Journal of General Virology (1999), **80**, 2329–2335. Printed in Great Britain SHORT COMMUNICATION

Phylogenetic analysis of GB viruses A and C: evidence for cospeciation between virus isolates and their primate hosts

Rémi N. Charrel,[1,2] Philippe De Micco[1] and Xavier de Lamballerie[1]

[1] Laboratoire de Virologie moléculaire, transfusionnelle et tropicale, Faculté de Médecine, 13385 Marseille, France
[2] Dept of Pathology, Center for Tropical Diseases, University of Texas Medical Branch, Galveston, TX 77555, USA

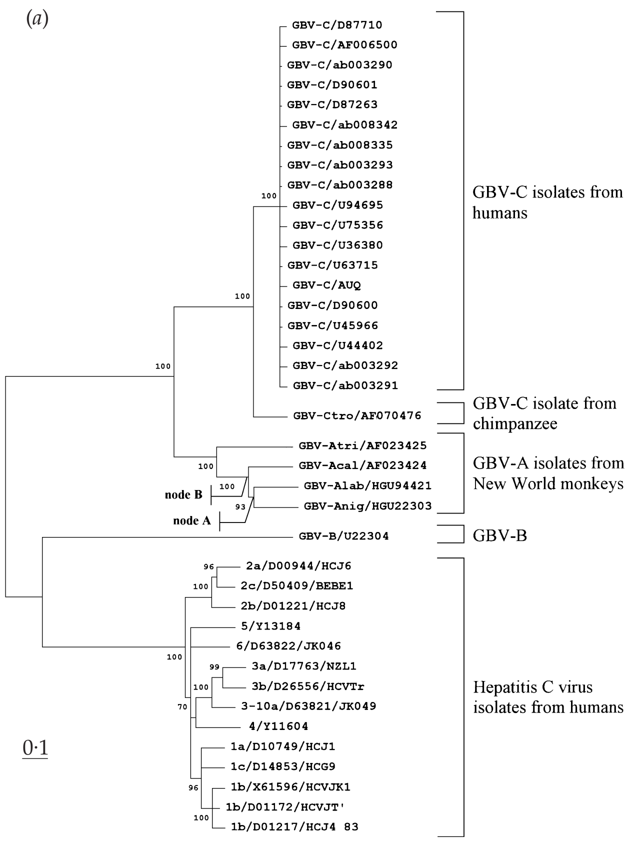
GB viruses A and C (GBV-A and GBV-C) have been isolated from humans and non-human primates. Phylogenetic analysis based on full-length polyproteins suggests that these two viruses have a common ancestor. It has now been determined that analysis of subgenomic amino acid sequences in the E2 and NS5 regions of GBV-A and a 345 nucleotide segment in the 5′ non-coding (5′NC) region was able to reproduce the phylogenetic relationships obtained by complete polyprotein sequences analysis. Using 5′NC sequences from databases, GBV-A isolates were discriminated into eight genetic groups, each one closely associated with specific primate hosts. Phylogenetic analyses performed on sequences from the ε-globin genes of primate hosts on one hand and complete polyprotein sequences from GBV-A and GBV-C isolates on the other suggest that a mechanism of cospeciation could be involved in virus evolution over a period of 35 million years.

Although the International Committee on Taxonomy of Viruses has not yet officially established their taxonomic status, the recently described GB viruses A, B and C (GBV-A, GBV-B and GBV-C) have been provisionally classified into the family *Flaviviridae* on the basis of their structure, genomic organization and sequences. Within this family, the most-closely related species is hepatitis C virus (HCV, genus *Hepacivirus*). GBV-C was isolated first from humans and more recently from chimpanzees (Adams *et al.*, 1998; Birkenmeyer *et al.*, 1998; Linnen *et al.*, 1996; Simons *et al.*, 1995*a*). Human isolates exhibit low genetic diversity and are more distantly related to simian isolates. GBV-A, originally discovered in a captive tamarin (Simons *et al.*, 1995*b*), has been isolated from different species of New World monkeys; four strains from distinct monkey species have been sequenced almost entirely (Erker *et al.*, 1998; Leary *et al.*, 1997; Simons *et al.*, 1995*b*).

Author for correspondence: Rémi Charrel (at the Laboratoire de Virologie moléculaire, transfusionelle et tropicale).
Fax +33 4 91 18 95 98. e-mail virophdm@lac2.gulliver.fr

GBV-A_{lab} (Leary *et al.*, 1997), GBV-A_{cal} (Erker *et al.*, 1998) and GBV-A_{tri} (Erker *et al.*, 1998) were isolated from *Saguinus labiatus* (tamarin), a *Callithrix jacchus* × *penicillata* hybrid (marmoset) and *Aotus trivirgatus* (owl monkey), respectively. The strain isolated from the *Callithrix* hybrid, originally named GBV-A_{mx} (Erker *et al.*, 1998), was renamed GBV-A_{cal} in this study to avoid any confusion with viruses isolated from *Saguinus mystax* monkeys. The original GBV-A strain (Simons *et al.*, 1995*b*) was recovered from an *S. labiatus* tamarin. However, on the basis of its high degree of genetic similarity to other virus strains isolated from *Saguinus nigricollis* tamarins (Fig. 2*a*), it was probably transmitted from an *S. nigricollis* tamarin during serial passage in different *Saguinus* species, as previously reported (Bukh & Apgar, 1997).

We submitted the complete amino acid sequences of these four strains to phylogenetic analysis together with complete sequences from GBV-C, GBV-B and HCV isolates (Fig. 1*a*). Because there was no evidence to allow us to assume that every site had evolved at the same rate (which is the prerequisite assumption for the maximum-likelihood and parsimony models), we tested the hypothesis of a common ancestor with a model incorporating a gamma distribution of rate variation among sites (Yang, 1993). The key to this model is the value of the shape parameter, α. We tested successive possibilities for α, such as 0·3, 0·5, 1, 2 and 10 (the smaller the number, the greater the degree of between-site rate variation and hence the greater the genetic distances obtained). Irrespective of the method used for the calculation of genetic distances (p-distance and Jukes–Cantor models were also tested) and of the value chosen for the shape parameter, α, analyses yielded similar groupings to those observed in Fig. 1(*a*). In particular, the hypothesis of a common ancestor for GBV-A and GBV-C was consistently supported by high bootstrap values. Within the cluster formed by the four GBV-A strains, GBV-A_{lab} grouped with GBV-A_{nig} and both of them clustered with GBV-A_{cal} (node A). The group constituted by these three virus strains grouped with GBV-A_{tri} (node B).



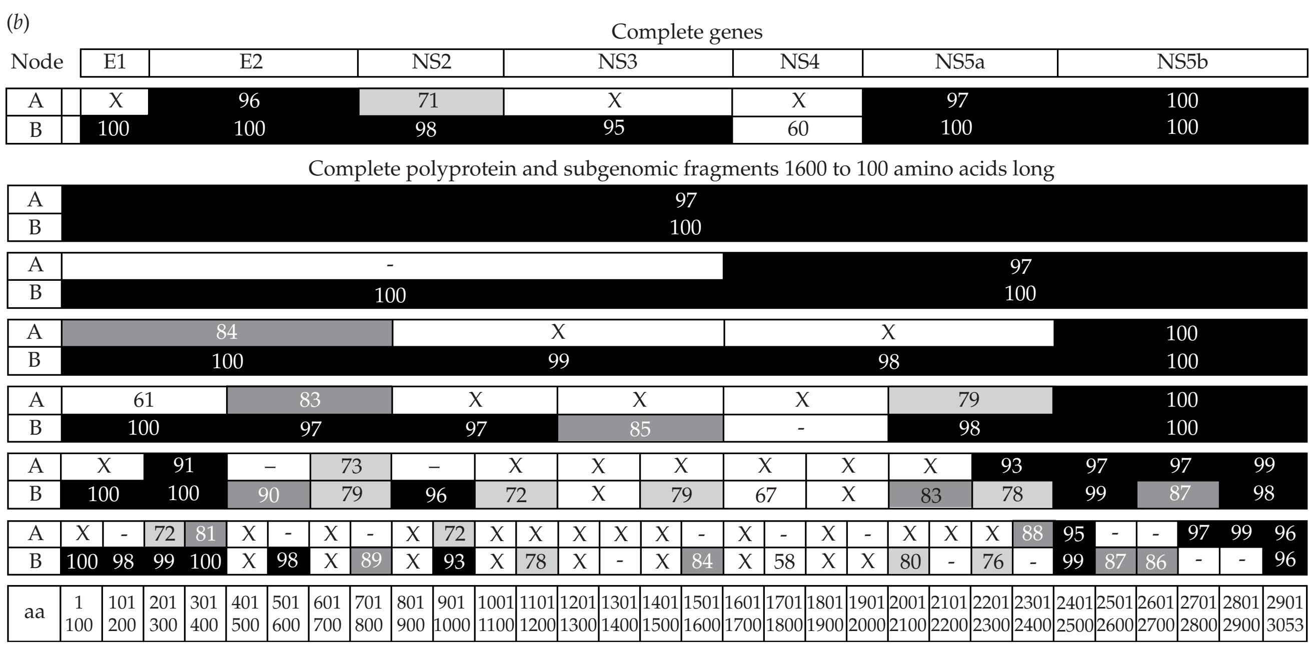


Fig. 1. Phylogenetic analysis of GBV-A on the basis of complete polyprotein sequences. (*a*) Phylogenetic tree constructed from complete polyproteins of representative isolates of GBV-A, GBV-B, GBV-C and HCV using the gamma distance algorithm ($\alpha = 2$) and the neighbour-joining method implemented in the MEGA software package (Kumar *et al.*, 1993). Bootstrap values corresponding to 500 replications are indicated. Grouping of GBV-A_{lab} and GBV-A_{nig} with GBV-A_{cal} corresponds to node A. The grouping of these three sequences with GBV-A_{tri} corresponds to node B. GenBank accession numbers are indicated after the first slash (/). For the HCV isolates, genotypes are indicated before the first slash. Bar represents an approximate distance of 0·1. (*b*) Bootstrap values (500 resamplings) for nodes A and B using subgenomic amino acid sequences. Phylogenetic analysis was performed on complete genes and on subgenomic fragments of 1600 down to 100 amino acids. aa, Amino acids. Bootstrap values $< 70\%$ are indicated by dashes. X indicates an incorrect branching pattern. Bootstrap support is indicated by shading and corresponds to 91–100% (black), 81–90% (dark grey), 71–80% (light grey) and $< 70\%$ (unshaded).

Nodes A and B were supported by 93 and 100% bootstrap values, respectively, after 500 cycles of resampling (Fig. 1*a*, *b*). GBV-A_{tri} was the most genetically divergent strain, with amino acid distances ranging from 39·5 to 40·2% when compared with the three remaining complete sequences. GBV-A_{lab} and GBV-A_{nig} exhibited 25·5% divergence and their genetic distance from GBV-A_{cal} ranged from 27·0 to 27·8%. By comparison, most of the GBV-C isolates exhibited lower

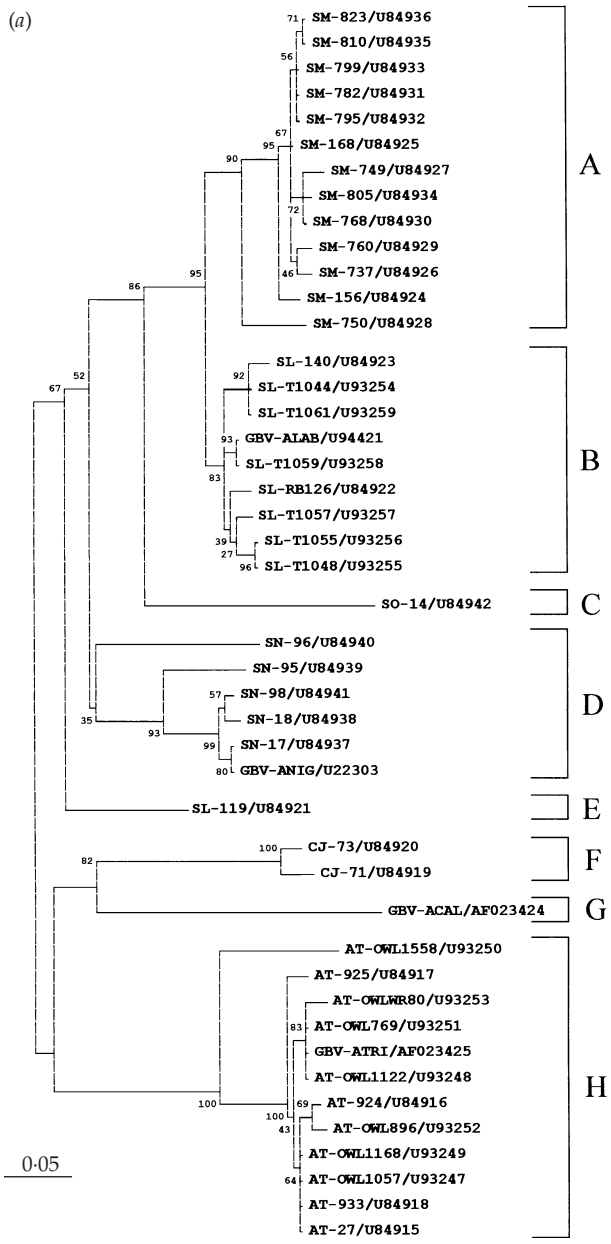
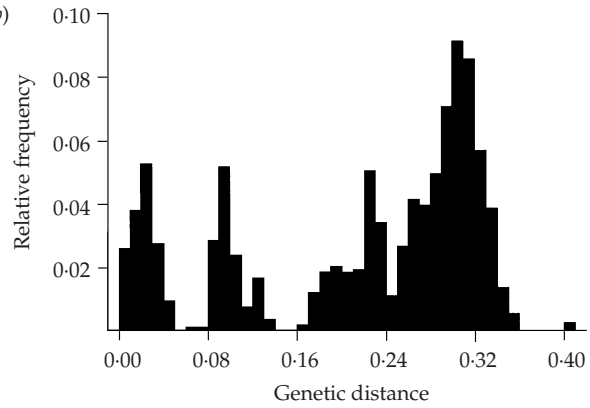
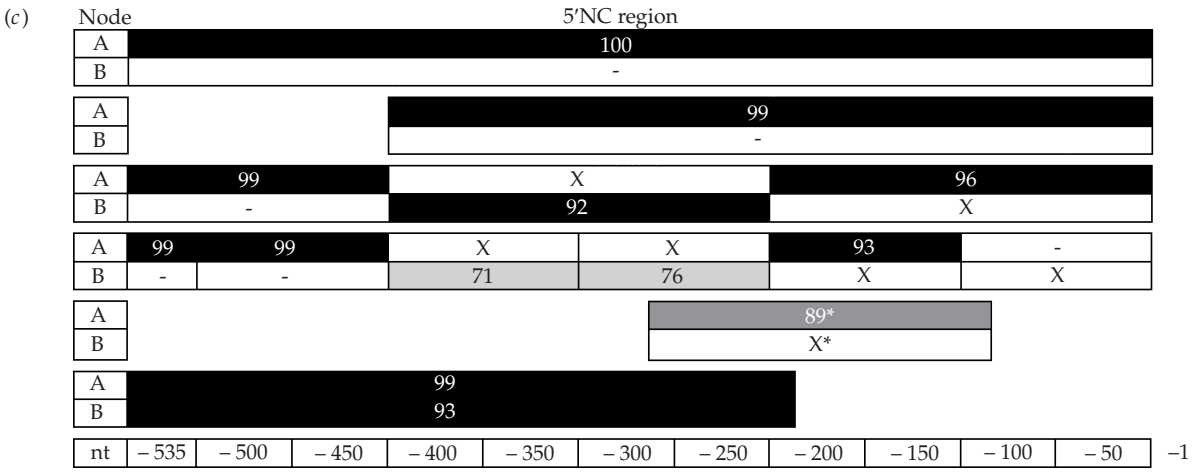


Fig. 2. Phylogenetic analysis of GBV-A isolates on the basis of nucleotide sequences in the 5′NC region. (*a*) Phylogenetic tree constructed from 45 185-nucleotide sequences using the Jukes–Cantor algorithm and the neighbour-joining method implemented in the MEGA software package (Kumar *et al.*, 1993). Bootstrap values corresponding to 500 replications are indicated. Abbreviations for the monkey species from which the virus isolates were recovered are: *S. mystax* (SM), *S. labiatus* (SL and GBV-ALAB), *S. oedipus* (SO), *S. nigricollis* (SN and GBV-ANIG), *C. jacchus* (GBV-ACAL), *C. jacchus* × *penicillata* hybrid (CJ) and *A. trivirgatus* (AT and GBV-ATRI). Genetic groups A to H are indicated. Sequences (GenBank accession nos U84915–U84942 and U93247–U93259) are located between positions −83 and −267, numbered backwards from the AUG of the prototype sequence of GBV-A (U22303). Bar represents an approximate distance of 0·05. (*b*) Distribution of evolutionary distances upon pairwise comparison. The genetic distance is reported on the *x*-axis. Frequency of genetic distances is recorded on the *y*-axis. (*c*) Bootstrap support for phylogenetic groupings based on analysis of fragments of the 5′NC region for the four strains GBV-A_{nig}, GBV-A_{lab}, GBV-A_{cal} and GBV-A_{tri}. The 5′ and 3′ boundaries of the regions analysed are indicated at the bottom of the figure. Positions are numbered relative to the initiator AUG of GBV-A_{nig}. Bootstrap values < 70% are indicated by dashes. X indicates an incorrect branching pattern. *, Bootstrap support obtained from analysis based on sequences reported previously (Bukh & Apgar, 1997; Leary *et al.*, 1996). Levels of bootstrap support are indicated by shading as outlined in Fig. 1.





genetic distances: below 5·4% between human isolates and 20·2–21·9% between human and chimpanzee isolates.

Because sequencing of complete genomes is an expensive and time-consuming procedure, studies of virus phylogeny are frequently performed by using subgenomic sequences. In the cases of HCV and GBV-C, subgenomic regions have been identified that reflect virus relationships observed by complete polyprotein analysis (Simmonds *et al.*, 1993, 1994; Smith *et al.*, 1995, 1997). We carried out a similar analysis of GBV-A isolates. The four complete GBV-A amino acid sequences and two GBV-C sequences (U36380 and U44402, acting as an outgroup) were aligned with the CLUSTAL W 1.7 program (Thompson *et al.*, 1994). Distances were calculated by using the gamma distance algorithm ($\alpha = 2$) and branching patterns were determined by the neighbour-joining method implemented in the MEGA software package (Kumar *et al.*, 1993) (Fig. 1*b*). Phylogenetic analyses were performed on the complete polyprotein, the different virus genes and sequences of 1600 down to 100 amino acids. The robustness of the resulting groupings were tested by 500 bootstrap replications. Discrepant groupings were considered to be invalid even when supported by high bootstrap values. The best results were provided by the complete E2, NS5a and NS5b genes and by partial sequences in the E2 and NS5 regions (Fig. 1*b*; see sequences giving bootstrap values $> 90\%$ at both nodes A and B).

A similar analysis was performed in the 5′ non-coding (5′NC) region, using progressively smaller nucleotide fragments, and distances were calculated by the Jukes–Cantor algorithm (Fig. 2*c*). Only analyses performed with a 345 nucleotide fragment covering positions -190 and -535 provided the appropriate branching pattern with bootstrap support above 90% (Fig. 2*c*). These results were confirmed when other distance algorithms (Kimura-2, Tamura, Tajima–Nei) and another phylogenetic method (unweighted pair group method with averages; UPGMA) were used.

In previous studies, 45 sequences of the 5′NC region of GBV-A were reported (Bukh & Apgar, 1997; Leary *et al.*, 1996). Although phylogenetic analysis of these sequences did not conform completely to the branching pattern observed by complete polyprotein analysis (Fig. 2*c*), the distribution of distances between these 45 sequences is interesting. It shows that viruses exhibiting genetic distances greater than 14% originated from different monkey species (Fig. 2*b*). The phylogenetic analysis (Fig. 2*a*) allows eight clusters (A to H) to be distinguished, each corresponding to sequences recovered from the same monkey species. Analyses performed with Jukes–Cantor or Kimura-2 algorithms and with UPGMA or maximum-parsimony methods showed similar results.

GBV-A sequences from *S. mystax* and *S. labiatus* are the most closely related. In nature, these two monkey species have contiguous distributions, but do not form mixed populations (C. Padua, personal communication). Although the possibility of cross-transmission of virus isolates cannot be completely excluded, it must be noted that these two species are genetically closer than any others in the genus *Saguinus* (Jacobs *et al.*, 1995). Thus, the low genetic distances observed between GBV-A viruses isolated from *S. mystax* and *S. labiatus* may reflect the recent split between these two species during the evolution of New World monkeys. Only isolate SL-119 did not group with the other virus isolates recovered from the same species (in this case, *S. labiatus*). This may constitute strong evidence against cospeciation. Nevertheless, SL-119 did not group with any of the other isolates, meaning that this isolate may have been acquired from a monkey belonging to another species with which *S. labiatus* populations could share a contiguous distribution in nature (C. Padua, personal communication) or from contacts during transportation or captivity.

The presence of different host species and the low genetic distance reported among viruses isolated from a unique monkey species suggested that these viruses shared a long-term evolutionary relationship with their respective primate hosts. Based on present-day genetic evidence, a cospeciation mechanism was evoked. To investigate this hypothesis, we compared the phylogenetic relationships observed among GBV-A and GBV-C isolates with the phylogeny of their primate hosts.

Phylogenetic analysis of the primate hosts was achieved by using the complete nucleotide sequences of the ε-globin gene (Fig. 3*a*). Because sequences of ε-globin have not been determined for *S. labiatus* and *A. trivirgatus*, sequences from surrogate species (*Saguinus midas* and *Aotus azarai*, respectively) were included for analysis. *S. labiatus* and *S. midas* belong to the same clade within the genus *Saguinus* (Jacobs *et al.*, 1995). *A. azarai* is currently the only species in the genus *Aotus* for which the ε-globin gene sequence is available from nucleotide databases.

The genera *Saguinus* and *Callithrix* are included in the same evolutionary lineage (callitrichines) of New World monkeys. The genus *Aotus* belongs to a different lineage, but is closer to the callitrichines than to the five remaining lineages of New World monkeys (Porter *et al.*, 1997). Chimpanzees (genus *Pan*) and humans (genus *Homo*) are located in a different branch (corresponding to the family *Hominidae* within the Old World anthropoids; Fig. 3*a*). The phylogenetic distribution of primate species infected by GBV-A or GBV-C isolates was compared with that of the different virus strains. The two phylogenetic trees presented in Fig. 3(*a*) show striking similarities between the branching pattern of viruses and that of their respective specific primate hosts. All branch-points are supported by very high bootstrap values, suggesting the existence of a co-speciation mechanism. However, further analyses including more taxa will be necessary to confirm this hypothesis.

The rate of evolution of GB viruses in their respective primate hosts was estimated (Fig. 3*b*, *c*). The GBV-A and GBV-C lineages were analysed separately. The different points on each curve reflect the correlation between the time of the

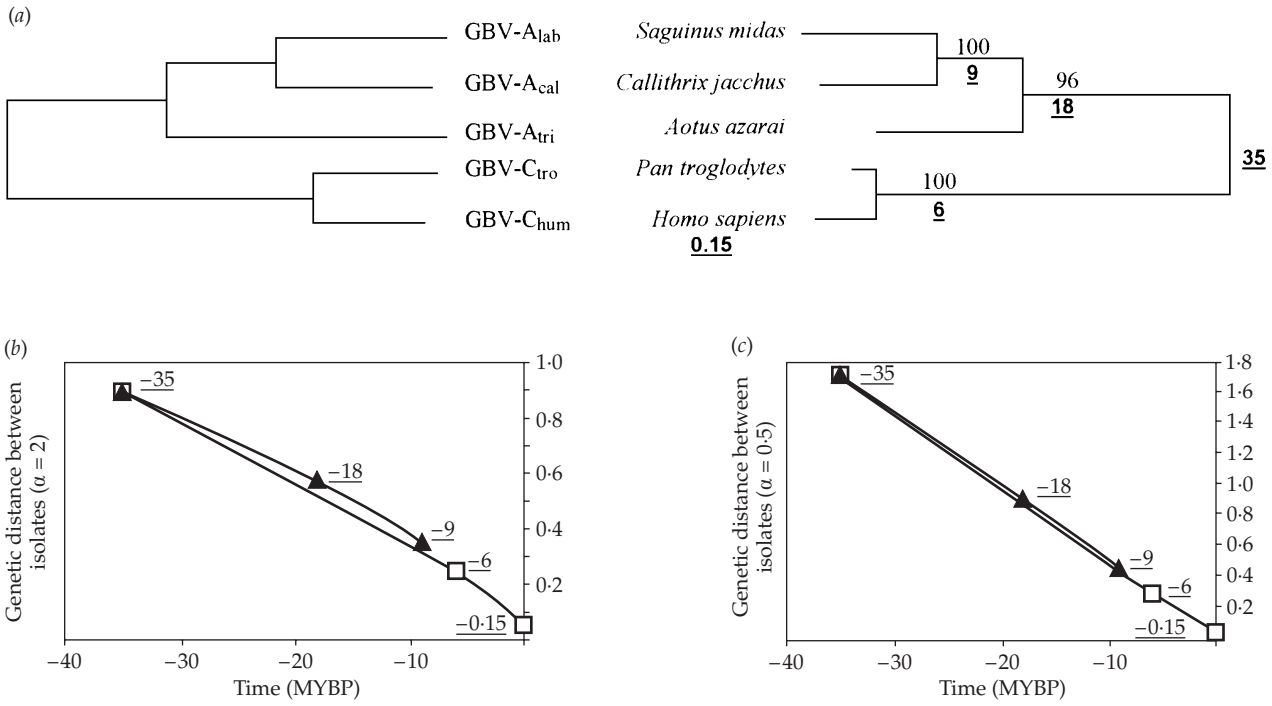


Fig. 3. Genetic evolution of GBV-A and GBV-C and of their primate hosts. (*a*) Phylogenetic trees constructed from complete GBV-A and GBV-C polyproteins (left) and from ε-globin nucleotide sequences of their primate hosts (right). The gamma distance algorithm ($\alpha = 2$) and neighbour-joining method were used. Bootstrap values are indicated above the branches. Times (in millions of years before the present) of splitting events during primate evolution are indicated below the branches and are underlined. *S. midas* is the surrogate species for *S. labiatus* and *A. azarai* is the surrogate species for *A. trivirgatus*. (*b*)–(*c*) Genetic evolution of GBV-A (▲) and GBV-C (□) over a period of 35 million years. Genetic distances between virus isolates (*y*-axis) are reported as a function of the times of splitting during primate evolution (*x*-axis). The genetic amino acid distances were calculated with the gamma distance algorithm and α values of 2 (*b*) or 0·5 (*c*).

splitting event between two given primate species during their evolution (*x*-axis) and the present-day genetic distances between virus strains that infect these two host species (*y*-axis). The genetic distance corresponding to the split between New and Old World primates, estimated at 35 million years before the present (MYBP) (Fleagle, 1988), has been determined by using the average genetic distance (gamma distance with α values of 2 and 0·5) observed between the complete GBV-C (human and chimpanzee isolates) and GBV-A sequences. Therefore, the GBV-A and GBV-C curves share the same origin. The evolutionary split between callitrichines and *Aotus* has been estimated at 18 MYBP (Porter *et al.*, 1997) and that between *Saguinus* (tamarins) and *Callithrix* (marmosets) at 9 MYBP (Porter *et al.*, 1997; Schneider *et al.*, 1993). On the basis of fossil records and molecular evidence, the split between *Pan* and *Homo* has been estimated at around 6 MYBP (Hill & Ward, 1988; Horai *et al.*, 1995) and the emergence of *Homo sapiens* occurred around 0·15 MYBP (Cann *et al.*, 1987). The calculation of the genetic distances between virus isolates was performed by using the gamma distance with α values of 2 (Fig. 3*b*) and 0·5 (Fig. 3*c*).

The resulting curves show that the rate of evolution in both virus lineages is globally a linear function of time. This suggests that the accumulation of genetic diversity has occurred at a constant rate and demonstrates that virus evolution has followed a similar model in the two lineages. This might not be expected in a case of cross-species virus transmission. By contrast, these data are consistent with the hypothesis of a cospeciation mechanism between the viruses and their respective primate hosts. This mechanism of evolution has been described for other viruses, such as arenaviruses (Bowen *et al.*, 1997) and hantaviruses (Morzunov *et al.*, 1998).

The curves illustrating the evolution of GBV-A and GBV-C lineages present similar slopes, suggesting that both viruses evolved at the same rate in their respective hosts. Between 35 and 6 MYBP, this rate is 3×10^{-8} amino acid substitutions per site per year. In the last 150000 years, the rate of evolution of human GBV-C increased to 3×10^{-7} amino acid substitutions per site per year. This phenomenon could be related to the important growth in the population of *Homo sapiens*. Numerous epidemiological studies have demonstrated that of the 6 billion individuals currently living on earth, more than 1 billion have been infected with GBV-C. This implies a formidable increase in the virus population of human GBV-C. According to the data obtained using a gamma distribution with $\alpha = 2$ for the

calculation of genetic distances, the increase in the rate of evolution may have begun with the emergence of the very first human ancestors of *Homo sapiens*, 6 MYBP.

Globally, these data support the hypothesis of a slow genetic evolution of the GBV-C lineage. This result is consistent with the findings of Adams *et al.* (1998), who studied the evolution of GBV-C in humans and chimpanzees. Nakao *et al.* (1997) reported a higher rate of evolution (2×10^{-4} amino acid substitutions per site per year) based on analysis of virus sequences obtained from a haemodialysed patient over an 8·4 year follow-up. However, it has been demonstrated that extrapolation of the rate of non-synonymous substitution over a short period of time to reconstruct the history of viruses may tend to underestimate the actual time of divergence (Smith *et al.*, 1997), probably because this kind of study does not take into account inter-individual virus transmission, which is a crucial parameter of virus dynamics.

This work was supported by a grant from the French Blood Agency (Agence Française du Sang). We thank Sophie Le Pogam, Houssam Attoui and Chuck Fulhorst for critical review of the manuscript and helpful discussion. R.N.C. is partly supported by grants from the French Foreign Affairs Ministry (Bourse Lavoisier), Servier Laboratories and the Philippe Foundation.

References

Adams, N. J., Prescott, L. E., Jarvis, L. M., Lewis, J. C., McClure, M. O., Smith, D. B. & Simmonds, P. (1998). Detection in chimpanzees of a novel flavivirus related to GB virus-C/hepatitis G virus. *Journal of General Virology* **79**, 1871–1877.

Birkenmeyer, L. G., Desai, S. M., Muerhoff, A. S., Leary, T. P., Simons, J. N., Montes, C. C. & Mushahwar, I. K. (1998). Isolation of a GB virus-related genome from a chimpanzee. *Journal of Medical Virology* **56**, 44–51.

Bowen, M. D., Peters, C. J. & Nichol, S. T. (1997). Phylogenetic analysis of the *Arenaviridae*: patterns of virus evolution and evidence for cospeciation between arenaviruses and their rodent hosts. *Molecular Phylogenetics and Evolution* **8**, 301–316.

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.

Cann, R. L., Stoneking, M. & Wilson, A. C. (1987). Mitochondrial DNA and human evolution. *Nature* **325**, 31–36.

Erker, J. C., Desai, S. M., Leary, T. P., Chalmers, M. L., Montes, C. C. & Mushahwar, I. K. (1998). Genomic analysis of two GB virus A variants isolated from captive monkeys. *Journal of General Virology* **79**, 41–45.

Fleagle, J. G. (1988). *Primate Adaptation and Evolution*. San Diego: Academic Press.

Hill, A. & Ward, S. (1988). Origin of the *Hominidae*: the record of African large hominoid evolution between 14 myr and 4 myr. *Yearbook of Physical Anthropology* **31**, 49–83.

Horai, S., Hayasaka, K., Kondo, R., Tsugane, K. & Takahata, N. (1995). Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs. *Proceedings of the National Academy of Sciences, USA* **92**, 532–536.

Jacobs, S. C., Larson, A. & Cheverud, J. M. (1995). Phylogenetic relationships and orthogenetic evolution of coat color among tamarins (genus *Saguinus*). *Systematic Biology* **44**, 515–532.

Kumar, S., Tamura, K. & Nei, M. (1993). MEGA: Molecular Evolutionary Genetics Analysis package, version 1.02. Pennsylvania State University, University Park, PA, USA.

Leary, T. P., Desai, S. M., Yamaguchi, J., Chalmers, M. L., Schlauder, G. G., Dawson, G. J. & Mushahwar, I. K. (1996). Species-specific variants of GB virus A in captive monkeys. *Journal of Virology* **70**, 9028–9030.

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.

Linnen, J., Wages, J., Jr, 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., Jr, Hoover, C., Fernandez, J., Chen, S., Zou, J. C., Morris, T., Hyams, K. C., Ismay, S., Lifson, J. D., Kim, J. P. and others (1996). Molecular cloning and disease association of hepatitis G virus: a transfusion-transmissible agent. *Science* **271**, 505–508.

Morzunov, S. P., Rowe, J. E., Ksiazek, T. G., Peters, C. J., St Jeor, S. C. & Nichol, S. T. (1998). Genetic analysis of the diversity and origin of hantaviruses in *Peromyscus leucopus* mice in North America. *Journal of Virology* **72**, 57–64.

Nakao, H., Okamoto, H., Fukuda, M., Tsuda, F., Mitsui, T., Masuko, K., Iizuka, H., Miyakawa, Y. & Mayumi, M. (1997). Mutation rate of GB virus C/hepatitis G virus over the entire genome and in subgenomic regions. *Virology* **233**, 43–50.

Porter, C. A., Czelusniak, J., Schneider, H., Schneider, M. P. C., Sampaio, I. & Goodman, M. (1997). Sequences of the primate ε-globin gene: implications for systematics of the marmosets and other New World primates. *Gene* **205**, 59–71.

Schneider, H., Schneider, M. P. C., Sampaio, I., Harada, M. L., Stanhope, M., Czelusniak, J. & Goodman, M. (1993). Molecular phylogeny of the New World monkeys (Platyrrhini, primates). *Molecular Phylogenetics & Evolution* **2**, 225–242.

Simmonds, P., Holmes, E. C., Cha, T.-A., Chan, S.-W., McOmish, F., Irvine, B., Beall, E., Yap, P. L., Kolberg, J. & Urdea, M. S. (1993). Classification of hepatitis C virus into six major genotypes and a series of subtypes by phylogenetic analysis of the NS-5 region. *Journal of General Virology* **74**, 2391–2399.

Simmonds, P., Smith, D. B., McOmish, F., Yap, P. L., Kolberg, J., Urdea, M. S. & Holmes, E. C. (1994). Identification of genotypes of hepatitis C virus by sequence comparisons in the core, E1 and NS-5 regions. *Journal of General Virology* **75**, 1053–1061.

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*a*). Isolation of novel virus-like sequences associated with human hepatitis. *Nature Medicine* **1**, 564–569.

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. and others (1995*b*). Identification of two flavivirus-like genomes in the GB hepatitis agent. *Proceedings of the National Academy of Sciences, USA* **92**, 3401–3405.

Smith, D. B., Mellor, J., Jarvis, L. M., Davidson, F., Kolberg, J., Urdea, M., Yap, P.-L., Simmonds, P. & The International HCV Collaborative Study Group (1995). Variation of the hepatitis C virus 5′ non-coding region: implications for secondary structure, virus detection and typing. *Journal of General Virology* **76**, 1749–1761.

Smith, D. B., Cuceanu, N., Davidson, F., Jarvis, L. M., Mokili, J. L. K., Hamid, S., Ludlam, C. A. & Simmonds, P. (1997). Discrimination of

hepatitis G virus/GBV-C geographical variants by analysis of the 5′ non-coding region. *Journal of General Virology* **78**, 1533–1542.

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.

Yang, Z. (1993). Maximum-likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites. *Molecular Biology and Evolution* **10**, 1396–1401.

Received 13 January 1999; Accepted 6 May 1999