Evidence of recombination in natural populations of hepatitis A virus

Mauro Costa-Mattioli, a,b Virginie Ferre, a Didier Casane, c Raoul Perez-Bercoff, d Marianne Coste-Burel, a Berthe-Marie Imbert-Marcille, a Elisabeth Claude Monique Andre, a Celine Bressollette-Bodin, a Sylviane Billaudel, a and Juan Cristina b,*

a Laboratoire de Virologie UPRES-EA1156, Institut de Biologie, Centre Hospitalier Regional Universitaire de Nantes, Rue Quai Moncousu, 9, 44093 Nantes, France
b Laboratorio de Virologia Molecular, Centro de Investigaciones Nucleares, Facultad de Ciencias, Universidad de la Republica, Iguazu 4225, 11400 Montevideo, Uruguay
c Phylogenie, Bioinformatique et Genome, UMR7622, Universite Pierre et Marie Curie, Paris, France
d Laboratoire de Genetique des Virus, CNRS, 1, Avenue de la Terrasse, 91198 Gif-sur-Yvette, France

Received 27 September 2002; returned to author for revision 15 November 2002; accepted 12 December 2002

Abstract

Genetic analysis of selected genome regions of hepatitis A virus (HAV) suggested that distinct genotypes of HAV could be found in different geographical regions. At least seven HAV genotypes have been identified all over the world, including four human genotypes (I, II, III, and VII) and three simian strains (IV, V, and VI). Phylogenetic analysis using full-length VP1 sequences revealed that human strain 9F94 has a close genetic relation with strain SLF-88 (sub-genotype VII). Nevertheless, the same analysis using full-length VP2 or VP3 sequences revealed that strain 9F94 has a close genetic relation with strain MBB (sub-genotype IB). To test the possibility of genetic recombination, phylogenetic studies were carried out, revealing that a crossing over had taken place in the VP1 capsid protein. These findings indicate that capsid-recombination can play a significant role in shaping the genetic diversity of HAV and, as such, can have important implications for its evolution, biology, and control.

© 2003 Elsevier Science (USA). All rights reserved.

Introduction

Human hepatitis A virus (HAV) is a hepatotropic member of the family Picornaviridae (Matthews, 1982; Melnick, 1982). Despite its overall physical and epidemiologic similarity to enteroviruses, the structural composition of HAV, its tissue tropism, and genetic distance from other members of the family indicate that HAV is unique within this family (Cuthbert, 2001; Hollinger and Ticehurst, 1996). Early comparative studies of the nucleotide sequences of different HAV strains suggested that isolates of diverse origin were closely related (Robertson et al., 1991). More recently nucleotide sequencing of variable genome regions, encoding the VP1 amino-terminus and the putative VP1/2A junction of wild-type HAV strains from different regions of the world, has demonstrated substantial sequence heterogeneity (Costa-Mattioli et al., 2001a,b, 2002; Jansen et al., 1990; Robertson et al., 1991, 1992). On the basis of the genetic variability observed within the putative VP1/2A junction, seven HAV genotypes have been identified worldwide (Robertson et al., 1992). However, an extensive molecular epidemiology study of HAV strains recovered from different countries revealed that HAV strains clustered in five distinct and well-supported genetic lineages (Costa-Mattioli et al., 2002).

RNA viruses exploit all known mechanisms of genetic variation to ensure their survival (Domingo and Holland, 1997). Their high rate of mutation and replication allow them to move through sequence space at a pace that often makes their DNA-based host’s evolution look glacial by comparison (Worobey and Holmes, 1999). Over the last two decades it has become increasingly clear that many RNA viruses add the capacity to exchange genetic material with one another. So, in addition to producing large amounts of
the raw material of evolution (mutations), these viruses also possess mechanisms that, in principle, allow them both to purge their genomes of accumulated deleterious changes (Muller, 1964) and to create or spread beneficial combinations of mutations in an efficient manner.

Genetic exchange among HAV strains has been observed in cell culture (Beard et al., 2001; Lemon et al., 1991). Recently, genetic recombination using an HAV subgenomic replicon constructed by replacing the P1 domain encoding region protein with the firefly luciferase sequence was shown (Gauss-Muller and Kusov, 2002). Here we report evidence of recombination in natural populations of HAV. Recombination involving two different strains can be detected when phylogenetic analysis reveals different evolutionary histories for different genome regions as, for example, in HIV (Robertson et al., 1995). In the present study, we have obtained sequence data corresponding to full-length coding sequences of VP1, VP2, and VP3 of HAV strain 9F94, recently isolated from a patient in Morocco, and compared them with corresponding sequences from HAV strains of all available genotypes. This work revealed a different evolutionary history for 9F94 VP1 than for VP2, VP3, and VP1/2A protein regions and a recombination event inside the VP1 protein. As consequence, genetic recombination provides an additional mechanism of evolution for HAV that was not previously appreciated and presents an added complication to the genetic classification of HAV.

Results

Phylogenetic analysis of HAV strains

HAV strains can be segregated into seven genotypes based on comparison of a short, selected, 168-long nucleic acid sequence of the HAV genome. Based on comparison of this short sequence, a recently recovered HAV isolate called 9F94 was classified as genotype II (Costa-Mattioli et al., 2002). We were interested in testing this classification by further sequence analysis, and therefore, complete sequences spanning the entire capsid coding region were obtained and used throughout all the recombination tests.

These sequences were aligned with sequences from 10 isolates of HAV strains isolated elsewhere for whom complete VP1, VP2, and VP3 sequences were available (for names, types, and accession numbers, see Table 1). Once aligned, using the Kimura two-parameters model and the neighbor-joining method, phylogenetic trees were obtained for VP1, VP2, and VP3 genes (Figs. 1A, B, and C, respectively). As can be seen in these figures, the phylogenetic analysis for the VP1 gene revealed that 9F94 (which was presumptively classified as genotype II based on the preliminary 168-nt VP/2A sequence; Costa-Mattioli et al., 2002) has a close genetic relation with strain SLF-88 (genotype VII, see Fig. 1A). However, the phylogenetic analysis done using full-length VP2 or VP3 sequences revealed that 9F94 was more genetically related to MBB (genotype IB) than to SLF-88 (see Figs. 1B and C). The apparent discrepancy among the results found with VP1 and those with VP2 and VP3 could be explained if a recombination event took place between putative parental strains comparable to MBB and SLF-88 (both isolated in the north of Africa).

Phylogenetic profile and RIP analysis

To gain insight into a possible recombination event, a phylogenetic profile analysis was carried out for all P1 HAV sequences included in this study. The phylogenetic profile derived from this data set is shown in Fig. 2. As can be seen in the figure, profile analysis of the putative parental-like (MBB, SLF88) and recombinant (9F94) strains show a clearly visible point of recombination on position 1461 of the analyzed sequences, corresponding to position 51 of the VP1 protein (see Fig. 2A). However, when the parental sequences SLF88, MBB, and the recombinant, 9F94, were omitted and the same studies were done with all the other sequences included in this work, no such obvious recombination junction was detected (see Fig. 2B). To confirm these results, we performed the same analysis, but this time used the recombination identification program (RIP) (Fig. 3). As can be seen in the figure, the data obtained from the RIP suggested that the 9F94 strain is the evolutionary product of a recombination event between the putative parent strain SLF88 and MBB.

Breakpoint analysis

To further characterize the putative recombinant site, we have employed the LARD method (Holmes et al., 1999). Simulations of sequence evolution under the null hypothesis (i.e., no recombination) gave strong statistical support for the alternative hypothesis of recombination ($P < 0.005$). The visual examination of the distribution of the likelihood ratio expected by chance under the null hypothesis indicates

<table>
<thead>
<tr>
<th>Strain</th>
<th>Genotype</th>
<th>Location</th>
<th>Year</th>
<th>Access No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>HM-175</td>
<td>IA</td>
<td>Australia</td>
<td>1976</td>
<td>M14707</td>
</tr>
<tr>
<td>FH2Japan</td>
<td>IA</td>
<td>Japan</td>
<td>2000</td>
<td>AB020568</td>
</tr>
<tr>
<td>HAF-203</td>
<td>IB</td>
<td>Brazil</td>
<td>2001</td>
<td>AF268396</td>
</tr>
<tr>
<td>MBB</td>
<td>IB</td>
<td>North Africa</td>
<td>1978</td>
<td>M20273</td>
</tr>
<tr>
<td>CR326</td>
<td>IA</td>
<td>Costa Rica</td>
<td>1976</td>
<td>M10033</td>
</tr>
<tr>
<td>SLF-88</td>
<td>VII</td>
<td>Sierra Leone</td>
<td>1988</td>
<td>AY032861</td>
</tr>
<tr>
<td>PA21</td>
<td>IIIA</td>
<td>Panama</td>
<td>1980</td>
<td>M45084</td>
</tr>
<tr>
<td>GA76</td>
<td>IIIA</td>
<td>USA</td>
<td>1976</td>
<td>M66695</td>
</tr>
<tr>
<td>AGM-27</td>
<td>V</td>
<td>Kenya</td>
<td>1985</td>
<td>D00926</td>
</tr>
<tr>
<td>CY-145</td>
<td>IV</td>
<td>Philippines</td>
<td>1988</td>
<td>M59286</td>
</tr>
<tr>
<td>P27</td>
<td>IIIA</td>
<td>France</td>
<td>1999</td>
<td>AI519486</td>
</tr>
<tr>
<td>9F94</td>
<td>II</td>
<td>Morocco</td>
<td>1994</td>
<td>AI519487</td>
</tr>
</tbody>
</table>

* According to Robertson et al. classification (Robertson et al., 1992).
that the likelihood ratio was obtained with the real data set (LR = 70 give strong evidence for recombination at position 1461) (see Fig. 4).

Profiles of synonymous and nonsynonymous substitution in VP1, VP2, and VP3 among parental and recombinant HAV

To gain insight into how recombination may affect the mode of evolution of HAV, the variation in the rates of synonymous (i.e., no amino acid coding change) and nonsynonymous (i.e., changes the amino acid coding assignment) substitutions among parental and the putative mosaic HAV strains were calculated for the VP1, VP2, and VP3 proteins (Fig. 5). Synonymous distances are clearly significantly higher than nonsynonymous ones for almost all pairwise comparisons (see Fig. 5). As a consequence, the ratio of nonsynonymous-to-synonymous amino acid substitutions \( \left( k_a/k_s \right) \) is very low along all proteins studied. This has usually been associated with purifying selection model acting at the level of amino acid conservation. There is no evidence for adaptive selection in any of the protein domains.

Interestingly, the rates of synonymous substitutions in MBB-9F94 comparisons for VP2 and VP3 proteins are significantly lower in both cases, while significantly higher rates are obtained when the VP1 analysis is performed (compare Fig. 5A and Figs. 5B and C). This likely is further evidence of a different evolutionary history for the VP1 protein. Nevertheless, these studies also show that even though recombination took place in the VP1 protein, it has not produced a drastic change in the mode of evolution of VP1 protein, since the nonsynonymous substitution rate was maintained at a very low rate, as in the cases of VP2 and VP3 (see Fig. 5). Thus, at least on this basis, the encoded VP1 protein does not appear to have been perturbed by the recombination event.

Discussion

HAV strains present in human populations have been demonstrated to have substantial sequence heterogeneity (Costa-Mattioli et al., 2001a,b, 2002; Jansen et al., 1990; Robertson et al., 1991, 1992). On the basis of the nucleic
acid heterogeneity, identified by sequencing discrete, partial, selected regions of the genome (similar to VP1 amino-terminal region and/or the putative VP1/2A region), HAV isolates have been differentiated into seven genotypes (Robertson et al., 1991).

In the present study, analysis of complete sequences for the VP1, VP2, and VP3 genes of recently isolated cases of HAV provided the opportunity to test whether recombination may play a role in HAV genetic diversity. Recently, an HAV subgenomic replicon was rescued by genetic recombination HAV genomes carrying lethal mutations in domain P3 (Gauss-Muller and Kusov, 2002). This HAV replicon does not contain structural proteins (P1) which were replaced by the luciferase gene. Although this work strongly indicates that HAV recombination took place, the replicon system is not sufficient for detecting a recombination event occurring in the capsid region.

Our results provide strong evidence that not only does recombination occur between two genotypes of HAV (see Figs. 2, 3, and 4), but also that it occurs in natural populations, since culture passage was not done in any of our experiments. The apparent inadequacy of the short sequence fragments providing a complete picture of recombinant HAV strains of genetic relationships demonstrates the value of obtaining complete sequences for multiple HAV genes.

Recombination in vivo in members of the family Picornaviridae has been previously observed (Furione et al., 1993; Lipskaya et al., 1991; Minor et al., 1986). Kew and Nottay (1984) reported the isolation of a recombinant polio virus which contained sequences derived from all three serotypes of poliovirus vaccine strains as a result of two crossovers. More recently, recombination among vaccine and wild-type polioviruses has been reported as a natural means of evolution of poliovirus (Guillot et al., 2000).

Recombination has also been reported in other members of the family, such as enteroviruses (Santti et al., 1999). Recombination in poliovirus or enterovirus (even between enterovirus; Kew et al., 2002) was first reported to occur outside the capsid region (Cuervo et al., 2001; Guillot et al., 2000; Kew et al., 2002; Kew and Nottay, 1984; Liu et al., 2000). Recently, polio (Liu et al., 2000; Martin et al., 2002) and foot and mouth disease virus capsid recombinants (Tosh et al., 2002) have been reported. These results are in agreement with those found in this HAV work. In the present study, we report a recombinant event among VP1 genes between two different genotypes of HAV. Recombination may serve two opposite purposes: exploration of a new combination of genomic region from different origins or rescuing of viable genomes from debilitated parental genomes (Domingo and Holland, 1997; Lai, 1992). The epidemiologic and clinical implication of HAV capsid region crossing over in particular are unknown. However, capsid recombinant may provide selective advantage to the virus; for example, the recombinant may be more transmissible than the preexisting lineages.

As methods of sequence analysis are increasingly able to distinguish recombination from other evolutionary processes, it is expected that recombination events will be found more often and in additional RNA viruses. Due to the increased prevalence and hyperendemicity of HAV in some regions of the world, such as north of Africa (where both the putative parental-like strains MBB and SLF-88 and the recombinant, 9F4, have been isolated), or South America, where recent studies suggest a changing epidemiologic pattern (Tanaka, 2000; Tapia-Conyer et al., 1999), it will be important to analyze additional isolates to investigate the frequency and impact of recombination in HAV.

Four human HAV genotypes (I, II, III, and VII) have...
been proposed to date on a basis of pairwise comparison of a short, selected, VP1/2A 168-long nucleic acid sequence of HAV genome (Robertson et al., 1992). One main assumption of this analysis is that the short sequence is representative of the genome. However, recombination invalidates this assumption by generating mosaic genomes, where different regions have different phylogenetic histories. If we follow the classification criteria cited above, strain 9F94 should be classified as type II (Costa-Mattioli et al., 2002). However, the complete sequence of VP1 showed that strain 9F94 more closely resembled that of strain SLF-88 of genotype VII, and analysis of VP2 and VP3 revealed the relatedness of these regions to strain MBB of genotype IB. Since only one other representative of type II has been identified to date, and since that identification is based on limited sequence analysis (Robertson et al., 1992), the possibility exists that type II strains may not be a distinct human genotype, but rather an artifact of classification based on insufficient sequence analysis.

The recognition of recombination is important not only for unraveling the phylogenetic history of genes, but also for molecular phylogenetic inference. By ignoring the presence of recombination, phylogenetic analysis may be severely compromised (Posada and Crandall, 2001; Schierup and Hein, 2000). For that reason, HAV phylogeny needs to be revised in light of recombination.

Although there is good evidence that human strains are closely related antigenically (Lemon, 1997), the possible implication of HAV capsid recombination events generating antigenic escape mutants cannot be ruled out. However the distribution of nonsynonymous substitutions shows extremely low rates when compared with synonymous ones (see Fig. 5). This suggests that the pattern of divergence observed in HAV capsid proteins is due to negative selection against substitutions at the amino acid level, which appears to be the main force shaping the pattern of nonsynonymous substitutions. In contrast, the antigenic immunogenic sites of multiple serotype viruses, such as the hemagglutinin gene of influenza virus (Ina and Gojobori, 1994), the complete capsid region of serotype A and C of FMDV (Haydon et al., 2001), and the VP3 region of HIV (Seibert et al., 1995), were subjected to positive selection by immune pressure. Whether new HAV variants may appear, as a result of recombination events for instance, it remains to be established if their fitness permits them to be selected in HAV population. However, the mode of HAV evolution appears, at least in part, to explain why there is only one serological group of HAV described so far.

Materials and methods

Virus strains

9F94 was recovered from a virus collection of the “Unité de Virologie” at the Centre Hospitalier Universitaire, Nantes. This strain comes from a 6-year-old French girl who was hospitalized at the Centre Hospitalier Universitaire in 1994, after a 3-month holiday in Morocco.

RNA extraction

HAV RNA was extracted from 140 μl serum samples with the QIAamp Viral RNA Kit (QIAGEN) and from 400 μl stool suspension (0.3 g stool diluted in 2 ml water) using the RNeasy mini kit (QIAGEN) according to instructions.
provided by the manufacturers. The RNA extracted was eluted from the columns with 50 μl RNase-free water.

**Reverse transcription**

A RNA/primer mixture was prepared in a sterile 0.2-ml tube as follows: 1 μl of 10 mM stock solution of dNTPs (Roche), 1 μl of oligo(dT)$_{12-18}$ (0.5 μg/μl) (Invitrogen), 5 μl of RNA, and 5 μl of sterile water. The RNA/primer mixture was thawed on ice, incubated 5 min at 65°C, and chilled on ice. From a master mix, 1 μl Rnasin (20-40 units/ml) (Promega), 2 μl of 0.1 M dithiothreitol (Invitrogen), 1.5 μg of T4 gene 32 (Roche), 1 μl (200 units) Supervest II reverse transcriptase (GIBCO), and 4 μl of 5× 1st Strand Synthesis buffer (Invitrogen) was added to the RNA/primer mixture. The reaction was incubated for 1 h at 42°C and the reverse transcriptase was inactivated by heating at 70°C for 15 min. One microliter of RNase H (1-4 units/ml) (Invitrogen) was added and the reaction was incubated for 20 min at 37°C. The cDNA was purified with the QiAquick PCR purification kit (QiAGEN) used according to the manufacturer’s recommendations. The cDNA was eluted in 30 μl of 10 mM Tris–Cl (pH 8.5) and kept on ice until added to the long PCR reaction mix.

**Long PCR**

The PCR reactions were performed in thin-walled PCR tubes in a total volume of 25 μl containing with a final concentration of 400 μM of each dNTPs, 400 μM of each positive (+757) and negative (−3749) primers (see Table 1), 2.5 μl 10× Advantage 2 PCR buffer (40 mM Tricine–KOH, pH 8.7 at 25°C, 15 mM K2Ac, 3.5 mM Mg(OAc)$_2$, 3.75 μg/ml BSA, 0.005% of Tween, and 0.005% Nonidet P-40) (Clontech), 2 μl purified template cDNA, 0.5 μl Advantage Polymerase Mix [Titanium TaqDNA polymerase, a small amount of proofreading polymerase, TaqStart Antibody (1.1 μg/μl) in the following storage buffer: 1% glycerol, 0.3 mM Tris–HCl, pH 8, 1.5 mM KCl, 1 μM EDTA]. PCR reactions were performed in a 9600 Thermal cycler (Applied Biosystems) with the following cycle conditions: hot start at 94°C for 2 min, denaturation for 30s, annealing at 52 for 45 s, extension at 68°C for 3 min (15 cycles), 68°C for 3 min 30 s (10 cycles), 68°C for 4 min (10 cycles). A final extension step was performed for 10 min at 68°C.

**Purification of amplicons**

The PCR reaction was electrophoresed on a 1% agarose gel. DNA bands were visualized with a UV transilluminator. The DNA was directly purified using the ExoSAP-IT enzymatic kit (Amersham).

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
<th>Position*</th>
</tr>
</thead>
<tbody>
<tr>
<td>+757</td>
<td>TTTCCAGACTGTGGGGAGTGG</td>
<td>758–778</td>
</tr>
<tr>
<td>+1194</td>
<td>GGTTCCTGTTGACCACAC</td>
<td>1187–1201</td>
</tr>
<tr>
<td>−1249</td>
<td>AACAACCATAGGATAACAGTC</td>
<td>1249–1226</td>
</tr>
<tr>
<td>+1470</td>
<td>ATGATGAGAATGAAATTAG</td>
<td>1470–1489</td>
</tr>
<tr>
<td>−1584</td>
<td>CAGATTTCACCTCTCCTGATC</td>
<td>1584–1563</td>
</tr>
<tr>
<td>+1794</td>
<td>TGTTCCTGAGGAGAGA</td>
<td>1794–1810</td>
</tr>
<tr>
<td>HAV1</td>
<td>GGTTCCTGCTTCTTATCTCATGATG</td>
<td>2167–2192</td>
</tr>
<tr>
<td>HAV2</td>
<td>AGTCACACTCTCCTCAGGAAACTT</td>
<td>3308–3285</td>
</tr>
<tr>
<td>HAV3</td>
<td>TTCTATATTTCATGGCTCCCTC</td>
<td>3286–3627</td>
</tr>
<tr>
<td>HAV4</td>
<td>ACAAAATAAACAATGAAAGACAA</td>
<td>2955–2934</td>
</tr>
<tr>
<td>HAV5</td>
<td>TTGTCTTTTAGTGTTATTGTC</td>
<td>2934–2956</td>
</tr>
<tr>
<td>HAV6</td>
<td>AGGAAATGTCTCAGGTACTTCTTCTTGGTAAACTG</td>
<td>2414–2380</td>
</tr>
<tr>
<td>HAV7</td>
<td>CTGGTGAACACAGCATGCACATC</td>
<td>2691–2667</td>
</tr>
<tr>
<td>HAVX8</td>
<td>AGAGCATATGTAATGCGAG</td>
<td>3106–3125</td>
</tr>
<tr>
<td>HAVX9</td>
<td>GAGCAGTTGAGATGTTTGGG</td>
<td>2845–2863</td>
</tr>
<tr>
<td>HAVX10</td>
<td>TTGTCTTTATTTCCGT</td>
<td>3674–3657</td>
</tr>
<tr>
<td>−3749</td>
<td>AACAGTCACAGGATGAAATC</td>
<td>3770–3778</td>
</tr>
</tbody>
</table>

*Position relative to the genome of HAV strain HM-175 (M14707).

**Sequencing**

The primers used to amplify and sequence the long PCR fragment are described in Table 2. The sequence reaction was carried out using the Big Dye DNA sequencing kit (Perkin–Elmer) on a 373 DNA sequencer apparatus (Perkin–Elmer).

**Sequence analysis**

The sequences compared came from a single DNA copy that spans the entire capsid coding region and in all cases the long DNA was used to determine the sequences that were entered in the recombination tests. The sequences were aligned using the CLUSTAL W program (Thompson et al., 1994). Using the software MEGA (Kumar et al., 1994), phylogenetic trees were generated by the neighbor-joining method applied to distance matrix obtained under the Kimura two-parameter model (Felsenstein, 1993). As a measure of the robustness of each node, we utilized the bootstrap method (1000 pseudo-replicates).

**Substitution rate analysis**

The substitution rate along the VP2, VP3, and VP1 genes were measured using a sliding window according to the procedure used by Alvarez-Valin et al. (2000). Pairwise nucleotide distances (synonymous and nonsynonymous) within each window were estimated by the method of Comeron (1995) as implemented in the computer program k-estimator (Ina and Gojobori, 1994). For those windows where the method was inapplicable, the Jukes–Cantor (Jukes and Cantor, 1969) method was used for correction.
for multiple hits. The window had a size of 30 codons and a movement of 15.

Detection of recombination by sequence comparison

To detect any possible events of recombination, the “phylogenetic-profile” method that was recently described (Weiller, 1998) was utilized. Briefly, the latter method constitutes a novel way of graphically displaying the coherence of the sequence relationships over the entire length of a set of aligned homologous sequences and have demonstrated sensitivity and accuracy for identifying recombinant sequences and their recombination junctions as well as detecting hot spots of recombinational activity. Alternatively, RIP (Siepel et al., 1995) was used to detect possible recombinant sequences. A sliding window of 200-nucleotide base-pairs in length was created and moved over the aligned query. The threshold for statistical significance was 90%.

Breakpoint analysis

Once the recombinant strain and strains representing possible parents were identified, the likely recombination breakpoint was determined using LARD (Holmes et al., 1999). Briefly, for every possible breakpoint, the sequence alignment was divided into two independent regions for which the branch lengths of a tree of the putative recombinant and its two parent sequences were optimized. The two results (likelihoods) obtained by using the separate regions were then combined to give a likelihood score for that breakpoint position and the breakpoint position that yielded the highest likelihood was identified. This “recombination model” likelihood then was compared, by using a likelihood ratio test, to the likelihood obtained from the same data under a model that permitted no recombination. To assess whether the recombination model gave a significantly better fit to the data than the null hypothesis of no recombination, the likelihood ratio obtained by using the real data were evaluated for significance against a null distribution of likelihood ratios produced by using Monte Carlo simulation of sequences generated without recombination. Sequences were simulated 200 times by using the maximum likelihood model parameters and sequence lengths from the real data using Seq-Gen (Rambaut and Grassly, 1997). The simulated sequences then were subjected to the same breakpoint analyses as the real data to produce the distribution of likelihood ratios expected if no recombination had occurred among them.

References
