Genomic analysis of two GB virus A variants isolated from captive monkeys

James C. Erker, Suresh M. Desai, Thomas P. Leary, Michelle L. Chalmers, Carla C. Montes and Isa K. Mushahwar

Virus Discovery Group, Experimental Biology Research, Abbott Laboratories, Dept 90D, Bldg L3, 1401 Sheridan Road, North Chicago, IL 60064, USA

The recent isolation of GB viruses A and B from GB agent infected tamarins and their lack of involvement in human hepatitis has sparked interest in the origin of these viruses. Several healthy non-human primate species have been shown to harbour sequences 52–79% identical to the GBV-A 5′ nontranslated region. In this paper we report the near genome length sequence of GBV-A_{mx} 70047 and GBV-A_{tri} 1122. These sequences support previous observations about the genomic organization of GBV-A and provide insight into the genomic variability within this virus genus. Although the GBV-A variant polyproteins possess many motifs conserved between other members of the *Flaviviridae*, they do not encode a basic core-like protein. Amino acid sequence comparisons and phylogenetic analysis demonstrate variability within the GBV-A genus similar to that observed between hepatitis C virus (HCV) types. However, genomic organization and disease association demonstrate a closer evolutionary relationship to GBV-C than to HCV.

The genomes of two novel viruses, GB virus A and GB virus B (GBV-A, GBV-B), were isolated from GB agent infected tamarins (Simons *et al.*, 1995*a*). These viruses resemble members of the *Flaviviridae* in their genomic organization; however, amino acid alignments indicate that they represent two new genera within this virus family (Muerhoff *et al.*, 1995). Successive passages of GBV-B suggested that this virus was associated with the onset of hepatitis in infected tamarins (Schlauder *et al.*, 1995*a*), although GBV-B sequences have yet to be identified in humans. Tamarins infected with GBV-A are persistently infected, yet hepatitis or a detectable immune response has not been observed (Schlauder *et al.*, 1995*a*; Pilot-Matias *et al.*, 1996*a*). RT–PCR studies using degenerate oligonucleotide primers have identified sequences 59% identical to GBV-A within the NS3 region in several human sera. These sequences were deduced to be part of a third novel virus termed GB virus C (GBV-C; Simons *et al.*, 1995*b*). Additionally, using GBV-A specific primers, sequences 79% identical to the GBV-A 5′ nontranslated region (NTR) were identified in seemingly healthy tamarins with no known exposure to the virus (Schlauder *et al.*, 1995*b*).

In order to investigate the origin of GBV-A, GBV-A specific primers designed to conserved regions within the 5′ NTR were used in RT–PCR studies of several animal species, including tamarin, marmoset, owl monkey, macaque and chimp. GBV-A related sequences were identified in several species of tamarins, marmosets and owl monkeys, many of which showed persistent viral infection. While sequences from animals of a given species were virtually identical, they were only 52–79% identical when compared between species (Leary *et al.*, 1996*a*; Bukh & Apgar, 1997). Analysis of these isolates suggests that species specific groupings of GBV-A variants break down into at least five distinct genotypes. Interestingly, GBV-A variant sequences from *Saguinus nigricollus*, a tamarin species used in the initial passages of the GB agent, were 91–99% identical to GBV-A (Bukh & Apgar, 1997), suggesting that this primate species may have been the source of the prototype strain of GBV-A.

The full-length genomes of GBV-A and GBV-A_{lab} (*Saguinus labiatus*) have been reported previously (Simons *et al.*, 1995*a*; Muerhoff *et al.*, 1995; Leary *et al.*, 1997). To determine the relatedness of the GBV-A variants, the 5′ NTR sequences of GBV-A_{mx} (*Callithrix* hybrid, GenBank accession no. U93245) and GBV-A_{tri} (*Aotus trivirgatus*, U93248) have been extended to near genome length by RT–PCR. Briefly, nucleic acids extracted from the sera of *Callithrix* hybrid 70047 (*jacchus*/*penicillata* cross) and *Aotus trivirgatus* 1122 were reverse transcribed (Leary *et al.*, 1996*a*) and used as template in PCR reactions involving the following methodologies: (i) anchored PCR utilizing a gene specific primer and a GBV-A

Author for correspondence: James C. Erker.
Fax +1 847 938 6042. e-mail James.Erker@add.ssw.abbott.com

GenBank accession numbers: GBV-A_{mx}, AF023424; GBV-A_{tri}, AF023425.

Table 1. Genomic characteristics

	GBV-A	GBV-Alab	GBV-Amx	GBV-Atri	GBV-C	GBV-B	HCV-1
Genomic length (nucleotides)	9653	9550	9586	9625	9377	9143	9401
Potential length (nucleotides)*	—	9693	9727	9795	9394	—	9499
Polyprotein (amino acids)	2954	2967	2970	3005	2843	2864	3011
Core-like protein (amino acids)	None	None	None	None	None	156	191
E1/E2 *N*-glycosylation sites	1/3	2/3	1/2	1/4	1/3	3/8	5/11
p7-like protein (amino acids)	185	185	186	185	81	†	63

* Additional sequence has been identified for one or more related isolates.
† The eukaryotic cleavage signal sequence between E2 and p7 is unclear.

specific primer designed within regions of identity between GBV-A and GBV-C (Leary *et al.*, 1997); (ii) anchored PCR utilizing a gene specific biotinylated primer and a random flanking primer (Sorenson *et al.*, 1993; Leary *et al.*, 1996*b*); (iii) 'touchdown' PCR (Roux, 1995) using GBV-A specific primers designed within regions of conservation between GBV-A and GBV-C; and (iv) conventional amplification with GBV-A variant specific primers. The final extension reactions used specifically primed cDNA as described (Leary *et al.*, 1997). PCR products were cloned into pT7Blue T-Vector (Novagen) and two or three clones sequenced using the Applied Biosystems model 373 DNA sequencer. Sequences were compiled, edited and analysed using the programs of the Wisconsin Sequence Analysis Package GCG, the Phylip package and TreeView as previously described (Erker *et al.*, 1996).

The GBV-A_{mx} and GBV-A_{tri} genomes were extended to near genome length (Table 1). GBV-A_{mx} is 9586 nucleotides with approximately 65·3% and 66·2% identity to GBV-A and GBV-A_{lab}, respectively, and with 48·7% identity to GBV-C. Comparisons of GBV-A_{mx} to GBV-A suggest that as many as 59 additional nucleotides may be missing from the 5′ end of the GBV-A_{mx} genome. As compared to GBV-A, the GBV-A_{mx} 3′ NTR appears to be missing 82 nucleotides, potentially the result of a mispriming event during the RT or PCR reaction. Thus, GBV-A_{mx} may potentially be 9727 nucleotides in length, 74 nucleotides longer than GBV-A. GBV-A_{tri} is 9625 nucleotides in length, with approximately 59% identity to GBV-A, GBV-A_{lab} and GBV-A_{mx} and with 47·9% identity to GBV-C. Alignments with GBV-A suggest 59 nucleotides have yet to be isolated at the 5′ end of the GBV-A_{tri} genome. The final extension product appears to be the result of a mispriming event during the PCR reaction resulting in the sense primer being located on each end of the product. Thus, GBV-A_{tri} may be missing up to 111 nucleotides from the 3′ NTR. With a potential length of 9795 nucleotides, GBV-A_{tri} would be 142 nucleotides longer than GBV-A.

Like other members of the *Flaviviridae*, the GBV-A variant genomes encode a single large polyprotein. The GBV-A_{mx} polyprotein is 2970 amino acids in length and possesses approximately 72% identity to GBV-A and GBV-A_{lab}. The GBV-A_{tri} polyprotein is 3005 amino acids with 60% identity to GBV-A, GBV-A_{lab} and GBV-A_{mx}. Though many conserved amino acid insertions and deletions exist between the GBV-A variants and GBV-C isolates (data not shown), the four GBV-A variant polyproteins are each only 48% identical to GBV-C. Amino acid sequence identity across the four GBV-A variant polyproteins ranges from 85 to 40%, with an average identity of 65% across the entire polyproteins (Fig. 1). Interestingly, the pattern and degree of variability observed between the GBV-A variants isolated from distinct primate species is similar to that of geographically distributed HCV types (Simmonds, 1995; Erker *et al.*, 1996). The highest degree of conservation is located within the NS3 protease/helicase and the NS5B replicase, presumably due to the functional importance of these enzymes in virus replication. Within these regions, amino acid residues used to define supergroup II helicases and supergroup II replicases, present in the GB viruses (Muerhoff *et al.*, 1995; Leary *et al.*, 1996*b*; Erker *et al.*, 1996; Leary *et al.*, 1997), are conserved in GBV-A_{mx} and GBV-A_{tri} (data not shown).

Despite the presence of more than 500 nucleotides upstream of the predicted translation start codon, GBV-A, GBV-A_{lab}, GBV-A_{mx} and GBV-A_{tri} do not appear to encode a core-like protein at the amino terminus of the polyprotein. Instead, the long GBV-A 5′ NTR possesses weak internal ribosome binding site activity with translation initiation occurring just upstream of the first envelope protein (Simons *et al.*, 1996). Similarity in sequence and potential secondary structure (data not shown) implies analogous function within

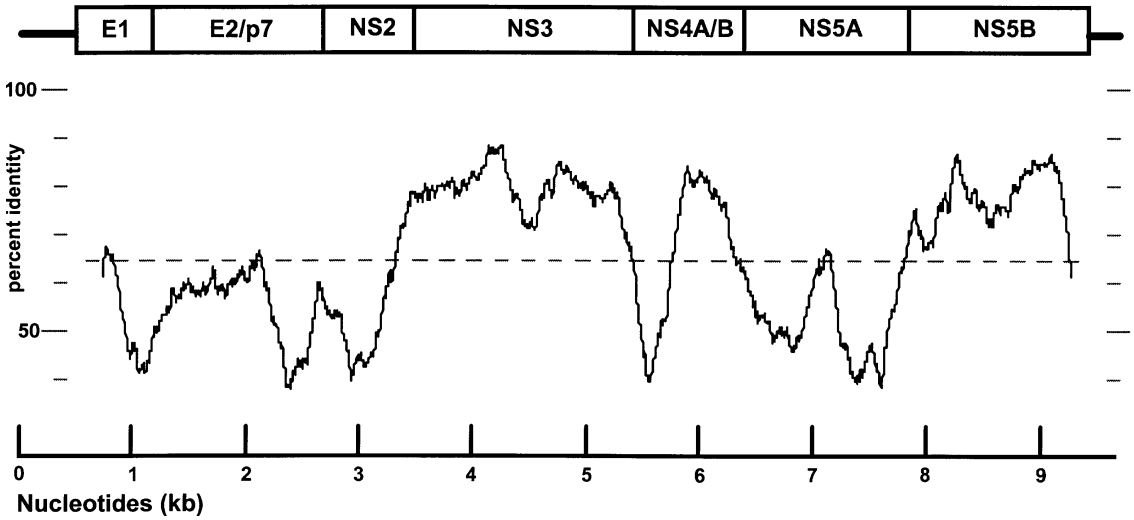

Fig. 1. The four GBV-A variant polyproteins were aligned and their identity plotted across a sliding 100 amino acid window of comparison using PLOTSIMILARITY. The polyprotein organization and the nucleotide scale are provided. The dashed line indicates the mean identity between the four polyproteins.

the 5′ NTR for the other GBV-A variants. The lack of a basic core-like protein is a characteristic held in common with GBV-C isolates, but is distinct from other members of the *Flaviviridae* (Simons *et al.*, 1996; Erker *et al.*, 1996).

Several conserved protein motifs suggest proteolytic processing of the GBV-A variant polyproteins occurs in a similar fashion to HCV. Each of the structural proteins is preceded by potential eukaryotic signal sequence cleavage sites allowing for host proteases to cleave the amino-terminal one-third of the polyprotein to produce the E1, E2 and p7 proteins. The structural proteins differ from those of HCV and GBV-B in that there are few *N*-linked glycosylation sites (Table 1). Despite the variability within the GBV-A variant envelope proteins (Fig. 1), several of these glycosylation sites are conserved. There are also numerous cysteine, glycine and proline residues conserved between the envelope proteins, many of which are also conserved with the GBV-C envelope proteins (data not shown), suggesting similar secondary and tertiary structures. Also of interest is the sequence variability and size of the p7 proteins (Fig. 1, Table 1). This cleavage product is almost three times larger than that of HCV and twice that of GBV-C. It is unclear what implications these variations may have as the role of p7 in HCV replication is unknown.

The non-structural proteins may be cleaved by one of two virally encoded proteases. A zinc-dependent thiol protease appears to be encoded within the NS2/NS3 gene product. While limited amino acid identity exists within this region (Fig. 1), histidine and cysteine residues necessary for autocatalytic cleavage of NS2 from NS3 (Grakoui *et al.*, 1993*a*) are conserved within the GBV-A_{mx} and GBV-A_{tri} proteins (data not shown). Located at the amino-terminal one-third of the NS3 protein is a serine protease presumably responsible for cleavage of downstream sequences. Amino acids previously shown to be required for catalytic activity of this protease; histidine, asparagine and serine (Grakoui *et al.*, 1993*b*), have been identified at appropriate positions in the GBV-A_{mx} and GBV-A_{tri} NS3 proteins (data not shown).

Amino acid sequence comparisons of the polyproteins suggest a close relationship between the GBV-A variants. In order to more accurately determine the relationship of the GB viruses to one another and HCV, phylogenetic distances have been determined as previously described (Leary *et al.*, 1997). The evolutionary distances of the GBV-A variants or GBV-C isolates to either GBV-B or HCV are all greater than 2·3 amino acid substitutions per position, while distances between the HCV types range from 0·174 to 0·396 substitutions per position. Thus, GBV-A and GBV-C isolates are clearly not members of the HCV or GBV-B groups, as indicated by the three major branches on the unrooted phylogenetic tree (Fig. 2). Three of the four GBV-A isolates fall within a range similar to that observed for HCV types; GBV-A, GBV-A_{lab} and GBV-A_{mx} have genetic distances ranging from 0·338 to 0·380 substitutions per position and group closely on the phylogenetic tree. Interestingly, GBV-A_{tri} has distances to the other GBV-A variants of 0·623 to 0·637 substitutions per position and branches closer to the node where GBV-C separates on the phylogenetic tree. However, the distance of the GBV-C isolates to any one of the GBV-A variants ranges from 0·965 to 1·000 amino acid substitutions per position, thereby distinguishing GBV-C isolates from the GBV-A variants. The groupings shown in Fig. 2 are supported by strong bootstrap values calculated using 100 resamplings of the data. In 100% of the trees, GBV-A variants, GBV-C isolates, HCV types and GBV-B are placed into four major groups. Additionally, the GBV-A variants group as indicated in greater than 93% of the trees.

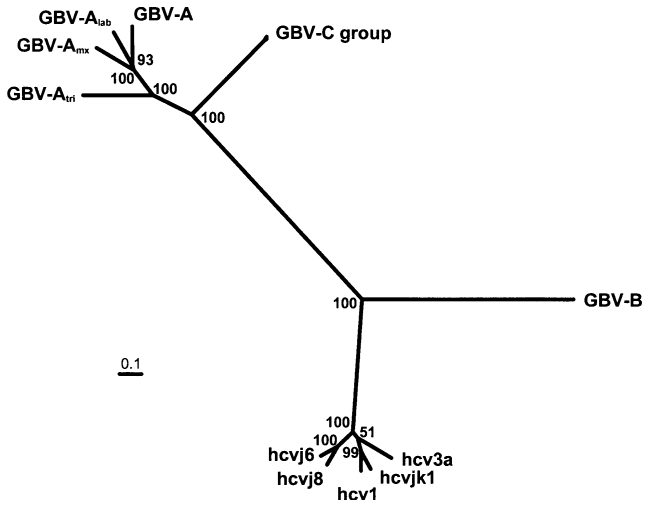

Fig. 2. Unrooted phylogenetic tree derived from alignments of flavivirus polyproteins. The bootstrap values obtained from 100 resamplings of the data are shown at the appropriate nodes. The scale to measure genetic distance is indicated at the lower left. The GBV-C sequences are indicated as a group due to the close relationship of the full-length isolates (within 0·016 amino acid substitutions per position). The following sequences were utilized for these analysis: GBV-A, U22303; GBV-A_{lab}, U94421; GBV-B, U22304; GBV-C, U36380; GBV-C(EA), U63715; PNF2161, U44402; R10291, U45966; GT110, D90600; GT230, D90601; HCV-1, M62321; HCVjk1, X61596; HCVj6, D00944; HCVj8, D10988; HCV3a, D28917. The variants reported here are GBV-A_{mx} 70047, AF023424, and GBV-A_{tri} 1122, AF023425.

Interestingly, many of the characteristics described above for the GBV-A variants are strikingly similar to those observed for GBV-C. The genomic organization of GBV-C resembles that of the GBV-A variants; ranging from the structure and function of the 5′ NTR, to the lack of a basic core-like protein, to the glycosylation and potential similarity in structure of the envelope proteins, to the size and composition of the 3′ NTR (Erker *et al.*, 1996). Infected patients show persistence of GBV-C RNA in serum and only become RNA negative with the development of an anti-E2 antibody response to a mammalian expressed GBV-C E2 recombinant protein (Pilot-Matias *et al.*, 1996*b*). This is the only serological marker to date which is consistently associated with GBV-C exposure. While GBV-C RNA has been identified in a small percentage of healthy blood donors, GBV-C prevalence among high risk groups, as well as acute and chronic non-A–E hepatitis cases, is potentially as high as 80% (Colombatto *et al.*, 1996; Fiordalisi *et al.*, 1996).

Also of interest is the phylogenetic relationship of the GBV-A variants to GBV-C. While genetic distances between the GBV-A variants and GBV-C isolates are close to 1·0 amino acid substitutions per position, those between the GBV-A variants are disperse, ranging from 0·338 to 0·637 substitutions. As additional GBV-A variants are identified, these genetic distances may continue to near 1·0 amino acid substitution per position and GBV-A variant branching on the phylogenetic tree may continue to approach the GBV-C node. All this taken together suggests that these viruses have diverged from a common ancestor and that GBV-C may be the human equivalent or species of GBV-A. GBV-A and GBV-C should thus be classified within a single genus within the *Flaviviridae*, separate from GBV-B and HCV.

Due to the close relationship of the GBV-A variants to GBV-C, primates naturally infected with a GBV-A variant could be an excellent animal model for GBV-C infection. However, this relationship might also be important when primates are used to study GBV-C. It has been shown that primates infected with GBV-A_{lab} may not be susceptible to GBV-A infection (Leary *et al.*, 1996*a*). Thus, the presence of a GBV-A variant or antibodies to GBV-A variant proteins may possibly protect an animal from GBV-C infection.

As additional full-length sequences become available, it will be of interest to determine the level of conservation between GBV-A variant genomes from a given primate species (GBV-A_{lab}, GBV-A_{mx}, etc.). Will sequence variability continue to reflect that of HCV allowing for classification into subtypes or, as suggested by the high degree of conservation seen within 5′ NTR sequences (Leary *et al.*, 1996*a*; Bukh & Apgar, 1997), will genomic sequences be highly conserved and cluster closely on a phylogenetic tree, as in the case of GBV-C sequences (Erker *et al.*, 1996)? These analyses may lend insight into GB virus evolution and the sequence constraints acting on GBV-A and GBV-C.

The authors wish to thank Julie Yamaguchi and Jennifer Lund for their excellent technical assistance throughout this work, and Dr A. Scott Muerhoff for critical review of the manuscript.

References

Bukh, J. & Apgar, C. L. (1997). Five new or recently discovered (GBV-A) virus species are indigenous to new world monkeys and may constitute a separate genus of the Flaviviridae. *Virology* **229**, 429–436.

Colombatto, P., Randone, A., Civitico, G., Monti Gorin, J., Dolci, L., Medaina, N., Oliveri, F., Verme, G., Marchiaro, G., Pagni, R., Karayiannis, P., Thomas, H. C., Hess, G., Bonino, F. & Brunetto, M. R. (1996). Hepatitis G virus RNA in the serum of patients with elevated gamma glutamyl transpeptidase and alkaline phosphatase: a specific liver disease. *Journal of Viral Hepatitis* **3**, 301–306.

Erker, J. C., Simons, J. N., Muerhoff, A. S., Leary, T. P., Chalmers, M. L., Desai, S. M. & Mushahwar, I. K. (1996). Molecular cloning and characterization of a GB virus C isolate from a patient with non-A–E hepatitis. *Journal of General Virology* **77**, 2713–2720.

Fiordalisi, G., Zanella, I., Mantero, G., Bettinardi, A., Stellini, R., Paraninfo, G., Cadeo, G. & 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.

Grakoui, A., McCourt, D. W., Wychowski, C., Feinstone, S. M. & Rice, C. M. (1993*a*). A second hepatitis C virus-encoded proteinase. *Proceedings of the National Academy of Sciences, USA* **90**, 10583–10587.

Grakoui, A., McCourt, D. W., Wychowski, C., Feinstone, S. M. & Rice, C. M. (1993*b*). Characterization of the hepatitis C virus-encoded serine proteinase: determination of protease-dependent polyprotein cleavage sites. *Journal of Virology* **67**, 2832–2843.

Leary, T. P., Desai, S. M., Yamaguchi, J., Chalmers, M. L., Schlauder, G.

G., Dawson, G. J. & Mushahwar, I. K. (1996*a*). Species-specific variants of GB virus A in captive monkeys. *Journal of Virology* **70**, 9028–9030.

Leary, T. P., Muerhoff, A. S., Simons, J. N., Pilot-Matias, T. J., Erker, J. C., Chalmers, M. L., Schlauder, G. G., Dawson, G. J., Desai, S. M. & Mushahwar, I. K. (1996*b*). The sequence and genomic organization of GBV-C: a novel member of the *Flaviviridae* associated with human non-A–E hepatitis. *Journal of Medical Virology* **48**, 60–67.

Leary, T. P., Desai, S. M., Erker, J. C. & Mushahwar, I. K. (1997). The sequence and genomic organization of a GB virus A variant isolated from captive tamarins. *Journal of General Virology* **78**, 2307–2313.

Muerhoff, A. S., Leary, T. P., Simons, J. N., Pilot-Matias, T. P., Erker, J. C., Chalmers, M. L., Schlauder, G. G., Dawson, G. J., Desai, S. M. & Mushahwar, I. K. (1995). Genomic organization of GBV-A and GBV-B: two new members of the *Flaviviridae* associated with GB-agent hepatitis. *Journal of Virology* **69**, 5621–5630.

Pilot-Matias, T. J., Muerhoff, A. S., Simons, J. N., Leary, T. P., Buijk, S. L., Chalmers, M. L., Erker, J. C., Dawson, G. J., Desai, S. M. & Mushahwar, I. K. (1996*a*). Identification of antigenic regions in the GB hepatitis viruses GBV-A, GBV-B and GBV-C. *Journal of Medical Virology* **48**, 329–338.

Pilot-Matias, T. J., Carrick, R. J., Coleman, P. F., Leary, T. P., Surowy, T. K., Simons, J. N., Muerhoff, A. S., Buijk, S. L., Chalmers, M. L., Dawson, G. J., Desai, S. M. & Mushahwar, I. K. (1996*b*). Expression of the GB virus C E2 glycoprotein using the Semliki Forest virus vector system and its utility as a serologic marker. *Virology* **225**, 282–292.

Roux, K. H. (1995). Using mismatched primer-template pairs in touchdown PCR. *BioTechniques* **16**, 812–814.

Schlauder, G. G., Dawson, G. J., Simons, J. N., Pilot-Matias, T. J., Guiterrez, R. A., Heynen, C. A., Knigge, M. F., Kurpiewski, G. S., Buijk, S. L., Leary, T. P., Muerhoff, A. S., Desai, S. M. & Mushahwar, I. K. (1995*a*). Molecular and serologic analysis in the transmission of the GB hepatitis agents. *Journal of Medical Virology* **46**, 81–90.

Schlauder, G. G., Pilot-Matias, T. J., Gabriel, G. S., Simons, J. N., Muerhoff, A. S., Dawson, G. J. & Mushahwar, I. K. (1995*b*). Origin of GB-hepatitis viruses. *Lancet* **346**, 447.

Simons, J. N., Pilot-Matias, T. J., Leary, T. P., Dawson, G. J., Desai, S. M., Schlauder, G. G., Muerhoff, A. S., Erker, J. C., Buijk, S. L., Chalmers, M. L., Van Sant, C. L. & Mushahwar, I. K. (1995*a*). Identification of two flavivirus-like genomes in the GB hepatitis agent. *Proceedings National Academy of Sciences, USA* **92**, 3401–3405.

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

Simons, J. N., Desai, S. M., Schultz, D. E., Lemon, S. M. & Mushahwar, I. K. (1996). Translation initiation in GB viruses A and C: evidence for internal ribosome entry and implications for genome organization. *Journal of Virology* **70**, 6126–6135.

Simmonds, P. (1995). Variability of hepatitis C virus. *Hepatology* **21**, 570–583.

Sorenson, A. B., Duch, M., Jorgensen, P. & Pedersen, F. S. (1993). Amplification and sequence analysis of DNA flanking integrated proviruses by a simple two-step polymerase chain reaction method. *Journal of Virology* **67**, 7118–7124.

Received 10 July 1997; Accepted 25 September 1997