Complete genome sequence of simian adenovirus 1: an Old World monkey adenovirus with two fiber genes

Gábor M. Kovács,† Balázs Harrach, Alexander N. Zakhartchouk and Andrew J. Davison

1Veterinary Medical Research Institute, Hungarian Academy of Sciences, PO Box 18, H-1581 Budapest, Hungary
2Vaccine and Infectious Disease Organization, University of Saskatchewan, Saskatoon, Saskatchewan, Canada S7N 5E3
3MRC Virology Unit, Institute of Virology, Church Street, Glasgow G11 5JR, UK

Simian adenovirus 1 (SAdV-1) is one of many adenovirus strains that were isolated from Old World monkey cells during poliomyelitis vaccine production several decades ago. Despite the availability of these viruses, knowledge of their genetic content and phylogeny is rudimentary. In the present study, the genome sequence of SAdV-1 (34 450 bp) was determined and analysed. In regions where genetic content varies between primate adenoviruses, SAdV-1 has a single virus-associated RNA gene, six genes in each of the E3 and E4 regions and two fiber genes. SAdV-1 clusters phylogenetically with HAdV-40, a member of human adenovirus species HAdV-F, which also has two fiber genes. However, based on phylogenetic distances and other taxonomic criteria, SAdV-1 is proposed to represent a novel adenovirus species.

The icosahedral, non-enveloped, double-stranded DNA viruses in the family Adenoviridae infect hosts from all of the main vertebrate lineages (Benkő & Harrach, 2003; Kovács et al., 2003). Most mammalian adenoviruses, including all six human adenovirus (HAdV) species (HAdV-A to HAdV-F), are members of the genus Mastadenovirus (Wadell, 2002). A sizeable number of simian adenoviruses (SAdVs) that infect Old World monkeys or chimpanzees have also been described in the literature. Many of these were isolated from cultures of simian cells during production and testing of poliomyelitis vaccine (Hull & Minner, 1957; Hull et al., 1956, 1958). These viruses were first grouped by Hull et al. (1956) on the basis of their cytopathic effects and some were later detected in monkeys (Hoffert et al., 1958). Rapoza (1967) subsequently grouped SAdVs into serotypes by using the haemagglutination-inhibition typing technique (Rosen, 1960). Early interest in SAdVs was due to three principal reasons: assessing the presence of neutralizing antibodies against SAdV-5 in human sera (Aulisio et al., 1964), investigating the oncogenic effect of certain SAdVs (including SAdV-5) in the hamster (Hull et al., 1965) and exploring the presence of small viral particles in SAdV samples (Hull et al., 1965), which later led to the discovery of adeno-associated viruses (Atchison et al., 1965; Rapoza & Atchinson, 1967). The original continuous numbering system applied to the isolated cytopathic agents, which did not refer to any virus group, was changed by Kalter et al. (1980) to the definition of 24 serotypes.

Although numerous SAdV strains have been deposited in culture collections, most studies of adenovirus genomics have focused on HAdVs (especially HAdV-5). Corresponding work on SAdVs is very limited and commenced with restriction-endonuclease analyses (Dimitrov et al., 1979; Naroditsky et al., 1978). Kidd et al. (1995) described the most significant, and probably the only, comprehensive report dealing with the classification of SAdVs based on DNA sequences, namely those of the virus-associated (VA) RNA genes. Other studies not focusing specifically on SAdVs also employed SAdV sequences, such as those of the inverted terminal repetition (Bailey & Mautner, 1994) or the E1A region (Avvakumov et al., 2001, 2004). For reasons associated with the development of alternative vector systems, interest in non-human primate adenoviruses has increased in recent years and has led to sequencing of the genomes of five chimpanzee adenoviruses (SAdV-21 to SAdV-25; Farina et al., 2001; Roy et al., 2004), which belong to recognized HAdV species (SAdV-21 to HAdV-B and the others to HAdV-E; Benkő et al., 2000). The Old World monkey adenoviruses are phylogenetically more distant from the HAdV species, although they appear to have closer
relationships to HAdV-A and HAdV-F (Bailey & Mautner, 1994; Kidd et al., 1995). Kovács et al. (2004) reported the first complete genome sequence of an adenovirus isolated from an Old World monkey (SAdV-3). Phylogenetic analyses showed that this virus represents a lineage that branched at an early stage of primate adenovirus evolution and is related marginally more closely to HAdV-A and HAdV-F (which are also early lineages in primate adenovirus evolution) than to other HAdV species.

The purpose of the present work was to determine the genetic content and phylogenetic relationships of a monkey adenovirus with properties different from those of SAdV-3, as we presumed that such a virus would represent a distinct lineage. SAdV-1 was chosen because it was isolated from the tissues of a cynomolgus monkey (Macaca fascicularis), whereas SAdV-3 was isolated from rhesus monkey (Macaca mulatta) cells (Hull et al., 1956). Both viruses were classified as belonging to group 1 based on cytopathic effects (Hull et al., 1956), but placed in different haemagglutination groups, with SAdV-1 in group 3 and SAdV-3 in group 2 (Rapoza, 1967).

SAdV-1 was purchased from ATCC (VR-195) and propagated on rhesus macaque kidney cells (LLC-MK2). By using methods described by Kovács et al. (2004), viral DNA was isolated and sequenced via a random M13 library, and the genetic content and phylogeny of the completed sequence were analysed. Each nucleotide in the genome was determined an average of nine times and the entire sequence was obtained on both strands. The purity of the SAdV-1 DNA preparation was estimated at 50% from the proportion of clones whose sequence matched that of the SAdV-1 genome.

Fig. 1 shows the genetic organization of SAdV-1. The genome (34,450 bp) is very similar in size to that of SAdV-3 (34,425 bp; Kovács et al., 2004) and marginally larger than those of HAdV-40 (34,214 bp; Davison et al., 1993) and HAdV-12 (34,125 bp; Sprengel et al., 1994). These four viruses have the smallest genomes characterized thus far among the primate adenoviruses. The G+C content of the SAdV-1 genome is 55·21 mol%, which is almost the same as that of SAdV-3 (55·25 mol%) and different from those of HAdV-40 (51·22 mol%) and HAdV-12 (46·52 mol%). As in other adenoviruses, G+C content varies along the genome and shows the same pattern as SAdV-3, with the highest value in the pVII gene and the lowest in the E3 region (data not shown).

The genetic layout of SAdV-1 exhibits the general characteristics of primate adenoviruses (Davison et al., 2003). One of the most striking features is the presence of two fiber genes (fiber-1 and fiber-2). The presence of a splice-acceptor site immediately preceding each fiber-coding region and the sequences of the encoded proteins suggest that both of these genes are expressed. To date, only two primate adenoviruses (HAdV-40 and HAdV-41, both belonging to species HAdV-F) have been shown to encode two fiber genes (Kidd et al., 1993; Pieniazek et al., 1990). Both HAdV-40 fibers are expressed during infection, with a single copy of one or the other type associated with each penton base at the capsid vertices (Kidd et al., 1993). This contrasts with the bird viruses in the genus Aviadenovirus, which have one (FAdV-9; Cao et al., 1998) or two (FAdV-1; Chiocca et al., 1996) fiber genes, but with two copies incorporated at each vertex (Gelderblom & Maichle-Lauppe, 1982). The SAdV-1 fiber-2 protein comprises 560 aa, which is similar in size to that of HAdV-41 with 562 aa. HAdV-40 isolates generally have a shorter fiber-2 protein, with one less copy of the repeated shaft motif compared with common HAdV-41 isolates (Tiemessen & Kidd, 1995). The length of the SAdV-1 protein is similar to that of HAdV-41 in this region (Fig. 2). The SAdV-1 fiber-1 protein consists of 363 aa, which is shorter than the 387 aa of HAdV-40 and HAdV-41. Sequences corresponding to two shaft motifs are absent from the SAdV-1 fiber-1 protein, which, as a consequence, contains 10 repeats rather than the 12 described for HAdV-40 and HAdV-41 (Fig. 2). Insertions and deletions also characterize the SAdV-1 fiber-1 sequences adjacent to and within the knob region.

SAdV-1 has a single VA-RNA gene, as do other monkey adenoviruses and members of species HAdV-A, HAdV-F
and a subgroup of HAdV-B (Kidd et al., 1995). The other HAdVs and chimpanzee adenoviruses have two adjacent VA-RNA genes. SAdV-1 was studied by Kidd et al. (1995), but its VA-RNA gene was not amplified successfully, even though the sequences of one primer pair (VA5 and VA6) match the SAdV-1 sequence. The SAdV-1 VA-RNA is predicted to be 164 nt. This is similar to the VA-RNAs of three other monkey adenoviruses (SAdV-16, -19 and -13, with 168, 168 and 146 nt, respectively; Kidd et al., 1995) and of HAdV-40 and HAdV-12 (159 and 145 nt, respectively). In contrast, the SAdV-3 VA-RNA is only 109 nt in size and similarly short VA-RNAs characterize most monkey adenoviruses (Kidd et al., 1995).

In addition to the fiber and VA-RNA genes, adenovirus genomes exhibit the greatest organizational differences in the E3 and E4 regions (Davison et al., 2003; Ursu et al., 2004). The SAdV-1 E3 and E4 regions each encode six genes corresponding to those in SAdV-3 (Kovács et al., 2004), the same arrangement as in HAdV-A. HAdV-40 has a similar gene layout in the E3 and E4 regions, but lacks one gene in each region (E3 12.5K and E4 ORF1), presumably through loss after divergence from SAdV-1. E4 ORF1 has transforming properties and may have an anti-apoptotic role (Frese et al., 2003; Leppard, 1997), while the function of E3 12.5K is not known (Russell, 2000).

During compilation of the genome sequence from random M13 clones, a deletion variant was identified. From the proportion of clones obtained, the mutant comprised approximately one-third of the genomes present in the DNA stock. The deletion spanned 659 bp (from nt 1138 to 1797) of the genome sequence, a region that contains the 3' end of E1A, the majority of E1B 19K and the overlapping 5' end of E1B 55K (Fig. 1). The presence of the mutant was confirmed by PCR analysis of the same stock of DNA used for sequencing, utilizing primers mapping in the regions flanking the deletion (data not shown). It is not known whether the mutant was present in the virus as originally isolated, or whether it was generated during passage in vitro. Mutants lacking sequences in the E1 region are often constructed in vitro specifically for development as vectors. It is possible that the SAdV-1 deletion mutant could be separated from parental virus by using an appropriate complementing cell line and utilized similarly.

Phylogenetic analyses were carried out for each SAdV-1 gene. Fig. 3 shows examples for the four genes employed for SAdV-3, HAdV-40 and HAdV-41. Angle brackets indicate repeated motifs in the shaft region as defined by Kidd et al. (1993). The commencement of the knob region is shown for the short fiber.

and a subgroup of HAdV-B (Kidd et al., 1995). The other HAdVs and chimpanzee adenoviruses have two adjacent VA-RNA genes. SAdV-1 was studied by Kidd et al. (1995), but its VA-RNA gene was not amplified successfully, even though the sequences of one primer pair (VA5 and VA6) match the SAdV-1 sequence. The SAdV-1 VA-RNA is predicted to be 164 nt. This is similar to the VA-RNAs of three other monkey adenoviruses (SAdV-16, -19 and -13, with 168, 168 and 146 nt, respectively; Kidd et al., 1995) and of HAdV-40 and HAdV-12 (159 and 145 nt, respectively). In contrast, the SAdV-3 VA-RNA is only 109 nt in size and similarly short VA-RNAs characterize most monkey adenoviruses (Kidd et al., 1995).

In addition to the fiber and VA-RNA genes, adenovirus genomes exhibit the greatest organizational differences in the E3 and E4 regions (Davison et al., 2003; Ursu et al., 2004). The SAdV-1 E3 and E4 regions each encode six genes corresponding to those in SAdV-3 (Kovács et al., 2004), the same arrangement as in HAdV-A. HAdV-40 has a similar gene layout in the E3 and E4 regions, but lacks one gene in each region (E3 12.5K and E4 ORF1), presumably through loss after divergence from SAdV-1. E4 ORF1 has transforming properties and may have an anti-apoptotic role (Frese et al., 2003; Leppard, 1997), while the function of E3 12.5K is not known (Russell, 2000).

During compilation of the genome sequence from random M13 clones, a deletion variant was identified. From the proportion of clones obtained, the mutant comprised approximately one-third of the genomes present in the DNA stock. The deletion spanned 659 bp (from nt 1138 to 1797) of the genome sequence, a region that contains the 3' end of E1A, the majority of E1B 19K and the overlapping 5' end of E1B 55K (Fig. 1). The presence of the mutant was confirmed by PCR analysis of the same stock of DNA used for sequencing, utilizing primers mapping in the regions flanking the deletion (data not shown). It is not known whether the mutant was present in the virus as originally isolated, or whether it was generated during passage in vitro. Mutants lacking sequences in the E1 region are often constructed in vitro specifically for development as vectors. It is possible that the SAdV-1 deletion mutant could be separated from parental virus by using an appropriate complementing cell line and utilized similarly.

Phylogenetic analyses were carried out for each SAdV-1 gene. Fig. 3 shows examples for the four genes employed for SAdV-3, HAdV-40 and HAdV-41. Angle brackets indicate repeated motifs in the shaft region as defined by Kidd et al. (1993). The commencement of the knob region is shown for the short fiber.

and a subgroup of HAdV-B (Kidd et al., 1995). The other HAdVs and chimpanzee adenoviruses have two adjacent VA-RNA genes. SAdV-1 was studied by Kidd et al. (1995), but its VA-RNA gene was not amplified successfully, even though the sequences of one primer pair (VA5 and VA6) match the SAdV-1 sequence. The SAdV-1 VA-RNA is predicted to be 164 nt. This is similar to the VA-RNAs of three other monkey adenoviruses (SAdV-16, -19 and -13, with 168, 168 and 146 nt, respectively; Kidd et al., 1995) and of HAdV-40 and HAdV-12 (159 and 145 nt, respectively). In contrast, the SAdV-3 VA-RNA is only 109 nt in size and similarly short VA-RNAs characterize most monkey adenoviruses (Kidd et al., 1995).

In addition to the fiber and VA-RNA genes, adenovirus genomes exhibit the greatest organizational differences in the E3 and E4 regions (Davison et al., 2003; Ursu et al., 2004). The SAdV-1 E3 and E4 regions each encode six genes corresponding to those in SAdV-3 (Kovács et al., 2004), the same arrangement as in HAdV-A. HAdV-40 has a similar gene layout in the E3 and E4 regions, but lacks one gene in each region (E3 12.5K and E4 ORF1), presumably through loss after divergence from SAdV-1. E4 ORF1 has transforming properties and may have an anti-apoptotic role (Frese et al., 2003; Leppard, 1997), while the function of E3 12.5K is not known (Russell, 2000).

During compilation of the genome sequence from random M13 clones, a deletion variant was identified. From the proportion of clones obtained, the mutant comprised approximately one-third of the genomes present in the DNA stock. The deletion spanned 659 bp (from nt 1138 to 1797) of the genome sequence, a region that contains the 3' end of E1A, the majority of E1B 19K and the overlapping 5' end of E1B 55K (Fig. 1). The presence of the mutant was confirmed by PCR analysis of the same stock of DNA used for sequencing, utilizing primers mapping in the regions flanking the deletion (data not shown). It is not known whether the mutant was present in the virus as originally isolated, or whether it was generated during passage in vitro. Mutants lacking sequences in the E1 region are often constructed in vitro specifically for development as vectors. It is possible that the SAdV-1 deletion mutant could be separated from parental virus by using an appropriate complementing cell line and utilized similarly.

Phylogenetic analyses were carried out for each SAdV-1 gene. Fig. 3 shows examples for the four genes employed for SAdV-3, HAdV-40 and HAdV-41. Angle brackets indicate repeated motifs in the shaft region as defined by Kidd et al. (1993). The commencement of the knob region is shown for the short fiber.
SAdV-1 fiber genes proved very sensitive to the parameters used, but was generally consistent with the proposition that the two genes arose by duplication in a common ancestor of SAdV-1 and HAdV-F, after this lineage had diverged from HAdV-A (Bailey & Mautner, 1994).

Although SAdV-1 and HAdV-40 form a monophyletic group, the extent of phylogenetic separation between them is similar to that between the HAdV species, thus indicating that SAdV-1 may represent a novel species. At least two of the species-demarcation criteria formulated by the International Committee on Taxonomy of Viruses must be met in order to classify a novel adenovirus species. These criteria include phylogenetic distance, DNA hybridization, restriction-endonuclease profiles, nucleotide composition, oncogenicity in rodents, growth characteristics, host range, cross-neutralization, lack of recombination, number of VA-RNA genes, haemagglutination and genetic organization of the E3 region (Benkő et al., 2000). SAdV-1 unambiguously fulfills the requirements for proposal as a novel species distinct from the other early primate adenovirus lineages represented by HAdV-40, SAdV-3 and HAdV-12. Its phylogenetic distance from HAdV-40, its closest relative, is much greater than 10% (the official differentiator value) and SAdV-1 differs from HAdV-40 in host species, nucleotide composition, organization of the E3 region and

Fig. 3. Amino acid distance-based neighbour-joining trees for four adenovirus genes. SAdV-1 is indicated by an arrow in each tree. The trees were prepared by using MEGA 2.1 (Kumar et al., 2001) and tested by bootstrapping, with values indicated where above 75%. The trees were rooted by using data for bovine adenovirus 1 (BAdV-1) or BAdV-3. Sequences other than those derived for SAdV-1 in the present work were obtained from GenBank (accession nos in Table 1 of Kovács et al., 2004; AY487947 for HAdV-4, Jacobs et al., 2004).

G. M. Kovács and others
restriction-endonuclease profiles (as derived from the genome sequence). SAdV-1 is distinct from SAdV-3 in phylogenetic distance, restriction-endonuclease profiles, host species, haemagglutination and cross-neutralization properties, and from HAdV-12 in phylogenetic distance, host species, nucleotide composition and restriction-endonuclease profiles.

The genealogies shown in Fig. 3 raise questions regarding the origins of the monkey adenovirus and HAdV lineages. The apparently closer relationship between monkey adenoviruses and HAdV-F has led to the suggestion of a transfer between hosts (Tiemessen & Kidd, 1995). Alternatively, it is possible that the earlier branches diverged in ancestral primates before the human lineage separated. Nonetheless, it seems likely that the genetic diversity of Old World monkey adenovirus species will turn out to be greater than that of HAdVs because of the longer evolutionary history of their hosts. This is likely to have an impact on the future of adenovirus taxonomy.

Acknowledgements

Technical advice contributed by C. Cunningham and A. Dolan is gratefully acknowledged. Part of this work was conducted during a period of study by G. M. K. at the MRC Virology Unit sponsored by the Federation of European Microbiological Societies. The work was also supported by Hungarian research grants OTKA T043422 and MEH 40.0499/2004. G. M. K. is a postdoctoral fellow of the Hungarian Research Scholarship.

References


