occurred that could not be linked to HCV infection. Efforts continue to isolate infectious agents that are responsible for these residual cases, and for sporadic community-acquired non-A to E hepatitis. In 1995, two groups reported the detection of viral sequences in such patients. These were later found

to be isolates of the same virus, named hepatitis G virus (HGV) or GB virus C (GBV-C) (Simons *et al.*, 1995; Linnen *et al.*, 1996). It is related to HCV, and is transmitted via bloodborne routes. A few studies have demonstrated a higher prevalence of HGV/GBV-C in patients with acute or chronic hepatitis than in healthy controls (Linnen *et al.*, 1996; Fiordalisi *et al.*, 1996) and linked it to some cases of fulminant hepatitis (Heringlake *et al.*, 1996). However, chronically infected patients have been found with normal aminotransferase levels, and HGV/GBV-C appears to have no effect on the progression to liver disease in patients who already carry another hepatitis virus (Alter *et al.*, 1997). Recent evidence suggests it replicates in lymphoid tissue, but not in hepatocytes (Hoffman *et al.*, 1998; Kobayashi *et al.*, 1999). Thus its role as a hepatitis agent is questionable.

The most recent hepatitis virus candidate was discovered in 1997, using the technique of representational difference analysis (RDA) (Nishizawa *et al.*, 1997). It was named TT virus (TTV) after the initials of the post-transfusion hepatitis patient from whom viral DNA was first isolated. TTV is discussed further in Section 1.3.

1.2 Hepatitis B Virus

1.2.1 Virion structure, genetic organisation and proteins

The HBV virion (Dane particle) consists of a 42nm double-shelled structure. The viral genome and DNA polymerase are encapsulated by a nucleocapsid made up of hepatitis B core antigen (HBcAg), which in turn is surrounded by an envelope comprising the predominant HBsAg (also known as the small S protein), and the middle and large S proteins.

Subviral structures (spheres and filaments), composed of the three hepatitis B S proteins, are usually present in the blood of HBV carriers at much higher concentrations than the infectious Dane particles. The exact function of these empty particles is unknown, but it has been postulated that they serve as immunological decoys to adsorb neutralising antibodies. A recent study suggests

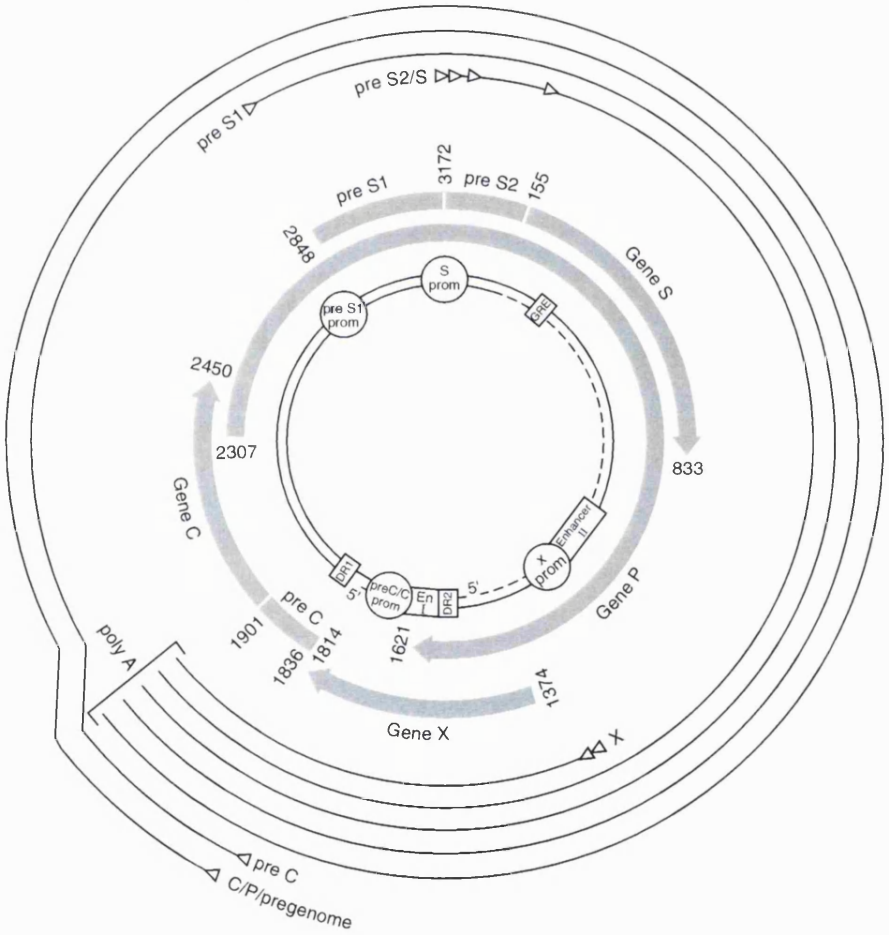
they are able to enhance intracellular viral replication at the early stage of infection (Bruns *et al.*, 1998).

HBV has a circular, partially double-stranded DNA genome. Its 3200 bp are economically organised, with four overlapping open reading frames (ORFs) and regulatory sequences contained within the genes (Figure 1.1). This genomic arrangement allows HBV to code for more protein per nucleotide than any other known virus, and imposes constraints on its evolution (Mizokami *et al.*, 1997). The longer minus strand has a 5' covalently liked terminal protein, or 'primase', which is part of the viral polymerase and is necessary for priming minus strand synthesis (Bosch *et al.*, 1988; Weber *et al.*, 1994). The 5' end of the plus strand has an overlap of around 200 complementary bases with the 5' end of the minus strand, enabling the DNA to circularise. This is helped by two 11bp direct repeat sequences, DR1 and DR2, which are also the initiation sites for minus strand and plus strand DNA synthesis respectively (Will *et al.*, 1987; Seeger and Maragos, 1990). The length of the 3' end of the plus strand is variable and is still attached to the viral polymerase, indicating that hepatitis B virions are secreted from infected hepatocytes before plus strand synthesis is completed (Landers *et al.*, 1977).

The four ORFs encode seven known viral proteins. Other than the three envelope (S) proteins, these are HBcAg, hepatitis B e antigen (HBeAg), viral polymerase (P) and the X protein. Four promoters and two enhancers provide transcriptional control. The preS1 promoter regulates transcription of a 2.4 kb mRNA, which codes for the large S protein, and contains binding sites for liver specific transcription factors. The preS2/S promoter is embedded within the preS1 region and controls transcription of a 2.1 kb mRNA, encoding both the middle S protein and HBsAg. The core promoter directs the production of two supergenomic transcripts: the 3.5kb pregenomic RNA, from which the HBcAg and polymerase are translated, and which serves as the RNA template during replication, and the precore mRNA, which codes for HBeAg, a secreted protein. The P ORF is the largest, encompassing about 80% of the genome. As well as the amino terminal primase protein and DNA polymerase it contains a reverse transcriptase domain and has RNaseH activity (Fallows and Goff, 1996; Kann and Gerlich, 1998). The

X promoter controls transcription and translation from the smallest ORF, producing the X protein, which has transactivating activity and many other regulatory functions (Murakami, 1999). All transcripts use the same polyadenylation site and are therefore coterminal.

Figure 1.1 The genetic organisation of the HBV genome. The partially double-stranded DNA is shown innermost, with promoters (prom), direct repeats (DR), enhancer regions (En) and the glucocorticoid regulatory element (GRE). Shaded arrows denote positions of ORFs. Outer circles show the transcribed mRNAs and pregenomic RNA.



1.2.2 Replication

Extensive experimental studies of duck hepatitis B virus (DHBV) have provided important insights into the key stages of replication of Hepadnaviridae. The initial host and organ-specific attachment of HBV to a hepatocyte is largely mediated by the large S protein, although cellular receptors have not been

defined (Bruss *et al.*, 1996). Similarly, the passage of the virion into the cell is not well understood. A recent study demonstrated that under conditions of low pH, a conformational change in the large S protein of DHBV occurs, which enables virions to bind to liposomes and intact cells (Grgacic and Schaller, 2000). However, previous studies have shown that uptake of the virus by hepatocytes is not dependent on an acidic endocytotic vesicle (Köck *et al.*, 1996; Hagelstein *et al.*, 1997).

The nucleocapsid is uncoated and directed to the nucleus where the DNA is released. The plus strand is filled by the DNA polymerase and the genome converted into a covalently closed circle (cccDNA) for transcription of genomic and subgenomic RNAs by cellular RNA polymerase II (Summers and Mason, 1982; Rall *et al.*, 1983; Miller *et al.*, 1984). These are transported to the cytoplasm for translation into the virion proteins and, in the case of the pregenomic RNA, for reverse transcription into the minus strand of the genome. This occurs following encapsidation of the pregenome. The HBV polymerase is thought to bind to a stem-loop structure at the 5' end (encapsidation signal ε) and this complex is then recognised by the self-assembling nucleocapsid (Knaus and Nassal, 1993). Due to its terminal redundancy, the pregenome also has an encapsidation signal structure at its 3' end. It has been hypothesised that the HBV polymerase must also require the 5' cap structure for encapsidation to occur (Jeong *et al.*, 2000). The pregenome is degraded by endogenous RNaseH activity during reverse transcription (Radziwill *et al.*, 1990).

Once plus strand synthesis has begun, the mature capsid is packaged in the envelope within the endoplasmic reticulum and disulphide bridges between the core and surface proteins maintain the virion structure. Again, the large S protein contains a sequence that is essential for virion assembly (Bruss, 1997). The mature capsids can follow a different route, back to the nucleus, if the large S protein is in short supply, thus increasing the pool of cccDNA molecules (Lenhoff and Summers, 1994). Virions are transported via the Golgi apparatus to the cell surface.

1.2.3 Subtypes and genotypes

The possible existence of antigenic heterogeneity within HBsAg was recorded soon after its discovery, as a result of immunodiffusion experiments (Raunio *et al.*, 1970). In addition to the *a* determinant, which is common to all variants, two pairs of mutually exclusive antigenic determinants were identified (*d/y* and *w/r*), giving rise to four defined subtypes (Le Bouvier, 1971; Bancroft *et al.*, 1972). This was increased to nine with the description of further subdeterminants. These are ayw_{1-4}, *ayr*, adw_2, adw_4, *adr q-* and *adrq+* (Magnius *et al.*, 1975; Couroucé *et al.*, 1976; Couroucé-Pauty *et al.*, 1978). DNA sequencing of the S gene has allowed the genetic basis of the subtypes to be determined (Magnius and Norder, 1995). However, antigenic subtypes do not reflect the phylogenetic relationships between HBV genomes and, therefore, classification into six genotypes (A-F) is now common (Okamoto *et al.*, 1988; Norder *et al.*, 1994). The designations are based on an intergroup divergence of at least 8%. Recently, using this criterion, a new genotype (G) has been identified (Stuyver *et al.*, 2000). Both genotypes and subtypes show clear differences in their geographical distribution, with genotypes A and D being the most prevalent in Europe (Magnius and Norder, 1995).

1.2.4 Epidemiology and transmission

As one of the most common infections in the world, hepatitis B is responsible for considerable levels of morbidity and mortality. It is currently estimated that there are 350 million chronic carriers of HBV, and one fifth of these are likely to die as a result of HBV induced liver disease (Kane, 1996). The global distribution is varied and, based on the prevalence of chronic carriers, countries are conventionally divided into areas of high (≥8%), medium (2-7%) and low (<2%) endemicity (Evans and London, 1998).

The predominant modes of HBV transmission are perinatal, parenteral and sexual. Although HBsAg or DNA has been detected in saliva, breast milk, urine and faeces, these are not thought to represent important vehicles of transmission (Boag, 1991). The relative importance of each transmission route depends upon the prevalence of chronic carriers in a particular area. Thus, in regions of high

endemicity, perinatal transmission maintains the pool of chronic carriers, and individuals have a greater than 60% lifetime risk of acquiring HBV (Alter, 1996). In the UK, where the prevalence is estimated to be <1%, sexual transmission and injecting drug use are the main routes of acquisition (Balogun *et al.*, 1999), with young adults aged 15-24 being at highest risk (Zuckerman, 1999*a*).

Needle exchange programmes and health campaigns encouraging safer sex were implemented in the mid-1980s in England and Wales to control the HIV epidemic; these measures also served to reduce the incidence of HBV infection. A recent survey of laboratory confirmed acute HBV infections revealed a large decrease from 1761 reported cases in 1985 to 583 in 1989 (Balogun *et al.*, 1999). The incidence remained stable until 1996, but has since risen significantly, with 719 cases reported in 1999. This rise has largely been attributed to an increase in transmissions amongst injecting drug users (IDUs) (Communicable Disease Surveillance Centre, CDSC, 1999). The true number of acute infections cannot be accurately assessed, as a large proportion of cases remain asymptomatic.

Although rare, nosocomial transmission continues to occur in the UK. Lapses in cross-infection control have resulted in the transmission of HBV in haemodialysis and organ transplantation units (Wreghitt, 1999; Tedder *et al.*, 1995). Similarly, surgeon to patient transmissions have been described, despite routine vaccination of all healthcare workers (The Incident Investigation Team and Others, 1997; Scottish Centre for Infection and Environmental Health, SCIEH, 1999).

Since the introduction of blood donation screening for HBsAg in the early 1970s, the risk of post-transfusion hepatitis caused by this agent has become very low (Soldan and Barbara, 1999). The introduction of diagnostic PCR has generated debate as to whether donations should now be screened for nucleic acids in an attempt to detect serologically negative cases (Roth and Seifried, 1999).

1.2.5 Outcomes of infection

The incubation period of HBV ranges from 45 to 120 days. The time between exposure and the first detectable serological response is known as the window period. This has been estimated to be in the region of 59 days to HBsAg positivity, with PCR to detect HBV DNA reducing it by around 1 week (Gerlich and Caspari, 1999; K. Soldan, pers. comm.). However, it is likely to vary widely between individuals. An acute phase follows, when HBsAg, HBeAg, HBV DNA and anti-HBc IgM antibodies are detected. The levels of aminotransferases, *eg.* alanine aminotransferase (ALT), in the blood are also raised, which provides a marker of liver injury. HBV infection can have a variety of clinical outcomes. Perinatally infected infants are generally asymptomatic, and only about one third of adults have specific symptoms. These include jaundice, malaise, nausea, and the excretion of dark urine and pale stools. HBV is not directly cytopathic: liver damage largely results from the cytotoxic T lymphocyte (CTL) response to infected hepatocytes (Chisari and Ferrari, 1995). In rare cases of acute HBV infection, liver necrosis is extensive and fulminant hepatitis develops.

Chronic HBV infection is operationally defined as the presence of HBsAg in serum for over 6 months. The major determinant of progression to chronic disease is the age of HBV acquisition. Around 90% of infants become chronic carriers, possibly due to neonatal tolerance to HBeAg epitopes. The risk of developing chronic infection drops to between 25 and 50% of children infected at age 1 to 5 years, and only 5 to 10% if infected as adolescents or adults (Beasley *et al.*, 1981; McMahon *et al.*, 1985).

The mechanisms for viral persistence in adults are under investigation. It has been noted that in adult immunocompetent patients who clear HBV after an acute infection there is a vigorous, polyclonal cellular immune response, involving CD4+ and CD8+ cells. This appears to be weak or interrupted in those who progress to chronic carriage. Other host and viral factors have been postulated to effect viral persistence, such as allelic differences in the HLA complex and the circulating HBV DNA level in early acute infection (Thio *et al.*, 1999; Roingeard. *et al.*, 1997). Mayerat *et al.* (1999) examined patients with acute and chronic hepatitis B infection in Switzerland, where genotypes A and D

predominate, and found that infection with genotype A was significantly associated with chronic outcome. However, no biological basis for this finding has been described.

The sequelae of HBV chronic infection can vary widely in severity, from chronic active hepatitis to an asymptomatic carrier, or tolerant, state. In general, carriers with detectable HBeAg have actively replicating HBV and are more likely to develop cirrhosis and hepatocellular carcinoma (HCC). Spontaneous seroconversion to anti-HBe positive has been reported in a varying proportion of carriers and is often associated with clearance of HBsAg and HBV DNA from the blood and remission of disease (Fattovich, 1996). However, using PCR, HBV DNA is still detectable in the liver of a proportion of chronic carriers who have cleared HBsAg, indicating that low level replication may continue (Mason *et al.*, 1998).

1.2.6 Prevention and treatment

Both passive and active immunisation are available for the prevention of HBV infection. Hepatitis B immunoglobulin (HBIG) was first prepared in the early 1970s by fractionation of pooled serum containing high titres of anti-HBs antibodies, and is effective in neutralising circulating virions if given soon after exposure (Grady *et al.*, 1978). Currently, owing to high costs, HBIG is given to specific risk groups only, such as infants born to HBeAg positive mothers (Wheeley *et al.*, 1991). The first plasma-derived vaccine of purified HBsAg subviral particles became commercially available in the UK in 1982, and was followed five years later by a recombinant yeast-derived HBsAg vaccine. When the recommended three doses are given at months 0, 1 and 6, a protective antibody response is induced in >90% of adults younger than 40 years (Andre, 1989).

In order to further improve the efficacy of vaccination and reduce the number of doses required, several new vaccine strategies are being examined. Small animal studies and clinical trials have shown that the inclusion of preS epitopes may be beneficial in terms of increased anti-HBs production, earlier onset of antibody appearance and response in previous non-responders to HBsAg vaccine only

(Suzuki *et al.*, 1994; Shouval *et al.*, 1994; Zuckerman *et al.*, 1997). Davis (1998) demonstrated the feasibility of DNA vaccines using eukaryotic expression vectors in mice and chimpanzees, and a novel approach using transgenic potatoes to produce oral HBV vaccine was recently reported (McNamee, 1999).

In 1992, the WHO recommended that all countries, regardless of endemicity, should introduce universal HBV vaccination into their national programmes of immunisation by 1997 (WHO, 1992). Over 80 countries have achieved this, although coverage varies widely (Kane, 1996). In the UK, control of HBV is based on the selective vaccination of high risk groups, such as injecting drug users, haemophiliacs, haemodialysis patients, homosexual men with multiple partners, prostitutes, infants born to HBsAg positive mothers, health care workers and contacts of patients with acute or chronic HBV infection (McMenamin, 1996). However, there is much evidence to suggest that this policy is failing as at-risk groups, particularly IDUs in the community and in prisons, are not being reached (Goldberg and McMenamin, 1998; Lamagni *et al.*, 1999). Universal antenatal screening for HBsAg is now recommended in an effort to reduce the pool of chronic carriers from whom perinatal transmission can ensue (Ramsay *et al.*, 1998).

Treatment strategies that aim to clear chronic carriers of the virus have had only limited success. Recombinant alpha interferon (IFN-α) was the first drug to be licensed for the treatment of chronic HBV infection. As a cytokine, it functions as an immune modulator and after 4-6 months of treatment can lead to a HBeAg to anti-HBe seroconversion rate of 30-40% (Wong *et al.*, 1993). However, it is associated with significant side effects, and a limited response rate in carriers who are in the tolerant (anti-HBe positive) phase (Brook *et al.*, 1989).

Several nucleoside analogues are currently under investigation for anti-HBV activity. Lamivudine has recently been licensed for clinical use. Lamivudine (or 3TC) is an enantiomer of 2'-deoxy-3'-thiacytidine, and thus competes with dCTP for incorporation into extending DNA chains. Once introduced, chain termination occurs, which *in vivo* results in a rapid decrease in the production of

new virions and a decrease in intracellular cccDNA (Severini *et al.*, 1995). Although less toxic than IFN-α, HBeAg seroconversion is only seen in approximately 16% of patients after 12 months of treatment, compared to 4% with placebo (Lai *et al.*, 1998). Combination therapy with other nucleoside analogues, such as famciclovir and adefovir dipivoxil, may improve outcome, although these agents act only to suppress viral replication and are therefore unlikely to be curative. Emerging therapies may prove to be more successful. For example, therapeutic vaccines may stimulate a HBV-specific CTL response, which is weak in chronic carriers (Pianko and McHutchison, 1999).

1.2.7 Variation in HBV

Many studies of HBV infection have focused on genomic variants in recent years in the hope that knowledge of their cause, prevalence and functional significance will aid the design of new therapies. Due to its unique method of replication, via an RNA intermediate, the HBV genome mutates at a higher rate than do those of other DNA viruses. The rate of nucleotide substitution per site per annum has been calculated to be 2×10^{-5} (Okamoto *et al.*, 1987), although this is likely to vary from patient to patient depending on their stage of infection. Also, variability is not seen randomly throughout the genome, due to the constraints imposed by the overlapping ORFs and the functions of the encoded proteins (Lauder *et al.*, 1993).

The observation that G→A hypermutation occurs suggests that mutations generally arise from a lack of proofreading by the P protein reverse transcriptase (Günther *et al.*, 1997). Hypermutation is a common feature associated with retroviral reverse transcriptases.

In addition to point mutations that occur during replication, variation can be generated in a number of ways, such as insertion and deletions due to RNA splicing (Terre *et al.*, 1991), template switching during reverse transcription, and recombination between different genotypes (Bollyky *et al.*, 1996; Bowyer and Sim, 2000).

Certain point mutations are epidemiologically significant. One of the most common is a G→A at position 1896, which produces a stop codon at residue 28 of the precore protein and thus prevents the translation of HBeAg. It was first detected in patients who had appeared to seroconvert to anti-HBe, but had severe liver disease (Carman *et al.*, 1989). Since then, the mutation has been found frequently in patients with long-standing infection, with and without active hepatitis. In such patients, the mutant often becomes the dominant circulating type after seroconversion (Hamasaki *et al.*, 1994). Many cases of acute fulminant hepatitis have been linked to infection with this variant, although it can also cause a self-limiting hepatitis, suggesting that other factors must be involved in the production of severe acute disease (Tanaka *et al.*, 1995; Karayiannis *et al.*, 1995; Mphahlele *et al.*, 1997).

That the G1896A mutation emerges at all is still not understood; possible explanations are enhanced replication, immune selection and hypermutation (Carman *et al.*, 1998; Ngui *et al.*, 1999). It appears mainly in genotypes B, C, D and F, due to structural constraints imposed by the RNA encapsidation signal, and its prevalence is therefore geographically restricted (Lok *et al.*, 1994).

HBV carrying the G1896A mutation has been associated in recent years with several transmissions of HBV from surgeons to patients in the UK (The Incident Investigation Team and Others, 1997; SCIEH, 1999). Recognition that anti-HBe positive carriers can have a high level of circulating virus has led to proposals that new guidelines be introduced to restrict practices of HBV positive surgeons. The challenge is to decide at which titre they become effectively uninfectious (Ballard and Boxall, 1999). New guidance from the Department of Health recommends that health care workers whose HBV DNA concentration in the blood exceeds 1000 genome equivalents per ml should not perform exposure-prone procedures (Department of Health, 2000).

Mutations in the S gene, which result in vaccination failure in infants born to HBsAg positive mothers, have been studied intensively in recent years. The most commonly described vaccine-escape mutant has a single nucleotide substitution

resulting in a G to R change at amino acid 145 (Carman *et al.*, 1990). Several other mutations have now been described within the major hydrophilic region (MHR) of HBsAg that also appear to result in vaccine failure, with concurrent circulation in the infected vaccinees of HBsAg and anti-HBs antibodies (Wallace and Carman, 1997; Chong-Jin *et al.*, 1999). On examination of maternal blood, the mutant is often absent, or present in very low amounts, suggesting that its appearance in the infant is due to a rapid selection (Gunther *et al.*, 1999*a*; Ngui *et al.*, 1997).

Transmission studies have demonstrated that the G145R mutant is stably transmitted and does not revert to wild type during chronic carriage (Ogata *et al.*, 1997), thus raising concerns that vaccine-escape mutants may emerge. Mathematical modelling predicts that this will be the case given certain different assumptions about the infectivity of the mutant strains, and the extent of protection afforded against them by current vaccines. However, it is forecast that emergence would not occur for at least 50 years (Wilson *et al.*, 1998). In order to monitor the prevalence of HBsAg mutants, modifications to current serological tests may be necessary, particularly where monoclonal antibodies are used, as altered epitopes can result in false negative results (Carman *et al.*, 1995; Zuckerman and Zuckerman, 1999).

The introduction of nucleoside analogues for the treatment of chronic HBV infection has been followed swiftly by the description of resistant mutants. Lamivudine resistance is mediated by mutations in the polymerase gene, most commonly in the YMDD locus. This locus falls within the catalytic region of the RTase/DNA polymerase domain of the HBV P protein, and is well conserved in all viral reverse transcriptases (Bartholomeusz *et al.*, 1998). The M residue, which is at codon 552, is substituted by V or, more rarely, I (Hussain and Lok, 1999). Identical mutations at the YMDD locus of the HIV reverse transcriptase have also been reported, in variants that are resistant to lamivudine (Tisdale *et al.*, 1993). In HBV, a further substitution at codon 528 from L to M has been reported in patients carrying the YVDD motif (Niesters *et al.*, 1998).

Transfection studies have confirmed that the YMDD locus mutations bestow lamivudine resistance in both HIV and HBV, but they also reduce replication competence, which may explain why wild type virus re-emerges on withdrawal of therapy (Back *et al.*, 1996; Ling and Harrison, 1999; Ono-Nita *et al.*, 1999). Das *et al.* (2000) have predicted that steric hindrance occurs between the β-methyl group of V or I, at codon 552 of HBV in the mutant polymerase, and the sulphur atom of lamivudine. This would account for the observed cross resistance to another nucleoside analogue, emtricitabine, which has a similar oxathiolane ring.

The L528M mutation has also been detected in patients with breakthrough infection following famciclovir therapy, suggesting a degree of cross resistance with lamivudine (Tillman *et al.*, 1999). The importance of other reported famciclovir-induced substitutions has not yet been established, although the YMDD locus does not appear to be involved (Gunther *et al.*, 1999*b*; Seignères *et al.*, 2000). Further work will be needed in order to devise effective combination therapies for the control of HBV replication in long term carriers.

1.3 TT Virus

1.3.1 Genetic organisation

TTV has been characterised as an non-enveloped single–stranded DNA virus with a caesium chloride density of 1.31-1.34 g/cm^3 (Mushahwar *et al.*, 1999). Filtration studies indicate that the virion is 30-50 nm in diameter, although it has yet to be visualised (Mushahwar *et al.*, 1999). Several complete TTV DNA sequences have now been published, revealing a circular genome that ranges from 3808 to 3852 nucleotides in length (Erker *et al.*, 1999; Okamoto *et al.*, 1999*b*; Hijikata *et al.*, 1999). Nuclease protection assays show that the genome is negative-stranded (Mushawar *et al.*, 1999).

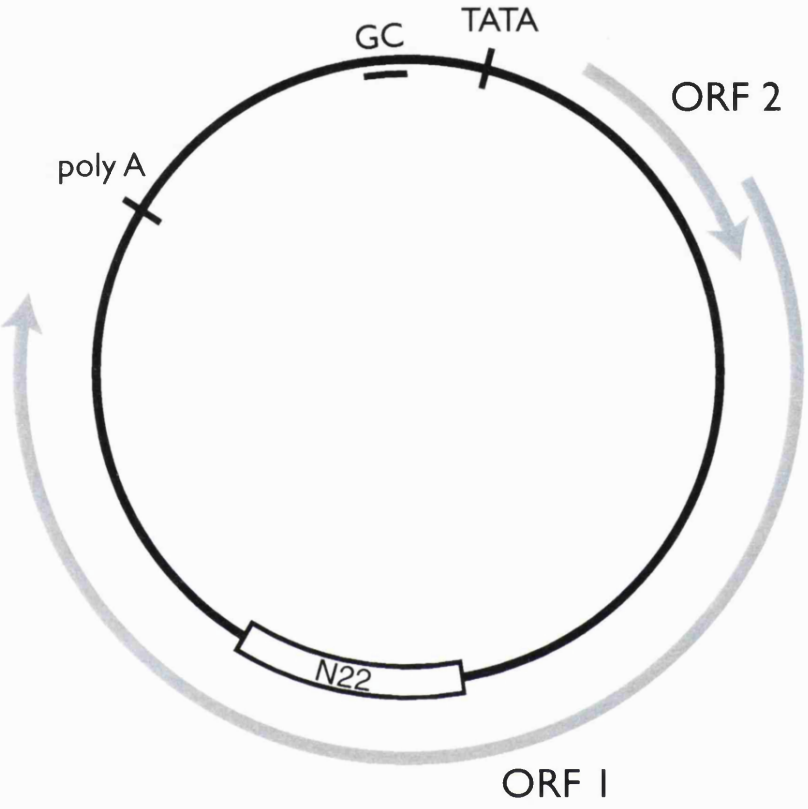
A number of regulatory sequences and two conserved ORFs have been identified on the positive strand of the genome-length sequences described so far (Miyata *et al.*, 1999; Mushahwar *et al.*, 1999; Erker *et al.*, 1999). The longest ORF, ORF-

1, codes for approximately 770 amino acids, with a smaller ORF upstream that overlaps ORF 1 by around 130 nucleotides (ORF2). Two putative third ORFs have been proposed, also on the positive strand (Erker *et al.*, 1999; Takahashi *et al.*, 2000*b*). Searches of protein databases have not revealed any significant similarities to known amino acid sequences. It is assumed therefore that TTV does not encode a polymerase and must rely on host cell polymerases for replication. The protein products of TTV have yet to be expressed in cell culture.

TTV was initially reported to share some physical characteristics with the *Parvoviridae*, although no nucleotide similarity was demonstrated. A resemblance between TTV and chicken anaemia virus (CAV), a circovirus, has been described by Takahashi *et al.* (1998*a*) on the basis of the presence of arginine-rich regions at the N-terminal of ORF-1 from both viruses. As ORF-1 is thought to code for the 51.7 kDa capsid protein in CAV, this similarity has led to the suggestion that ORF-1 of TTV is the capsid gene. Furthermore, CAV contains motifs at the carboxyl end of ORF-1 that may be involved in rolling circle replication, two of which have been identified in ORF-1 of TTV (Niagro *et al.*, 1998; Mushahwar *et al.*, 1999). A likeness in the genomic arrangements of the two viruses has also been noted (Miyata *et al.*, 1999). However, no significant sequence similarity between TTV and any other member of the *Circoviridae* has been shown, and this has led Mushahwar *et al.* (1999) to propose that TTV be placed in a new virus family called *Circinoviridae*.

Recently, three novel human DNA virus sequences (CBD203, CBD231 and CBD279) have been described that appear to be intermediately related to TTV and CAV (Takahashi *et al.*, 2000*a*). They have been named TTV-like mini virus (TLMV). It is not clear from these reports whether TTV, CAV and TLMV belong to one family of viruses.

Figure 1.2 The genetic organisation of the TTV genome. The N22 region was the first TTV DNA to be discovered using RDA (Nishizawa *et al.*, 1997). Conserved features shown are ORFs 1 and 2 (shaded arrows), TATA box (TATA), polyadenylation signal (polyA) and GC-rich region (GC).



1.3.2 Genotypes

Considerable genetic variability of TTV has been demonstrated, with a 222-bp fragment of the genome from ORF-1 showing as much as 65% divergence between sequences. Several groups have assigned up to 16 genotypes based on the sequence variability in short TTV PCR products (Y. Tanaka *et al.*, 1998; Okamoto *et al.*, 1999*c*; Khudyakov *et al.*, 2000). In contrast, phylogenetic analysis of full-length, or near full-length TTV sequences reveals division into three main types, represented by the prototype TA278 in group 1, US35 in group 2 and JA10 in group 3 (Erker *et al.*, 1999). Most of the sequence variability

occurs in the coding regions. In addition to these three TTV groups there are other highly divergent complete sequences (Okamoto *et al.*, 1999*b*; Hijikata *et al.*, 1999; Takahashi *et al.*, 2000*b*; Ukita *et al.*, 2000). PCR cloning studies have found mixed infections, with up to seven different genetic types, to be a common event (Takayama *et al.*, 1999; Niel *et al.*, 2000). Different TTV sequences do not appear to show geographical segregation, as is seen with other bloodborne hepatitis viruses (Mushahwar *et al.*, 1999).

1.3.3 Epidemiology and transmission

TTV was initially presented as a candidate hepatitis virus, as TTV DNA titres were found to correlate closely with aminotransferase levels in three post-transfusion non A-G hepatitis patients (Nishizawa *et al.*, 1997). TTV DNA has since been amplified from a high percentage of patients with post-transfusion hepatitis and those with chronic liver disease (Charlton *et al.*, 1998; H. Tanaka *et al.*, 1998). It has also been detected in the liver of infected patients at 10-100 times the concentration in serum, suggesting that the virus is hepatotropic (Okamoto *et al.*, 1998*a*). However, more extensive studies have revealed that TTV is highly prevalent in healthy individuals in many parts of the world, so its ability to cause disease is far from certain (Prescott and Simmonds, 1998; Hsieh *et al.*, 1999; Maggi *et al.*, 1999; Abe *et al.*, 1999). PCR assays with primers directed against more conserved genomic regions have led to reports of prevalences approaching 100% in some population groups (Takahashi *et al.*, 1998*b*; Leary *et al.*, 1999*b*). In the UK, the prevalence of TTV DNA in blood donors was 18% using two primer systems (F. Bobet, pers. comm.).

TTV DNA has been detected in blood products that have undergone viral inactivation procedures. Thus, its prevalence in haemophilia patients is high (Gerolami *et al.*, 1998). Transmission by blood transfusion has also been reported (Fujiwara *et al.*, 1998). However, TTV DNA has been found in faeces (Okamoto *et al.*, 1998*b*), breast milk (Toyoda *et al.*, 1999), semen (Inami *et al.*, 2000) and saliva (Ross *et al.*, 1999) and the age-distribution of infected individuals suggests that TTV is mainly a community-acquired virus (Yamada-Osaki *et al.*, 1998; Davidson *et al.*, 1999). Evidence for transplacental transmission has also been documented (Morrica *et al.*, 2000). Non-parenteral

modes of transmission may therefore account for its high global prevalence (Zuckerman, 1999*b*). The detection of human-like TTV sequences in domestic farm animals raises the possibility that they may serve as a source of human infection (Leary *et al.*, 1999*a*). TTV-like DNA sequences have also been identified in nonhuman primates (Verschoor *et al.*, 1999; Okamoto *et al.*, 2000*a*; Abe *et al.*, 2000; Romeo *et al.*, 2000; Cong *et al.*, 2000).

Both acute and chronic TTV infections have been documented, with DNA persisting for up to 22 years in one study (Matsumoto *et al.*, 1999; Lefrere *et al.*, 2000). On examining the sequence of fragments of the genome, several groups have found it to remain very stable over a number of years (Biagini *et al.*, 1999; Ball *et al.*, 1999). This would be expected if TTV were using host cell polymerases to replicate, due to the low mutation rate of human DNA polymerase enzymes. However, it suggests that in order to generate the observed extensive genetic variability, TTV must have been present in the human population for many thousands of years. This has led to the consideration that TTV is part of the normal human flora (Simmonds *et al.*, 1999).

1.3.4 Disease association

It is now accepted that TTV is unlikely to cause liver disease in most infected individuals. Numerous studies have failed to find any association between TTV DNA positivity and acute sporadic hepatitis, fulminant hepatitis, HCC, or exacerbations of chronic hepatitis (Parquet *et al.*, 1999; Skelton *et al.*, 2000; Huang *et al.*, 2000; Giménez-Barcons *et al.*, 1999). Preliminary attempts to detect anti-TTV antibodies have used immunoprecipitation techniques, and, as antigen, an expressed truncated ORF-1 protein product (Tsuda *et al.*, 1999; Handa *et al.*, 2000). These methods indicate that antibody is formed and can be detected in patients with and without TTV viraemia. In an attempt to find the site of replication of TTV, Okamoto *et al.* (2000*b* and *c*) have identified double-stranded forms of the viral genome in hepatocytes and bone marrow cells, which they suggest are the replicative intermediates. TTV DNA has also been detected in liver biopsy material by *in situ* hybridisation (Rodríguez-Iñigo *et al.*, 2000). A pathogenic role for TTV has yet to be described. It is possible that certain genetic

types are able to cause disease, or that certain hosts, *eg.* immunocompromised patients, are susceptible to potential pathogenic effects of the virus.

1.4 Molecular epidemiological tools

1.4.1 Gel-based mutation screening

Numerous gel-based assays have been developed in order to screen PCR products for mutations, without the need for sequencing. These have many applications, but in virology they are employed for genotyping, investigating transmission events, screening for epidemiologically important mutants, and for the examination of quasispecies evolution (Cotton, 1993; Arens, 1999). Some methods are sensitive to single base pair mutations. The most commonly used are described here:

i) RFLP - restriction fragment length polymorphism

This method employs restriction endonucleases to cleave PCR products and the resulting fragments are separated by gel electrophoresis. It is often used for genotyping, where a specific pattern of digested fragments is obtained from each genotype. For this reason it is important to amplify a relatively conserved region of the viral genome where the restriction sites are likely to remain stable (Pohjanpelto *et al.*, 1996; Lindh *et al.*, 1998) A large amount of sequence data is required in order to design a RFLP assay, and the panel of endonucleases must be customised. RFLP can also be used to screen for a known mutation.

ii) HMA - heteroduplex mobility assay

When a mixed population of PCR products are denatured and allowed to reanneal, a mixture of homoduplex (both strands have the same sequence) and heteroduplex (there is one or more mismatch) double strands are formed. When electrophoresed in a non-denaturing gel, the heteroduplexes are retarded in relation to the homoduplexes and the distance between the two species is related to the number of mismatched base pairs (Barlow *et al.*, 2000). This technique has been used to detect single base pair mutations in the human cystic fibrosis gene (White *et al.*, 1992), although in other applications, only about 80% of mutations are said to be detectable (Cotton, 1993). HMA has been successfully used for

genotyping HIV, where tester PCR products of known genotype are mixed with a PCR product amplified from a sample of unknown genotype (Tatt *et al.*, 2000).

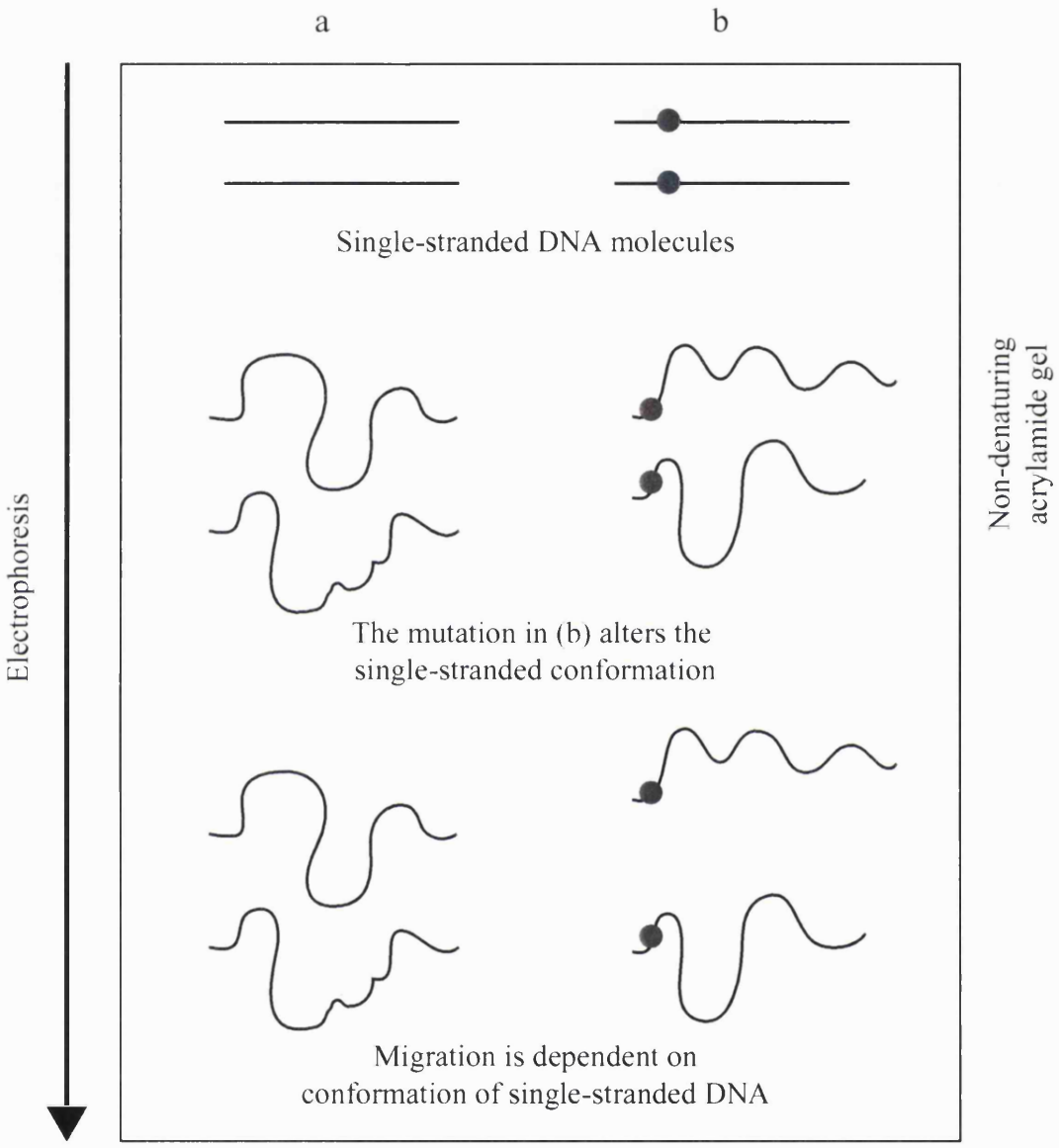
iii) BESS - base excision sequence scanning

This is a relatively new technique that uses automated sequencing technology. PCR products are amplified with one fluorescently labelled primer in a PCR mix containing dUTP. They are then subjected to digestion with uracil-N-glycosylase and endonuclease IV, generating a defined series of fragments which can be separated and detected on a standard sequencing gel (Hawkins and Hoffman, 1997). Approximately 95% of mutations can be detected and localised when both strands of the PCR product are analysed. The data are suitable for phylogenetic analysis (see section 1.4.2), which is useful for grouping specimens *eg.* from an outbreak or from different geographical locations (Charrel *et al.*, 1999).

iv) SSCP - single strand conformation polymorphism

Many studies of sequence polymorphisms in eukaryotic and prokaryotic genes have utilised this method, which is based on the principle that in a non-denaturing gel single-stranded PCR products will fold into distinct secondary structures depending on their nucleotide sequence. The differences in conformation alter electrophoretic mobility, which is detected as different banding patterns (Orita *et al.*, 1989). This is depicted in Figure 1.3. SSCP assays have been used to investigate nosocomial outbreaks of HBV (Yusof *et al.*, 1994; Hardie *et al.*, 1996), although detection of single base pair mutations is again not always possible (Cotton, 1993).

Figure 1.3 Schematic diagram demonstrating the principle of SSCP analysis.



v) DGGE - denaturing gradient gel electrophoresis

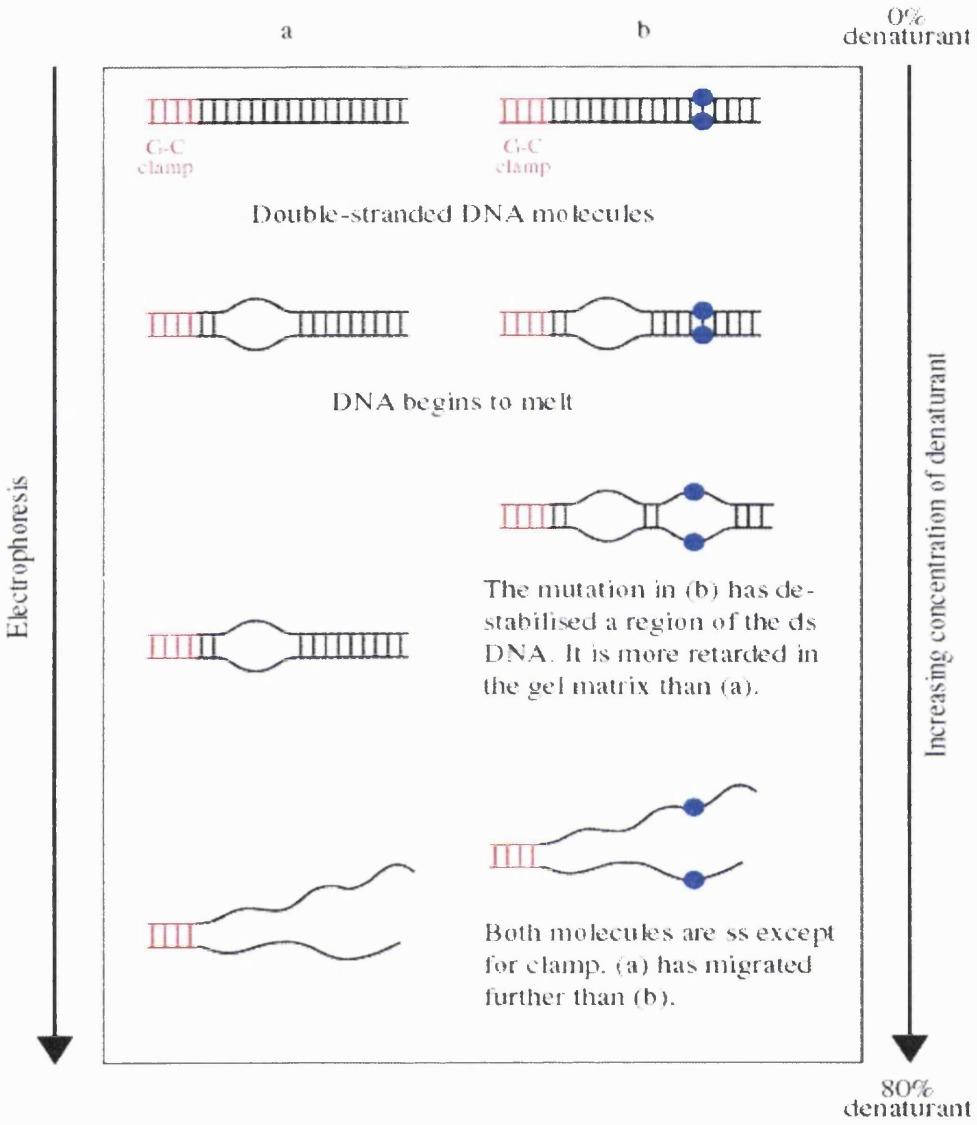
This method was developed primarily for the analysis of genetic mutations in human disease and can be used to detect single base pair mutations. It is based on the differential migration of double stranded DNA through a gel containing

increasing concentrations of the denaturing chemicals formamide and urea (Myers *et al.*, 1987). As electrophoresis progresses, each DNA fragment will reach a position where the portion of the fragment with the lowest melting temperature (T_m) will begin to unravel, causing a significant decrease in electrophoretic mobility. The T_m of any domain is dependent on its nucleotide sequence and therefore mutations are detected as differences in band positions on the gel. If a mutation is present in the domain with the highest T_m, it will not be detected, as sequence-dependent migration is lost when the strands become completely dissociated. To improve sensitivity a 'GC clamp' is usually added to one end of the PCR product to provide a domain with a very high T_m that will not become denatured under DGGE conditions (Sheffield *et al.*, 1989; Fodde and Losekoot, 1994). The DGGE principle is shown in Figure 1.4. It has been used recently in the study of JC virus molecular epidemiology, and to examine HCV quasispecies (De Santis and Azzi, 2000; Harris and Teo, 2001).

1.4.2 Phylogenetic analysis

DNA sequencing inevitably provides more information about a viral genomic fragment than any gel-based assay. Sequence data can be submitted to phylogenetic analyses in order to answer questions about the genetic relatedness of homologous DNA. In virology, it aids in the genotyping of virus families, enables molecular epidemiological studies to be carried out, and allows the evolution of viruses to be investigated (Holmes and Garnett, 1994). Of relevance to public health is the application of phylogenetic analyses to the investigation of microbial transmissions, such as that of HIV from a dentist (Crandall, 1995), HBV from a blood donor (Zuckerman *et al.*, 1995), and HCV from contaminated blood (McAllister *et al.*, 1998).

Figure 1.4 Schematic diagram demonstrating the principle of DGGE analysis.



Phylogenetic methods were originally developed and applied to the study of eukaryotic evolution using data such as morphological characters and gene frequency data (Edwards and Cavalli-Sforza, 1964; Camin and Sokal, 1965). The advent of DNA sequencing has resulted in a rapid accumulation of molecular data, which can be analysed in a similar way to answer questions about the relatedness of particular sequences to others, and to understand evolutionary relationships. The algorithms used to infer phylogenies are continually being refined and updated, and are now applied regularly to viral nucleotide and amino acid sequences.

There are several points to consider when drawing and interpreting a phylogenetic tree, perhaps the most important being that all trees represent only estimates of the real genetic relationship between sequences. In order to maximise the accuracy of a tree, it is important to use models that approximate the true behaviour of DNA. Similarly, the quality of the data entered will affect the tree topology obtained. Sequences should come from homologous regions, and be as similar in length as possible. If too short, or lacking in variation, there will be insufficient 'phylogenetic information', but if sequences are too variable the true relationships may be masked by background noise (Clewley, 1998).

The first stage in the analysis of nucleotide or amino acid data is to produce an alignment of a set of homologous sequences. Several algorithms are available for this purpose, the most useful for evolutionary studies being the global methods (Weiller *et al.*, 1995). These determine an optimal alignment, according to their parameters, across the whole length of the sequences by carrying out pairwise alignments, identifying the two most closely related sequences, and adding the rest in order of similarity. Gaps are introduced where necessary to accommodate insertions or deletions.

There are many methods for constructing a phylogenetic tree from a sequence alignment. The most commonly used programs fall into three major categories:

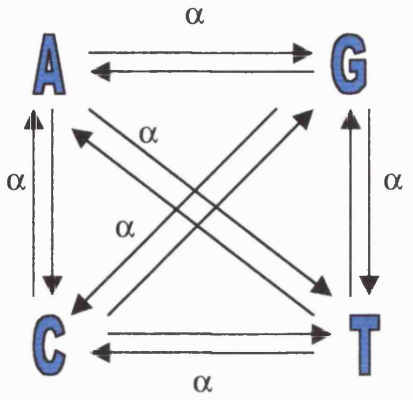
i) Maximum Parsimony – These methods search for the tree that requires the minimum number of changes at each character (nucleotide or amino acid) site to explain the evolution of the group, and only phylogenetically informative sites are considered. The main failure of parsimony is that it cannot easily allow for homoplasy, *i.e.* convergent mutations, parallel mutations and reversals, which result from the independent acquisition of the same properties along different evolutionary lineages. This can lead to unrelated sequences being clustered together (Stewart, 1993). To help address this problem, weights can be applied to rare substitution events, or to different codon positions to improve the model parameters.

ii) Distance methods – These involve two main stages; firstly a distance matrix is calculated by carrying out pairwise comparisons of the input sequences. There are several available models to estimate genetic distances, many of which consider transitions and transversions separately (2-parameter models), and can add a correction for the possible occurrence of multiple substitutions at the same site (Jukes and Cantor, 1969; Kimura, 1980; Felsenstein, 1981; Hasegawa *et al.*, 1985; Yang *et al.*, 1994). The principles of some of the commonly used nucleotide substitution models are represented in Figure 1.5.

Secondly, a clustering algorithm is applied to the data in the matrix to produce a tree. The simplest program of this type is UPGMA (unweighted pair group method with arithmetic means), which looks for the smallest pairwise distance, then considers these two most related sequences as a single taxonomic unit. The distances are recalculated between this cluster of two and each of the remaining sequences, and the process is repeated (Sneath and Sokal, 1973). The topology of the tree is an arbitrary result of the clustering algorithm. Other distance methods, such as the Fitch and Margoliash method (1967) consider the difference between the expected distance (that in the matrix) and the observed distance (the sum of the branch lengths) between any two sequences, and search possible trees in order to choose the one with the smallest difference. Another commonly used method, the neighbour-joining method, uses a clustering algorithm which adds pairs of sequences (neighbours) to a star-like tree, minimising the internal branch length at each addition (Saitou and Nei, 1987). Thus, the most parsimonious tree that fits the distance matrix is drawn.

iii) Maximum Likelihood – The starting point when using these methods is a probabilistic model of sequence evolution. The programs then compute the likelihood of any tree, which is the probability of the data, given the tree, and therefore search for the tree with the maximum likelihood (Felsenstein, 1988). These methods allow statistical comparisons between generated trees more easily than the other methods, although they are not generally used for large data sets, being computationally very demanding. For example, for ten sequences the possible number of rooted trees is 34,459,425 (Li, 1997), and for each tree this method calculates the likelihood value.

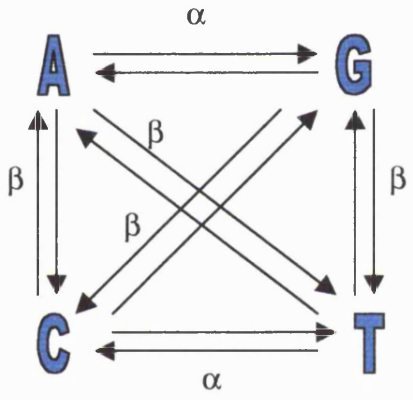
Figure 1.5 Different models of nucleotide substitution



All substitutions are equally likely (rate = α).

Equal base frequencies = Jukes-Cantor model (JC)

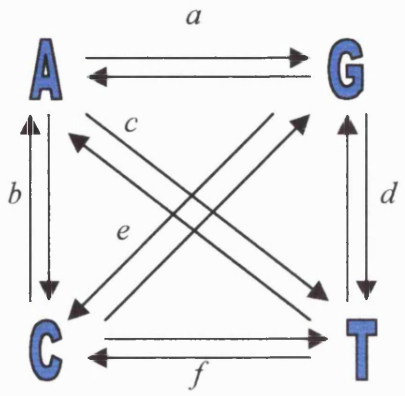
Unequal base frequencies = Felsenstein model (F81)



Transitions and transversions have different rates (α and β).

Equal base frequencies = Kimura 2 parameter model (K2P)

Unequal base frequencies = Hasegawa, Kishino and Yano model (HKY85)



All six pairs of substitutions have different rates (*a*, *b*, *c*, *d*, *e* and *f*).

Unequal base frequencies = General reversible model (REV)

Therefore, when more than 10 sequences are entered it is necessary to apply a search mechanism to reduce the number of trees considered. As there is no natural way to arrange trees in order to search, this is often carried using 'perturbation' methods, such as taxon swap or tree bisection and reconnection (TBR). These take a possible tree, rearrange it and then recalculate the likelihood to see if the new tree is an improvement. One downfall is the potential for ending up with a locally optimum tree that is neither the true tree nor the most likely.

Most of the phylogenetic methods can be shown to provide the correct tree, given a large amount of data that does not breach any of the assumptions in the model. However, real data sets often contain complexities that cannot be readily handled by current phylogenetic methods. For example, there may be a base composition bias, which would lead to sequences with a similar composition being grouped together, even though they may not be the most closely related. A difference in frequency of types of mutation events, and an unequal mutation frequency at different sites are commonly seen in DNA data. It is therefore important to use the weighting facilities in many programs to counteract these occurrences. The simplest clustering methods, such as UPGMA, do not perform well if the rate of evolution varies along different lineages in the same tree. Also, the assumption of independence among sites is likely to be inaccurate, particularly in viral sequence data. Structures within DNA, such as stem-loops, and the existence of overlapping open reading frames mean that nucleotide substitutions at affected sites will be constrained (Hillis *et al.*, 1994).

Resampling methods, such as bootstrapping and jackknifing can be used to examine the robustness of a tree, and statistical methods have been developed to test the reliability of a phylogeny (Felsenstein, 1988). More sophisticated techniques are becoming available for inferring phylogenies. For example, the use of split decomposition can take into account conflicting alternative groupings by displaying results as a network, rather than the typical bifurcating tree (Bandelt and Dress, 1992; Dopazo *et al.*, 1993). This technique has a particular application when examining sequences arising from recombination events. To increase the flexibility of sequence alignment, Hidden Markov Models are being

developed, which can 'learn' from a given data set and assess the probability that a substitution, insertion or deletion will occur at each site (Weiller *et al.*, 1995).

1.4.3 Recombination detection

The exchange of genetic material between viruses of the same species is increasingly thought to play a role in the evolution of many virus groups, and methods derived from phylogenetic analysis programs are being employed to study its occurrence. The simplest of these start with a multiple alignment of homologous sequences and use a "sliding window" of around 50-200 bp to scan the alignment for conflicting relationships. For example, if sequence *a* is highly related to sequence *b* over 80% of the fragment, but is shown to be more closely related to sequence *c* over the remaining 20%, this would suggest that recombination has occurred. Methods, such as bootscanning, TOPAL, RIP, and RDP, can give an indication of the breakpoints and often draw simple dendrograms using the putative recombinant section to graphically represent the change in branch position that would occur in comparison to a tree produced from the whole sequence alignment (Salminen *et al.*, 1995; McGuire and Wright, 1998; Worobey and Holmes, 1999). If a potential recombinant and both parental sequences can be identified in an alignment, more sophisticated tests can then be applied to test the significance of suggested breakpoints, such as LARD, which is based on a maximum likelihood model (Holmes *et al.*, 1999).

1.5 Aims

The broad aim of this thesis was to use appropriate molecular epidemiological tools to answer specific questions about defined variants of two hepatotropic viruses; HBV and TTV.

1.5.1 Molecular epidemiology of two variants of HBV

A screening assay was developed, using DGGE, to investigate the prevalence of a common HBV variant in two population groups throughout the 1990s. The hypothesis tested was that the variant had spread from prisons into the community, resulting in an increased prevalence in a population local to the prisons.

The same assay was used to examine the prevalence of a second variant that was thought to be responsible for a point source outbreak of acute hepatitis B. Social and behavioural factors that may explain the contrasting epidemiologies of the two variants were then analysed.

1.5.2 Molecular characterisation of a variant of TTV

A novel divergent TTV-like sequence, identified in a patient with non-A to E hepatitis, was characterised and compared with previously described TTV sequences. After sequencing the complete genome, phylogenetic techniques were used to examine the evolutionary relationship between this virus and other TTV isolates. Its potential place within the *Circoviridae* was also investigated. PCR and sequencing were applied to search for this TTV-like virus in blood donors and other patients with non-A to E hepatitis in an attempt to define its hepato-pathogenic role.

Chapter 2

Materials and Methods

2.1 Amplification of HBV surface and core fragments

2.1.1 DNA extraction

Nucleic acid extraction from 100 µl serum or plasma was carried out using a guanidinium thiocyanate-silica method (Boom *et al.*, 1990). Buffers L6, L2 and silica suspensions were prepared by Severn Biotech Ltd. (Kidderminster, UK). Nucleic acids were eluted in 50 µl nuclease-free water.

2.1.2 PCR conditions

The primary PCR mix for the amplification of surface gene and core gene fragments contained 1.5 mM $MgCl_2$, 200 µM each dNTP, 1 unit *Taq* polymerase (Life Technologies, Paisley, UK), 20 pmoles each outer primer, 10 µl extracted DNA and nuclease-free water to give a final volume of 50 µl. Amplification was carried out by heating to 94°C for 2 m, followed by 30 cycles of 94°C for 30 s, 55°C for 30 s, 72°C for 60 s, and a final extension at 72°C for 2 m. Secondary PCRs were carried out under the same conditions, except for the addition of 2 µl primary PCR product as template, and 20 pmoles each inner primer. The primer sequences are shown in Table 2.1. The outer surface primers amplified a 501 bp fragment between positions 236 and 736 of the HBV genome, and the inner primers amplified a 477 bp fragment between positions 236 and 712. The outer core primers amplified a 270 bp product between positions 1891 and 2160, and the inner primers amplified a 258 bp product between positions 1897 and 2154. Nucleotide positions are numbered from the *Eco*R1 site (Ono *et al.*, 1983).

Table 2.1 Oligonucleotide primers used for the amplification of HBV surface and core fragments

Region	Position	Name	Primer Sequence
Surface	Outer – sense	S5*	5' ATACCACAGAGTCTAGACTC
	Outer – antisense	S6*	5' AACTGAAAGCCAAACAGTGG
	Inner – sense	109‡	5'ATACCACAGAGTCTAGACTCGTGGTGGACT
	Inner – antisense	585R‡	5' AAGCCCTACGAACCACTGAACAAATGGCAC
Core	Outer – sense	1763‡	5' GCTTTGGGGCATGGACATTGACCCGTATAA
	Outer – antisense	2032R‡	5' CTGACTACTAATTCCCTGGATGCTGGGTCT
	Inner – sense	1778-E‡	5' GACGAATTCCATTGACCCGTATAAAGAATT
	Inner – antisense	2017R-B‡	5' ATGGGATCCCTGGATGCTGGGTCTTCCAAA

*Indicates taken from Yusof *et al.* (1994)

‡Indicates taken from Kaneko *et al.* (1989)

2.1.3 Detection of PCR products

PCR products were mixed with 2 µl loading buffer (10% Ficoll-400 in 10 mM Tris-HCL pH 7.5, 50 mM EDTA, containing 0.25% bromophenol blue and 0.25% orange G) and electrophoresed through a 2% agarose gel (SB fine gel, Severn Biotech Ltd.) in 1 X TRIS-borate-EDTA (TBE) buffer. One µg of a 1 kb molecular weight marker (Life Technologies) was loaded next to test samples in order to estimate the size of PCR products in base pairs. To stain the DNA, gels were incubated in a solution of ethidium bromide in TBE buffer (concentration 5 µg/ml) for 20 m. PCR products were visualised using a short wave ultraviolet transilluminator.

2.2 Single strand conformation polymorphism (SSCP) assay

2.2.1 Radioisotopic PCR

Primary PCR for the amplification of HBV surface and core fragments was carried out as described in section 2.1.2. The secondary PCR mix for both

surface and core fragment amplification contained 1.5 mM $MgCl_2$, 70 μM each dNTP, 1 unit *Taq* polymerase, 20 pmoles each inner primer, 2 μCi ^{32}P-deoxycytosine 5'-phosphate (Amersham Pharmacia Biotech, Little Chalfont, UK), 2 μl primary PCR product and nuclease-free water to give a final volume of 20 μl. Cycling conditions were as described in section 2.1.2.

2.2.2 Preparation of non-denaturing polyacrylamide gel

A Macrophor sequencing tank (LKB Bromma, Sweden) was used for the SSCP gel. The water-jacketed thermostatic plate was first coated with 10 ml "repel-silane" (2% solution of dimethyldichlorosilane in 1,1,1-trichloroethane, BDH, Poole, UK). The glass plate was coated in 10 ml "bind-silane" mix (10 ml ethanol, 300 μl 10% acetic acid and 30 μl γ-methacryloxypropyltrimethoxysilane (Sigma-Aldrich, Poole, UK). The acrylamide gel mix contained 0.5 X MDE (Flowgen, Ashby de la Zouch, UK), 0.6 X TBE, 0.05% glycerol and distilled water to give a final volume of 80 ml. Polymerisation was achieved by adding 36 μl N,N,N',N'-tetramethylethylenediamine (TEMED, Life Technologies) and 360 μl 10% ammonium persulphate (APS, Sigma-Aldrich) to the gel mix immediately before pouring using a 50 ml syringe. Wells were formed with a 20 or 26 well comb, and the gel was allowed to polymerise for at least 1 h.

2.2.3 Loading and electrophoresis

The surface and core radioactive PCR products from each sample were pooled for simultaneous analysis, and mixed with an equal volume of loading buffer (98% deionised formamide, 20 mM EDTA, 0.05% bromophenol blue and 0.05% xylene cyanol) in 0.6 ml tubes. The samples were denatured at 95°C for 5 m and then placed on ice before loading.

Electrophoresis was carried out in 0.6 X TBE buffer at 15°C and at constant wattage (8 W) for 18 h. The temperature was maintained using a thermostatic circulator (Pharmacia LKB Biotechnology, Uppsala, Sweden). On completion, the thermostatic plate was removed, the gel covered with a transparent plastic wrap and exposed to autoradiographic film (Hyperfilm-MP, Amersham Pharmacia Biotech).

2.2.4 Development of the autoradiograph

The autoradiographic film was immersed sequentially in developer, fixative (Kodak, Hemel Hempstead, UK) and water and, when dry, was examined on a X-ray viewer.

2.3 Sequencing PCR products

2.3.1 Automated sequencing

Secondary PCR products were electrophoresed through a 2% agarose gel, and stained with ethidium bromide. Bands were excised and DNA recovered using the Igenie DNA extraction kit (Helena Biosciences, Sunderland, UK). Products were sequenced using both inner primers with the ABI Prism DNA sequencing kit or the ABI BigDye DNA sequencing kit (PE Applied Biosciences, Warrington, UK). Reactions were electrophoresed on either an ABI 373 or an ABI 377 automated sequencer (PE Applied Biosciences).

2.3.2 Analysis of sequence data

Chromatogram data was analysed using SeqEd™ v1.0.3 (PE Applied Biosciences) to obtain a consensus sequence from both primers. The consensus was trimmed to remove the primer sequences, then multiple alignments and phylogenetic trees were drawn in the MegAlign program from the LASERGENE package (DNASTAR Inc., Madison, WI, USA). This program uses the Clustal V algorithm to align sequences, and a neighbour-joining method to produce a phylogenetic tree. Sequence distances are expressed as percent nucleotide divergence.

2.4 Denaturing gradient gel electrophoresis (DGGE) screening

2.4.1 GC-clamp PCR

The primary PCR mix for both surface and core fragment amplification contained 1 X EXPAND PCR buffer (which included 1.5 mM $MgCl_2$), 200 µM each dNTP, 0.7 unit EXPAND high fidelity polymerase mix (Roche Diagnostics Ltd., Lewes, UK), 20 pmoles each outer primer (Table 2.1), 10 µl extracted DNA

and nuclease-free water to give a final volume of 50 µl. PCR cycling conditions were as described in 2.1.2. Secondary PCRs were carried out under the same conditions, except for the addition of 2 µl of primary PCR product as template, and 20 pmoles each inner primer. A 40 bp GC-clamp was attached to the 5' end of the inner sense primers (Sheffield *et al.*, 1989). Therefore their sequences were as follows:

Surface: 109-clamp
5'CGCCCGCCGCGCCCCGCGCCCGTCCCGCCGCCCCCGCCCGATACCACAGAGTCTAGACTCGTGGTGGAC
Core: 1778-E-clamp
5'CGCCCGCCGCGCCCCGCGCCCGTCCCGCCGCCCCCGCCCGGACGAATTCCATTGACCCGTATAAAGAATT

The inner antisense primers were as shown in Table 2.1. The secondary surface and core 'clamped' PCR products were 517 bp and 298 bp respectively, and they were initially visualised on agarose gels as described in 2.1.3.

2.4.2 Preparation of denaturing gradient polyacrylamide gels

Stock solutions of each acrylamide mix were prepared in deionised water. For the analysis of HBV clamped surface gene PCR products, a gradient of 10% to 60% denaturants was used. The 10% denaturant mix contained 8% acrylamide (National Diagnostics, Hull, UK), 0.6 X TRIS-acetate-EDTA (TAE) buffer (Life Technologies), 4% formamide (BDH) and 0.7 M urea (Sigma-Aldrich). The 60% denaturant mix contained 8% acrylamide, 0.6 X TAE buffer, 24% formamide and 4.2 M urea. For the analysis of HBV clamped core gene PCR products, a gradient of 10% to 60% denaturants was also used, but both stock solutions contained 12% acrylamide.

The gradient gel was poured using 30ml each stock solution and a gravity driven gradient maker with pump drive (GRI, Braintree, UK). Polymerisation was achieved by adding 13 µl TEMED and 250 µl 10% APS per 30 ml stock solution before pouring. Wells were formed with a 48 well comb, and the gel was allowed to polymerise for at least 1 h.

2.4.3 Loading and electrophoresis

Clamped surface and core PCR products (2-7µl of each) were mixed with an equal volume of loading buffer (10% Ficoll-400 in 10 mM Tris-HCL pH 7.5, 50 mM EDTA and 0.25% xylene cyanol FF) and loaded into the wells. Electrophoresis was carried out in 0.6 X TAE buffer at 60°C and at 100 V for 18 h. Buffer was continuously circulated to maintain constant temperature throughout the electrophoresis tank (Ingeny PhorU, GRI).

2.4.4 Staining the gels

Gels were stained with SYBR Green I (Flowgen) at a concentration of 0.5 X in 200 ml of 0.6 X TAE. Bands were visualised after 1 h by ultraviolet transillumination. Gel images were obtained using a Kodak Digital Science DC40 camera.

2.5 Statistical analysis

The chi square test (uncorrected) and chi square test for linear trend were performed, using StatCalc in the Epi-Info 2000 package (CDC, Atlanta, USA).

2.6 Amplification of TTV N22 and non-coding region (NCR) fragments

2.6.1 DNA extraction

Nucleic acid extraction from 100 µl serum or plasma was carried out as described in section 2.1.1. DNA was also extracted from archived paraffin–embedded liver biopsy samples (Coombs *et al.*,1999). Small sections were agitated in 100 µl Tween-20 (Sigma-Aldrich) before being heated to 90°C for 10 m. They were then cooled to 55°C and 2 µl of 10 mg/ml Proteinase K (Life Technologies) was added. The sections were incubated at 55°C for 4 h with gentle agitation every hour. After digestion, the sections were heated to 99°C with 100 µl 5% Chelex-100 resin (Bio-Rad Laboratories, Hemel Hempstead, UK) in 50 mM Tris-HCL pH 7.5, 1 mM EDTA for 10 m. They were agitated

again and spun in a microcentrifuge for 15 m at 10500 g before being placed on ice. This allowed the paraffin wax to harden in a disk above the digested section and be removed with a pipette tip. After heating to 45°C, 100 µl chloroform was added. Following gentle agitation and a further centrifugation for 15 m at 10500 g, the upper phase (approximately 150 µl) was removed.

2.6.2 PCR conditions

The primary PCR mix for the N22 region (within ORF-1) amplification contained 2.0 mM $MgCl_2$, 200 µM each dNTP, 1 unit *Taq* polymerase (Life Technologies), 25 pmoles each outer primer, 10 µl extracted DNA and nuclease free water to give a final volume of 50 µl. Amplification was carried out by heating to 95°C for 30 s, followed by 35 cycles of 95°C for 30 s, 60°C for 45 s, 72°C for 45 s, and a final extension at 72°C for 7 m. Secondary PCRs were carried out for 25 cycles under the same conditions, except for the addition of 2 µl of primary PCR product as template, and 25 pmoles of each inner primer. The primer sequences are shown in Table 2.2. The outer N22 primers amplified a 286 bp fragment between positions 1900 and 2185 of the prototype TTV sequence (TA278, GenBank accession no. AB017610). The semi-nested inner primers amplified a 271 bp fragment between positions 1915 and 2185.

The primary PCR mix for the NCR amplification contained 2.0 mM $MgCl_2$, 200 µM each dNTP, 1 unit *Taq* polymerase, 25 pmoles each outer primer, 10 µl extracted DNA and nuclease-free water to give a final volume of 50 µl. Amplification was carried out by heating to 94°C for 30 s, followed by 35 cycles of 94°C for 30 s, 55°C for 30 s, 72°C for 30 s, and a final extension at 72°C for 7 m. Secondary PCRs were carried out for 35 cycles under the same conditions, except for the addition of 2 µl of primary PCR product as template, and 25 pmoles of each inner primer. The primer sequences are shown in Table 2.2. The outer NCR primers amplified a 306 bp fragment between positions 3087 and 3392, and the inner primers amplified a 243 bp fragment between positions 3120 and 3362.

Table 2.2 Oligonucleotide primers used for the amplification of TTV N22 and NCR fragments

Region	Position	Name	Primer Sequence
N22	Outer – sense	NG059*	5' ACAGACAGAGGAGAAGGCAACATG
	Inner – sense	NG061*	5' GGCAACATGYTRTGGATAGACTGG
	Outer/Inner – antisense	NG063*	5' CTGGCATTTTACCATTTCCAAAGTT
NCR	Outer – sense	BF1‡	5' GTGGGACTTTCACTTGTCGGTGTC
	Outer – antisense	BR1‡	5' GACAAATGGCAAGAAGATAAAGGCC
	Inner – sense	BF2‡	5' AGGTCACTAAGCACTCCGAGCG
	Inner – antisense	BR2‡	5' GCGAAGTCTGGCCCCACTCAC

*Indicates taken from Okamoto *et al.* (1998*a*)

‡Indicates taken from Leary *et al.* (1999*a*)

2.6.3 Detection of PCR products

PCR products were detected by agarose gel electrophoresis and ethidium bromide staining, as described in section 2.1.3.

2.7 Cloning N22 PCR products

Secondary N22 PCR products were cloned using a TOPO TA Cloning Kit (Invitrogen BV, De Schelp, The Netherlands). This method worked on the principle that PCR products generated by *Taq* polymerase possess a single adenosine (A) overhang at the 3' end, due to the template independent terminal transferase activity of the enzyme (Marchuk *et al.*, 1991). Therefore, the plasmid vector was supplied as a linear molecule with complementary 3' thymidine (T) overhangs. Ligation of the PCR products into the vector was achieved by covalently bound topoisomerase I (Shuman, 1994). Transformation was carried out in chemically competent *E.coli*, and standard blue/white colony screening was employed after an overnight incubation on L agar (Luria agar: 1.0% tryptone, 5.0% yeast extract, 1.0% 0.17 M NaCl, 1.5% agar, pH 7.0), containing

50 mg/ml ampicillin. Colony PCRs were carried out using the secondary PCR conditions for amplification of this region.

2.8 Sequencing PCR products

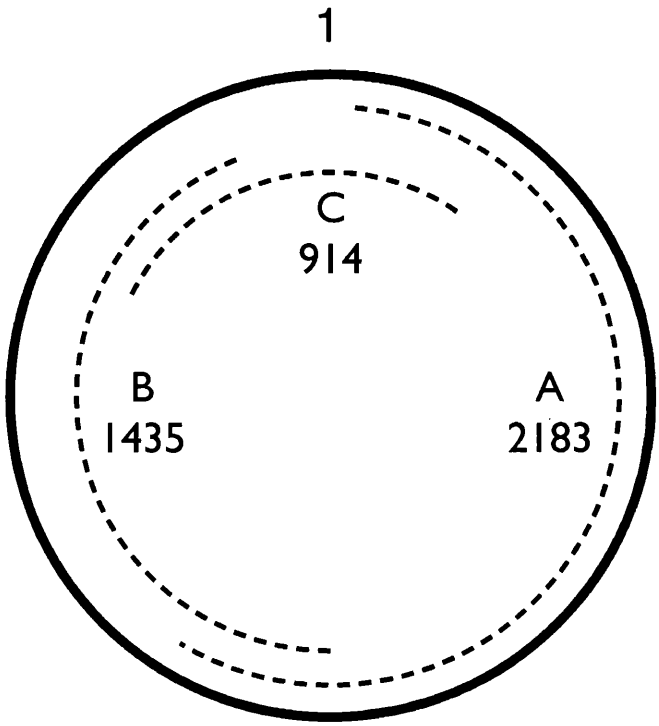
2.8.1 Automated sequencing

PCR products (either amplified directly from the sample or from colonies after TA cloning) were sequenced as described in section 2.3.

2.8.2 TTV genome sequencing

PCR products covering the whole genome of a divergent type of TTV (named PMV) were generated using an EXPAND High Fidelity PCR System (Roche Diagnostics Ltd.) according to the manufacturer's protocol for amplification of products up to 3 kb, and with a Clontech Advantage-GC 2 PCR kit (Clontech Laboratories Ltd., Basingstoke, UK) for the GC-rich fragment C. All PCRs were carried out in a 50 µl volume. Primary or single round PCRs contained 10 µl of extracted DNA as template; secondary PCRs used 2 µl of primary PCR product. The PCR primers used to generate the entire PMV genome sequence were: for fragment A, T801 and NG063 (Takahashi *et al.*, 1998*b*; Okamoto *et al.*, 1998*a*); for fragment B, NG059, NG061 and BR1 (Okamoto *et al.*, 1998*a*; Leary *et al.*, 1999*a*); for fragment C,
outer sense INV1 (5'CCTTACAGACACCCCTTACTACCCT),
with outer antisense INV2 (5' CAGTGGCACTTTCCTTTCTTTC),
and inner sense INV3 (5' ACTAAGCACTCCGAGCGAAGC)
with inner antisense INV4 (5' ATAACCCTAAGAGCCTTGCCCATAG).
The positions of the three PCR products are indicated in Figure 2.1. All three were cloned using an Invitrogen TOPO TA Cloning Kit. Plasmid DNA containing the inserted PCR products was purified using a QIAfilter Plasmid Midi Kit (Qiagen Ltd. Crawley, UK) and sequencing was carried out by Cambridge Bioscience Ltd. (Cambridge, UK). The resulting three fragments were assembled using programs EditSeq and MegAlign in the LASERGENE package (DNASTAR Inc.) to construct the complete genome sequence of 3736 nucleotides.

Figure 2.1 Positions of the 3 overlapping PCR products in the PMV genome (fragments A, B and C), including their length in bp.



2.8.3 Analysis of sequence data

Additional phylogenetic software packages were used for the analysis of sequence data. Multiple alignments were performed using Clustal W (Thompson *et al.*, 1994), which aligns nucleotides or amino acids and produces guide trees using the UPGMA cluster analysis method of Sneath and Sokal (1973). PUZZLE v4.0.2 (Strimmer & von Haeseler, 1997) was used to estimate the transition/transversion (ts/tv) ratio, and base frequencies, from nucleotide alignments. These values were inputted into subsequent programs to improve the accuracy of the tree-drawing. Phylogenetic analysis was performed using the following programs in the PHYLIP package: Seqboot, Dnadist, Protdist, Neighbor, Fitch, Dnaml, Consense (Felsenstein, 1993).

In certain cases, the program Seqboot was first applied to an alignment in order to test the robustness of the final tree. Seqboot samples columns from the alignment randomly with replacement, so that the resulting alignments are the same length as the original, with some columns excluded and others duplicated.

Between 100 and 1000 new alignments are typically created, depending on the number and length of the sequences. This type of resampling method is likely to simulate the amount of variation expected if new data sets were used, although one of the assumptions is that evolution occurs independently at each site, which is not always the case.

Dnadist was used to calculate genetic distances between sequences in an alignment, the output taking the form of a distance matrix, which was then used in two further programs, Neighbor and Fitch, to draw a tree. Three different models of nucleotide substitution were available: Jukes and Cantor (1969), Kimura (1980) and a maximum likelihood model (Felsenstein, 1993). The Kimura 2-parameter model was used most frequently, which allows for a difference between transition and transversion rates. The distance for each pair of sequences estimates the total branch length between the two, which in turn is an estimate of the divergence time between those two sequences.

Genetic distances generated by Dnadist are approximately comparable to the percent nucleotide divergence figures obtained using the MegAlign program (see Section 2.3.2). For example, a genetic distance of 0.3 is equivalent to a nucleotide divergence of about 30%. It is possible to obtain estimates of over 100% divergence, or genetic distances of greater than 1, as most programs that estimate evolutionary distances add a correction for the occurrence of multiple nucleotide substitutions at the same site (multiple hits).

The program Protdist also creates a distance matrix, but uses an amino acid alignment, and three different models of amino acid replacement to do so. The most widely used, and that used in this study, is the Dayhoff PAM model, which allows prediction of the probability of changing from any one amino acid to any other, based on an empirically derived model (Dayhoff, 1978). Again, the distance matrix was used as input data in the clustering programs Neighbor and Fitch.

In Neighbor it is possible to implement both the neighbor-joining algorithm (Saitou and Nei, 1987) and the UPGMA algorithm (Sneath and Sokal, 1973),

whereas Fitch carries out the Fitch-Margoliash and least-squares methods (Fitch, and Margoliash, 1967). Fitch is the more time-consuming program to run, but also the more accurate. It contains a function which allows "global rearrangements" to be carried out, enabling each possible group on the final tree to be removed and re-added to assess possible improvements to the tree topology. This function was selected each time Fitch was used.

Dnaml is a maximum likelihood method written by Felsenstein (1993). It was used to generate trees to compare with those resulting from the distance methods described above. If similar topology and branch lengths are obtained using these two different methods, the accuracy of the tree is thought to be highly supported. The transition/transversion ratio and base frequencies calculated in PUZZLE were entered, and the global rearrangements option was selected. However, bootstrapping was not carried out prior to using Dnaml, as the length of time required to analyse multiple data sets was too great.

The program Consense was used when bootstrapping had been performed to draw a consensus tree from those produced by the multiple data sets. The consensus tree has a number at each node, which indicates how many times the group descended from that node occurred. This was converted into a percentage and shown on the tree. All tree diagrams were generated using Treeview (Page, 1996).

In order to find homologous DNA or protein sequences, the GenBank/EMBL and SwissProt databases were searched using BLASTN, BLASTP and FASTA (Altschul *et al.*, 1997; Pearson and Lipman, 1988).

Secondary structure predictions were made using the program Mulfold (Jaeger *et al.*, 1989), which uses energy minimisation criteria to predict optimal and suboptimal secondary structures for both RNA and DNA.

To plot sequence similarity across multiple alignments the GCG program Plotsimilarity (Genetics Computer Group, WI, USA) was used. This program calculates the average similarity of all nucleotide or amino acid sequences in an

alignment at each character position, using a sliding window of comparison. The window is moved along all sequences, one position at a time, and the average similarity over the window is plotted at the middle position of the window. For DNA alignments, a sliding window of 50 nucleotides was employed, whereas a sliding window of 20 amino acids was used when carrying out the analysis on protein sequences. The average similarity across the entire alignment is plotted as a dotted line.

The Recombinant Identification Program (RIP) was used to investigate the possibility that PMV contained mosaic regions. This computer program was developed at the HIV Database to identify recombinant HIV-1 sequences and is available at their website (http://hiv-web.lanl.gov/, Los Alamos National Laboratory, Los Alamos, NM, USA). However, by entering representative TTV sequences as background reference sequences, it was adapted to scan PMV for regions of similarity with each. On entering an alignment of PMV with the background sequences, the program proceeds by using a sliding window to assess the similarity (the percentage of identical base pairs) of the query sequence (PMV) to each background sequence. The values are retained and the window is moved along the alignment by one nucleotide. The program then displays either a graphical or alignment output revealing which background sequence the query sequence most resembles at all possible positions. So called "best matches" are marked if they are significant according to a statistical test. The threshold for statistical significance was set at 95%. Various window sizes were used in order to maximise the sensitivity of the result.

Chapter 3

The molecular epidemiology of two variants of HBV

3.1 Introduction

Acute hepatitis B virus infections in England and Wales are reported to the Public Health Laboratory Service Communicable Disease Surveillance Centre (PHLS CDSC). A laboratory confirmed acute case is defined by a positive test result for anti-HBc IgM antibodies, or a positive test result for HBsAg with typical hepatitis symptoms and, where available, a negative result for anti-HAV IgM antibodies. These cases, along with those identified during contact tracing, are used to estimate the incidence of HBV infection and to monitor epidemiological trends (Balogun *et al.*, 1999).

Symptomatic infections can only be used as a proxy marker of incident infections. The true incidence is likely to be much greater than the incidence of reported cases, as up to two thirds of infections in adults are either asymptomatic or cause general flu-like symptoms (McMahon *et al.*, 1985). These cases are therefore unlikely to be investigated for possible infection with hepatitis B virus.

A survey of laboratory confirmed hepatitis B infections over an 11 year period revealed a large decrease from 1761 reported in 1985 to 583 in 1989, which has largely been attributed to a change in injecting drug use behaviour as a result of the HIV epidemic (Balogun *et al.*, 1999). However, since the mid-1990s the incidence has shown a steady increase: 719 cases were reported in 1999. Several factors may be facilitating this rise. For example, heroin and crack use is known to have increased throughout the 1990s, particularly in north and south west England (Dean, 1999), with 1 in 5 IDUs reporting having shared equipment in 1995 (Department of Health, 1996). The failure of the selective HBV vaccination program to reach IDUs has ensured that they remain a largely susceptible, high risk group (CDSC, 1999; Goldberg and McMenamin, 1998; Lamagni *et al.*, 1999). Moreover, the number of gonorrhoea diagnoses in England rose substantially between 1995 and 1999, in both males and females. This indicates that high risk sexual behaviour potentially plays an important role in maintaining HBV endemicity (Fenton *et al.*, 2000).

Inside prisons, the problems in preventing bloodborne virus transmission are magnified. Several studies have shown that IDUs often continue to inject whilst in prison, that they are more likely to share equipment and that up to a quarter of IDU inmates begin injecting during a period of incarceration (Pickering and Stimson, 1993; Gore *et al.*, 1995; Weild *et al.*, 2000). It is clear that the transmission of HBV can occur in this setting; 274 cases of acute infection were diagnosed in prisons in England and Wales between 1985 and 1996 (Balogun *et al.*, 1999). However, few outbreaks have been described, the most comprehensive report being of a HBV and HIV outbreak within a Scottish prison (Taylor *et al.*, 1995; Hutchinson *et al.*, 1998).

Clusters of HBV infection are more often reported from nosocomial and community-based settings. In these situations PCR and DNA sequence analysis are frequently applied to investigate chains of transmission and to try and link infections to a point source. This can enable failures in cross-infection control to be identified. A transmission is said to have taken place if the same, or a very closely related, HBV sequence is recovered from both putative transmitter and recipients. For example, this has occurred in the UK during surgery from four carrier surgeons to their patients (The Incident Investigation Teams, 1997), and also via a contaminated cryopreservation tank (Hawkins *et al.*, 1996).

As a reference laboratory, CPHL frequently receives specimens that require specialist testing. The Hepatitis Unit has offered a DNA sequencing service for a number of years to aid in the investigation of potential nosocomial transmission of HBV and to examine unusual clusters of acute hepatitis B infection in the community. PCR products are generated and sequenced (as described in Sections 2.1 and 2.3) and then added to a multiple alignment for phylogenetic comparison. If identical sequences are retrieved, and a recent transmission is therefore supported, appropriate measures are recommended to prevent further infections.

Samples are usually received from hospital laboratories and information supplied typically includes the patient's name, date of birth, date of onset, potential risk factors and reason why sequencing is requested. The results of serological tests

already carried out may also be included. However, not all of this data may be available. A large database of HBV sequences has now been established in the Hepatitis Unit, against which all new sequence requests are compared.

Techniques that were developed to screen for mutations in eukaryotic genes, e.g., RFLP, SSCP and HMA, may be adapted to facilitate molecular epidemiological studies of HBV without resort to nucleotide sequencing. This chapter describes the application of a SSCP assay and the development and application of a DGGE procedure to study the molecular epidemiology of two defined variants of HBV. One variant was first identified as being responsible for multiple outbreaks of acute hepatitis B in prisons, and was subsequently detected in a number of settings during routine molecular investigations, carried out during the 1990s. The DGGE approach was successfully applied to track its spread into the wider community. The other variant caused a large outbreak of acute hepatitis B in a nosocomial context. Application of the DGGE procedure allowed the spread of this variant to be closely studied. These molecular epidemiological studies permitted conclusions to be drawn regarding the roles of socio-cultural settings and lifestyles on the further transmission of HBV, following its introduction into the community.

3.2 The emergence and spread of the HBV prison variant (HBV^{PV})

3.2.1 HBV outbreaks in prisons in the north of England

Between 1993 and 1995, serum specimens were received from cases of acute hepatitis B infection in inmates of three prisons in the north of England. Five samples were received from a prison in Durham, six from Hull and four from Lancaster. The samples were sent by the local Public Health Laboratories, with requests from the prison doctors to investigate possible chains of transmission. It was thought that the clusters were unrelated. Two fragments from the surface and core genes of the HBV genome were amplified by PCRs described in Section 2.1.2. One sample (prisoner 6 HUL) was negative in the core region PCR. On DNA sequence analysis (as detailed in Section 2.3), eleven of the 15 surface gene fragments, originating from all three prisons, were found to be identical. Ten of these also demonstrated complete sequence identity in the core

gene fragment; the remaining sequence had a single nucleotide difference. It was therefore concluded that there was a transmission link between the prisons. The multiple alignments of nucleotide sequences are shown in Figure 3.1.

In 1994, 18 HBsAg positive serum samples were received from a prison in Liverpool. Fourteen were positive by PCR (Section 2.1.2), seven of which originated from known IDUs. Serological tests, performed by Liverpool Public Health Laboratory, revealed that nine of the PCR positive inmates were undergoing acute HBV infection at the time of serum sampling, three were at the recovery stage of acute infection, one was a known carrier and there was one for whom no information was provided.

To screen for nucleotide sequence identity among the PCR products, radioisotopic SSCP analysis was carried out as described in Section 2.2. This technique can distinguish between most PCR products with different sequences as, on denaturation, the single strands adopt distinct 3-dimensional conformations, which can be resolved following electrophoresis through a non-denaturing polyacrylamide gel (Yusof *et al.*, 1994). A representative autoradiograph derived from the SSCP gel is shown in Figure 3.2.

Figure 3.1a Nucleotide alignment of surface gene PCR product sequences amplified from prison samples. DUR=Durham; HUL=Hull; LAN=Lancaster.

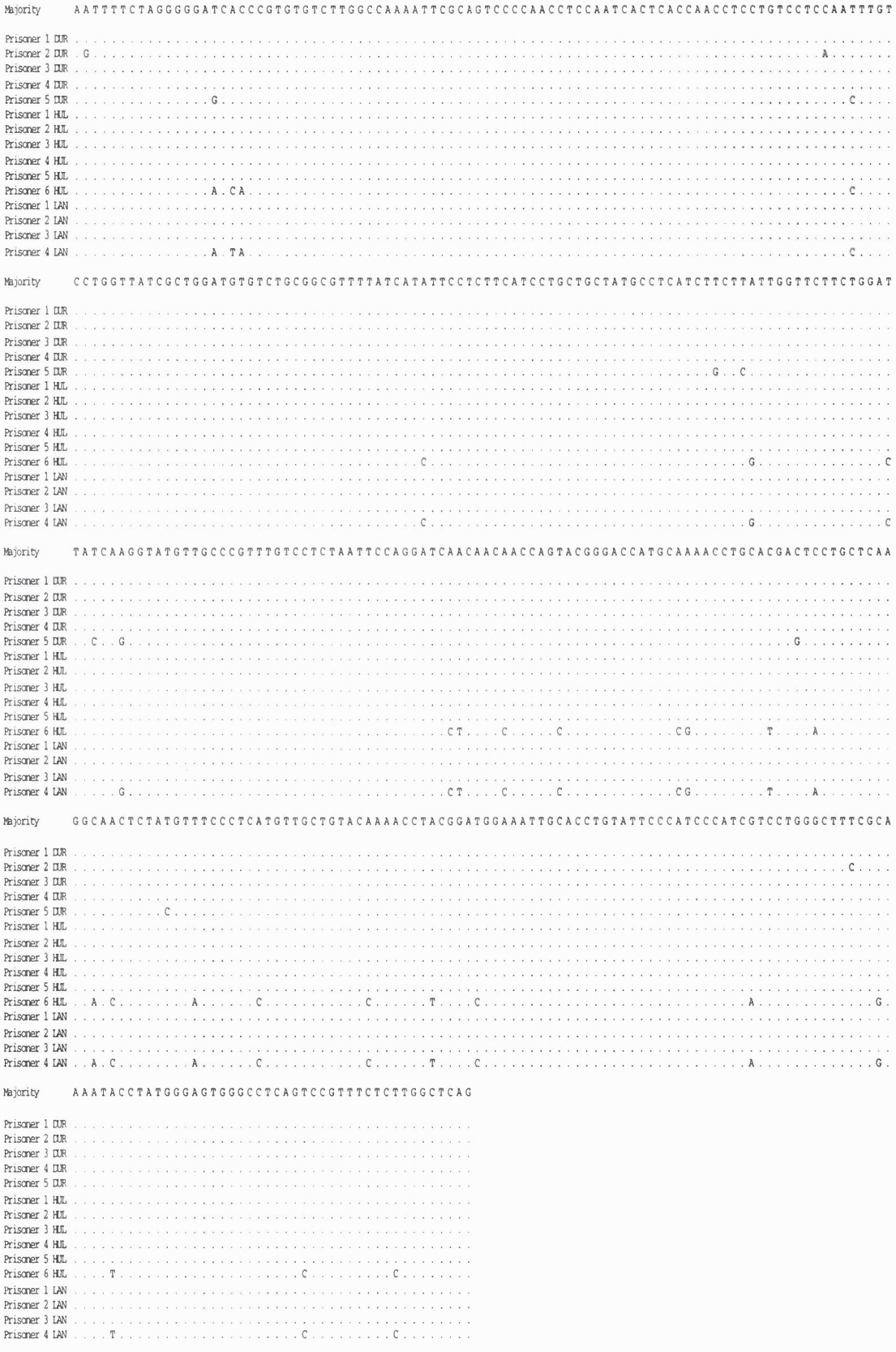


Figure 3.1b Nucleotide alignment of core gene PCR product sequences amplified from prison samples. DUR=Durham; HUL=Hull; LAN=Lancaster.

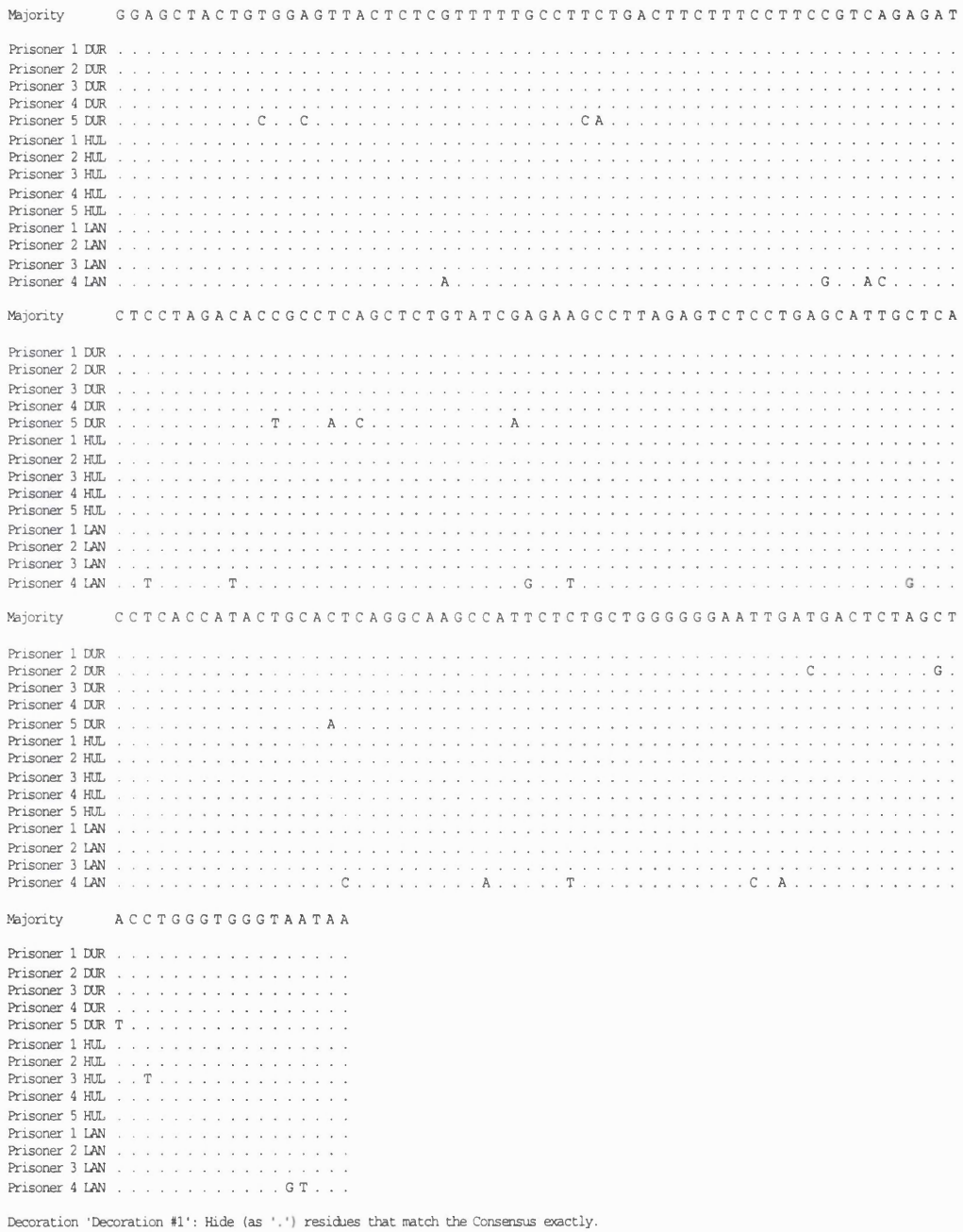
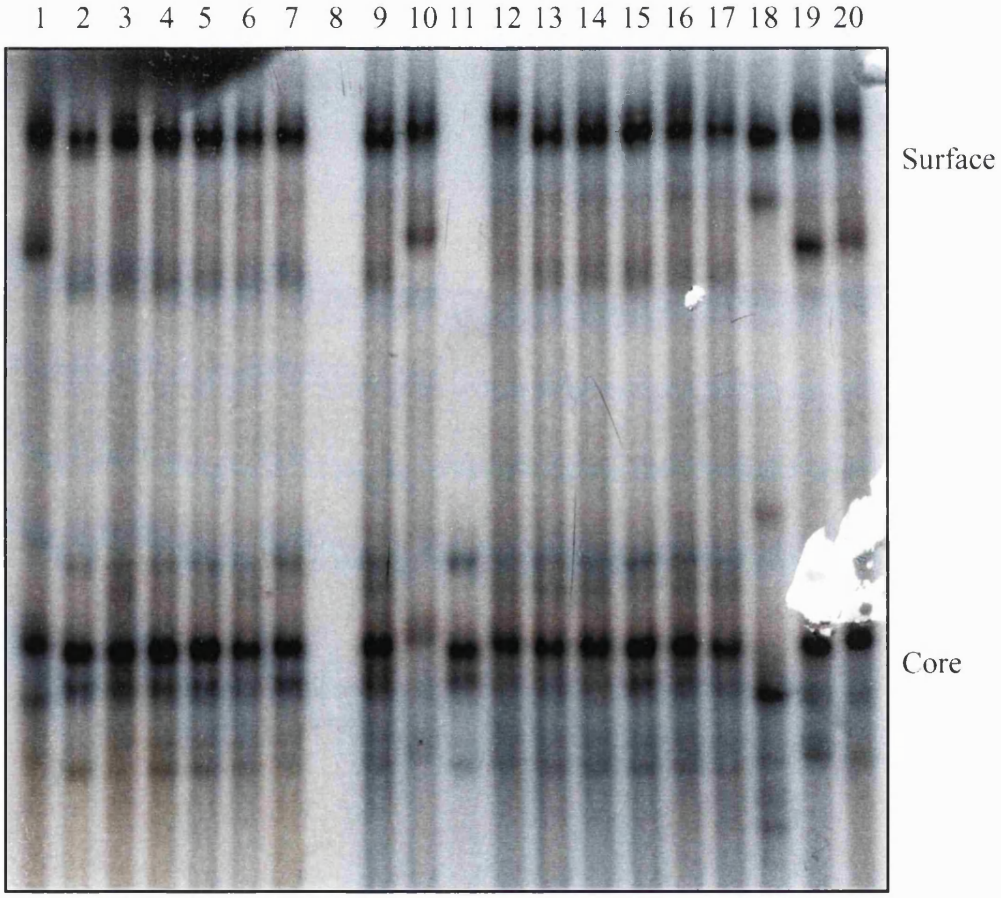


Figure 3.2 Autoradiograph of SSCP assay on samples from a cluster of acute HBV infections in a Liverpool prison, and controls. The HBV surface and core PCR products derived from each sample were run in the same lane as indicated. Lanes 1, 10 and 20 were amplified from the plasmid control pHBV130; Lanes 2-7, 9, 11-17 were amplified from 14 inmates and lanes 18 and 19 originate from epidemiologically unlinked controls.



As shown in Figure 3.2, 12 samples, in lanes 2-7, 9 and 13-17, appeared identical in both regions of the genome analysed. The other two were identical in the core region only; one was negative in the surface region PCR (lane 11) and the other had a different banding pattern in the surface region (lane 12).

Three of the 12 samples with both matching fragments were sequenced to confirm the SSCP result (prisoners x, y and z, corresponding to lanes 6, 9 and 15 on the autoradiograph in Fig. 3.2). They were added to the multiple alignments of 15 samples received from the three original prison outbreaks described above (Fig. 3.1), and the resulting dendrograms are shown in Figure 3.3. It was evident that the 12 sequences from Liverpool were identical in both subgenomic regions to the common "prison variant". The deduced amino acid sequence of the surface gene fragment indicated that the prison variant (from now on referred to as HBV^{PV}) was of subtype *adw2*, genotype A (Norder *et al.*, 1994).

In the above investigations, where DNA sequences were generated, the analysis was carried out using the CLUSTAL V and neighbour-joining algorithms available in MegAlign. Further phylogenetic techniques were not required as theprimary aim was to search for identical sequences, rather than inferring evolutionary relationships between them. This information could have been extracted from the nucleotide alignments, although the dendrograms provided a convenient graphical representation, which was easy to interpret.

3.2.2 HBV^{PV} in the community: Warrington

The HBV^{PV} sequences were subsequently found in a number of samples derived from the community. In 1996 and 1997, 12 samples from patients acutely infected with hepatitis B were received. These patients were residents in and around the Warrington area. Molecular investigation was requested as the patients reported no risk factors, and their infections were apparently unrelated. Following DNA sequencing, multiple sequence alignments were created. The resulting dendrograms are shown in Figure 3.4.

Figure 3.3 Dendrograms from multiple alignments of HBV PCR product sequences amplified from inmates of four prisons in the north of England. DUR = Durham; HUL = Hull; LAN = Lancaster; LIV = Liverpool. Prisoners x, y and z from Liverpool correspond to lanes 6, 9 and 15 on the SSCP autoradiograph in Figure 3.1.

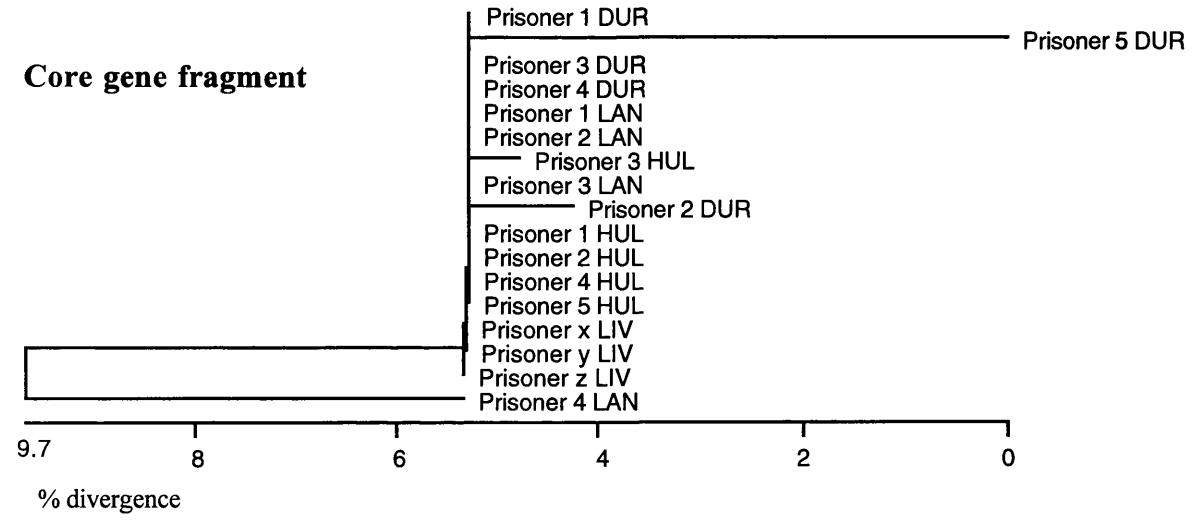
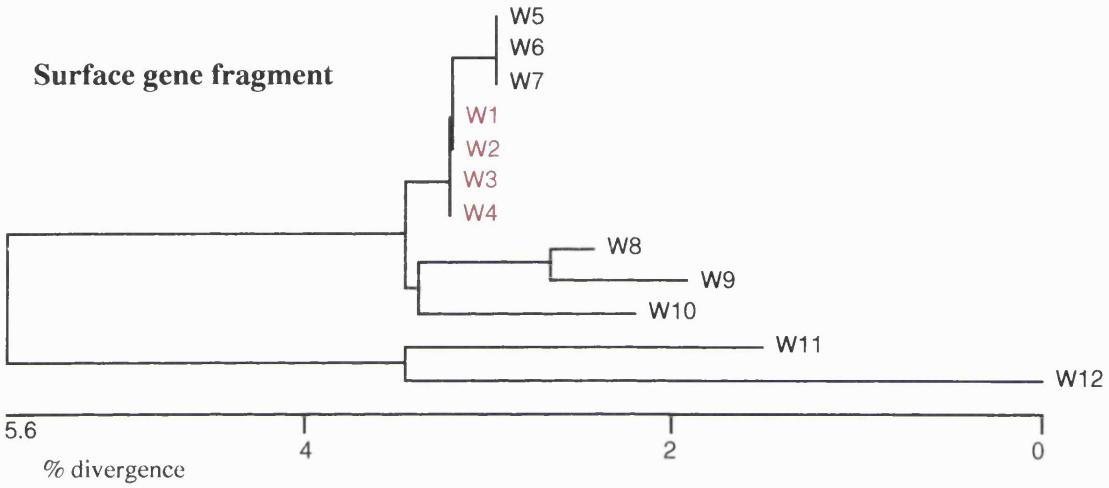
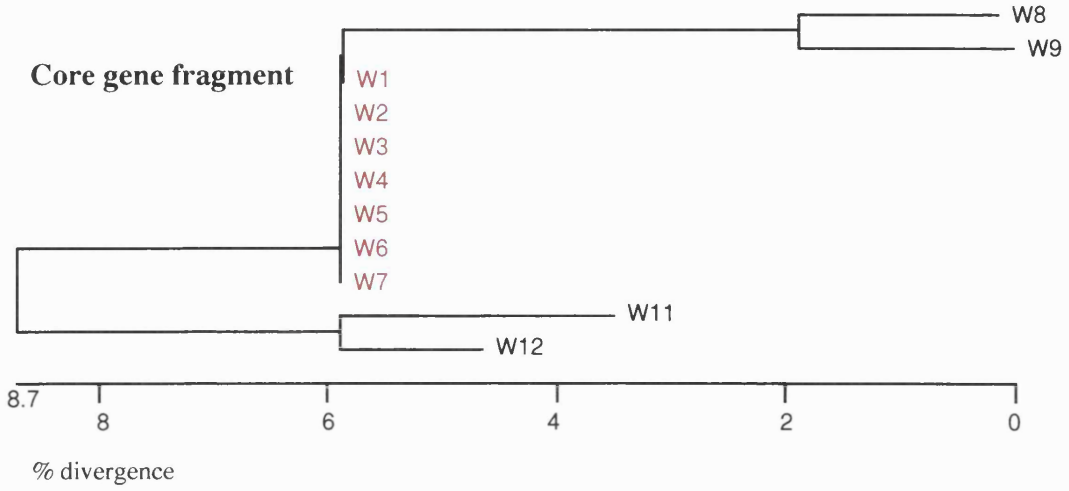


Figure 3.4 Dendrograms from multiple alignments of HBV PCR product sequences amplified from a cluster of acute infections in the Warrington area. Those matching the HBV^{PV} sequence are shown in red.





Seven of the 12 samples (W1 to W7) carried the HBV^{PV} core gene sequence. Four of these (W1 to W4) also shared the HBV^{PV} surface gene sequence, whilst the other 3 (W5 to W7) had the same single nucleotide substitution in this fragment. Thus W1 to W4 carried HBV^{PV}, while W5 to W7 carried a variant that was very closely related to HBV^{PV}. The remaining 5 samples contained different HBV sequences, 3 of genotype A and 2 of genotype D (one was negative in the core gene PCR).

3.2.3 HBV^{PV} in the community: Hull

The HBV^{PV} sequence was further identified in 1996 during the routine investigation of clusters of acute HBV infection in Hull, the location of one of the prisons. Risk factor information suggested that the infections were associated with attendance at a tattoo parlour (3 cases) and injecting drug use (2 cases). Again, these were isolated requests and were not thought at the time to be linked to any other outbreaks.

3.2.4 HBV^{PV} in the community: London

In 1998, a female patient had surgery in a London hospital and later developed acute hepatitis B. An investigation was initiated, which revealed that both she and a male patient at the same hospital shared the HBV^{PV} sequence in both surface and core PCR products. A transmission link between them could not be found, although they were treated in the same intensive care unit. The only HBsAg positive health care worker involved in the female patient's operation was carrying a different variant of genotype A. The possibility therefore existed that transmission could have occurred from the male patient to the female patient, perhaps due to inadequate implementation of cross-infection procedures.

3.2.5 Questions raised by the prevalence of HBV^{PV}

Although HBV is not a rapidly mutating virus, compared to RNA viruses, it was unexpected to find these 2 sequences (surface and core fragments, comprising a total of 601bp, or approximately one fifth of the genome) in such a large number of apparently unlinked infections. As the earliest identification of the variant was in prison outbreaks of HBV infection, it was hypothesised that the variant might have spread from prisons to the outside community, particularly in the north of

England. It is possible that in the early 1990s, the prevalence of the variant was low, but it was permitted to spread following its introduction into a prison. Frequent transfers of HBV^{PV} carrying inmates then led to the occurrence of outbreaks in multiple prisons. Hence the infected inmates were able, upon release, to transmit the virus to the outside community, for example to other IDUs or to sexual partners.

To substantiate this hypothesis, a study was initiated to look for HBV^{PV} in serum samples from community cases of acute HBV infection. Three questions were addressed:

1) What is the prevalence of HBV^{PV} in the general population?

2) Has its prevalence been increasing in recent years?

3) If so, could its spread in prisons in the north of England have contributed to the increase?

3.3 Tracking the spread of HBV^{PV}

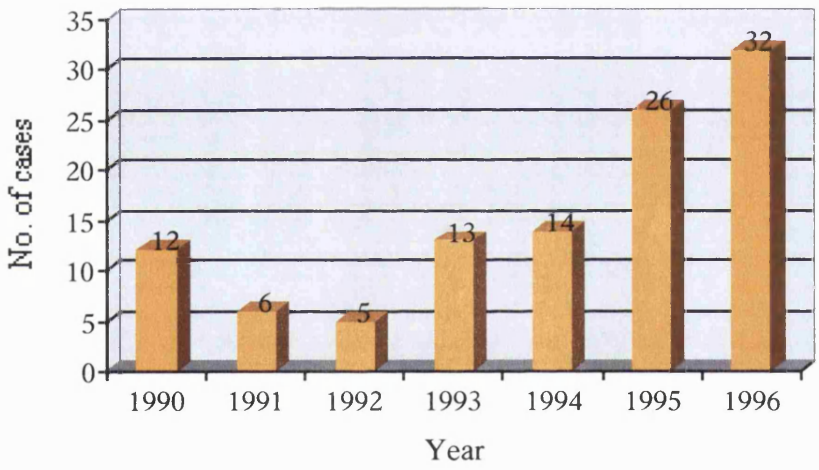
3.3.1 Choice of DGGE as the molecular screening method

In order to establish the prevalence of this variant in a large number of samples, a rapid and sensitive screening method was required. DNA sequencing was considered inappropriate as, despite automated improvements, it remains a lengthy and expensive procedure. Also, the amount of information provided by sequencing was not required in this study; it was sufficient to be able to distinguish HBV^{PV} sequences from non-HBV^{PV} sequences. Similarly, the radioisotopic SSCP technique described in Section 2.2 was considered unsuitable for large-scale screening. It too is a lengthy procedure, capable of analysing only 26 samples at a time on the available apparatus, as band resolution diminished if more samples were loaded onto the gel. Moreover, this technique suffers from the safety and disposal problems inherent in working with radioactive labels. The DGGE method was chosen, and developed to screen both surface and core gene PCR products for sequence identity.

3.3.2 Samples used for screening

The Hull region was selected as the test site, as this was the location of one of the prisons where HBV^{PV} was first identified, and where five community cases were subsequently detected. The Public Health Laboratory in Hull receives samples for HBV testing from an area with a population of approximately 460,000. These samples are referred from local GPs and hospitals. The catchment area also includes three prisons (R.E Meigh, pers. comm.). Serum samples from all laboratory confirmed acute HBV infections from 1990 to 1996 were screened. The total number of acute cases that occurred during that time was 108. The number of cases per year is shown in Figure 3.5.

Figure 3.5 Number of acute cases of HBV infection occurring in Hull per year



A risk factor for infection was recorded for 60/108 cases (55.6%). Eleven of 60 cases (18.3%) originated from countries of high endemicity, 16/60 cases (26.7%) were in prison at the time of diagnosis and 33/60 cases (55.0%) were IDUs. Other possible risk factors, such as sexual contact or surgery, were not made known.

As a comparison, two groups of samples were also chosen from those referred to CPHL and confirmed as having acute HBV infection. Thirty sera were selected at random from those received in 1990 and a further 30 from 1996. All samples

originated from hospitals and Public Health Laboratories located in the south east of England. Risk factor information was only available for 16/60 cases (26.7%).

3.3.3 Optimisation of DGGE conditions

The DGGE technique was developed from that described by Myers *et al.*, (1987) and Sheffield *et al.*, (1989). In order to optimise the gel and electrophoresis conditions, the first 18 samples from Hull (those from patients diagnosed in 1990 and 1991) were used to constitute an optimisation panel. DNA was extracted as described in Section 2.1.1 and GC-clamp PCRs were used to amplify the surface and core gene fragments as described in Section 2.4.1. These PCRs amplified DNA fragments corresponding to those in the 2 standard HBV PCRs (Section 2.1.2). However, a 40 bp GC-rich sequence was added to the 5' end of the inner sense primers, which resulted in its incorporation into the PCR products. The function of this high melting temperature domain was to prevent complete denaturation of the products on electrophoresis and thereby increase the sensitivity of the DGGE technique (Sheffield *et al.*, 1989).

One sample, obtained in 1990 (1990/7), was negative by PCR. This may have been due to target DNA loss from suboptimal storage of the serum. The remaining 17 samples were positive for both HBV surface and core gene amplification, and the PCR products were sequenced as described in Section 2.3. The nucleotide alignments are shown in Figure 3.6. Three samples (1990/3, 1991/1 and 1991/2) were found to be identical to HBV^{PV} in the surface region. These samples also matched HBV^{PV} in the core region. One further sample (1991/3) carried the HBV^{PV} core sequence, but showed 1 nt difference in the surface region. The remaining sequences contained up to 28/404 nt differences (median=7) from HBV^{PV} in the surface region, and up to 24/197 nt differences (median=13) in the core region.

Figure 3.6a Nucleotide alignment of surface gene PCR product sequences amplified from optimisation panel. PV Surface indicates the HBV^{PV} sequence for comparison.

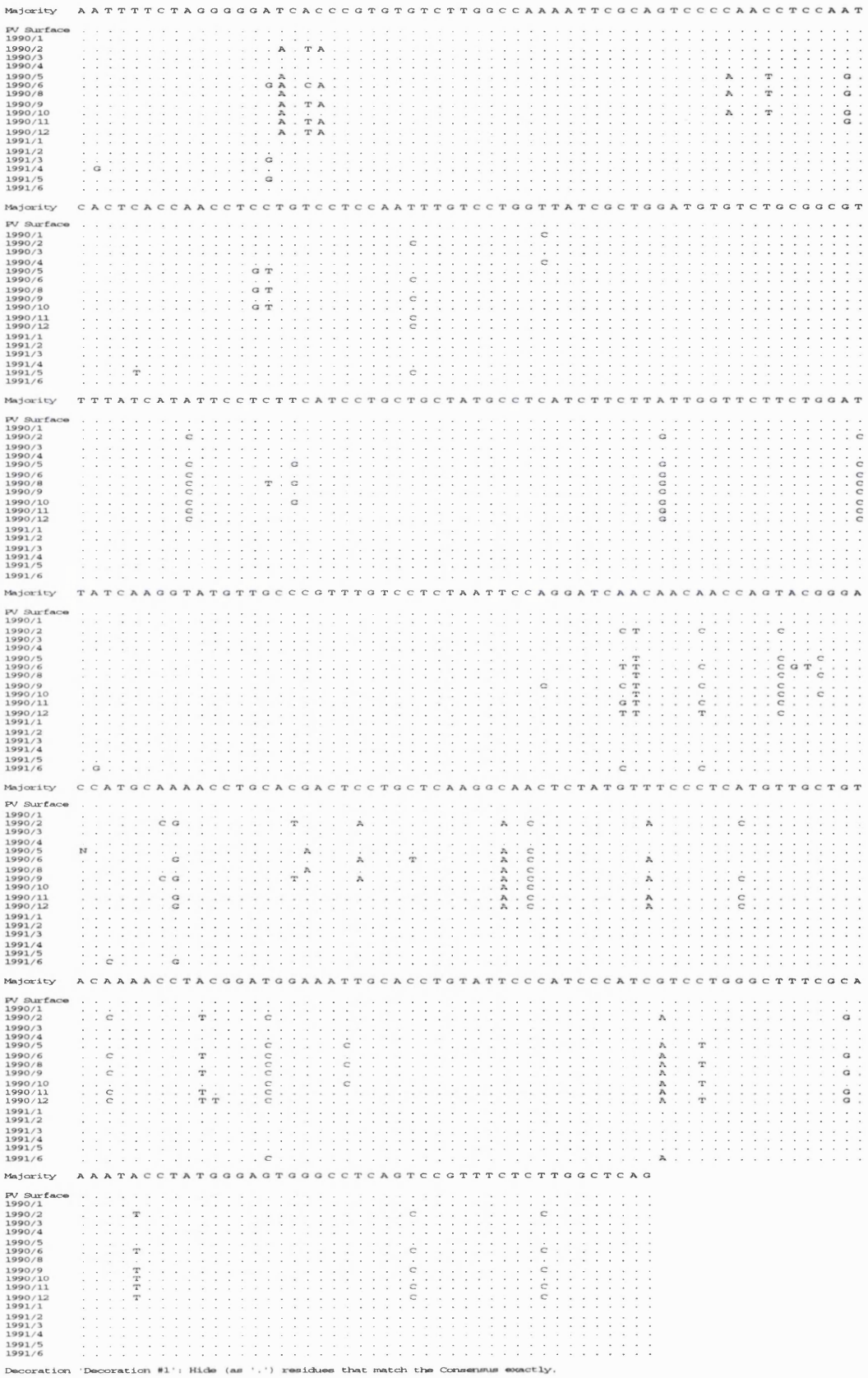
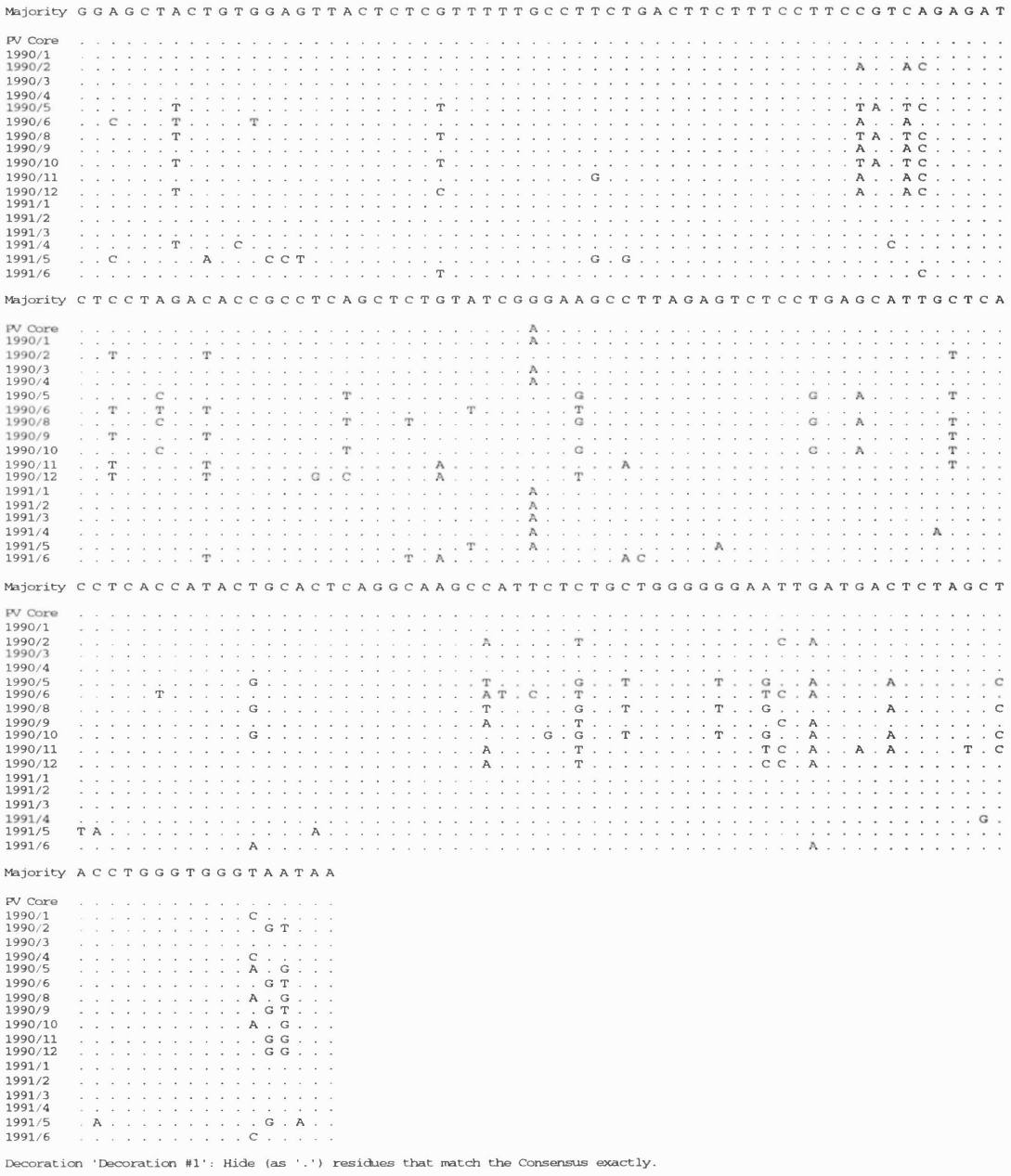


Figure 3.6b Nucleotide alignment of core gene PCR product sequences amplified from optimisation panel. PV Core indicates the HBV^{PV} sequence for comparison.



Using this panel of 17 samples, several DGGE runs were carried out. The concentration of acrylamide and the gradient of denaturants were varied in each, in order to optimise the discriminatory power of the technique. Different voltage and electrophoresis times were also examined. As migration of the product effectively ceases once all but the GC-clamp region has denatured, the electrophoresis time could be varied with no effect on the resulting gel. Electrophoresis was therefore performed overnight (18 h) at a low voltage (100 V) to make the procedure more streamlined. The final conditions are described in Section 2.4.

Figure 3.7 shows the gel pictures of the optimisation panel DGGEs, along with the dendrograms resulting from the sequence alignments. As Figure 3.7a shows, the surface region DGGE was not able to discriminate single nucleotide mutations, most likely due to the large size of the PCR product (Fodde and Losekoot, 1994). Thus, while samples 1990/1, 1990/4, 1991/3 and 1991/4 all showed a single nucleotide difference from HBV^{PV} in this region, and sample 1991/5 had 3 different nucleotides, they all migrated to the same position as those that exactly matched the HBV^{PV} surface sequence. These sequences are denoted by * in Figure 3.7a. The discrimination was not improved by lowering the acrylamide concentration or by altering the denaturant gradient. However, the core region DGGE did distinguish single nucleotide substitutions. Samples 1990/1 and 1990/4 both differed from HBV^{PV} by 1 nucleotide in this region, and both showed different band positions on DGGE (Figure 3.7b).

The core region of sample 1990/5 contained two bands on the gel picture, one of which migrated to the same position as the HBV^{PV} control, although the sequence of this sample was very different as shown on the dendrogram. It was thus decided that in order to be considered as carrying HBV^{PV} a sample must match the HBV^{PV} control band migration in *both* surface and core DGGEs. By employing this criterion the anomalous band in sample 1990/5 core region, where a divergent HBV sequence happened to melt at the same temperature as the HBV^{PV} control, would not be classified as HBV^{PV}, as the surface region band clearly did not match HBV^{PV} (Fig. 3.7a). Sequences that matched in only one region were therefore considered to originate from different variants.

Figure 3.7a DGGE gel picture (upper panel) and dendrogram (lower panel) of optimisation panel surface gene PCR products.

Upper panel: Lanes marked with * are those with identical band positions, which match or are similar to the HBV^{PV} surface sequence.

Lower panel: * denotes the corresponding sequences on the dendrogram.

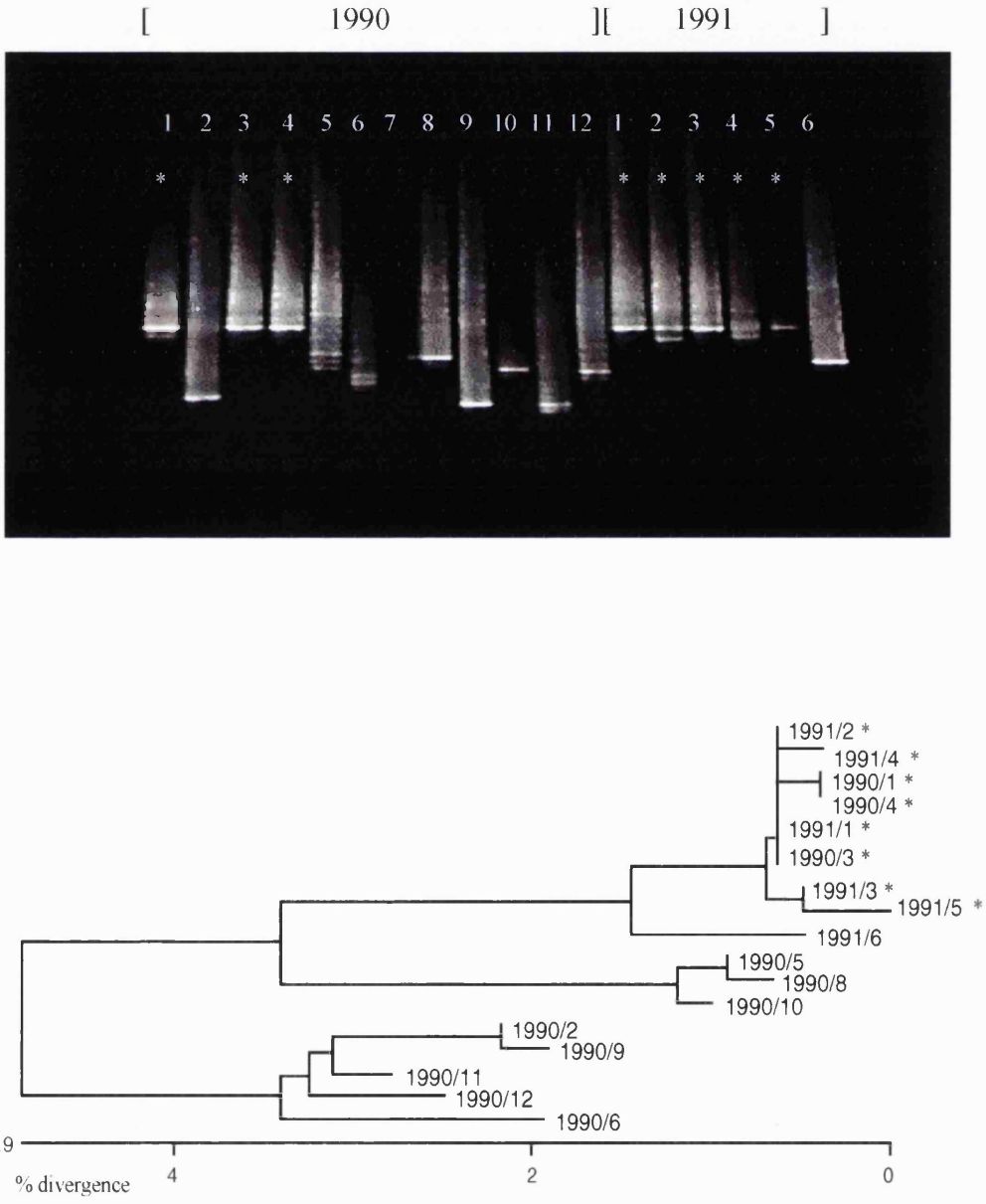
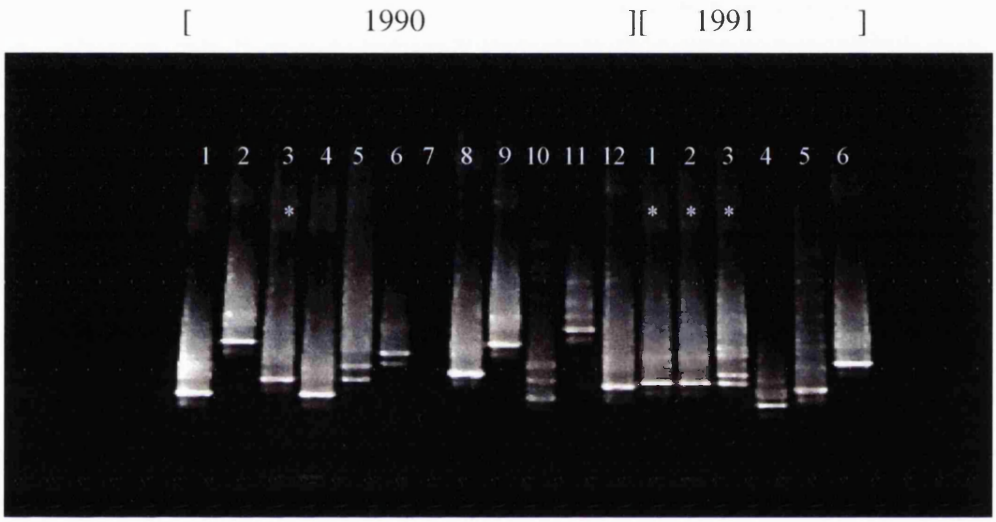
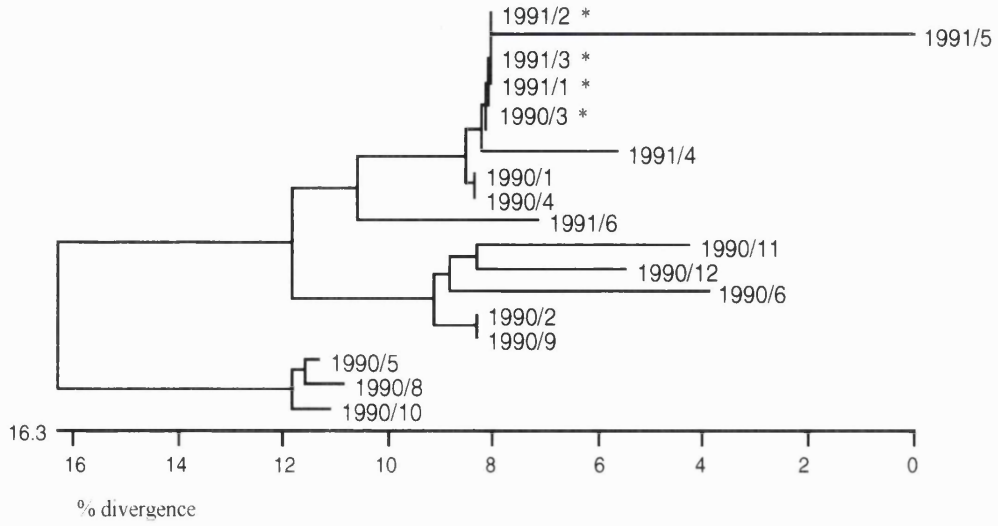


Figure 3.7b DGGE gel picture (upper panel) and dendrogram (lower panel) of optimisation panel core gene PCR products.

Upper panel: Lanes marked with * are those with identical band positions, which match the HBV^{PV} core sequence.

Lower panel: * denotes the corresponding sequences on the dendrogram.





Further analysis of the melting temperatures (T_m) of surface and core PCR products with theoretical point mutations was carried out using the MELT87 program (Lerman and Silverstein, 1987). It was confirmed that the calculated T_m of these hypothetical PCR products were sufficiently different to be distinguished from the HBV^{PV} PCR products if there was a single substitution in the core region or more than 3 substitutions in the surface region. Furthermore, discrimination was not dependent upon the position of the substitution within the PCR product.

Surface and core gene secondary PCR products that had been shown to match HBV^{PV} on DNA sequencing were cloned using the TOPO TA Cloning Kit in order to make a stock of control HBV^{PV} GC-clamped PCR products. Three colony PCRs were performed on each, using the secondary HBV PCR conditions (Section 2.1.2) and the colony PCR products were sequenced to confirm their identity to HBV^{PV}. Plasmid DNA was purified using a QIAfilter Plasmid Midi Kit and this was used as template material to make the surface and core controls using the secondary GC-clamp PCR described in Section 2.4.1. Three controls were run on each DGGE for comparison.

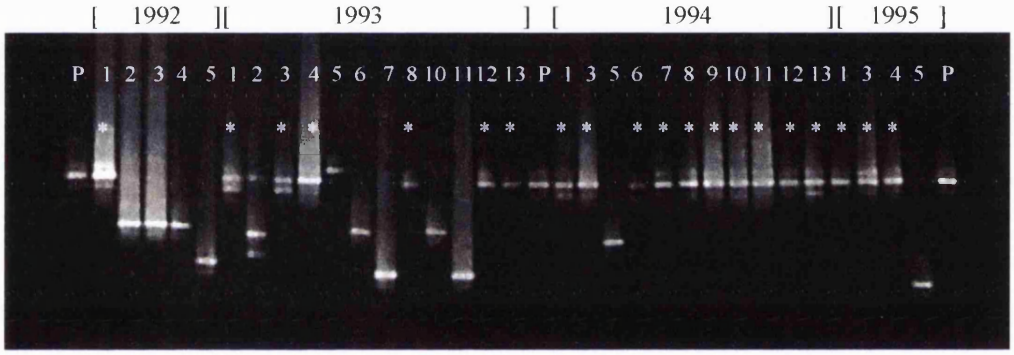
The DNA from all remaining Hull serum samples (n=90) and south east England serum samples (n=60) was extracted as described in Section 2.1.1 and GC-clamp PCRs used to amplify the surface and core gene fragments as described in Section 2.4.1. Thirteen samples from Hull were negative in one of the two PCRs and were excluded from further analyses. Eight DGGE gels were run (two at a time) with 28-35 samples and three HBV^{PV} controls loaded on each (Sections 2.4.2 - 2.4.4).

3.3.4 DGGE results

The gel pictures showing DGGE analysis of the Hull samples surface and core PCR products are shown in Figure 3.8. The comparison samples from the south east of England are shown in Figure 3.9. Those that migrated to the same position as the HBV^{PV} control sequences are denoted by *.

Figure 3.8 DGGE results of PCR products from Hull samples (1992-1995). Lanes marked with P contain control HBV^{PV} sequences for comparison. * denotes lanes where the band migration matches that of the HBV^{PV} control. When this occurs in both gels (surface and core) the corresponding sample is considered to be carrying HBV^{PV}.

Surface fragment



Core fragment

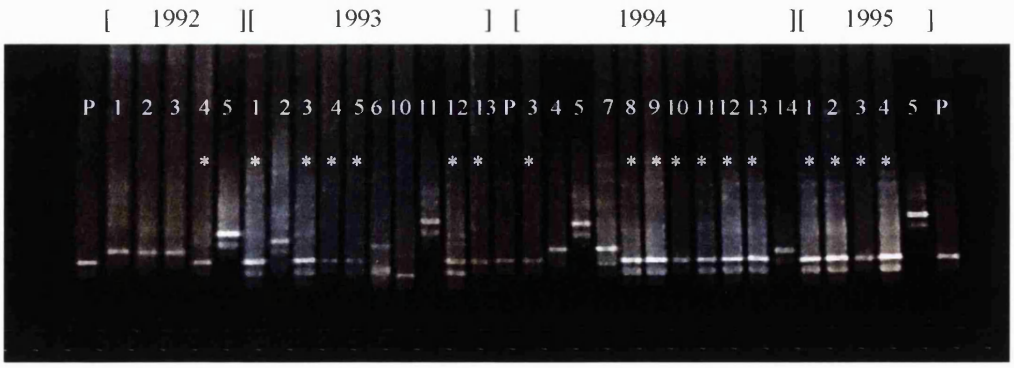
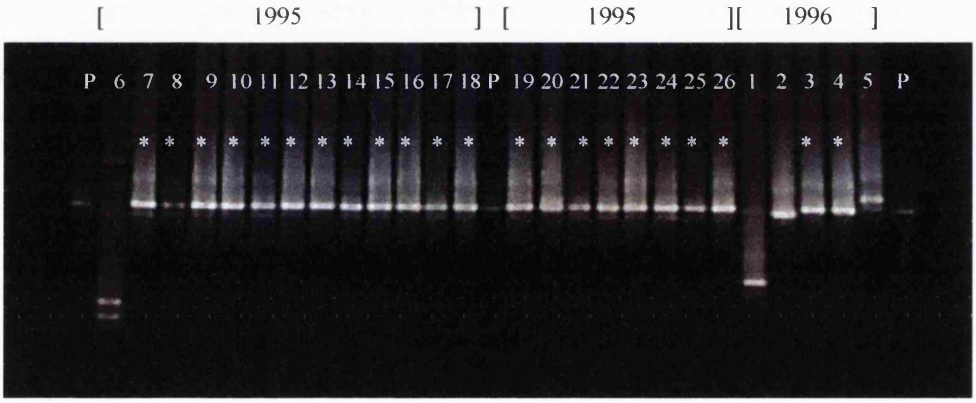


Figure 3.8 (cont.) DGGE results of PCR products from Hull samples (1995-1996).

Surface fragment



Core fragment

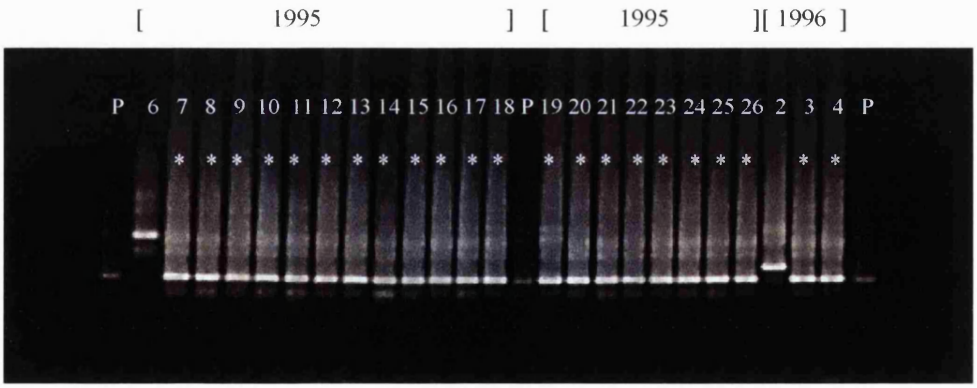
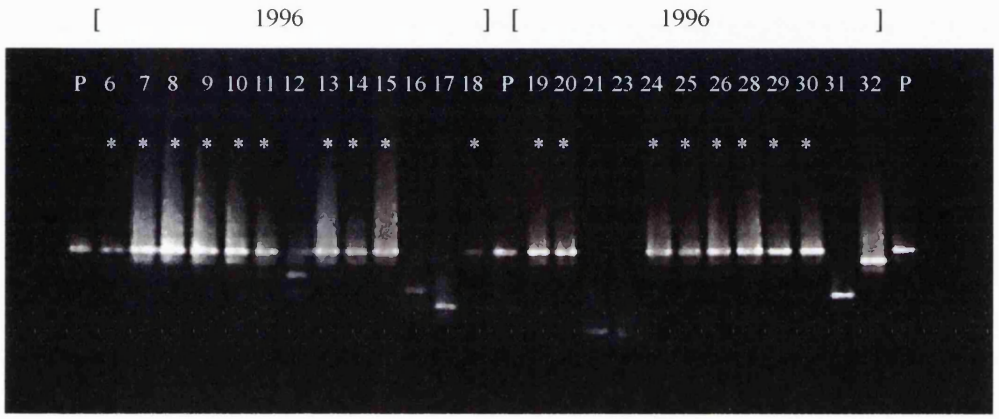


Figure 3.8 (cont.) DGGE results of PCR products from Hull samples (1996).

Surface fragment



Core fragment

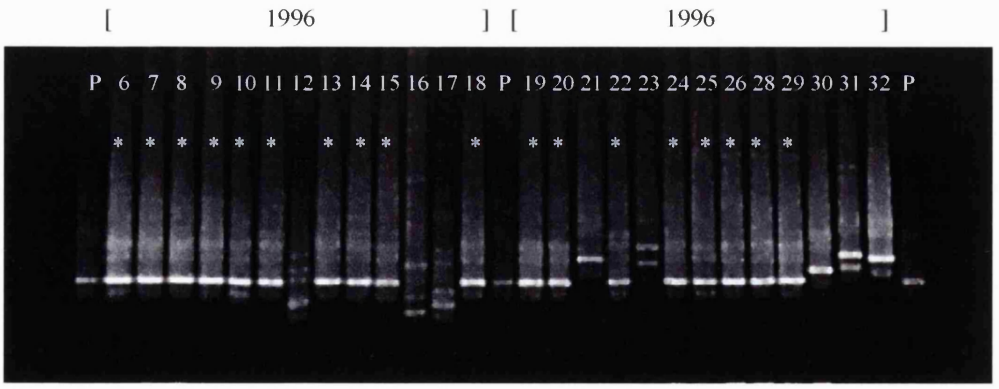
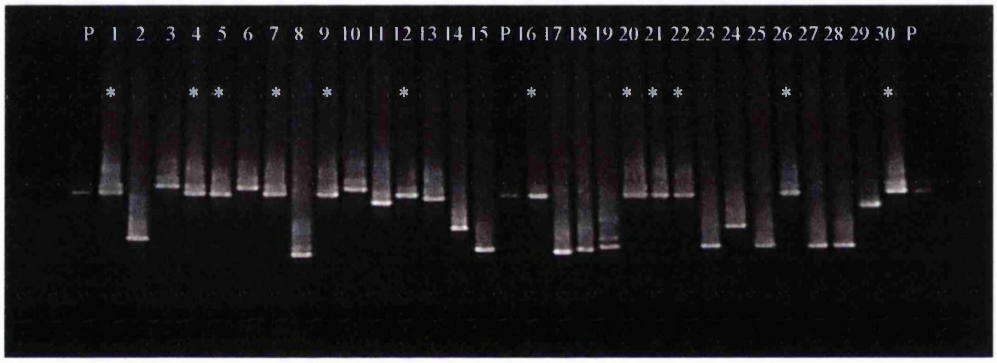


Figure 3.9a DGGE results of PCR products from south east England samples (1990). Lanes marked with P contain control HBV^{PV} sequences for comparison. * denotes lanes where the band migration matches that of the HBV^{PV} control. When this occurs in both gels (surface and core) the corresponding sample is considered to be carrying HBV^{PV}.

Surface fragment



Core fragment

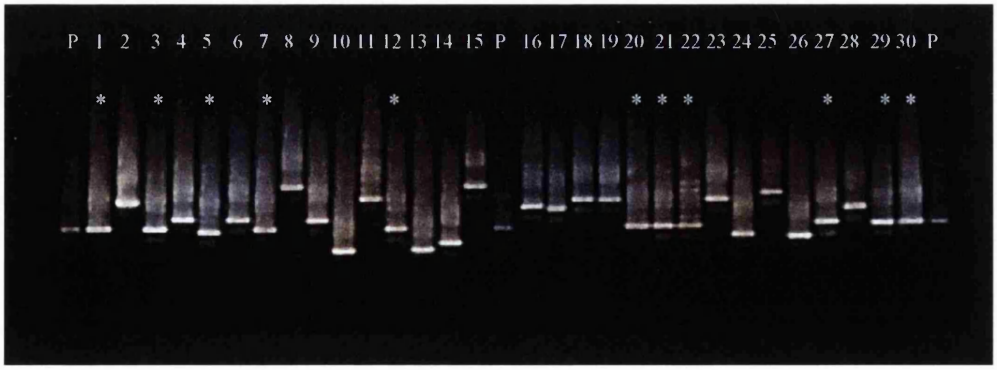
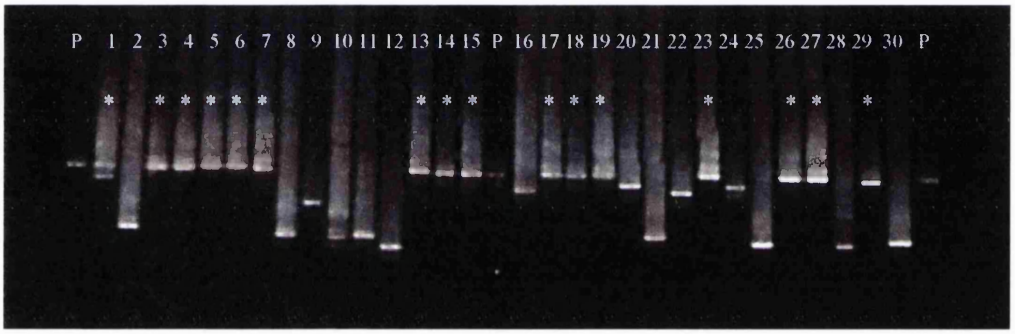
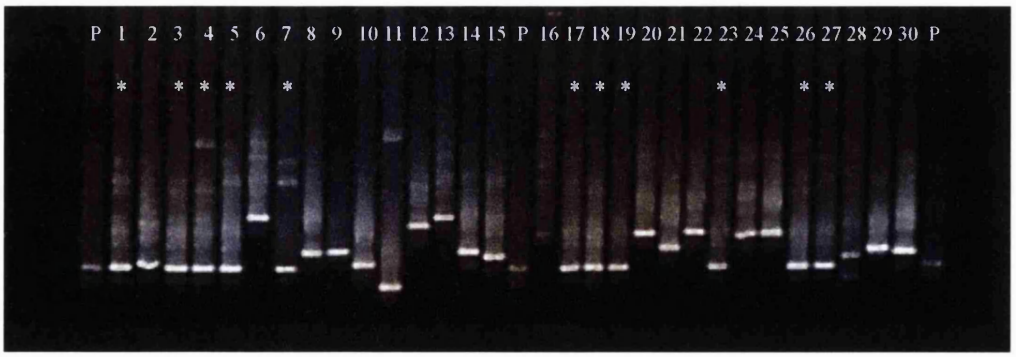


Figure 3.9b DGGE results of PCR products from south east England samples (1996). Lanes marked with P contain control HBV^{PV} sequences for comparison. * denotes lanes where the band migration matches that of the HBV^{PV} control. When this occurs in both gels (surface and core) the corresponding sample is considered to be carrying HBV^{PV}.

Surface fragment



Core fragment



Some samples produced more than one band following DGGE (eg., Fig. 3.7b, sample 1990/5, Fig. 3.8, samples 1993/2 and 1995/6 surface fragment). This may have been due to the presence of more than one sequence in the PCR products, resulting in the formation of homoduplex and heteroduplex bands. The amplification of more than one sequence can occur if HBV exists as a heterogeneous quasispecies population in the original serum specimens. However, this is likely to occur rarely as specimens from acutely infected individuals usually contain a homogeneous HBV population at this stage (S. Whalley, pers. comm.). The sequence chromatogram data from sample 1990/5 core region did not contain double peaks, which are characteristic of a mixed population of PCR product sequences.

Another potential mechanism for the generation of heteroduplexes is the misincorporation of one or more nucleotides during PCR by *Taq* polymerase. The probability that this would occur was reduced in this screening process by using the EXPAND PCR System. This contains a mix of two polymerase enzymes, *Taq* polymerase and *Pwo* polymerase. *Pwo* polymerase has 3'-5' exonuclease proof-reading activity, which enables the majority of *Taq* polymerase-induced errors to be corrected. Consequently, the EXPAND PCR system has a fidelity three times higher than that of *Taq* polymerase (8.5×10^{-6} errors per base pair per PCR cycle). It has been estimated that, for a 30 cycle PCR amplifying a 200 bp product, the proportion of PCR products containing a misincorporated nucleotide would be 5% using *Taq* polymerase alone, or 2% using EXPAND. Therefore, even when a secondary PCR is performed, the proportion of products containing a mutation would be insufficient to be visualised as separate homoduplex and heteroduplex bands following DGGE.

A final possible cause of the appearance of multiple bands is the presence of excess primer dimers, or, more likely, of non-specific PCR products. Although these fragments were invisible on standard agarose gel electrophoresis, they may have been visualised more readily using this more sensitive polyacrylamide gel system followed by SYBR Green staining. This would seem to be the most likely cause as, in later gels, less PCR product was loaded and the appearance of double bands diminished (Figure 3.9).

Initial examination of the DGGE pictures yielded the observation that HBV^{PV} was the dominant HBV type in both the Hull and comparison (SE England) populations. The large increase in cases of acute HBV in Hull in 1995 and 1996 was predominantly due to this variant. In south east England, however, the HBV^{PV} prevalence appeared to be constant, constituting about one third of the samples studied at both time points (1990 and 1996).

Table 3.1 summarises the data obtained from DGGE analysis of both sets of samples. A chi square test for linear trend was carried out on the Hull data, which showed a significant increase in the proportion of samples with HBV^{PV} over the seven years studied ($p<0.01$). It can be seen that a peak of acute HBV infection caused by HBV^{PV} occurred in 1995, when 23 of the 25 PCR positive samples (92%) matched the HBV^{PV} control band in both surface and core fragment DGGEs.

Table 3.1 Summary of DGGE results from both patient groups.

Samples from Hull

* S = surface, C = core

Year	No. of samples tested	No. of samples positive in both S and C* PCR	No. of samples matching HBV^{PV} in both S and C DGGEs
1990	12	11	1
1991	6	6	3
1992	5	5	0
1993	13	10	5
1994	14	9	7
1995	26	25	23
1996	32	28	19

Samples from SE England

1990	30	30	8
1996	30	30	11

When the proportion of samples with HBV^{PV} in the samples from south east England was compared, no difference between 1990 and 1996 was found (p=0.405). Comparisons were then made between the Hull samples received in 1990 and the SE England samples received in the same year. Again, no difference in the proportion with HBV^{PV} was found (p=0.228). However, when a chi square test was carried out for the 1996 samples from both populations, the difference in HBV^{PV} proportion was significant (p=0.0175). These analyses suggest that whilst HBV^{PV} had become the dominant variant in the Hull population, it had remained at a constant level in the south east of England. It is noted, however, that the number of samples received from Hull, particularly in the early 1990s, was small and no south east England samples were tested from the years between 1990 and 1996. The results of statistical analyses must therefore be interpreted with caution.

Table 3.2 shows the Hull data analysed according to risk groups. This breakdown was not carried out for the south east England comparison samples, as risk factor information was recorded in only a minority of cases.

Table 3.2 Summary of DGGE results from Hull patient group, analysed according to risk groups.

Risk category	No. of samples positive in both S and C PCR	No. of samples matching the PV in both S and C DGGEs	% of samples with PV
Ethnic	8	0	0.0%
Prison	15	12	80.0%
IDU	29	26	89.7%
Unknown	42	20	47.6%
Total	**94**	**58**	**61.7%**

Several studies have shown that the major risk factor for the transmission of bloodborne viruses within prisons is the sharing of needles and other equipment during injecting drug use. Approximately 60% of admissions to one of the prisons in the Hull area have some connection with drug use (R.E.Meigh, pers.

comm.). Sexual activity is thought to be responsible for a comparatively small number of infections acquired by inmates (Bird *et al.*, 1993; Curtis and Edwards, 1995; Weild *et al.*, 2000). The data shown in Table 3.2 indicate that infection with HBV^{PV} is strongly associated with the prison/IDU risk group in this population.

3.4 The emergence and spread of the HBV autohaemotherapy variant (HBV^{AV})

3.4.1 HBV outbreak in an alternative therapy clinic

A large outbreak of hepatitis B occurred in 1998, which was associated with attendance at an alternative therapy clinic. The first report was received from a Consultant in Communicable Disease Control in the Midlands who had identified a symptomatic case of acute hepatitis B infection in a 43 year old woman. Her only reported risk factor was attendance at an alternative therapy clinic in London. Case finding procedures identified a further nine serologically confirmed acute hepatitis B infections associated with the clinic, including two of the clinic workers. All had received a treatment known as autohaemotherapy, which involved the drawing of 1-2 ml of the patient's blood, mixing of the blood with an equal volume of saline in the same syringe and then injection of the mixture into the patient's buttocks or acupuncture points.

This mode of treatment was strongly implicated as being responsible for HBV transmission. During the outbreak investigation it was found that, although new needles and syringes were used for each patient, saline was drawn from a multi-use bottle directly after venepuncture, when the syringe also contained the patient's blood. The bottle was topped up from a larger container of saline and was not made available for analysis. It was not possible to identify the source patient, as most patients visited the clinic more than once before and during the outbreak.

A lookback investigation was launched and, including the original 10 cases described above, a total of 399 patients were identified as having been treated at the clinic between January 1997 and February 1998. Of these, 352 provided

serum samples for testing (216 females; mean age 42 years, range 1-86 years). Serological assays revealed that 299 patients had no markers of HBV infection, 24 showed evidence of resolved HBV infection and 33 were HBsAg positive. All serum samples were tested by PCR as described in Section 2.1. HBV DNA was amplified from 32 of the 33 HBsAg positive samples, and from 1 secondary infection case, identified during contact tracing. The results of serological testing on these samples are shown in Table 3.3.

Table 3.3 Summary of serological results on 33 HBV DNA positive samples from 32 clinic attendees and 1 secondary case.

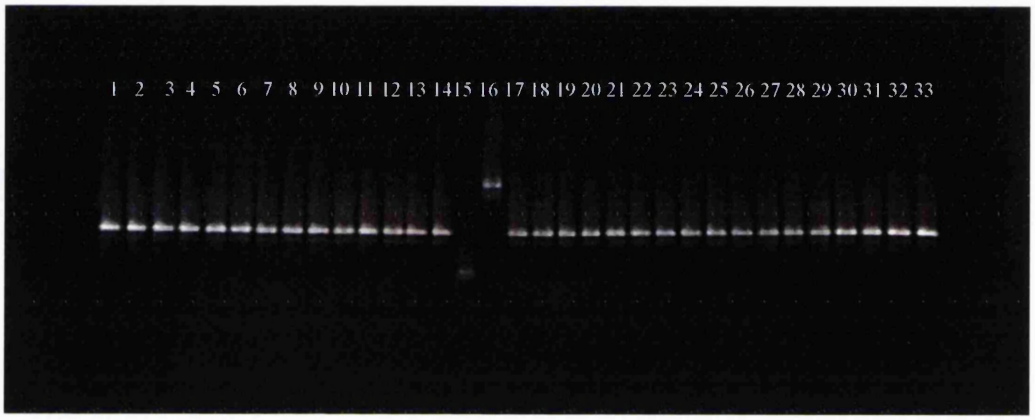
Serological testing results	Stage of infection	Number of samples
HBsAg+, anti-HBc IgM-, HBV DNA+	Incubation	9
HBsAg+, anti-HBc IgM+, HBV DNA+	Acute	20
HBsAg+, anti-HBc IgM-, total anti-HBc+, HBeAg+, HBV DNA+	Resolving/Carrier	2
HBsAg+, anti-HBc IgM-, total anti-HBc+, anti-HBe+, HBV DNA+	Resolving/Carrier	2

3.4.2 Identification and tracking the spread of HBV^{AV}

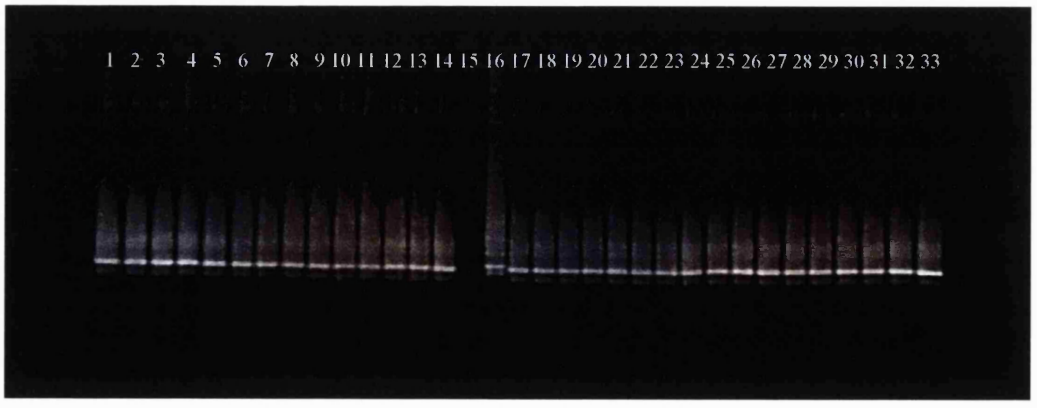
Using the methods developed for HBV^{PV} screening, described above and in Section 2.4, DGGE was applied to the 32 HBV DNA positive samples from clinic attendees (which included two clinic workers) and the single secondary infection case. As before, both surface and core gene fragments were analysed. The gel pictures, shown in Figure 3.10, clearly demonstrate that in 30 of the 32 clinic attendees identical banding patterns occurred in both genomic regions. The sample from the secondary case, who was a sexual partner of one of the clinic attendees, also matched in both regions (lane 33). This strongly suggested that these individuals had been infected by the same HBV variant, from now on referred to as HBV^{AV}. The remaining two samples (lanes 15 and 16) had different band positions on the surface fragment DGGE. One of these also showed a different band position on the core fragment DGGE; the other was negative in this PCR.

Figure 3.10 DGGE results of PCR products from samples associated with the autohaemotherapy outbreak. Lanes 1 and 2 contain PCR products from clinic workers, 3-32 are from clinic attendees and lane 33 is from a secondary infection case.

Surface fragment



Core fragment



These two different samples were derived from patients with a resolving or chronic carrier serological profile. Both patients originated from geographical regions of medium endemicity for HBV infection.

To confirm that the bands migrating to identical positions on the DGGE gels were derived from PCR products with identical or very similar sequences, DNA sequencing was carried out on the surface and core fragments from 3 representative samples (patients 5, 22 and 30) with matching DGGE bands, and from the two non-matching samples. The resulting nucleotide alignments are shown in Figure 3.11, which demonstrate that the 3 samples with matching band patterns on DGGE had identical sequences in both regions. The deduced amino acid sequence of the surface gene fragment indicated that HBV^{AV} was of subtype *ayw3*, genotype D (Norder *et al.*, 1994). Patient 15 was also infected with a genotype D variant and patient 16 carried a genotype A variant; these were 2.3% and 7.6% divergent from HBV^{AV} respectively, in the surface gene region.

DGGE was also employed to track HBV^{AV} in cases of acute HBV infection (n=20) that were received in CPHL during the same time period (January to June, 1998), but were not epidemiologically linked to the alternative therapy clinic. These randomly selected samples did not originate from any specific geographical location. Three HBV^{AV} samples and two HBV^{PV} samples were also included for comparison. PCR and DGGE conditions were as described in Section 2.4, and the gel pictures are shown in Figure 3.12. On both gels, none of the randomly selected samples migrated to the same position as HBV^{AV}, suggesting that it was not a common variant circulating in the community at that time. In contrast, four new examples of HBV^{PV} were identified in patients with no epidemiological link (patients c, e, n and t). Furthermore, a cluster of four IDUs infected by a common variant was revealed, in that they had identical band positions on both surface and core DGGEs (patients g, h, i and j). These IDUs were residents of a town in the Midlands.

Figure 3.11a Nucleotide alignment of the surface gene PCR product sequences amplified from five clinic attendees. Patient numbers correspond to the lanes on the DGGE in Figure 3.9.

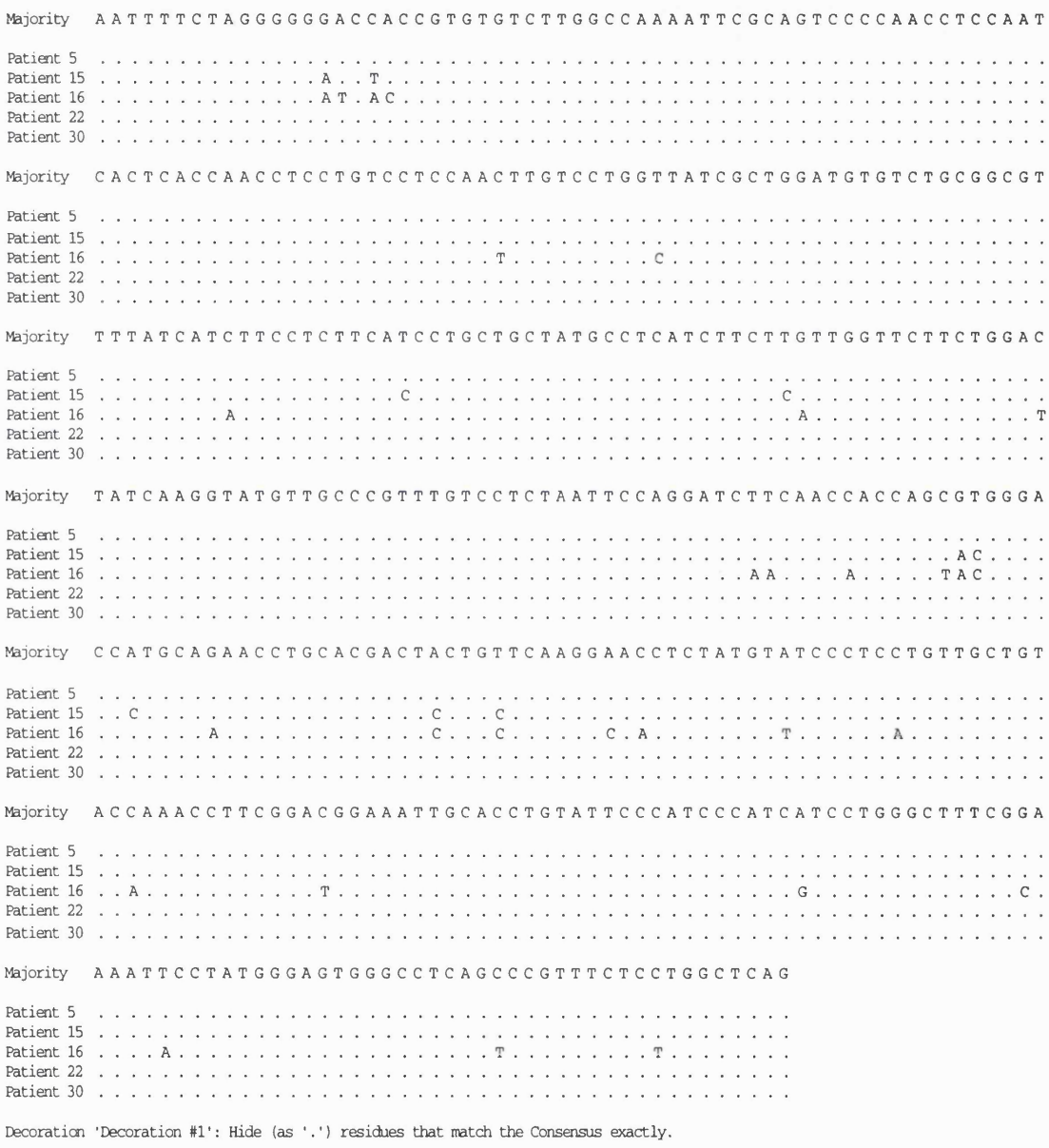
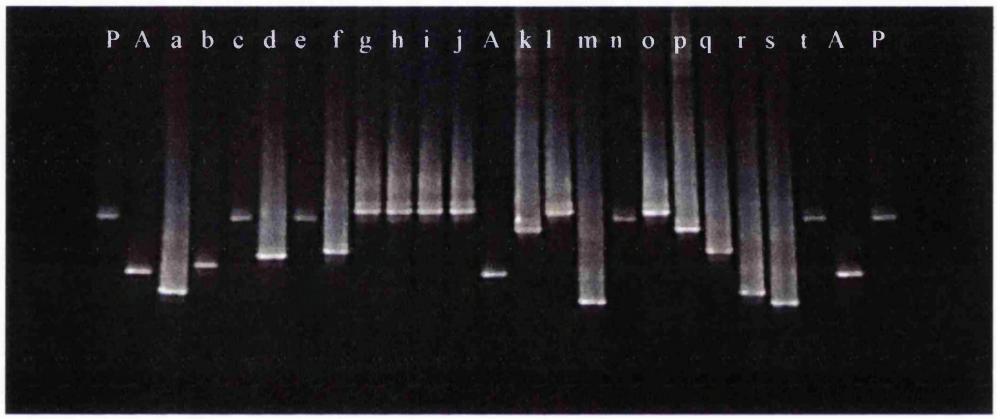


Figure 3.11b Nucleotide alignment of the core gene PCR product sequences amplified from five clinic attendees. Patient numbers correspond to the lanes on the DGGE in Figure 3.9.

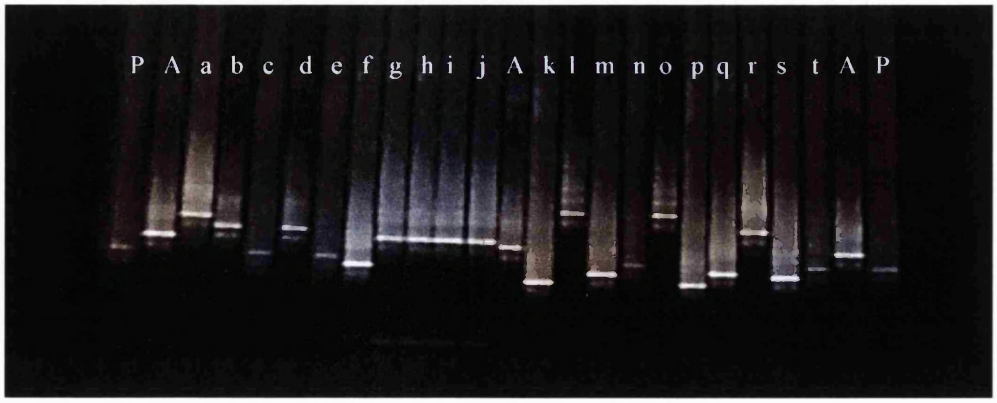
```
Majority   GGAGCTACTGTGGAGTTACTCTCGTTTTTGCCTTCTGACTTCTTTCCTTCAGTACGAGAT

Patient 5  ............................................................
Patient 16 ..........................T.........A..................C..CA.....
Patient 22 ............................................................
Patient 30 ............................................................

Majority   CTTCTTGATACCGCCTCAGCTCTGTATCGGGAAGCCTTAGAGTCTCCTGAGCATTGTTCA

Patient 5  ............................................................
Patient 16 ..CA.A..C................T...A............................C...
Patient 22 ............................................................
Patient 30 ............................................................

Majority   CCTCATCATACTGCACTCAGGCAAGCCATCCTTTGCTGGGGGGAGCTAATGACTCTAGCT

Patient 5  ............................................................
Patient 16 .....C.......................T..C...........AT.G............
Patient 22 ............................................................
Patient 30 ............................................................

Majority   ACCTGGGTGGGTGTTAA

Patient 5  .................
Patient 16 ............AA...
Patient 22 .................
Patient 30 .................

Decoration 'Decoration #1': Hide (as '.') residues that match the Consensus exactly.
```

Figure 3.12 DGGE results of PCR products from patients with acute HBV infection, not associated with the autohaemotherapy outbreak (lanes a-t). Lanes marked with P at either end contain control HBV^{PV} sequences. Lanes marked with A are samples linked to the autohaemotherapy outbreak (HBV^{AV}), loaded onto the gel for comparison.

Surface fragment



Core fragment



To further confirm the DGGE results, the surface region PCR products from the 20 randomly selected samples (a-t) were sequenced. These were added to the alignment of HBV surface sequences derived from autohaemotherapy patients 5, 15, 16, 22 and 30, shown in Figure 3.11a. The resulting dendrogram is shown in Figure 3.13. This correlates exactly with the DGGE in Figure 3.12. All samples with the same band migration on DGGE have identical or very closely related HBV surface sequences, and none of the randomly selected samples match the HBV^{AV} sequence (as represented by patients 5, 22 and 30).

3.5 Discussion

Studies of HBV transmission have conventionally been conducted on the basis of gathering epidemiological information, serological testing and antigenic subtyping (Stevens *et al.*, 1975; Lok *et al.*, 1987; Magnius and Norder, 1995; Tobe *et al.*, 2000). These investigations rarely allow definitive links to be made due to the lack of discrimination between viral strains or variants. More recently it has been common practice to use sequencing or other gel-based analysis methods to establish the source of a cluster of infections and to track chains of transmission. For example, using sequence analysis, an outbreak of hepatitis B in German cardiac transplant patients was traced to contaminated biopsy instruments or medication vials (Petzold *et al.*, 1999). Sequencing was also used to demonstrate a link between patients attending an electroencephalogram clinic in Toronto, Canada (Hepatitis B Outbreak Investigation team, 2000) and to support the occurrence of intra-familial transmission in a study of immigrants to Australia (McIntosh *et al.*, 1998). A SSCP assay was successfully applied to investigate multiple episodes of cross-infection in an oncology unit in South Africa (Hardie *et al.*, 1996). Some groups have used whole genome sequencing to increase the discrimination between variants in an outbreak (Petzold *et al.*, 1999; Petrosillo *et al.*, 2000). Where the same or very similar subgenomic sequences are identified, it is generally concluded that recent transmissions have occurred between infected individuals.

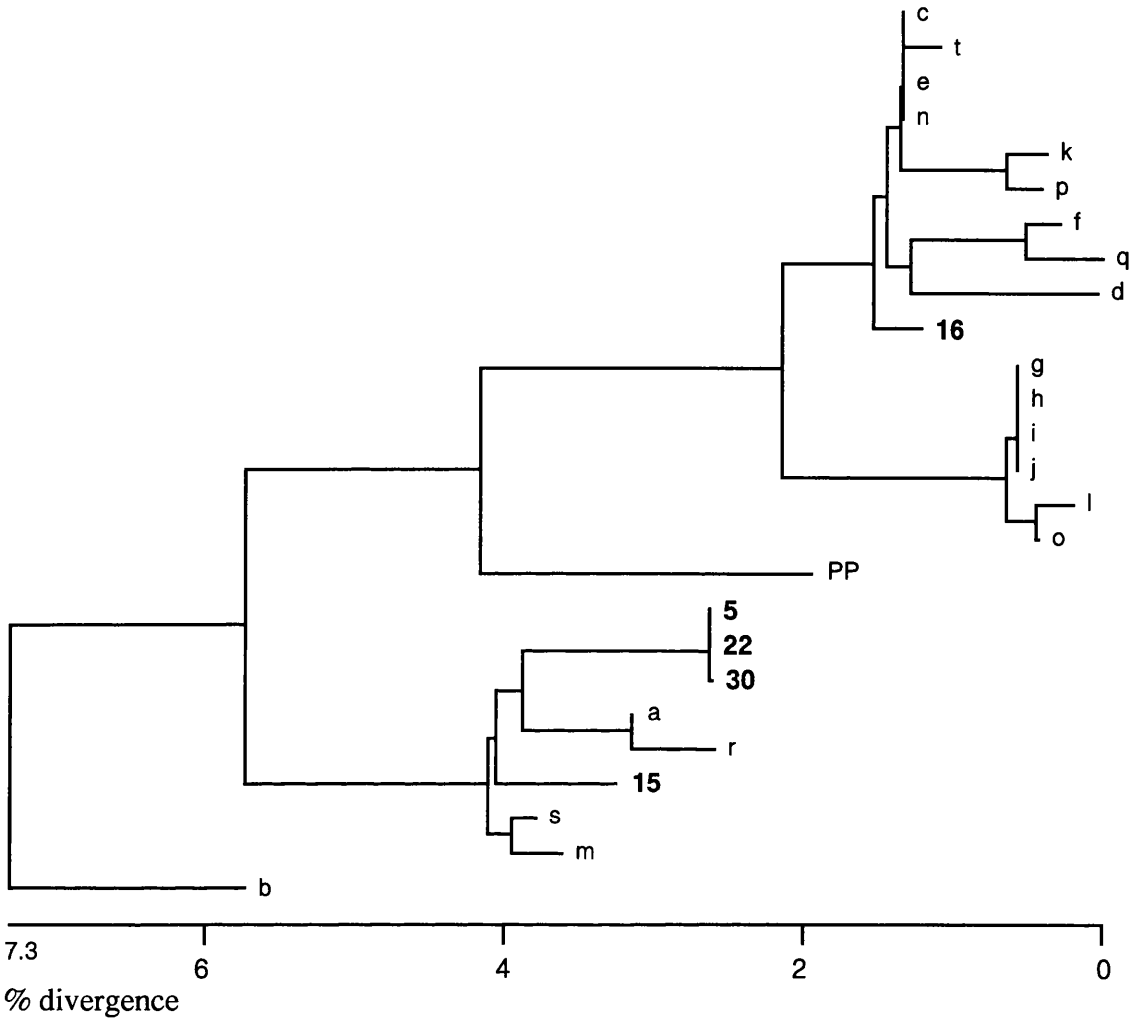


Figure 3.13 Dendrogram of HBV surface PCR product sequences. Samples 5, 15, 16, 22 and 30 (shown in green) correspond to those shown on the DGGE gel pictures in Fig. 3.9. These are from acutely infected patients who had autohaemotherapy. Samples a to t correspond to those shown on the DGGE gel pictures in Fig. 3.11. These are from acutely infected patients unrelated to the autohaemotherapy outbreak. PP designates the sequence of the plasmid control used routinely in PCR studies.

A high-throughput DGGE screening method was developed to analyse serum samples collected from cases of acute HBV infection in the north of England over a period of 7 years. This enabled the increasing prevalence of HBV^{PV}, predominantly in IDUs and prison inmates, to be quickly assessed and compared to a population group in the south east.

Two different regions of the genome were chosen for DGGE analysis. The surface gene fragment codes for amino acids 40 to 174 of HBsAg, which encompasses the major hydrophilic region (amino acids 99 to 169). It is in this domain that the mutations associated with vaccine escape and reduced detection in serological assays are localised. The HBV^{PV} surface sequence was examined for reported mutations, but appeared to be a wild type. This fragment is also relatively conserved. It completely overlaps the polymerase gene and its evolution is thus constrained by both surface protein structural requirements and by the necessity to encode a functioning polymerase enzyme. The nucleotide sequence of the core gene fragment is not similarly restricted, which is reflected in its greater diversity. Serological testing had indicated that several of the inmates infected with HBV^{PV} were HBeAg positive so the presence of precore mutations, which abrogate the production of HBeAg, were not further investigated (Carman *et al.*, 1989).

In this study, the presence of identical HBV sequences in inmates from four different prisons in the north of England between 1993 and 1995 indicated that there was a transmission link between them. There is considerable movement of inmates between prisons on a countrywide scale, so it is likely that one or more inmates became acutely infected in one prison and initiated clusters of infection in another, following a transfer. This spread would have been aided by the length of time taken for an individual to clear HBV infection; the 95% of adults infected with HBV who do not become chronic carriers can remain infectious for up to six months (Evans and London, 1998).

In one of the prisons (Hull) the cluster of acute HBV infections occurred during 1994, when five inmates were found to be infected with the HBV^{PV} sequence (Section 3.2.1, Figures 3.1 and 3.3). This was the year before a large increase in

the number of acute HBV infections in the local Hull community was seen, caused predominantly by HBV^{PV} (Figure 3.8). Thirteen of these were diagnosed in IDUs and a further four in prison inmates. Both this study and previous studies of risk behaviours in inmates suggest that the rapid increase of HBV^{PV} infection was brought about by the release of one or more infected inmates into the Hull region, when outbreaks in local drug users ensued. It has also been shown that many inmates in the UK change sexual partners at a high rate, in between prison stays, and so the potential for the sexual transmission of HBV in the outside community is also great (Bird *et al.*, 1993; Curtis and Edwards, 1995; Weild *et al.*, 2000). The spread of HBV^{PV} in Hull was therefore most likely associated with its prior dissemination in a local prison. The findings highlight the public health importance of implementing preventative measures to curb HBV spread in prisons.

Prisons provide a unique environment for the transmission of bloodborne infections, as high risk behaviours such as injecting drug use, and to a lesser extent, sexual intercourse, occur inside (Gill *et al.*, 1995; Weild *et al.*, 2000). In the past there has been a lack of facilities for UK inmates to minimise the risk of infection from bloodborne pathogens; for example, clean needles, sterilising equipment, condoms, and drug treatment programs were not routinely offered. In December 1993, the Scottish Prison Service introduced the provision of sterilising tablets to enable prisoners to clean injecting equipment (Taylor *et al.*, 1995). However, there are no plans to extend this to prisons in England and Wales, despite a successful pilot study in England in 1998, (Longfield, 1999). Moreover, HBV vaccination programmes in prisons have not been well implemented. A study of 429 inmates in 2 Scottish prisons was carried out in 1996, and revealed that only 4% had been offered vaccination (Bird *et al.*, 1997). Similarly, a HBV vaccine initiative in prisons in England and Wales during 1996 reached only about 15% of IDU inmates (Weild *et al.*, 2000).

As well as in three other prisons in the north of England, HBV^{PV} was identified in community acquired cases in neighbouring Warrington. The chronology of events suggested that the Warrington cluster may, like Hull, have been related to the introduction of HBV^{PV} into the Warrington area by contacts of infected

prisoners, or by prisoners themselves, following release. However, the subsequent identification of HBV^{PV} in a nosocomial transmission incident in London compromised such an inference, since it implied that HBV^{PV} was more prevalent than previously thought.

A BLAST search of the GenBank/EMBL databases using the HBV^{PV} surface and core PCR product sequences returned a large number of identical or very similar sequences from various studies, indicating that HBV^{PV} is a common variant worldwide. For example, HBV^{PV} -like sequences have been isolated from two heart transplant patients in France with chronic hepatitis B, from a long-term immunosuppressed HBV carrier in Germany, from two Nicaraguan chronic HBV carriers and from two South African patients (Stuyver *et al.*, 1999; Preikschat *et al.*, 1999; Arauz-Ruiz *et al.*, 1997; Bowyer *et al.*, 1997). Therefore, while there is a strong suggestion that in Hull, HBV^{PV} spread from the prisons to the community, the link between HBV^{PV} from inmates of northern English prisons to Warrington, and particularly London, is more tenuous.

The HBV outbreak amongst attendees of an alternative therapy clinic further demonstrated the utility of the DGGE method over the more expensive and labour intensive DNA sequencing. It was possible to rapidly identify 30/32 HBV DNA positive clinic attendees who were infected with an identical HBV variant (HBV^{AV}), which included two unvaccinated clinic staff. The epidemic nature of the infections (most of the cases being identified within a 6 month period) and the lack of other risk factors lent strong support to the inference that these individuals had been infected as a result of receiving autohaemotherapy treatment. The two further clinic attendees who were carrying different HBV variants were most likely chronic carriers. A large number of randomly selected acute HBV samples, from epidemiologically unrelated cases, were also screened by DGGE to determine whether HBV^{AV} had been circulating in the wider community. None were found to be carrying a variant identical to HBV^{AV}.

The epidemiology of HBV^{AV} contrasted markedly with that of HBV^{PV}. HBV^{PV} was highly prevalent in the population groups tested. This was probably due to its introduction and continuing person-to-person transmission amongst

individuals engaged in high-risk activities (predominantly prisoners and IDUs), as seen in Hull. This behaviour enabled spill over into other sections of the community.

In contrast, HBV^{AV} did not appear to be a prevalent variant in the UK. A search for HBV^{AV} and HBV^{AV}-like sequences in local HBV sequences database did not reveal HBV^{AV} as having been characterised prior to the autohaemotherapy outbreak investigation. Similarly, a BLAST search of the GenBank/EMBL databases did not return any sequences identical to the 601 bp examined. This rare variant was able to cause an explosive outbreak in individuals who would not normally be associated with high risk of bloodborne virus infection because it originated from a point source, most likely a contaminated bottle of saline.

In further contrast to HBV^{PV}, HBV^{AV} did not appear to be transmitted secondarily at a high rate. This was because the HBV^{AV} outbreak was quickly identified and vaccination made available to family and sexual contacts. Thus only one case of secondary infection was identified. Furthermore, the patients were not engaged in high-risk activities for bloodborne virus transmission. Prospective monitoring will reveal to what extent, if any, HBV^{AV} penetrates the general population.

Estimations of the mutation rate of HBV range from 1 to 8 x 10^{-5} substitutions per site per year in immunocompetent hosts (Okamoto *et al.*, 1987; Orito *et al.*, 1989; Lin *et al.*, 1991). This is approximately 10^4 times faster than the mutation rate in many other DNA viruses. It has previously been accepted that such a rate of drift in the sequence of circulating HBV genomes would be sufficiently rapid to result in the emergence of a highly heterogeneous population, and that this would allow discrimination between related and unrelated infections in a transmission investigation. However, previous studies have shown that identical fragments of the HBV genome can be recovered from individuals with no known transmission link between them. When Uy *et al.*, (1992) analysed 22 sera of the same subtype from German patients, seven different chains of transmission were suspected based on epidemiological data. They sequenced 126 bp covering the preS1 region and discovered identical sequences in 15 samples, and in a further 5

control samples. This level of conservation was unexpected, and thus they were unable to discriminate between different chains of transmission using the sequence data. Similarly, in a nosocomial cluster of infections in Sweden, two non-IDU patients were carrying a HBV variant previously found only in local IDUs (Kidd-Ljunggren *et al.*, 1999). Long-term genomic stability has also been demonstrated in analyses of intrafamilial transmission (Lin *et al.*, 1990; McIntosh *et al.*, 1998), and in a recent study of chronic carriers the same viral genome-length sequence could be recovered from an individual over periods of up to 35 years, when the patients were in the HBeAg positive phase and their immune response was consequently weak (Hannoun *et al.*, 2000).

This study of the molecular epidemiology of HBV^{PV} confirms the stability of the HBV genome. Identical surface and core gene sequences (representing, in total, one fifth of the viral genome) were found in samples from epidemiologically-unlinked patients, taken up to 8 years apart. This observation is likely due to the maintenance of the HBV^{PV} sequence by continuous transmission. It would be interesting, however, to discover whether it contains mutations in other regions of the genome not examined here that would render it particularly 'fit' or confer a replicative advantage.

In summary, the DGGE assay has been successfully used to study a commonly circulating HBV variant which had caused several clusters of infections in various population groups, and a rarely circulating variant which led to a localised outbreak of HBV infection. Close scrutiny of their molecular epidemiologies revealed sharp contrasts between transmission patterns of HBV in England. The DGGE method provided significant cost and time-saving advantages over DNA sequencing in these studies.

Chapter 4

The molecular characterisation of a variant of TTV

4.1 Introduction

The discovery of TT virus DNA in patients with post transfusion hepatitis prompted the search for further evidence that TTV is the causative agent. However, initial reports of an association between TTV DNA positivity and various forms of liver disease may have been due to the exposure of these individuals to the virus via past bloodborne transmission, as subsequent studies have failed to implicate TTV in the aetiology of hepatitis (Kanda *et al.*, 1999; Hsieh *et al.*, 1999; Fukuda *et al.*, 1999; Giménez-Barcons *et al.*, 1999).

It is now clear that TTV displays a high level of genetic heterogeneity. Full-length genome sequencing studies showed that the first primers to be used for PCR detection were situated in a particularly variable region (N22) within the longest ORF, ORF-1 (Okamoto *et al.*, 1999*c*; Erker *et al.*, 1999). Therefore, the initial prevalence studies were likely to have underestimated the true extent of TTV infections. Newer primer sets, directed against the more conserved non-coding region, have revealed that TTV is a ubiquitous infection in many population groups (Takahashi *et al.*, 1998*b*; Leary *et al.*, 1999*b*; Biagini *et al.*, 2000).

The genetic diversity seen even within short fragments of the TTV genome has led several groups to propose a system of genotyping based on phylogenetic analysis (Y. Tanaka *et al.*, 1998; Okamoto *et al.*, 1999*c*; Khudyakov *et al.*, 2000). However, the examination of full-length genomes suggests that these classification schemes are premature, and that TTV may possess diversity beyond that of a single virus species (Erker *et al.*, 1999; Okamoto *et al.*, 1999*b*; Hijikata *et al.*, 1999). It has recently been suggested that extensive recombination has occurred among divergent TTV types, which may help to explain the origin of such high levels of genetic diversity (Worobey, 2000). There is also speculation that certain genetic types of TTV may cause disease.

As TTV was the first virus with a circular single-stranded DNA genome to be detected in humans, its taxonomic status remains unclear. Viruses with a similar genomic structure are found in the *Circoviridae* family, which comprises two

genera: *Gyrovirus*, containing CAV, and *Circovirus*, with porcine circovirus (PCV) and beak and feather disease virus (BFDV) as members (Pringle, 1999). However, the genomes of TTV are almost twice as long as those of the *Circoviridae*, and little sequence homology has been demonstrated. The unique characteristics of TTV and the high level of sequence heterogeneity shown between types led Mushahwar *et al.* (1999) to suggest that TTV be classified as a new family of viruses called the *Circinoviridae.* With the subsequent identification of a human DNA virus (TLMV) that shares similarities with both TTV and CAV, it is clear that the final classification of these viruses will require further examination.

In this study a previously undescribed divergent TTV-like sequence was identified, its full genome sequence was characterised and its relationship to other TTV sequences was examined using phylogenetic techniques. Similarly, comparisons were made between representative TTV sequences, TLMV and CAV genomes in an attempt to provide a rational basis for their classification. Finally, PCR and sequencing were used to screen sera for the divergent TTV-like sequence in order to assess its role as a hepatitis agent.

4.2 Detection of a divergent TTV-like sequence

4.2.1 Survey of TTV sequences present in two patient groups

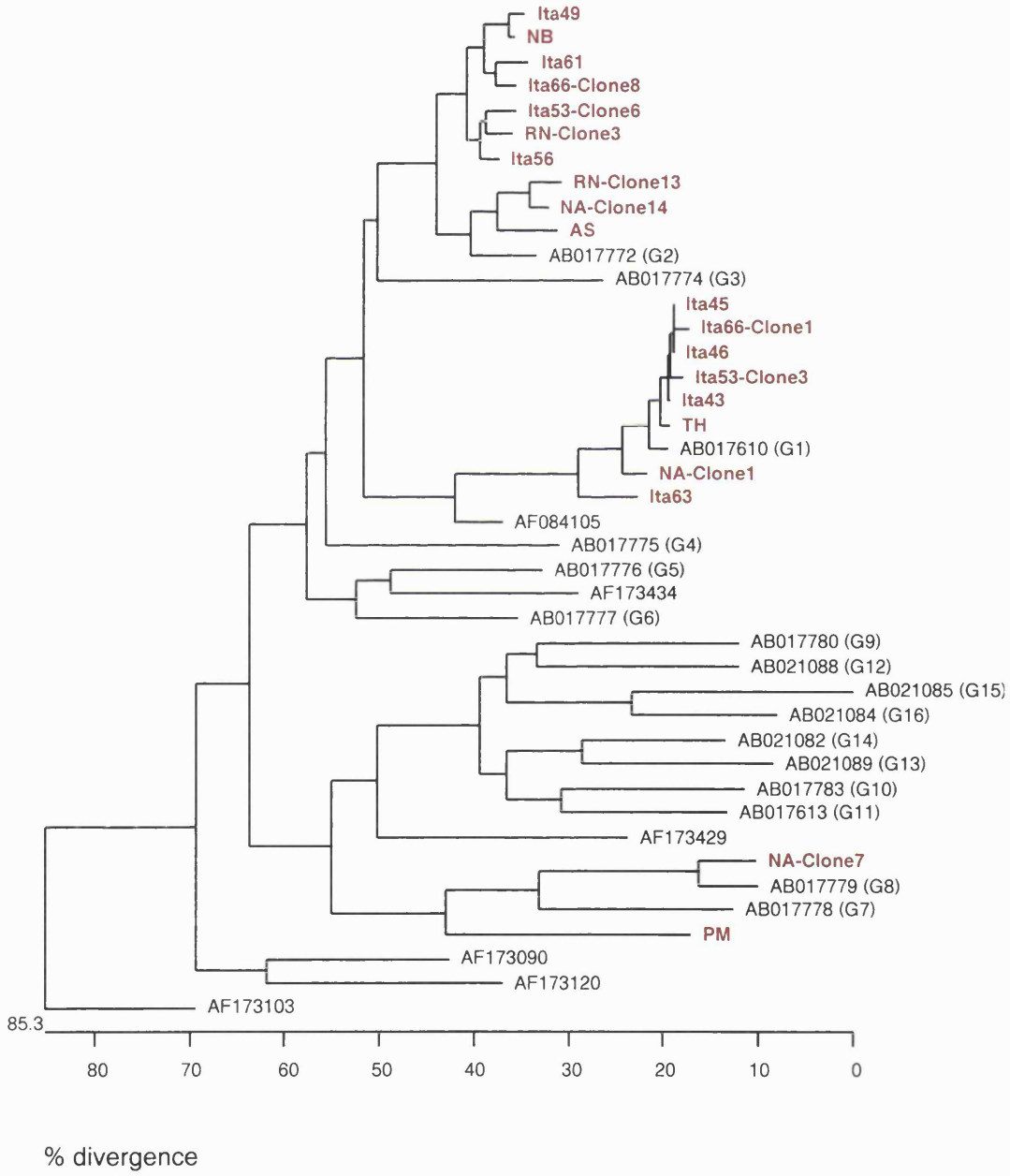
Serum samples from two patient groups, British (n=43) and Italian (n=24), were tested for TTV DNA. The former were referred to CPHL for routine acute non-A to E hepatitis confirmatory testing. The latter were from 20 patients with various forms of chronic hepatitis related to infection with HCV (three with steatotic hepatitis, ten with chronic active hepatitis, five with asymptomatic HCV infection and two with cirrhosis), and from four with no history of hepatitis. All samples were tested for TTV DNA following extraction of DNA from 100 µl of serum as described in Section 2.1.1. A 271-bp fragment from the N22 region of the TTV genome was amplified from 15 samples using PCR conditions described in Section 2.6.2. Of these 15, six were British samples and nine were Italian. The PCR products were sequenced in both directions (Section 2.3).

Four of the 15 positive samples, two from each patient group, appeared to contain more than one sequence, as evidenced by multiple double peaks on the sequencing chromatograms. This suggested that the four patients were infected with more than one genetic type of TTV. The secondary PCR products from these samples were cloned (Section 2.7) and 15 colony PCR products from each were sequenced. The resulting sequences were trimmed to 201 bp in order to align them with 22 TTV sequences available in the GenBank/EMBL databases. These were chosen to represent the most diverse TTV sequences available in order to assess the genetic types circulating in the 15 positive patients. Sixteen of the reference sequences had been assigned a genotype by Okamoto *et al.* (1999*c*) on the basis that they were at least 30% divergent from each other. The remaining six reference sequences were at least 30% divergent from each of the sixteen genotypes and could therefore be considered as further genotypes according to Okamoto's classification system. Multiple alignments and phylogenetic analysis were carried out as described in Section 2.3.2 and the resulting dendrogram is shown in Figure 4.1.

Twenty different sequences were identified. Eleven were obtained directly from PCR products (samples Ita43, Ita45, Ita46, Ita49, Ita56, Ita61, Ita63, TH, AS, NB and PM). The colony PCR products from the four putative mixed infections revealed two different TTV types in three of the samples (Ita53, Ita66 and RN) and three different types in sample NA. Within each type, some of the clones differed by single nucleotide substitutions from the consensus sequence. However, only the consensus sequence is shown in Figure 4.1; thus a further nine sequences were obtained from these four samples.

Using the alignment and neighbour-joining algorithms available in MegAlign (DNASTAR Inc.), eight of the 20 sequences fell into one group, clustering with the genotype 1 described by Okamoto *et al.* (1999*c*). Ten of the sequences fell into the genotype 2 cluster. Two further sequences (PM and NA-Clone7) were very divergent and did not fall into either of these two clusters. NA-Clone7 was most closely related to genotype 8 (11.9% divergence from AB017779). Sequence PM was at least 55.8% divergent from all other TTVs shown in Figure 4.1, and was thus considered the prototype sequence of a further genotype.

Figure 4.1 Dendrogram produced from a nucleotide alignment of 20 TTV N22 region sequences identified in this study (shown in red) and 22 representative TTV sequences from GenBank/EMBL, which are indicated by their accession numbers. Genotypes are assigned according to the system of Okamoto *et al.* (1999*c*) and are shown in parentheses.



This sequence originated from patient PM, a four year old girl who had presented with acute non-A to E hepatitis. She had no risk factors for bloodborne infection or travel outside the country.

A homology search was carried out, using both the nucleotide and the amino acid sequences of the PM fragment. Sequences contained in the databases GenBank/EMBL and SwissProt were searched using the programs described in Section 2.8.3. No significant similarity was revealed between PM and any of the database sequences.

Figure 4.2 shows an alignment of the predicted amino acid sequences of the reference sequences and PM. This clearly demonstrates the extensive heterogeneity in this subgenomic fragment. Only 8 of the 72 amino acid sites (11.1%) showed complete identity.

4.2.2 Assessment of the stability of TTV DNA

The mutation rate of TTV DNA in a patient with acute non-A to E hepatitis (sample TH) was investigated. Four serum samples were taken over a 13 month period and the patient remained TTV DNA positive throughout this time. The secondary PCR products, amplified from samples taken at months 0 and 13, were cloned as described in Section 2.7. Sequence analysis revealed that the dominant sequence was the same in both samples, and was identical to the sequence obtained directly from the secondary PCR products. It was found in 6/10 colonies from month 0 and 7/10 colonies from month 13. The remaining colony sequences contained either a unique single nucleotide substitution or deletion.

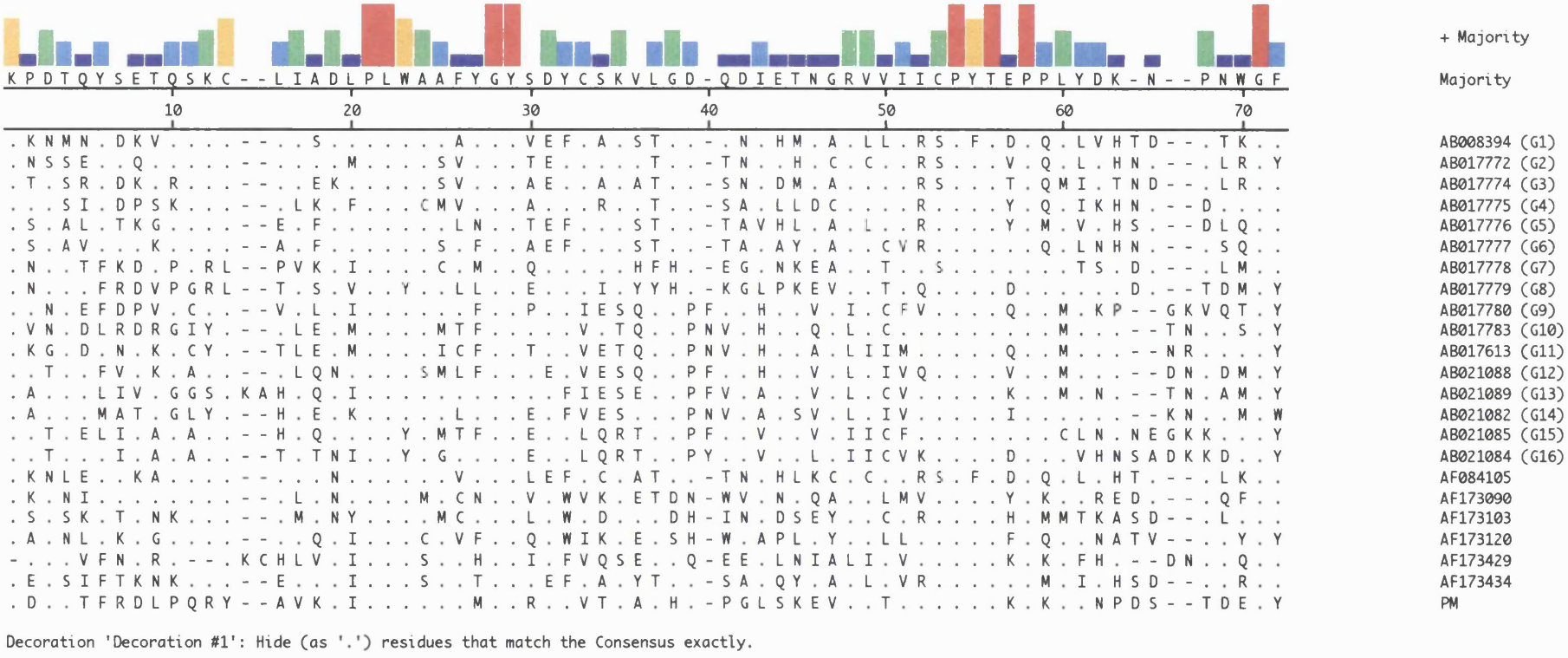


Figure 4.2 Deduced amino acid alignment of 22 reference TTV sequences from the N22 region of ORF1 and PM.

4.2.3 Cloning PM PCR products

In order to investigate the potential presence of other TTV types circulating in patient PM, the secondary PCR products were cloned as described in Section 2.7, and ten colony PCR products were sequenced. These were trimmed to 216 bp and aligned with the direct sequence obtained from PM PCR products (Figure 4.3). Eight of the ten colony PCR products were identical to the direct sequence. Each of the remaining two contained a single A→G substitution, which were synonymous mutations, although the possibility that they were a result of *Taq* polymerase misincorporation could not be excluded. The data indicate that patient PM was not infected with more than one genetic type of TTV at the time of sampling.

4.2.4 Detection of PM sequence in archived liver tissue

DNA was extracted from an archived paraffin–embedded liver biopsy sample obtained from patient PM at the time of acute hepatitis, as described in Section 2.6.1, and used as template for the TTV N22 PCR. The expected 271 bp fragment was amplified and sequenced, and is shown in the alignment in Figure 4.3. It was identical to the direct sequence obtained from the serum sample, which provides evidence that the virus from which sequence PM is derived can infect the liver.

4.3 Characterisation of the PM virus (PMV) genome

4.3.1 Initial sequence analysis and comparison with TTV genomes

The sequence of the PM fragment was extended to the full-length genome by producing three overlapping PCR products, as described in Section 2.8.2. This full-length sequence is subsequently referred to as PMV. PCR products A and B (Figure 2.1) were generated using standard long range amplification conditions. However, PCR product C was not amplifiable in this way due to the presence of a GC-rich region in the template DNA, and a PCR under conditions specific for GC templates was necessary to produce this fragment (Section 2.8.2). The full-length PMV sequence was 3736 nucleotides long and is shown in Figure 4.4.

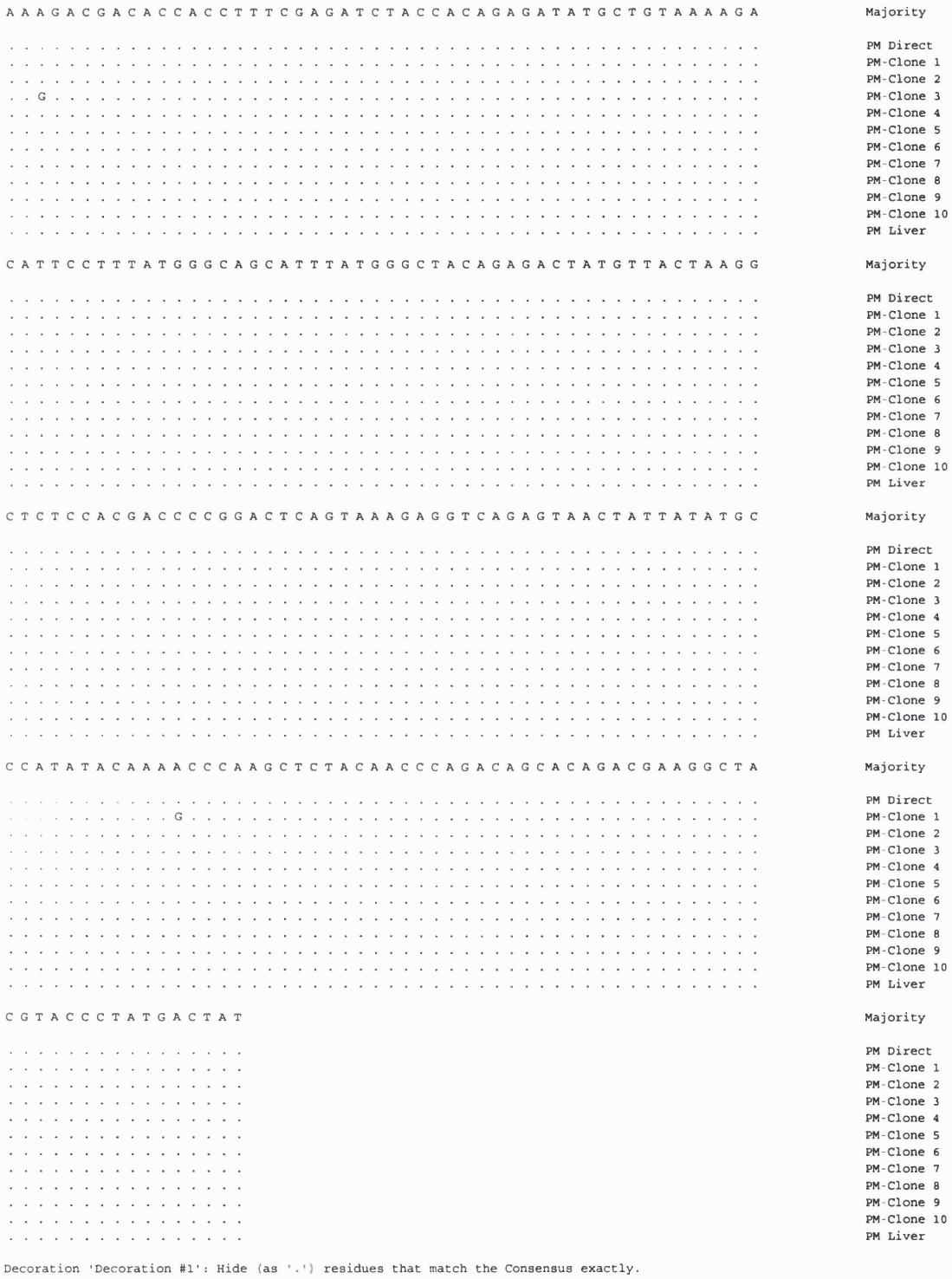


Figure 4.3 Nucleotide alignment of PM sequences obtained directly from the PCR products amplified from a serum DNA extract (PM Direct), from 10 colony PCRs (PM-Clone 1-10) and from PCR products amplified from a liver DNA extract (PM Liver).

Figure 4.4 Nucleotide sequence of PMV genome

Putative coding regions are indicated as: ORF 1 ORF 2 **ORF 3**

GCTACGTCACTAACCACGTGCACTTCACAGGCCAACCAGAGCCTATG
TCGTGCACTTCCTGGGCCGGGTCTACATCCTAATATAAGTAACTGCAC
TTCCGAATGGCTGAGTTTTCCACGCCCGTCCGCGGCGAGAACACCAC
GGAGGGGAGTCCGCGCGTCCCGTGGGCGGGTGCCGAAGGTGAGTTTA
CACACCGAAGTCAAGGGGCAATTCGGGCACGGGACTGGCTGGGCTAT
GGGCAAGGCTCTTAGGGTTATCATTCTTAAAACATGTTTTTTGGCAGG
CCGTGGAGAAAGAAAAGGAAAGTGCCACTGCCGCCTCTGCAGACTG
AACCGCTACAACCACCTATACACATGAGCCTGTGGAGACCCCCGGTA
TACAATGTCCCGGGACAAGAAAGGCAGTTTTATGAAGCCTGCCTGCA
TGCTCACGATGCTTTTTGTGGTTGTGGCGATTTTGTTGCTCATATTAAC
AGTGTGGCGGCTCGTGTTGGACGTCCTTCAGACAATCAACCCCCGAG
ACCCCAGAGGCCACCTGCAATAAGATCCATGAGAGCTCTACCAGCTC
CTCCAAGTAGCGGCACATCCACAGACAACCAGGGAGCACAATGGCCT
GGAGATGGTGGTGGAGAAGGCGCCGAAGGTGGCCCAGACGACGCAG
GACCACCTGGAGACGGAGACCTCGCCCCAGAAGACGTAGAAGAACT
GCTAGAACTCGTCGAAGAGGCAGAGTAAGGAGATGGAGGCGGAGAG
GGAGAGGGTGGCGGAGAAGGACCTATATAAGACGTAGAAGGCGCCG
CAAAAAGAAAATTAACATACAACAGTGGAACCCCGCCACTGTAAAG
AAATGCGTGGTCACCGGATACGTACCCCTCCTCATCTGTGGAACGGG
TACCACGGGAACCACCTACAAAAACTATGGCAGCCACATAAATGACT
ATACAAAGTTTGACCCCTTTGGAGGTGGGTTTAGTACTTTAATGTTTA
ACCTAAGAATACTATTTGATGAATATAAAAAGCAAAGATGTCGCTGG
AGCAGATCTAACGATGAACTAGAACTTGTAAGGTACCTAGGCTGTAG
CTTTACCTTGTACAGAGAGCAAAATGTTGACTTTTTATTTAAATATAA
TAGAAGACAGCCATTTTCAGACAGCCAACTTACAGGACCTAGCTTAC
ACCCAGGCATAATAATGAAACAAAAAAGAAAAGTAATAGTGCCTAG
CTACAAAACTAAACCTAAAGGCAGACCCGTAAAAAGAATAAGAATA
AAGCCCCCAACTCTCTTTACAGACCGTTGGTACTTTCAAAAGGACTTT
AGTAACTTCCCCCTAGTAACTATCAGTGCCTCTGCGGCTAGCCTGCGG
TTTCCGTTCTGCTCACCACAAACTGGCAACATTTGCATATACTTCCAG
ATTTTAGATCCCTACTGGTATAACCGCCGGATGTCCATAACCCCAGAC
TTACTAAAGTCAAACTATGAAAGTTTTTTTAGTGTATTAAACACTAAA
TTTAGTAGCAGCCCCCTTTATCCAAATGTAGCCACCACAGTACCATCA
GGACCAATAGGCACTGTATTTAACACATTTAAAACCCAAGAACACGT
AGTAGACCCCAGATGGGATACTGTAAAAAAACAGAGCACAGGCAGT
CAGTGGTCTAATACTTTTGTAGACTCACATTGGGGAGACCACATATAT
GAAAACACCAGCAGTAAATCAATACTAACTGCTATGAAAGACAATGC
AACTAAAATGTATGACAGAAGAAAAAGAGATACATACCTAGGCAGT
AAATATTTAAACTATAGAACAGGACTATACAGCAGCATATTTTTAGC
TAATGAAAGAACCAGCCCTGACTTTCCAGGCCTATACCAAGAAGTAG
TTTACAACCCTTTAGTAGATGAAGGAGAAGGCAACATAGTATGGATA
GACTGGTGCTCTAAAGACGACACCACCTTTCGAGATCTACCACAGAG
ATATGCTGTAAAAGACATTCCTTTATGGGCAGCATTTATGGGCTACAG
AGACTATGTTACTAAGGCTCTCCACGACCCCGGACTCAGTAAAGAGG
TCAGAGTAACTATTATATGCCCATATACAAAACCCAAGCTCTACAAC
CCAGACAGCACAGACGAAGGCTACGTACCCTATGACTATAACTTTGG

AAAAGGGAAAATGCCAGACGGCAACGGGTACATACCCATAGCCTAC
AGATTTCAGTGGTACCCTTGCATGTTCCACCAACAGAACTTTATGAAT
GACATTGTACAGTCAGGACCCTTTGCATACCACGGAGAAGAAAAAAG
CTGCACTCTAACTGCAAAGTACCGCTTCAGATTTTTATTTGGGGGCAA
TCCTATATCTCAACAGGTCATTAAAGATCCCTCTAAACAACCCGACTT
CCAAATACCCGGAGCCCGTGACCTCTTTAGCACAGTACAAGTCACGA
ACCCGAAACTCATCGACGAGGGATACTTCTTCAACGCCTGGGACATC
AGACGTGGGCTATTTGGCTTCTCAGCTATTAAAAGAATGCAAAACCA
ACAAATCCCTACTAAATATTTTACAGGCCCACCAAAGCGCCCGCGAT
TCGAGGTACCCGCCATTGCAAACGCAGGCTCAGATTTACACCCTACC
AAATGGCACCCCTGGAACGAAACCAGCGAGGAGGAAGAAGACCAGA
AAGAAGACTCCCCCTCGCCGCAAACGAAGGCACCATTACAGCTCATC
CTCAAACAGCAGCTCAGGGAGCAAAGAGAGCTCAAGCGACGAATCC
AATTCCTAGTAAAACAGCTAGTAAAGACCCAATATCACCTCCACGCA
CCCATTATCCCTAGTAGGACCTCCAGCCTTACTGTTCCCAGAG**ATGAC**
CAAAAAAATATTTTCCAGCAACGACTGGAAATGTGAGTATGAATC
CTGTAAGCAGTGGGACAGGCCAGCTAGACAAAGCCTTACAGACA
CCCCTTACTACCCTTGGATGCTACCCCCAAAACAGCCATTCTCTG
TAACATTTAAACTAGGCTTCAAATAAACAAGGCCGTGGGAGTTTCA
CTTGTCGGTGTCAGCTTATTAAGGTCACTAAGCACTCCGAGCGAAGC
GAGGAGTGCGACCCTTACTGTGGGGCCCACTTCTTCGGAGCCGCGCG
CTACGCCTTCGGCTGCGCGCGGCACCTCAGACCCCCGCTCGTGCTGAC
ACGCTTGCGCGTGTCAGACCACTTCGGGCTCGCGGGGGTCGGATGCT
TTATTAAACAGACTCCGAGGTTGTCTTGGACACTGAGGGGGTGAACA
GCAACGTAAGTGAGTGGGGCCAGACTTCGCCATAAGGCCTTTATCTT
CTTGCCATTTGTCAGTAACTTAAATTGCCATAAACTTCGACCTCAATT
TTAGGCCTTCCGGACTACAAAAATCGCCATTTTAGTGACGTCACAGC
AGCCATTTTAAGTAAGGCGGAAGAGCTCTAGCTATACAAAATGGCGG
CGGAGCACTTCCGCTTTGCCCAAAATGGTGGGCAAGCACATCCGGGT
CAAAGGTCATGCCTACGTCACAAGTCACGTGGGGAGGGTTGCCGTGT
AACCCGGAAGTCAATCCTCCCACGTGGCCTGTCACGTGACGCATACG
TCACGGCGGTCATTTTGTTTTACAAAATGGCGGACTTCCTTCCGCTTT
TTTGAAAATAACGGCCCAGCGGCGGCGCGCGCTACGCGCGCGCGC
CGGGGGGCTACCGCCCCCCCCGT

Other TTV-like sequences that were at least 90% of full genome length were sought in order to make comparisons with PMV, and 23 were identified in GenBank/EMBL. These are listed in Table 4.1. A database homology search was carried out using the nucleotide sequence of the PMV genome. It did not reveal any significant nucleotide similarity between the PMV sequence and other sequences, apart from TTV sequences.

Table 4.1 Full and near full-length TTV sequences

Accession number	Sequence name	Genome length (nt)	Reference	Genotype[c]
AB017610	TA278	3853	Miyata *et al.* (1999)	1
AF122913	GH1	3852	Mushahwar *et al.* (1999)	1
AF122914	JA20	3853	Erker *et al.* (1999)	1
AF122915	JA9	3852	Erker *et al.* (1999)	1
AF122916	JA1	3839	Erker *et al.* (1999)	2
AF122917	JA4	3537[a]	Erker *et al.* (1999)	3
AF122918	JA2B	3840	Erker *et al.* (1999)	2
AF122919	JA10	3539[a]	Erker *et al.* (1999)	3
AF122920	US35	3839	Erker *et al.* (1999)	2
AF122921	US32	3539[a]	Erker *et al.* (1999)	2
AF116842	BDH1	3739	He *et al.* (1999)[b]	1
AF079173	TTVCHN1	3739	Yusen *et al.* (1998)[b]	1
AF129887	TTVCHN2	3739	Huang *et al.* (1999)[b]	1
AB017613	TUS01	3818	Okamoto *et al.* (1999*b*)	Unclassified
AB025946	SANBAN	3808	Hijikata *et al.* (1999)	SANBAN
AB038619	TTVsan-IR1031	3795	Takahashi *et al.* (2000*b*)	SANBAN
AB038620	TTVsan-S039	3792	Takahashi *et al.* (2000*b*)	SANBAN
AB038621	TTVyon-KC009	3676[a]	Takahashi *et al.* (2000*b*)	YONBAN
AB038622	TTVyon-LC011	3675[a]	Takahashi *et al.* (2000*b*)	YONBAN
AB038623	TTVyon-KC186	3676[a]	Takahashi *et al.* (2000*b*)	YONBAN
AB038624	TTVyon-KC197	3676[a]	Takahashi *et al.* (2000*b*)	YONBAN
AB028668	TJN01	3787	Ukita *et al.* (2000)	Unclassified
AB028669	TJN02	3794	Ukita *et al.* (2000)	SANBAN
AF261761	PMV	3736	This study	Unclassified

[a] Incomplete genome sequence

[b] Published in GenBank/EMBL only

[c] Genotype assigned, based on full sequence analysis, according to Erker *et al.* (1999: 1, 2 and 3) or Takahashi *et al.* (2000*b*: SANBAN and YONBAN)

In a comparison with these full or near full-length TTV genomes it was found that PMV shared several features, including the positions of putative ORFs 1 and 2 and the non-coding region, a GC-rich string of bases and a TATA box. A predicted third ORF was not conserved among all TTV genomes. A polyadenylation signal (AATAAA), starting at position 3028 in PMV, was found in all full and near full-length genomes with the exception of the YONBAN group (Takahashi *et al.*, 2000*b*).

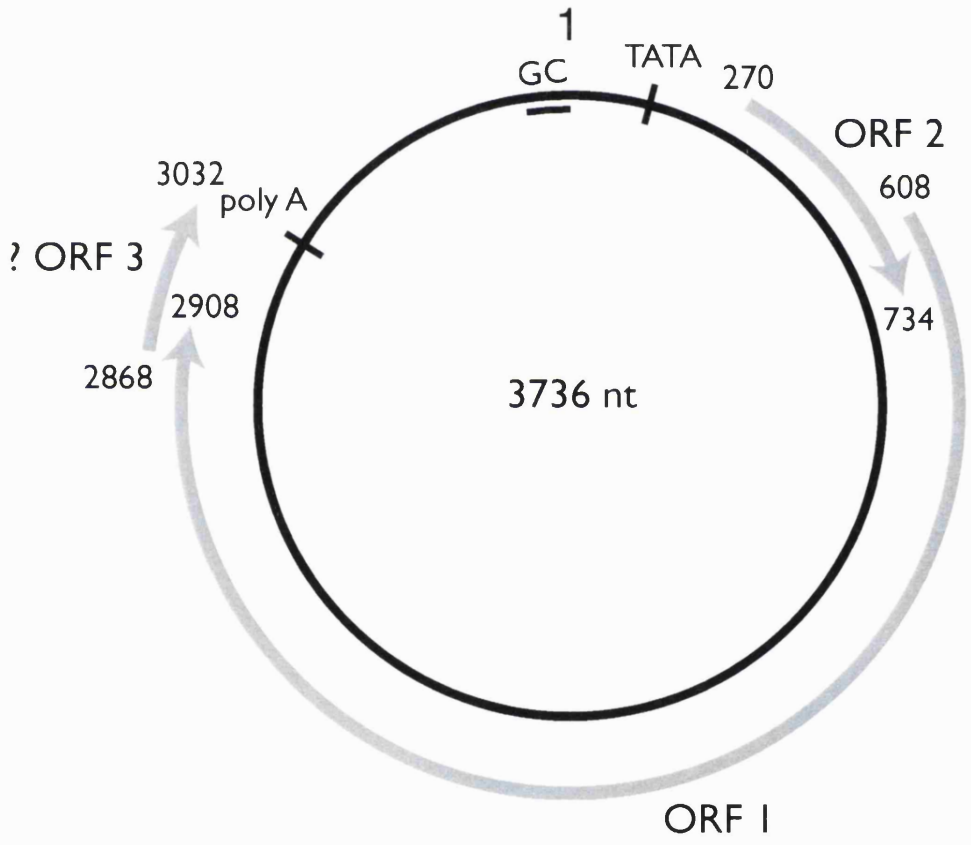


Figure 4.5 Organisation of the PMV genome. The relative positions of the putative ORFs, TATA box (TATA), polyadenylation signal (poly A) and GC-rich region (GC) are indicated. Nucleotides are numbered according to the start position of the prototype TTV genome sequence, TA278 (GenBank accession no. AB017610).

A nucleotide alignment of the 23 full and near full-length TTV sequences and PMV was carried out as described in Section 2.8.3, after trimming the full-length sequences. Genetic distances were estimated using Dnadist, under Kimura's 2-parameter model of nucleotide substitution, and a tree was produced from this using the program Fitch (in PHYLIP, Felsenstein, 1993). This is shown in Figure 4.6. Five major groups were identified, corresponding to genotypes 1, 2 and 3 described by Erker *et al.* (1999), and the SANBAN and YONBAN groups (Hijikata *et al.*, 1999; Takahashi *et al.*, 2000*b*). Sequences TUS01, TJN01 and PMV formed distinct, highly divergent branches.

In subsequent analyses, one representative sequence from each of the five major groups was used, as well as TUS01, TJN01 and PMV. These were:

TA278	(genotype 1)
JA1	(genotype 2)
JA4	(genotype 3)
SANBAN	(SANBAN group)
TTVyon-KC186	(YONBAN group)

When the phylogenetic analysis was repeated using just these 8 trimmed sequences, a different branching order was evident. However, a similar tree was recovered when using both a distance method, under all available models of substitution (Dnadist and Fitch) and a maximum likelihood method (Dnaml). These are shown in Figure 4.7. Bootstrap values from 1000 replicates are indicated on the Fitch tree. A low value for the branching position of sequences TUS01 and TJN01 suggests that their relationship to the other TTVs was not easily resolvable. The calculated genetic distances are given in Figure 4.8. The mean genetic distance between each sequence was 0.61 (range 0.34 - 0.77).

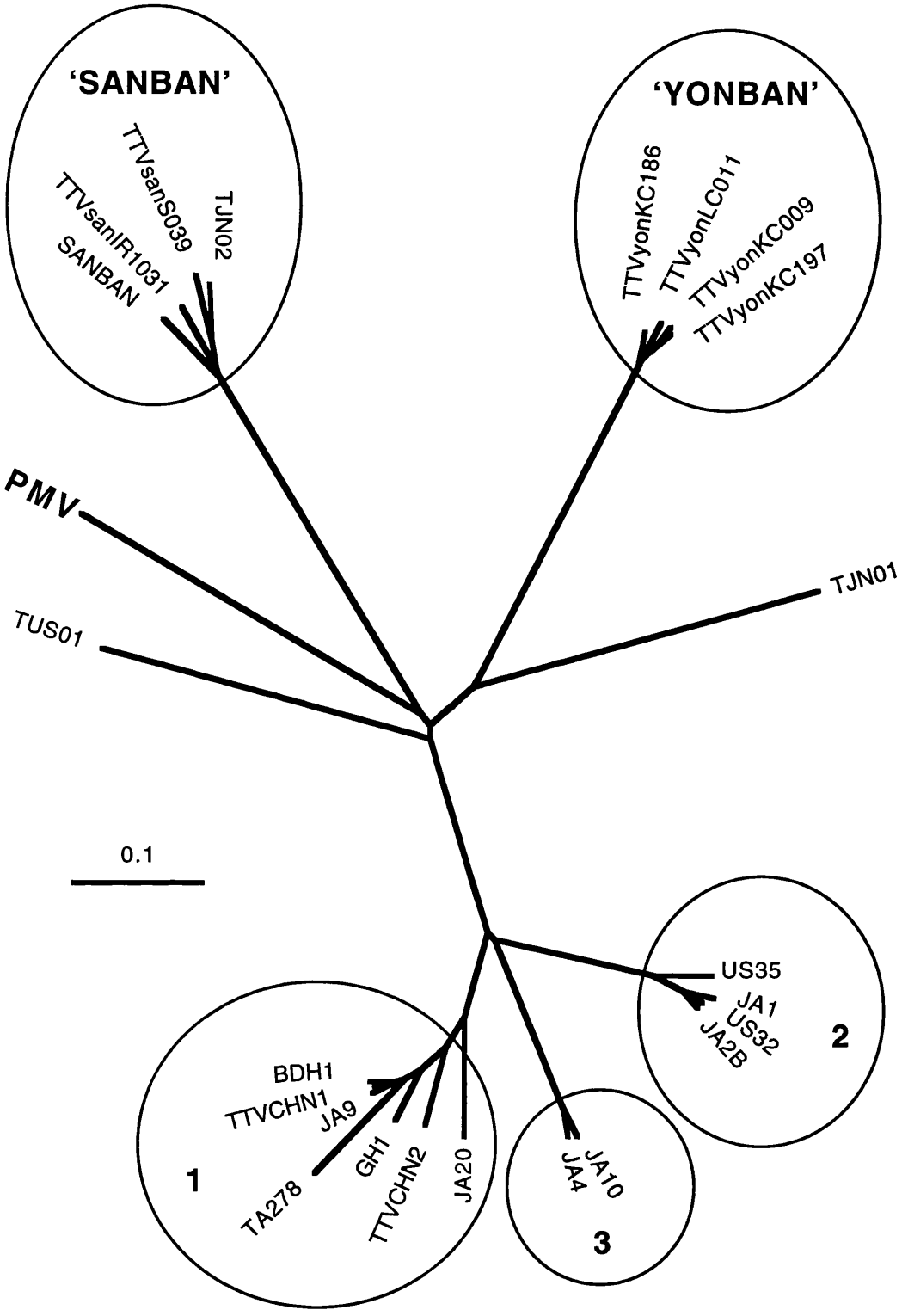


Figure 4.6 Phylogenetic tree of 23 trimmed TTV sequences from GenBank and PMV (shown in bold). Genotypes 1, 2 and 3 have been assigned by Erker *et al.*, (1999). The scale bar indicates the proportion of nucleotides substituted for a given branch length.

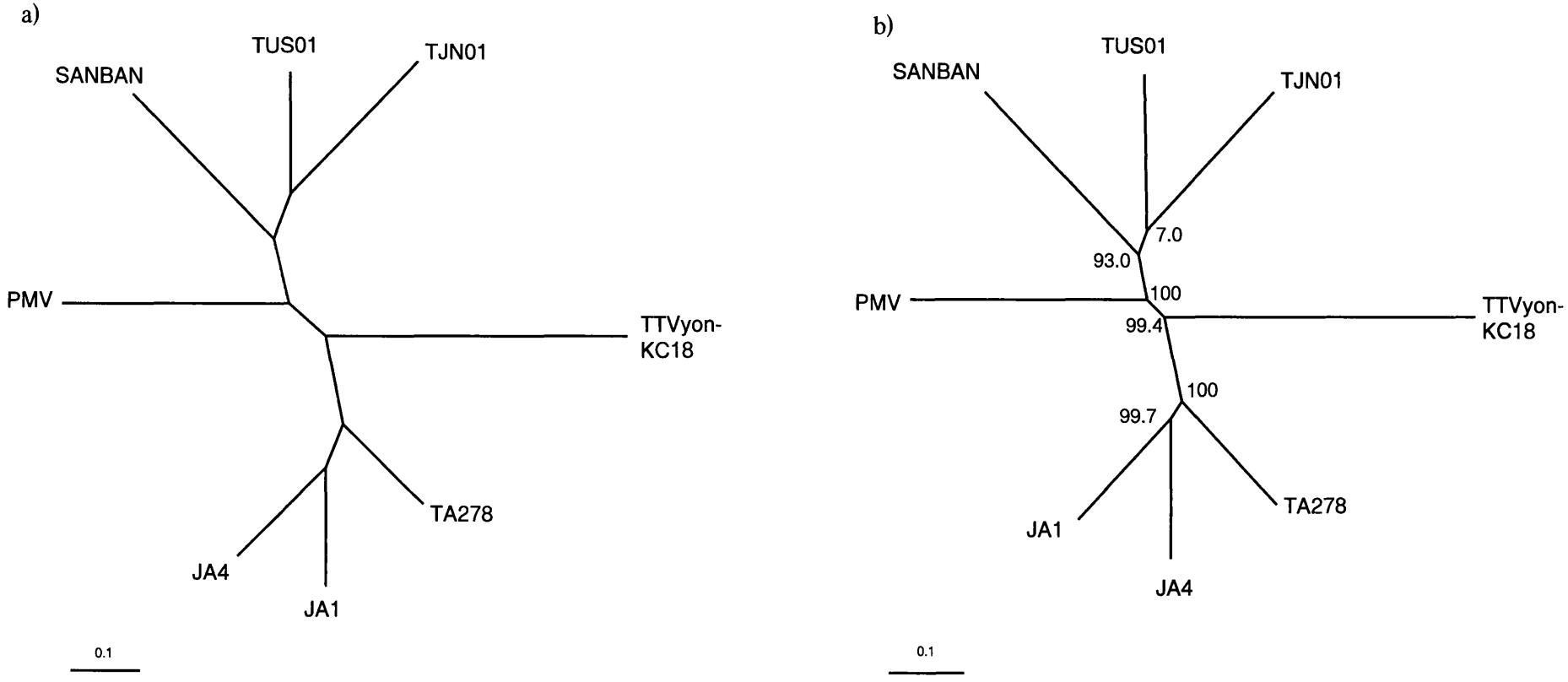


Figure 4.7 Phylogenetic trees of seven representative trimmed TTV sequences and PMV, produced using a) Dnaml, a maximum likelihood method, and b) Dnadist and Fitch, a distance method. Tree b) is the consensus tree obtained after 1000 bootstrap replicates, and the percentage occurrences of each node is indicated. The scale bars indicate the proportion of nucleotides substituted for a given branch length.

Figure 4.8 Genetic distances used to produce the Fitch tree in Figure 4.7 b.

	TA278						
JA1	0.3700	JA1					
JA4	0.3780	0.3375	JA4				
TJN01	0.6574	0.6192	0.6172	TJN01			
TUS01	0.5919	0.6167	0.5961	0.4236	TUS01		
SANBAN	0.6708	0.6775	0.6681	0.5790	0.4734	SANBAN	
TTVyon-KC186	0.6471	0.7382	0.7279	0.7584	0.7057	0.7536	TTVyon-KC186
PMV	0.6123	0.6431	0.6393	0.6307	0.6070	0.6127	0.7665

The similarity of PMV to the seven representative sequences was plotted across the complete nucleotide alignment (Figure 4.9a). This analysis was carried out using PlotSimilarity as described in Section 2.8.3, and reveals domains of conservation, primarily in the non-coding region, and variability across the 3.7 kb studied. The average nucleotide sequence similarity by this method was 0.51.

4.3.2 Open reading frames

The longest putative ORF, ORF-1, of the PMV genome coded for 767 amino acids and had an average amino acid identity of only 32.5% with the corresponding ORFs of 23 other TTVs (Table 4.1). It had an N-terminal arginine-rich region, noted previously in other TTVs (Takahashi *et al.*, 1998*a*). Also, three N-glycosylation motif sites were present (NTS, NAT and NET), suggesting that the product may be a glycoprotein. When the amino acid sequence of ORF-1 was submitted to a FASTA search of the SwissProt database, the N-terminal was found to be more than 50% identical to 20 different arginine-rich sperm protamine proteins, over a 53-62 amino acid stretch.

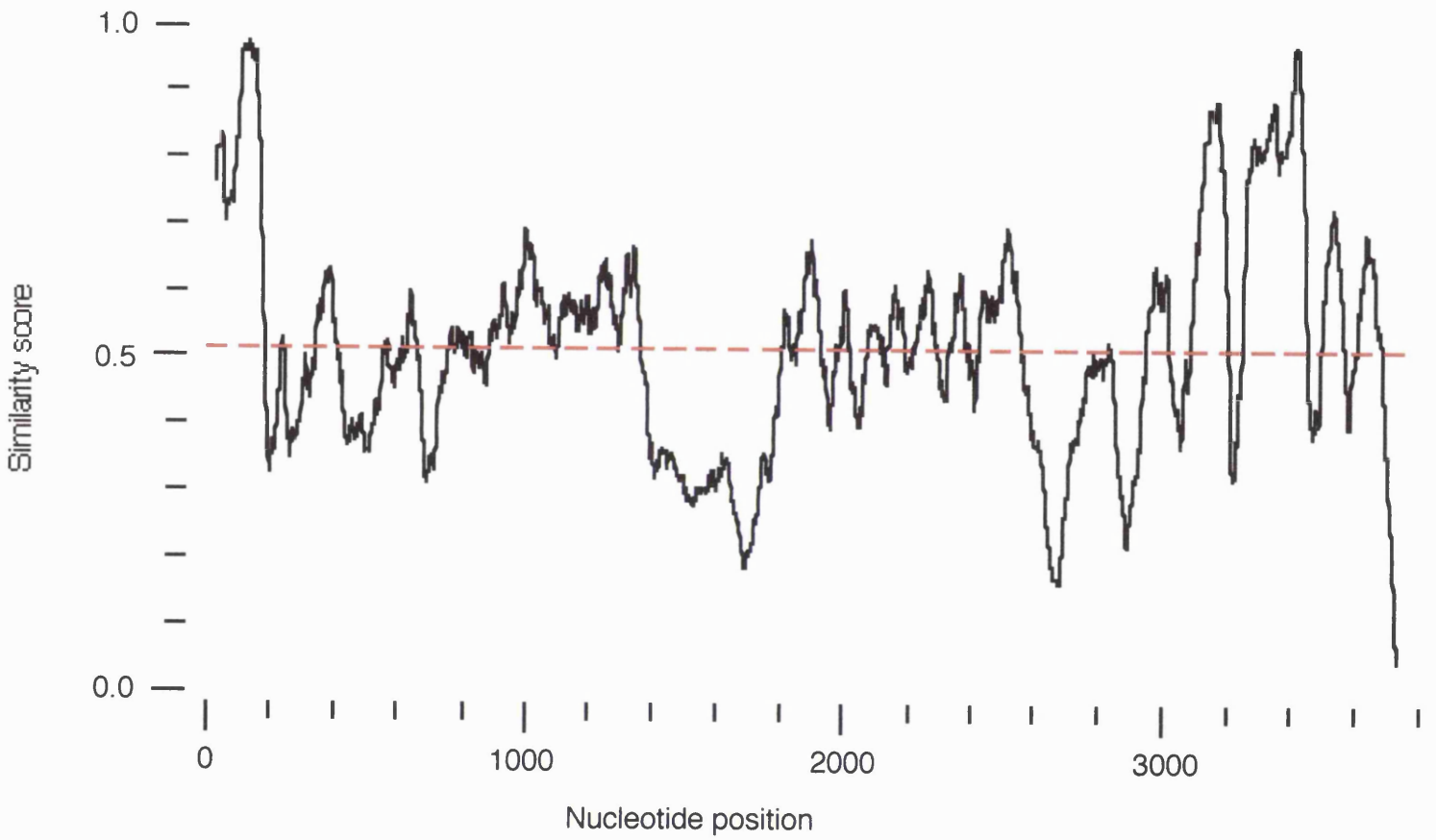


Figure 4.9a Similarity plot of an alignment of the PMV genome sequence with seven representative TTV genomes. The average nucleotide sequence similarity is indicated by the red dashed line.

A shorter ORF, ORF-2, could be identified in PMV, but again there was a high degree of sequence divergence from other TTVs, with the average amino acid identity being 32.3%. The sequence motif containing the initiation codon of ORF-1 in PMV had a single nucleotide difference from that described by Kozak (1986) as the optimal sequence for translation initiation (ACCATGG). The ORF-2 initiation motif possessed two suboptimal nucleotides. These were ACAATGG and AACATGT respectively (suboptimal nucleotides underlined). Conserved binding sites for the transcription initiators Sp1 (GGCGGG) and Cap (CAATTC), immediately upstream of the ORF-2 initiation codon, were also identifiable in PMV (Hijikata *et al.* 1999).

An amino acid similarity plot for ORF-1 revealed hypervariable regions between amino acid positions 300 and 440, and between 700 and 770 in a multiple alignment of PMV with the seven representative sequences. The amino acid plot for ORF-2 showed a similar level of overall similarity to ORF-1 (Figure 4.9b).

The putative ORF-3, proposed by Erker *et al.* (1999), was 165 nucleotides long in PMV. However, it overlapped with the C-terminal end of ORF-1 by 41 nucleotides, a feature not seen in TTV genome sequences so far described. Erker *et al.* (1999) reported ORF-3 to be highly conserved, but when the full or near full-length TTV sequences from their study were analysed with sequences TUS01, SANBAN, TJN01 and PMV, ORF-3 was revealed to be variable. The four sequences comprising the 'YONBAN' group had a similar coding region, downstream of ORF 1, but they lacked an initiation codon. The average amino acid sequence identity between PMV and all other ORF-3 sequences was 45.3%. The Kozak optimal initiation motif could not be identified at the PMV start codon.

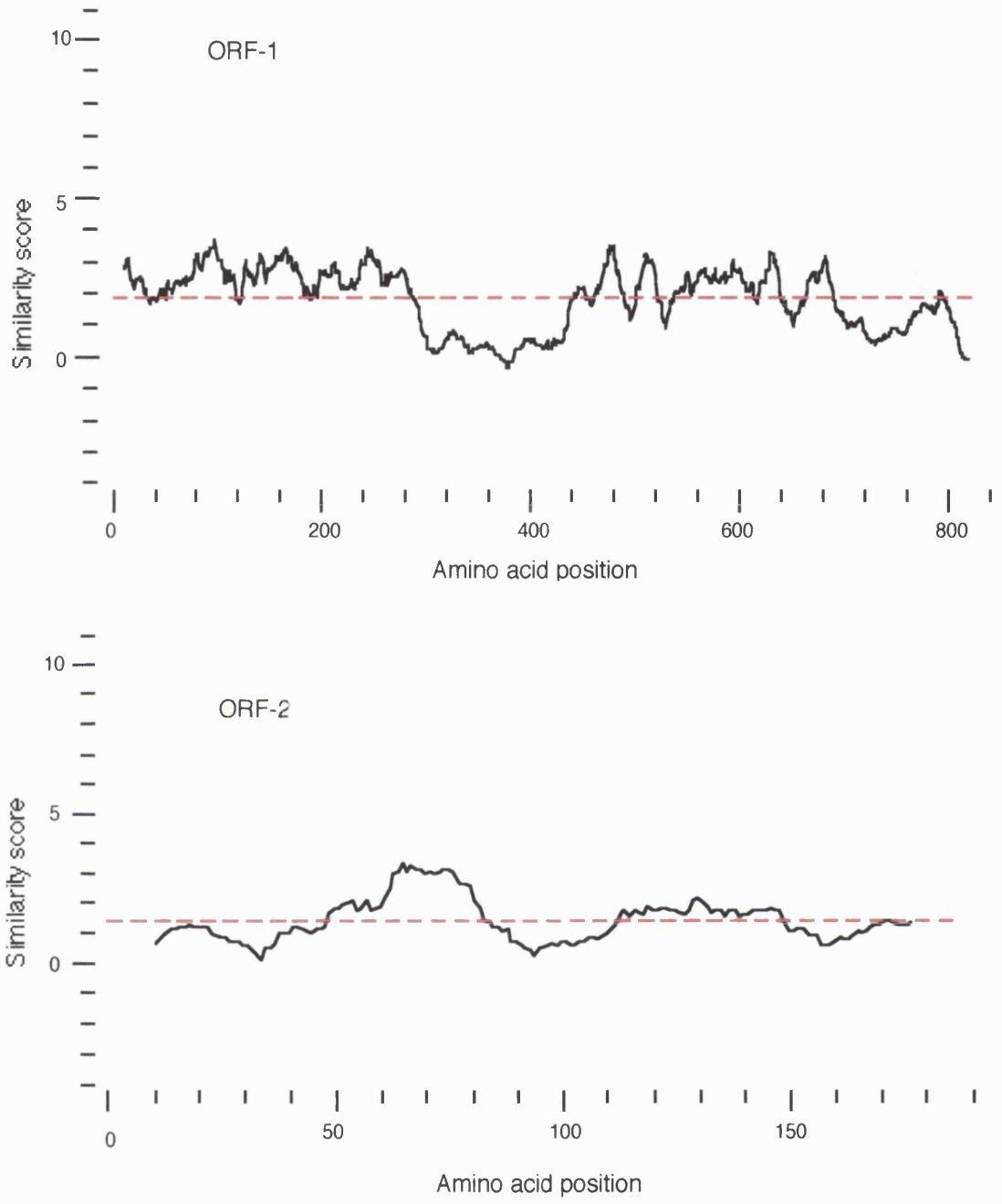


Figure 4.9b Amino acid similarity plots of alignments of ORFs 1 and 2 from PMV and seven representative TTVs. The average amino acid similarity is indicated by the red dashed line.

4.3.3 Non-coding region

A conserved non-coding region was found to stretch from the polyadenylation signal to the initiation codon of ORF-2, encompassing the GC-rich region. Comparisons were carried out between PMV and five full-length TTV sequences (TA278, JA1, TUS01, SANBAN and TJN01). There were no full-length sequences available for any genotype 3 or 'YONBAN' sequences. In three TTV genomes, the non-coding region was about 1040 nucleotides long; in TUS01 and TJN01 it was shorter at 1014 and 1011 nucleotides respectively. The equivalent region in PMV was only 972 nucleotides long, the most noticeable truncation being in the GC-rich region, which was 58 nucleotides in length compared to 117 in the prototype TTV genome TA278 (Miyata *et al.*, 1999).

The mean nucleotide divergence between PMV and the five analysed sequences in this region was 20.6%, (range 12.4% - 29.2%), verifying that the non-coding region is more conserved than the ORFs. A maximum likelihood tree and a distance method tree (Fitch), based on an alignment of the non-coding region, are shown in Figure 4.10. It can be seen that the branching positions of TA278 and TUS01 differ between the two, and the low values obtained on bootstrapping indicate that there was little support for any one topology. However, in both analyses, PMV and SANBAN were closely related, an observation not apparent following the phylogenetic analysis of near full-length sequences (Figure 4.7).

Hijikata *et al.* (1999) proposed that the domain immediately downstream of the polyadenylation signal in the non-coding region of TTV forms a stem–loop structure. The corresponding region in PMV, analysed in this way, predicted a secondary structure that was very similar to other TTV genomes (Figure 4.11). By contrast, the secondary structure of the GC-rich region of the PMV genome is likely to be affected by deletions: Figure 4.12 shows how the GC-rich region of PMV might be different from the GC-rich stretch in sequences TA278 and JA1. Sequences SANBAN, TUS01 and TJN01 also displayed altered folding properties in this region, compared to the prototype TA278 (data not shown).

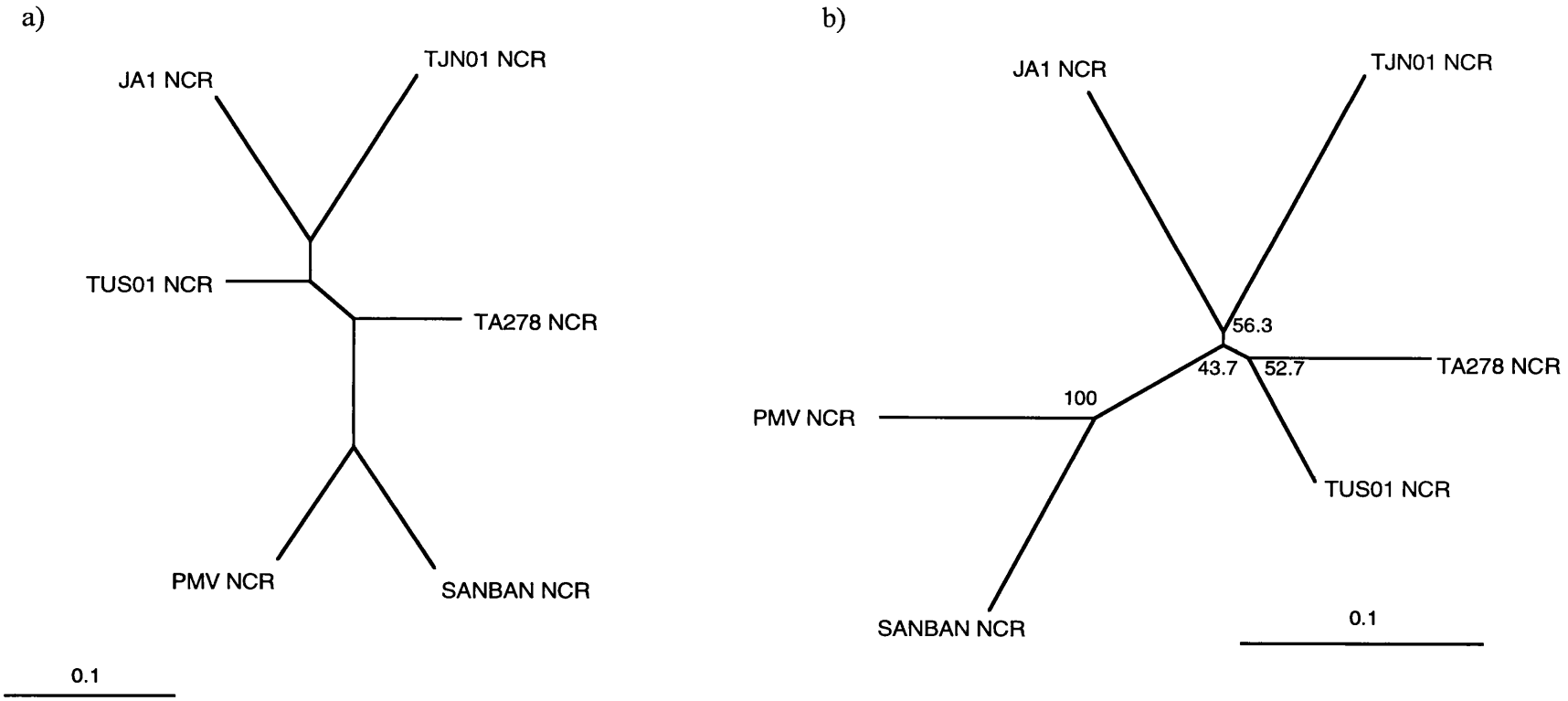
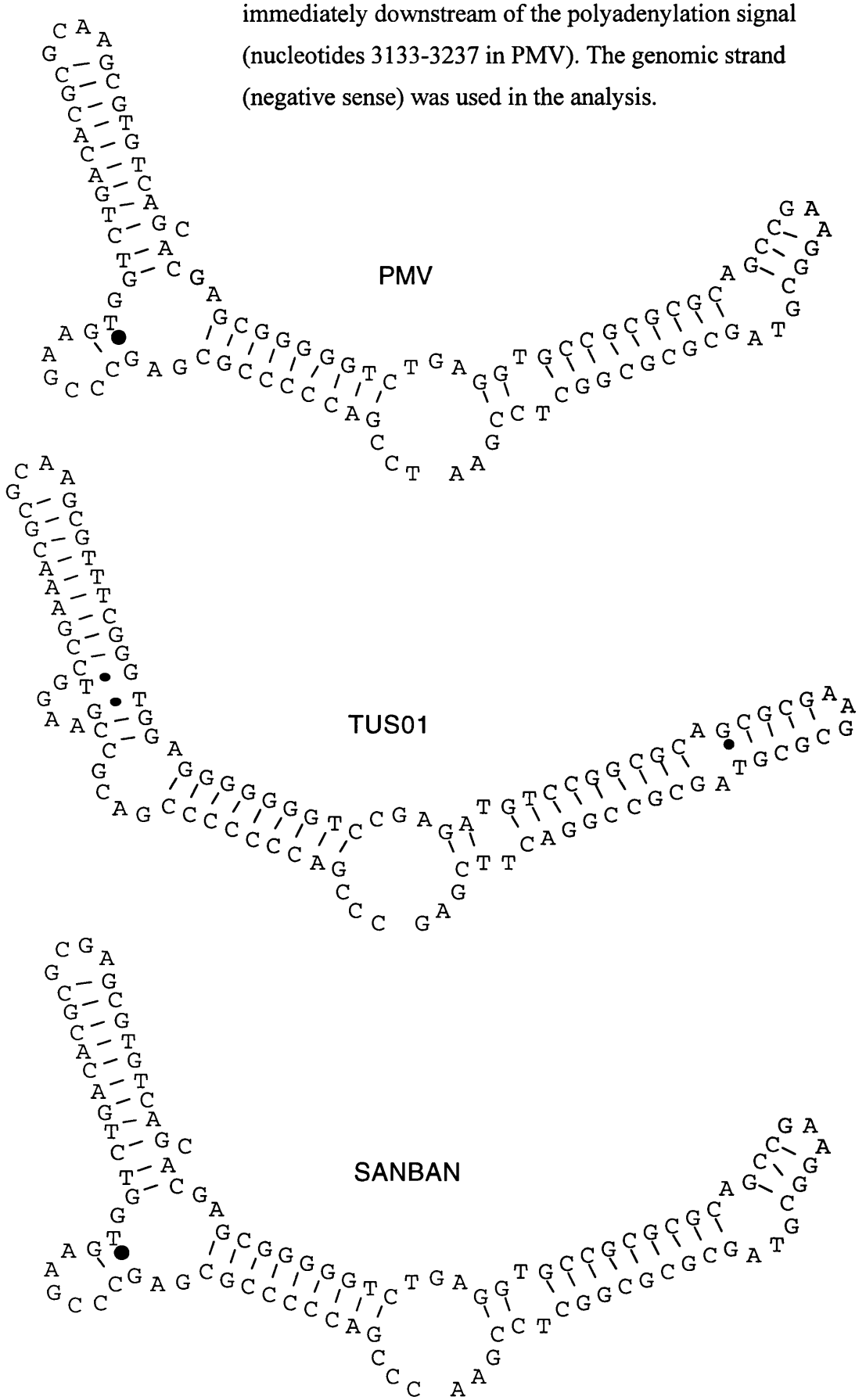


Figure 4.10 Phylogenetic trees produced from an alignment of the non-coding region (NCR) of five TTV sequences and PMV, Using a) Dnaml, a maximum likelihood method, and b) Dnadist and Fitch, a distance method. Tree b) is the consensus tree obtained after 1000 bootstrap replicates, and the percentage of occurrences of each node is indicated. The scale bars denote the proportion of nucleotides substituted for a given branch length.

Figure 4.11 Predicted secondary structure of the region immediately downstream of the polyadenylation signal (nucleotides 3133-3237 in PMV). The genomic strand (negative sense) was used in the analysis.



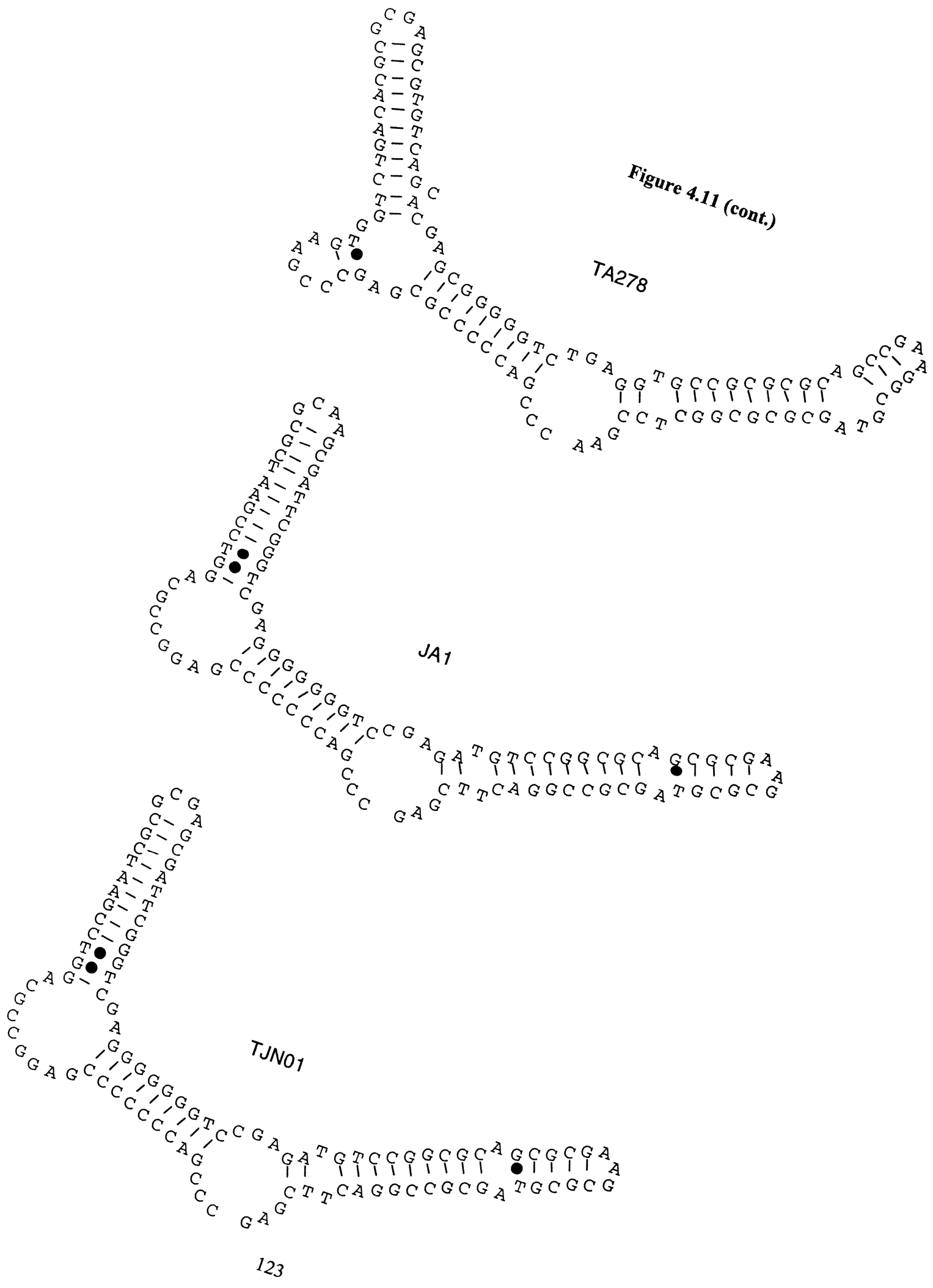
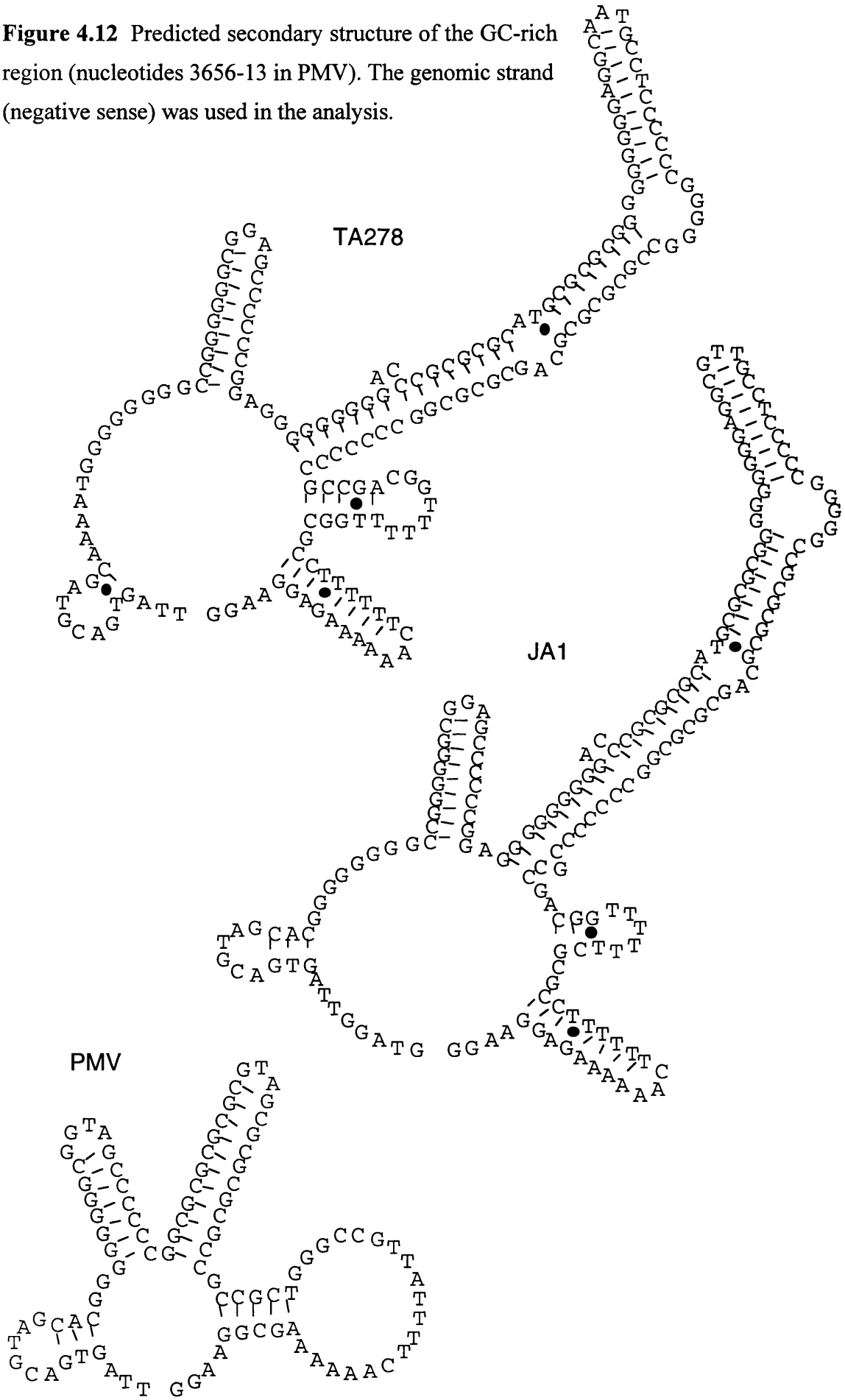


Figure 4.11 (cont.)

Figure 4.12 Predicted secondary structure of the GC-rich region (nucleotides 3656-13 in PMV). The genomic strand (negative sense) was used in the analysis.



4.4 Relationship of PMV and other TTVs to TLMV and CAV

The recent discovery of TLMV has added to the debate regarding the classification of TTV and related viruses. TLMV has a similar genetic organisation to TTV, with a circular single-stranded DNA genome containing three putative ORFs and a conserved non-coding region. The lengths of the nine complete genomes currently deposited in GenBank/EMBL range from 2765-2952 nt, placing them midway between TTV and CAV in size. When these TLMV genomes are subjected to phylogenetic analysis, they display a similar level of diversity to TTV sequences, the mean nucleotide divergence between each being 55.8%. The genomes of CAV isolates are, however, relatively conserved (Meehan *et al.*, 1992).

Little sequence homology is evident between these three virus groups. In their description of TLMV, Takahashi *et al.* (2000*a*) identified an amino acid motif that was present in ORF 2 of representatives of all three: W X_7 H X_3 C X C X_5 H, where X is any amino acid. The motif was sought in PMV, and an amino acid alignment (40 residues) of the ORF 2 fragment encompassing the motif was carried out using two CAV sequences (CAE26P4 and CAU66304), nine TLMV sequences (TLMV-CBD203, CBD231, CBD279, CLC062, CLC138, CLC156, CLC205, NLC023 and NLC026), the seven representative TTVs (Figure 4.7) and PMV. This alignment is shown in Figure 4.13, and demonstrates that the motif was interrupted in PMV, TLMV-CLC156 and TLMV-CLC062.

The 120 nucleotides that coded for this region were aligned, genetic distances were calculated using Dnadist and Fitch analysis was applied in PHYLIP. The resulting phylogenetic tree (Figure 4.14) clearly demonstrates the high level of variability between TTV sequences and between TLMV sequences in this small fragment of ORF 2. By contrast, the CAV sequences were almost identical. Interestingly, in this region the CAV sequences were more closely related to the TTV sequences than to the TLMVs. PMV again demonstrated a close relationship to SANBAN.

Figure 4.13 Amino acid alignment of a conserved motif within ORF-2 of TTV, TLMV and CAV, according to Takahashi *et al.*, (2000*a*). The motif is indicated in red.

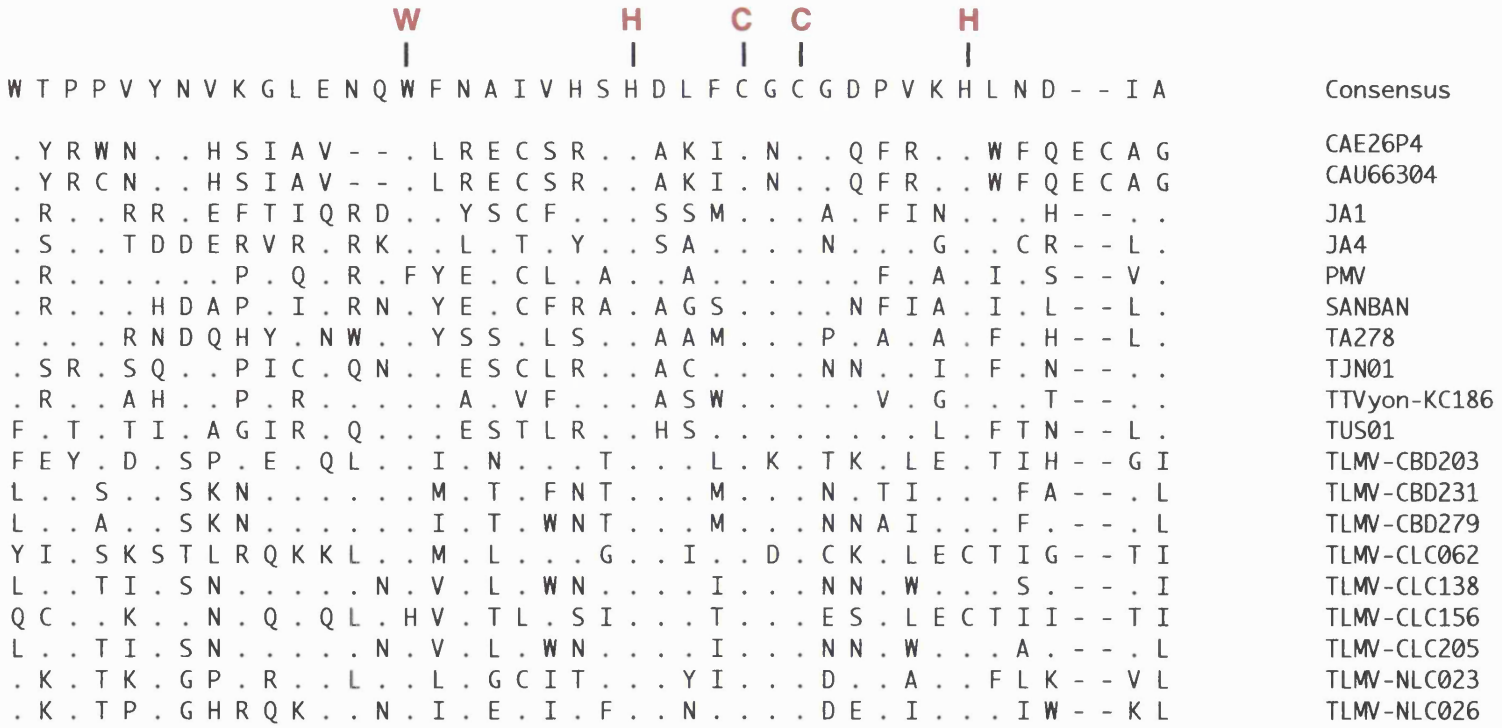
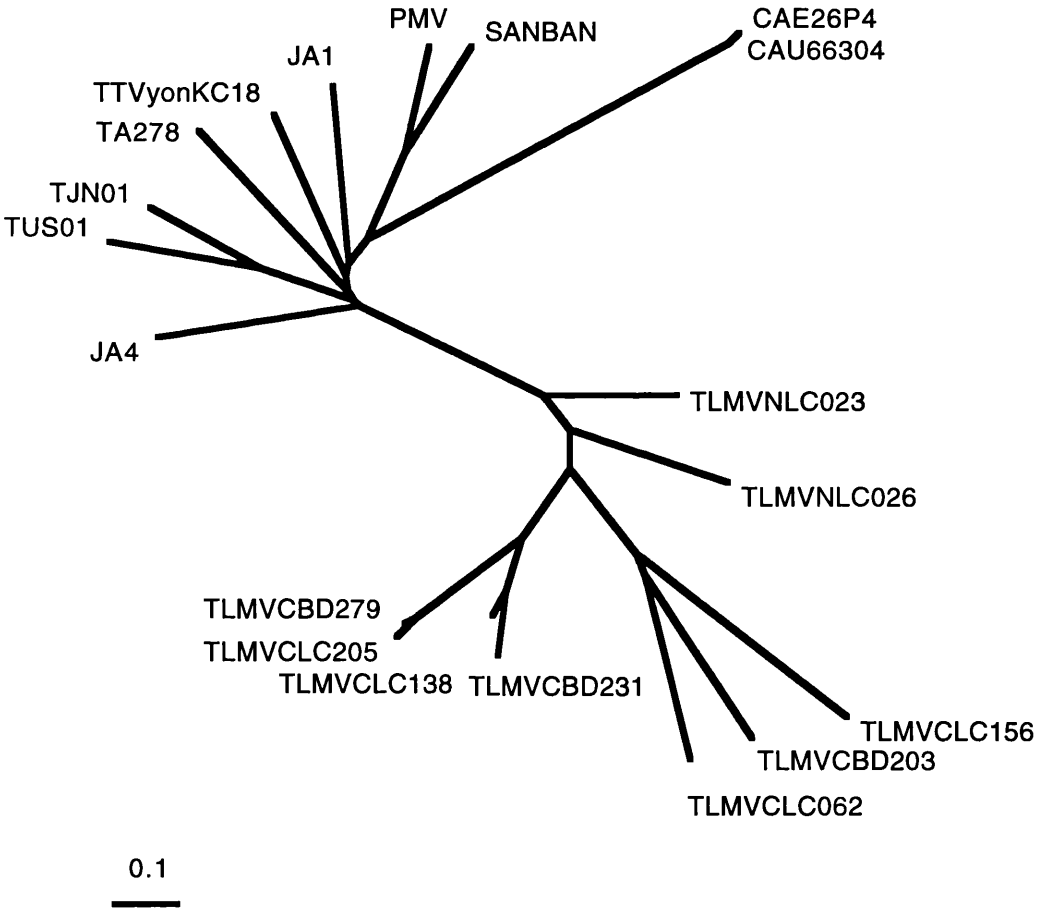


Figure 4.14 A distance method phylogenetic tree (Fitch) produced from an alignment of the 120 nucleotides coding for the conserved ORF-2 amino acid motif. The scale bar indicates the proportion of nucleotides substituted for a given branch length.



4.5 Recombination in PMV

In order to test whether PMV was a recombinant genome, the RIP program was used, as described in section 2.8.3, to identify potential breakpoints in an alignment of PMV with the seven representative TTV sequences. The trimmed alignment that was used to generate the phylogenetic trees in Figure 4.7 and the PlotSimilarity profile in Figure 4.9a was also used as input to RIP. A sliding window size of 200 nucleotides was selected and the resulting plot is shown in Figure 4.15. At various points along the alignment, PMV was most similar to all of the representative sequences, except TTVyon-KC186. However, no strong evidence for the existence of recombinant domains within the PMV genome was apparent.

4.6 Screening for PMV-like sequences

PMV-like sequences were searched for in 197 sera belonging to two patient groups, UK blood donors (n=99) and acute non-A to E hepatitis patients (n=98). A PCR assay with primers situated in the non-coding region was used to amplify TTV DNA as described in Section 2.6.2 (Leary *et al.*, 1999*a*). These 'set B' primers were demonstrated to amplify the expected region of PMV by using cloned PMV DNA as template (fragment B, Figure 2.1). A total of 49 (17 blood donors and 32 hepatitis patients) were positive and of these, 36 gave readable DNA sequence data. The remaining 13 were presumed to contain mixed infections, due to the appearance of multiple double peaks on the sequence data chromatograms.

A CLUSTAL W alignment was performed, using these 36 sequences with the equivalent region of PMV and the 23 TTVs shown in Table 4.1. A neighbour-joining tree was then drawn using MegAlign (Figure 4.16). Although many sequences identified in this study were similar to the other TTVs in this fragment, it was seen that PMV remained phylogenetically distinct, with the most closely related sequences showing around 10% divergence. This suggested that PMV may be a rare TTV type.

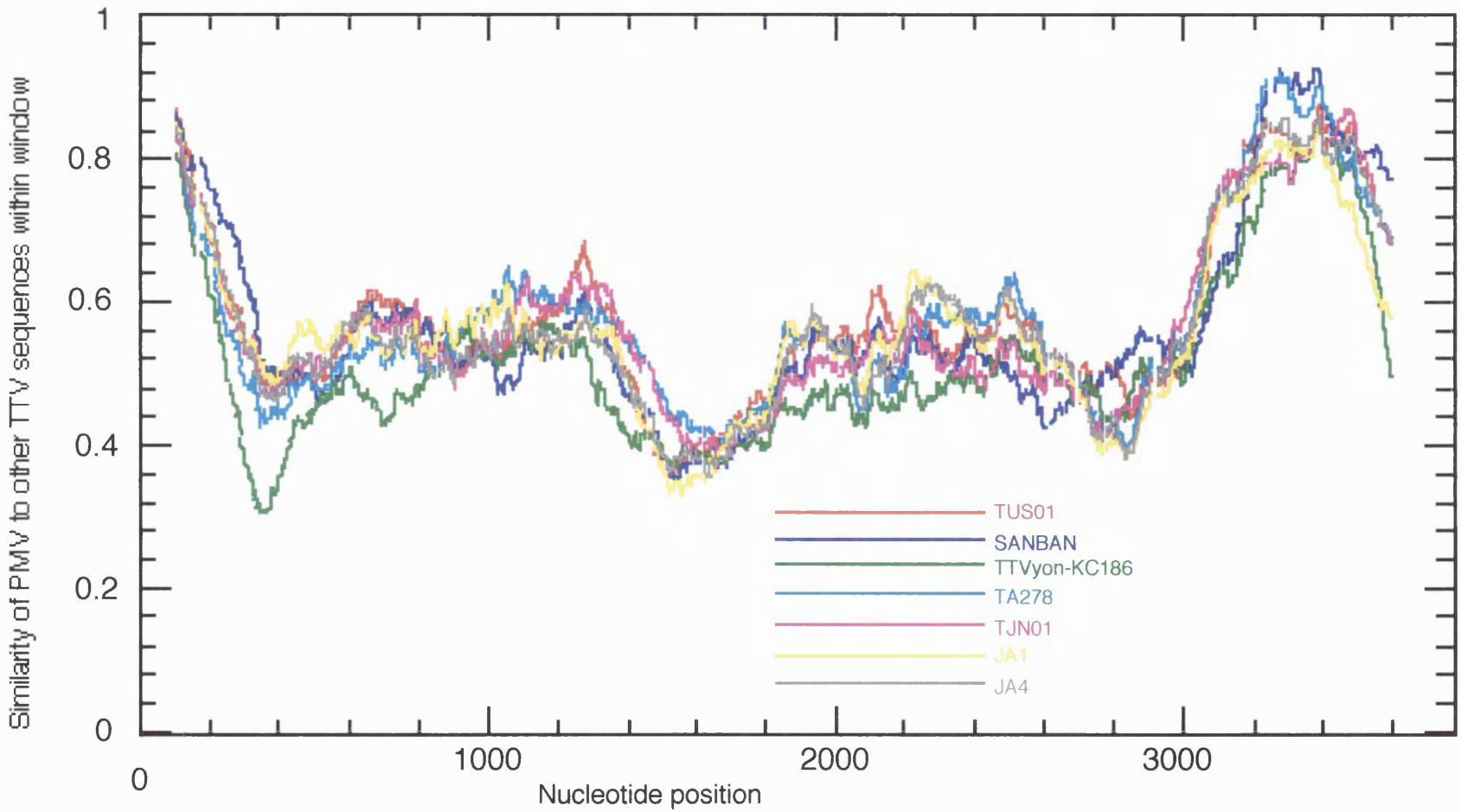
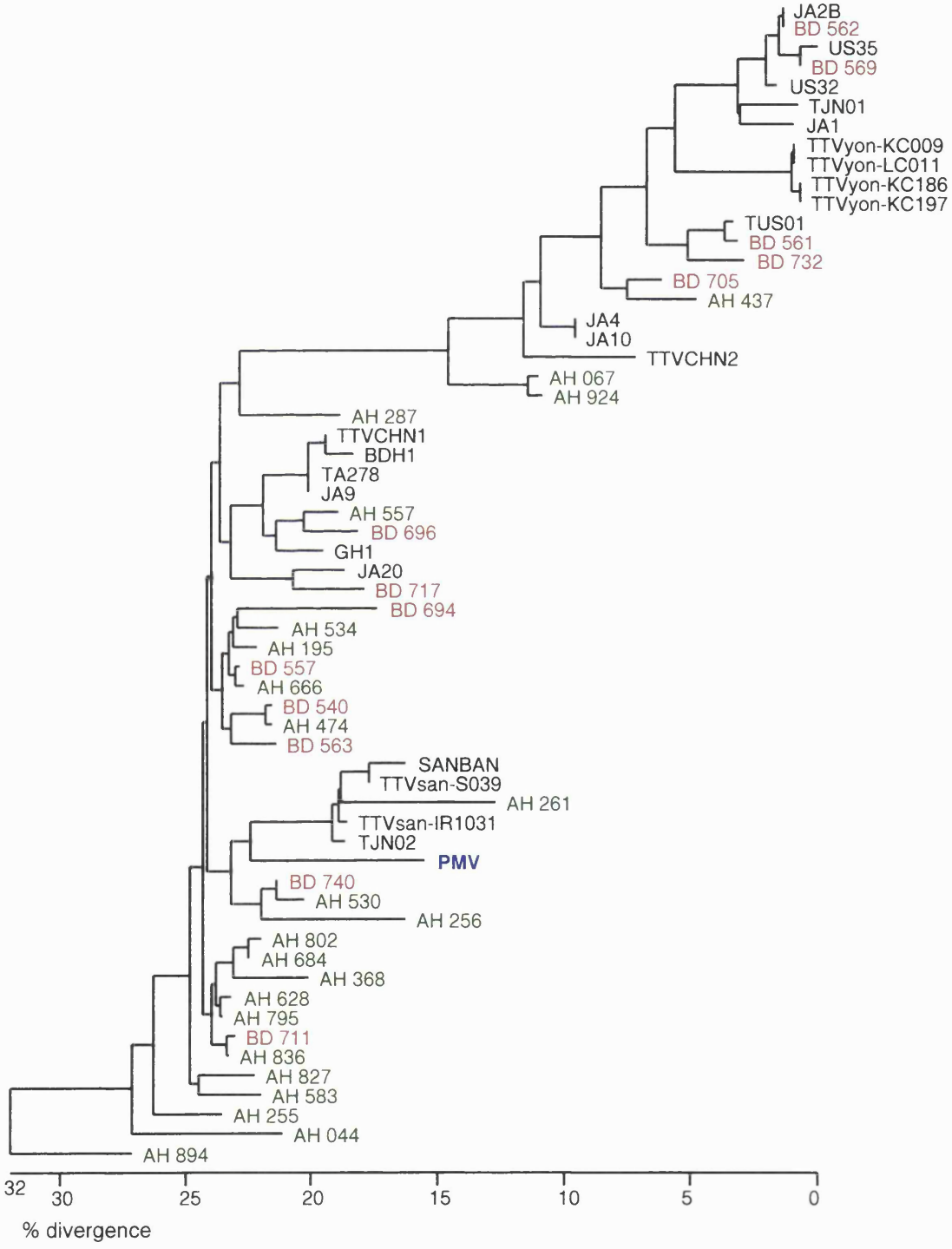


Figure 4.15 Similarity plot output from Recombinant Identification Program (RIP). This shows the similarity of the query sequence, PMV, to each of the seven representative TTV sequences across a nucleotide alignment.

Figure 4.16 Dendrogram produced from an alignment of set B region PCR product sequences. The prefix BD indicates sequences derived from blood donors (shown in red); AH indicates sequences derived from patients with acute hepatitis (shown in green).



4.7 Discussion

TTV DNA sequences were detected in both acute non-A to E hepatitis patients and UK blood donors in this study, and the existence of multiple infections in a single host was confirmed using PCR-cloning and sequencing. Most of the sequences identified using the N22 primers were closely related to those already described. However, a highly divergent fragment was detected and the full genome was then sequenced and characterised.

PMV shared its genome organisation and a moderately conserved non-coding region with other TTV sequences. Its genome also contained the typical motifs identified in most other TTVs, such as a conserved TATA box and polyadenylation signal. It is the shortest TTV-like genome reported so far, the largest deletions being in the GC-rich region. A similarity plot of an alignment of PMV and seven representative TTV sequences showed a high degree of sequence heterogeneity, with genetic distances between them varying from 0.34 to 0.77. Sequence variation was more common in the coding regions, which may lead to major antigenic differences between the viruses.

The protein encoded by the ORF-1 of PMV contained a region that was similar to several sperm protamine protein sequences. These proteins are conserved among different animals, such as cuttlefish, wallabies and quail, where they function in the same way as histones. During the haploid phase of spermatogenesis, they compact sperm DNA into a highly condensed stable complex. This is consistent with the ORF-1 product being a virion capsid protein that enwraps the viral DNA genome. Nishizawa *et al.* (1999) and Erker *et al.* (1999) have previously proposed that the ORF-1 of TTV codes for the viral capsid protein on the basis that its N-terminal arginine-rich region is shared with the circovirus capsid proteins. The functions of the proteins encoded by ORF-2 and ORF-3 are not known. No significant similarity to proteins in the Swissprot database was evident.

The non-coding region was relatively conserved, suggesting that it plays an important regulatory role in viral replication. The effect that the truncated GC-

rich stretch may have on the replication of PMV is yet to be determined. The non-coding region of CAV also contains a GC-rich stretch and several binding sites for transcription factors. A tandem array of four 19 nt repeats, thought to function as an enhancer, was not present in PMV or other full-length TTV sequences. As CAV infects lymphoid cells, causing anaemia, lymphoid depletion and haemorrhages in chicks, sequences that are recognised by lymphoid-specific proteins can be identified in the non-coding region of CAV. For example, the sequence GGGGATTCCCC is recognised by NF-κB, and motif CTATTC is thought to be involved in globin gene expression (Meehan *et al.*, 1992; Plumb *et al.*, 1989). These were not present in any TTV sequences.

The binding sites of liver-enriched transcription factors were also sought, such as hepatocyte nuclear factors 1,3 and 4, and the CCAAT/enhancer-binding protein alpha (Tronche *et al.*, 1997; Overdier *et al.*, 1994; Rastegar *et al.*, 2000). These sequences are abundantly present in the HBV genome. However, none were found in the non-coding regions of the full length TTVs or PMV, apart from one CCAAT motif in the non-coding region of SANBAN.

The circoviruses PCV and BFDV have a stem-loop structure with a conserved nonanucleotide motif at the tip, within the non-coding region. This is found in plant nanoviruses (recently re-classified from plant circoviruses), such as coconut foliar decay virus (CFDV) and banana bunchy top virus (BBTV) and in the related geminiviruses, which are also plant pathogens containing a small, circular single-stranded genome (Bassami *et al.*, 1998; Niagro *et al.*, 1998). It functions as a *cis*-acting element required for rolling-circle replication (RCR). CAV, however, does not have this motif and the mechanism for CAV replication is not fully understood (Noteborn and Koch, 1995). The amino acid sequence of CAV ORF-1 possesses three motifs found in the replication associated (Rep) proteins of the other circoviruses (Niagro *et al.*, 1998). It is also known that, in culture, CAV-infected cells generate double stranded replicative DNA forms, although Todd *et al.* (1996) found no evidence for rolling-circle replication in CAV, using plasmid constructs containing CAV DNA.

TTV sequences so far described, including PMV, do not contain the nonanucleotide motif, nor do they possess an ORF that resembles the Rep proteins encoded by PCV and BFDV. Mushahwar *et al.* (1999) reported that two TTV sequences (TA278 and GH1) each contained two of the three amino acid motifs found in CAV ORF-1, which are associated with RCR (FTL and YuxK, where u is a bulky, hydrophobic residue). These were also found in PMV ORF-1. Moreover, double-stranded forms of TTV have been detected in hepatocytes and bone marrow cells (Okamoto *et al.*, 2000*b* and *c*). This suggests that their method of replication may be more like that of CAV then the other circoviruses.

Twenty-three full-length and near full-length genomes of TTV-like viruses fell into eight major groups on phylogenetic analysis, with a mean genetic distance of 0.61 between them (Figure 4.6). The branching order observed on analysis of PMV with seven representative trimmed nucleotide sequences (Figure 4.7) was also seen when their ORF-1 amino acid sequences were aligned. However, the branching patterns observed on phylogenetic analysis of the ORF 2 and putative ORF 3 amino acid sequences differed in the relative positions of TUS01, TJN01, SANBAN, TTVyon-KC186 and PMV (data not shown). When different models of DNA substitution were applied during phylogenetic analysis, it was found that branch lengths showed some variation, although tree topology did not change. It has been suggested that, as sequence divergence between TTV types is so extreme, current methods for estimating evolutionary distance are inadequate and basic, uncorrected *p*-distances should be used (where *p* is the proportion of different nucleotides between 2 sequences) (Prescott *et al.*, 1999).

The eight TTV groups corresponded to the previously defined genotypes 1, 2 and 3 (Erker *et al.*, 1999), with TUS01, TJN01, the SANBAN clade, the YONBAN clade and PMV forming five distinct branches. Following the nomenclature of Erker *et al.* (1999), these five could be considered new 'genotypes'. However, the level of genetic divergence within the TTV cluster was higher than would be expected if all these sequences were from viruses of the same species.

Historically, virus species have been typed according to differences in serological reactivity. With the use of DNA sequencing and phylogenetic

analysis it is now increasingly common to segregate a virus species into genotypes. However, there is no standard definition of the level of divergence required to separate genotypes. Using other DNA viruses as examples, the information in Table 4.2 demonstrates that the way in which viruses are genetically typed varies considerably.

Table 4.2 Examples of the criteria used to type DNA viruses

Virus	Number of types	Criteria for typing	Reference
JC Virus (JCV)	>7 genotypes	Nucleotide divergence of 1.2% to 2.3% between each, on phylogenetic analysis of coding region sequences.	Jobes *et al.* (1998) Agostini *et al.* (1998)
Human papillomavirus (HPV)	> 80 types	Nucleotide identity of <90% between L1 ORF sequences.	Delius *et al.* (1998)
Adeno-associated virus (AAV)	6 serotypes	Nucleotide identity of 82% or less across whole genome.	Bantel-Schaal *et al.* (1999)
Hepatitis B virus (HBV)	6/7 genotypes	Nucleotide divergence of >8% across whole genome.	Okamoto *et al.* (1988) Norder *et al.* (1994)

For circoviruses, although there is very little sequence similarity between the three (CAV, PCV and BFDV), individual isolates of each virus are moderately conserved. For example, the maximum nucleotide divergence shown between six full-length CAV genomes, available in GenBank/EMBL, was 5% (not shown). This may reflect a more ancient origin of TTV.

The analysis presented here, using full and near full-length genomes and amino acid sequences, confirms that until the full extent of divergence within the TTVs is known, it is inappropriate to determine genotypes based on short TTV-like sequences.

Short stretches of sequence similarity between TTV and CAV have been reported (Miyata *et al.*, 1999; Hijikata *et al.*, 1999). Furthermore, TLMV was recently discovered in human serum and described as an intermediate relative of TTV and CAV (Takahashi *et al.*, 2000*a*). In order to further assess the relationships between these viruses, the nucleotide sequences that code for a semi-conserved amino acid motif in ORF-2 of all three virus groups were compared. Phylogenetic analysis demonstrated that in this short fragment, CAV was more related to TTV sequences than to TLMV sequences. The PMV fragment grouped with SANBAN, although the eight TTVs maintained distinct branching positions. The nine TLMV sequences also showed extensive heterogeneity and formed seven main branches. By contrast, the CAV sequences were nearly identical.

It is questionable whether, in the absence of 'ancestor' sequences for comparison, the phylogenetic methods currently available are sufficiently complex to accurately infer evolutionary relationships between the virus groups studied in this chapter. The genetic distances between TTV types are very large over much of the genome, and regions of potential homology between TTV, TLMV and CAV are short. The simple corrections used for multiple hits are likely to be inadequate. It is also becoming increasingly clear that viruses evolve in a way which is very unlike their hosts, due to the unique constraints of single-stranded genomes, overlapping ORFs and vast population sizes (Simmonds, 2001). Therefore, phylogenetic methods developed to study eukaryotic evolution may need substantial revision before they can be meaningfully applied to viral sequences.

Despite this, phylogenetic analyses remain useful for assessing relatedness between groups of sequences and thus it is still appropriate to use them for taxonomic purposes. Both the analysis presented in Figure 4.14 and the near-full length alignment of the eight representative TTVs suggest that they should have equal taxonomic rank with each other and with the gyrovirus genus of *Circoviridae*. It is proposed that TTV be classified as a new virus family, and that the eight representative TTV sequences are each assigned a different genus.

How is the wide diversity, observed between TTV genomes, achieved? Studies of long term infection indicate that the mutation rate in chronically infected individuals is low (Biagini *et al.*, 1999; Ball *et al.*, 1999). This is supported by the sequence analysis of a persistently infected patient's TTV sequence in this study, which remained very stable over a 13 month period. Low mutation rates are consistent with the assumption that TTV uses a host cell polymerase during replication.

A plausible explanation for the extensive heterogeneity between TTV types has been that they are derived from an ancient virus lineage that has infected and co-evolved with humans over many millions of years (Prescott *et al.*, 1999). This also provides a possible answer to the question regarding its apparent lack of pathogenicity; as the parasite-host relationship has evolved, the virus has become so adapted to the host as to persist with minimum harm, thereby increasing its chances of survival and propagation.

Homologous recombination between widely divergent types may also be a mechanism by which TTV heterogeneity is created and maintained (Worobey, 2000). This is supported by the finding that an individual can be infected with up to seven different genetic types at any one time (Ball *et al.*, 1999; Niel *et al.*, 2000). Sequences that change positions in trees based on different subgenomic regions may indicate the existence of recombinant genomes. PMV was revealed to have a relatively close relationship to SANBAN, both in the tree produced from the non-coding region alignment (Figure 4.10) and from the ORF 2 motif alignment (Figure 4.14), in contrast to their apparent distance in the tree produced from a near full-length alignment (Figure 4.7). The low bootstrap values obtained for the non-coding region tree may also be indicative of conflicting phylogenetic signal in this region, which may be due to the presence of recombinant sequences.

The potential mosaic structure of PMV was further explored using the program RIP. The resulting plot did not give a clear indication that recombination with any of the seven representative TTVs had occurred. Additional analysis would be required to substantiate the evidence for recombination in PMV. If repeated

recombination is an evolutionary feature of the TTVs, methods such as RIP, which recognise stretches of nucleotides from a different lineage to others in the same sequence, will not be effective at detecting it (Maynard Smith, 1999). The possibility of frequent recombination occurring in the geminiviruses was investigated using the homoplasy test, which is a more appropriate method for the analysis of potential multiple recombination events (Padidam *et al.*, 1999; Maynard Smith and Smith, 1998). However, it has been suggested that the method is only useful if intra-group nucleotide divergences are < 5%. It may be helpful to wait until the full extent of TTV diversity is known before applying further recombination analysis methods.

The existence of extensive recombination would help to explain why TTV types do not appear to be geographically segregated in the same way as, for example, HCV and HIV types, where opportunities for mixed infections to occur are restricted by the requirement for them to be transmitted by bloodborne and sexual routes. Thus, for HCV and HIV, few population groups are multiply exposed. In contrast, many transmission routes probably exist for TTV, as indicated by the detection of TTV DNA in several body fluids (Okamoto *et al.*, 1998*b*; Toyoda *et al.*, 1999; Inami *et al.*, 2000; Ross *et al.*, 1999) and it may be transmitted transplacentally (Morrica *et al.*, 2000). Thus, studies of the age-distribution of TTV prevalence in endemic areas have suggested that it is acquired early in life (Yamada-Osaki *et al.*, 1998; Davidson *et al.*, 1999), and multiple infections are known to be common (Takayama *et al.*, 1999; Niel *et al.*, 2000; Ball *et al.*, 1999).

The ability of the two sets of primers used to detect TTV DNA differed. During the initial detection of TTV sequences, the primers used were directed against the variable N22 region within ORF-1 (Okamoto *et al.*, 1998*a*). They detected 6 TTV infections in 43 acute non-A to E hepatitis patients from the UK (14%). When 'set B' primers (Leary *et al.*, 1999*a*), which amplify a fragment of the more conserved non-coding region, were used to screen 98 similar patients, they were able to detect 32 positive serum samples (32.7%). This discordance in detection rates by different primer sets has also been noted in several other studies, and it has been suggested that whilst the full extent of divergence between TTV

sequences is unknown, multiple sets of primers should be used to estimate the prevalence of TTV infection (Desai *et al.*, 1999; Leary *et al.*, 1999*b*).

The presence of TTV in a very large proportion of healthy individuals worldwide suggests that, at least in the majority of cases, it is not causing disease and that the concept of a normal viral flora should now be considered (Simmonds *et al.*, 1999). The detection of TLMV, another potential family of benign viruses, strengthens this notion. The possibility that non-pathogenic circovirus-like viruses may be found in other species was predicted by Tischer *et al.* (1995) when they discovered antibodies that cross-reacted to PCV antigen in human, mice and cattle sera. Furthermore, TTV-like DNA sequences have been discovered in domestic farm animals and several primate species (Leary *et al.*, 1999*a*; Verschoor *et al.*, 1999; Okamoto *et al.*, 2000*a*; Abe *et al.*, 2000; Romeo *et al.*, 2000). In at least 2 species (chimpanzees and macaques), some of these sequences appear to be genetically distinct from those circulating in humans. Others, however, are highly related to human TTV sequences, raising the possibility that extensive cross-species infection occurs. In contrast, all three animal circoviruses and the related plant nanoviruses have restricted host ranges.

It is perhaps too soon to conclude that TT viruses have no pathogenic effect. Up to 10^4 copies of TTV can circulate in the blood of persistently infected individuals. Also, viral DNA is present in many extracellular body fluids, liver tissue and bone marrow cells. It is thought that the human host recognises it as foreign, as indirect immunoprecipitation experiments have indicated that an antibody response can be elicited, although this apparently does not clear infection, or prevent reinfection with many different TTV types (Tsuda *et al.*, 1999).

Is PMV, in particular, pathogenic? PMV may be hepatotropic, as its DNA sequence was also detected in liver. However, since a pre-hepatitis serum specimen from the patient was not available, it was impossible to determine if PMV carriage occurred from the onset of disease. It is noted that viruses showing less sequence divergence than the TTV group can cause different clinical syndromes. For example, a strain of PCV, reported to be associated with

a disease in pigs known as postweaning multisystemic wasting syndrome, shows only 26% divergence from the genome of the prototype PCV, which is common in domestic herds and apparently does not cause disease (Morozov *et al.*, 1998). Similarly, JC virus is an established causative agent of progressive multifocal leukoencephalopathy (Padgett and Walker, 1973), while another polyoma virus, BK, which shows around 30% nucleotide divergence from JC, is associated with haemorrhagic cystitis (Hashida *et al.*, 1976; Gluck *et al.*, 1994). Therefore, the pathogenic capacity of PMV cannot be totally ruled out.

Further study is required to unravel the precise evolutionary and taxonomic connections between the circular single stranded genomes examined here. It is clear that current phylogenetic techniques may not adequately depict the true relationship between them, as models of DNA substitution are too simplistic to cope with such extensive genetic heterogeneity, and the possibility of recombination cannot easily be accommodated. While PMV may be unique genetically and might be hepatotropic, it appears to be a rare virus in humans. The emerging pattern of infection with TTV-like viruses is that of a collection of related but different viruses, with varying prevalences. Their pathogenic potential has yet to be elucidated.

Chapter 5

Conclusions and Further Work

The use of appropriate genetic methodologies in the study of viral variants can dramatically increase the amount of information gained regarding their epidemiology and relationships to other similar viruses. Several procedures were developed and applied during the course of this investigation, including standard PCR and gel-based methodologies, and computer-based analysis programs, in order to examine the epidemiology and evolutionary history of three significant viral variants.

5.1 HBV^{PV} and HBV^{AV}

In the initial work on HBV, a gel-based screening assay was chosen in preference to DNA sequencing. As the question asked was "are these PCR products the same or different?", sequencing was considered both too slow and unnecessary. DGGE was applied to the study of two defined variants in different contexts. It was thus possible to track a variant common to the UK, HBV^{PV}, over time and describe its increasing prevalence in one location. DGGE was also applied to establish which patients were and were not infected in a point source outbreak, and this led to the conclusion that the variant involved, HBV^{AV}, was not prevalent in the UK. The rapid screening of large numbers of samples in the study of both variants allowed their contrasting epidemiologies to be described.

Despite the widespread use of PCR fragment analysis for the investigation of clusters of HBV infection, there are few descriptions in the literature of genetic studies of defined HBV variants circulating in a population. The closest example is the study of Flodgren *et al.* (2000). Here, the pre-S2 region was sequenced from 30 epidemiologically unlinked patients with acute HBV infection in a Russian city. Twelve samples were found to have identical sequences in this fragment, and a further 8 shared a second identical sequence. The analysis was not extended to determine how the predominantly circulating variants were transmitted over time.

This study further demonstrated that the relative genetic stability of the HBV genome during continual transmission amongst high-risk populations can lead to the situation where one variant becomes dominant. Using DGGE, it was possible to observe the increase in homogeneity of PCR product sequences yielded by the

infecting HBV genomes in the Hull community, and to correlate this with a previous outbreak in a local prison. It is evident that attempts to identify specific transmission events of HBV^{PV} using sequence-based analysis of small PCR products would not be possible, due to the confounding effect of high HBV^{PV} prevalence.

Further work: In the light of the very high prevalence of HBV^{PV}, how can a basis be formulated for assessing its role in future transmission incidents? It seems clear that it must continue to rest on epidemiological data. Thus, if two people with no known connection are determined to be acutely infected with HBV^{PV}, the probability that their infections have arisen by chance can be estimated only if a large study of HBV^{PV} prevalence is undertaken in the vicinity of the potential transmission. The use of a high-throughput assay such as DGGE would allow rapid screening of sera from unlinked acute hepatitis cases to be carried out, thereby enabling such a statistical analysis.

One possible bias in estimating HBV^{PV} prevalence when using selected study groups arises from its potential to cause symptomatic infection. All samples tested from Hull and south east England came from acutely infected patients with specific symptoms, such as jaundice and passage of dark urine, which warranted testing for hepatitis. As approximately two thirds of infected adults do not experience such symptoms, the prevalence of HBV^{PV} in these cases remains unknown. To circumvent this possible source of bias, it would be useful to screen a number of asymptomatic HBsAg-positive specimens to determine if they are also carrying HBV^{PV}.

5.2 PMV

The third variant identified in this study, PMV, was a novel and divergent type of TTV. It was fortuitously amplified using prototype primers directed towards the variable N22 region of ORF-1 (Okamoto *et al.*, 1998*a*). For PMV, unlike the situation with the HBV variants, PCR and sequencing were the only feasible methods to use in the study of its epidemiology and relationship with other similar viruses, due to the high level of genetic heterogeneity amongst the TTVs

so far described. Even a PCR fragment from the relatively conserved non-coding region showed such variability that a close relative of PMV might have differed by several nucleotides, and would therefore appear different using a DGGE or SSCP assay. Using the PCR-sequencing approach, acute non-A to E hepatitis patients and UK blood donors were screened for PMV, but no identical or closely related sequences were found, leading to the conclusion that PMV rarely circulates in the UK.

In order to make full comparisons with the few full-length or near full-length TTV sequences available in the GenBank/EMBL databases, the sequence of the entire PMV genome was determined. It was found to share many features with other TTVs, such as the positions of putative ORFs, a polyadenylation signal and a relatively conserved non-coding region.

Two procedures based on multiple alignment scanning with a sliding window were used to plot nucleotide and amino acid similarity between PMV and representative TTVs, and to assess the possibility of recombination amongst them. The first identified regions of conservation and hypervariability, and demonstrated that, surprisingly, the putative coding regions were the most heterogeneous between types. The second showed no clear evidence for recombination between PMV and any other genome. However, the phylogenetic analysis of subgenomic regions suggested that certain domains of the PMV genome were quite closely related to homologous domains of the SANBAN genome, indicating that these domains may share a relatively recent common ancestry.

The extreme variability of the TTV genomes may confound attempts to unravel the true relationships between them, particularly if multiple recombination events are a common feature of TTV evolution. The use of parsimony was considered inappropriate for phylogenetic analysis due to the probable presence of a high level of homoplasy. Distance methods and maximum likelihood did not always result in a similar tree, as demonstrated by the NCR dendrograms (Figure 4.10). When sequences are so different, it is questionable whether current models of DNA substitution are sufficiently complex to reconstruct an accurate

evolutionary tree. However, phylogenetic analysis remains useful in this situation to group similar sequences. For example, the 23 full-length and near full-length TTV sequences analysed (including PMV) fell into 8 major clades (Figure 4.6). The use of newer techniques such as split decomposition may be advocated for the analysis of the TTVs, where conflicting phylogenetic signals in an alignment are interpreted as a network, thus displaying alternative possible relationships between the sequences.

The methods used in this study enabled the genomic characteristics of what may be a new species of TTV to be evaluated, and will be a useful starting point for the further examination of this newly discovered virus family. The final classification of TTV types awaits a decision by the International Committee on Taxonomy of Viruses (ICTV). It seems clear that they possess a level of divergence beyond that of genotypes, and that further TTV types may yet be discovered.

Further work: The potential hepatopathic role of PMV could not be confirmed by this work. In order to implicate PMV as an agent of liver disease, it would be necessary to correlate infection with subsequent liver injury, and to demonstrate replication in the liver. Further screening, using either PCR and sequencing of a conserved region of the genome, or PCR using PMV-specific primers, could be employed to search for more examples of this virus in hepatitis patients and compare its prevalence with that in healthy controls. To this end, it would also be worthwhile searching for PMV in different blood compartments or even different body fluids. Recent studies by Okamoto *et al.* (1999*a* and 2000*d*) showed that, in the same individuals, the peripheral blood mononuclear cells contained genetic types of TTV different to those in the plasma.

Several questions remain as to the origin of the TTVs and the mechanisms by which they have developed such extensive genomic variability. Also, the presence of TTV-like sequences in domestic farm animals, and in a number of nonhuman primate species requires further investigation. Many of the sequences detected are similar to those already found in human hosts, leading to the conclusion that TTV has an extended host range and that cross-species

transmission may be common. However, a small number of TTV types appear at present to be species specific. A more extensive study of the TTVs of other nonhuman species is required in order to further our understanding of the origin and evolutionary history of this intriguing group of viruses.

5.3 Ethical considerations

The ethical issues surrounding the use of archived specimens for research, teaching and quality control have been debated by several professional clinical bodies in recent years, particularly those concerning patient consent and confidentiality. Difficulties in this area have arisen when trying to strike a reasonable balance between what is in the best interests of individual patients, whose tissues are used, and what is in the best public interest in terms of health benefits accruing from medical research.

Archived samples that are retained after surgery, or after diagnostic tests have been carried out, can be a very valuable resource for new research projects. However, consent would not have been given when samples were taken, for research that had not even been considered. Contacting patients months, or more usually years, after samples are collected to seek consent is very difficult or impossible.

To ensure that existing archives are not wasted most professional bodies have reached a consensus, which is that the use of anonymised left-over tissue for research is ethically acceptable, and that consent from the original donors is not required (Royal College of Pathologists, RCP, 1999; Medical Research Council, MRC, 2000). In 1996, the Royal College of Pathologists guidelines also stated that research on anonymised left-over tissue would not need to be submitted to a research ethics committee. However, this is now considered unacceptable (Ashcroft, 2000; MRC, 2000). The trend is towards full anonymisation. This ensures patient confidentiality and exempts research projects from the Data Protection Act of 1998 (General Medical Council, GMC, 2000; Furness, 2001). If patient data remains associated with the specimens, a decision by a Local Research Ethics Committee (LREC) would be required, which would consider

whether or not the likely benefits of the research outweigh the undesirable outcome of loss of complete confidentiality and the lack of specific consent.

It has also been suggested that new archives should be established under much more stringent conditions than in the past (Ashcroft, 2000). In order to ensure that left-over tissues can still be made available for research The Medical Research Council recently proposed that a two-stage consent form be used, so that patients can be made aware that their tissues may be used for currently unspecified research and be given the option to refuse to this without affecting their treatment (MRC, 2000; RCP, 1999). It is then recommended that samples are anonymised and unlinked whenever possible, or at least coded when it is essential to keep patient data linked to the samples. All research projects will be subject to LREC approval. However, these guidelines do not allow for the possibility that the findings of future research may have implications for individual donor's health or treatment, in which case it would be useful to be able to feedback information to the donor's clinicians. This would clearly be impossible if samples were irreversibly anonymised and unlinked. It may be desirable in certain cases for donors to have access to the results of future research projects that have involved their samples. Also, the value of certain research projects may be diminished if there is no possibility to collect follow-up samples. These issues have yet to be resolved.

With regard to this thesis, the specimens used in the study of HBV variants described in Chapter 3 were anonymised, archived samples. Although no application was made to a research ethics committee to carry out the DGGE screening, and no consent was sought, it was considered that the specimens were studied within the framework of recent guidelines, as patient confidentiality was protected and the research could not have provided information that would have had clinical implications for the patients.

The research with TTV described in Chapter 4 was more speculative in that a new virus was tested for in an attempt to discover the cause of cryptic acute hepatitis. Certain details about the medical history of the patients were retained and some samples were identified by the patients initials. The recent guidelines

indicate that such practices will no longer be acceptable. This will be particularly important for future virological research when new disease-causing agents are sought. It is clear that if and when the research conducted could have an impact on an individual's clinical care, such as with the discovery of a new virus, informed consent and LREC approval must be obtained before projects can begin.

References

Abe, K., Inami, T., Asano, K., Miyoshi, C., Masaki, N., Hayashi, S., Ishikawa, K.-I., Takebe, Y., Win, K.M., El-Zayadi, A.R., Han, K.-H. and Zhang, D.Y. (1999) TT virus is widespread in the general populations from different geographic regions. *Journal of Clinical Microbiology* **37**, 2703-2705.

Abe, K., Inami, T., Ishikawa, K., Nakamura, S. and Goto, S. (2000) TT virus infection in nonhuman primates and characterization of the viral genome: identification of simian TT virus isolates. *Journal Of Virology* **74**, 1549-1533.

Agostini, H.T., Shishido-Hara, Y., Baumhefner, R.W., Singer, E.J., Ryschkewitsch, C.F. and Stoner, G.L. (1998) JC virus type 2: definition of subtypes based on DNA sequence analysis of ten complete genomes. *Journal of General Virology* **79**, 1143-1151.

Alter, M. (1996) Epidemiology and disease burden of hepatitis B and C. *Antiviral Therapy* **1**, 9-15.

Alter, M.J., Gallagher, M., Morris, T.T., Moyer, L.A., Meeks, E.L., Krawczynski, K., Kim, J.P. and Margolis, H.S. (1997) Acute non-A-E hepatitis in the United States and the role of hepatitis G virus infection. *New England Journal of Medicine* **336**, 741-746.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**, 3389-3402.

Andre, F.E. (1989) Summary of safety and efficacy data on a yeast-derived hepatitis B vaccine. *American Journal of Medicine* **87**, 14S-20S.

Arauz-Ruiz, P., Norder, H., Visona, K.A. and Magnius, L.O. (1997) Molecular epidemiology of hepatitis B virus in Central America reflected in the genetic variability of the small S gene. *Journal of Infectious Diseases* **176**, 851-858.

Arens, M. (1999) Methods for subtyping and molecular comparison of human viral genomes. *Clinical Microbiology Reviews* **12**, 612-626.

Ashcroft, R. (2000) The ethics of reusing archived tissue for research. *Neuropathology and Applied Neurobiology* **26**, 408-411.

Back, N.K., Nijhuis, M., Keulen, W., Boucher, C.A., Oude Essink, B.O., van Kuilenburg, A.B., van Gennip, A.H. and Berkhout, B. (1996) Reduced replication of 3TC-resistant HIV-1 variants in primary cells due to a processivity defect of the reverse transcriptase enzyme. *EMBO Journal* **15**, 4040-4049.

Balayan, M.S., Andjaparidze, A.G., Savinskaya, S.S., Ketiladze, E.S., Braginsky, D.M., Savinov, A.P. and Poleschuk, V.F. (1983) Evidence for a virus in non-A, non-B hepatitis transmitted via the fecal-oral route. *Intervirology* **20**, 23-31.

Ball, J.K., Curran, R., Berridge, S., Grabowska, A.M., Jameson, C.L., Thomson, B.J., Irving, W.L. and Sharp, P.M. (1999) TT virus sequence heterogeneity *in vivo*: evidence for co-infection with multiple genetic types. *Journal of General Virology* **80**, 1759-1768.

Ballard, A.L. and Boxall, E.H. (1999) Assessing the infectivity of hepatitis B carriers. *Communicable Disease and Public Health* **2**, 178-183.

Balogun, M.A., Ramsay, M.E., Fairley, C.K., Collins, M. and Heptonstall, J. (1999) Acute hepatitis B infection in England and Wales: 1985-96. *Epidemiology and Infection* **122**, 125-131.

Bancroft, W.H., Mundon, F.K. and Russell, P.K. (1972) Detection of additional antigenic determinants of hepatitis B antigen. *Journal of Immunology* **109**, 842-848.

Bandelt, H.J. and Dress, A.W. (1992) Split decomposition: a new and useful approach to phylogenetic analysis of distance data. *Molecular Phylogenetics and Evolution* **1**, 242-252.

Bantel-Schaal, U., Delius, H., Schmidt, R. and zur Hausen, H. (1999) Human adeno-associated virus type 5 is only distantly related to other known primate helper-dependent parvoviruses. *Journal Of Virology* **73**, 939-947.

Barlow, K.L., Green, J. and Clewley, J.P. (2000) Viral genome characterisation by the heteroduplex mobility and heteroduplex tracking assays. *Reviews in Medical Virology* **10**, 321-335.

Bartholomeusz, A., Schinazi, R.F. and Locarnini, S.A. (1998) Significance of mutations in the hepatitis B virus polymerase selected by nucleoside analogues and implications for controlling chronic disease. *Viral Hepatitis Reviews* **4**, 167-187.

Bassami, M.R., Berryman, D., Wilcox, G.E. and Raidal, S.R. (1998) Psittacine beak and feather disease virus nucleotide sequence analysis and its relationship to porcine circovirus, plant circoviruses, and chicken anaemia virus. *Virology* **249**, 453-459.

Beasley, R.P., Hwang, L.Y., Lin, C.C. and Chien, C.S. (1981) Hepatocellular carcinoma and hepatitis B virus. A prospective study of 22 707 men in Taiwan. *Lancet* **2**, 1129-1133.

Biagini, P., Gallian, P., Attoui, H., Cantaloube, J.-F., de Micco, P. and de Lamballerie, X. (1999) Determination and phylogenetic analysis of partial sequences from TT virus isolates. *Journal of General Virology* **80**, 419-424.

Biagini, P., Gallian, P., Touinssi, M., Cantaloube, J.-F., Zapitelli, J., de Lamballerie, X. and de Micco, P. (2000) High prevalence of TT virus infection in French blood donors revealed by the use of three PCR systems. *Transfusion* **40**, 590-595.

Bird, A.G., Gore, S.M., Burns, S.M. and Duggie, J.G. (1993) Study of infection with HIV and related risk factors in young offenders' institution. *British Medical Journal* **307**, 228-231.

Bird, A.G., Gore, S.M., Hutchinson, S.J., Lewis, S.C., Cameron, S. and Burns, S. (1997) Harm reduction measures and injecting inside prison versus mandatory drugs testing: results of a cross sectional anonymous questionnaire survey. *British Medical Journal* **315**, 21-24.

Blumberg, B.S., Gerstley, B.J., Hungerford, D.A., London, W.T. and Sutnick, A.I. (1967) A serum antigen (Australia antigen) in Down's syndrome, leukemia, and hepatitis. *Annals of Internal Medicine* **66**, 924-931.

Boag, F. (1991) Hepatitis B: heterosexual transmission and vaccination strategies. *International Journal of STD and AIDS* **2**, 318-324.

Bollyky, P.L., Rambaut, A., Harvey, P.H. and Holmes, E.C. (1996) Recombination between sequences of hepatitis B virus from different genotypes. *Journal of Molecular Evolution* **42**, 97-102.

Boom, R., Sol, C.J.A., Salimans, M.M.M., Jansen, C.L., Wertheim-van Dillen, P.M.E. and van der Noordaa, J. (1990) Rapid and simple method for purification of nucleic acids. *Journal of Clinical Microbiology* **28**, 495-503.

Bosch, V., Bartenschlager, R., Radziwill, G. and Schaller, H. (1988) The duck hepatitis B virus P-gene codes for protein strongly associated with the 5'-end of the viral DNA minus strand. *Virology* **166**, 475-485.

Bowyer, S.M., van Staden, L., Kew, M.C. and Sim, J.G. (1997) A unique segment of the hepatitis B virus group A genotype identified in isolates from South Africa. *Journal of General Virology* **78**, 1719-1729.

Bowyer, S.M. and Sim, J.G. (2000) Relationships within and between genotypes of hepatitis B virus at points across the genome: footprints of recombination in certain isolates. *Journal of General Virology* **81**, 379-392.

Brook, M.G., Karayiannis, P. and Thomas, H.C. (1989) Which patients with chronic hepatitis B virus infection will respond to alpha-interferon therapy? A statistical analysis of predictive factors. *Hepatology* **10**, 761-763.

Bruns, M., Miska, S., Chassot, S. and Will, H. (1998) Enhancement of hepatitis B virus infection by noninfectious subviral particles. *Journal of Virology* **72**, 1462-1468.

Bruss, V., Gerhardt, E., Vieluf, K. and Wunderlich, G. (1996) Functions of the large hepatitis B virus surface protein in viral particle morphogenesis. *Intervirology* **39**, 23-31.

Bruss, V. (1997) A short linear sequence in the pre-S domain of the large hepatitis B virus envelope protein required for virion formation. *Journal of Virology* **71**, 9350-9357.

Camin, J.H. and Sokal, R.R. (1965) A method for deducing branching sequences in phylogeny. *Evolution* **19**, 311-326.

Carman, W.F., Jacyna, M.R., Hadziyannis, S., Karayiannis, P., McGarvey, M.J., Makris, A. and Thomas, H.C. (1989) Mutation preventing formation of hepatitis B e antigen in patients with chronic hepatitis B infection. *Lancet* **2**, 588-591.

Carman, W.F., Zanetti, A.R., Karayiannis, P., Waters, J., Manzillo, G., Tanzi, E., Zuckerman, A.J. and Thomas, H.C. (1990) Vaccine-induced escape mutant of hepatitis B virus. *Lancet* **336**, 325-329.

Carman, W.F., Korula, J., Wallace, L., MacPhee, R., Mimms, L. and Decker, R. (1995) Fulminant reactivation of hepatitis B due to envelope protein mutant that escaped detection by monoclonal HBsAg ELISA. *Lancet* **345**, 1406-1407.

Carman, W.F., Thomas, H.C., Zuckerman, A.J. and Harrison, T.J. (1998) Molecular variants of hepatitis B virus. In: Zuckerman, A.J. and Thomas, H.C., (Eds.) *Viral Hepatitis*, 2 edn. pp. 141-172. London, UK: [Churchill Livingstone]

Charlton, M., Adjei, P., Poterucha, J., Moore, B., Therneau, T., Krom, R. and Wiesner, R. (1998) TT-virus infection in North American blood donors,

patients with fulminant hepatic failure, and cryptogenic cirrhosis. *Hepatology* **28**, 839-842.

Charrel, R.N., Levy, N., Tesh, R.B. and Chandler, L.J. (1999) Use of base excision sequence scanning for detection of genetic variations in St. Louis encephalitis virus isolates. *Journal of Clinical Microbiology* **37**, 1935-1940.

Chisari, F.V. and Ferrari, C. (1995) Hepatitis B virus immunopathogenesis. *Annual Reviews of Immunology* **13**, 29-60.

Chong-Jin, O., Wei, N.C., Shiuan, K. and Gek, K.L. (1999) Identification of hepatitis B surface antigen variants with alterations outside the "a" determinant in immunized Singapore infants. *Journal of Infectious Diseases* **179**, 259-263.

Choo, Q.L., Kuo, G., Weiner, A.J., Overby, L.R., Bradley, D.W. and Houghton, M. (1989) Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. *Science* **244**, 359-362.

Clewley, J.P. (1998) A user's guide to producing and interpreting tree diagrams in taxonomy and phylogenetics. Part 2. The multiple alignment of DNA and protein sequences to determine their relationships. *Communicable Disease and Public Health* **1**, 132-134.

Cong, M.-E., Nichols, B., Dou, X.-G., Spelbring, J.E., Krawczynski, K., Fields, H.A. and Khudyakov, Y.E. (2000) Related TT viruses in chimpanzees. *Virology* **274**, 343-355.

Coombs, N.J., Gough, A.C. and Primrose, J.N. (1999) Optimisation of DNA and RNA extraction from archival formalin-fixed tissue. *Nucleic Acids Research* **27**, e12.

Cotton, R.G. (1993) Current methods of mutation detection. *Mutation Research* **285**, 125-144.

Couroucé, A.M., Holland, P.V., Muller, J.Y. and Soulier, J.P. (1976) HBs antigen subtypes. *Bibliotheca Haematologica* **42**.

Couroucé-Pauty, A.M., Lemaire, J.M. and Roux, J.F. (1978) New hepatitis B surface antigen subtypes inside the *ad* category. *Vox Sanguinis* **35**, 304-308.

Crandall, K.A. (1995) Intraspecific phylogenetics: support for dental transmission of human immunodeficiency virus. *Journal of Virology* **69**, 2351-2356.

Curtis, S.P. and Edwards, A. (1995) HIV in UK prisons: a review of seroprevalence, transmission and patterns of risk. *International Journal of STD and AIDS* **6**, 387-391.

Dane, D.S., Cameron, C.H. and Briggs, M. (1970) Virus-like particles in serum of patients with Australia-antigen-associated hepatitis. *Lancet* **1**, 695-698.

Das, K., Sarafianos, S.G., Arnold, E., Westland, C., Yang, H., Gibbs, C. and Xiong, X. (2000) Structural basis of HBV resistance to lamivudine (3TC) and emtricitabine (FTC) (oral presentation). 10th International Symposium on Viral Hepatitis and Liver Disease, April 9th-13th, 2000, Atlanta, USA.

Davidson, F., MacDonald, D.M., Mokili, J.L.K., Prescott, L.E., Graham, S. and Simmonds, P. (1999) Early acquisition of TT virus (TTV) in an area endemic for TTV infection. *Journal of Infectious Diseases* **179**, 1070-1076.

Davis, H.L. (1998) DNA-based immunization against hepatitis B: experience with animal models. *Current Topics in Microbiology and Immunology* **226**, 57-68.

Dayhoff, M.O. (1978) *Atlas of Protein Sequence and Structure, Volume 5, Supplement 3*, Washington, DC: National Biomedical Research Foundation.

De Santis, R. and Azzi, A. (2000) Use of denaturing gradient gel electrophoresis for human polyomavirus JC sequence analysis. *Journal of Virological Methods* **85**, 101-108.

Dean, M. (1999) Britain on the brink of new heroin epidemic. *Lancet* **353**, 1947

Delius, H., Saegling, B., Bergmann, K., Shamanin, V. and de Villiers, E.-M. (1998) The genomes of three of four novel HPV types, defined by differences of their L1 genes, show high conservation of the E7 gene and the URR. *Virology* **240**, 359-365.

Desai, S.M., Muerhoff, A.S., Leary, T.P., Erker, J.C., Simons, J.N., Chalmers, M.L., Birkenmeyer, L.G., Pilot-Matias, T.J. and Mushahwar, I.K. (1999) Prevalence of TT virus infection in US blood donors and populations at risk for acquiring parenterally transmitted viruses. *Journal of Infectious Diseases* **179**, 1242-1244.

Dopazo, J., Dress, A. and von Haeseler, A. (1993) Split decomposition: a technique to analyze viral evolution. *Proceedings of the National Academy of Sciences USA* **90**, 10320-10324.

Edwards, A.W.F. and Cavalli-Sforza, L.L. (1964) Reconstruction of evolutionary trees. In: Heywood, V.H. and McNeill, J., (Eds.) pp. 67-76. London, UK: [Systematics Association Publ. No. 6]

Erker, J.C., Leary, T.P., Desai, S.M., Chalmers, M.L. and Mushahwar, I.K. (1999) Analyses of TT virus full-length genomic sequences. *Journal of General Virology* **80**, 1743-1750.

Evans, A.A. and London, W.T. (1998) Epidemiology of hepatitis B. In: Zuckerman, A.J. and Thomas, H.C., (Eds.) *Viral Hepatitis* , pp. 107-114. London: [Churchill Livingstone]

Fallows, D.A. and Goff, S.P. (1996) Hepadnaviruses: current models of RNA encapsidation and reverse transcription. *Advances in Virus Research* **46**, 165-194.

Fattovich, G. (1996) Natural course and prognosis of chronic hepatitis type B. *Viral Hepatitis Reviews* **2**, 263-276.

Feinstone, S.M., Kapikian, A.Z. and Purceli, R.H. (1973) Hepatitis A: detection by immune electron microscopy of a virus-like antigen associated with acute illness. *Science* **182**, 1026-1028.

Felsenstein, J. (1981) Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of Molecular Evolution* **17**, 368-376.

Felsenstein, J. (1988) Phylogenies from molecular sequences: inference and reliability. *Annual Reviews of Genetics* **22**, 521-565.

Felsenstein, J. (1993) PHYLIP (phylogeny inference package) v3.5c. Department of Genetics, University of Washington, Seattle, USA.

Fenton, K.A., Rogers, P.A., Simms, I., Maguire, H. and Catchpole, M. (2000) Increasing gonorrhoea reports - not only in London. *Lancet* **355**, 1907.

Fiordalisi, G., Zanella, I., Mantero, G., Bettinardi, A., Stellini, R., Paraninfo, G., Cadeo, G. and Primi, D. (1996) High prevalence of GB virus C infection in a group of Italian patients with hepatitis of unknown etiology. *Journal of Infectious Diseases* **174**, 181-183.

Fitch, W.M. and Margoliash, E. (1967) Construction of phylogenetic trees. *Science* **155**, 279-284.

Flodgren, E., Bengtsson, S., Knutsson, M., Strebkova, E.A., Kidd, A.H., Alexeyev, O.A. and Kidd-Ljunggren, K. (2000) Recent high incidence of fulminant hepatitis in Samara, Russia: molecular analysis of prevailing hepatitis B and D virus strains. *Journal of Clinical Microbiology* **38**, 3311-3316.

Fodde, R. and Losekoot, M. (1994) Mutation detection by denaturing gradient gel electrophoresis (DGGE). *Human Mutation* **3**, 83-94.

Fujiwara, T., Iwata, A., Iizuka, H., Tanaka, T. and Okamoto, H. (1998) Transfusion transmitted virus. *Lancet* **352**, 1310-1310.

Fukuda, Y., Nakano, I., Katano, Y., Kumada, T., Hayashi, K., Nakano, S. and Hayakawa, T. (1999) TT virus (TTV) is not associated with acute sporadic hepatitis. *Infection* **27**, 125-127.

Furness, P. (2001) The data protection act and human tissues. Royal College of Pathologists, London (http://www.rcpath.org).

Gerlich, W.H., and Caspari, G. (1999) Hepatitis viruses and the safety of blood donations. *Journal of Viral Hepatitis* **6 (Suppl. 1)**, 6-15.

Gerolami, V., Halfon, P., Chambost, H., Thuret, I. and Halimi, G. (1998) Transfusion transmitted virus. *Lancet* **352**, 1309-1309.

Gill, O.N., Noone, A. and Heptonstall, J. (1995) Imprisonment, injecting drug use, and bloodborne viruses. *British Medical Journal* **310**, 275-276.

Giménez-Barcons, M., Forns, X., Guilera, M., Soler, M., Soguero, C., Sánchez-Tapias, J.-M., Rodés, J. and Saiz, J.-C. (1999) Infection with a novel human DNA virus (TTV) has no pathogenic significance in patients with liver diseases. *Journal of Hepatology* **30**, 1028-1034.

Gluck, T.A., Knowles, W.A., Johnson, M.A., Brook, M.G. and Pillay, D. (1994) BK virus-associated haemorrhagic cystitis in an HIV-infected man. *AIDS* **8**, 391-392.

Goldberg, D. and McMenamin, J. (1998) The United Kingdom's hepatitis B immunisation strategy--where now? *Communicable Disease and Public Health* **1**, 79-83.

Gore, S.M., Bird, A.G. and Ross, A.J. (1995) Prison rites: starting to inject inside. *British Medical Journal* **311**, 1135-1136.

Grady, G.F., Lee, V.A., Prince, A.M., Gitnick, G.L., Fawaz, K.A., Vyas, G.N., Levitt, M.D., Senior, J.R., Galambos, J.T., Bynum, T.E., Singleton, J.W., Clowdus, B.F., Akdamar, K., Aach, R.D., Winkelman, E.I., Schiff, G.M. and Hersh, T. (1978) Hepatitis B immune globulin for accidental exposures among medical personnel: final report of a multicenter controlled trial. *Journal of Infectious Diseases* **138**, 625-638.

Grgacic, E.V.L. and Schaller, H. (2000) A metastable form of the large envelope protein of duck hepatitis B virus: low-pH release results in a transition to a hydrophobic, potentially fusogenic conformation. *Journal Of Virology* **74**, 5116-5122.

Gunther, S., Sommer, G., Plikat, U., Iwanska, A., Wain-Hobson, S., Will, H. and Meyerhans, A. (1997) Naturally occurring hepatitis B virus genomes bearing the hallmarks of retroviral G→A hypermutation. *Virology* **235**, 104-108.

Gunther, S., Fischer, L., Pult, I., Sterneck, M. and Will, H. (1999*a*) Naturally occurring variants of hepatitis B virus. *Advances in Virus Research* **52**, 25-137.

Gunther, S., von Breunig, F., Santantonio, T., Jung, M.C., Gaeta, G.B., Fischer, L., Sterneck, M. and Will, H. (1999*b*) Absence of mutations in the YMDD motif/B region of the hepatitis B virus polymerase in famciclovir therapy failure. *Journal of Hepatology* **30**, 749-754.

Hagelstein, J., Fathinejad, F., Stremmel, W. and Galle, P.R. (1997) pH-independent uptake of hepatitis B virus in primary human hepatocytes. *Virology* **229**, 292-294.

Hamasaki, K., Nakata, K., Nagayama, Y., Ohtsuru, A., Daikoku, M., Taniguchi, K., Tsutsumi, T., Sato, Y., Kato, Y. and Nagataki, S. (1994) Changes in the prevalence of HBeAg-negative mutant hepatitis B virus during the course of chronic hepatitis B. *Hepatology* **20**, 8-14.

Handa, A., Dickstein, B., Young, N.S. and Brown, K.E. (2000) Prevalence of the newly described human circovirus, TTV, in United States blood donors. *Transfusion* **40**, 245-251.

Hannoun, C., Horal, P. and Lindh, M. (2000) Long-term mutation rates in the hepatitis B virus genome. *Journal of General Virology* **81**, 75-83.

Hardie, D.R., Kannemeyer, J. and Stannard, L.M. (1996) DNA single strand conformation polymorphism identifies five defined strains of hepatitis B virus (HBV) during an outbreak of HBV infection in an oncology unit. *Journal of Medical Virology* **49**, 49-54.

Harris, K.A. and Teo, C.G. (2001) Diversity of hepatitis C virus quasispecies evaluated by denaturing gradient gel electrophoresis. *Clinical and Diagnostic Laboratory Immunology* **8**, 62-73.

Hasegawa, M., Kishino, H. and Yano, T. (1985) Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution* **22**, 160-174.

Hashida, Y., Gaffney, P.C. and Yunis, E.J. (1976) Acute hemorrhagic cystitis of childhood and papovavirus-like particles. *Journal of Pediatrics* **89**, 85-87.

Havens, W.P. (1946) Period of infectivity of patients with experimentally induced infectious hepatitis. *Journal of Experimental Medicine* **83**, 251-258.

Hawkins, A.E., Zuckerman, M.A., Briggs, M., Gilson, R.J., Goldstone, A.H., Brink, N.S. and Tedder, R.S. (1996) Hepatitis B nucleotide sequence analysis: linking an outbreak of acute hepatitis B to contamination of a cryopreservation tank. *Journal of Virological Methods* **60**, 81-88.

Hawkins, G.A. and Hoffman, L.M. (1997) Base excision sequence scanning. *Nature Biotechnology* **15**, 803-804.

Heringlake, S., Osterkamp, S., Trautwein, C., Tillmann, H.L., Boker, K., Muerhoff, S., Mushahwar, I.K., Hunsmann, G. and Manns, M.P. (1996) Association between fulminant hepatic failure and a strain of GBV virus C. *Lancet* **348**, 1626-1629.

Hijikata, M., Takahashi, K. and Mishiro, S. (1999) Complete circular DNA genome of a TT virus variant (isolate name SANBAN) and 44 partial ORF2 sequences implicating a great degree of diversity beyond genotypes. *Virology* **260**, 17-22.

Hillis, D.M., Huelsenbeck, J.P. and Cunningham, C.W. (1994) Application and accuracy of molecular phylogenies. *Science* **264**, 671-677.

Hoffman, A., Tucker, T., Zachoval, R., Nitschko, H., Breitschop, H., Frosner G.G. (1998) Localisation of replication site of GB virus A, GB virus B and GB virus C by in-situ hybridization (poster presentation). European Society for Clinical Virology - Progress in Clinical Virology IV, August 28th-September 2nd, 1998, Hamburg, Germany.

Holmes, E.C. and Garnett, G.P. (1994) Genes, trees and infections: molecular evidence in epidemiology. *Tree* **9**, 256-260.

Holmes, E.C., Worobey, M. and Rambaut, A. (1999) Phylogenetic evidence for recombination in dengue virus. *Molecular Biology and Evolution* **16**, 405-409.

Hsieh, S.-Y., Wu, Y.-H., Ho, Y.-P., Tsao, K.-C., Yeh, C.-T. and Liaw, Y.-F. (1999) High prevalence of TT virus infection in healthy children and adults and in patients with liver disease in Taiwan. *Journal of Clinical Microbiology* **37**, 1829-1831.

Huang, Y.-H., Wu, J.-C., Chiang, T.-Y., Chan, Y.-J., Huo, T.-I., Huang, Y.-S., Hwang, S.-J., Chang, F.-Y. and Lee, S.-D. (2000) Detection and viral nucleotide sequence analysis of transfusion-transmitted virus infection in acute fulminant and non-fulminant hepatitis. *Journal of Viral Hepatitis* **7**, 56-63.

Hussain, M. and Lok, A.S. (1999) Mutations in the hepatitis B virus polymerase gene associated with antiviral treatment for hepatitis B. *Journal of Viral Hepatitis* **6**, 183-194.

Hutchinson, S.J., Goldberg, D.J., Gore, S.M., Cameron, S., McGregor, J., McMenamin, J. and McGavigan, J. (1998) Hepatitis B outbreak at Glenochil prison during January to June 1993. *Epidemiology and Infection* **121**, 185-191.

Inami, T., Konomi, N., Arakawa, Y. and Abe, K. (2000) High prevalence of TT virus DNA in human saliva and semen. *Journal of Clinical Microbiology* **38**, 2407-2408.

Jaeger, J.A., Turner, D.H. and Zuker, M. (1989) Improved predictions of secondary structures for RNA. *Proceedings of the National Academy of Sciences USA* **86**, 7706-7710.

Jeong, J.-K., Yoon, G.-S. and Ryu, W.-S. (2000) Evidence that the 5'-end cap structure is essential for encapsidation of hepatitis B virus pregenomic RNA. *Journal Of Virology* **74**, 5502-5508.

Jobes, D.V., Chima, S.C., Ryschkewitsch, C.F. and Stoner, G.L. (1998) Phylogenetic analysis of 22 complete genomes of the human polyomavirus JC virus. *Journal of General Virology* **79**, 2491-2498.

Jukes, T.H. and Cantor, C.R. (1969) Evolution of protein molecules. In: Munro, H.N., (Ed.) *Mammalian Protein Metabolism*, pp. 21-123. New York: [Academic Press]

Kanda, T., Yokosuka, O., Ikeuchi, T., Seta, T., Kawai, S., Imazeki, F. and Saisho, H. (1999) The role of TT virus infection in acute viral hepatitis. *Hepatology* **29**, 1905-1908.

Kane, M. (1996) Global strategies for the control of hepatitis B. In: Zuckerman, A.J., (Ed.) *Prevention of hepatitis B in the newborn, children and adolescents*, pp. 87-96. London: The Royal College of Physicians of [London]

Kaneko, S., Feinstone, S.M. and Miller, R.H. (1989) Rapid and sensitive method for the detection of serum hepatitis B virus DNA using the polymerase chain reaction technique. *Journal of Clinical Microbiology* **27**, 1930-1933.

Kann, M. and Gerlich, W. (1998) Hepadnaviridae: structure and molecular virology. In: Zuckerman, A.J. and Thomas, H.C., (Eds.) *Viral Hepatitis*, 2 edn. pp. 77-105. London, UK: [Churchill Livingstone]

Karayiannis, P., Alexopoulou, A., Hadziyannis, S., Thursz, M., Watts, R., Seito, S. and Thomas, H.C. (1995) Fulminant hepatitis associated with hepatitis B virus e antigen-negative infection: importance of host factors. *Hepatology* **22**, 1628-1634.

Khudyakov, Y.E., Cong, M.-E., Nichols, B., Reed, D., Dou, X.-G., Viazov, S.O., Chang, J., Fried, M.W., Williams, I., Bower, W., Lambert, S., Purdy, M., Roggendorf, M. and Fields, H.A. (2000) Sequence heterogeneity of TT virus and closely related viruses. *Journal Of Virology* **74**, 2990-3000.

Kidd-Ljunggren, K., Broman, E., Ekvall, H. and Gustavsson, O. (1999) Nosocomial transmission of hepatitis B virus infection through multiple-dose vials. *Journal of Hospital Infection* **43**, 57-62.

Kimura, M. (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* **16**, 111-120.

Knaus, T. and Nassal, M. (1993) The encapsidation signal on the hepatitis B virus RNA pregenome forms a stem-loop structure that is critical for its function. *Nucleic Acids Research* **21** , 3967-3975.

Kobayashi, M., Tanaka, E., Nakayama, J., Furuwatari, C., Katsuyama, T., Kawasaki, S. and Kiyosawa, K. (1999) Detection of GB virus-C/hepatitis G virus genome in peripheral blood mononuclear cells and liver tissue. *Journal of Medical Virology* **57**, 114-121.

Kozak, M. (1986) Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* **44**, 283-292.

Köck, J., Borst, E.M. and Schlicht, H.J. (1996) Uptake of duck hepatitis B virus into hepatocytes occurs by endocytosis but does not require passage of the virus through an acidic intracellular compartment. *Journal of Virology* **70**, 5827-5831.

Krawczynski, K. and Bradley, D.W. (1989) Enterically transmitted non-A, non-B hepatitis: identification of virus-associated antigen in experimentally infected cynomolgus macaques. *Journal of Infectious Diseases* **159**, 1042-1049.

Krugman, S., Giles, J.P. and Hammond, J. (1967) Infectious hepatitis. Evidence for two distinctive clinical, epidemiological, and immunological types of infection. *Journal of the American Medical Association* **200**, 365-373.

Lai, C.L., Chien, R.N., Leung, N.W., Chang, T.T., Guan, R., Tai, D.I., Ng, K.Y., Wu, P.C., Dent, J.C., Barber, J., Stephenson, S.L. and Gray, D.F. (1998) A one-year trial of lamivudine for chronic hepatitis B. *New England Journal of Medicine* **339**, 61-68.

Lamagni, T.L., Davison, K.L., Hope, V.D., Luutu, J.W., Newham, J.A., Parry, J.V. and Gill, O.N. (1999) Poor hepatitis B vaccine coverage in injecting drug users: England, 1995 and 1996. *Communicable Disease and Public Health* **2**, 174-177.

Landers, T.A., Greenberg, H.B. and Robinson, W.S. (1977) Structure of hepatitis B Dane particle DNA and nature of the endogenous DNA polymerase reaction. *Journal of Virology* **23**, 368-376.

Lauder, I.J., Lin, H.J., Lau, J.Y., Siu, T.S. and Lai, C.L. (1993) The variability of the hepatitis B virus genome: statistical analysis and biological implications. *Molecular Biology and Evolution* **10**, 457-470.

Le Bouvier, G.L. (1971) The heterogeneity of Australia antigen. *Journal of Infectious Diseases* **123**, 671-675.

Leary, T.P., Erker, J.C., Chalmers, M.L., Desai, S.M. and Mushahwar, I.K. (1999*a*) Improved detection systems for TT virus reveal high prevalence in humans, non-human primates and farm animals. *Journal of General Virology* **80**, 2115-2120.

Leary, T.P., Erker, J.C., Chalmers, M.L., Desai, S.M. and Mushahwar, I.K. (1999*b*) Optimized PCR assay for the detection of TT virus. *Journal of Virological Methods* **82**, 109-112.

Lefrere, J.J., Roudot-Thoraval, F., Lefrere, F., Kanfer, A., Mariotti, M., Lerable, J., Thauvin, M., Lefevre, G., Rouger, P. and Girot, R. (2000) Natural history of the TT virus infection through follow-up of TTV DNA-positive multiple transfused patients. *Blood* **95**, 347-351.

Lenhoff, R.J. and Summers, J. (1994) Coordinate regulation of replication and virus assembly by the large envelope protein of an avian hepadnavirus. *Journal of Virology* **68**, 4565-4571.

Lerman, L.S. and Silverstein, K. (1987) Computational simulation of DNA melting and its application to denaturing gradient gel electrophoresis. *Methods in Enzymology* **155**, 482-501.

Li, W.-H. (1997) *Molecular Evolution*, Massachusetts, USA: Sinauer Associates, Inc.

Lin, H.J., Lai, C.-L., Lau, J.Y.N., Chung, H.-T., Lauder, I.J. and Fong, M.-W. (1990) Evidence for intrafamilial transmission of hepatitis B virus from sequence analysis of mutant HBV DNAs in two Chinese families. *Lancet* **336**, 208-212.

Lin, H.J., Lai, C.L., Lauder, I.J., Wu, P.C., Lau, T.K. and Fong, M.W. (1991) Application of hepatitis B virus (HBV) DNA sequence polymorphisms to the study of HBV transmission. *Journal of Infectious Diseases* **164**, 284-288.

Lindh, M., Gonzalez, J.E., Norkrans, G. and Horal, P. (1998) Genotyping of hepatitis B virus by restriction pattern analysis of a pre-S amplicon. *Journal of Virological Methods* **72**, 163-174.

Ling, R. and Harrison, T.J. (1999) Functional analysis of mutations conferring lamivudine resistance on hepatitis B virus. *Journal of General Virology* **80**, 601-606.

Linnen, J., Wages, J.J., Zhang-Keck, Z.Y., Fry, K.E., Krawczynski, K.Z., Alter, H., Koonin, E., Gallagher, M., Alter, M., Hadziyannis, S., Karayiannis, P., Fung, K., Nakatsuji, Y., Shih, J.W., Young, L., Piatak, M.J., Hoover, C., Fernandez, J., Chen, S., Zou, J.C., Morris, T., Hyams, K.C., Ismay, S., Lifson, J.D. and Kim, J.P. (1996) Molecular cloning and disease association of hepatitis G virus: a transfusion-transmissible agent. *Science* **271**, 505-508.

Lok, A.S., Lai, C.L., Wu, P.C., Wong, V.C., Yeoh, E.K. and Lin, H.J. (1987) Hepatitis B virus infection in Chinese families in Hong Kong. *American Journal of Epidemiology* **126**, 492-499.

Lok, A.S., Akarca, U. and Greene, S. (1994) Mutations in the pre-core region of hepatitis B virus serve to enhance the stability of the secondary structure of the pre-genome encapsidation signal. *Proceedings of the National Academy of Sciences USA* **91**, 4077-4081.

Longfield, M. (1999) Annual report of the director of health care, 1997-1998. DD604393 CCN077828, London: The Stationery Office.

MacCallum, F.O. and Bauer, D.J. (1947) Homologous serum hepatitis. *Lancet* **2**, 691-692.

Maggi, F., Fornai, C., Morrica, A., Casula, F., Vatteroni, M.L., Marchi, S., Ciccorossi, P., Riente, L., Pistello, M. and Bendinelli, M. (1999) High prevalence of TT virus viremia in Italian patients, regardless of age, clinical diagnosis, and previous interferon treatment. *Journal of Infectious Diseases* **180**, 838-842.

Magnius, L.O., Kaplan, L., Vyas, G.N. and Perkins, H.A. (1975) A new virus specified determinant of hepatitis B surface antigen. *Acta Pathologica et Microbiologica Scandinavica* **83B**, 295-297.

Magnius, L.O. and Norder, H. (1995) Subtypes, genotypes and molecular epidemiology of the hepatitis B virus as reflected by sequence variability of the S-gene. *Intervirology* **38**, 24-34.

Mahoney, F.J. (1999) Update on diagnosis, management, and prevention of hepatitis B virus infection. *Clinical Microbiology Reviews* **12**, 351-366.

Marchuk, D., Drumm, M., Saulino, A. and Collins, F.S. (1991) Construction of T-vectors, a rapid and general system for direct cloning of unmodified PCR products. *Nucleic Acids Research* **19**, 1154

Mason, A.L., Xu, L., Guo, L., Kuhns, M. and Perrillo, R.P. (1998) Molecular basis for persistent hepatitis B virus infection in the liver after clearance of serum hepatitis B surface antigen. *Hepatology* **27**, 1736-1742.

Matsumoto, A., Yeo, A.E.T., Shih, J.W.K., Tanaka, E., Kiyosawa, K. and Alter, H.J. (1999) Transfusion-associated TT virus infection and its relationship to liver disease. *Hepatology* **30**, 283-288.

Mayerat, C., Mantegani, A. and Frei, P.C. (1999) Does hepatitis B virus (HBV) genotype influence the clinical outcome of HBV infection? *Journal of Viral Hepatitis* **6**, 299-304.

Maynard Smith, J. and Smith, N.H. (1998) Detecting recombination from gene trees. *Molecular Biology and Evolution* **15**, 590-599.

Maynard Smith, J. (1999) The detection and measurement of recombination from sequence data. *Genetics* **153**, 1021-1027.

McAllister, J., Casino, C., Davidson, F., Power, J., Lawlor, E., Yap, P.L., Simmonds, P. and Smith, D.B. (1998) Long-term evolution of the hypervariable region of hepatitis C virus in a common-source-infected cohort. *Journal of Virology* **72**, 4893-4905.

McGuire, G. and Wright, F. (1998) TOPAL: recombination detection in DNA and protein sequences. *Bioinformatics.* **14**, 219-220.

McIntosh, E.D., Givney, R., Zhang, S.S., Courouce, A.M., Burgess, M. and Cossart, Y.E. (1998) Molecular epidemiology and variation of hepatitis B in recent immigrant families to Australia. *Journal of Medical Virology* **56**, 10-17.

McMahon, B.J., Alward, W.L., Hall, D.B., Heyward, W.L., Bender, T.R., Francis, D.P. and Maynard, J.E. (1985) Acute hepatitis B virus infection: relation of age to the clinical expression of disease and subsequent development of the carrier state. *Journal of Infectious Diseases* **151**, 599-603.

McMenamin, J. (1996) Hepatitis B in the UK: who is at risk, who succumbs? Role of screening. In: Zuckerman, A.J., (Ed.) *Prevention of hepatitis B in the newborn, children and adolescents*, pp. 1-8. London: [The Royal College of Physicians of London]

McNamee, D. (1999) Transgenic potatoes produce oral HBV vaccine. *Lancet* **354**, 1707-1707.

Meehan, B.M., Todd, D., Creelan, J.L., Earle, J.A.P., Hoey, E.M. and McNulty, M.S. (1992) Characterization of viral DNAs from cells infected with chicken anaemia agent: sequence analysis of the cloned replicative form and transfection capabilities of cloned genome fragments. *Archives of Virology* **124**, 301-319.

Miller, R.H., Marion, P.L. and Robinson, W.S. (1984) Hepatitis B viral DNA-RNA hybrid molecules in particles from infected liver are converted to viral DNA molecules during an endogenous DNA polymerase reaction. *Virology* **139**, 64-72.

Miyata, H., Tsunoda, H., Kazi, A., Yamada, A., Khan, M.A., Murakami, J., Kamahora, T., Shiraki, K. and Hino, S. (1999) Identification of a novel GC-rich 113-nucleotide region to complete the circular, single-stranded DNA genome of TT virus, the first human circovirus. *Journal Of Virology* **73**, 3582-3586.

Mizokami, M., Orito, E., Ohba, K., Ikeo, K., Lau, J.Y. and Gojobori, T. (1997) Constrained evolution with respect to gene overlap of hepatitis B virus. *Journal of Molecular Evolution* **44**, S83-S90

Morozov, I., Sirinarumitr, T., Sorden, S.D., Halbur, P.G., Morgan, M.K., Yoon, K.-J. and Paul, P.S. (1998) Detection of a novel strain of porcine circovirus in pigs with postweaning multisystemic wasting syndrome. *Journal of Clinical Microbiology* **36**, 2535-2541.

Morrica, A., Maggi, F., Vatteroni, M.L., Fornai, C., Pistello, M., Ciccorossi, P., Grassi, E., Gennazzani, A. and Bendinelli, M. (2000) TT virus: evidence for transplacental transmission. *Journal of Infectious Diseases* **181**, 803-804.

Mphahlele, M.J., Shattock, A.G., Boner, W., Quinn, J., McCormick, P.A. and Carman, W.F. (1997) Transmission of a homogenous hepatitis B virus

population of A1896-containing strains leading to mild resolving acute hepatitis and seroconversion to hepatitis B e antigen antibodies in an adult. *Hepatology* **26**, 743-746.

Murakami, S. (1999) Hepatitis B virus X protein: structure, function and biology. *Intervirology* **42**, 81-99.

Mushahwar, I.K., Erker, J.C., Muerhoff, A.S., Leary, T.P., Simons, J.N., Birkenmeyer, L.G., Chalmers, M.L., Pilot-Matias, T.J. and Dexai, S.M. (1999) Molecular and biophysical characterization of TT virus: evidence for a new virus family infecting humans. *Proceedings of the National Academy of Sciences USA* **96**, 3177-3182.

Myers, R.M., Maniatis, T. and Lerman, L.S. (1987) Detection and localization of single base changes by denaturing gradient gel electrophoresis. *Methods in Enzymology* **155**, 501-527.

Ngui, S.L., O'Connell, S., Eglin, R.P., Heptonstall, J. and Teo, C.G. (1997) Low detection rate and maternal provenance of hepatitis B virus S gene mutants in cases of failed postnatal immunoprophylaxis in England and Wales. *Journal of Infectious Diseases* **176**, 1360-1365.

Ngui, S.L., Hallett, R. and Teo, C.G. (1999) Natural and iatrogenic variation in hepatitis B virus. *Reviews in Medical Virology* **9**, 183-209.

Niagro, F.D., Forsthoefel, A.N., Lawther, R.P., Kamalanathan, L., Ritchie, B.W., Latimer, K.S. and Lukert, P.D. (1998) Beak and feather disease virus and porcine circovirus genomes: intermediates between the geminiviruses and plant circoviruses. *Archives of Virology* **143**, 1723-1744.

Niel, C., Saback, F.L. and Lampe, E. (2000) Co-infection with multiple TT virus strains belonging to different genotypes is a common event in healthy Brazilian adults. *Journal of Clinical Microbiology* **38**, 1926-1930.

Niesters, H.G., Honkoop, P., Haagsma, E.B., de Man, R.A., Schalm, S.W. and Osterhaus, A.D. (1998) Identification of more than one mutation in the hepatitis B virus polymerase gene arising during prolonged lamivudine treatment. *Journal of Infectious Diseases* **177**, 1382-1385.

Nishizawa, T., Okamoto, H., Konishi, K., Yoshizawa, H., Miyakawa, Y. and Mayumi, M. (1997) A novel DNA virus (TTV) associated with elevated transaminase levels in posttransfusion hepatitis of unknown etiology. *Biochemical and Biophysical Research Communications* **241**, 92-97.

Nishizawa, T., Okamoto, H., Tsuda, F., Aikawa, T., Sugai, Y., Konishi, K., Akahane, Y., Ukita, M., Tanaka, T., Miyakawa, Y. and Mayumi, M. (1999) Quasispecies of TT virus (TTV) with sequence divergence in hypervariable regions of the capsid protein in chronic TTV infection. *Journal Of Virology* **73**, 9604-9608.

Norder, H., Courouce, A.M. and Magnius, L.O. (1994) Complete genomes, phylogenetic relatedness, and structural proteins of six strains of the hepatitis B virus, four of which represent two new genotypes. *Virology* **198**, 489-503.

Noteborn, M.H.M. and Koch, G. (1995) Chicken anaemia virus infection: molecular basis of pathogenicity. *Avian Pathology* **24**, 11-31.

Ogata, N., Zanetti, A.R., Yu, M., Miller, R.H. and Purcell, R.H. (1997) Infectivity and pathogenicity in chimpanzees of a surface gene mutant of hepatitis B virus that emerged in a vaccinated infant. *Journal of Infectious Diseases* **175**, 511-523.

Okamoto, H., Imai, M., Kametani, M., Nakamura, T. and Mayumi, M. (1987) Genomic heterogeneity of hepatitis B virus in a 54-year-old woman who contracted the infection through materno-fetal transmission. *Japanese Journal of Experimental Medicine* **57**, 231-236.

Okamoto, H., Tsuda, F., Sakugawa, H., Sastrosoewignjo, R.I., Imai, M., Miyakawa, Y. and Mayumi, M. (1988) Typing hepatitis B virus by homology in nucleotide sequence: comparison of surface antigen subtypes. *Journal of General Virology* **69**, 2575-2583.

Okamoto, H., Nishizawa, T., Kato, N., Ukita, M., Ikeda, H., Iizuka, H., Miyakawa, Y. and Mayumi, M. (1998*a*) Molecular cloning and characterisation of a novel DNA virus (TTV) associated with posttransfusion hepatitis of unknown etiology. *Hepatology Research* **10**, 1-16.

Okamoto, H., Akahane, Y., Ukita, M., Fukuda, M., Tsuda, F., Miyakawa, Y. and Mayumi, M. (1998*b*) Fecal excretion of a nonenveloped DNA virus (TTV) associated with posttransfusion non A-G hepatitis. *Journal of Medical Virology* **56**, 128-132.

Okamoto, H., Kato, N., Iizuka, H., Tsuda, F., Miyakawa, Y. and Mayumi, M. (1999*a*) Distinct genotypes of a nonenveloped DNA virus associated with posttransfusion non-A to G hepatitis (TT virus) in plasma and peripheral blood mononuclear cells. *Journal of Medical Virology* **57**, 252-258.

Okamoto, H., Nishizawa, T., Ukita, M., Takahashi, M., Fukuda, M., Iizuka, H., Miyakawa, Y. and Mayumi, M. (1999*b*) The entire nucleotide sequence of a TT virus isolate from the United States (TUS01): comparison with reported isolates and phylogenetic analysis. *Virology* **259**, 437-448.

Okamoto, H., Takahashi, M., Nishizawa, T., Ukita, M., Fukuda, M., Tsuda, F., Miyakawa, Y. and Mayumi, M. (1999*c*) Marked genomic heterogeneity and frequent mixed infection of TT virus demonstrated by PCR with primers from coding and noncoding regions. *Virology* **259**, 428-436.

Okamoto, H., Fukuda, M., Tawara, A., Nishizawa, T., Itoh, Y., Hayasaka, I., Tsuda, F., Tanaka, T., Miyakawa, Y. and Mayumi, M. (2000*a*) Species-specific TT viruses and cross-species infection in nonhuman primates. *Journal Of Virology* **74**, 1132-1139.

Okamoto, H., Takahashi, M., Nishizawa, T., Tawara, A., Sugai, Y., Sai, T., Tanaka, T. and Tsuda, F. (2000*b*) Replicative forms of TT virus DNA in bone marrow cells. *Biochemical and Biophysical Research Communications* **270**, 657-662.

Okamoto, H., Ukita, M., Nishizawa, T., Kishimoto, J., Hoshi, Y., Mizuo, H., Tanaka, T., Miyakawa, Y. and Mayumi, M. (2000*c*) Circular double-stranded forms of TT virus DNA in the liver. *Journal Of Virology* **74**, 5161-5167.

Okamoto, H., Takahashi, M., Kato, N., Fukuda, M., Tawara, A., Fukuda, S., Tanaka, T., Miyakawa, Y. and Mayumi, M. (2000*d*) Sequestration of TT virus of restricted genotypes in peripheral blood mononuclear cells. *Journal of Virology* **74**, 10236-10239.

Ono-Nita, S.K., Kato, N., Shiratori, Y., Masaki, T., Lan, K.H., Carrilho, F.J. and Omata, M. (1999) YMDD motif in hepatitis B virus DNA polymerase influences on replication and lamivudine resistance: A study by in vitro full-length viral DNA transfection. *Hepatology* **29**, 939-945.

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

Orita, M., Suzuki, Y., Sekiya, T. and Hayashi, K. (1989) Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction. *Genomics* **5**, 874-879.

Orito, E., Mizokami, M., Ina, Y., Moriyama, E.N., Kameshima, N., Yamamoto, M. and Gojobori, T. (1989) Host-independent evolution and a genetic classification of the hepadnavirus family based on nucleotide sequences. *Proceedings of the National Academy of Sciences USA* **86**, 7059-7062.

Overdier, D.G., Porcella, A. and Costa, R.H. (1994) The DNA-binding specificity of the hepatocyte nuclear factor 3/forkhead domain is influenced by amino-acid residues adjacent to the recognition helix. *Molecular and Cellular Biology* **14**, 2755-2766.

Padgett, B.L. and Walker, D.L. (1973) Prevalence of antibodies in human sera against JC virus, an isolate from a case of progressive multifocal leukoencephalopathy. *Journal of Infectious Diseases* **127**, 467-470.

Padidam, M., Sawyer, S. and Fauquet, C.M. (1999) Possible emergence of new geminiviruses by frequent recombination. *Virology* **265**, 218-225.

Page, R.D.M. (1996) Treeview: an application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences* **12**, 357-358.

Parquet, M.C., Yatsuhashi, H., Koga, M., Inoue, O., Matsumoto, T., Hamada, R., Igarashi, A. and Yano, M. (1999) Prevalence and clinical characteristics of TT virus (TTV) in patients with sporadic acute hepatitis of unknown etiology. *Journal of Hepatology* **31**, 985-989.

Paver, W.K. and Mortimer, P.P. (1996) An overview of the hepatitis viruses. *Clinical Microbiology and Infection* **2** , 132-141.

Pearson, W.R. and Lipman, D.J. (1988) Improved tools for biological sequence comparison. *Proceedings of the National Academy of Sciences USA* **85**, 2444-2448.

Petrosillo, N., Ippolito, G., Solforosi, L., Varaldo, P.E., Clementi, M. and Manzin, A. (2000) Molecular epidemiology of an outbreak of fulminant hepatitis B. *Journal of Clinical Microbiology* **38**, 2975-2981.

Petzold, D.R., Tautz, B., Wolf, F. and Drescher, J. (1999) Infection chains and evolution rates of hepatitis B virus in cardiac transplant recipients infected nosocomially. *Journal of Medical Virology* **58**, 1-10.

Pianko, S. and McHutchison, J. (1999) Chronic hepatitis B: new therapies on the horizon? *Lancet* **354**, 1662-1663.

Pickering, H. and Stimson, G.V. (1993) Syringe sharing in prison. *Lancet* **342** , 621-622.

Plumb, M., Frampton, J., Wainwright, H., Walker, M., Macleod, K., Goodwin, G. and Harrison, P. (1989) GATAAG: a *cis*-control region binding an erythroid-specific nuclear factor with a role in globin and non-globin gene expression. *Nucleic Acids Research* **17**, 73-92.

Pohjanpelto, P., Lappalainen, M., Widell, A., Asikainen, K. and Paunio, M. (1996) Hepatitis C genotypes in Finland determined by RFLP. *Clinical and Diagnostic Virology* **7**, 7-16.

Preikschat, P., Meisel, H., Will, H. and Gunther, S. (1999) Hepatitis B virus genomes from long-term immunosuppressed virus carriers are modified by specific mutations in several regions. *Journal of General Virology* **80**, 2685-2691.

Prescott, L.E. and Simmonds, P. (1998) Global distribution of transfusion-transmitted virus. *New England Journal of Medicine* **339**, 776-777.

Prescott, L.E., MacDonald, D.M., Davidson, F., Mokili, J., Pritchard, D.I., Arnot, D.E., Riley, E.M., Greenwood, B.M., Hamid, S., Saeed, A.A., McClure, M.O., Smith, D.B. and Simmonds, P. (1999) Sequence diversity of TT virus in geographically dispersed human populations. *Journal of General Virology* **80**, 1751-1758.

Pringle, C.R. (1999) Virus taxonomy at the XIth International Congress of Virology, Sydney, Australia, 1999. *Archives of Virology* **144**, 2065-2070.

Provost, P.J. and Hilleman, M.R. (1979) Propagation of human hepatitis A virus in cell culture in vitro. *Proceedings of the Society for Experimental Biology and Medicine* **160**, 213-221.

Radziwill, G., Tucker, W. and Schaller, H. (1990) Mutational analysis of the hepatitis B virus P gene product: domain structure and RNase H activity. *Journal of Virology* **64**, 613-620.

Rall, L.B., Standring, D.N., Laub, O. and Rutter, W.J. (1983) Transcription of hepatitis B virus by RNA polymerase II. *Molecular and Cellular Biology* **3**, 1766-1773.

Ramsay, M., Gay, N., Balogun, K. and Collins, M. (1998) Control of hepatitis B in the United Kingdom. *Vaccine* **16**, S52-S55

Rastegar, M., Rousseau, G.G. and Lemaigre, F.P. (2000) CCAAT/enhancer-binding protein-alpha is a component of the growth hormone-regulated network of liver transcription factors. *Endocrinology* **141**, 1686-1692.

Raunio, V.K., London, W.T., Sutnick, A.I., Millman, I. and Blumberg, B.S. (1970) Specificities of human antibodies to Australia antigen (34833). *Proceedings of the Society for Experimental Biology and Medicine* **134**, 548-557.

Rizzetto, M., Canese, M.G., Arico, S., Crivelli, O., Trepo, C., Bonino, F. and Verme, G. (1977) Immunofluorescence detection of new antigen-antibody system (delta/anti-delta) associated to hepatitis B virus in liver and in serum of HBsAg carriers. *Gut* **18**, 997-1003.

Rodríguez-Iñigo, E., Casqueiro, M., Bartolomé, J., Ortiz-Movilla, N., López-Alcorocho, J.M., Herrero, M., Manzarbeitia, F., Oliva, H. and Carreño, V. (2000) Detection of TT virus DNA in liver biopsies by *in situ* hybridization. *American Journal of Pathology* **156**, 1227-1234.

Roingeard, P., Dubois, F., Bacq, Y., Barin, F. and Goudeau, A. (1997) Could serum HBV DNA level in the early phase of acute hepatitis B predict evolution to chronic HBV carriage? *Journal of Hepatology* **26**, 953-954.

Romeo, R., Hegerich, P., Emerson, S.U., Colombo, M., Purcell, R.H. and Bukh, J. (2000) High prevalence of TT virus (TTV) in naive chimpanzees and in hepatitis C virus-infected humans: frequent mixed infections and identification of new TTV genotypes in chimpanzees. *Journal of General Virology* **81**, 1001-1007.

Ross, R.S., Viazov, S., Runde, V., Schaefer, U.W. and Roggendorf, M. (1999) Detection of TT virus DNA in specimens other than blood. *Journal of Clinical Virology* **13**, 181-184.

Roth, W.K. and Seifried, E. (1999) Routine PCR screening of blood. *Lancet* **353**, 1799-1800.

Saitou, N. and Nei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4**, 406-425.

Salminen, M.O., Carr, J.K., Burke, D.S. and McCutchan, F.E. (1995) Identification of breakpoints in intergenotypic recombinants of HIV type 1 by bootscanning. *AIDS Research and Human Retroviruses* **11**, 1423-1425.

Seeger, C. and Maragos, J. (1990) Identification and characterization of the woodchuck hepatitis virus origin of DNA replication. *Journal of Virology* **64**, 16-23.

Seignères, B., Pichoud, C., Ahmed, S.S., Hantz, O., Trepo, C. and Zoulim, F. (2000) Evolution of hepatitis B virus polymerase gene sequence during famciclovir therapy for chronic hepatitis B. *Journal of Infectious Diseases* **181**, 1221-1233.

Severini, A., Liu, X.Y., Wilson, J.S. and Tyrrell, D.L. (1995) Mechanism of inhibition of duck hepatitis B virus polymerase by (-)-beta-L-2',3'-dideoxy-3'-thiacytidine. *Antimicrobial Agents and Chemotherapy* **39**, 1430-1435.

Sheffield, V.C., Cox, D.R., Lerman, L.S. and Myers, R.M. (1989) Attachment of a 40-base-pair G + C-rich sequence (GC-clamp) to genomic DNA fragments by the polymerase chain reaction results in improved detection of single-base changes. *Proceedings of the National Academy of Sciences USA* **86**, 232-236.

Shouval, D., Ilan, Y., Adler, R., Deepen, R., Panet, A., Even-Chen, Z., Gorecki, M. and Gerlich, W.H. (1994) Improved immunogenicity in mice of a mammalian cell-derived recombinant hepatitis B vaccine containing pre-

S1 and pre-S2 antigens as compared with conventional yeast-derived vaccines. *Vaccine* **12**, 1453-1459.

Shuman, S. (1994) Novel approach to molecular cloning and polynucleotide synthesis using vaccinia DNA topoisomerase. *Journal of Biological Chemistry* **269**, 32678-32684.

Simmonds, P., Prescott, L.E., Logue, C., Davidson, F., Thomas, A.E. and Ludlam, C.A. (1999) TT virus - part of the normal human flora? *Journal of Infectious Diseases* **180**, 1748-1749.

Simmonds, P. (2001) The origin and evolution of hepatitis viruses in humans. *Journal of General Virology* **82**, 693-712.

Simons, J.N., Leary, T.P., Dawson, G.J., Pilot-Matias, T.J., Muerhoff, A.S., Schlauder, G.G., Desai, S.M. and Mushahwar, I.K. (1995) Isolation of novel virus-like sequences associated with human hepatitis. *Nature Medicine* **1**, 564-569.

Skelton, M., Kew, M.C., Yu, M.C., Crookes, R.L., Swanevelder, J.P. and Hodkinson, J. (2000) Transfusion-transmissible virus and hepatocellular carcinoma: a case-control study. *Journal of Viral Hepatitis* **7**, 230-234.

Sneath, P.H.A. and Sokal, R.R. (1973) *Numerical Taxonomy*, San Francisco, USA: W.H.Freeman.

Soldan, K. and Barbara, J. (1999) The risks of infection transmission by blood transfusion in England. *Journal of Clinical Pathology* **52**, 405-408.

Stevens, C.E., Beasley, R.P., Tsui, J. and Lee, W.C. (1975) Vertical transmission of hepatitis B antigen in Taiwan. *New England Journal of Medicine* **292**, 771-774.

Stewart, C.B. (1993) The powers and pitfalls of parsimony. *Nature* **361**, 603-607.

Strimmer, K. and von Haeseler, A. (1997) PUZZLE maximum likelihood analysis for nucleotide, amino acid, and two state data v4.0. Zoologisches Institut, Universitaet Muenchen, Meunchen, Germany.

Stuyver, L., De Gendt, S., Cadranel, J.F., Van Geyt, C., Van Reybroeck, G., Dorent, R., Gandjbachkh, I., Rosenheim, M., Charlotte, F., Opolon, P., Huraux, J.M. and Lunel, F. (1999) Three cases of severe subfulminant hepatitis in heart-transplanted patients after nosocomial transmission of a mutant hepatitis B virus. *Hepatology* **29**, 1876-1883.

Stuyver, L., De Gendt, S., Van Geyt, C., Zoulim, F., Fried, M., Schinazi, R.F. and Rossau, R. (2000) A new genotype of hepatitis B virus: complete genome and phylogenetic relatedness. *Journal of General Virology* **81**, 67-74.

Summers, J. and Mason, W.S. (1982) Replication of the genome of a hepatitis B-like virus by reverse transcription of an RNA intermediate. *Cell* **29**, 403-415.

Suzuki, H., Iino, S., Shiraki, K., Akahane, Y., Okamoto, H., Domoto, K. and Mishiro, S. (1994) Safety and efficacy of a recombinant yeast-derived pre-S2 + S-containing hepatitis B vaccine (TGP-943): phase 1, 2 and 3 clinical testing. *Vaccine* **12**, 1090-1096.

Takahashi, K., Ohta, Y. and Mishiro, S. (1998*a*) Partial ~2.4-kb sequences of TT virus (TTV) genome from eight Japanese isolates: diagnostic and phylogenetic implications. *Hepatology Research* **12**, 111-120.

Takahashi, K., Hoshino, H., Ohta, Y., Yoshida, N. and Mishiro, S. (1998*b*) Very high prevalence of TT virus (TTV) infection in general population of Japan revealed by a new set of PCR primers. *Hepatology Research* **12**, 233-239.

Takahashi, K., Iwasa, Y., Hijikata, M. and Mishiro, S. (2000*a*) Identification of a new human DNA virus (TTV-like mini virus, TLMV) intermediately related to TT virus and chicken anemia virus. *Archives of Virology* **145**, 979-993.

Takahashi, K., Hijikata, M., Samokhvalov, E.I. and Mishiro, S. (2000*b*) Full or near full length nucleotide sequences of TT virus variants (types SANBAN and YONBAN) and the TT virus-like mini virus. *Intervirology* **43**, 119-123.

Takayama, S., Yamazaki, S., Matsuo, S. and Sugii, S. (1999) Multiple infection of TT virus (TTV) with different genotypes in Japanese hemophiliacs. *Biochemical and Biophysical Research Communications* **256**, 208-211.

Tanaka, H., Okamoto, H., Luengrojanakul, P., Chainuvati, T., Tsuda, F., Tanaka, T., Miyakawa, Y. and Mayumi, M. (1998) Infection with an unenveloped DNA virus (TTV) associated with posttransfusion non-A to G hepatitis in hepatitis patients and healthy blood donors in Thailand. *Journal of Medical Virology* **56**, 234-238.

Tanaka, S., Yoshiba, M., Iino, S., Fukuda, M., Nakao, H., Tsuda, F., Okamoto, H., Miyakawa, Y. and Mayumi, M. (1995) A common-source outbreak of fulminant hepatitis B in hemodialysis patients induced by precore mutant. *Kidney International* **48**, 1972-1978.

Tanaka, Y., Mizokami, M., Orito, E., Ohno, T., Nakano, T., Kato, T., Kato, H., Mukaide, M., Park, Y.-M., Kim, B.-S. and Ueda, R. (1998) New

genotypes of TT virus (TTV) and a genotyping assay based on restriction fragment length polymorphism. *FEBS Letters* **437**, 201-206.

Tatt, I.D., Barlow, K.L. and Clewley, J.P. (2000) A gag gene heteroduplex mobility assay for subtyping HIV-1. *Journal of Virological Methods* **87**, 41-51.

Taylor, A., Goldberg, D., Emslie, J., Wrench, J., Gruer, L., Cameron, S., Black, J., Davis, B., McGregor, J. and Follett, E. (1995) Outbreak of HIV infection in a Scottish prison. *British Medical Journal* **310**, 289-292.

Tedder, R.S., Zuckerman, M.A., Goldstone, A.H., Hawkins, A.E., Fielding, A., Briggs, E.M., Irwin, D., Blair, S., Gorman, A.M. and Patterson, K.G. (1995) Hepatitis B transmission from contaminated cryopreservation tank. *Lancet* **346**, 137-140.

Terre, S., Petit, M.A. and Brechot, C. (1991) Defective hepatitis B virus particles are generated by packaging and reverse transcription of spliced viral RNAs in vivo. *Journal of Virology* **65**, 5539-5543.

Thio, C.L., Carrington, M., Marti, D., O'Brien, S.J., Vlahov, D., Nelson, K.E., Astemborski, J. and Thomas, D.L. (1999) Class II HLA alleles and hepatitis B virus persistence in African Americans. *Journal of Infectious Diseases* **179**, 1004-1006.

Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22**, 4673-4680.

Tillmann, H.L., Trautwein, C., Bock, T., Boker, K.H., Jackel, E., Glowienka, M., Oldhafer, K., Bruns, I., Gauthier, J., Condreay, L.D., Raab, H.R. and

Manns, M.P. (1999) Mutational pattern of hepatitis B virus on sequential therapy with famciclovir and lamivudine in patients with hepatitis B virus reinfection occurring under HBIg immunoglobulin after liver transplantation. *Hepatology* **30**, 244-256.

Tischer, I., Bode, L., Apodaca, J., Timm, H., Peters, D., Rasch, R., Pociuli, S. and Gerike, E. (1995) Presence of antibodies reacting with porcine circovirus in sera of humans, mice, and cattle. *Archives of Virology* **140**, 1427-1439.

Tisdale, M., Kemp, S.D., Parry, N.R. and Larder, B.A. (1993) Rapid *in vitro* selection of human immunodeficiency virus type 1 resistant to 3'-thiacytidine inhibitors due to a mutation in the YMDD region of reverse transcriptase. *Proceedings of the National Academy of Sciences USA* **15**, 5653-5656.

Tobe, K., Matsuura, K., Ogura, T., Tsuo, Y., Iwasaki, Y., Mizuno, M., Yamamoto, K., Higashi, T. and Tsuji, T. (2000) Horizontal transmission of hepatitis B virus among players of an American football team. *Archives of Internal Medicine* **160**, 2541-2545.

Todd, D., Creelan, J.L., Meehan, B.M. and McNulty, M.S. (1996) Investigation of the transfection capability of cloned tandemly-repeated chicken anaemia virus DNA fragments. *Archives of Virology* **141**, 1523-1534.

Toyoda, H., Naruse, M., Yokozaki, S., Morita, K., Nakano, I., Itakura, A., Okamura, M., Fukuda, Y. and Hayakawa, T. (1999) Prevalence of infection with TT virus (TTV), a novel DNA virus, in healthy Japanese subjects, newborn infants, cord blood and breast milk. *Journal of Infection* **38**, 198-199.

Tronche, F., Ringeisen, F., Blumenfeld, M., Yaniv, M. and Pontoglio, M. (1997) Analysis of the distribution of binding sites for a tissue-specific

transcription factor in the vertebrate genome. *Journal of Molecular Biology* **266**, 231-245.

Tsuda, F., Okamoto, H., Ukita, M., Tanaka, T., Akahane, Y., Konishi, K., Yoshizawa, H., Miyakawa, Y. and Mayumi, M. (1999) Determination of antibodies to TT virus (TTV) and application to blood donors and patients with post-transfusion non-A to G hepatitis in Japan. *Journal of Virological Methods* **77**, 199-206.

Ukita, M., Okamoto, H., Nishizawa, T., Tawara, A., Takahashi, M., Iizuka, H., Miyakawa, Y. and Mayumi, M. (2000) The entire nucleotide sequences of two distinct TT virus (TTV) isolates (TJN01 and TJN02) remotely related to the original TTV isolates. *Archives of Virology* **145**, 1543-1559.

Uy, A., Wunderlich, G., Olsen, D.B., Heerman, K.-H., Gerlich, W.H. and Thomssen, R. (1992) Genomic variability in the preS1 region and determination of routes of transmission of hepatitis B virus. *Journal of General Virology* **73**, 3005-3009.

Verschoor, E.J., Langenhuijzen, S. and Heeney, J.L. (1999) TT viruses (TTV) of non-human primates and their relationship to the human TTV genotypes. *Journal of General Virology* **80**, 2491-2499.

Wallace, L.A. and Carman, W.F. (1997) Surface gene variation of HBV: scientific and medical relevance. *Viral Hepatitis Reviews* **3**, 5-16.

Weber, M., Bronsema, V., Bartos, H., Bosserhoff, A., Bartenschlager, R. and Schaller, H. (1994) Hepadnavirus P protein utilizes a tyrosine residue in the TP domain to prime reverse transcription. *Journal of Virology* **68**, 2994-2999.

Weild, A.R., Gill, O.N., Bennett, D., Livingstone, S.J., Parry, J.V. and Curran, L. (2000) Prevalence of HIV, hepatitis B, and hepatitis C antibodies in prisoners in England and Wales: a national survey. *Communicable Disease and Public Health* **3**, 121-126.

Weiller, G.F., McClure, M.A. and Gibbs, A.J. (1995) Molecular phylogenetic analysis. In: Gibbs, A., Calisher, C.H. and Garcia-Arenal, F., (Eds.) *Molecular Basis of Viral Evolution*, pp. 553-585. Cambridge, UK: [Cambridge University Press]

Wheeley, S.M., Jackson, P.T., Boxall, E.H., Tarlow, M.J., Gatrad, A.R., Anderson, J. and Bissenden, J. (1991) Prevention of perinatal transmission of hepatitis B virus (HBV): a comparison of two prophylactic schedules. *Journal of Medical Virology* **35**, 212-215.

White, M.B., Carvalho, M., Derse, D., O'Brien, S.J. and Dean, M. (1992) Detecting single base substitutions as heteroduplex polymorphisms. *Genomics* **12**, 301-306.

Will, H., Reiser, W., Weimer, T., Pfaff, E., Buscher, M., Sprengel, R., Cattaneo, R. and Schaller, H. (1987) Replication strategy of human hepatitis B virus. *Journal of Virology* **61**, 904-911.

Wilson, J.N., Nokes, D.J. and Carman, W.F. (1998) Current status of HBV vaccine escape variants--a mathematical model of their epidemiology. *Journal of Viral Hepatitis* **5 Suppl 2**, 25-30.

Wong, D.C., Purcell, R.H., Sreenivasan, M.A., Prasad, S.R. and Pavri, K.M. (1980) Epidemic and endemic hepatitis in India: evidence for a non-A, non-B hepatitis virus aetiology. *Lancet* **2**, 876-879.

Wong, D.K., Cheung, A.M., O'Rourke, K., Naylor, C.D., Detsky, A.S. and Heathcote, J. (1993) Effect of alpha-interferon treatment in patients with hepatitis B e antigen-positive chronic hepatitis B. A meta-analysis. *Annals of Internal Medicine* **119**, 312-323.

Worobey, M. and Holmes, E.C. (1999) Evolutionary aspects of recombination in RNA viruses. *Journal of General Virology* **80**, 2535-2543.

Worobey, M. (2000) Extensive homologous recombination among widely divergent TT viruses. *Journal Of Virology* **74**, 7666-7670.

Wreghitt, T.G. (1999) Blood-borne virus infections in dialysis units--a review. *Reviews in Medical Virology* **9**, 101-109.

Yamada-Osaki, M., Sumazaki, R., Noguchi, E., Shibasaki, M. and Matsui, A. (1998) Transfusion transmitted virus. *Lancet* **352**, 1309-1310.

Yang, Z., Goldman, N. and Friday, A. (1994) Comparison of models for nucleotide substitution used in maximum likelihood phylogenetic estimation. *Molecular Biology and Evolution* **11** , 316-324.

Yusof, J.H., Flower, A.J. and Teo, C.G. (1994) Transmission of hepatitis B virus analyzed by conformation-dependent polymorphisms of single-stranded viral DNA. *Journal of Infectious Diseases* **169**, 62-67.

Zuckerman, A.J. (1999*a*) More than third of world's population has been infected with hepatitis B virus. *British Medical Journal* **318**, 1213.

Zuckerman, A.J. (1999*b*) The acronym TTV. *Lancet* **353**, 932.

Zuckerman, A.J. and Zuckerman, J.N. (1999) Molecular epidemiology of hepatitis B virus mutants. *Journal of Medical Virology* **58**, 193-195.

Zuckerman, J.N., Sabin, C., Craig, F.M., Williams, A. and Zuckerman, A.J. (1997) Immune response to a new hepatitis B vaccine in healthcare workers who had not responded to standard vaccine: randomised double blind dose-response study. *British Medical Journal* **314**, 329-333.

Zuckerman, M.A., Hawkins, A.E., Briggs, M., Waite, J., Balfe, P., Thom, B., Gilson, R.J. and Tedder, R.S. (1995) Investigation of hepatitis B virus transmission in a health care setting: application of direct sequence analysis. *Journal of Infectious Diseases* **172**, 1080-1083.

Communicable Disease Surveillance Centre (1999) Acute hepatitis B in injecting drug users is increasing as initiative to improve vaccine coverage begins. *Communicable Disease Report (CDR) Weekly* **9**, 183-186.

Department of Health (1996) Unlinked anonymous HIV seroprevalence monitoring programme in England and Wales: Report from the unlinked anonymous surveys steering group. Data to the end of 1995.

Department of Health (2000) Health Service Circular: hepatitis B infected health care workers (HSC 2000/020).

General Medical Council (2000) Confidentiality: protecting and providing information. (http://www.gmc-uk.org).

Hepatitis B Outbreak Investigation Team (2000) An outbreak of hepatitis B associated with reusable subdermal electroencephalogram electrodes. *Canadian Medical Association Journal* **162**, 1127-1131.

Medical Research Council (2000) Human tissue and biological samples for use in research: operational and ethical guidelines (http://www.mrc.ac.uk).

Royal College of Pathologists (1999) Consensus statement of recommended policies for uses of human tissue in research education and quality control, with notes reflecting UK law and practices prepared by a working party of the Royal College of Pathologists and the Institute of Biomedical Science. Royal College of Pathologists, London.

Scottish Centre for Infection and Environmental Health (1999) Surgeon to patient transmission of hepatitis B. Weekly report **33**, 149.

The Incident Investigation Teams and Others (1997) Transmission of hepatitis B to patients from four infected surgeons without hepatitis B e antigen. *New England Journal of Medicine* **336**, 178-184.

WHO (1992) Expanded programme on immunisation global advisory group. *Weekly Epidemiology Record* **67**, 11-16.

Journal of General Virology (2000), **81**, 2273–2279. Printed in Great Britain

Characterization of a highly divergent TT virus genome

R. L. Hallett,[1] J. P. Clewley,[1] F. Bobet,[1] P. J. McKiernan[2] and C. G. Teo[1]

[1] Hepatitis and Retrovirus Laboratory, Central Public Health Laboratory, 61 Colindale Avenue, London NW9 5HT, UK
[2] The Liver Unit, Birmingham Children's Hospital, Whittal Street, Birmingham B4 6NH, UK

A novel TT virus (TTV)-like DNA sequence was detected in the serum of a patient (PM) with acute non-A–E hepatitis. The full-length genome sequence, referred to here as PM virus (PMV), was obtained and its relationship to other full or near full-length TTV sequences examined. Although it shares a common genomic arrangement and short conserved regions, the majority of the genome is extremely divergent, displaying an average genetic distance of 0·60 from all other TTV sequences. By comparing PMV with TTV genomes representing the most divergent types so far described, six major groups can be distinguished. The level of genetic diversity seen between these genomes is higher than would be expected within a single virus species. Indeed, PMV could be considered the prototype of an independent taxonomic group within the *Circoviridae* family. A genoprevalence study of sera from blood donors and patients with acute hepatitis suggests that PMV is rare.

Introduction

Efforts continue to isolate agents responsible for viral hepatitis that cannot be attributed to hepatitis viruses A to E. As part of these investigations, DNA from a novel virus was identified in the serum of a patient with post-transfusion hepatitis of unknown aetiology (Nishizawa *et al.*, 1997). The virus was named TT virus, and has subsequently been characterized as an unenveloped single-stranded DNA virus with a caesium chloride density of 1·31–1·34 g/cm^3 (Mushahwar *et al.*, 1999). Several complete TTV DNA sequences have now been published, revealing a circular genome which ranges from 3808 to 3852 nucleotides in length (Erker *et al.*, 1999; Okamoto *et al.*, 1999*b*; Hijikata *et al.*, 1999). Nuclease protection assays suggest that the genome is negative-stranded (Mushahwar *et al.*, 1999).

TTV was initially reported to share some physical characteristics with the *Parvoviridae*, although no nucleotide similarity was demonstrated. A resemblance between TTV and chicken anaemia virus (CAV), a circovirus, has been described by Takahashi *et al.* (1998*a*) on the basis of arginine-rich regions in the open reading frames (ORFs) from both viruses, and on similarities in their genomic arrangements (Miyata *et al.*, 1999).

Author for correspondence: Rachel Hallett.
Fax +44 20 8200 1569. e-mail rachel.hallett@virgin.net

The GenBank accession number of the sequence reported in this paper is AF261761.

However, no significant sequence similarity to any other member of the *Circoviridae* has been shown, and this has led Mushahwar *et al.* (1999) to propose that TTV be placed in a new virus family called *Circinoviridae*. Recently, three novel human DNA virus sequences (CBD203, CBD231 and CBD279) have been described that appear to be intermediately related to TTV and CAV (Takahashi *et al.*, 2000). They have been named TTV-like mini virus (TLMV). It is not clear from these reports whether TTV, CAV and TLMV belong to one family of viruses.

Considerable genetic variability of TTV has been demonstrated, with a 222 bp fragment of the genome from the longest ORF (ORF-1) showing as much as 65% divergence between sequences. Several groups have assigned up to 16 genotypes based on the sequence variability in short TTV PCR products (Tanaka *et al.*, 1998; Okamoto *et al.*, 1999*b*). In contrast, phylogenetic analysis of full-length or near full-length TTV sequences reveals division into three main types, represented by the prototype TA278 in group 1, US35 in group 2 and JA10 in group 3 (Erker *et al.*, 1999). Most of the sequence variability occurs in the coding regions. In addition to these three TTV groups there are two other highly divergent complete sequences, TUS01 and SANBAN (Okamoto *et al.*, 1999*b*; Hijikata *et al.*, 1999).

In this study we have sequenced the full genome of another TTV-like virus, named PMV, with a view to characterizing its genetic organization and to ascertaining its taxonomic status. We have also conducted a survey of its prevalence.

Methods

■ **Specimens.** Serum samples referred to our laboratory from patients with acute non-A–E hepatitis and UK blood donors were examined for the presence of TTV DNA by PCR. In some cases of acute non-A–E hepatitis, archived liver samples were also investigated.

■ **DNA extraction.** DNA was extracted from serum (100 µl volumes) by a guanidinium thiocyanate–silica method (Boom *et al.*, 1990). DNA was also extracted from paraffin-embedded liver biopsy samples by sequential treatments with Tween 20, proteinase K, Chelex-100 and chloroform, as described by Coombs *et al.* (1999).

■ **PCR.** A semi-nested PCR was used to amplify part of ORF-1 as previously described (Okamoto *et al.*, 1998). A PCR using primers located within the conserved non-coding region (set B; Leary *et al.*, 1999) was also used. All positive PCR products were gel-purified and sequenced as described below.

■ **PCR product cloning and sequencing.** PCR products were recovered from 2% agarose gels after staining with ethidium bromide and visualization under UV, and were purified with an Igenie DNA Extraction Kit (Helena BioSciences). They were sequenced in both directions using an ABI Prism DNA Sequencing Kit and an ABI 373 automated sequencer (PE Applied Biosystems). ORF-1 PCR products were cloned using an Invitrogen TOPO TA Cloning Kit and colony PCRs performed using the TTV-specific inner primers NG061 and NG063 (Okamoto *et al.*, 1998). Ten colony PCR products from each sample were sequenced.

■ **PMV genome sequencing.** PCR products covering the whole genome were generated using an Expand High Fidelity PCR System (Roche Diagnostics) according to the manufacturer's protocol for amplification of products up to 3 kb, and with a Clontech Advantage-GC 2 PCR kit. All PCRs were carried out in a 50 µl volume. First or single round PCRs contained 10 µl of extracted DNA as template; second round PCRs used 2 µl of first round product. The PCR primers used to generate the entire PMV genome sequence were: for fragment A, T801 and NG063 (Takahashi *et al.*, 1998*b*; Okamoto *et al.*, 1998); for fragment B, NG059, NG061 and BR1 (Okamoto *et al.*, 1998; Leary *et al.*, 1999); for fragment C, first round sense INV1 (5′ CCTTACAGACACCCCTTACTACCCT), first round antisense INV2 (5′ CAGTGGCACTTTCCTTTTCTTTC), and second round sense INV3 (5′ ACTAAGCACTCCGAGCGAAGC) with second round antisense INV4 (5′ ATAACCCTAAGAGCCTTGCCCATAG).

The positions of the three PCR products are indicated in Fig. 1. All three were cloned using an Invitrogen TOPO TA Cloning Kit. Plasmid DNA containing the inserted PCR products was purified using a QIAfilter Plasmid Midi Kit (Qiagen) and sequencing was carried out by Cambridge Bioscience. The resulting three fragments were assembled using programs EditSeq and MegAlign in the Lasergene Navigator package (Dnastar) to give the complete genome sequence of 3736 nucleotides.

■ **Sequence analysis.** Sequences contained in the GenBank/EMBL and SWISS-PROT databases were searched using BLASTN, BLASTP and FASTA (Altschul *et al.*, 1997; Pearson & Lipman, 1988). DNA and protein alignments were carried out using the ClustalW algorithm (Thompson *et al.*, 1994). Pairwise distances were calculated and phylogenetic analysis performed using programs in the Phylip package (Seqboot, DNAdist and Fitch, Neighbor and DNAml) (Felsenstein, 1993). Tree diagrams were generated using Treeview (Page, 1996). Secondary structure predictions of the non-coding regions in the PMV genome were made using the GCG program Mulfold (Jaeger *et al.*, 1989). By using the GCG program Plotsimilarity, with a sliding window of 50 nucleotides, sequence similarity was plotted across a multiple alignment of PMV and six full or near full-length TTV genomes representing the most divergent genetic types. This analysis was also carried out on the amino acid sequences of ORFs 1 and 2 using the same genomes and a sliding window of 20 amino acids.

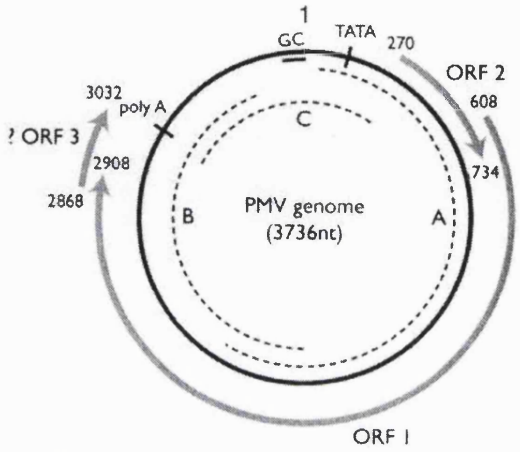


Fig. 1. Putative structure of the PMV genome. Dotted lines (A, B and C) show the positions of the three PCR products that were assembled to provide the complete genome sequence. Shaded arrows indicate the conserved ORFs, with the numbers showing their nucleotide positions. Other conserved characteristics are a TATA box, polyadenylation signal (polyA) and GC-rich region (GC).

Results

Initial detection and cloning

The initial TTV PCR, carried out on serum from a 4-year old girl (PM), revealed a 222 bp fragment of DNA from ORF-1 that was highly divergent from all other equivalent TTV sequences deposited in GenBank. This sequence was recovered from eight of ten colonies of cloned PCR products (the remaining two colonies contained one point mutation in each), and also from an archived liver sample from the same patient. This UK patient had acute non-A–E hepatitis and coagulopathy. She had no risk factors for bloodborne infection or travel outside the country.

Genome analysis

The entire genome of PMV was sequenced by producing three overlapping PCR products, as shown in Fig. 1. PCR products A and B were generated with standard long-range amplification conditions. However, PCR product C was not amplifiable this way due to the presence of a GC-rich region in the template DNA, and a PCR under conditions specific for GC templates was necessary to amplify this fragment (see Methods). A database homology search was carried out using the nucleotide sequence of the PMV genome. It did not reveal any significant nucleotide similarity between the PMV sequence and other sequences, other than TTV sequences.

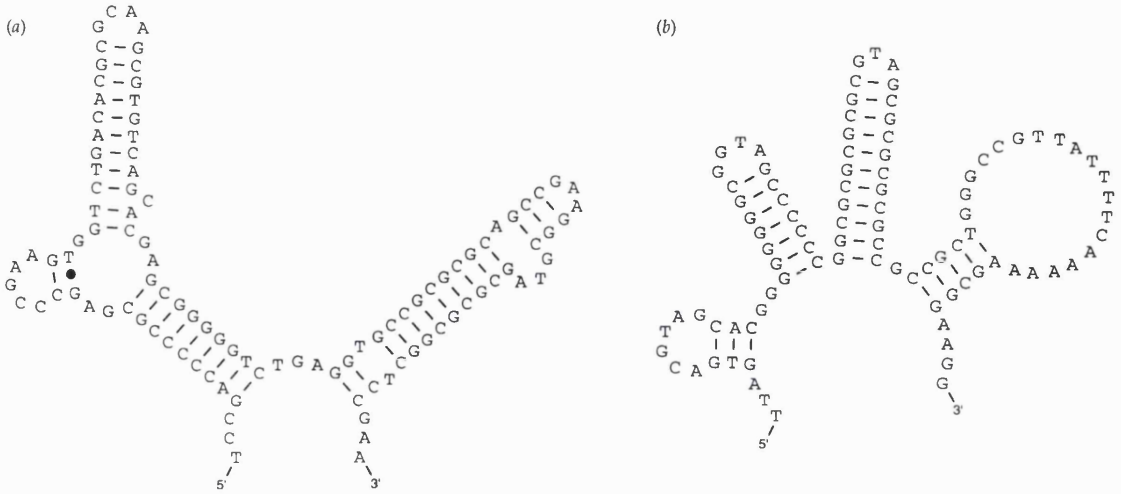


Fig. 2. Predicted secondary structure of two areas of the PMV non-coding region. The genomic strand (negative sense) was used in the analysis: (*a*) region immediately downstream of the polyadenylation signal (nucleotides 3133–3237); (*b*) truncated GC-rich region (nucleotides 3656–13).

In a comparison with full or near full-length TTV genomes it was found that PMV shared characteristics, including the positions of ORFs 1 and 2 and the non-coding regions, a GC-rich string of bases and a TATA box (Fig. 1). A polyadenylation signal (AATAAA), starting at position 3028 in PMV, was found in all full and near full-length genomes with the exception of SANBAN (Hijikata *et al.*, 1999).

Open reading frames

The longest ORF, ORF-1, of the PMV genome encodes 767 amino acids and has an average amino acid identity of only 31·1% with the corresponding ORFs of other TTVs. It has an N-terminal arginine-rich region, noted previously in other TTVs (Takahashi *et al.*, 1998*a*). When the amino acid sequence of ORF-1 was used to BLAST search the SWISS-PROT database, 37% identity to a cuttlefish protamine protein (PRT2 SEPOF) was revealed.

A shorter ORF, ORF-2, was found in PMV, but again there was a high degree of sequence divergence from other TTVs, the average amino acid identity being 32·7%. The sequence motif (ACCATGG) containing the initiation codon of ORF-1 in PMV had a single nucleotide difference from that described by Kozak (1986) as the optimal sequence for translation initiation. The ORF-2 initiation motif had two suboptimal nucleotides. Conserved Sp1 (GGCGGG) and Cap (CAATTC) sites, immediately upstream of the ORF-2 initiation codon, were also present in the PMV genome (Hijikata *et al.*, 1999).

The putative ORF-3 proposed by Erker *et al.* (1999) was 165 nucleotides long in PMV. However, it overlapped with the C-terminal end of ORF-1 by 40 nucleotides, a feature not seen in TTV genome sequences so far described. Erker *et al.* (1999) reported ORF-3 to be highly conserved, but when the full or near full-length TTV sequences from their study were analysed with sequences TUS01, SANBAN and PMV, ORF-3 was revealed to be variable. The average amino acid sequence identity between PMV and these ORF-3 sequences was 46·3%. The Kozak initiation motif was not identified at the start codon in PMV.

Non-coding region

A conserved non-coding region was found to stretch from the polyadenylation signal to the initiation codon of ORF-2, encompassing the GC-rich region. Comparisons were carried out between PMV and five full-length TTV sequences (TA278, JA20, US35, TUS01 and SANBAN). For four of these, the non-coding region was about 1040 nucleotides long; in TUS01 it was shorter at 1014 nucleotides. The equivalent region in PMV was only 972 nucleotides long, the most noticeable truncation being in the GC-rich region, which was 58 nucleotides in length compared to 117 in the prototype TTV genome TA278 (Miyata *et al.*, 1999). The average nucleotide identity between PMV and the five analysed sequences in this region was 69·8%.

Hijikata *et al.* (1999) proposed that the domain immediately downstream of the polyadenylation signal in the non-coding region of TTV forms a stem–loop structure. The corresponding region in PMV, analysed in this way, predicted a secondary structure that was very similar to the TTV structure (Fig. 2*a*). By contrast, the secondary structure of the GC-rich region of the PMV genome is likely to be affected by deletions: Fig. 2(*b*) shows how the GC-rich region of PMV might be different from the GC-rich stretch in isolates TA278 and TUS01 (Okamoto *et al.*, 1999*b*).

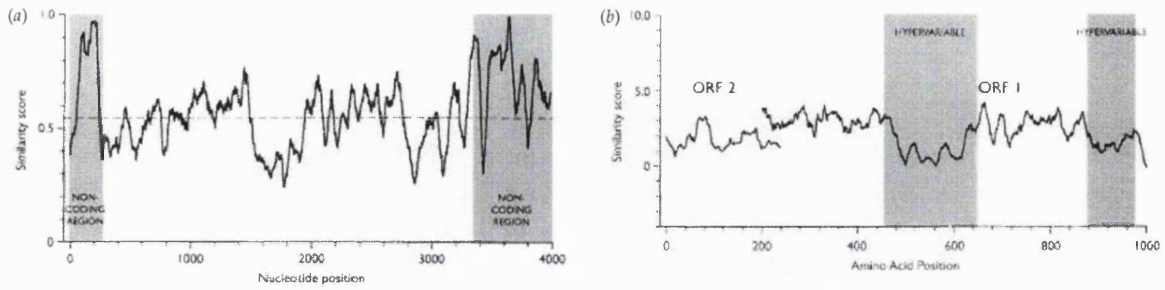
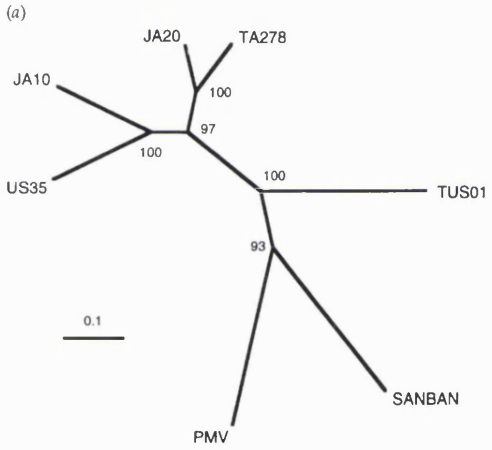


Fig. 3. (*a*) Similarity plot of an alignment of the PMV genome sequence with six representative TTV genomes. The average nucleotide sequence identity is indicated by the dashed line. (*b*) Amino acid similarity plot of an alignment of ORFs 1 and 2 from the same seven sequences. The bars denote hypervariable regions within ORF-1. The six genomes used are TA278, JA20, US35, JA10, TUS01 and SANBAN (Miyata *et al.*, 1999; Erker *et al.*, 1999; Okamoto *et al.*, 1999*b*; Hijikata *et al.*, 1999).



(*b*)

	TA278	JA20	US35	JA10	TUS01	SANBAN	PMV
TA278							
JA20	0.17						
US35	0.35	0.33					
JA10	0.35	0.34	0.34				
TUS01	0.54	0.53	0.58	0.57			
SANBAN	0.64	0.62	0.68	0.63	0.58		
PMV	0.57	0.57	0.65	0.60	0.63	0.58	

Fig. 4. (*a*) A neighbour-joining phylogenetic tree demonstrating the genetic relatedness between PMV and six representative full or near full-length TTV genomes, with (*b*) the corresponding genetic distances. Bootstrap percentages from 1000 replicates are shown on the branches. The scale bar indicates the proportion of nucleotides substituted for a given branch length.

Relationship to other TTV-like genomes

The similarity of PMV to the overlapping part of the six representative sequences is shown in Fig. 3(*a*), which reveals regions of conservation and variability across the 3·7 kb studied. The average nucleotide sequence identity was 55% (by Plotsimilarity). The amino acid similarity plot for ORF-1 revealed hypervariable regions between amino acid positions 300 and 440, and between 710 and 760 in the multiple alignment. For ORF-2, the amino acid plot was smoother, indicating that a high level of variability occurs across the whole ORF, rather than being concentrated into hypervariable zones.

The phylogenetic tree and corresponding genetic distances produced from a multiple alignment of PMV with the six representative TTV sequences are shown in Fig. 4. It is clear that the sequences of TUS01, SANBAN and PMV are highly divergent from the prototype TTV sequence, TA278, and from each other. Six major groupings were observed, consisting of TA278 and JA20 as one group, with the other five sequences each representing a separate group. The average genetic distance between each group was 0·48. Strong support for these groupings was obtained from bootstrap resampling.

Phylogenetic sequence comparisons were also made between PMV, TTV, and complete CAV and TLMV sequences (CAE26P4 and CBD279 respectively), by first carrying out individual pairwise alignments to identify homologous regions, and then performing a multiple alignment using gap-stripped 2·5 kb sequences. Fig. 5 shows that there is one cluster of sequences that includes the TTVs (TA278, JA20, JA10 and US35), and five other sequences (PMV, TUS01, SANBAN, TLMV and CAV), each of which is divergent from the others.

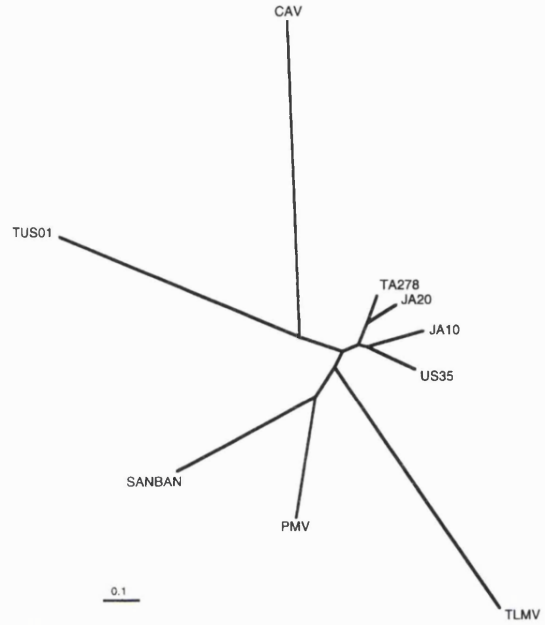


Fig. 5. A neighbour-joining phylogenetic tree showing the relationship between 2·5 kb homologous regions of TTV genomes, a CAV genome (CAE26P4) and a TLMV genome (CBD279). A multiple alignment was made of the homologous regions to the CAV genome, which were identified by individual pairwise alignments with CAV. The alignment was used in DNAdist and Neighbor with a transition/transversion ratio of 1·1. The scale bar indicates the proportion of nucleotides substituted for a given branch length.

PMV was about equally closely related to TA278 and JA20, and to SANBAN.

Prevalence of PMV

PMV-like sequences were searched for in 197 sera belonging to two patient groups, UK blood donors ($n = 99$) and acute non-A–E hepatitis patients ($n = 98$). A PCR assay with set B primers situated in the non-coding region was used to amplify TTV DNA (Leary *et al.*, 1999). Cloned PMV DNA (fragment B, Fig. 1) was used as a positive control. A total of 49 (17 blood donors and 32 hepatitis patients) were positive and, of these, 36 gave readable DNA sequence data. Similar to TTV, but not PMV. On alignment with the set B PCR product from PMV it was seen that PMV remained phylogenetically unique, with the most closely related sequences showing around 10% divergence.

Discussion

The circular genome of a novel TTV-like virus, PMV, has been characterized. PMV shares its genome organization and a moderately conserved non-coding region with other TTV sequences. Its genome also contains the typical motifs identified in other TTVs. It is the shortest TTV-like genome reported so far, the largest deletions being in the GC-rich region. The similarity plot of an alignment of PMV and the six most divergent TTV sequences shows a high degree of sequence heterogeneity, with genetic distances between them varying from 0·17 to 0·68. Sequence variation is more common in the coding regions, which may lead to major antigenic differences between the viruses. The protein encoded by the ORF-1 of PMV was homologous to a cuttlefish protamine protein that is involved in the folding and packaging of sperm DNA in the nucleus. This is consistent with the ORF-1 product being a virion capsid protein that enwraps the viral DNA genome. Nishizawa *et al.* (1999) and Erker *et al.* (1999) have previously proposed that ORF-1 of TTV encodes the viral capsid protein on the basis of its N-terminal arginine-rich region, which is shared with the circovirus capsid proteins. The functions of the proteins encoded by ORF-2 and ORF-3 are not known.

The non-coding region is relatively conserved, suggesting that it plays an important regulatory role in virus replication. The effect that the truncated GC-rich stretch may have on the replication of PMV is yet to be determined.

Representative full-length and near full-length genomes of TTV-like viruses fall into six major groups on phylogenetic analysis, with an average genetic distance of 0·48 between them (Fig. 4). The groupings observed on analysis of the overlapping nucleotide sequences are also seen when the amino acid sequences of ORF-1 are aligned (data not shown). The six groups correspond to the previously defined groups 1a (the prototype TA278) and 1b (JA20), 2 (US35) and 3 (JA10) (Erker *et al.*, 1999), with TUS01, SANBAN and PMV forming three new branches. Following the nomenclature of Erker *et al.* (1999), TUS01, SANBAN and PMV could be considered new 'genotypes'. However, the level of genetic divergence within the TTV cluster is higher than would be expected if all these sequences were from viruses of the same species. Several groups have postulated a genotyping system based on the phylogenetic analysis of short PCR products (Tanaka *et al.*, 1998; Okamoto *et al.*, 1999*a*; Mulyanto *et al.*, 2000). The analysis presented here, using full and near full-length genomes, suggests that until the full extent of divergence within the TTVs is known, it is premature to classify short TTV-like sequences in this way. The possibility that further divergent genomes remain to be discovered cannot be excluded.

Several groups have reported short stretches of sequence similarity between TTV and CAV (Miyata *et al.*, 1999; Hijikata *et al.*, 1999). Furthermore, TLMV was recently discovered in human serum and reported as an intermediate relative of TTV and CAV (Takahashi *et al.*, 2000). In order to further assess the relationships between these viruses, we compared representative TTV sequences, a CAV genome and a TLMV genome (Fig. 5). Phylogenetic analysis shows that six lineages can be

discerned: the lineage to which the TTV cluster of groups 1, 2 and 3 belongs, and five others (PMV, SANBAN, TLMV, TUS01 and CAV).

The *Circoviridae* has recently been divided into two genera: *Gyrovirus*, containing CAV; and *Circovirus*, with porcine circovirus (PCV) and beak and feather disease virus (BFDV) as members (Pringle, 1999). Although there is very little sequence similarity between them, individual isolates of each virus are moderately conserved; for example, the maximum divergence shown between six full-length CAV genomes, available in GenBank, was 5% (not shown). By this criterion, TUS01, SANBAN, PMV and TLMV individually, and TA278, JA20, US35 and JA10 as a cluster, should have equal taxonomic rank with each other and with the *Gyrovirus* and *Circovirus* genera of *Circoviridae*.

Although TTV was initially considered a hepatitis virus, this is not yet substantiated. PMV may be hepatotropic as its sequence was also detected in liver, but since a pre-hepatitis serum specimen from the patient was not available we were unable to determine if PMV infection was associated with onset of disease. It is noted that circoviruses showing less sequence divergence than the TTV group can cause different clinical syndromes. For example, a strain of PCV, reported to be associated with a disease in pigs known as postweaning multisystemic wasting syndrome, shows only 26% divergence from the genome of the prototype PCV, which is common in domestic herds and apparently does not cause disease (Morozov *et al.*, 1998).

Further study is required to unravel the precise evolutionary and taxonomic connections between the circular single-stranded genomes examined here. While PMV may be unique genetically and might be hepatotropic, it appears to be a rare virus in humans. The emerging pattern of infection with TTV-like viruses is that of a collection of related but different viruses, with varying prevalences. Their pathogenic potential has yet to be elucidated.

We thank Dr Jim Gray for provision of serum and archived liver specimens.

References

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**, 3389–3402.

Boom, R., Sol, C. J. A., Salimans, M. M. M., Jansen, C. L., Wertheim-van Dillen, P. M. E. & van der Noordaa, J. (1990). Rapid and simple method for purification of nucleic acids. *Journal of Clinical Microbiology* **28**, 495–503.

Coombs, N. J., Gough, A. C. & Primrose, J. N. (1999). Optimization of DNA and RNA extraction from archival formalin-fixed tissue. *Nucleic Acids Research* **27**, 12–14.

Erker, J. C., Leary, T. P., Desai, S. M., Chalmers, M. L. & Mushahwar, I. K. (1999). Analyses of TT virus full-length genomic sequences. *Journal of General Virology* **80**, 1743–1750.

Felsenstein, J. (1993). PHYLIP (phylogeny inference package) v3.5c. Department of Genetics, University of Washington, Seattle, USA.

Hijikata, M., Takahashi, K. & Mishiro, S. (1999). Complete circular DNA genome of a TT virus variant (isolate name SANBAN) and 44 partial ORF2 sequences implicating a great degree of diversity beyond genotypes. *Virology* **260**, 17–22.

Jaeger, J. A., Turner, D. H. & Zuker, M. (1989). Improved predictions of secondary structures for RNA. *Proceedings of the National Academy of Sciences, USA* **86**, 7706–7710.

Kozak, M. (1986). Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* **44**, 283–292.

Leary, T. P., Erker, J. C., Chalmers, M. L., Desai, S. M. & Mushahwar, I. K. (1999). Improved detection systems for TT virus reveal high prevalence in humans, non-human primates and farm animals. *Journal of General Virology* **80**, 2115–2120.

Miyata, H., Tsunoda, H., Kazi, A., Yamada, A., Khan, M. A., Murakami, J., Kamahora, T., Shiraki, K. & Hino, S. (1999). Identification of a novel GC-rich 113-nucleotide region to complete the circular, single-stranded DNA genome of TT virus, the first human circovirus. *Journal of Virology* **73**, 3582–3586.

Morozov, I., Sirinarumitr, T., Sorden, S. D., Halbur, P. G., Morgan, M. K., Yoon, K.-J. & Paul, P. S. (1998). Detection of a novel strain of porcine circovirus in pigs with postweaning multisystemic wasting syndrome. *Journal of Clinical Microbiology* **36**, 2535–2541.

Mulyanto, M., Hijikata, M., Matsushita, M., Ingkokusmo, G., Widjaya, A., Sumarsidi, D., Kanai, K., Ohta, Y. & Mishiro, S. (2000). TT virus (TTV) genotypes in native and non-native prostitutes of Irian Jaya, Indonesia: implication for non-occupational transmission. *Archives of Virology* **145**, 63–72.

Mushahwar, I. K., Erker, J. C., Muerhoff, A. S., Leary, T. P., Simons, J. N., Birkenmeyer, L. G., Chalmers, M. L., Pilot-Matias, T. J. & Desai, S. M. (1999). Molecular and biophysical characterization of TT virus: evidence for a new virus family infecting humans. *Proceedings of the National Academy of Sciences, USA* **96**, 3177–3182.

Nishizawa, T., Okamoto, H., Konishi, K., Yoshizawa, H., Miyakawa, Y. & Mayumi, M. (1997). A novel DNA virus (TTV) associated with elevated transaminase levels in posttransfusion hepatitis of unknown etiology. *Biochemical and Biophysical Research Communications* **241**, 92–97.

Nishizawa, T., Okamoto, H., Tsuda, F., Aikawa, T., Sugai, Y., Konishi, K., Akahane, Y., Ukita, M., Tanaka, T., Miyakawa, Y. & Mayumi, M. (1999). Quasispecies of TT virus (TTV) with sequence divergence in hypervariable regions of the capsid protein in chronic TTV infection. *Journal of Virology* **73**, 9604–9608.

Okamoto, H., Akahane, Y., Ukita, M., Fukuda, M., Tsuda, F., Miyakawa, Y. & Mayumi, M. (1998). Fecal excretion of a nonenveloped DNA virus (TTV) associated with posttransfusion non A-G hepatitis. *Journal of Medical Virology* **56**, 128–132.

Okamoto, H., Takahashi, M., Nishizawa, T., Ukita, M., Fukuda, M., Tsuda, F., Miyakawa, Y. & Mayumi, M. (1999*a*). Marked heterogeneity and frequent mixed infection of TT virus demonstrated by PCR with primers from coding and noncoding regions. *Virology* **259**, 428–436.

Okamoto, H., Nishizawa, T., Ukita, M., Takahashi, M., Fukuda, M., Iizuka, H., Miyakawa, Y. & Mayumi, M. (1999*b*). The entire nucleotide sequence of a TT virus isolate from the United States (TUS01): comparison with reported isolates and phylogenetic analysis. *Virology* **259**, 437–448.

Page, R. D. M. (1996). Treeview: an application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences* **12**, 357–358.

Pearson, W. R. & Lipman, D. J. (1988). Improved tools for biological sequence comparison. *Proceedings of the National Academy of Sciences, USA* **85**, 2444–2448.

Pringle, C. L. (1999). Virus taxonomy at the XIth International Congress of Virology, Sydney, Australia, 1999. *Archives of Virology* **144**, 2065–2068.

Takahashi, K., Ohta, Y. & Mishiro, S. (1998*a*). Partial ~ 2·4-kb sequences of TT virus (TTV) genome from eight Japanese isolates: diagnostic and phylogenetic implications. *Hepatology Research* **12**, 111–120.

Takahashi, K., Hoshino, H., Ohta, Y., Yoshida, N. & Mishiro, S. (1998*b*). Very high prevalence of TT virus (TTV) infection in general population of Japan revealed by a new set of PCR primers. *Hepatology Research* **12**, 233–239.

Takahashi, K., Iwasa, Y., Hijikata, M. & Mishiro, S. (2000). Identification of a new human DNA virus (TTV-like mini virus: TLMV) intermediately related to TT virus and chicken anaemia virus. *Archives of Virology* (in press).

Tanaka, Y., Mizokami, M., Orito, E., Ohno, T., Nakano, T., Kato, T., Kato, H., Mukaide, M., Park, Y.-M., Kim, B.-S. & Ueda, R. (1998). New genotypes of TT virus (TTV) and a genotyping assay based on restriction fragment length polymorphism. *FEBS Letters* **437**, 201–206.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22**, 4673–4680.

Received 15 May 2000; Accepted 19 June 2000

Molecular epidemiology of a large outbreak of hepatitis B linked to autohaemotherapy

George J M Webster, Rachel Hallett, Simon A Whalley, Margie Meltzer, Koye Balogun, Dave Brown, Conor P Farrington, Sukhdev Sharma, Gervase Hamilton, Stephen C Farrow, Mary E Ramsay, Chong-Gee Teo, Geoffrey M Dusheiko

Summary

Background Unregulated skin-piercing procedures potentially facilitate the transmission of bloodborne pathogens. In February, 1998, a patient who had recently received autohaemotherapy at an alternative medicine clinic in the UK was diagnosed with acute hepatitis B. The autohaemotherapy procedure involved the drawing of 1 mL of the patient's blood, mixing with saline, and reinjection of the autologous blood mixture. We investigated the extent of hepatitis B virus (HBV) infection in patients and staff of the clinic.

Methods Patients who had attended the clinic between January, 1997, and February, 1998, were tested for serological markers of HBV, and for HBV DNA by PCR. HBV DNA was sequenced to assess the relatedness of the virus identified in the cases. We analysed the number and dates of visits with regard to HBV status.

Findings Serum samples were received from 352 patients and four staff members. Serological evidence of exposure to HBV was found in 57 (16%). Of the 33 patients and staff who were positive for hepatitis B surface antigen, 30 (91%) showed complete nucleotide identity in the DNA segments derived from the surface and core genes. Five patients with linked infection had markers of chronic hepatitis B, and one of these was regarded as the likely source of the outbreak. The attack rate was associated with the number of visits ($p<0.0001$) and the week of visit ($p=0.011$). Contaminated saline in a repeatedly used bottle was the probable vehicle of transmission.

Interpretation We have described a large community-based outbreak of hepatitis B due to transmission by a single HBV variant. Our findings emphasise the continuing risk of transmission of bloodborne viruses in all health-care settings where skin-piercing procedures are used.

Lancet 2000; **356:** 379–84

Centre for Hepatology, Department of Medicine, Royal Free Campus, Royal Free and University College Medical School, London NW3 2PF, UK (G J M Webster MRCP, S A Whalley MRCP, D Brown MPhil, Prof G M Dusheiko FRCP); **Public Health Laboratory Service Hepatitis and Retrovirus Laboratory** (R Hallett BSc, C-G Teo FRCPath); **Public Health Laboratory Service Communicable Disease Surveillance Centre** (K Balogun PhD, G Hamilton FFPHM, C P Farrington PhD, M E Ramsay FFPHM); **and Barnet Health Authority, London, UK** (M Meltzer FRCPath, S Sharma MFPHM, S C Farrow FFPHM)

Correspondence to: Prof Geoffrey M Dusheiko (e-mail: g.dusheiko@rfc.ucl.ac.uk)

Introduction

In countries of high endemicity, hepatitis B virus (HBV) infection is a major cause of morbidity and mortality,[1] but the incidence of acute hepatitis B is low in the UK[2] and USA.[3] Most infected adults develop an acute symptomatic hepatitis followed by loss of circulating hepatitis B surface antigen (HBsAg). However, about 5% of adults and 90% of infants do not become HBsAg-negative after infection. Most sporadic infections in non-endemic regions occur after high risk behaviour, including unprotected sexual intercourse and injecting drug use. After the introduction of widespread screening of blood products, the immunisation of health-care workers, and the implementation of infection-control regulations,[4] hospital outbreaks in more developed countries are now infrequent. Community-acquired outbreaks of HBV infection due to skin-piercing procedures, such as tattooing[5] and acupuncture,[6,7] were reported in the 1970s and 1980s, and in the UK led to the introduction of licensing and hygiene regulations.[8] We report a large outbreak of acute hepatitis B linked to autohaemotherapy done at an alternative medicine clinic in the UK.

Methods

Background and epidemiological investigation

In February, 1998, a 43-year-old woman living in central England developed acute icteric hepatitis B. She had attended a clinic in London in October and November, 1997, where she had received a treatment known as autohaemotherapy. This technique involved the drawing of about 1 mL of the patient's blood with a needle and syringe, mixing of this blood with an equal volume of saline, then injection of the mixture into the buttocks or acupuncture points. After the case came to light, the Consultant in Communicable Disease Control (CCDC) with responsibility for the local authority in which the clinic was based advised the clinic staff to cease autohaemotherapy. Directors of the public health laboratories and CCDCs were sent an electronic message asking them to report any cases of hepatitis B possibly linked to this clinic. Sexual and household contacts of index cases were tested as recommended.[9] Through these case-finding procedures, seven further HBsAg-positive cases were identified over the next 4 weeks. An investigation was started to identify other cases arising from this clinic, to characterise the epidemiology of the outbreak, and to limit secondary spread of the disease. In March, 1998, a patient was referred to one of us by the North London Blood Transfusion Service, with evidence of hepatitis C virus (HCV) seroconversion. She had received autohaemotherapy at the clinic in March, 1997. As a result, the study was extended to include HCV.

Patients who attended the clinic between January, 1997, and February, 1998, were identified, either

through their clinic records, or after media coverage. A lookback investigation was carried out between February and September, 1998, by the local health authority and the Public Health Laboratory Service Communicable Disease Surveillance Centre. Each patient was asked to give a serum sample for serological testing, and to complete a questionnaire detailing the dates on which they received autohaemotherapy. Clinic staff were tested for markers of HBV and HCV infection.

Methods

Serum samples were tested for HBsAg by ELISA (Hepanostika Uni-Form II, Organon Teknika, Boxtel, Netherlands), and for total antibody against hepatitis B core antigen (HBc) with an in-house competitive radioimmunoassay. Samples with positive or equivocal results for HBsAg were tested for hepatitis B e antigen (HBeAg), antibody against HBe (Hepanostika HBe, Organon Teknika), IgM antibody to HBc (Murex Biotech, Dartford, UK), and antibody to HBsAg (IMX Abbott GmbH Diagnostika, Wiesbaden-Delkenheim, Germany). If negative for HBsAg, samples were tested for total antibody to HBc, IgM antibody to HBc, and antibody to HBsAg. Antibody to hepatitis delta virus (HDV; ETI-AB-DELTAK-2, Diasorin SrI, Saluggia, Italy) was tested for in all patients who were HBsAg-positive.

Patients who were HBsAg-positive were defined as having active HBV infection. Those who were also positive for IgM against HBc were defined as having acute infection, and those who were negative for IgM against HBc, or who remained HBsAg-positive for more than 6 months, were defined as having chronic hepatitis. Patients who were HBsAg-negative but positive for total antibody against HBc and HBsAg were defined as having a resolved HBV infection, as were those who were negative for antibody against HBsAg, but positive for antibody against HBc on sequential sampling. In the latter category, patients who were positive for IgM against HBc were defined as having a recently resolved infection. Patients were classed as having no serological markers of HBV infection if they were negative for HBsAg and antibody against HBc and HBsAg, or if they were negative for HBsAg and antibody to HBsAg but positive for antibody against HBc in a single specimen.

All samples were tested for HBV DNA after extraction of DNA from serum samples by the guanidium-thiocyanate-silica method.[10] Fragments from the hepatitis B core and surface genes were amplified by nested PCR.[11] The PCR products from all HBV-DNA-positive samples were sequenced with both sets of inner primers. Sequencing reactions were done with the ABI Prism DNA Sequencing kit, and reactions were electrophoresed on an ABI 373 automated sequencer (PE Applied Biosystems, Warrington, UK). The resulting sequences (197 nucleotides from the core gene and 404 nucleotides from the surface gene) were aligned by use of the CLUSTAL V algorithm in the Lasergene Program (DNASTAR Inc, USA).[12] HBV DNA from patients with acute hepatitis B, whose serum had been referred to the Hepatitis and Retrovirus Laboratory during 1997, was also sequenced and was used as a control. The genetic relatedness of aligned sequences was compared by phylogenetic analysis.[13] Serum HBV DNA was quantified with the AMPLICOR HBV MONITOR assay (Roche Diagnostic Systems, NJ, USA).

Samples were analysed for antibody to HCV with one of two ELISAs (MONOLISA anti-HCV PLUS, Sanofi Diagnostics Pasteur, Marnes-la-Coquette, France; and Ortho HCV 3.0, Ortho-Clinical Diagnostics, Neckargemund, Germany). A supplemental recombinant immunoblot assay (RIBA; Chiron RIBA HCV 3.0 SIA, Chiron Corporation, Emeryville, CA, USA), and repeat ELISAs were done where equivocal or positive results were found. Patients were defined as serologically positive for HCV if they were ELISA-positive and RIBA-positive. HCV RNA detection was carried out by reverse-transcription PCR. RNA was extracted from serum by use of the Roche AMPLICOR Extraction Kit. Amplification of the 5′ non-coding region and genotyping was done with a restriction fragment length polymorphism assay.[14]

All patients with serological markers of HBV or HCV infection were referred for clinical assessment. A history of acute hepatitis, and enquiry into other risk factors for exposure to viral hepatitis was sought. Household and sexual contacts of these patients were identified and counselled, and, when appropriate, offered serological testing and HBV vaccination.[9]

Statistical analysis

If the number of treatment visits was known, the relation between the proportion of patients with serological markers of HBV and the number of visits was assessed by means of a χ^2 test for trend. Trends over time in the probability of HBV infection per patient visit were investigated by use of data on patients with known visit dates. This analysis was complicated by the fact that many patients visited on several occasions; hence for these patients, the precise date of infection is not known. This analysis was therefore done with a likelihood method which allowed for multiple visits. In each week (w) of the study period, we assumed that all patients undergoing treatment in that week had the same probability (P_w) of acquiring infection. All patients were regarded as equally susceptible to infection at the outset, and the outcome for different patients was assumed to be independent. Given a patient's sequence of visits (eg, weeks $w_1, \ldots w_k$), the patient's probability of escaping infection is $\pi\ (1\text{-}P_{wi})$. The overall likelihood was obtained by multiplying the outcome probabilities for all the patients. The weekly infection probabilities, P_w, were estimated by maximisation of the likelihood. The hypothesis of a constant infection probability $P_w=P$ was tested by the likelihood ratio test. If information on the date of onset of clinical hepatitis was available, this was incorporated by exclusion of visits in weeks in which infection could not have occurred (ie, more than 6 months before onset of clinical disease).

Results

HBV infection

399 patients at the clinic between January, 1997, and February, 1998, were identified. Serum samples were received from 352 (216 women, 136 men; mean age 42 years, range 1–86 years). Including the original cases, 54 patients (15%) had evidence of HBV infection (figure 1). 23 had markers of resolved infection, six of whom had a recently resolved infection. 31 had evidence of active hepatitis B on initial blood testing. Antibodies against HDV were negative in all tested patients.

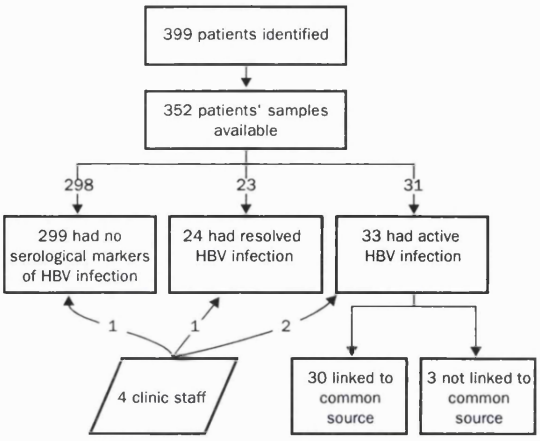


Figure 1: **HBV serological status at the time of initial blood testing in clinic staff and patients who received autohaemotherapy**

The clinic was staffed by four members of the same family, one of whom was a registered physician. Two staff members developed acute hepatitis B in March 1998, of whom one was also positive for antibodies against HCV, but HCV RNA negative. Another had resolved hepatitis B, and was positive for antibodies against HCV, but HCV RNA negative. The fourth staff member had no markers of HBV or HCV infection. The two staff members with acute hepatitis B cleared HBsAg after an icteric illness. Autohaemotherapy had been done by several of the staff members on each other during the previous year.

HBV DNA could be amplified and sequenced from 32 of the 33 HBsAg-positive patients and staff members. There was complete nucleotide identity in both the surface and core genes in 30 cases (91% of HBsAg-positive cases). The sequences had nucleotide motifs characteristic of variants belonging to genotype D. The variant found in these 30 linked cases was different from control samples from acutely infected patients, as shown by phylogenetic analysis (figure 2).

25 of the 30 cases linked by HBV DNA sequence analysis cleared circulating HBsAg during the first 6 months of assessment. Clinical hepatitis was documented in 23, and jaundice occurred in 16. Six of the 30 linked cases were identified in the preclinical phase of acute hepatitis, when they were HBsAg-positive, with high serum concentrations of HBV DNA, but normal or slightly raised concentrations of serum aminotransferases. All six patients developed a typical biochemical and serological pattern of acute hepatitis, followed by clearance of HBsAg.

Five of the patients with linked infection did not clear HBsAg over a 6-month period, and so were classified as having chronic hepatitis B. None had a history of acute clinical hepatitis, but three were positive for IgM against HBc at presentation. The remaining two patients were negative for this antibody, but positive for IgG against HBc when first tested. One was a 79-year-old man (patient 16), born in India, who was HBsAg-positive, HBeAg-positive, and with HBV DNA concentrations of $1{\cdot}2\times10^{10}$ copies/mL. The other was a 79-year-old woman (patient 18) with diabetes mellitus, also born in India, who remained HBsAg-positive and HBeAg-positive. Of the three patients with IgM against HBc at presentation who became chronically infected, two were taking oral steroid therapy. One was a 72-year-old woman (patient 6), who was taking prednisolone for cryptogenic organising pneumonitis, and who remained HBeAg-positive with very high HBV DNA concentrations (10^8–10^{10} copies/mL) and slightly raised concentrations of serum aminotransferases. The other was a 19-year-old woman (patient 11) who was being treated with prednisolone and azathioprine for a relapse of ulcerative colitis. Patient 22, an 18-year-old woman who had no other significant medical problems and took no regular medication, also failed to clear HBsAg.

Two of the HBsAg-positive patients in whom HBV-DNA was amplified (patients 24 and 27) were not linked by sequence analysis (figure 2). They had serological evidence of chronic hepatitis B (negative for IgM but positive for IgG against HBc), with no history of clinical acute hepatitis, and both were born in countries with increased endemic rates of HBV carriage. Consequently, there was no evidence to link definitively their infection to autohaemotherapy.

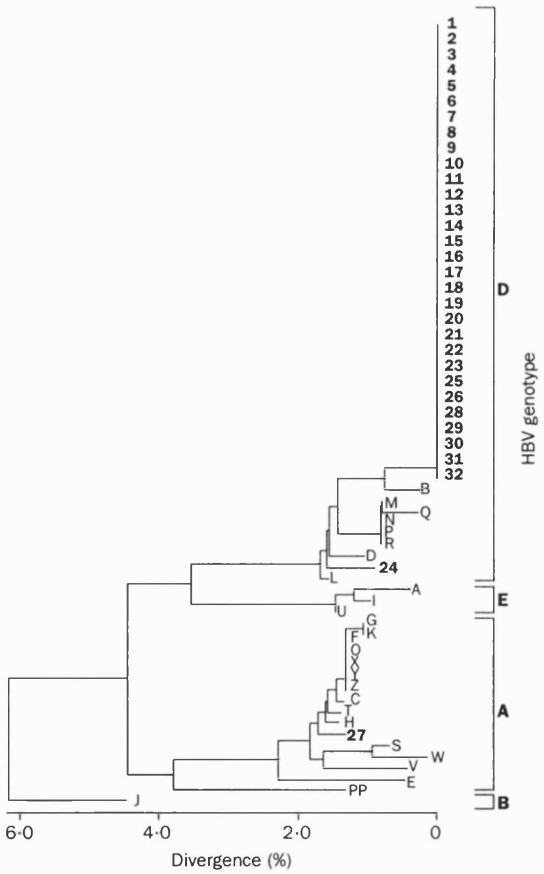


Figure 2: **Genetic relatedness of HBV identified in the outbreak**

Analysis of 66 nucleotide sequences from a 404 bp fragment of the HBV surface gene. Cases 1 and 2 are clinic workers; 3–32 are patients at the clinic. Sequences M, N, P, and R are from an outbreak of acute hepatitis B among injecting drug users. Sequences F, O, X, Y, and Z are from another outbreak involving tattoo clients, prison inmates, and injecting drug users. Other sequences are from single transmission incidents not known to be part of any outbreak. PP designates the sequence of the plasmid control used in PCR studies.

Treatments received	Number of patients	Active infection	Recently resolved infection	Past resolved infection	Serological markers of HBV infection
1–4*	194	6	3	12	21 (11%)
5–9	84	12	1	4	17 (20%)
10–14*	27	9	1	1	11 (41%)
15–29	10	3	1	0	4 (40%)
Total	315	30	6	17	53

*Including one patient with unlinked active infection.

Frequency of serological markers of hepatitis B virus infection in patients with a known number of autohaemotherapy treatments

Relation between autohaemotherapy and HBV infection

The dates of 1546 clinic visits in the period January, 1997, to February, 1998, were available for 315 patients (53 with markers of hepatitis B and 262 who were negative). The proportion of patients with markers of HBV infection increased with the number of treatments received (χ^2=18, p<0·0001; table). The proportion of patients with markers of HBV was higher among those treated in the months after June, 1997, than among those treated before this date (figure 3). Analysis by period of visit, taking account of multiple visits, showed a significant variation in the estimated probability of infection per visit by week of visit (p=0·011), which increased over the study period.

Except for the 23 patients with acute clinical hepatitis in whom the onset of clinical disease was known (figure 4), no other cases of acute clinical hepatitis B were reported in clinic attendees over the investigation period. We postulated that one of the patients with chronic hepatitis B and phylogenetically linked infection was likely to be the original source. Several factors implicated HBV transmission from patient 16, the 79-year-old man who only attended the clinic in June, 1997, as the source of the outbreak. No patient completing treatment in the 5 months before he attended in June, 1997, was infected with the common source virus or had a recent history of clinical hepatitis. All patients with linked active infection or recently resolved infection received autohaemotherapy at least once after patient 16 had attended. Assuming a maximum incubation period for hepatitis B of 6 months, none of the patients with known onset dates for clinical hepatitis had a predicted earliest date of infection before June, 1997 (figure 4). The hepatitis B virus characterised

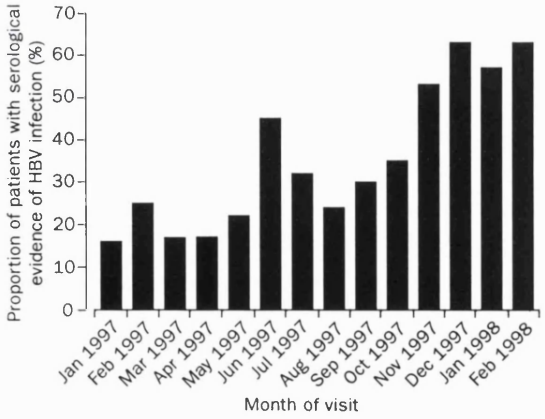


Figure 3: **Proportion of patients attending during each month who were subsequently found to have serological markers of HBV infection**

Patients may have attended in more than 1 month.

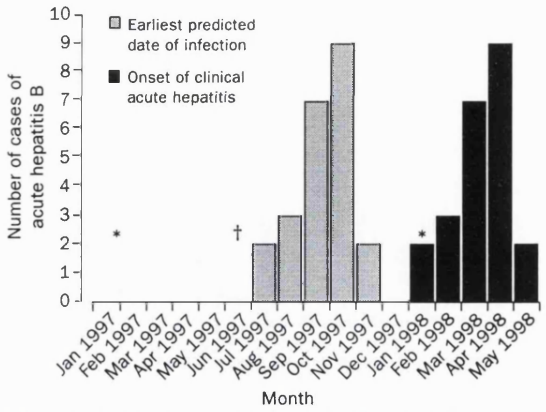


Figure 4: **Month of onset of acute clinical hepatitis B and earliest predicted month of infection**

Dates of autohaemotherapy of patient 18(*) and patient 16 (†), who were possible sources of transmission.

in this outbreak belonged to genotype D, which circulates widely in various geographical regions, including India,[15] from where patient 16 originated. The patient had chronic hepatitis B, with a high level of circulating HBV DNA. Although transmission from this patient cannot be proven, we believe that transmission from the other patients with chronic hepatitis B was less likely. Patient 18 had received treatment in January, 1997, and again in January, 1998. Since no cases of clinical hepatitis occurred for more than 12 months after this first date of treatment, we thought it more likely that she became infected in January, 1998. She was not tested until June, 1998, when she was negative for IgM against HBc. The three remaining patients with chronic hepatitis B (patients 6, 11, and 22) were positive for IgM against HBc at initial testing in March, 1998, suggesting recent infection. These three patients all began treatment after patient 16 stopped attending, and so were not regarded as the likely source of the outbreak.

A review of the technique used by the clinic staff, and detailed reporting from patients, revealed that new needles and syringes were used for each patient. However, saline was drawn from a multiuse bottle directly after venepuncture, with the syringe containing the patient's blood. This bottle was refilled from a central source. The bottle was not available for analysis.

HCV infection

Five of the 352 patients were positive for antibodies against HCV. The index case was a 54-year-old woman who was found during blood donation in October, 1997, to be positive for antibodies against HCV, and for HCV RNA. She was infected with HCV genotype 2b. She last donated blood in January, 1997, when she was negative for antibodies against HCV, and for HCV RNA. Her only risk factor for HCV infection was having received autohaemotherapy in March, 1997. Four other patients were known to have chronic hepatitis C before they attended the clinic. Further testing revealed that all four were infected with a different HCV genotype.

Secondary transmission of HBV

50 household or sexual contacts of patients exposed to HBV were identified and tested. 24 contacts of patients with active hepatitis B, who had no markers of previous

HBV infection, were vaccinated. The sexual contact of one patient, who had been lost to follow up, developed acute icteric hepatitis B. HBV genetic analysis linked the infection to her partner.

Discussion

We have described a large community-based outbreak of hepatitis B involving cases definitively linked to a single HBV variant. Our findings strongly implicate autohaemotherapy as having facilitated viral transmission. Many people in the USA and Europe consult alternative therapists in addition to, or instead of, seeking conventional medical help.[16,17] There has been debate about the validity of some alternative therapies, and the need for greater clinical analysis of their efficacy and safety.[18] Autohaemotherapy is available in clinics worldwide. The reinjection of the patient's own blood, often after oxygenation or ultraviolet irradiation, is thought by its proponents to invigorate the immune system. It is offered for the treatment of a wide range of ailments, including allergies, malignancy, viral hepatitis, and herpes zoster.[19,20]

The point source of infection in this outbreak cannot be unequivocally identified. The observation that all the patients with linked acute hepatitis B had received treatment during the latter months of 1997 may be biased because blood sampling was done in early 1998, when recently infected patients would be more likely to be HBsAg-positive than those infected earlier in the year. However, the finding that all cases of symptomatic acute hepatitis B in this outbreak developed clinical disease after December, 1997, and that the proportion of patients with markers of HBV infection increased in the last 6 months of that year, strongly suggests that this outbreak began during the middle of 1997. Since patient 16, who had chronic hepatitis B, only attended in June, 1997, the epidemiological data suggest that virus transmission occurred from that time. The possibility that he became infected at this time from another patient with symptomless acute or chronic hepatitis B, who was not identified in the outbreak investigation, cannot be excluded. The source of infection in previous health-care-related outbreaks has often been attributed to practitioners infected with HBV.[21] However, although the staff in the present outbreak were infected, they presented with clinical acute hepatitis several weeks after the first case of acute hepatitis was identified, and had themselves also received autohaemotherapy.

The risks of bloodborne infection from autohaemotherapy have been reported from Germany, where several cases of HCV and HIV infection were attributed to viral contamination of inadequately sterilised and repeatedly used equipment.[22,23] Transmission of HBV through the use of multidose containers has been previously described, in both hospital[24] and community-based[25] health-care settings. Since patients infected at this clinic had received treatment up to 16 weeks apart, whether transmission occurred from viable virus that persisted in the bottle from initial inoculation, or whether repeated contamination of the container occurred from the pool of reattending infected patients, is difficult to determine.

In previous outbreaks of hepatitis related to skin-piercing procedures, cases have been linked on the basis of epidemiological data, a history of clinical hepatitis, and serological markers of recent infection.[6,7] HBsAg subtyping has been used to provide further evidence for the relatedness of infections. However, HBV sequence amplification followed by phylogenetic analysis is an even more powerful tool with which to link infections to a common source.[26,27]

The documentation of HCV seroconversion in a patient with no other apparent risk factors suggests that HCV was transmitted through a similar mechanism to HBV. The finding of many fewer cases of hepatitis C than hepatitis B is of interest, but remains unexplained. HCV is less easily spread during parenteral health-care procedures than HBV,[28] and the low rate of hepatitis C in this outbreak may be a reflection of the small size of the inoculum, or of the poor ex-vivo viability of HCV. That the patient who was thought to be the source of HBV was not infected with HCV is of note.

This report emphasises that HBV infection poses a continuing risk in non-endemic countries where universal vaccination has not been instituted. Our findings are a reminder to patients and health practitioners to be aware of the hazards of bloodborne virus transmission, and the need for aseptic practice during all skin-piercing procedures. In the absence of the strict regulations that are applied to conventional medicine, further outbreaks of this nature may continue to be associated with this type of alternative therapy.

Contributors

Margie Meltzer, Sukhdev Sharma, Stephen Farrow, Koye Balogun, Gervase Hamilton, and Mary Ramsay did the public-health investigation; Rachel Hallett and Chong-Gee Teo did the virological and molecular epidemiological analysis; Conor Farrington did the statistical analysis; George Webster, Simon Whalley, Dave Brown, and Geoffrey Dusheiko were responsible for clinical case management; and George Webster, Mary Ramsay, Rachel Hallett, Chong-Gee Teo, and Geoffrey Dusheiko were involved with the production of the paper.

Acknowledgments

We thank G Colucci of Roche Diagnostic Systems for the supply of the AMPLICOR HBV MONITOR assays, and the following for collection, and analysis of data: D Power, M Collins, C Parker, S Harbour, F Mohamed, K Harris, D Taylor, D Banin, F Mayet, S Shafi, and P Mortimer. We also thank D Mutimer, E Ong, the Consultants in Communicable Disease Control, and the general practitioners concerned for individual patient care, and J Hoofnagle for advice on the paper.

References

1 Lee WM. Hepatitis B infection. *N Engl J Med* 1997; **337:** 1733–45.
2 Balogun MA, Ramsay ME, Fairley CK, Collins M, Heptonstall J. Acute hepatitis B infection in England and Wales: 1985–96. *Epidemiol Infect* 1999; **122:** 125–31.
3 Centers for Disease Control and Prevention. Hepatitis Surveillance Report No. 56. Atlanta: Centers for Disease Control and Prevention, 1996: 36.
4 UK Health Departments. Protecting health care workers and patients from hepatitis B: recommendations of the Advisory Group on Hepatitis. London: Her Majesty's Stationery Office, 1993.
5 Limentani AE, Elliott LM, Noah ND, Lamborn JK. An outbreak of hepatitis B from tattooing. *Lancet* 1979; ii: 86–88.
6 Boxall EH. Acupuncture hepatitis in the West Midlands. *J Med Virol* 1978; **2:** 377–79.
7 Stryker WS, Gunn RA, Francis DP. Outbreak of hepatitis B associated with acupuncture. *J Fam Pract* 1986; **22:** 155–58.
8 Local Government (Miscellaneous Provisions) Act. London: Her Majesty's Stationery Office, 1982.
9 UK Health Departments. Immunisation against infectious disease, 1996. London: Her Majesty's Stationery Office, 1996.
10 Boom R, Sol CJA, Salimans MMM, Jansen CL, Wertheim-van Dillen PME, van der Noordaa J. Rapid and simple method for purification of nucleic acids. *J Clin Microbiol* 1990; **28:** 495–503.
11 Kaneko S, Feinstone SM, Miller RH. Rapid and sensitive method for the detection of serum hepatitis B virus DNA using the polymerase chain reaction technique. *J Clin Microbiol* 1989; **27:** 1930–33.
12 Higgins DG, Sharp PM. Fast and sensitive multiple sequence alignments on a microcomputer. *Comput Appl Biosci* 1989; **5:** 151–53.
13 Saitou N, Nei M, The neighbour-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 1987; **4:** 406–25.

14 Harris KA, Gilham C, Mortimer PP, Teo CG. The most prevalent hepatitis C virus genotypes in England and Wales are 3a and 1a. *J Med Virol* 1999; **58:** 127–31.

15 Norder H, Hammas B, Lee S-D, et al. 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* 1993; **74:** 1341–48.

16 Eisenberg DM, Kessler RC, Foster C, Norlock FE, Calkins DR, Delbanco TL. Unconventional medicine in the United States—prevalence, costs, and patterns of use. *N Engl J Med* 1993; **328:** 246–52.

17 Thomas KJ, Carr J, Westlake L, Williams BT. Use of non-orthodox and conventional health care in great Britain. *BMJ* 1991; **302:** 207–10.

18 Angell M, Kassirer JP. Alternative medicine—the risks of untested and unregulated remedies. *N Engl J Med* 1998; **339:** 839–40.

19 Bocci V. Is ozonetherapy therapeutic? *Perspect Biol Med* 1998; **42:** 131–43.

20 Olwin JH, Ratajcak HV, House RV. Successful treatment of herpetic infections by autohaemotherapy. *J Altern Complement Med* 1997; **3:** 155–58.

21 The Incident Investigation Teams and Others. Transmission of hepatitis B to patients from four infected surgeons without hepatitis B e antigen. *N Engl J Med* 1997; **336:** 178–84.

22 Daschner FD. Hepatitis C and human immunodeficiency virus infection following ozone autohaemotherapy. *Eur J Clin Microbiol Infect Dis* 1997; **16:** 620.

23 Gabriel C, Blauhut B, Greul R, Schneeweis B, Roggendorf M. Transmission of hepatitis C by ozone enrichment of autologous blood. *Lancet* 1996; **347:** 541.

24 Alter MJ, Ahtone J, Maynard JE. Hepatitis B virus transmission associated with a multiple-dose vial in a haemodialysis unit. *Ann Intern Med* 1983; **99:** 33–33.

25 Canter J, Mackey K, Good LS, et al. An outbreak of hepatitis B associated with jet injections in a weight reduction clinic. *Arch Intern Med* 1990; **150:** 1923–27.

26 Yusof JHM, Flower AJE, Teo CG. Transmission of hepatitis B virus analysed by conformation-dependent polymorphisms of single-stranded DNA. *J Infect Dis* 1994; **169:** 62–67.

27 Tedder RS, Zuckerman MA, Goldstone AH, et al. Hepatitis B transmission from contaminated cryopreservation tank. *Lancet* 1995; **346:** 137–40.

28 Petrosillo N, Puro V, Ippolito G, et al. Hepatitis B virus, hepatitis C virus and human immunodeficiency virus infection in health care workers: a multiple regression analysis of risk factors. *J Hosp Infect* 1995; **30:** 273–81.