Complete sequence determination and genetic analysis of Banna virus and Kadipiro virus: proposal for assignment to a new genus (Seadornavirus) within the family Reoviridae

Houssam Attoui,1 Frédérique Billoir,2 Philippe Biagini,1 Philippe de Micco1,2 and Xavier de Lamballerie2

1 Laboratoire de Virologie Moléculaire, EFS Alpes-Méditerranée, 149 Boulevard Baille, 13005 Marseille cedex 5, France
2 Unité des Virus Emergents, Laboratoire de Virologie Moléculaire, Tropicale et Transfusionnelle, Faculté de Médecine de Marseille, 27 Boulevard Jean Moulin, 13005 Marseille cedex 5, France

Arboviruses with genomes composed of 12 segments of double-stranded (ds) RNA have previously been classified as members or probable members of the genus Coltivirus within the family Reoviridae. A number of these viruses have been isolated in North America and Europe and are serologically and genetically related to Colorado tick fever virus, the Coltivirus type species. These isolates constitute subgroup A of the coltiviruses. The complete genome sequences are now presented of two Asian arboviruses, Kadipiro virus (KDV) and Banna virus (BAV), which are currently classified as subgroup B coltiviruses. Analysis of the viral protein sequences shows that all of the BAV genome segments have cognate genes in KDV. The functions of several of these proteins were also indicated by this analysis. Proteins with dsRNA-binding domains or with significant similarities to polymerases, methyltransferases, NTPases or protein kinases were identified. Comparisons of amino acid sequences of the conserved polymerase protein have shown that BAV and KDV are only very distantly related to the subgroup A coltiviruses. These data demonstrate a requirement for the subgroup B viruses to be reassigned to a separate new genus, for which the name Seadornavirus is proposed.

Introduction

In the sixth report of the International Committee for the Taxonomy of Viruses (ICTV), a number of arboviruses with 12-segment double-stranded (ds) RNA genomes were assigned to the genus Coltivirus, family Reoviridae (Holmes et al., 1995). These coltiviruses included two distinct virus species. The first, identified as Colorado tick fever virus (CTFV), contains two serotypes: the original CTFV isolate, from an American patient with a febrile syndrome (Florio et al., 1946, 1950), and coltivirus S6-14-03 (CTFV S6-14-03), also isolated in the USA (Lane et al., 1982). The second coltivirus species was isolated in Europe and is identified as Eyach virus (EYAV) (Chastel et al., 1984; Rehse-Küpper et al., 1976). All of these viruses have been isolated from ticks and CTFV can be transmitted to humans by the bite of adult Dermacentor species, in particular Dermacentor andersoni (Emmons, 1988). The antigenic relationship between CTFV and EYAV was initially established by complement fixation (Chastel et al., 1984; Karabatsos et al., 1987; Rehse-Küpper et al., 1976).

A number of other arthropod-borne virus isolates from South-East Asia, Indonesia and China (Brown et al., 1993; Chen & Tao, 1996), containing 12-segment dsRNA genomes, were also provisionally included as coltiviruses. At least one of these viruses (Banna virus; BAV) has been isolated from humans suffering from meningoencephalitis (Chen & Tao, 1996; Xu et al., 1990). Previous investigations showed the existence of two distinct groups among the virus isolates from South-East Asia. However, serological and genomic-hybridization investigations did not identify any significant relationship between either of these two groups and the American or European Coltivirus species (Brown et al., 1993; Chen & Tao, 1996).

Since 1997, the genomic RNA sequences of CTFV, EYAV, CTFV S6-14-03 and a number of Asian virus isolates have been partially characterized (Attoui et al., 1997, 1998). This genetic
analysis has permitted taxonomic classification of different coltivirus isolates based on phylogenetic comparisons of their RNA sequences. These methods of distinguishing and identifying different virus species have been formally accepted by the ICTV in their seventh report on virus taxonomy (Mertens et al., 2000) and are included as part of the ‘species parameters’ for members of the family Reoviridae. The genus Coltivirus is currently divided into two subgroups: subgroup A includes the original American and European isolates, while subgroup B contains the Asian isolates. Subgroup B is further subdivided in two species, Kadiiro virus (KDV) and BAV. However, the incomplete nature of the RNA sequence data did not clarify fully the genetic relationship between the two Asian virus species (KDV and BAV), and the validity of including the Asian, American and European viruses within a single genus was uncertain.

We report the RNA sequences for the complete genomes of KDV and BAV. We have also sequenced a 2973 bp region of the polymerase gene of CTFV. Comparisons of these genetic data have allowed us to define precisely the taxonomic status of virus isolates previously included within the genus Coltivirus. The need to place KDV and BAV within a separate new genus of the family Reoviridae is discussed.

**Methods**

- **Virus strains.** The Florio strain (N-7180) of CTFV was purchased from the ATCC. The French isolate of EYAV (EYAV-Fr578) was provided by C. Chastel. BAV isolates from Indonesia (BAV-In6423, BAV-In6969 and BAV-In7043) and from China (BAV-Ch) and KDV from Java (KDV-Ja7075) were provided by R. B. Tesh and R. E. Shope.

- **Cell lines.** BHK-21 cells were cultured as monolayers or suspensions, as described previously (Attoui et al., 1998). C6/36 mosquito cells were cultured as monolayers in Leibovitz’s L-15 medium supplemented with 5% FBS, 2% tryptose phosphate broth, penicillin G (100 IU/ml) and streptomycin (100 µg/ml).

- **Virus propagation.** The Florio strain of CTFV was propagated in suspensions of BHK-21 cells, as described elsewhere (Attoui et al., 1998). EYAV-Fr578 was propagated by intracerebral inoculation of suckling mice and infected brains were recovered at day 7 post-infection. The BAV and KDV isolates were propagated in C6/36 mosquito cells. Infected cells were harvested at day 4 post-infection.

- **Isolation and purification of nucleic acids.** Viral dsRNA was extracted from infected cell cultures or suckling mouse brains by using a guanidinium isothiocyanate-based procedure (RNA NOW, Biogentex). Viral dsRNA from cell cultures was purified further by precipitation of single-stranded RNA in the presence of 2 M LiCl and electrophoresis on a 10% polyacrylamide gel, as described previously (Attoui et al., 1998).

- **Cloning of cDNA genome segments.** cDNA copies of the dsRNA genome segments from BAV and KDV were synthesized, cloned and sequenced by using a modification of the Lambda method (Attoui et al., 1998; Lambden et al., 1992). The key step for cDNA synthesis and cloning from these dsRNA segments was the ligation of a 3’-amino-blocked oligodeoxyribonucleotide (primer A: 5’ PO4-AGGTCTCG-TAGACGGTCACC-NH2 3’) to both 3’-OH termini of the dsRNA with 10 U T4 RNA ligase (Boehringer Mannheim). The tailed dsRNA was recovered by using the RNAid kit and denatured by heating at 99 °C for 1 min in the presence of 15% DMSO. cDNA copies of the genomic RNA were synthesized by using a complementary primer (primer B: 5’ GGTGCACGGTCTACGAGACCT 3’) and 200 U SuperScript reverse transcriptase (Gibco BRL). The cDNA was amplified by using primer B (Attoui et al., 1998) according to the single-primer amplification technique (Lambden et al., 1992). The amplicons were analysed by agarose gel electrophoresis and then ligated into the pGEM-T cloning vector. The recombinant vector was transfected into competent E. coli XL1-Blue and the insert sequence was determined by using M13 universal primers, the D-Rhodamine DNA sequencing kit and an ABI Prism 377 sequence analyser (Perkin Elmer). PCRs were carried out with 2.5 U Taq polymerase (Gibco BRL) and 0.5 µM of each primer. Thermal cycling parameters were as follows: one cycle of denaturation (90 °C, 10 min) followed by 35 cycles of denaturation (94 °C, 50 s) and extension (72 °C, 3 min). The cycling program was ended by an extension step at 72 °C for 10 min. Where the PCR amplification produced truncated products, identification of conserved terminal sequences allowed them to be mapped to either the 5’ or 3’ end of the dsRNA segment (Attoui et al., 1998). The remainder of the segment was amplified (by using primer B and an appropriate sequence-specific primer) and sequenced as described above.

- **RT–PCR amplification of partial genomic sequences.** The viral dsRNAs extracted from the BAV isolates BAV-In6969, BAV-In7043 and BAV-Ch were copied into cDNA in the presence of random hexanucleotides and SuperScript reverse transcriptase as described previously (Attoui et al., 1998). PCR primers were designed from the sequences of the first, second and sixth genomic segments of BAV, with the help of the Oligo software (National Biosciences). Primer sequences are displayed in Table 1. The protocols used for PCR amplification, cloning and sequencing are described above.

**Table 1. Primers used in the amplification of segments 1, 2 and 6 of BAV isolates**

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence (5’ → 3’)</th>
<th>Segment</th>
<th>Map position</th>
<th>Orientation</th>
</tr>
</thead>
<tbody>
<tr>
<td>BAV1S</td>
<td>TGAATCACAATCGTCTTGGCAAGGAC</td>
<td>1</td>
<td>964–988</td>
<td>Sense</td>
</tr>
<tr>
<td>BAV1R</td>
<td>CTAATCTACATGGGAGTTGACCCTTAG</td>
<td>1</td>
<td>3253–3229</td>
<td>Sense</td>
</tr>
<tr>
<td>BAV2S</td>
<td>CCTGATACCTGCTCGACAAACCTG</td>
<td>2</td>
<td>184–207</td>
<td>Sense</td>
</tr>
<tr>
<td>BAV2R</td>
<td>CAAGATATAATTCCCTGGCTAATCAC</td>
<td>2</td>
<td>2643–2620</td>
<td>Antisense</td>
</tr>
<tr>
<td>BAV6S</td>
<td>TCCCACTATATCGATGACGATTCTG</td>
<td>6</td>
<td>72–96</td>
<td>Sense</td>
</tr>
<tr>
<td>BAV6R</td>
<td>TTAATCGTGACGTCTGATGACGCC7</td>
<td>6</td>
<td>1426–1402</td>
<td>Antisense</td>
</tr>
</tbody>
</table>
Genome segment 1 dsRNA of CTFV was gel-purified and transcribed into cDNA by using random hexanucleotides. The cDNA was amplified by PCR with a series of non-specific and highly degenerate primers (derived from the polymerase sequences of flaviviruses) at a low hybridization temperature (40°C for 2 min). The longest amplicon obtained was cloned and sequenced as described above.

**Comparison of polymerase sequences.** Sequences of VP1 of BAV and KDV and a partial sequence of the CTFV VP1 were compared with the putative RNA-dependent RNA polymerases from representative members of six genera of the family *Reoviridae*. These included sequences from the following genera. Genus *Orthoreovirus* (10 dsRNA segments): mammalian reovirus serotypes 1 (Lang strain, GenBank accession number M24734), 2 (Jones strain, M31057) and 3 (Dearing strain, M31058). Genus *Orbivirus* (10 dsRNA segments): African horsesickness virus serotype 9 (U94887), bluetongue virus serotypes 2 (L20508), 10 (X12819), 11 (L20445), 13 (L20446) and 17 (L20447) and Palyam virus (BAA76549, isolate Chuzan). Genus *Fijivirus* (X12819), 11 (L20445), 13 (L20446) and 17 (L20447) and Palyam virus (BAA76549, isolate Chuzan). Genus *Rotavirus* (11 dsRNA segments) group A: bovine rotavirus strains RF (J04346) and UK (X55444), simian rotavirus strains SA11–both (X16830) and SA11 (AF019955), porcine rotavirus strain Gottfried (M32805) and avian rotavirus (BAA24146). Genus *Rotavirus* group B: human/marine rotavirus strain IDR (M97203). Genus *Rodavirus* group C: porcine rotavirus Cowden strain (M74216). Genus *Fijivirus* (10 dsRNA segments): Nilaparvata lugens reovirus (NLRV) Izumo strain (D49693). Genus *Phyloreovirus* (12 dsRNA segments): rice dwarf virus (RDV) Chinese strain (U73201), strain H (D10222) and strain A (D90198). Genus *Oryzavirus* (10 dsRNA segments): rice ragged stunt virus (RRSV) Thai strain (U66714).

**Sequence analysis.** All sequence alignments were generated by the CLUSTAL W software (Thompson et al., 1994) and the NCBI BLAST 2 program (http://www3.ncbi.nlm.nih.gov/BLAST). Phylogenetic analyses were performed with the program MEGA (Kumar et al., 1993) using the p-distance determination algorithm and sequence relatedness was reported as percentage identity or percentage genetic distance. Tree drawing was performed with the help of the TreeView program (Page, 1996). Comparison of our sequences with those available from nucleic acid and protein databases was performed by using the NCBI gapped BLAST program (http://www3.ncbi.nlm.nih.gov/BLAST). The program Pfam (http://www.sanger.ac.uk/Pfam/search.shtml) was used to search for previously described protein family domains. The program Motif (http://www.motif.genome.ad.jp) was used to analyze the theoretical protein sequences for the presence of known functional amino acid motifs.

**Results**

**Sequence determination**

The complete genome sequences of BAV-In6423 and KDV-In7075 have been characterized. Sequences of genome segments 1–12 of BAV are deposited in the GenBank database under accession numbers AF133430, AF134514–AF134518, AF052018–AF052014 and AF019908, respectively. Sequences of segments 1–12 of KDV are deposited under accession numbers AF133429, AF134509–AF134513, AF052023–AF052019 and AF019909, respectively. The total RNA segment length, the longest open reading frame, the lengths of the 5′ and 3′ non-coding regions (NCR) and the length of the putative protein encoded were identified for each segment and are shown in Table 2. The viral proteins are identified as ‘VPX’, where X refers to the number of the RNA segment based on its relative electrophoretic mobility during agarose gel electrophoresis.

Partial sequences of BAV-In6969, BAV-In7043 and BAV-Ch segments 1, 2 and 6 were also determined and have been deposited in the GenBank database (BAV-In6969, accession numbers AF134522–AF13524; BAV-In7043, AF134519–AF134521; BAV-Ch, AF134525–AF134527). The PCR products from genome segments 1 and 2 of isolates BAV-Ch, BAV-In6969 and BAV-In7043 were each 12 and three nucleotides longer, respectively, than the corresponding PCR products from BAV-In6423. The sequence of the amplicon from segment 1 of CTFV, obtained by non-specific PCR amplification, was also determined and found to contain a 2973 bp coding sequence (991 aa) (accession number AF134529).

**Sequence analysis**

Analysis of the NCRs of different genome segments has identified conserved motifs located at the termini. The motifs 5′ GUAGAA(A/U)/(A/U)/(A/U)(A/U)U 3′ and 5′ (A/U)(A/C)(C/U)GAC 3′ were found in the 5′ and 3′ NCRs, respectively, of all positive strands from the KDV RNA segments. However, the positive RNA strands of BAV contained the motifs 5′ GUAU(A/U)/(A/U)/(A/U)/(A/U)(A/U)/(A/U)U 3′ and 5′ (A/G)(A/C)(C/U)GAC 3′ in their 5′ and 3′ NCRs. BAV and KDV share the first three nucleotides at both termini (5′ GUU and GAC 3′) and the 5′ and 3′ terminal dinucleotides of all segments are inverted complements. Comparison of amino acid sequences showed that all of the genome segments of BAV have cognate genes in KDV. However, the equivalent viral proteins are not necessarily encoded by RNA segments with the same size or order of electrophoretic migration. Fig. 1 shows the relationships between the cognate segments of BAV and KDV. Calculations of phylogenetic distances showed that the amino acid identity between homologous proteins ranged from 24 to 42% with similarity ranging from 38 to 57% (Fig. 1). The BLAST 2 program did not detect any significant similarity between BAV VP12 and KDV VP8. However, the amino acid identity calculated from the alignment produced by the CLUSTAL W program was 27%.

The partial sequences of VP1, VP2 and the full-length VP6 from BAV-In6969, BAV-In7043 and BAV-Ch were compared with their homologous proteins from BAV-In6423. This revealed a high degree of conservation, with overall amino acid identities ranging from 94 to 99% for VP1 (aa 338–808), from 93 to 99% for VP2 (aa 40–851) and from 93 to 99% for VP6. Four additional amino acids (RIGT) were detected at position 605 in VP1 and an additional glutamine was detected at position 423 of VP2 from BAV-In6969, BAV-In7043 and BAV-Ch.

The partial amino acid sequence of VP1 (991 aa) was also deduced from the nucleotide sequence of CTFV genome.
Table 2. Lengths of dsRNA segments 1–12, putative encoded proteins and 5’ and 3’ NCRs of BAV and KDV

<table>
<thead>
<tr>
<th>Segment</th>
<th>Length (bp)</th>
<th>Protein (aa)</th>
<th>5’ NCR</th>
<th>Terminal sequence</th>
<th>3’ NCR</th>
<th>Length (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BAV</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>3747</td>
<td>1214</td>
<td>23</td>
<td>5’GUAAuuaaaoU...</td>
<td>...aaCuGAC3’</td>
<td>79</td>
</tr>
<tr>
<td>2</td>
<td>3048</td>
<td>954</td>
<td>93</td>
<td>5’GUAAuuaaaaoU...</td>
<td>...acCcGAC3’</td>
<td>90</td>
</tr>
<tr>
<td>3</td>
<td>2400</td>
<td>720</td>
<td>23</td>
<td>5’GUAAuuaaaaoU...</td>
<td>...aaCuGAC3’</td>
<td>214</td>
</tr>
<tr>
<td>4</td>
<td>2038</td>
<td>576</td>
<td>26</td>
<td>5’GUAAuuaaaaoU...</td>
<td>...acCcGAC3’</td>
<td>281</td>
</tr>
<tr>
<td>5</td>
<td>1716</td>
<td>508</td>
<td>31</td>
<td>5’GUAAuuaaaaoU...</td>
<td>...aaGAC3’</td>
<td>158</td>
</tr>
<tr>
<td>6</td>
<td>1671</td>
<td>425</td>
<td>111</td>
<td>5’GUAAuuaaaaoU...</td>
<td>...gaCuGAC3’</td>
<td>285</td>
</tr>
<tr>
<td>7</td>
<td>1136</td>
<td>306</td>
<td>63</td>
<td>5’GUAAuuaaaaoU...</td>
<td>...aaCuGAC3’</td>
<td>152</td>
</tr>
<tr>
<td>8</td>
<td>1119</td>
<td>302</td>
<td>32</td>
<td>5’GUAAuuaaaaoU...</td>
<td>...gaCcGAC3’</td>
<td>178</td>
</tr>
<tr>
<td>9</td>
<td>1101</td>
<td>283</td>
<td>23</td>
<td>5’GUAAuuaaaaoU...</td>
<td>...gaGAC3’</td>
<td>226</td>
</tr>
<tr>
<td>10</td>
<td>977</td>
<td>249</td>
<td>28</td>
<td>5’GUAAuuaaaaoU...</td>
<td>...aaGAC3’</td>
<td>199</td>
</tr>
<tr>
<td>11</td>
<td>867</td>
<td>180</td>
<td>75</td>
<td>5’GUAAuuaaaaoU...</td>
<td>...aaGAC3’</td>
<td>249</td>
</tr>
<tr>
<td>12</td>
<td>862</td>
<td>207</td>
<td>43</td>
<td>5’GUAAuuaaaaoU...</td>
<td>...aaGAC3’</td>
<td>195</td>
</tr>
<tr>
<td>Consensus</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>5’GUAAWWWwwwU...</td>
<td>…RMCYGAC3’</td>
<td>–</td>
</tr>
<tr>
<td>KDV</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>3774</td>
<td>1223</td>
<td>26</td>
<td>5’GUAGAAaauu...</td>
<td>...aaacuGAC3’</td>
<td>76</td>
</tr>
<tr>
<td>2</td>
<td>3035</td>
<td>968</td>
<td>44</td>
<td>5’GUAGAAaauu...</td>
<td>...aaacuGAC3’</td>
<td>84</td>
</tr>
<tr>
<td>3</td>
<td>2415</td>
<td>739</td>
<td>44</td>
<td>5’GUAGAAaauu...</td>
<td>...aaacuGAC3’</td>
<td>151</td>
</tr>
<tr>
<td>4</td>
<td>2121</td>
<td>631</td>
<td>25</td>
<td>5’GUAGAAaauu...</td>
<td>...aaacuGAC3’</td>
<td>200</td>
</tr>
<tr>
<td>5</td>
<td>1947</td>
<td>573</td>
<td>101</td>
<td>5’GUAGAAaauu...</td>
<td>…aaacuGAC3’</td>
<td>124</td>
</tr>
<tr>
<td>6</td>
<td>1670</td>
<td>506</td>
<td>25</td>
<td>5’GUAGAAaauu...</td>
<td>…aaacuGAC3’</td>
<td>124</td>
</tr>
<tr>
<td>7</td>
<td>1259</td>
<td>347</td>
<td>32</td>
<td>5’GUAGAAaauu...</td>
<td>…aaacuGAC3’</td>
<td>183</td>
</tr>
<tr>
<td>8</td>
<td>1114</td>
<td>302</td>
<td>86</td>
<td>5’GUAGAAaauu...</td>
<td>…aaacuGAC3’</td>
<td>119</td>
</tr>
<tr>
<td>9</td>
<td>1054</td>
<td>303</td>
<td>23</td>
<td>5’GUAGAAaauu...</td>
<td>…aaacuGAC3’</td>
<td>119</td>
</tr>
<tr>
<td>10</td>
<td>946</td>
<td>260</td>
<td>17</td>
<td>5’GUAGAAaauu...</td>
<td>…aaacuGAC3’</td>
<td>146</td>
</tr>
<tr>
<td>11</td>
<td>894</td>
<td>263</td>
<td>23</td>
<td>5’GUAGAAaauu...</td>
<td>…aaacuGAC3’</td>
<td>79</td>
</tr>
<tr>
<td>12</td>
<td>756</td>
<td>190</td>
<td>102</td>
<td>5’GUAGAAaauu...</td>
<td>…aaacuGAC3’</td>
<td>81</td>
</tr>
<tr>
<td>Consensus</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>5’GUAGAAWWWwwwU...</td>
<td>…WVMMYGAC3’</td>
<td>–</td>
</tr>
</tbody>
</table>

Highly conserved terminal sequences are shown in upper-case. In consensus sequences, M represents A or C; R represents A or G; V represents A, C or G; W represents A or U and Y represents C or U.

Comparison of the BAV and KDV sequences with the available sequences of CTFV (segments 9, 10, 11 and 12) and EYAV (segment 12) showed a maximum value for amino acid identity of 15%. This level of similarity is comparable to values found between viral proteins from different genera within the family Reoviridae. The amino acid sequence identity between homologous proteins from distinct species within a single genus is defined as between 20 and 45% (Mertens et al., 2000).

Functional analysis

This analysis was carried out by using two different approaches. The first approach was to characterize functional amino acid motifs in the viral proteins. The second was to search for similarities with proteins that have known functional activities. The first structural feature noted in VP1 was a GDD motif at aa 755–757 of BAV and positions 748–750 of KDV. This motif is found in polymerase proteins from all members of the family Reoviridae (Mertens et al., 2000). Sequence analysis comparing the protein sequences with the database revealed the existence of a potyvirus-like (positive-sense single-stranded RNA virus) RNA-dependent RNA polymerase domain in the VP1 of KDV (aa 28–663). A search for sequence homologies with the gapped BLAST program revealed a partial match between VP1 of KDV and the RNA-dependent RNA polymerases of rotavirus C (accession number M74216) and avian rotavirus (AB009629). Similarities were also detected between VP1 of BAV and the polymerase of RDV (D10222). These observations strongly suggest that the first genome segment of both BAV and KDV encodes the viral RNA-dependent RNA polymerase (VP1(Pol)).

VP2 of BAV contains an RGD motif (aa 48–50) that is characteristic of integrin-binding proteins. The RGD motif of...
BAV VP2 aligns with an SGD motif (aa 51–53) in KDV VP2. Hydropathy analysis of VP2 using the Kyte and Doolittle program (Kyte & Doolittle, 1982), showed that the BAV and KDV proteins are both hydrophilic and that the RGD and SGD motifs are present in highly hydrophilic domains. The gapped BLAST program indicated that VP2 of both viruses also resembles several other nucleotide-binding proteins, including dynein, actin, a number of helicases, ATPases, methyltransferases and the S7 protein of NLRV.

Analysis of VP3 from BAV and KDV did not identify any specific functional motifs. However, VP3 of BAV (aa 274–335) partially matched the adenine-specific methyltransferase of Haemophilus influenzae HincII (51% similarity), while that of KDV partially matched the adenine-specific methyltransferase of E. coli Eco57I (47% similarity). KDV VP3 (aa 268–327) also exhibited 49% similarity to the helicase protein E1 of human papillomavirus type 45 (accession no. P36728) and 68% similarity (aa 6–30) to the probable helicase (UL5 protein) of human herpes simplex virus type 1 (accession no. X14112). Alignment of VP4 of BAV and KDV identified a highly conserved amino acid sequence [AADG(S/V)GVVVF/GFPS(V/I)ANG] containing the signature sequence motif [VVPV-FGPS(V/I)ANG] of the methyltransferase family UbiE (Barkovitch et al., 1997).

Analysis of BAV VP5 and the corresponding KDV protein (VP6) did not identify any functional motif or any significant similarity to protein sequences in the databases. A leucine zipper was identified by using the program Pfam in the homologous proteins VP6 of BAV (aa 36–57) and VP5 (aa 106–134) of KDV. Database searching showed that these proteins exhibit similarities to bacterial purine NTPases. VP6 of BAV (aa 124–579) has 49% similarity to a purine NTPase of Methanococcus jannaschii (U67572), while VP5 (aa 70–212) of KDV exhibits 47% similarity to a purine NTPase of Pyrococcus horikoshii (AP000004).

**Phylogenetic analysis based on the amino acid sequence of the viral Pol**

Sequence analysis of the Pol proteins from viruses within single genera in the family Reoviridae showed 21–97% aa identity between different rotaviruses, 97–99% for phytoreoviruses, 56–98% for orbiviruses and 92–99% for orthoreoviruses. However, between different genera, amino acid identity in the Pol protein was less than or equal to 18%. Within the genus Coltivirus, the amino acid identity between the BAV and KDV polymerases was 42%. However, this contrasts with amino acid identities of only 7–9 and 9–7% between CTFV and KDV and BAV, respectively. The results of the Pol sequence analysis, illustrated by a radial neighbour-joining tree, are displayed in Fig. 2.

The genomes of BAV and KDV are 20682 and 20985 nucleotides long. These values are in agreement with those calculated by Brown et al. (1993) from estimates of segment sizes after agarose gel electrophoresis and show a very significant difference from the genome of CTFV, estimated by the same authors to be 28530 bp (approximately 36% larger). Comparison with viruses belonging to other genera of the family Reoviridae showed that the BAV and KDV genomes are longer than those of the orbiviruses (around 19000 bp) and rotaviruses (around 18500 bp), but shorter than those of fijiviruses (around 28500 bp), phytoreoviruses (around 25500 bp) and orthoreoviruses (around 23500 bp). These total genome sizes alone demonstrate that there are major
H. Attoui and others

Fig. 2. Radial tree constructed with the help of the TreeView program and based on the VP1 (Pol) polymerase sequences of representative members of different genera of family Reoviridae. Abbreviations: RVA, RVB and RVC, Rotavirus groups A, B and C; BTV, bluetongue virus; AHSV, African horsesickness virus; PALV-Chu, Palyam virus isolate Chuzan; Reo, mammalian reovirus.

Discussion

The genus Coltivirus includes a number of arthropod-borne viruses that are known or suspected to be human pathogens. Since 1997, the partial sequences of American, European and Asian coltivirus isolates have been reported (Attoui et al., 1997, 1998). We present full-length genome sequences for two representative Asian viruses. These include BAV-In6423, isolated in 1980 in central Java (in Klaten, Semarang) by the late J. D. Converse (US Naval Medical Research Unit no. 2, Jakarta detachment) from Culex vishnui mosquitoes and identified by Brown et al. (1993), which was shown in previous investigations to be almost identical to BAV-Ch (Attoui et al., 1998), isolated from serum and cerebrospinal fluid of patients with meningoencephalitis (Chen & Tao, 1996). BAV-Ch is the representative isolate of the BAV species within Coltivirus subgroup B (Mertens et al., 2000). KDV-Ja7075 was also isolated in central Java, in Bantul (Kadipiro), Yogyakarta, from Culex fuscocephalus mosquitoes in 1981 by J. D. Converse and was identified by Brown et al. (1993). Recently, isolates with a broadly similar electropherotype (6-5-1) have been reported from the north-eastern part of China (B. Chen, personal communication). KDV is classified as a distinct species within subgroup B of the genus Coltivirus (Attoui et al., 1998; Mertens et al., 2000).

As previously observed with other members of the family Reoviridae (Mertens et al., 2000), all 12 segments of a single
virus isolate were found to have conserved terminal sequences. Such conserved motifs may act as sorting signals, bringing a single copy of each genome segment into the nascent viral capsid (Anzola et al., 1987; Xu et al., 1989). Moreover, in both group B viruses, the terminal dinucleotides in the 5′ and 3′ NCRs of all segments are inverted complements. Together with a number of segment-specific inverted terminal repeats, these terminal dinucleotides could interact by homologous base pairing, thus holding the RNA transcripts in a circular form (Attoui et al., 1997, 1998). This circular form has been described as a ‘panhandle’ structure that might possibly function as a guiding site for the RNA-dependent RNA polymerase. Comparison of the terminal regions of coltivirus genome segments showed that the terminal sequences of KDV and BAV are dissimilar to those of CTFV and that BAV and KDV do not share the same terminal trinucleotides with CTFV (Attoui et al., 1998; Mertens et al., 2000). These data also indicate that CTFV is more distantly related to KDV and BAV than the latter two viruses are to each other.

The amino acid sequence alignments showed that the proteins of BAV and KDV are homologous in the order given in Fig. 1. The presence of characteristic polymerase domains and the similarity to polymerases of other members of the family Reoviridae indicate that the RNA-dependent RNA polymerase of these viruses, as well as that of CTFV, is VP1. To agree with the unified protein nomenclature adopted in the Reoviridae section of the seventh ICTV report (Mertens et al., 2000), these proteins should be therefore identified as VP1(Pol).

RGD and SGD tripeptides, respectively, were identified in VP2 of BAV and KDV. VP2 of both viruses also showed similarities to several nucleotide-binding proteins. RGD motifs have been described in proteins of several other viruses, including enteroviruses (VP1), orbiviruses [VP7(T13)] and phage φ29 (terminal protein) (Basak et al., 1996; Illana et al., 1998; Roivainen et al., 1991). RGD-containing peptides inhibit enterovirus attachment to the cell backbone (Roivainen et al., 1991). These experiments suggest a receptor-binding site function for this motif and its interaction with cellular integrins. Crystallographic analysis of orbivirus protein VP7(T13) suggests a similar function for the RGD motif that is exposed on the core particle surface (Basak et al., 1996). In the case of phage φ29, site-directed mutagenesis has shown that this sequence is involved in interactions between the terminal protein and the viral polymerase (Illana et al., 1998). Further investigations with synthetic peptides and crystallographic analysis will be necessary to determine the conformation, structural location and functional role played by the RGD sequence in the VP2 of BAV.

VP12 of BAV and VP8 of KDV contain dsRNA-binding domains. The possible function of these domains in replication has been discussed elsewhere (Attoui et al., 1998). Other BAV and KDV proteins show similarities to nucleic acid modifying enzymes. The members of the family Reoviridae are known to posses proteins with enzymatic activities that can cap and methylate mRNA synthesized by the viral polymerase, including nucleotide phosphohydrolases (NTPases), guanylyltransferases and methyltransferases I and II (Juuti et al., 1998; Seliger et al., 1987; Stauber et al., 1997; Yue & Shatkin, 1998).

The VP3 proteins of BAV and KDV show some similarity to the adenosine-specific methyltransferases of E. coli and H. influenzae, respectively. Such similarity was reported previously for the rotavirus C methyltransferase, VP3, and the adenine-specific methyltransferase FOKI (Bremont et al., 1992). The capping enzyme VP4(Cap) of bluetongue virus contains guanylyltransferase as well as methyltransferase I and II activities (Ramadevi et al., 1998). Accordingly, VP3 or VP4 of BAV and KDV may carry out these reactions. However, database searching did not identify a putative guanylyltransferase directly, which has usually been identified biochemically by GTP-binding analysis (Mertens et al., 2000; Ramadevi et al., 1998).

VP6 of BAV and VP5 of KDV exhibit similarities to purine NTPases. These proteins may therefore represent NTPases involved in RNA translocation (Juuti et al., 1998), helicase activity (Stauber et al., 1997) or the processing and synthesis of viral mRNA cap structures (Yue & Shatkin, 1998).

Finally, a protein kinase domain was found in the VP7 of both BAV and KDV (Attoui et al., 1998).

On the basis of preliminary data, two subgroups, A and B, were created within the genus Coltivirus (Attoui et al., 1997, 1998; Mertens et al., 2000). Subgroup A includes CTFV and EYAV, while subgroup B contains BAV and KDV. Sequence analysis corroborated serological data regarding the antigenic relationship between CTFV and EYAV and confirmed differences from BAV and KDV. While the antigenic relationship between CTFV and EYAV was confirmed by both neutralization and complement fixation assays (Chastel et al., 1984; Karabatsos et al., 1987; Rehse-Küpper et al., 1976), the antigenic relationship between BAV and KDV was not assayed. However, antiseras against CTFV, EYAV and BAV were used to analyse the relationship between the three viruses (Chen & Tao, 1996). This analysis failed to demonstrate any significant cross-reaction between BAV and either CTFV or EYAV.

Characterization of the complete genome sequences of BAV and KDV and of a 2973 nucleotide (991 aa) region of the polymerase gene of CTFV has allowed us to reassess the pre-existing taxonomic status within the family Reoviridae. Amino acid sequence alignments and the common features that were detected in the viral proteins demonstrate significant homologies between the genes of BAV and KDV that were not detected in previous cross-hybridization studies (Brown et al., 1993). In the genome segment that encodes VP1(Pol), the genetic distance (amino acid sequence difference) is 83%. This value can be compared to the genetic distance between rotaviruses of groups A and C (51%) and confirms that BAV and KDV belong to a single genus. However, comparison with
the CTFV polymerase showed that the genetic distance between the Asian viruses and CTFV is approximately 90%. This does not support the current grouping of Asian viruses and CTFV in the same genus. This genetic distance is comparable to that existing between the polymerases of viruses from different genera of the family Reoviridae (> 82%).

We propose the creation of a new genus to include BAV and KDV. This new classification is based on genetic distances between viral isolates. However, it is also supported by a number of other molecular and biological observations. (i) The genomes of BAV and KDV have almost identical lengths and are very significantly shorter than that of CTFV. (ii) The G+C content of the BAV and KDV genomes is 37–39 mol%, which is lower than that calculated from the available sequences of CTFV (around 50%). (iii) The conserved terminal sequences of BAV and KDV show some similarity and are completely different from those of CTFV. (iv) BAV and KDV were isolated from mosquitoes (Culex species), while CTFV was isolated from ticks (Dermacentor species). (v) BAV and KDV were isolated in South-East Asia, while CTFV and related viruses were isolated in America and Europe.

We propose the name Sealdornavirus for the new genus (sigla from 'South-East Asian dodeca RNA viruses'), where the Latin prefix dodeca refers to the 12 genome segments. A formal proposal has been made to the ICTV.

The authors are grateful to Professor C. Chastel for providing EYAV-Fr578, to Drs R. B. Tesh and R. E. Shope for providing BAV and KDV, to Professors B. Q. Chen and S. J. Tao for their personal communications and to Dr Peter P. C. Mertens for critical review of the manuscript.

References


