Genome analysis of four Old World monkey adenoviruses supports the proposed species classification of primate adenoviruses and reveals signs of possible homologous recombination

Iva I. Podgorski, Laura Pantó, Tibor Papp, Balázs Harrach and Mária Benkő

Institute for Veterinary Medical Research, Centre for Agricultural Research, Hungarian Academy of Sciences, Budapest, Hungary

Within the family Adenoviridae, presently Simian mastadenovirus A is the single species approved officially for monkey adenoviruses (AdVs), whilst the establishment of six further species (Simian mastadenovirus B to Simian mastadenovirus G) has been proposed in the last few years. We examined the genetic content and phylogenetic relationships of four Old World monkey (OWM) AdV types [namely simian AdV (SAdV)-8, -11, -16 and -19] for which it had been proposed that they should be classified into different AdV species: SAdV-11 to Human mastadenovirus G, and the other three viruses into three novel species. By full genome sequencing, we identified gene contents characteristic for the genus Mastadenovirus. Among the 36 ORFs, 2 genes of different lengths, predicted to encode the adenoviral cellular attachment protein (the fibre), were found. The E3 regions contained six genes, present in every OWM AdV, but lacked the E3 19K gene, which has seemingly appeared only in the ape (hominid) AdV lineages during evolution. For the first time in SAdVs, two other exons belonging to the gene of the so-called U exon protein were also predicted. Phylogenetic calculations, based on the fibre-1 and the major capsid protein, the hexon, implied that recombination events might have happened between different AdV species. Phylogeny inference, based on the viral DNA-dependent DNA polymerase and the penton base protein, further supported the species classification proposed earlier.

INTRODUCTION

Adenoviruses (AdVs) are dsDNA viruses widespread among humans and vertebrate animals, mostly non-pathogenic for their hosts although, in rare cases, they can cause infections with significant consequences (Benkő, 2015). Primate AdVs are members of the genus Mastadenovirus, and while relatively we have a great deal of knowledge about human and ape (chimpanzee, gorilla and bonobo) AdVs, the AdVs of the more ancient primate lineages, such as Old World monkeys (OWMs), New World monkeys and prosimians, are hardly known. Although OWM AdVs were discovered more than 50 years ago (Hull et al., 1956), and were found in monkeys of many different species (macaques, grivet, black and white colobuses, red colobuses, hamadryas baboons and yellow baboons), more than half of the described types have not been studied in detail. The interest in more ancient simian AdVs (SAdVs) has risen with the awareness of the potential risk they may pose for humans if host switching occurred (Benkő et al., 2014). In addition, there is an increasing interest in gene delivery vectors derived from non-human AdVs (Lopez-Gordo et al., 2014), especially from SAdVs, since they are the closest relatives to human AdVs (HAdVs), but evolutionally still far enough apart for them not to be influenced by pre-existing immunity in the human population.

In the following text, informal abbreviations will be used for the species names, e.g. Human mastadenovirus A, HAdV-A, and Simian mastadenovirus A, SAdV-A. While all known HAdV types are grouped unambiguously into seven established species, HAdV-A to HAdV-G, there is only one species accepted officially for the classification of monkey AdVs, SAdV-A, containing OWM AdVs exclusively. The very first phylogenetic analysis of SAdVs (SAdV-1 to -25) was based on the very short sequences of the virus-associated RNA (VA-RNA) genes, studied in

The GenBank/EMBL/DDBJ accession numbers for the full genome sequences of the SAdVs studied are KP329561 (SAdV-8), KP329562 (SAdV-11), KP329564 (SAdV-16) and KP329565 (SAdV19).
both OWM (SAdV-1 to -20) and chimpanzee (SAdV-21 to -25) AdVs (Kidd et al., 1995). More than 10 years ago, the first full sequence of an OWM AdV (SAdV-3) was published (Kovács et al., 2004). The phylogenetic distance of SAdV-3 from all HAdVs known at that time warranted a novel species (established later as SAdV-A), which would contain OWM AdVs only. The next fully sequenced OWM AdV was SAdV-1, being the first SAdV recognized to have two fibre genes (Kovács et al., 2005). Phylogenetically, SAdV-1 has been found to be closer to HAdVs with two fibre genes (in species HAdV-F) than to SAdV-3 (Kovács et al., 2005). A few years ago, a new species, SAdV-B, was proposed for OWM AdVs (Roy et al., 2012). Additional OWM AdVs have been described in olive baboons [baboon AdV (BaAdV)-1, BaAdV-2/4 and BaAdV-3)], along with a proposal for BaAdV-2/4 and -3 to form a new species, SAdV-C (Chiu et al., 2013). Most recently, another SAdV type (strain 23336) from rhesus macaque has been proposed to form a new species, SAdV-D (Malouli et al., 2014). Subsequently, we proposed a species designation for all the 20 known serotypes of OWM AdVs. According to our proposal, serotype SAdV-13 would be the sole member of species SAdV-D (Pantó et al., 2015), whereas, in the future, an additional novel monkey AdV species might be needed for the non-serotyped SAdV strain 23336 (Malouli et al., 2014). The establishment of three additional species (SAdV-E to SAdV-G) was proposed in the same paper.

The main purpose of the present study was to provide further support, by full genomic sequence analyses, for the proposed new SAdV species. To this end, four OWM AdVs (SAdV-8, -11, -16 and -19) were sequenced, and their genetic content and phylogenetic relationships scrutinized. Our results indeed warranted the formation of species SAdV-B, SAdV-C and SAdV-E. Furthermore, for what is believed to be the first time from SAdVs, all the three exons of the putative gene of the ‘U exon protein’ (UXP) were discerned. This might help the complete identification of this spliced gene in other AdVs in the future.

**RESULTS**

**General features**

The DNA of four SAdV strains was fully sequenced and analysed. The main characteristics of the new genome sequences are summarized in Table 1, in comparison with the range of the examined values in known members of each SAdV species where the newly sequenced viruses are proposed to belong. SAdV-16 is an exception as being the first representative of a newly proposed species, SAdV-E. Genomic assemblies revealed that all four genomes contained 36 putative coding regions characteristic for mastadenoviruses. These included two genes, of different lengths, predicted to encode the cellular attachment protein (the fibre) of AdVs. No homologue of the E3 19K gene, usually present in members of the ape (hominid) AdV lineages, was found in any of the studied genomes, and each of them contained only a single copy of the VA-RNA gene.

The majority of the deduced protein sequences of SAdV-8, -11 and -19 were very similar to (sharing >95% identity with) their counterparts in other members of the AdV species (namely SAdV-B, HAdV-G and SAdV-C, respectively) where they are proposed to be classified. In the SAdV-8 genome, there is one exception to this, the UXP, with 88% identity. For SAdV-11, lower identity values were seen in the E1A gene (88%), E3 region genes (in some cases as low as 40%) and UXP, (85%). The deduced protein sequences of SAdV-16 did not exhibit particular similarity to any known AdV, except the hexon, penton base and pVIII. In the SAdV-19 genome, there are several exceptions, including the hexon (89% identity), the proteins encoded by genes in the E3 region (in some cases the identity was as low as 64%), UXP (66% identity), fibre-1 (only 27% identity to SAdV-C members, while 46% to HAdV-G members), fibre-2 (66% identity) and ORF6/7 (85% identity).

**Phylogeny inference**

Phylogeny reconstructions, performed with the amino acid sequences deduced from the DNA-dependent DNA
Fig. 1. Phylogenetic tree based on the full amino acid sequence of (a) the DNA-dependent DNA polymerase, (b) the penton base, (c) fibre-1 and (d) fibre-2. The black arrow indicates the node that separates the group of AdVs with two fibre genes (except SAdV-18). VA-RNA and E3 19K labels mark the nodes after which either one or two copies of these genes appeared during the evolution.
polymerase (pol), penton base, fibre-1 and fibre-2 genes are presented in Fig. 1. SADV-8 and -11 always clustered clearly with the SAdVs that had been proposed to form species SADV-B (Roy et al., 2012) and HADV-G (Jones et al., 2007), respectively. SADV-16 formed an independent branch proposed to be accepted as a new species, SADV-E (Panto et al., 2015). SADV-19 appeared among the members of the previously proposed species SADV-C (Chiu et al., 2013), except on the tree based on fibre-1 (Fig. 1c). Interestingly, the phylogeny inference based on the hexon amino acid sequences (Fig. 2) implied divergent relationships among several AdVs. These contradictions could be explained by the results of recombination analysis of the SADV-19 genome in comparison to that of members of species HADV-G, HADV-F and SADV-C (Fig. 3a, b). The SimPlot and BootScan analyses indicated that recombination event(s) might have happened in the hexon gene. Nonetheless, our conclusion is that the results of phylogeny inference based on these proteins need to be handled with care.

**UXP**

The UXP sequences of the four SAdVs were compared to their counterparts in members of species HADV-C in order to determine the position and splicing sites of all the three UXP exons (Tollefson et al., 2007). In this study, the UXP sequences of three ape AdVs, sequenced by others earlier, namely chimpanzee (SADV-34), gorilla (SADV-43) and bonobo (SADV-44) AdV (Roy et al., 2009), were identified in this study as well by comparison of the sequences to members of the species HADV-C. The main characteristics of the UXPs are shown in Table 2. UXP sequences were aligned to compare the degree of conservation of the three exons in different AdVs (Fig. 4). The putative positions of the three UXP exons in the genomes, derived from the presence of splice donor and acceptor sites, are summarized in Fig. 5.

**DISCUSSION**

Here, we report the genomic characterization of four OWM AdVs and discuss their taxonomical classification. One virus (SADV-11) was found to belong to the previously established species HADV-G, whereas each of the other three viruses seemed to represent a different species. Thus, the proposals for the establishment of SADV-B, SADV-C and SADV-E were supported. In most cases, the taxonomic classification could be decided unequivocally. SADV-16 was an exception inasmuch as several additional aspects had to be considered for its species allocation.

The size of the SADV genomes sequenced to date range between 31 045 (SADV-7; Roy et al., 2011) and 36 838 bp (SADV-20; Roy et al., 2012). The four newly sequenced SADV genomes fall within this range, and have 36 putative genes characteristic for members of the genus *Mastadenovirus*, including the presence of two fibre genes, a feature recognized in members of species HADV-G and HADV-F and in several monkey AdVs before (Fig. 2) (Alonso-Padilla et al., 2016; Panto et al., 2015). Besides mastadenoviruses, the presence of two fibre genes has also been found in many representatives of different avian adenovirus species (Kaján et al., 2010, 2012; Marek et al., 2014a, b; Zhao et al., 2015), but only in two members of the genus *Atadenovirus* (Pénèzes et al., 2014; To et al., 2014). The G+C content, an important AdV species demarcation criterion (Harrach, 2014) among SAdVs, varies between 47.8 (SADV-20) and 65.7 % (SADV strain A1139; Roy et al., 2012). However, within each of the species the differences in the base composition do not exceed 3 % (Panto et al., 2015). The G+C content of the four SAdVs studied here conformed to this rule, and usually corresponded well to the narrow range of G+C content of the species they were proposed to belong to (Table 1).

The existence of a protein encoded by the U exon was predicted more than 20 years ago (Davison et al., 1993), but the entire gene with its three exons has been described in members of the species HADV-C only (Tollefson et al., 2007). Deletion or truncation of the U exon results in impaired virus replication and causes aberrant localization of the DNA-binding protein (DBP) in the nucleus of infected cells (Tollefson et al., 2007). Amino acid sequence alignments of the predicted UXP s of SAdVs and HAdVs (Fig. 4) from species HADV-C showed that the first and second exons are relatively well conserved, even in AdVs of different species, whilst the third exon is extremely variable in both sequence content and length (Table 2). Nonetheless, the genomic localization of all the three exons in the studied SAdVs was comparable (Fig. 5). The third exon of the UXP gene overlaps with the coding region of the DBP gene. Interestingly, in the individual AdV species, the position of the third exon relative to the DBP gene seems to be well conserved. More precisely, the splice acceptor site of the third exon was found to be 6, 9 or 15 nt upstream from the start codon of the DBP (Fig. 5). Thus, the third exon was always in the next frame when compared to that of the DBP. The previously described and newly revealed putative UXP sequences might be of help in defining the complete gene sequences in other AdVs in the future. However, we can expect that defining all the three exons of the UXP in non-primate AdVs might be more challenging due to the increase of divergence in the sequences of evolutionally more distant AdV species.

All OWM AdVs studied earlier have been found to have one VA-RNA gene only. However, human and chimpanzee AdVs (in species HADV-B to HADV-E) have been shown to possess two VA-RNA genes (Kidd et al., 1995; Larsson et al., 1986). In certain primate AdVs, the VA-RNA genes have not been studied yet, as the PCR used for their amplification has failed either due to the high specificity of the primers, or because the genes were missing from the genomes of some viruses (Kidd et al., 1995). Nonetheless, the VA-RNA gene of SADV-16 (strain SA7) was characterized almost 30 years ago, proving that this OWM AdV has only one such gene (Larsson et al., 1986). Our SADV-16 sequence confirmed the presence of this VA-RNA gene, albeit in one of the
earlier studies its PCR amplification had failed, probably for reasons described above (Kidd et al., 1995). The sequence of the VA-RNA genes of SAdV-11 and SAdV-19 has been published earlier (Kidd et al., 1995), and here we report their exact position in the genomes. However, our analysis revealed a longer (164 versus 104 nt) VA-RNA gene in SAdV-11. This size difference was identified as a 60 nt ‘deletion’ between the 68 nt long 5' and 36 nt long 3' ends, which were completely identical in our sequence and in the VA-RNA gene reported earlier (Kidd et al., 1995). The reason for the gene fragment missing from the previously reported sequence might be the formation of secondary structures interfering with the PCR amplification of the given locus. In general, VA-RNAs of OWM AdVs seem to be shorter (93 to 104 nt) than those of HAdVs. Among the few exceptions are SAdV-11 (164 nt) and SAdV-19 (146 nt), as well as SAdV-16 and -19 (168 nt in both). All these viruses are proposed to be members of different species: HAdV-G, SAdV-D, SAdV-E and SAdV-C, respectively (Pantó et al., 2015). Furthermore, we also proved for what is believed to be the first time the presence of a VA-RNA in SAdV-8. This gene is longer (159 nt) than the VA-RNAs of most OWM AdVs. SAdV-8 is a member of the proposed species SAdV-B, from which we do not have information about the VA-RNA gene.

![Phylogenetic tree based on full hexon amino acid sequences.](image-url)

**Fig. 2.** Phylogenetic tree based on full hexon amino acid sequences. The black arrow indicates the node that separates the group of AdVs with two fibre genes (except SAdV-18). VA-RNA and E3 19K labels mark the nodes after which either one or two copies of these genes appeared during the evolution.
Fig. 3. (a) SimPlot and (b) BootScan analyses of Simian mastadenovirus C, Human mastadenovirus G and Human mastadenovirus F members relative to SAdV-19. The annotated genome of SAdV-19 is shown between the two graphs to allow easier observation of the genomic locus where the putative recombination might have taken place. The black arrows indicate the assumed possible recombination spots in the genome, i.e. in the hexon and fibre gene regions (genes highlighted in red in the annotated genome).
of any other member. All OWM AdVs studied to date contain only one VA-RNA gene, confirming the results published in earlier studies (Kidd et al., 1995; Larsson et al., 1986).

Phylogenetic trees, based on the full pol (Fig. 1a), penton base (Fig. 1b) and fibre-2 (Fig. 1d) amino acid sequences, also confirmed that SAdV-11 belongs to the previously established species HAdV-G, whereas the classification of SAdV-8, -16 and -19 required the establishment of the previously proposed species SAdV-B, SAdV-C and SAdV-E, respectively (Chiu et al., 2013; Pantó et al., 2015; Roy et al., 2012). Interestingly, however, the tree based on the fibre-1 amino acid sequence completely separates SAdV-19 not only from its proposed species SAdV-C, but also from all the other species we know (Fig. 1c). Comparison of this fibre with the available sequences revealed that it shares only 46% or less amino acid identity with the fibre sequence of members of species HAdV-G. With the recombination analyses, we were unable to find any AdV in the known species that could be supposed as the origin of this fibre gene (Fig. 3). Nonetheless, the evolutionary tree indicates that it might be the most ancient of all the known primate AdVs with two fibre genes (Fig. 1c).

The hexon-based tree shows different relationships among several SAdVs and HAdVs (Fig. 2). This might be the result of some recombination events. SimPlot analysis of SAdV-19 clearly indicated the probability of a homologous recombination event in the hexon gene, and the BootScan analysis suggested that the SAdV-19 hexon gene resulted from a recombination event between yet unknown members of species HAdV-F and HAdV-G (Fig. 3a). The hexon-based phylogenetic tree (Fig. 2) could not separate the two species from species SAdV-C. This also supports the probability of several recombination events among these AdVs. Homologous recombination events, most often in the hexon and fibre genes, have been described in many primate AdVs (Chiu et al., 2013; Crawford-Miksza & Schnurr, 1996; Dehghan et al., 2013a, b; Walsh et al., 2011). Consequently, a divergent topology of the other AdV lineages could also be observed on the hexon tree. For example, HAdV-4 and -16 are separated from other members of their species HAdV-E and HAdV-B, respectively. This is not surprising since it has been shown that the two viruses share very high overall nucleotide sequence identity in the hexon gene (Pring-Akerblom et al., 1995). Furthermore, the hexon of chimpanzee AdV-63 (ChAd-63; proposed but officially not classified into species HAdV-C) is very similar to that of SAdV-36 (species HAdV-E), indicating the possibility of an interspecies homologous recombination. On most trees, SAdV-16 appeared far enough from the species SAdV-B to be considered as representative of a new species, SAdV-E (Figs 1 and 2). However, the penton base and hexon trees show that SAdV-16 is close to, or falls within, species SAdV-B, respectively. A homologous recombination event in the hexon gene of SAdV-16 is, therefore, very likely. Some other properties of SAdV-16, such as the host origin, the G+C content and the results of the haemagglutination-inhibition tests (Rapoza, 1967), also support that it should be considered as a new species, distinct from SAdV-B.

A certain ambiguity arises with the classification of SAdV-19 as well. The inverted terminal repeats (ITRs) of SAdV-19 (127 bp) are longer than those of the other BaAdVs from the species SAdV-C (87 bp). SAdV-19 has an overall longer genome, and has been isolated from a different baboon species. Nonetheless, other features such as the G+C content (Table 1), as well as the phylogeny inference based on the pol and penton base, give support to the placing of SAdV-19 into the species SAdV-C.

The results of the present study further support the need for establishing three new SAdV species: SAdV-B, SAdV-C and SAdV-E. The organization of the genomes of all newly sequenced SAdVs was comparable, and very similar to that of the previously sequenced SAdVs. The gene of the UXP homologue was identified in all the four SAdV genomes based on comparison with the UXP gene of members of species HAdV-C. However, mRNA studies would be essential for the ultimate confirmation of these predictions. By further screening of primates, especially the more ancient New World monkeys and prosimians, numerous additional AdV types and lineages would likely be discovered in the future.

**METHODS**

**Cells and virus stocks.** Samples of the prototype SAdV strains originating from the American Type Culture Collection (SAdV-8, ATCC VR-1539, strain P-5, from crab-eating macaque; SAdV-11, ATCC VR-206, strain P-10, from rhesus macaque; SAdV-16, ATCC VR-941, strain C-8, from grivet; SAdV-19, ATCC VR-275, strain AA153, from yellow baboon) were used either directly in PCRs (SAdV-16 and -19), or for the inoculation of Vero E6 cells (SAdV-8 and -11) to propagate the viruses for next generation sequencing (NGS). After a few passages, seven 175 cm² tissue culture flasks were used for virus production. The tissue culture supernatants and the cells (disrupted by three freezing and thawing cycles) were clarified with low-speed centrifugation. Then, the virions were sedimented in a Beckman ultracentrifuge and the viral DNA was isolated with the phenol/chloroform extraction method.

**PCR methodology.** The DNA of SAdV-16 and -19 was sequenced by PCR combined with traditional Sanger sequencing, with consensus and specific primers designed as described previously (Kovács & Benko, 2009). To amplify the first fragments of the viral genome, a nested PCR with degenerate primers targeting the gene of the pol (Wellehan et al., 2004) and that of the IVA2 protein (Pantó et al., 2015; Vidovszky et al., 2015) were used. Subsequently, standard PCR with degenerate primers targeting the hexon gene was used (Kiss et al., 1996). Specific primers, based on partial sequences, were designed with the use of the Primer Designer program version 2.0. Dream Taq DNA polymerase (Fermentas) was found to be optimal for the PCR amplification of the shorter fragments. The PCRs were performed as described previously (Doszpoly et al., 2013). For the amplification of longer (>1000 bp) fragments, the Takara PrimeSTAR Max DNA polymerase was used according to the manufacturer’s recommendations. The PCR products were purified from agarose gels with the use of the MEGApure-spin total fragment DNA purification kit (iNtRON Biotechnology).
Sequencing and genome assembly. We used two different sequencing approaches. SaDV-8 and -11 were sequenced by the NGS method. Their purified genomic DNA was sent to a commercial service (BGI in China), where paired-end sequence reads were generated using the Illumina HiSeq 2500 system. The quality of the FASTQ sequences was enhanced by trimming off low-quality bases using the ‘Trim sequences’ option of the CLC Genomics Workbench version 7.0.4. The quality-filtered sequence reads were arranged into a number of contig sequences. The analysis was performed using the ‘de novo assembly’ option of the CLC Genomics Workbench version 7.0.4. The remaining

Table 2. The size and position of the three putative exons of the UXP identified previously in four HAdVs and in seven SaDV types in this study (highlighted in bold)

<table>
<thead>
<tr>
<th>Virus</th>
<th>Exon position (no. of amino acids encoded)</th>
<th>1st exon</th>
<th>2nd exon</th>
<th>3rd exon</th>
</tr>
</thead>
<tbody>
<tr>
<td>SaDV-8</td>
<td>29531–29691 (54)</td>
<td>23643–23719 (25)</td>
<td>22766–23040 (91)</td>
<td></td>
</tr>
<tr>
<td>SaDV-11</td>
<td>28624–28781 (53)</td>
<td>22905–22981 (25)</td>
<td>21970–22388 (139)</td>
<td></td>
</tr>
<tr>
<td>SaDV-16</td>
<td>29161–29321 (54)</td>
<td>23442–23518 (25)</td>
<td>22590–22870 (93)</td>
<td></td>
</tr>
<tr>
<td>SaDV-19</td>
<td>28507–28673 (56)</td>
<td>22855–22931 (25)</td>
<td>21935–22349 (131)</td>
<td></td>
</tr>
<tr>
<td>HAdV-1</td>
<td>30927–31090 (55)</td>
<td>24725–24801 (25)</td>
<td>23704–24098 (131)</td>
<td></td>
</tr>
<tr>
<td>HAdV-2</td>
<td>30856–31019 (55)</td>
<td>24715–24791 (25)</td>
<td>23676–24088 (137)</td>
<td></td>
</tr>
<tr>
<td>HAdV-5</td>
<td>30868–31031 (55)</td>
<td>24668–24744 (25)</td>
<td>23629–24041 (137)</td>
<td></td>
</tr>
<tr>
<td>HAdV-6</td>
<td>30835–30998 (55)</td>
<td>24700–24776 (25)</td>
<td>23661–24073 (137)</td>
<td></td>
</tr>
<tr>
<td>SaDV-34</td>
<td>32643–32803 (54)</td>
<td>25085–25161 (25)</td>
<td>23701–24380 (226)</td>
<td></td>
</tr>
<tr>
<td>SaDV-43</td>
<td>32048–32211 (55)</td>
<td>24955–25031 (25)</td>
<td>23811–24274 (154)</td>
<td></td>
</tr>
<tr>
<td>SaDV-44</td>
<td>32686–32846 (54)</td>
<td>25127–25203 (25)</td>
<td>23761–24245 (221)</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 4. Alignment of the amino acid sequences of the UXPs identified in four HAdVs previously and in seven SaDV types in this study. Red color indicates high consensus value (90 % and higher), and blue color indicates lower consensus value (50–90 %).

http://jgv.microbiologyresearch.org
and purification of the virions. The genome fragments were obtained from the NGS data. The genome of SAdV-16 and -19 was sequenced directly without prior large-scale propagation and purification of the virions. The genome fragments were obtained by PCR and sequenced with the PCR primers on both strands. For the larger fragments, a genome walking strategy was applied. The sequences of the genome ends of SAdV-16 and -19 were also determined by PCR. Based on our former experience that every ITR in members of the species HAdV-F, HAdV-G, SAdV-A, SAdV-B and SAdV-C started with the same octamer (CATCATCA), a primer was designed with a long 5′ extension (5′-CAGCTCGGATTTCATCATCA-3′). This primer was used paired with a specific outward-oriented primer in each of the four cases, resulting in an ~500 bp fragment. Thus, the ITR sequences could be obtained except for the very conserved 8 nt motif. The conditions of the sequencing reactions and nucleotide sequence assembly have been described in detail previously (Pénzes, et al., 2014; Tarján, et al., 2014).

The genome sequences were annotated with the web-accessible annotation tool Artemis (Berriman & Rutherford, 2003; Marek, et al., 2013). Genomic sequence fragments were queried systematically against the non-redundant database of the National Center for Biotechnology Information, using the BLASTX program online. The sequences of the genes, which are known to contain introns in other AdVs, were checked for the presence of putative splice donor and acceptor sites. Splice sites in the genomes were determined by manual search by comparison with the corresponding regions of the earlier described SAdVs and HAdVs.

The UXP gene sequence and location in the genome were determined by comparison with the HAdV-5 UXP sequence (Tollefson et al., 2007). The VA-RNA gene sequence of SAdV-8 was determined by comparison with the available VA-RNA sequences of primate AdVs, with special focus on the criteria for identification of VA-RNA genes and fully conserved nucleotide positions determined in the study by Kidd et al. (1995). Similarity plots and bootscanning analyses were performed with Simplot 3.5.1 with window size 1000 bp, step size 50 bp (Lole, et al., 1999).

Phylogenetic calculations. Phylogenetic calculations were carried out essentially by a scheme described earlier (Pantó, et al., 2015). Analyses were based on full amino acid sequences deduced from the pol, penton base, hexon, fibre-1 and fibre-2 proteins of all primate AdVs sequenced to date. The tree shrew AdV (TSAdV) was also included. In the calculations based on fibre-1 and -2 sequences, the sequence of fowl AdV type 1 (FAdV-1) was also used. For the different proteins, the following models were applied: JTT+I+G for Pol, WAG+I+G for penton base, LG+I+G for hexon, LG+G for fibre-1 and CpREV+G for fibre-2.


Fig. 5. Genomic position and splicing pattern of the UXP gene of SAdVs identified in this study, in comparison to those in HAdV-5. The reading frame of the DNA-binding protein (DBP) gene, overlapping the third exon, is also shown.
AKNOWLEDGEMENTS

The authors gratefully acknowledge the financial support provided by the Hungarian Scientific Research Fund OTKA K100163, as well as the EU grant ADVance FP7-290002, and the Balassi Institute (Temps Public Foundation).

REFERENCES


adenovirus types reveal the existence of two unknown lineages that merit the establishment of novel species within the genus *aviadenovirus*. *J Gen Virol* 95, 156–170.


