Szent István University

Postgraduate School of Veterinary Science

Molecular characterisation of simian adenoviruses

Ph.D. dissertation

Iva Podgorski

2016

Supervisors and consultants:

……..
Prof. Dr. Mária Benkő, D.Sc.
Institute for Veterinary Medical
Research
Centre for Agricultural Research
Hungarian Academy of Sciences
Supervisor

……
Prof. Dr. Balázs Harrach, D.Sc.
Institute for Veterinary Medical Research
Centre for Agricultural Research
Hungarian Academy of Sciences
Supervisor

Dr. Győző Kaján
Institute for Veterinary Medical Research
Centre for Agricultural Research
Hungarian Academy of Sciences
Consultant

Copy …. of eight

……..
Iva Podgorski

Table of contents

Abbreviations **5**

1. Summary **6**

2. Introduction **8**

3. Review of literature **10**

3.1 Family *Adenoviridae* 10

3.2 Structure of adenoviruses 12

3.3 Non-human primate adenoviruses 15

3.3.1 Ape adenoviruses 15

3.3.2 Old World monkey adenoviruses 16

3.3.3 New World monkey adenoviruses 18

4. Aims of this study **19**

5. Materials and methods **20**

5.1 Strains from the American Type Culture Collection (ATCC) 20

5.1.1 Viruses 20

5.1.2 DNA extraction 22

5.1.3 PCR and DNA sequencing 22

5.1.4 ATCC mixtures – end-point dilution assay 25

5.1.5 Molecular cloning of fibre-1 knobs 25

5.1.6 Fluorescence-activated cell sorting 26

5.2 Screening for new adenoviruses 27

5.2.1 Samples for screening 27

5.2.2 DNA extraction 27

5.2.3 PCR and DNA sequencing 27

5.2.4 Cell culture methods 27

5.2.5 Full genome sequencing of red-handed tamarin adenovirus 1 29

5.3 Bioinformatics 29

6. Results **31**

6.1 Strains from the ATCC 31

6.1.1 Production of virions and viral DNA for next-generation sequencing 31

6.1.2 Partial genome sequencing and ATCC mixtures 33

6.1.3 Full genome sequencing 38

6.1.4 Receptor studies 46
6.2 Novel adenoviruses 47
6.2.1 Novel adenoviruses detected by PCR 47
6.2.2 Virus isolation attempts 47
6.2.3 Phylogeny inference 47
6.2.4 Red-handed tamarin adenovirus 1 50

7. Discussion 52

7.1 Strains from the ATCC 52
7.1.1 Production of virions and viral DNA for next-generation sequencing 52
7.1.2 Partial genome sequencing and ATCC mixtures 53
7.1.3 Full genome sequencing 59
7.1.4 Receptor studies 65
7.2 Novel adenoviruses 65
7.2.1 Novel adenoviruses detected by PCR 65
7.2.2 Virus isolation attempts 66
7.2.3 Phylogeny inference 67
7.2.4 Red-handed tamarin adenovirus 1 69

8. New scientific results 71

9. References 72

10. Publications 80

11. Acknowledgements 81

12. Appendix 82

12.1 Tables 82
12.2 Accession numbers of GenBank retrieved sequences 85

Abbreviations

aa	amino acid(s)
AAV	adeno-associated virus
AdV	adenovirus
ATCC	American Type Culture Collection
BaAdV	baboon adenovirus
bp	base pair(s)
CHO	Chinese hamster ovary (cells)
CPE	cytopathic effect
DMEM	Dulbecco's modified Eagle's medium
dNTP	deoxyribonucleotide triphosphate mix
FACS	fluorescence-activated cell sorting
FAdV	fowl adenovirus
FBS	fetal bovine serum
FCS	fetal calf serum
FITC	fluorescein isothiocyanate
HAdV	human adenovirus
HAG	haemagglutination group
HEK	human embryonic kidney (cells)
ICTV	International Committee on Taxonomy of Viruses
ITR	inverted terminal repeat
kb	kilobasepair(s)
MFC	microcebus fibroblast cells
NCBI	National Center for Biotechnology Information
NGS	next-generation sequencing
nt	nucleotide(s)
NWM	New World monkey
ORF	open reading frame
OWM	Old World monkey
PCR	polymerase chain reaction
pol	DNA-dependent DNA polymerase
RNA	ribonucleic acid
RPMI	Roswell Park Memorial Institute (medium)
SAdV	simian adenovirus
SDS PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
suniv	simian universal
TC	tissue culture
TE	Tris-EDTA
TMAdV	titi monkey adenovirus
TP	terminal protein
UXP	U-exon protein
VA RNA	virus-associated ribonucleic acid
wt	wild type

1. Summary

Among primates, adenoviruses (AdVs) were identified in humans, apes (chimpanzees, bonobos and gorillas), several species of Old World monkeys (OWMs) and a few species of New World monkeys (NWMs). Based on molecular characterization, we propose a species classification for the (serotyped) OWM AdV prototypes (SAdV-1 to 20). Majority of these SAdVs was successfully propagated on Vero cells. Based on partial sequences of the IVa2, DNA-dependent DNA polymerase (*pol*), penton base and hexon genes acquired by consensus PCR from all non-sequenced SAdV types, we found most of them to belong to one or the other of the two earlier accepted species that contain earlier studied monkey AdV serotypes. Species *Human mastadenovirus G* has been established for HAdV-52, but SAdV-1, -2, -7, -11, -12, and -15 also belong to it. The species *Simian mastadenovirus A* includes SAdV-3, -4, -6, -9, -10, -14, and -48. Several SAdVs (SAdV-5, -8, -49, -50) together with baboon AdV-1, and nine rhesus monkey AdV strains seemed to be members of the species *Simian mastadenovirus B* approved officially by the ICTV during my study. *Simian mastadenovirus C*, officially accepted during this study, should contain SAdV-19, together with strains baboon AdV-2/4 and -3. Our study revealed the existence of five further virus lineages, eventually proposed as species to the ICTV. These candidate species are *Simian mastadenovirus D* (SAdV-13), *Simian mastadenovirus E* (SAdV-16), *Simian mastadenovirus F* (SAdV-17, -18), *Simian mastadenovirus G* (SAdV-20) and *Simian mastadenovirus H* (strain SAdV-23336, proposed SAdV-54). Several biological and genomic properties such as the host origin, hemagglutination panel, number of fibre genes, and GC content of the DNA support this proposed classification. Three SAdV strains, originating from the American Type Culture Collection turned out to be the mixtures of at least two virus types, either of the same or even of two different species. These prototype strains are SAdV-12 and -15, containing viruses belonging to *Human mastadenovirus G*, and SAdV-5 containing viruses belonging to *Human mastadenovirus G* and *Simian mastadenovirus B*. Seven of the studied SAdVs were fully sequenced, and all of them shared genetic composition characteristic for mastadenoviruses, with difference seen in the fibre region: some of them contain one, two, or even three fibre genes. This is the first time we detected three fibre genes in an AdV, which might give a great potential to these AdVs in vectorizing. 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. Also, for the first time in SAdVs, the two downstream exons belonging to the gene of the so-called U exon protein could also be predicted. SAdV-2 differed from the other SAdVs in the E1B region of the genome: in place of the E1B 19K gene, it had an ITR repetition and another copy of the E3 ORF1 gene. Phylogenetic calculations, based on the fibre-1 and the major capsid protein,

the hexon, implied that recombination events might have happened between members of different AdV species. Molecular cloning of fibre-1 knobs of the SAdVs belonging to *Human mastadenovirus G* was performed in order to use them for receptor binding studies on A549 cells (used with or without neuraminidase pre-treatment). The fluorescence activated cell sorting results indicated that these knobs use sialic acid-containing glycans as receptors. Analyses of the full OWM AdV sequences further supported the theory on virus−host co-evolution, clustering together with other OWM AdVs.

In order to find novel non-human primate AdVs, we screened by PCR 138 organ and fecal samples representatives of different simian and prosimian species, and 18 new AdVs were detected: 8 in prosimians, 7 in NWMs, 1 in OWMs, and 2 in apes. This was the first study in which AdVs were detected in prosimians, gibbons and orangutans. A nested PCR targeting the IVa2 gene of mastadenoviruses proved to be the most powerful method in our hands for the detection of SAdVs, therefore the phylogenetic analyses were based on the short fragment obtained from the IVa2 gene. In spite of the successful PCRs, our attempts to isolate the detected viruses on different cell lines (Vero E6, HEK293, A549, CHO-K1, 3T6, MFC, cmt93) remained futile. New species was proposed to the ICTV for the titi monkey AdV described previously, putatively named *Platyrrhini mastadenovirus A*. Analysis of the almost fully sequenced NWM AdV, the red handed tamarin AdV-1, revealed genetic content similar to that of the titi monkey AdV, but less similar to OWM AdVs, confirming the NWM origin of the virus. On the phylogenetic trees, the AdVs retrieved from the both groups never examined before (prosimians and NWMs) appeared indeed on novel branches. The theory on virus−host co-evolution was supported by comparing the phylogeny of the primate hosts with that of their AdVs.

2. Introduction

Adenoviruses (AdVs) were discovered more than 60 years ago (Rowe *et al.*, 1953) and have been described in representatives of many vertebrate species since (Harrach, 2014). Most of the AdVs are apathogenic in healthy individuals, but some can cause disease and in rare cases even death. On the other hand, because of their biological characteristics, AdVs have a growing popularity as gene delivery and vaccination vectors and also as possible anti-cancer therapeutic agents. However, the human population is widely infected by AdVs and the existing specific antibodies significantly limit the medical applicability of human AdVs (HAdVs). Consequently, there is an increased interest in the possible use of non-human, especially simian AdVs (SAdVs; Alonso-Padilla *et al*., 2016; Harrach & Podgorski, 2014). Non-human primate AdVs, as the closest relatives to HAdVs, have great potential for use in medicine. Since gorilla and chimpanzee AdVs might be too close relatives of HAdVs, we search for AdVs in the more ancient primate species: orangutans and gibbons, Old World monkeys (OWMs), New World monkeys (NWMs) and prosimians, about which there is only limited information available. Chimpanzee AdVs have already been used as vaccine vectors in humans (reviewed by Capone *et al.*, 2013), but many recent studies raised the question of possible host switching of ape AdVs to humans, and the safety of such vectors (Benkő *et al.*, 2014; Dehghan *et al*., 2013a; Mennechet *et al*., 2015). The ideal vector virus should be evolutionarily and characteristically close enough to HAdVs to be molecularly handled in the same ways, but still far enough to prevent the possibility of crossing the species barrier and infecting humans.

More than 50 HAdV types are classified into seven species (*Human mastadenovirus A* to *Human mastadenovirus G*, HAdV-A to HAdV-G) within the genus *Mastadenovirus*. Certain HAdV species do contain also chimpanzee and/or other ape AdVs; HAdV-G contains HAdV-52 and several monkey AdVs. The first classification of SAdVs was based on hemagglutination-inhibition test as a tool of taxon demarcation (Rapoza, 1967). Nowadays, the recognized diversity of SAdVs is approaching that of the HAdVs. Species *Simian mastadenovirus A* (SAdV-A) was, until this study, the only species officially approved for monkey AdVs exclusively. Phylogeny analysis of the 25 recognized SAdV serotypes, SAdV-1 to 20, isolated from Old World monkeys (OWM), and SAdV-21 to 25 from chimpanzees has been performed by PCR amplification and sequencing of the virus-associated (VA) RNA gene(s) (Kidd *et al.*, 1995). However, the short sequences acquired from the VA RNA are not ideal for comparative analysis. By now, the ape AdVs are well characterized and fully classified, but most of the monkey AdVs still await classification, and only very short or no sequence is published from their genome. In the past decades, there have been many AdVs found in OWMs, but the majority of them has not been sequenced yet.

Our knowledge concerning the NWM AdVs is even more limited. The only fully sequenced such virus is the titi monkey AdV (TMAdV; Chen *et al.*, 2011). There were a few reports on NWM AdVs (Gál *et al.*, 2013; Hall *et al.*, 2012; Shroyer *et al.*, 1979; Wevers *et al.*, 2011), but with hardly any details about them. In prosimian hosts, there have not been any AdVs reported as yet. The comparative genome analysis of different AdVs would help us understand the viral diversity and evolution, the function of certain genes. We could also assess their possible use in medicine for which the isolation and efficient replication in cell culture would be essential. The capacity of the viruses to reach high titres shows us if they could be used as promising delivery vectors at all.

We were interested also in the amelioration of the taxonomy of the family *Adenoviridae* by establishing correct novel species for the newly characterized primate AdVs. The aim of this study was to find and characterize novel AdVs in prosimians, NWMs, OWMs, orangutans and gibbons, as well as to get more data about the OWM AdVs isolated previously, with the main focus on the host specificity, virus−host evolution, AdV variability, and differences between them on molecular level.

3. Review of literature

3.1 Family *Adenoviridae*

Adenoviruses are medium-sized (70-90 nm), non-enveloped viruses which were first isolated from human adenoid tissue, hence the name (Rowe *et al.*, 1953). AdVs are mostly not pathogenic for their hosts, although in rare cases they can cause diseases with fatal outcome (Benkő, 2015). Because of their widespread occurrence in a variety of vertebrate hosts (Harrach *et al.*, 2011), they make an ideal model for the study of viral evolution (Figure 1). Earlier the classification of AdVs was based on biological characteristics (host origin, erythrocyte agglutination, oncogenicity, and neutralization by specific antisera), but later on, development of DNA sequencing and bioinformatics enabled phylogenetic relationships to be revised and clarified on the basis of gene and genome similarity (Bailey & Mautner, 1994; Davison *et al.*, 2003).

The icosahedral capsid of AdVs contains double-stranded, linear DNA which can range from 26.1 (frog AdV-1) to 48.4 kb (white sturgeon AdV-1) in size, with nucleotide composition of 33.6% (ovine AdV-7) to 67.6% G+C (turkey AdV-1; Harrach, 2014). Ends of the DNA contain inverted terminal repeats (ITRs) which encompass 30 (atadenovirus) to 371 (mastadenovirus) base pairs (bp), covalently linked with 5' ends to a terminal protein (TP).

There are five accepted genera within the family *Adenoviridae*: *Mastadenovirus*, *Aviadenovirus*, *Siadenovirus*, *Atadenovirus* and *Ichtadenovirus* (Harrach, 2014). One more genus was proposed, named *Testadenovirus* (Doszpoly *et al.*, 2013), but is not officially accepted yet. Most representatives of each genus are presented in Figure 1. The genera *Mastadenovirus*, *Aviadenovirus*, *Ichtadenovirus* and *Testadenovirus* were named after the hosts in which the AdVs were found: mammals, birds, fish and testudinoid turtles, respectively. The *Atadenovirus* genus was named after a bias towards high A+T content (Benkő & Harrach, 1998; Dán *et al.*, 1998), whereas *Siadenovirus* was named after the presence of a gene encoding a sialidase (Davison *et al.*, 2003). More than 40 species have been established within the mentioned genera, based on several species demarcation criteria such as the host origin, phylogenetic distance (>5-15% amino acid sequence divergence of the DNA-dependent DNA polymerase by distance matrix or maximum likelihood analysis), genome organization differences, G+C content, oncogenicity in rodents, growth characteristics, host range, cross-neutralization, ability to recombine, number of VA-RNA genes and hemagglutination properties (Harrach, 2014; Harrach *et al.*, 2011).

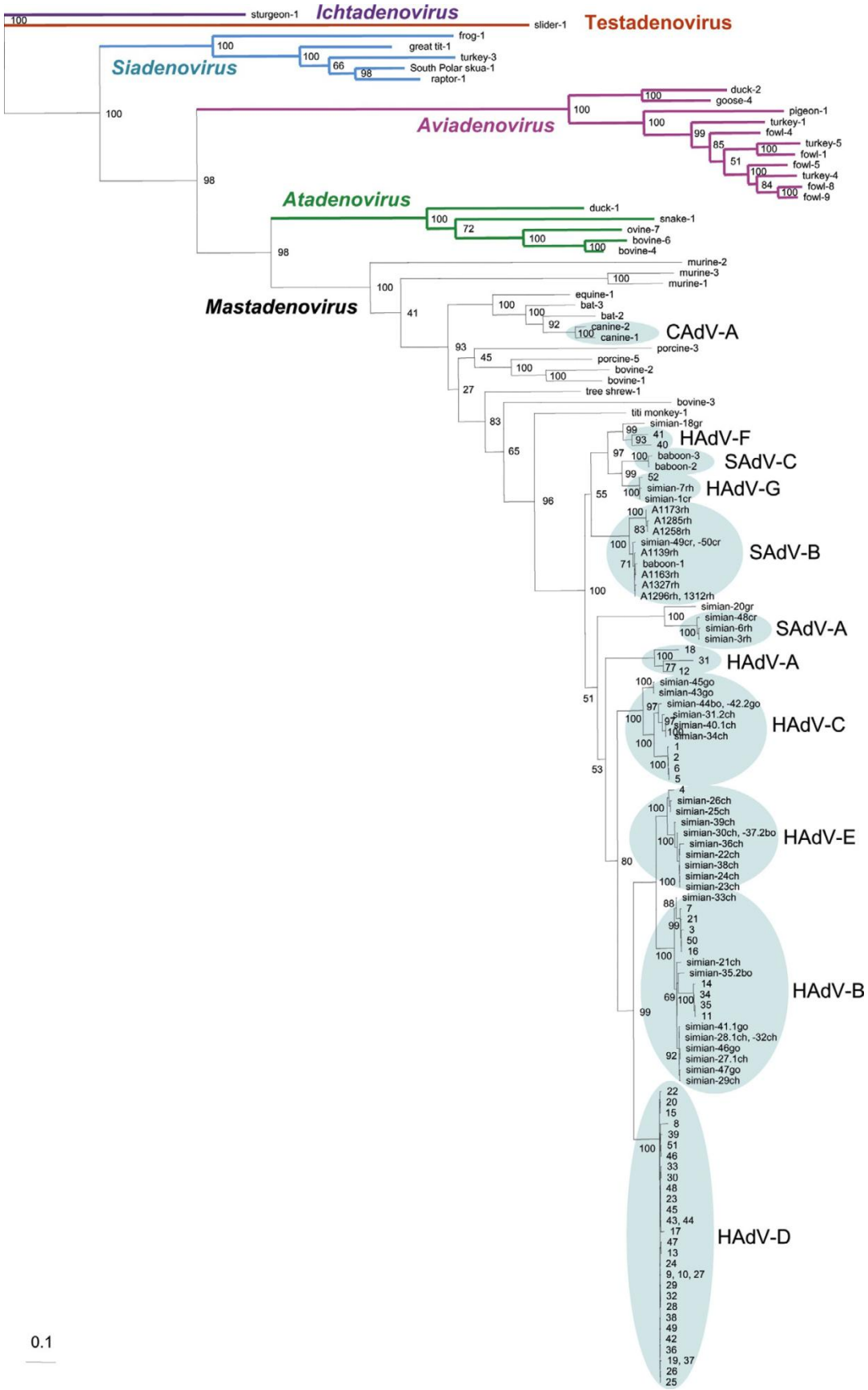


Figure 1. Phylogenetic tree of adenoviruses based on maximum likelihood analysis of available DNA-dependent DNA polymerase amino acid sequences (Harrach, 2014).

3.2 Structure of adenoviruses

The tropism of AdVs is mostly determined by the capsid proteins, which interact with the host cell surface receptors. The capsid consists primarily of hexon (polypeptide II) and penton base (polypeptide III) proteins that form pentameric structures at each of the 12 vertices, securing a trimeric fibre (polypeptide IV) that projects outward. The homotrimeric fibre protein has three structural domains. The proximal tail, the middle part called the shaft, and the distal head or knob domain which recognizes the cell surface receptors for the initial attachment to the host cell. The flexibility of fibre has an important role during this interaction (Wu *et al.*, 2003). The penton base of mastadenoviruses has an RGD motif to bind integrins thus triggering internalization by endocytosis (Wickham *et al.*, 1993). Additional structural proteins include cement proteins (minor coat proteins IIIa, VI, IX and VIII) and the core proteins (TP, V, VII, and X) that associate with the AdV genome (Figure 2; San Martín, 2012).

AdVs encode several precursor proteins which have to be cleaved by the also virus-coded protease for the immature particle to become infectious. These include three capsid proteins (pIIIa, pVI and pVIII), and three core proteins (pVII, pX and pTP; Diouri *et al.*, 1996). Core protein V and minor coat protein IX are present only in the members of the genus *Mastadenovirus*. Members of some species and genera differ in some of the minor capsid components, however the general virion architecture is conserved (Figure 2). Protein IX is located on the outer surface of the capsid, and has a capsid stabilizing role (Colby & Shenk, 1981), and a role in the viral entry (Strunze *et al.*, 2011). Its exposed domain is available for interaction with the host cells, and this makes it popular for adenoviral vector modification (Parks, 2005). The IIIa protein might have a role in stabilizing the vertex region and the packaged genome upon assembly or in signalling for vertex and genome release during uncoating (Abrescia *et al.*, 2004; Ma & Hearing, 2011; San Martín *et al.*, 2008). Protein VI has a role in virus escape into cytosol (Moyer *et al.*, 2011), in facilitating trafficking to the nucleus along the microtubular network (Wodrich *et al.*, 2010), as an activator of the adenoviral gene expression (Schreiner *et al.*, 2012), in promoting transport of newly synthesized hexon to the nucleus (Wodrich *et al.*, 2003), and as a substrate and cofactor of the adenoviral protease to yield the infectious viral particle (Mangel *et al.*, 1993; Mangel *et al.*, 1996). Polypeptide VIII, beside its architectural contribution to the capsid (Liu *et al.*, 1985), might have a role in genome packaging as well, since it interacts with the putative packaging protein IVa2 (Singh *et al.*, 2005).

General genome organization of AdVs can be divided to a conserved middle and variable terminal parts (Figure 3; Harrach *et al.*, 2011). E1A, E1B 19K, the genes for protein IX and V, VA-RNA gene, and the E3 and E4 region genes (with the exception of 34K) occur in mastadenoviruses only.

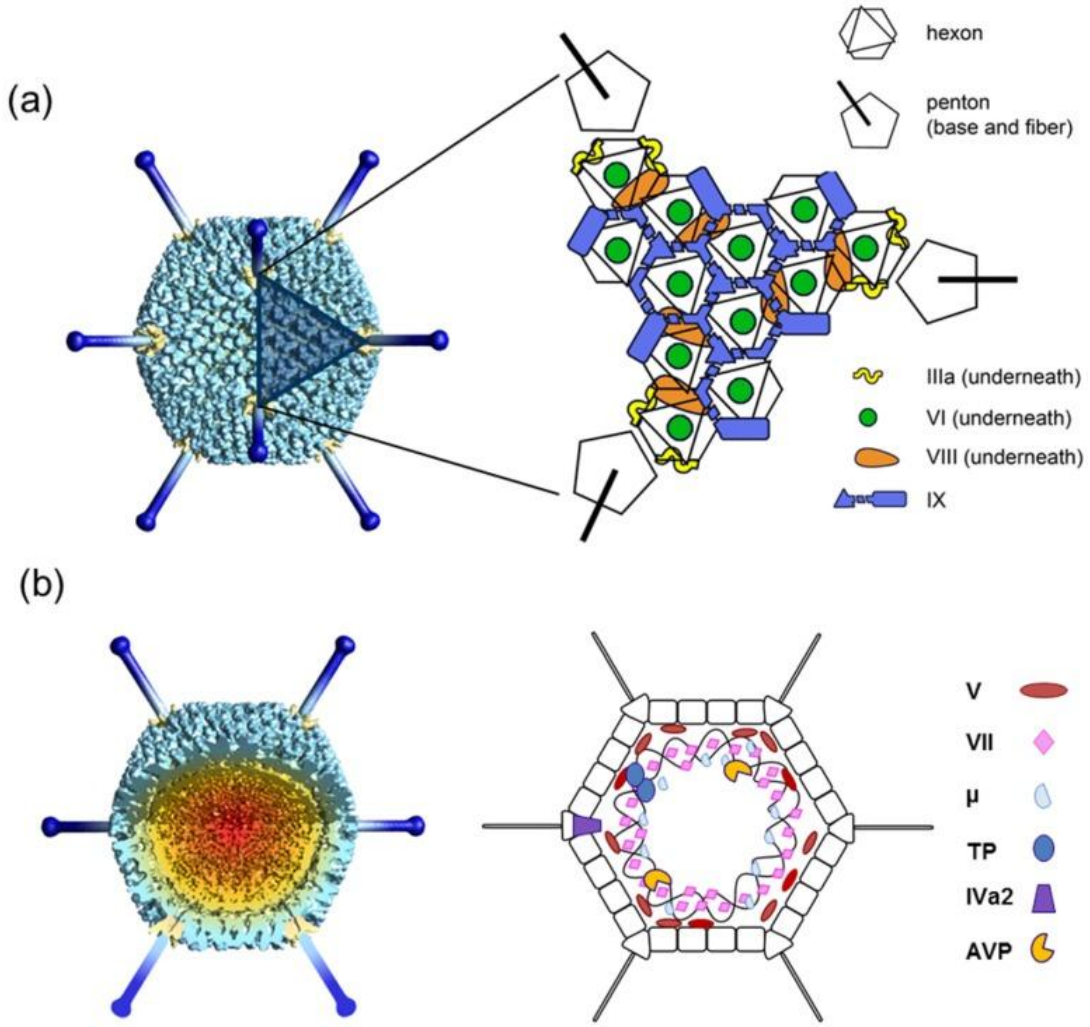


Figure 2. Structure and components of adenovirus: (a) left - penton bases are highlighted in yellow and the shaded triangle indicates one facet; (b) non-icosahedral components (San Martín, 2012). AVP: adenoviral protease; μ (mu) protein is also called protein X.

Most AdVs contain one fibre at each vertex, except fowl AdVs (genus *Aviadenovirus*) which have two fibres per vertex (Gelderblom & Maichle-Lauppe, 1982), and lizard AdV-2 (genus *Atadenovirus*) proposed to have even three fibres on some of its vertices (Pénzes *et al*., 2014). Interestingly, the number of coding genes is not always in accordance with the number of protruding fibres. Two fibre genes can be found in a few HAdVs (HAdV-40, -41 and -52; Favier *et al.*, 2002; Jones *et al.*, 2007; Kidd *et al.*, 1993), most of the OWM AdVs (Abbink *et al.*, 2015; Chiu *et al.*, 2013; Kovács *et al.*, 2005; Roy *et al.*, 2009, 2011, 2012), several aviadenoviruses (Griffin & Nagy, 2011; Kaján *et al.*, 2010, 2012; Marek *et al.*, 2014b), and two atadenoviruses (Pénzes *et al.*, 2014; To *et al.*, 2014). However, for most of the AdVs, the number of protruding fibres is not known yet.

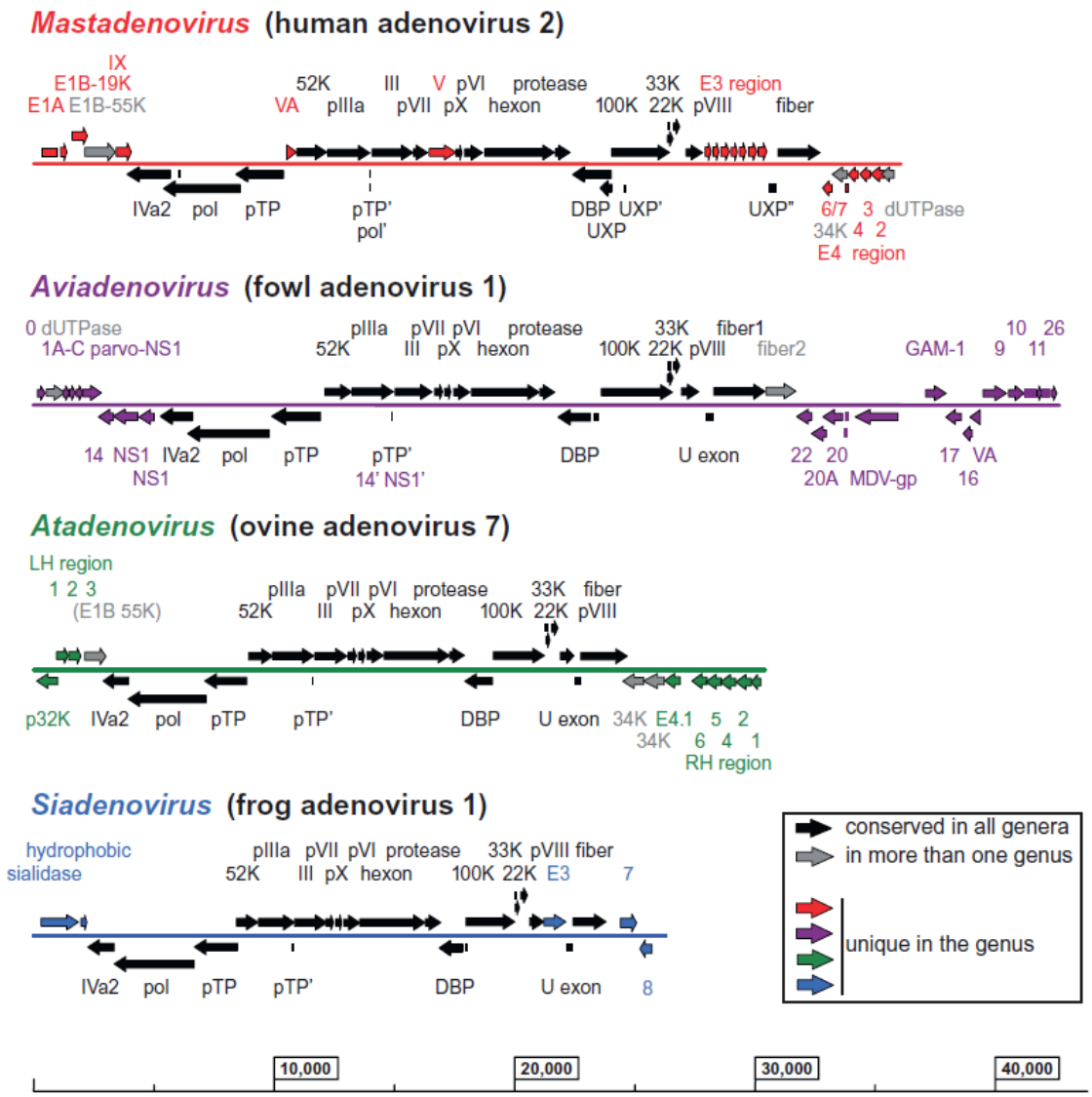


Figure 3. Schematic illustration of different genome organizations in members of four adenovirus genera. Black arrows depict genes conserved in every genus, gray arrows show genes present in more than one genus, and coloured arrows show genus-specific genes (Harrach *et al.*, 2011).

All HAdVs possess at least one VA-RNA gene (Ma & Mathews, 1996; Mathews & Shenk, 1991), and approximately 80% of them has even two of them (reviewed by Vachon & Conn, 2016). VA RNA is essential for efficient virus replication (Bhat & Thimmappaya, 1984; Thimmappaya *et al.*, 1982). Some OWM AdVs, studied earlier, have been found to have one VA-RNA gene only (Kidd *et al.*, 1995). On the other hand, chimpanzee AdVs, similar to the majority of HAdVs (in species HAdV-B to E), have been described 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 had been missing indeed from the genomes of some viruses (Kidd *et al*, 1995).

3.3 Non-human primate adenoviruses

Based on the molecular phylogeny, primates can be grouped into two big suborders, namely Strepsirrhini (containing the prosimians) and Haplorrhini (containing NWMs, OWMs, gibbons, great apes, and humans; Perelman *et al.*, 2011). Primate AdVs, assigned to eight species previously established within the genus *Mastadenovirus*, represent the best studied AdVs today. From the representatives of species in the Haplorrhini suborder, more than 150 AdV types are known (http://www.vmri.hu/~harrach/ADENOSEQ.HTM), albeit with an overwhelming majority of ape and human AdVs compared to the more ancient primates (NWMs and prosimians) with very limited or no information about the prevalence, evolution, and genome characteristics of their AdVs. More than 50 HAdV types are classified into seven HAdV species (HAdV-A to G). Certain HAdV species do also contain chimpanzee or other ape AdVs; HAdV-G contains HAdV-52 and several monkey AdVs. Despite the fact that OWM AdVs were discovered more than 50 years ago (Hull *et al.*, 1956), and found in many different species (macaques, grivets, black and white colobuses, red colobuses, hamadryas baboons, yellow baboon), more than half of them were not studied in detail. Species *Simian mastadenovirus A* (SAdV-A) was, until this study, the only species officially approved for monkey AdVs exclusively. The interest in more ancient SAdVs is rising with the awareness of the risk they may pose for humans in case of host switching (Benkő *et al.*, 2014). On the other hand, there is an increasing interest in vectors derived from non-human AdVs (Lopez-Gordo *et al.*, 2014), especially in SAdVs since they are the closest relatives to HAdVs, but still evolutionarily far enough not to be influenced by the pre-existing specific immunity in humans. SAdVs have been found to be associated with several diseases in primates, including diarrhoea, pneumoenteritis, conjunctivitis, and hepatitis (Bányai *et al.*, 2010; Kim *et al.*, 1967; Vasileva *et al.*, 1978; Zöller *et al.*, 2008), and some of them have been reported to induce tumours when injected into neonate rodents (Hull *et al.*, 1965). Among NWMs, the titi monkey AdV (TMAdV) outbrake caused high fatality case rate (83%), most probably because the titi monkey is not the original host of this virus (Chen *et al.*, 2011). Short sequences from various genes of monkey AdVs have often been reported from colonies of captive macaques that had either been suffering from diarrhoea (Wang *et al.*, 2007) or not showing any clinical signs ascribed to AdVs (Lu *et al.*, 2011; Wevers *et al.*, 2011).

3.3.1 Ape adenoviruses

The first description of a SAdV in the literature was that of a chimpanzee AdV (Rowe *et al.*, 1956), today known as SAdV-21 classified into the species HAdV-B. Later on, when investigating chimpanzees suffering from kuru, experiments resulted in the discovery of four

novel ape AdVs (Rogers *et al.*, 1967). The similarity of chimpanzee AdVs to HAdV-4 strains of the species HAdV-E was recognised quite early (Li & Wadell, 1988). The first complete chimpanzee AdV genome sequence was that of SAdV-25 (Farina *et al.*, 2001). Soon after, development of vector vaccines from chimpanzee AdVs started (Xiang *et al.*, 2002), and it has been the subject of growing interest (Capone *et al.*, 2013). Consequently, there was a growing number of different chimpanzee AdV isolates that had been studied, resulting in fully sequenced genomes of the other four chimpanzee AdV types, SAdV-21 to 24 (Roy *et al.*, 2004) and those of two additional chimpanzee AdVs (under the strain names of ChAd3 and ChAd6; Colloca *et al.*, 2012; Peruzzi *et al.*, 2009). Ape AdVs have been isolated not only from chimpanzees but also from bonobos and gorillas (Roy *et al.*, 2009). These viruses have been proposed to be members of species HAdV-B, -C and -E, respectively, as they are closest to those genetic lineages (Roy *et al.*, 2009). Partial genome analysis of gorilla AdVs confirmed the theory on the mixed host origin of members of the species HAdV-B (Wevers *et al.*, 2010). Furthermore, Colloca and co-workers (2012) screened more than a thousand faecal samples from chimpanzees and bonobos and isolated AdVs from around 50% of them. The full sequences of some of these viruses indicated that they are closest to members of the species HAdV-B, -C or –E (Colloca *et al.*, 2012). Almost all viruses in the species HAdV-D are from human sources. However, a recent study suggested that some chimpanzee AdVs might belong to species HAdV-D (Wevers *et al.*, 2011). The same study also described an AdV found in gorilla closely related to species HAdV-F (Wevers *et al.*, 2011). Recently, the genome of an AdV isolated from chimpanzee has been fully sequenced, and was found to cluster with species HAdV-A (Zhou *et al.*, 2014).

3.3.2 Old World monkey adenoviruses

The study of monkey AdVs is lagging behind that of the ape AdVs. The first monkey AdVs, together with many other, mainly enteric, simian viruses from divergent families were discovered while testing poliomyelitis vaccines prepared on monkey kidney cell cultures made from macaques of two species (Hull *et al.*, 1956). These novel virus isolates were typed later, and many but not all of them turned out to be adenovirus. The AdV isolates got new name labels and numbers according to their serologic distinctness (SAdV-1 to 21 etc.). Additional monkey AdV serotypes, characterised by the lack of cross-neutralisation (Benkő *et al.*, 2000), were found by screening other macaques and monkeys from two additional species: grivet and baboon (Fuentes-Marins *et al.*, 1963; Hull & Minner, 1957; Hull *et al.*, 1958; Kim *et al.*, 1967; Malherbe & Harwin, 1963). The first grouping of monkey AdVs was based on their ability to haemagglutinate erythrocytes of different host origin (Rapoza, 1967). With the use of this biological assay, 16 monkey AdV strains were divided into four

haemagglutination groups (HAG I–IV). By analysing the nucleotide (nt) sequences of the left genome ends, researchers have inferred that SAdV-16 (originally named SA7) shares a similar organisation with HAdV-12, a member of the species HAdV-A (Kimelman *et al.*, 1985). A comparative sequence analysis of 25 distinct SAdV serotypes, SAdV-1 to 20, isolated from OWMs, and SAdV-21 to 25 from chimpanzees was performed by PCR amplification and sequencing of the VA-RNA genes (Kidd *et al.*, 1995). All chimpanzee AdVs were proved to have two (tandem) VA-RNA genes in their genomes. However, in monkey AdVs, only one VA-RNA gene was detected (or none). These data were applied in making the first phylogenetic tree of SAdVs (Kidd *et al.*, 1995). The short VA RNA sequences and the first partial hexon sequences, obtained from chimpanzee AdVs in our lab, prompted us to place SAdV-21 into the species HAdV-B, and SAdV-22 to 25 into species HAdV-E (Benkő *et al.*, 2000). These assumptions regarding the taxonomic place of the chimpanzee AdVs were confirmed later by phylogenetic analysis of other longer sequences (Benkő & Harrach, 2003; Farina *et al.*, 2001; Purkayastha *et al.*, 2005; Roy *et al.*, 2004).

The first full monkey AdV genome sequence published was that of SAdV-3 (isolated from a rhesus macaque; Kovács *et al.*, 2004). It was proposed to be the first member of a new species, SAdV-A. This species was approved by the ICTV and until my study was the only species that included OWM AdVs exclusively (Harrach, 2014). The next sequenced OWM AdV genome was that of SAdV-1 (isolated from a crab-eating macaque). This virus has been found to belong to species HAdV-G (Kovács *et al.*, 2005), together with a HAdV type, HAdV-52 (Jones *et al.*, 2007). Further full-genome sequences were published from SAdV-48, -49 and -50 (Roy *et al.*, 2009), as well as some partial sequences (Bányai *et al.*, 2010; Maluquer de Motes *et al.*, 2011). As the interest in OWM AdVs as potential gene delivery tools increased, additional SAdV genomes were fully sequenced. These included SAdV-7 (Roy *et al.*, 2011), followed by SAdV-6, -18 and -20 (Roy *et al.*, 2012). Phylogenetic analysis of AdVs that were newly isolated from rhesus macaques (Roy *et al.*, 2012) indicated that they belonged to a common lineage with SAdV-49 and -50, which had been sequenced previously (Roy *et al.*, 2009). For the classification of these SAdVs, the establishment of the species *Simian mastadenovirus B* (SAdV-B) was proposed, whereas SAdV-48 was described to belong to the species SAdV-A (Roy *et al.*, 2012; Roy *et al.*, 2009). Four novel AdV strains, found in olive baboons, have been sequenced recently. One strain, baboon AdV-1 (BaAdV-1), was proposed to be a member of the candidate species SAdV-B, while strains BaAdV-2, -3 and -4 were found to form a separate clade, representing the proposed new species *Simian mastadenovirus C* (SAdV-C; Chiu *et al.*, 2013). Most recently, another SAdV type (strain 23336) from rhesus macaque has been proposed to form a new species, “Simian mastadenovirus D” (SAdV-D; Malouli *et al.*, 2014). Several additional OWM AdVs

were described last year, belonging to the previously established species HAdV-G (Abbink *et al.*, 2015).

3.3.3 New World monkey adenoviruses

Much fewer AdVs have been detected in NWMs. The very first NWM AdVs were isolated from owl monkeys, *Aotus* sp. (Shroyer *et al.*, 1979). It took more than three decades until the next one was reported in red titi monkeys (*Callicebus cupreus*; TMAdV) with fatal pneumonia. This was the first (and still the only) NWM AdV that has been sequenced completely (Chen *et al.*, 2011). Furthermore, this was the first report about the potential cross-species transmission of an AdV from NWMs to humans. A few years later, the same group demonstrated that TMAdV was able to cause cross-species infection in common marmosets (*Callithrix jacchus*), but caused mild, self-limiting disease only (Yu *et al.*, 2013), similar to the infection in humans seen previously. Beside captive animals, AdVs were identified also in a few wild-living NWMs, including a white-lipped tamarin (*Saguinus labiatus*) and three common marmosets (Wevers *et al.*, 2011). Unfortunately, the partial sequences gained from these NWMs have not been submitted to GenBank. Nonetheless, the NWM AdVs were found to be phylogenetically well separated from, and more ancient than all the other primate AdVs. Finally, two more AdVs were detected in cotton-top tamarin (*Saguinus oedipus*; Hall *et al.*, 2012) and pygmy marmoset (*Cebuella pygmaea*; Gál *et al.*, 2013), respectively. Interestingly, the phylogenetic calculations, based on the partial adenoviral *pol* amino acid (aa) sequences, showed that they grouped with some other, non-primate mastadenoviruses. The latter virus turned out to be very similar to certain bat AdVs implying the possibility of interspecies host switch (Vidovszky *et al*., 2015).

4. Aims of this study

The aims of my studies were:

1. to obtain at least partial sequences from multiple genes of every known OWM SAdV serotype (available in ATCC) in order to explore their phylogenetic relationships and to establish the putative species they would belong to;

2. to obtain full genome sequences of some of the OWM SAdVs in ATCC in order to compare the genome organization with human and other non-human primate AdVs, to obtain information about eventual new or spliced genes, and to find vector candidates;

3. to improve the taxonomy of the family *Adenoviridae* and to establish correct novel species for the newly characterized primate AdV lineages;

4. to study the receptor binding properties of the fibre-1 of selected primate AdVs that contain at least two fibre genes;

5. to survey captive and wild OWMs, NWMs, prosimians, and orangutans and different species of gibbons (apes) for the presence of AdVs in order to find new AdV types and to characterise them preliminarily by partial sequencing;

5. Materials and methods

5.1 Strains from the American Type Culture Collection (ATCC)

5.1.1 Viruses

OWM SAdV strains (SAdV-1 to 20) deposited in the American Type Culture Collection (ATCC) were studied by PCR, DNA sequencing and/or bioinformatics (Table 1). For the purpose of the next generation sequencing (NGS), Vero cells E6 (derived from kidney epithelial cells of African green monkey, *Chlorocebus* sp.; the E6 lineage shows some contact inhibition) were infected with all the non-sequenced OWM SAdVs originating from the ATCC. For the initiation of infection, 5 µl of the virus suspension was inoculated into 25-cm^2 tissue culture (TC) flask with confluent Vero E6 cells. Some of the strains (SAdV-2, -8, -11, and -17) replicated well on this cell line and could be propagated in larger volumes (each in eight 175-cm^2 TC flasks) to obtain concentrated viral DNA. The cells were grown in RPMI (Roswell Park Memorial Institute) medium (Biosera, Boussens, France) without L-Glutamine, supplemented with antibiotics and 10% fetal bovine serum (FBS). Each of the 175-cm^2 flasks was inoculated by 0.5 ml of cell culture containing the virus. The cultures were examined every day for the cytopathic effects (CPE) with an inverted light microscope. When full CPE reached, the flasks were frozen and thawed three times. Low speed centrifugation (1300 × g, 30 min, 4°C) was used to remove the cell debris. Viruses were pelleted from the supernatants by ultracentrifugation (41,000 × g, 90 min, 4°C, Beckman rotor 60-Ti). The sediment was soaked in 50 µl of 1x TE buffer for several hours. The bottoms of the ultracentrifuge tubes were washed two more times with 50 µl of 1x TE buffer, and the suspensions were merged.

For the receptor binding studies, A549 cells were infected with SAdV-1, -2, -11 and HAdV-5, as a positive control. The virions were subjected to purification with CsCl gradient (adjusted protocol by Wadell *et al*., 2002). After obtaining ~90% CPE, the cells were harvested with scraper in 50 ml Falcon tubes and centrifuged for 5 min at 130 × g. The supernatant was discarded and the cells were resuspended in max. 6 ml of DMEM (Sigma Aldrich, Stockholm, Sweden), then frozen and thawed 3 times to release the virions. An equal volume (max. 6 ml) of Vertrel XF (Sigma Aldrich, St. Louis, MO, USA) was added to the samples and they were shaken by hands for ~10 min. This step provided an approximately 100x concentration of the virions. The suspension was then exposed to centrifugation at 3300 × g for 7 min. CsCl gradient was made with different density solutions (pH=7.5) in Beckman tubes: 1.27 g/ml (1.5 ml), 1.32 g/ml (2 ml), and 1.37 g/ml (2 ml). Top

Table 1. Names of the studied SAdVs and related information (Pantó *et al.*, 2015)

Name	Old name	Host species	Source	ATCC strain	Acc. number	Ref.
SAdV-1	SV1	Crab-eating macaque, *Macaca fascicularis*	rectal swab	VR-195	AY771780	(Kovács *et al.*, 2005)
SAdV-2	SV11	Rhesus macaque, *Macaca mulatta*	tissue culture	VR-196	KP853120, KP853125, KP853112	(Pantó *et al*., 2015)
SAdV-3	SV15			VR-1449	AY598782	(Kovács *et al.*, 2004)
SAdV-4	SV17			VR-198	KP853121, KP853126, KP853113	(Pantó *et al*., 2015)
SAdV-5	SV20		rectal swab	VR-199	KP853111, KP853127, KP853128, KP853114	(Pantó *et al*., 2015)
SAdV-6	SV39	Macaque, *Macaca sp.*	tissue culture	VR-200	CQ982401	(Roy *et al.*, 2012)
SAdV-7	SV25	Rhesus macaque, *Macaca mulatta*	rectal swab	VR-201	DQ792570	(Roy *et al.*, 2009)
SAdV-8	SV30	Crab-eating macaque, *Macaca fascicularis*		VR-1539	KP329561	(Podgorski *et al*., 2016)
SAdV-9	SV31	Macaque, *Macaca sp.*		VR-204	KP853122, KP853129, KP853115	(Pantó *et al*., 2015)
SAdV-10	SV32			VR-205	KP853110, KP853130, KP853116	(Pantó *et al*., 2015)
SAdV-11	SV33	Rhesus macaque, *Macaca mulatta*		VR-206	KP329562	(Podgorski *et al*., 2016)
SAdV-12	SV34		tissue culture (CNS)	VR-207	KP853123, KP853131, KP853132, KP853117	(Pantó *et al*., 2015)
SAdV-13	SV36	Macaque, *Macaca sp.*	tissue culture	VR-208	KP329563	(Pantó *et al*., 2015)
SAdV-14	SV37	Rhesus macaque, *Macaca mulatta*		VR-209	KP853124, KP853133, KP853118	(Pantó *et al*., 2015)
SAdV-15	SV38		cervical cord	VR-355	KP853109, KP853134, KP853135, KP853119	(Pantó *et al*., 2015)
SAdV-16	SA7	Grivet, *Chlorocebus aethiops*	rectal swab	VR-941	KP329564	(Podgorski *et al*., 2016)
SAdV-17	SA17		unknown	VR-942	KP329566 (unreleased)	(Pantó *et al*., 2015)
SAdV-18	SA18			VR-943	CQ982407	(Roy *et al.*, 2012)
SAdV-19	AA153	Yellow baboon, *Papio cynocephalus*	stool	VR-275	KP329565	(Podgorski *et al*., 2016)
SAdV-20	V340	Grivet, *Chlorocebus aethiops*	fatal pneumoenteritis	VR-541	HQ605912	(Roy *et al.*, 2012)
SAdV-48		Crab-eating macaque, *Macaca fascicularis*	stool	-	HQ241818	(Roy *et al.*, 2009)
SAdV-49					HQ241819	
SAdV-50					HQ241820	
BaAdV-1		Olive baboon, *Papio hamadryas anubis*	nasal swab		KC693021	(Chiu *et al.*, 2013)
BaAdV-2/4					KC693022	
BaAdV-3					KC693023	
A1139[a]		Rhesus macaque, *Macaca mulatta*	stool		JN880448	(Roy *et al.*, 2012)
A1163[a]					JN880449	
A1173[a]					JN880450	
A1258[a]					JN880451	
A1285[a]					JN880452	
A1296[a]					JN880453	
A1312[a]					JN880454	
A1327[a]					JN880455	
A1335[a]					JN880456	
SAdV-54 strain 23336[a]					KM190146	(Malouli *et al.*, 2014)

[a]strain name

layer with the virus was taken from the vertrel/cell suspension and put onto the CsCl gradient, followed by ultracentrifugation with Beckman SW41 rotor at 77,000 × g for 90 min at 4°C. The lower band was collected in the smallest volume possible and the density was checked with a refractometer. The virions were purified on illustra NAP-5 or NAP-10 columns (GE Healthcare Life Sciences, Uppsala, Sweden), depending on the volume collected. Glycerol was added to a final concentration of 10%, and the virus concentration was measured by spectrophotometer. Purified viruses were stored at -80°C.

5.1.2 DNA extraction

For NGS purposes, the DNA of SAdV-2, -8, -11, and -17 was extracted from the concentrated virions: 200 µl of the virus suspension was mixed with 5.4 µl of 20% sodium dodecyl sulfate (SDS) and 2 µl Proteinase K (20 mg/ml; Sigma-Aldrich, St. Louis, MO, USA) and incubated at 55°C for 1 h, followed by addition of 35 µl of 5 M NaCl and overnight incubation at 4°C. The mix was then centrifuged for 15 min at 13,000 × g and the supernatant was moved to a new tube and mixed with 400 µl mili-q water. One volume of phenol/chloroform/IAA mixture (25:24:1 ratio; pH 7,7-8,3; Sigma-Aldrich, St. Louis, MO, USA) was added and mixed, and centrifuged for 15 min at 13,000 × g. The upper phase was transferred to a new tube, and one volume of chloroform was added and mixed followed again by centrifugation for 15 min at 13,000 × g. The upper phase was moved to a new tube and mixed with 0.1 volume of 3 M NaAc (pH 5.2) and 2.5 volumes of absolute ethanol. The mixture was incubated for 1−2 h at -80°C, followed by centrifugation for 15 min at 13,000 × g. The supernatant was removed and the pellet was washed with 500 µl of 70% ethanol, followed by centrifugation for 15 min at 13,000 × g. The supernatant was discarded and the pellet was air-dried then dissolved in 50 µl of 1x TE buffer. The quality of DNA was checked by gel-electrophoresis on 1% agarose gel, and the concentration of DNA was measured by NanoDrop spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA).

5.1.3 PCR and DNA sequencing

Shorter or longer fragments of the genes of four well-conserved adenoviral proteins (namely IVa2, *pol*, penton base and hexon) were obtained by PCR from 14 SAdVs, the full genome sequence of which has not been published previously (SAdV-2, -4, -5, -8 to 17, and -19). Later on, some of these AdVs were fully sequenced by NGS (SAdV-2, -8, -11 and -17) or by traditional methods (SAdV-13, -16 and -19). The primer sequences and the estimated sizes of the expected PCR products are presented in Table 2. Fragments from the *pol* and the hexon genes were obtained by PCR methods published by others (Kiss *et al.*, 1996; Wevers *et al.*, 2011). The IVa2 gene fragment was amplified with consensus degenerate nested

primers, designed in-house on the basis of highly conserved aa motifs taken from an alignment containing different mastadenovirus sequences. Similarly, consensus nested primers targeting the gene of penton base were also based on an aa alignment containing proteins from SAdVs only (Table 2; A. Doszpoly, personal communication). To acquire the genome fragments between the PCR-amplified parts of the adjacent genes IVa2 and *pol*, degenerate primers (designated as simian universal; ‘‘suniv’’) were designed from nt sequences of SAdVs exclusively. For a primer-walking approach, several additional consensus suniv primers were prepared (Table 3).

Table 2. PCR primers used for amplification of different gene fragments (Pantó *et al.*, 2015)

Name	Target gene	Sequence (5' → 3')	Product size[a]	Position[b]	References
HexAdB	Hexon (mastadenoviruses)	GCCGCARTGGTCYTACATGCACATC	301	17558–17809	(Kiss *et al.*, 1996)
HexAdJ		CAGCRYRCCGCGGATGTCAAART			
4431s	DNA-dependent DNA polymerase (primate AdVs)	GTNTWYGAYATHTGYGGHATGTAYGC	999	5269–6220	(Wevers *et al.*, 2010)
4428as		GAGGCTGTCCGTRTCNCCGTA			
IVa2 outfo	IVa2 (mastadenoviruses)	CCNNSNCCNGARACNGTNTTYTT	397	3998–4348	(Pantó *et al*., 2015)
IVa2 outre		GGRTTCATRTTRTGNARNACNAC			
IVa2 info		CCNCARRTNGAYATGATHCCNCC	302	4067-4319	
IVa2 inre		TTNSWNGGRAANGCRTGRAARAAYTT			
penton outfo	Penton base (SAdVs)	ACNCARACNATHAAYTTYGAYGA	363	13461–13778	(Pantó *et al*., 2015)
penton outre		GTRTANACNCCNGGCATNAC			
suniv4617F	IVa2-*pol* (SAdVs)	CARATYTGCATYTCCCASGC	1201	4307–5467	(Pantó *et al*., 2015)
suniv5821R		TACACHTACAAGCCAATCAC			

[a] Full length of the PCR product

[b] Position of the useful sequence (without the primers) according to SAdV-1 (AY771780) genome numbering

Table 3. Sequencing primers

Name	Sequenced PCR product	Sequence (5' → 3')	Position (in SAdV-1)	Reference
4466	4431s-4428as	CGTGRSHTACACHTAYAARCCAA	5470	(Pantó *et al*., 2015)
suniv5040F	IVa2-*pol*	ATCTCGATCCARCARRYYTC	4729	
suniv5040R		GARRYYTGYTGGATCGAGAT	4707	
suniv5330R		TCCAARGGMAARCTKCGCGCC	4994	

The PCRs were performed in 50 µl volume with the following ingredients (final concentration): 3 mM $MgCl_2$, 0.2 mM dNTP, 1 µM each primer, GoTaq Buffer, and 1.5 unit of the GoTaq DNA polymerase enzyme (Promega Corp., Mannheim, Germany). If applicable, SAdV-24 was used as a positive control. The PCR programs consisted of an initial denaturation step at 94°C for 5 min followed by 45 cycles (94°C, 30 s; 46°C, 60 s; 72°C, 60

s) and a final elongation step at 72°C for 3 min. The program of PCRs with suniv primers was modified to fewer cycles (35), annealing at 52°C for 30 s and elongation at 72°C for 90 s, with a final elongation of 7 min. The size, quality and amount of PCR products were checked by loading 10 µl of the completed reaction mixtures on agarose gels. Amplified fragments were purified using a Nucleospin Extract II Kit (Macherey-Nagel, Dűren, Germany) and sequenced directly on both strands using Big Dye Terminator v3.1 Cycle Sequencing Kit (Life Technologies Inc., Warrington, UK). Capillary electrophoresis was performed by a commercial service on a 3500 Series Genetic Analyzer (Life Technologies Inc., Warrington, UK).

The full genome sequence was determined from the prototype strains of six SAdVs by the classical Sanger method and capillary sequencing (SAdV-16 and -19) or NGS (SAdV-2, -8, -11 and -17). Most of the sequencing of SAdV-13 was performed by Laura Pantó presently studying at the Hokkaido University (Japan) with consecutive PCRs and Sanger capillary sequencing, only some gaps had to be filled in by custom designed PCRs. For the Sanger sequencing/primer walking, specific primers (Appendix Suppl. Table 1 and 2) based on partial sequences were designed with the use of the Primer Designer program version 2.0. For the amplification of longer fragments (>1000 bp), the Takara PrimeSTAR® Max DNA polymerase (Takara, Saint-Germain-en-Laye, France) was used according to the manufacturer recommendations. PCR products were purified on agarose gel with MEGAquick-spin Total Fragment DNA Purification Kit (iNtRON Biotechnology, Kyungki-Do, Korea). The genome fragments were sequenced with the PCR primers on both strands. For the larger fragments primer walking strategy was applied. 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). The sequences of the genes 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 manually by comparison with the earlier described SAdVs and HAdVs. The UXP sequence and location in the genome was determined by comparison with HAdV-5 UXP sequence (Tollefson *et al.*, 2007). For the NGS, the purified genomic DNA was sent to a commercial service (BGI in China or BaseClear in Leiden, The Netherlands), where paired-end sequence reads were generated using the Illumina HiSeq2500 system. FASTQ sequence reads were generated using Illumina Casava pipeline version 1.8.3. Initial quality assessment was based on data passing the Illumina Chastity filtering. Subsequently, reads containing adapters and/or PhiX control signal were removed using an in-house filtering protocol. The second quality assessment was based on the remaining reads using the FASTQC quality control tool version 0.10.0. 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 puzzled 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 gaps were filled by PCR using specific primers and sequenced by traditional (Sanger) method.

5.1.4 ATCC mixtures – end-point dilution assay

Three of the studied SAdVs (SAdV-5, -12 and -15) originating from the ATCC were not clean, but mixtures of two or more types. Consequently, if heterogeneous nt sequences were obtained with the PCRs amplifying shorter fragments, the PCR products were molecularly cloned using a CloneJET^{TM} PCR Cloning Kit (Thermo Fisher Scientific, Waltham, MA, USA) according to the manufacturers’ recommendations. Chemically competent DH5α *E. coli* cells were transformed with the ligated vector by heat shock (90 s, 42°C), and plasmids were purified by the alkaline lysis method.

The three virus mixtures were subjected to end-point dilution assay on Vero E6 cells in order to isolate at least one type from each mixture. For each mixture, a 96-well plate containing monolayer Vero E6 cells was infected with serial dilutions of a virus stock. The plates were checked every day for the CPE. After seven days, the last wells which showed CPE were marked and the plates were frozen and thawed three times, after which the cell-virus mixture was taken from the marked wells. New 96-well plates were prepared and infected with the cell-virus mixture from the previous plate. The same procedure was repeated one more time, so for each mixture three plates were used. The cell-virus mixture from the third plate, which still showed CPE, was subjected to nested PCR and sequencing with penton base targeting primers to determine if the virus type is clean or not. In case it was clean (i.e., the sequence was uniform), the virus was produced in larger amounts (in the same way as SAdV-2, -8, -11 and -17) and was sent to a commercial NGS service (BGI, China). In case the gained “isolate” was still not clean, it was sent to a partner company Batavia (Leiden, The Netherlands) where it was subjected to plaque purification by Hungarian secondee Mónika Ballmann.

5.1.5 Molecular cloning of fibre-1 knobs

Knobs of the fibre-1 genes of several SAdVs (SAdV-1, -7, -11 and -19) which contain two fibre genes were cloned into the pQE-30 Xa plasmid (kind gift from Niklas Arnberg, Umea, Sweden) in order to express the knobs and study the cellular receptors they can attach to. Primers with restriction enzyme recognition sequences were designed for each of the knobs (Appendix Suppl. Table 3), and the PCR was performed with the viral DNAs in order to

obtain the fragments for cloning. The REDTaq DNA polymerase (Sigma Aldrich, St. Louis, MO, USA) enzyme was found to be optimal for the amplification of the fragments according to the manufacturers' recommendations. PCR products were cleaned by gel-electrophoresis and were subjected to restriction endonuclease digestion using appropriate enzymes (Appendix Suppl. Table 3) according to the manufacturers' recommendation. At the same time, the pQE-30 Xa plasmid was digested with the same combination of restriction enzymes. Digested products were cleaned again by gel-electrophoresis and were subjected to ligation reactions. Ligation was performed in 20 µl end-volume using T4 DNA ligase enzyme with buffer (Thermo Fisher Scientific, Waltham, MA, USA), 30 min at room temperature. Chemically competent *E. coli* TOP10F strain (Thermo Fisher Scientific, Waltham, MA, USA) was transformed with 5 µl of the ligated mixture by heat-shock at 42°C for 90 s. Transformed bacteria were plated on standard selective LB agar containing ampicillin (100 µg/ml) and incubated overnight at 37°C. Colonies which appeared were subjected to PCR reaction in order to check for the presence of the plasmid. Positive colonies were propagated in selective LB broth and mini-preparations of the plasmid DNA were obtained with Nucleospin® Plasmid (Macherey-Nagel, Dűren, Germany). Direct sequencing was used for the confirmation of fibre knob presence in the plasmid. Niklas Arnberg's group (Umea University, Sweden) expressed the knobs and produced them in larger amounts for future experiments.

5.1.6 Fluorescence-activated cell sorting

In collaboration with Niklas Arnberg's group (Umea University, Sweden), a few receptor studies were made with the fibre-1 knobs using A549 (human alveolar basal epithelial adenocarcinoma) cells. To investigate the ability of the studied fibre knobs to use the sialic acid-containing glycans as receptors, cells were treated with neuraminidase (cleaves poly-sialic acids) from *V. cholerae.* The A549 cells were detached, centrifuged for 5 min at room temperature at 5200 × g, resuspended with binding buffer (medium without FBS) and split into two falcon tubes. In one falcon tube, neuraminidase was added (20 mU/ml) and the cells were incubated for 1 h at 37°C on rocking table. Afterwards the cells were plated in V-bottomed 96-well plates and centrifuged at 4°C for 4 min at 470 × g. The cells were washed with binding buffer, and fibre knobs diluted in binding buffer added, followed by incubation on ice for 1 h on rocking table. Cells were washed with ice cold FACS buffer (PBS, 2% FCS, 0.01% NaN_3), and resuspended with primary RGS-His, mouse IgG antibody in FACS buffer (1:200), followed by incubation on ice and rocking table for 30 min. Cells were washed with FACS buffer, and resuspended with secondary Polyclonal Rabbit Anti-mouse Immunoglobulin/FITC antibody in FACS buffer (1:40), followed by incubation on ice for 30

min on rocking table, and then washed again with the FACS buffer. Cells were transferred to FACS tubes, and the binding was measured on FACS LSR II machine (Becton Dickinson).

5.2 Screening for new adenoviruses

5.2.1 Samples for screening

A total of 138 fecal or organ samples were screened for AdVs (Table 4): 10 samples from apes, 11 samples from OWMs, 19 samples from NWMs, and 98 samples from prosimians. Samples originated from several different locations: Hungarian zoological garden (zoo) No. 1 (18 samples), Hungarian zoo No. 2 (5 samples), Hungarian zoo No. 3 (1 sample), a Croatian zoo (15 samples), a French zoo (15 samples), the animal collection of a French university (12 samples), and Madagascar (72 samples from several areas: Nosy Be Island, Nosy Komba (Nosy Ambariovato) Island, Ankarana Reserve, Ankarafantsika Nature Reserve, Andasibe, Kirindy Mitea National Park, and Ranomafana National Park). Fecal samples were obtained from healthy animals with no sign of AdV infection, and the organ samples were obtained from animals which died for reasons not associated with AdV infection.

5.2.2 DNA extraction

To find novel AdVs, DNA was extracted from fecal samples with the E.Z.N.A.® Stool DNA Kit (OMEGA bio-tek) according to the manufacturers' instructions, and from the organ samples as described earlier (Kovács & Benkő, 2009).

5.2.3 PCR and DNA sequencing

Dream Taq® DNA Polymerase (Thermo Fisher Scientific, Waltham, MA, USA) and degenerate primers targeting IVa2 gene (Table 2) of mastadenoviruses were used for nested PCR (Wellehan *et al.*, 2004). Positive PCR products were purified and sequenced by using Big Dye Terminator v3.1 Cycle Sequencing Kit (Life Technologies Inc., Warrington, UK) according to the manufacturers' instructions. Capillary electrophoresis was performed by a commercial service on a 3500 Series Genetic Analyzer (Life Technologies Inc., Warrington, UK).

5.2.4 Cell culture methods

To isolate newly detected AdVs, samples positive by PCR (with primers targeting the IVa2 gene) were prepared as follows: 200 mg of a positive fecal sample was mixed in 1.5 ml

Table 4. Samples screened for adenoviruses by nested PCR

Primate group	Species (English name)	Species (Latin name)	Origin of sample*	# of samples
Apes	Sumatran orangutan	*Pongo abelii*	1	7
	Lar gibbon	*Hylobates lar*	2	1
	Siamang	*Symphalangus syndactylus*	1	2
OWMs	Eastern black-and-white colobus	*Colobus guereza*	2	1
	Northern plains gray langur	*Semnopithecus entellus*	2	1
	Diana monkey	*Cercopithecus diana*	2	1
	Hamadryas baboon	*Papio hamadryas*	1	1
	Javan langur	*Trachypithecus auratus*	1	2
	Golden-bellied mangabey	*Cercocebus chrysogaster*	1	1
	Mandrill	*Mandrillus sphinx*	1	1
	Moustached guenon	*Cercopithecus cephus*	3	2
	Barbary macaque	*Macaca sylvanus*	3	1
NWMs	Gray-bellied night monkey	*Aotus lemurinus griseimembra*	3	2
	Pygmy marmoset	*Cebuella pygmaea*	2, 4	2
	Red-bellied tamarin	*Saguinus labiatus*	2	1
	Emperor tamarin	*Saguinus imperator*	2, 3	2
	Tufted capuchin	*Cebus apella*	2, 4	2
	Common Squirrel monkey	*Saimiri sciureus*	1, 4	4
	Cotton-top tamarin	*Saguinus oedipus*	4	1
	Golden-headed lion tamarin	*Leontopithecus chrysomelas*	4	1
	Red-handed tamarin	*Saguinus midas*	3, 7	2
	Black howler	*Alouatta caraya*	3	1
	Three-striped night monkey	*Aotus trivirgatus*	2	1
Prosimians	White-headed lemur	*Eulemur albifrons*	2	1
	Greater slow loris	*Nycticebus coucang*	2	1
	Brown greater galago	*Otolemur crassicaudatus*	2	1
	Ring-tailed lemur	*Lemur catta*	2, 3	4
	Black-and-white ruffed lemur	*Varecia variegata*	2	1
	Red lemur	*Eulemur rufus*	5	20
	Red-bellied lemur	*Eulemur rubriventer*	5	13
	Sanford's brown lemur	*Eulemur sanfordi*	5	1
	Crowned lemur	*Eulemur coronatus*	3, 5	5
	Common brown lemur	*Eulemur fulvus*	5	10
	Mongoose lemur	*Eulemur mongoz*	3, 5	6
	Black lemur	*Eulemur macaco*	5	9
	Unknown lemur species	-	5	6
	Indri	*Indri indri*	5	4
	Red-fronted lemur	*Eulemur rufifrons*	1	1
	Mouse lemur	*Microcebus murinus*	3, 6	13
	Red ruffed lemur	*Varecia rubra*	3	1
	Eastern lesser bamboo lemur	*Hapalemur griseus*	3	1

*1 – Hungarian zoo 1; 2 – Croatian zoo; 3 – French zoo; 4 – Hungarian zoo 2; 5 – Madagascar; 6- French University; 7 – Hungarian zoo 3

eppendorf tube with 250 ml of DMEM (for mouse rectum carcinoma cell line (cmt93), human lung adenocarcinoma epithelial cell line (A549) and human embryonic kidney cell line (HEK293)) or RPMI (for mouse embryonic fibroblast cell line (3T6), African green monkey kidney cell line (Vero E6), Chinese hamster ovary cell line (CHO-K1) and mouse lemur fibroblast cell line (MFC)) medium supplemented with antibiotics; 200 mg of an organ sample was placed in 2-ml tube containing metal ball and 250 ml of medium. Samples were vortexed

for 10 min at 50 Hz with TissueLyser LT (Qiagen, Hilden, Germany), followed by three cycles of freezing and thawing. Supernatant, gained after centrifugation for 2 minutes at 18,000 × g, was placed in a new 1.5-ml tube and centrifuged again. Fifty µl of the supernatant was used for infection of the cells subcultivated a day before and plated in a 24-wells plate. Cells were incubated at 37°C, and examined every 24 hours for the CPE for maximum of 10 days. The plates were frozen and thawed three times, suspension of the infected cells collected and used for blind passage if the CPE did not occur. Again, 50 µl of the supernatant was used, and blind passages were repeated for up to seven times, after which the suspension was subjected to PCR with IVa2 primers.

5.2.5 Full genome sequencing of red-handed tamarin adenovirus 1

One of the NWM AdVs (red-handed tamarin AdV-1; RHTAdV-1), detected earlier by Laura Pantó, was subjected to sequencing of longer fragments of the genome. Degenerate primers targeting IVa2 (Table 2), *pol* (Wellehan *et al.*, 2004), penton-base (outer primers in Table 2), and hexon genes (Table 2) were used for gaining the first sequences from RHTAdV-1. For the penton base gene, additional inner primers were used: 5'-TCRAAYTTNACNCCDATRTC-3' (reverse) and 5'-CAYACNAAYATGCCNAAYGTNGA-3' (forward). Gained gene fragments were connected with specific primers (Appendix Suppl. Table 4).

5.3 Bioinformatics

The identities of the newly gained sequences were examined using the BLASTX program at the website of the NCBI. Primate AdV sequence alignments were prepared with the ClustalW program of the MEGA6 package (Tamura *et al.*, 2013). Phylogenetic calculations, based on aa alignments of partial (IVa2 gene for NWM and prosimian AdVs, *pol* for ATCC strains) or full gene (hexon and fibre-1 for ATCC strains) sequences available from every studied virus, were performed using the ProtDist and PhyML algorithms provided by the Mobyle portal of the Pasteur Institute (Paris), and the ProtTest program (Darriba *et al.*, 2011). The ProtDist analyses were run with the JTT substitution model, followed by Fitch-Margoliash analysis, applying the global rearrangements option. PhyML calculations (Guindon & Gascuel, 2003) were based on a user tree obtained using ProtDist and a substitution model determined using ProtTest. Phylogenetic analysis, based on nt alignment of partial penton base gene sequences (for ATCC strains) available from every studied virus, were performed using PhyML in TOPALi v2 platform with model proposed by the Model Selection module of TOPALi. Bootstrap analysis with 100 sampling replicates was applied for every tree. The trees were visualized using the MEGA6 program. Tree shrew AdV-1 (TSAdV-1) served as an

outgroup on all the trees except the fibre-1 based tree, on which the fowl AdV-1 (FAdV-1) was used. 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).

6. Results

6.1 Strains from the ATCC

6.1.1 Production of virions and viral DNA for next-generation sequencing

We were able to grow 12 of the 14 non-sequenced SAdVs on the Vero E6 cells, eight of them were produced in large amounts and their DNA was extracted (Table 5) for NGS.

Table 5. Summary of newly studied ATCC strains

Virus	Growth on Vero cells	DNA concentration	Virus gained	Full genome sequenced
SAdV-2	+	493 ng/µl	adenovirus	NGS
SAdV-4	+	302 ng/µl	parvovirus	-
SAdV-5	+	2,1 µg/µl	adenovirus mix	NGS (one type)
SAdV-8	+	165 ng/µl	adenovirus	NGS
SAdV-9	+	-	-	
SAdV-10	-			
SAdV-11	+	1 µg/µl	adenovirus	NGS
SAdV-12	+	1,6 µg/µl	adenovirus mix	NGS (mixture)
SAdV-13	-			Sanger (by colleague)
SAdV-14	+	-	-	
SAdV-15	+	1,6 µg/µl	adenovirus mix	NGS (mixture)
SAdV-16	+			Sanger
SAdV-17	+	90 ng/µl	adeno+parvovirus	NGS
SAdV-19	+			Sanger

After the ultracentrifugation and DNA extraction, agarose gel electrophoresis indicated that in one case (SAdV-4) parvovirus was in majority (confirmed by partial sequencing), whereas in one case (SAdV-17) equal amounts of both AdV and parvovirus were found (Figure 4). Pure AdV bands were seen in case of the SAdV-2, -5, -8, -11, -12 and -15 (Figure 4). Concentration of extracted DNA varied from strain to strain (Table 5), but in all cases was sufficient for the NGS purpose (>50ng/µl). Extracted DNAs were examined on the agarose gel in several dilutions: 50-fold, 10-fold and undiluted DNA. In Figure 4 we show only one dilution of each virus. The sizes of the bands are comparable to the concentrations of the DNA (Table 5). SAdV-5 had the most concentrated DNA, which can be easily observed if the 10-fold dilutions are compared (Figure 4). Equal amounts of DNA were gained in the case of SAdV-12 and -15, from which we present the undiluted and the 50-fold diluted DNA bands in Figure 4, respectively.

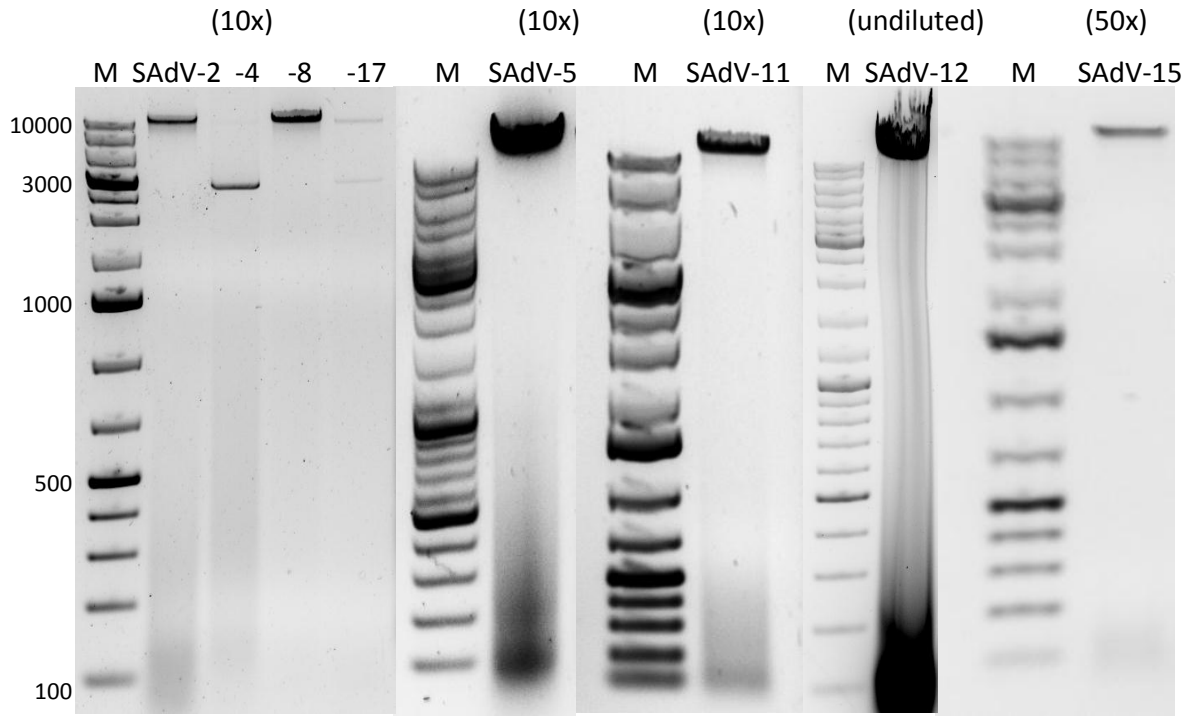


Figure 4. Results of the DNA extraction of SAdVs propagated on Vero E6 cells. Bands appearing close to the 10,000 bp label correspond to AdV, whereas those appearing close to 3000 bp label correspond to parvovirus. Different dilutions (undiluted, 10-fold and 50-fold) of the DNAs are shown for different viruses.

Several SAdVs (SAdV-1, -2 and -11) were propagated on A549 cells to have larger amounts of them for the possible receptor-binding studies (in collaborating laboratory in Umea, Sweden), and the summary of the virion production is shown in Table 6. The refractory index of AdVs is 1.3662. The bands gained after the CsCl density-gradient centrifugation are shown in Figure 5. For SAdV-11, there was no AdV band after the CsCl density-gradient centrifugation. The concentration of SAdV-1 and -2 virions was much lower than that of HAdV-5 (Table 6, Figure 5).

Table 6. CsCl density-gradient centrifugation results

Virus	CsCl density-gradient centrifugation success	Refractory index	Virion concentration
HAdV-5	+	3.661	898.8 µg/ml
SAdV-1	+	3.660	61.6 µg/ml
SAdV-2	+	3.661	137.2 µg/ml
SAdV-11	-	-	-

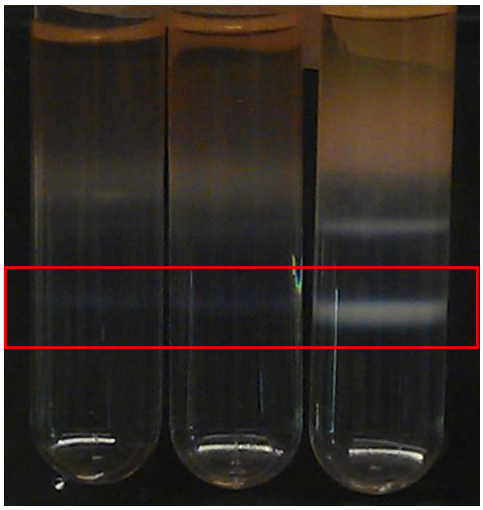


Figure 5. Bands of SAdV-1, -2 and HAdV-5 visible after CsCl density-gradient centrifugation. Tubes from left to the right: SAdV-1, SAdV-2, HAdV-5.

6.1.2 Partial genome sequencing and ATCC mixtures

From four partial genes which were used for phylogeny analysis (IVa2, *pol*, penton base and hexon) of 14 non-sequenced SAdVs (Pantó *et al.*, 2015), here we present the phylogeny reconstructions, performed with the aa sequences, deduced from the *pol* (JTT model with the invariable sites and gamma distribution options used; Figure 6).

SAdV-2, -11, -12 and -15 clustered clearly with the SAdVs that had been proposed to form species HAdV-G (Jones *et al.*, 2007). Two OWM ATCC strains (SAdV-5 and -8) together with the earlier described lineage involving SAdV-49, -50 (Roy *et al.*, 2009), nine AdVs isolated from rhesus macaques (Roy *et al.*, 2012) and one AdV isolated from olive baboon (BaAdV-1; Chiu *et al.*, 2013) appeared on the tree as a monophyletic clade corresponding to the recently accepted species SAdV-B (Podgorski & Harrach, 2015). Several SAdVs, SAdV-4, -9, -10 and -14, clustered with members of the species SAdV-A. SAdV-13 alone seemed to represent an independent lineage, candidate species *Simian mastadenovirus D* (SAdV-D). A sister clade was formed by the closest virus, a novel AdV strain (23336) reported from rhesus macaque recently (Malouli *et al.*, 2014), and the distance between these two viruses seems to warrant a separate species classification for both. The later species could be named *Simian mastadenovirus H* (SAdV-H). Similarly, SAdV-20 also formed alone an independent branch most closely diverging from the clade species SAdV-A. In this case, the establishment of a novel species *Simian mastadenovirus G* (SAdV-G) seems to be justified. SAdV-16 appeared closest to the clade of species SAdV-B, but as a long distinct branch, it is proposed to represent an independent species: *Simian mastadenovirus E* (SAdV-E). SAdV-17 and -18 were sister clade closest to the HAdV-F

species members, but well separated from them as a species *Simian mastadenovirus F* (SAdV-F). SAdV-19 appeared on a branch together with other baboon AdVs (Chiu *et al.*, 2013) of the SAdV-C species, recently accepted by the ICTV. Considering the tree topology as well as other features of the examined OWM AdVs (Table 7) the existence of minimum five distinct lineages, most probably meriting the species-level demarcation, was revealed besides the already established HAdV and SAdV species (Table 7).

Direct sequencing of the PCR products obtained from the penton base gene showed that three viruses, namely SAdV-5, -12 and -15, were mixtures (Pantó *et al.*, 2015). Molecular cloning of the amplicons resulted in the separation of two different sequences from each mixture. The sequences from samples of SAdV-12 and SAdV-15 indicated the presence of different viruses in each strain, yet all of the four viruses appeared to be members of the species HAdV-G (Figure 7). The prototype strain of SAdV-5 was also found to contain two AdVs, but only one of them clustered with members of HAdV-G, whereas the other clustered with the members of the recently established species SAdV-B (Podgorski & Harrach, 2015). PCR products obtained from the penton base gene after the end-point dilution assay on the Vero E6 cells showed that one type from each mixture was obtained in a clean form. The cleaned types were produced in larger volumes on the Vero E6 cells and the DNA was extracted for the NGS purpose. Unfortunately, the NGS showed that SAdV-12 and -15 are still not clean types, but they contain also a small amount of the second type. Consequently, we were unable to separate the DNA sequence of one type only. In case of SAdV-5, we got partial sequence (about half of the genome) of a clean type, but it showed to be exactly the same as SAdV strain A1335 (Roy *et al.*, 2012) submitted earlier to the GenBank (acc. number JN880456), belonging to the species SAdV-B. Therefore, we did not sequence the rest of the genome of that type. All the three types were sent to the collaborating company (Batavia) in Leiden (Netherlands) which performed the plaque purification assay with SAdV-12, and both types of SAdV-12 were isolated and are currently being sequenced. Plaque purification of SAdV-5 was unsuccessful (again the same type was gained, belonging to the SAdV-B species), and SAdV-15 has not been plaque-purified yet.

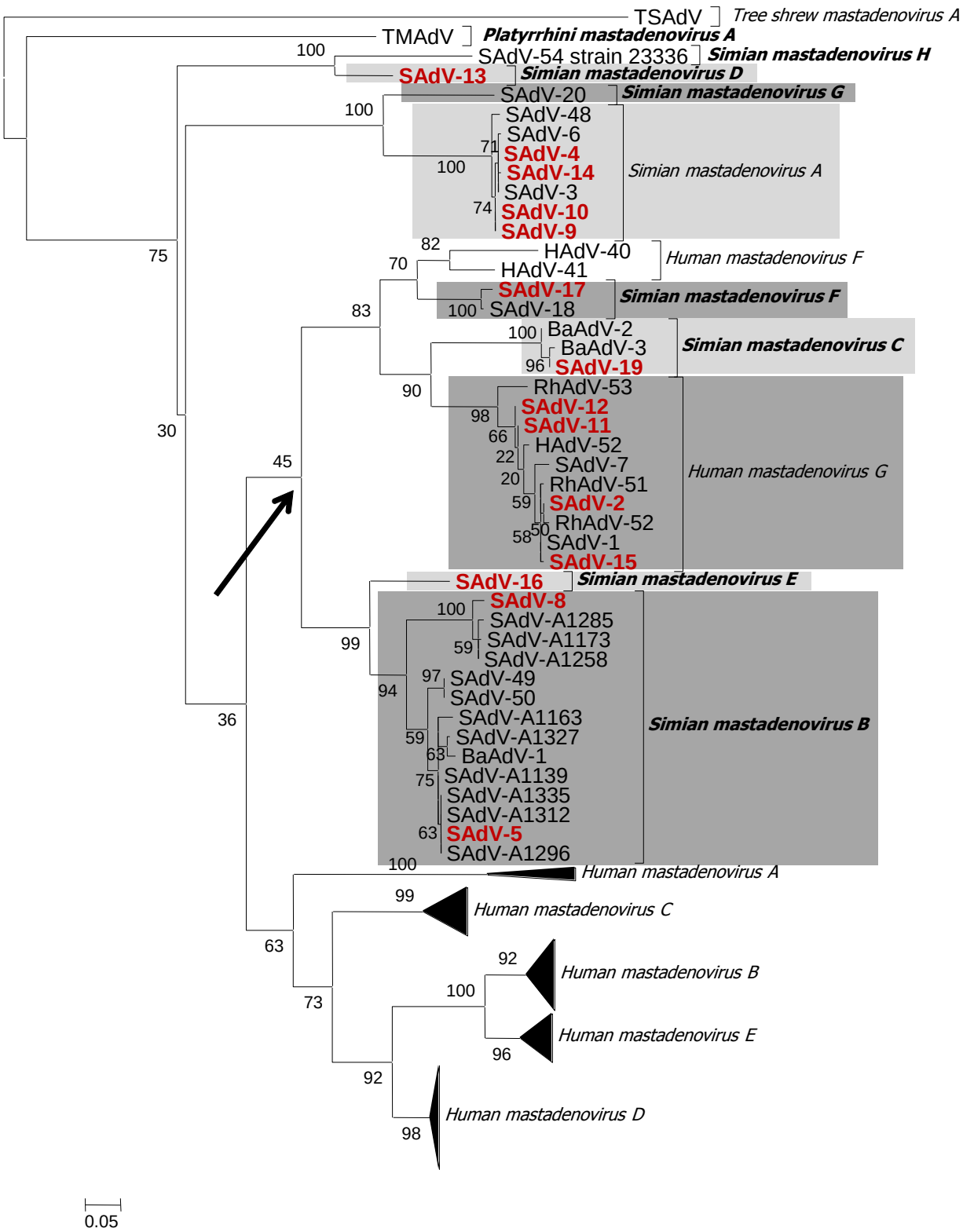


Figure 6. Phylogeny reconstruction based on partial aa sequences from the DNA-dependent DNA polymerase (*pol*) gene. Newly studied SAdVs are highlighted with bold red, and newly established or proposed species with bold black. Black arrow indicates the node which separates the group of AdVs with two or three fibre genes (except SAdV-18). Abbreviations: HAdV, human AdV; BaAdV, baboon AdV; TSAdV, tree shrew AdV; TMAdV, titi monkey AdV.

Table 7. Proposal of taxonomy of OWM AdVs based on phylogenetic analyses and other criteria (edited after Pantó *et al.*, 2015)

Adenovirus	Existing/ Proposed species	Host species	HAG	Fibre genes	GC content[b]
SAdV-1	HAdV-G	Crab-eating macaque	III	2	56.4 (55.2)
SAdV-2		Rhesus macaque		3	58.0 (54.7)
SAdV-7				2	56.8 (56.3)
SAdV-11					57.9 (55.0)
SAdV-12[a]					58.1
SAdV-15[a]				2[c]	57.5
SAdV-51/RhAdV-51				2	57.8 (55.0)
SAdV-52/RhAdV-52					57.1 (54.9)
SAdV-53/RhAdV-53					58.0 (54.6)
HAdV-52		Human			57 (55.1)
SAdV-3	SAdV-A	Rhesus macaque	II	1	58.1 (55.3)
SAdV-4[a]					59.6
SAdV-6		Macaque		1	58.3 (55.8)
SAdV-9[a]					59.4
SAdV-10[a]					59.8
SAdV-14[a]		Rhesus macaque			59.6
SAdV-48		Crab-eating macaque		1	57 (54.4)
SAdV-5[a]	SAdV-B	Rhesus macaque	III		65.3
SAdV-8		Crab-eating macaque		2	63.1 (60.3)
SAdV-49					65.6 (62.8)
SAdV-50					65.4 (62.6)
BaAdV-1		Olive baboon			65.4 (62.7)
A1139[d]		Rhesus macaque			65.7 (62.6)
A1163[d]					65.1(62.0)
A1173[d]					63.0 (61.1)
A1258[d]					63.0 (60.1)
A1285[d]					62.7 (61.0)
A1296[d]					65.5 (62.6)
A1312[d]					65.5 (62.6)
A1327[d]					65.5 (62.9)
A1335[d]					65.6 (62.8)
SAdV-19	SAdV-C	Yellow baboon		2	53.7 (52.2)
BaAdV-2		Olive baboon		2	52.3 (52.6)
BaAdV-3					52.3 (52.3)
SAdV-13	SAdV-D	Macaque	I	1	50.0 (49.9)
SAdV-16	SAdV-E	Grivet	IV	2	63.6 (57.9)
SAdV-17	SAdV-F	Grivet		3	64.5 (58.8)
SAdV-18				1	63.4 (61.4)
SAdV-20	SAdV-G	Grivet		1	47.1 (47.8)
SAdV-54 strain 23336[d]	SAdV-H	Rhesus macaque		1	47.4 (46.7)

[a] full genome sequence not available

[b] GC content of the amplified fragments or corresponding region of the full genome; GC content of the full genome is shown in brackets ( )

[c] unpublished data

[d] strain name

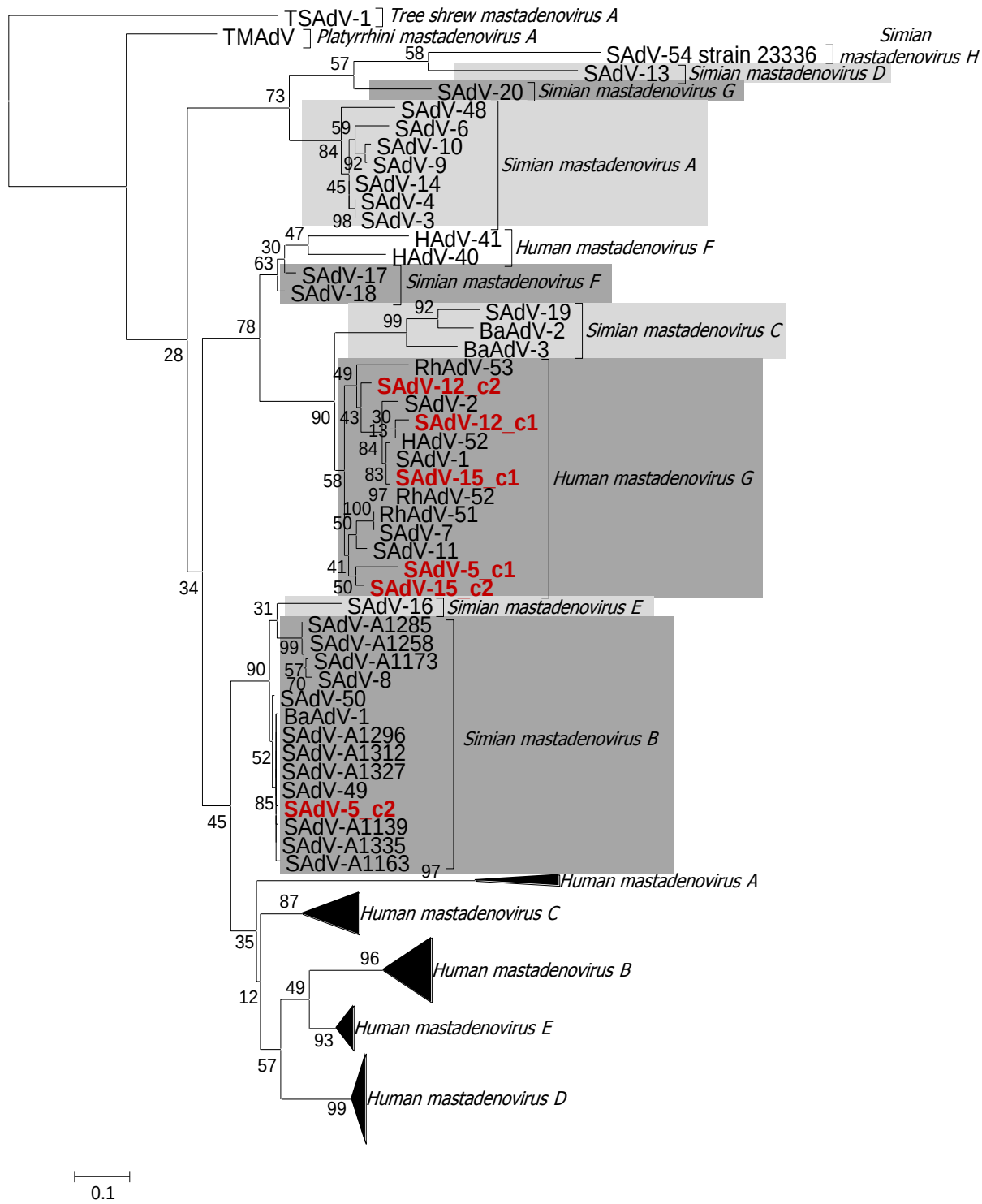


Figure 7. Phylogeny reconstruction based on the partial nucleotide sequence from the penton base gene (TrNef model with invariable sites and gamma distribution options). The three strains containing at least two different types of AdVs are highlighted with bold red.

6.1.3 Full genome sequencing

Four SAdVs (SAdV-2, -8, -11 and -17) were successfully sequenced by NGS, and two (SAdV-16 and -19) with the traditional Sanger sequencing (Podgorski *et al*., 2016). The main characteristics of the new genome sequences are summarised in Table 8, 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 to.

Table 8. Genome characteristics of the simian adenovirus (SAdV) types studied in this work

Virus	Proposed species	Genome length (bp)	ITRs length (bp)	ITRs length range in the proposed species (bp)	G+C content (%)	G+C range in the proposed species (%)	Acc. number
SAdV-2	HAdV-G	35,481	81	60-133	54.7	54.6-56.3	**-**
SAdV-8	SAdV-B	35,685	188	166-220	60.3	60.1-62.9	**KP329561**
SAdV-11	HAdV-G	34,510	71	60-133	55.0	54.6-56.3	**KP329562**
SAdV-16	SAdV-E	35,159	181	-	57.9	-	**KP329564**
SAdV-17	SAdV-F	35,670	59	59-180	58.8	58.8-61.4	**KP329566**[a]
SAdV-19	SAdV-C	34,604	127	87-127	52.2	52.2-52.6	**KP329565**

[a]unreleased

Genomic assemblies revealed that all six genomes contain 36 or 37 putative coding regions characteristic for mastadenoviruses. These included two (SAdV-8, -11, -16, -19) or three (SAdV-2, -17) genes, of different lengths, predicted to code for 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 VA-RNA gene. SAdV-2 is an exception from all the studied SAdVs as it does not contain E1A 19K gene, but has an extra copy of ITR and E4 ORF1 gene in this region instead (Figure 8). SAdV-8 genome is shown in Figure 9 as a representative of other studied SAdVs.

The predicted UXP sequences of the studied SAdVs were compared to their counterparts in HAdV-C types 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 earlier by others, namely those of chimpanzee (SAdV-34), gorilla (SAdV-43) and bonobo (SAdV-44) AdV (Roy *et al.*, 2009) were identified as well by comparison of the sequences to members of HAdV-C. Main characteristics of the UXPs are shown in Table 9. UXP sequences were aligned to compare the degree of conservation of the three exons in different AdVs (Figure 10). Derived from the presence of splice donor and acceptor sites, the putative positions of the three UXP exons in the genomes, including that

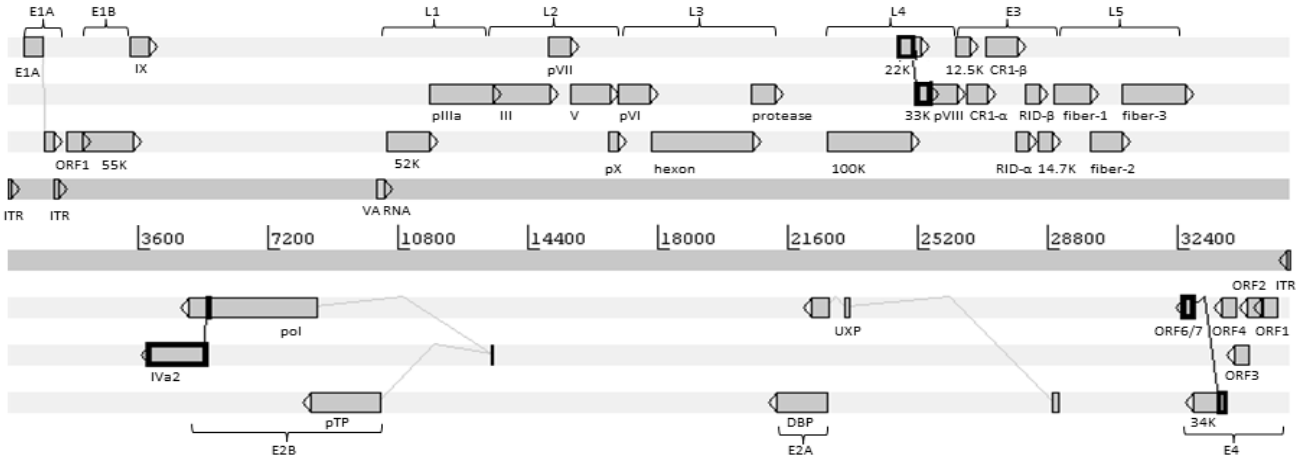


Figure 8. Annotated genome map of SAdV-2. Spliced genes sharing some parts with another ORF are highlighted with thick line. Note the missing E1B 19K and the duplicated ORF1 in its place, and the three fibre genes.

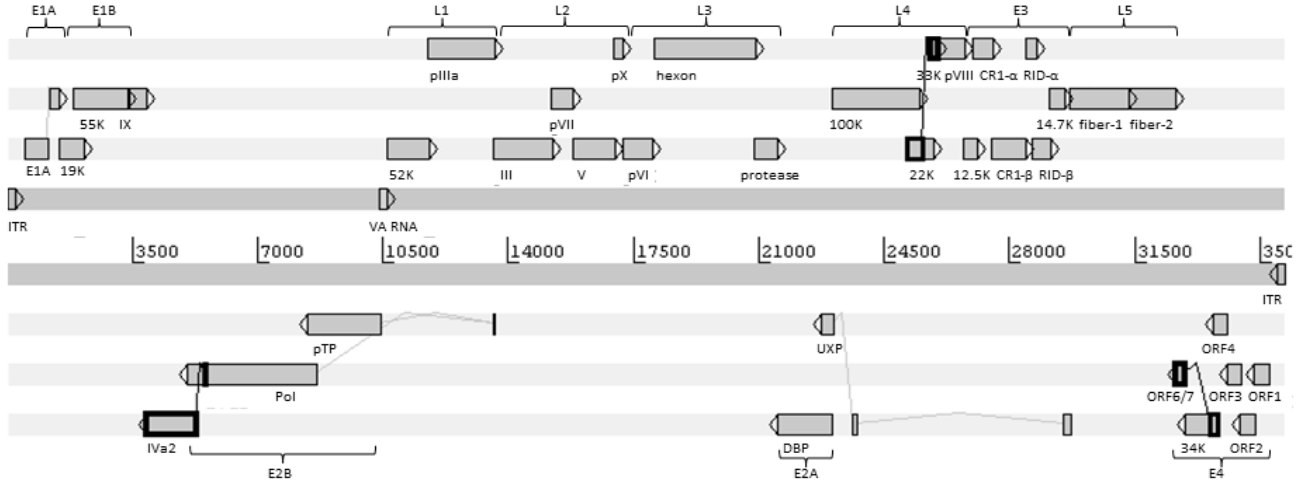


Figure 9. Annotated genome map of SAdV-8. Spliced genes sharing some parts with another ORF are highlighted with thick line.

Table 9. Main characteristics of the gene of the U exon protein (UXP) identified in four HAdVs (Tollefson *et al*., 2007) and in ten SAdV types (Podgorski *et al*., 2016)

Virus (species)	1st exon position	2nd exon position	3rd exon position
	(the number of aa coded by each exon is in brackets)		
SAdV-8 (SAdV-B)	29531-29691 (54)	23643-23719 (25)	22766-23040 (91)
SAdV-19 (SAdV-C)	28507-28673 (56)	22855-22931 (25)	21955-22349 (131)
SAdV-13 (SAdV-D)	29111-29274 (55)	23179-23252 (25)	22296-22588 (96)
SAdV-16 (SAdV-E)	29161-29321 (54)	23442-23518 (25)	22590-22870 (93)
SAdV-17 (SAdV-F)	28849-29006 (53)	23375-23451 (25)	22499-22833 (111)
SAdV-2 (HAdV-G)	28916-29073 (53)	23200-23276 (25)	22256-22707 (150)
SAdV-11 (HAdV-G)	28624-28781 (53)	22905-22981 (25)	21970-22388 (139)
HAdV-1 (HAdV-C)	30927-31090 (55)	24725-24801 (25)	23704-24098 (131)
HAdV-2 (HAdV-C)	30856-31019 (55)	24715-24791 (25)	23676-24088 (137)
HAdV-5 (HAdV-C)	30868-31031 (55)	24668-24744 (25)	23629-24041 (137)
HAdV-6 (HAdV-C)	30835-30998 (55)	24700-24776 (25)	23661-24073 (137)
SAdV-34	32643-32803 (54)	25085-25161 (25)	23701-24380 (226)
SAdV-43	32048-32211 (55)	24955-25031 (25)	23811-24274 (154)
SAdV-44	32686-32846 (54)	25127-25203 (25)	23761-24425 (221)

of SAdV-13 sequenced by a colleague (unpublished), are summarised in Figure 11. The majority of the deduced protein sequences of SAdV-2, -8, -11, -17 and -19, was found very similar to (sharing >95% identity with) their counterparts in other members of the same species (Table 8). In the SAdV-2 genome, there are a few exceptions to this, the UXP (81%), CR1-α (64%), CR1-β (50%), RID-α (93%), RID-β (88%), and all the three fibre genes (45-65%). In the SAdV-8 genome, there is one exception to this, the UXP with only 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-13 and SAdV-16 did not exhibit particular similarity to any known AdVs. Exceptions to this lack of high similarity in SAdV-13 were seen in the IX and 34K genes, whereas in SAdV-16 in hexon, penton base and pVIII genes. SAdV-17 shared lower identity with SAdV-18 in hexon (90%), DBP (91%), ORF6/7 (87%), ORF4 (87%), and in almost the whole E3 region, which was comparable to species HAdV-G and HAdV-F members only, with 37-80% identity. The three fibre genes of the SAdV-17 showed very different identity values: fibre-1 46% to HAdV-F members, fibre-2 62% to SAdV-C members, and fibre-3 99% to SAdV-18. In the SAdV-19 genome, which was comparable to other baboon AdVs from SAdV-C, there are several exceptions: the hexon (89%), the proteins coded by genes in the E3 region (in some cases the identity found as low as 64%), UXP (66%), fibre-1 (only 27% identity to SAdV-C members, while 46% to HAdV-G members), fibre-2 (66% identity), and ORF6/7 (85% identity).

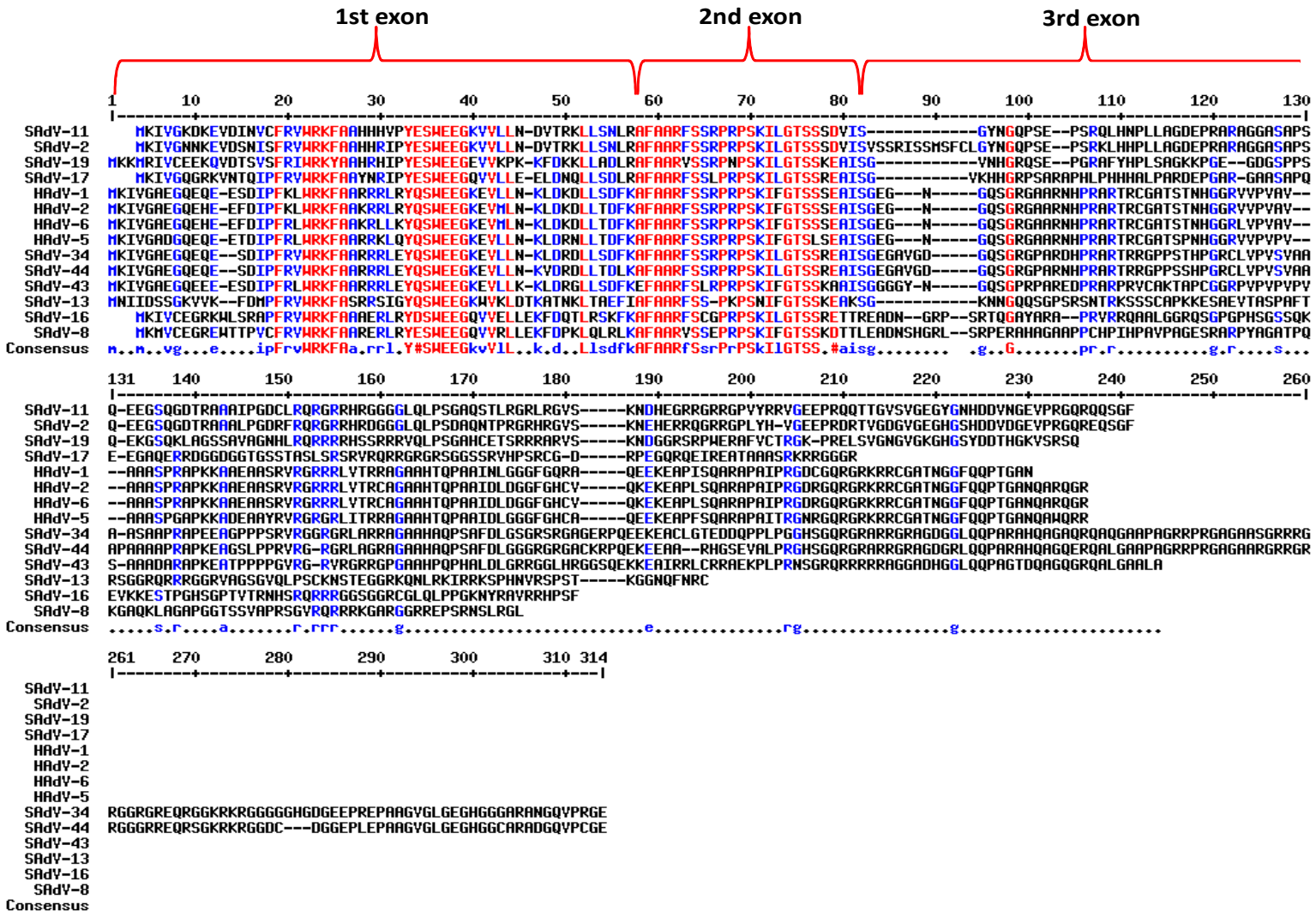
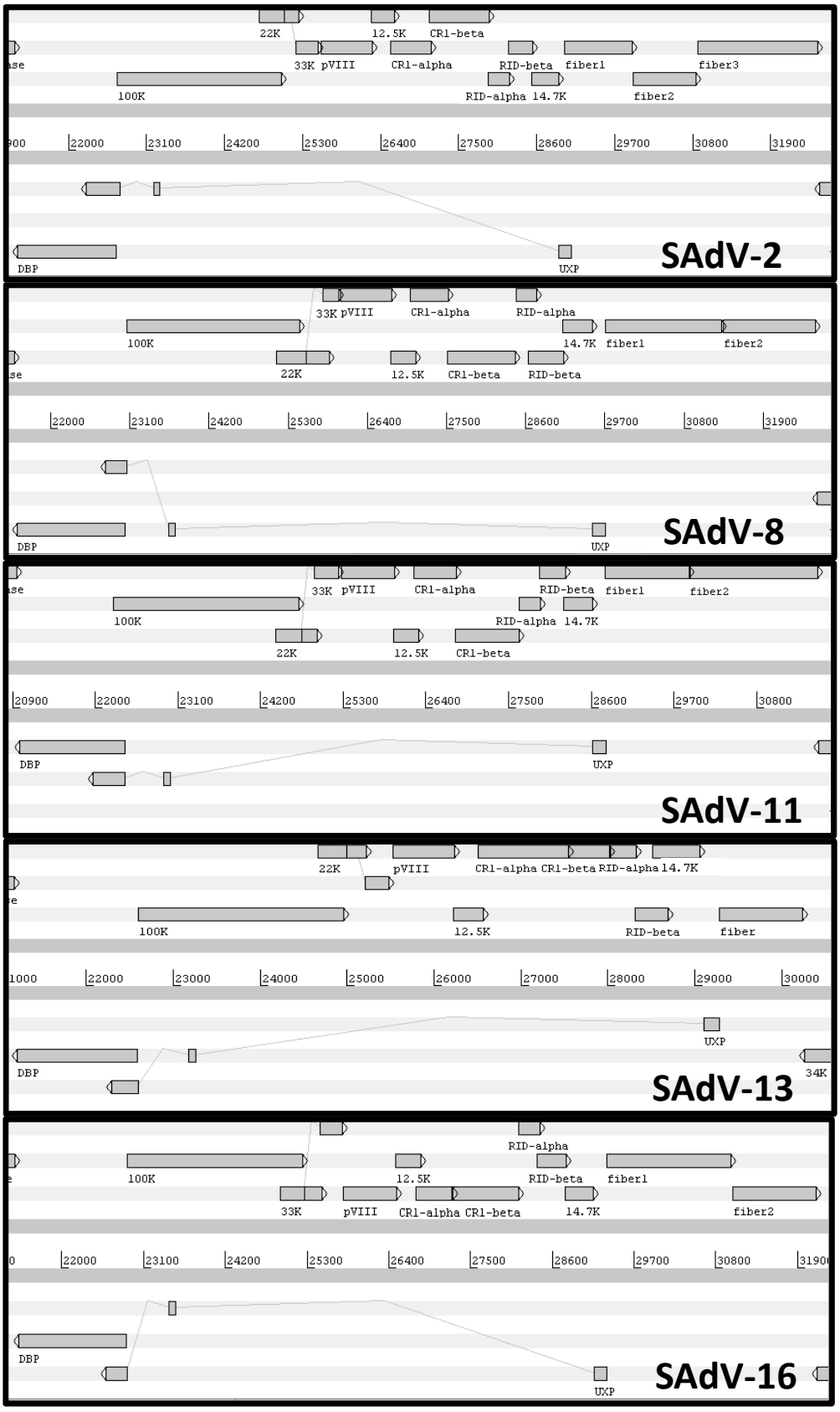


Figure 10. Alignment of the aa sequences of UXP identified in four HAdVs previously and in ten SAdV types in this study.


22K
12.5K
CR1-beta
33K pVIII
CR1-alpha
RID-beta
fiber1
fiber3
100K
RID-alpha 14.7K
fiber2
22000 23100 24200 25300 26400 27500 28600 29700 30800 31900
DBP
UXP
SAdV-2
33K pVIII
CR1-alpha
RID-alpha
100K
14.7K
fiber1
fiber2
22K
12.5K
CR1-beta
RID-beta
22000 23100 24200 25300 26400 27500 28600 29700 30800 31900
DBP
UXP
SAdV-8
33K pVIII
CR1-alpha
RID-beta
fiber1
fiber2
100K
RID-alpha 14.7K
22K
12.5K
CR1-beta
20900 22000 23100 24200 25300 26400 27500 28600 29700 30800
DBP
UXP
SAdV-11
22K
pVIII
CR1-alpha CR1-beta
RID-alpha 14.7K
100K
12.5K
RID-beta
fiber
1000 22000 23000 24000 25000 26000 27000 28000 29000 30000
UXP
DBP
34K
SAdV-13
RID-alpha
100K
12.5K
RID-beta
fiber1
33K
pVIII
CR1-alpha CR1-beta
14.7K
fiber2
22000 23100 24200 25300 26400 27500 28600 29700 30800
DBP
UXP
SAdV-16

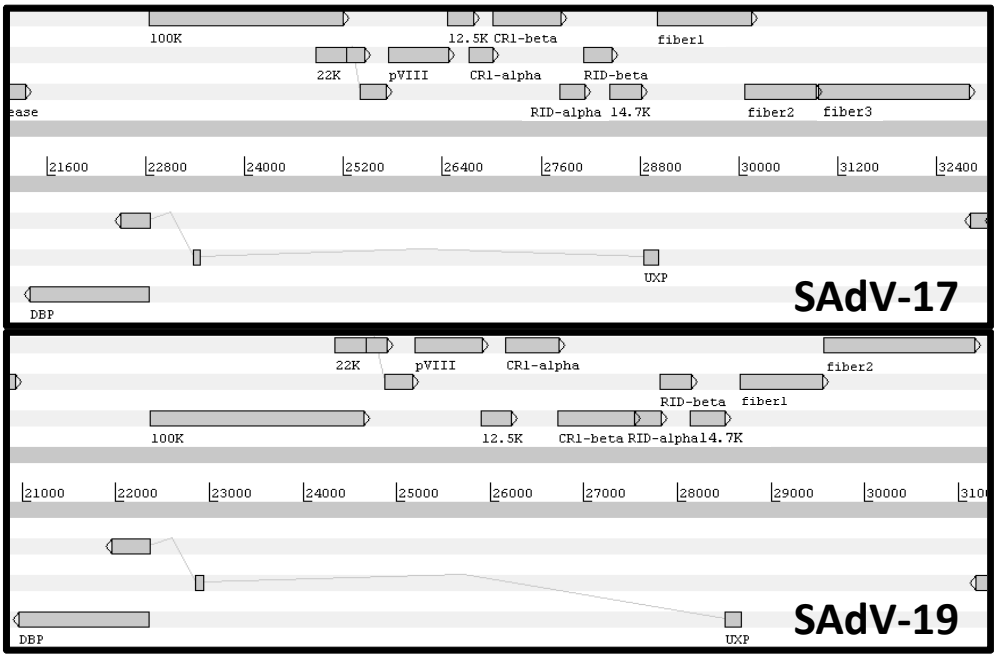


Figure 11. Annotated part of SAdV genomes in which the genomic position and splicing pattern of the UXP, as well as variability in the number of fibre genes can be seen.

Homologous recombinations have been often described in the hexon genes in many primate AdVs (Dehghan *et al.*, 2013a; Dehghan *et al.*, 2013b; Walsh *et al.*, 2009, 2011), therefore we made a phylogeny analysis based on the full hexon aa sequences of all the primate AdVs sequenced to date. The phylogeny inference (Figure 12) pointed out divergent relationships among several AdVs compared to the results gained by other proteins. One of these contradictions could be explained by the results of recombination analysis of the SAdV-19 genome in comparison to that of members of the species HAdV-G, HAdV-F and SAdV-C (Figure 13a and b). The SimPlot and BootScan analyses indicated that recombination event(s) have happened in the hexon gene of SAdV-19.

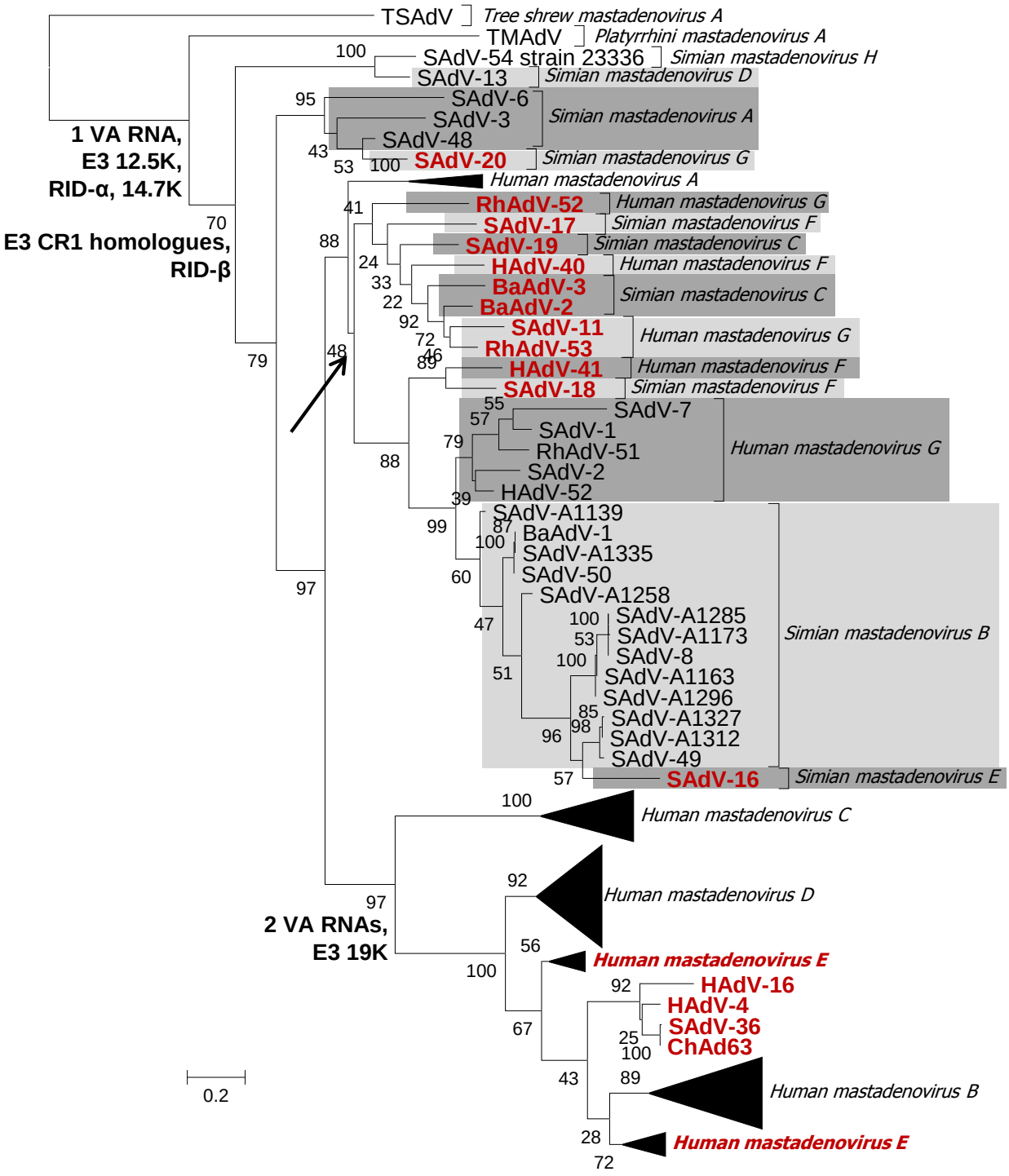


Figure 12. Phylogenetic tree based on full hexon aa sequences. Red letters sign the "problematic" types where the strains are not monophyletic with the other members of the same species or not distinct from the types of other species. Black arrow indicates the node which separates the group of AdVs with two fibre genes (except SAdV-18). Virus associated RNA (VA RNA) and E3 gene labels mark the nodes after which one or more copies of these genes appeared during the evolution.

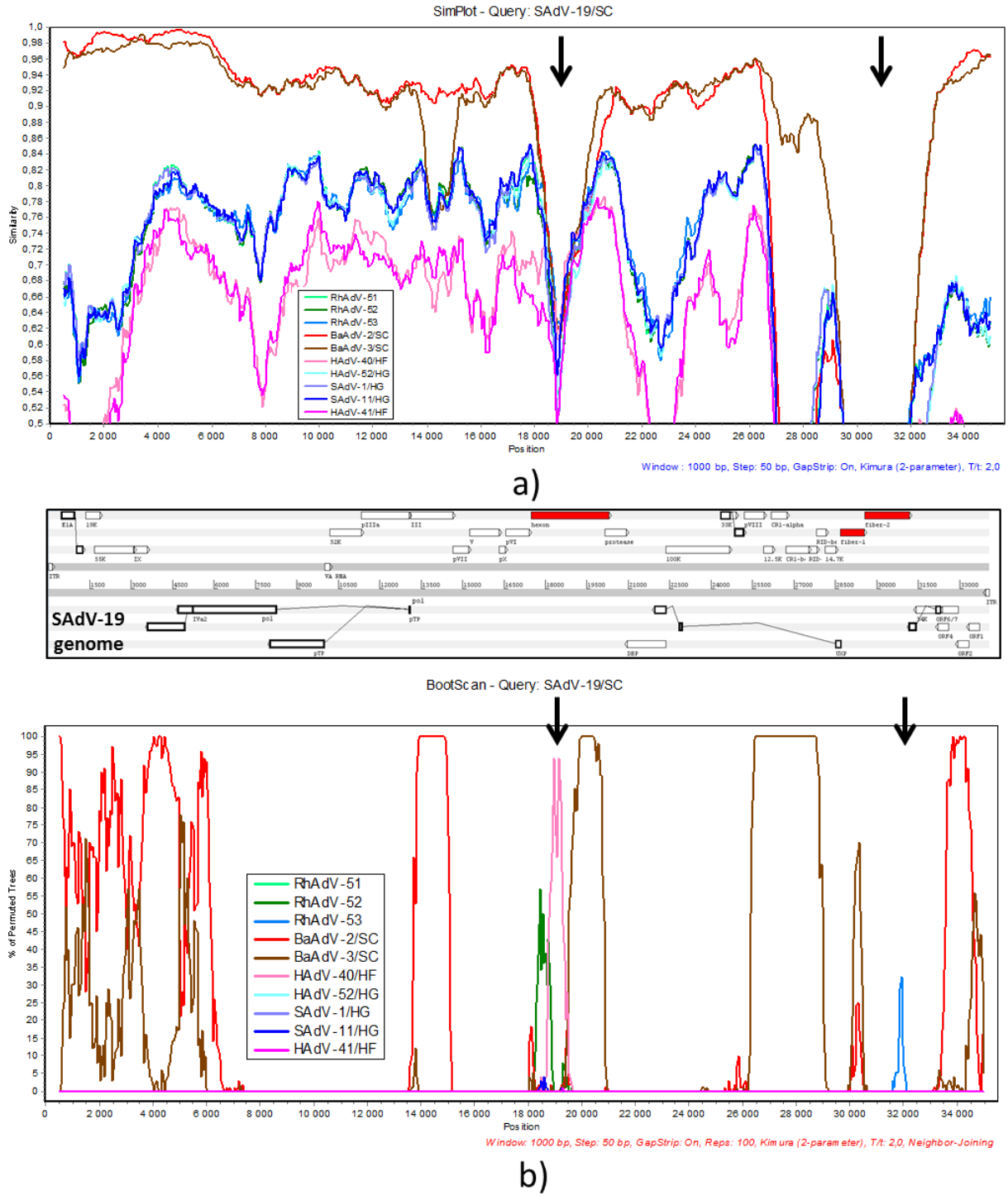


Figure 13. (a) SimPlot and **(b)** BootScan analyses of *Simian mastadenovirus C*, *Human mastadenovirus G* and *Human mastadenovirus F* members relative to SAdV-19. The annotated genome of SAdV-19 is shown between the two graphs to allow easier observation of the genomic locus. Arrows depict the main interspecies recombinations. Red colouring of the genes show the hexon and fibre genes involved in the recombination of SAdV-19. BaAdV-3 strain seems to have a recombination in its penton base gene.

6.1.4 Receptor studies

Three controls were used in the fibre-1 knobs binding studies: HAdV-5 fibre knob which does not bind the sialic acid, but binds the coxsackie-adenovirus receptor, CAR (Bergelson *et al.*, 1997); HAdV-37 fibre knob and HAdV-52 fibre-1 knob, which both bind the sialic acid (Arnberg *et al.*, 2000; Lenman *et al.*, 2015). Pre-treatment of the A549 cells with neuraminidase reduced the binding of fibre knob of HAdV-37 and HAdV-52 fibre-1 knob, as expected (Table 10; Figure 14). Binding of the fibre knob of HAdV-5 was not significantly influenced. Furthermore, the binding was reduced for the fibre-1 knobs of all the studied members of species HAdV-G, as well as for the SAdV-19, to background levels.

Table 10. Results of the FACS analysis of the fibre-1 knobs binding to the sialic acid (bold numbers depict the reduced binding)

Virus	Untreated				Neuraminidase			
SAdV-1	734.0	363.5	418.0	271.0	**22.0**	**-0.5**	**-17.5**	**-11.0**
SAdV-7	3754.0	2238.5	1719.0	1662.0	**119.5**	**68.5**	**16.5**	**93.0**
SAdV-11	222.5	246.0	106.5	188.0	**9.5**	**7.0**	**-27.5**	**-5.5**
SAdV-19	630.5	179.0	139.5	124.5	**142.5**	**5.5**	**-22.0**	**-7.5**
HAdV-5	541.5	409.0	401.0	409.0	359.5	393.5	407.0	388.5
HAdV-37	1685.5	1125.0	1499.0	2077.0	**629.0**	**474.5**	**406.0**	**488.0**
HAdV-52	710.5	424.5	409.0	484.5	**-3.5**	**13.5**	**-31.0**	**-8.5**

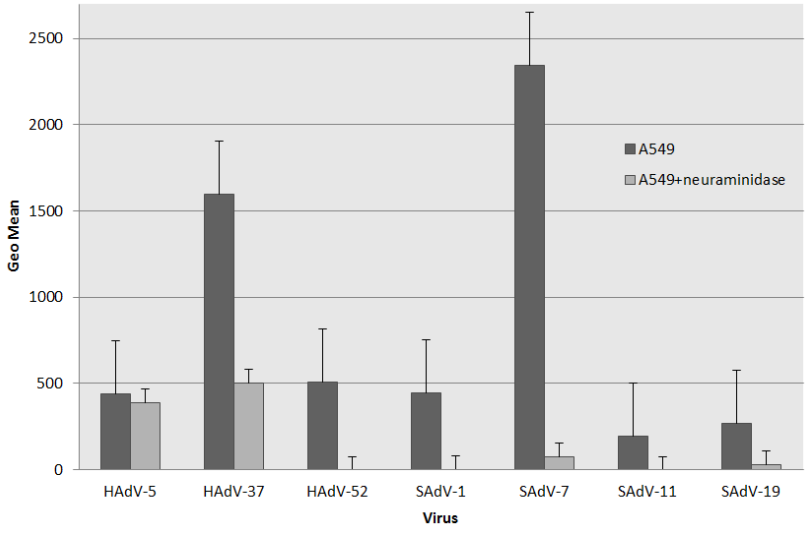


Figure 14. Fibre knobs of HAdV-5 and -37, and fibre-1 knobs of HAdV-52 and SAdVs binding to A549 cells with or without pre-treatment with neuraminidase. Error bars represent mean ± SED.

6.2 Novel adenoviruses

6.2.1 Novel adenoviruses detected by PCR

Extracted DNAs of 138 samples were used in nested PCR reactions with degenerate primers targeting IVa2 gene of the mastadenoviruses. The PCRs resulted in an approximately 300 bp amplicon from 23 samples (16.6%; Table 11): two from 10 ape samples (20%), one from 11 OWM samples (9%), seven from 19 NWM samples (36.8%), and 13 from 98 prosimian samples (13.3%). The PCR products were sequenced and the primer sequences were excluded from the analyses, resulting in 253-bp long useful sequences. The AdV sequence gained from the mandrill sample was not clean, therefore it was not included in the phylogenetic analysis. AdVs detected earlier by other colleagues (also listed in Table 11, but not highlighted with bold) were also included in the analysis.

Nucleotide and aa sequences of the amplified IVa2 region of all the samples were compared to each other, showing that some of the newly detected prosimian AdVs are identical with each other, as well as some NWM AdVs (mostly on the aa level, but sometimes even on nt level). Sequence comparisons showed that there were eight different AdVs detected in prosimians, and seven in NWMs. If we include the AdVs detected by other colleagues of the laboratory, then we got all together 10 different AdVs in prosimians and 15 in NWMs (Figure 15). Unique name is given for the identical AdVs found in different hosts (Table 11) and are represented on the tree (Figure 15) under that name. Identical AdVs were found in hosts from the same (black lemur AdV-3, lemur AdV-1) or different place of origin (lemur AdV-2, marmoset AdV, tufted capuchin AdV-3, red-handed tamarin AdV-2).

6.2.2 Virus isolation attempts

All the samples which were positive by PCR for AdVs were prepared for the infection of different cell lines (Vero E6, A549, HEK293, CHO-K1, 3T6, cmt93). Unfortunately, none of the viruses propagated on any of these cells (all the PCRs performed after the 7^{th} blind passage were negative for the AdVs). MFC cells are still being used in trials of the prosimian AdVs isolation.

6.2.3 Phylogeny inference

Maximum likelihood analysis of partial adenoviral IVa2 aa sequences (84 aa) separated novel AdVs from all the other, earlier established species of primate AdVs (Figure 15). Two new basal primate AdV lineages appeared on the tree: prosimian AdVs as the most ancient ones of primate AdVs, and NWM AdVs, grouped together with previously described TMAdV.

Table 11. Novel adenoviruses detected in this study by PCR (highlighted with bold)

	Host	Proposed adenovirus name	Place of sample origin
Prosimians	**Crowned lemur**	crowned lemur AdV	French zoo
	Eastern lesser bamboo lemur	Eastern lesser bamboo lemur AdV	
	Ring-tailed lemur	ring-tailed lemur AdV	Hungarian zoo 1
	Black-and-white ruffed lemur	black-and-white ruffed lemur AdV	
	Black lemur 1	black lemur AdV-1	Nosy Komba Island, Madagascar
	Black lemur 2	black lemur AdV-2	
	Black lemur 3	black lemur AdV-3	
	Black lemur 4		
	Red lemur 1	red lemur AdV	Ranomafana National Park, Madagascar
	Red lemur 2	lemur AdV-1	
	Red-bellied lemur 1		
	Red-bellied lemur 2		
	Red lemur 3	lemur AdV-2	Kirindy Mitea National Park, Madagascar
	Red lemur 4		
	Red-fronted lemur		Hungarian zoo 7
NWMs	**Gray-bellied night monkey**	gray-bellied night monkey AdV	French zoo
	Red-faced spider monkey	red-faced spider monkey AdV	Hungarian zoo 1
	Golden-headed lion tamarin	golden-headed lion tamarin AdV	Hungarian zoo 5
	Red-bellied tamarin	red-bellied tamarin AdV	Hungarian zoo 3
	Cotton-top tamarin	cotton-top tamarin AdV	Hungarian zoo 5
	Common squirrel monkey 1	common squirrel monkey AdV-1	
	Common squirrel monkey 2	common squirrel monkey AdV-2	Hungarian zoo 2
	Common squirrel monkey 3	common squirrel monkey AdV-3	Hungarian zoo 4
	Common marmoset 1	common marmoset AdV	Hungarian zoo 1
	Common marmoset 2	marmoset AdV	Hungarian zoo 2
	Common marmoset 3		
	White-headed marmoset		Hungarian zoo 3
	Tufted capuchin 1	tufted capuchin AdV-1	Croatian zoo
	Tufted capuchin 2	tufted capuchin AdV-2	Hungarian zoo 4
	Tufted capuchin 3	tufted capuchin AdV-3	Hungarian zoo 1
	Tufted capuchin 4		Hungarian zoo 5
	Red-handed tamarin 1	red-handed tamarin AdV-1	Hungarian zoo 6
	Red-handed tamarin 2	red-handed tamarin AdV-2	Hungarian zoo 1
	Red-handed tamarin 3		French zoo
OWMs	**Mandrill**	mandrill AdV	Hungarian zoo 7
Apes	**Sumatran orangutan**	Sumatran orangutan AdV	Hungarian zoo 7
	Siamang	siamang AdV	Hungarian zoo 7

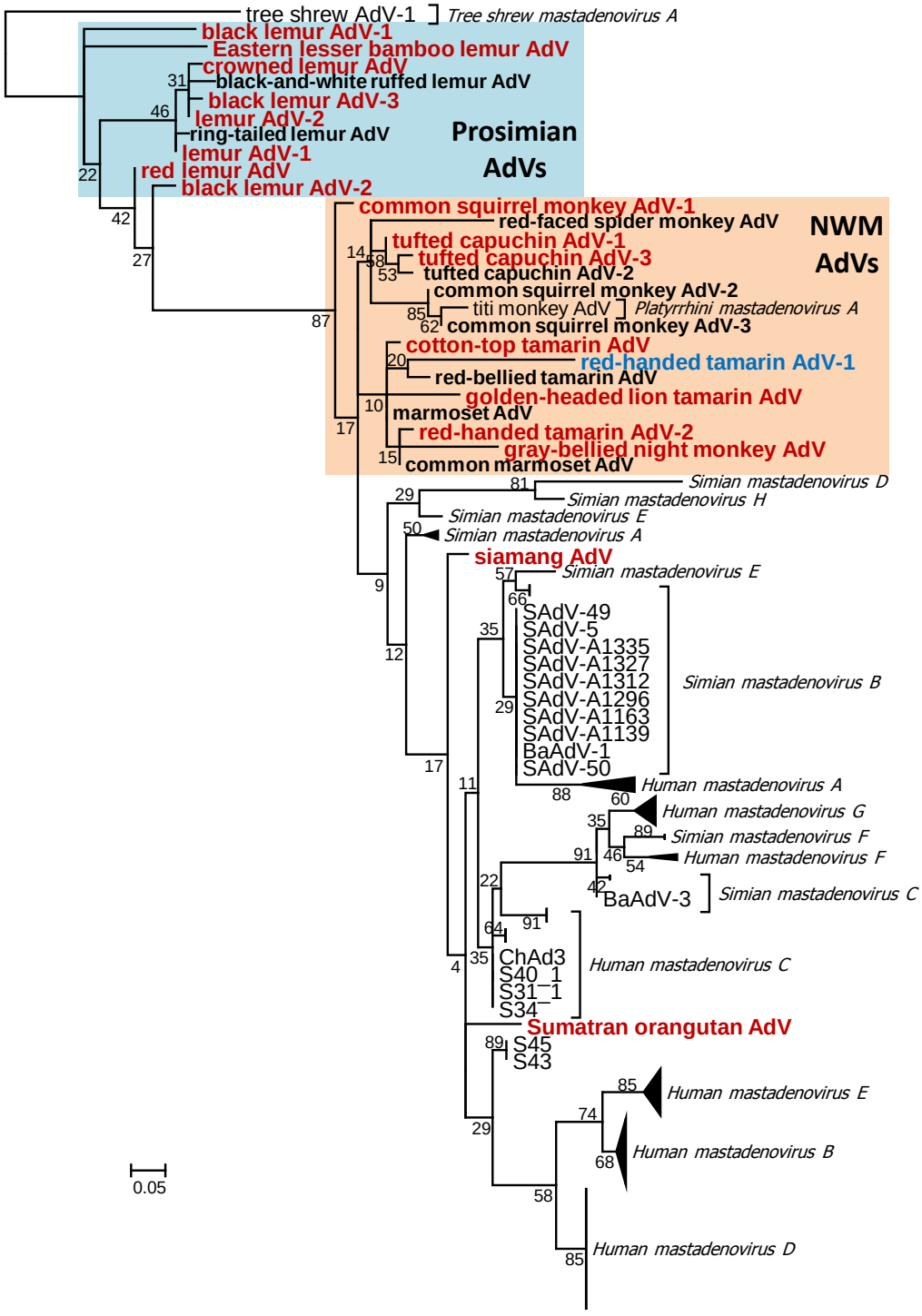


Figure 15. Phylogeny reconstruction based on partial aa sequences from the IVa2 protein of primate AdVs and tree shrew AdV 1 (TSAdV-1) as an outgroup. New AdVs identified in this study are highlighted with bold red. AdVs identified in the laboratory by others (not published) are shown in bold black. Genome of red-handed tamarin AdV-1 (highlighted with blue) is almost fully sequenced.

6.2.4 Red-handed tamarin adenovirus 1

With the traditional (Sanger) method we sequenced 25,885 bp of the RHTAdV-1 genome, ranging from the E1B 55K gene in the left end to the UXP gene in the right part of the genome (Table 12, Figure 16). The sequenced region of the RHTAdV-1 shared genetic composition characteristic for mastadenoviruses, with difference seen in the E3 region: two of the five detected ORFs in this region represent novel genes, not seen in any other AdV before. Majority of the other genes was comparable to TMAdV (Table 12), sharing 30% (UXP) to 82% (protease) identity on the aa level. 33K gene shared more identity with HAdV-4 (48%) than with any other AdV. Most of the genes were shorter (even up to a difference of 88 aa in protein 100K) than those in the TMAdV, except the RID-α which was 21 aa longer. Phylogeny inference placed the RHTAdV-1 among other NWM AdVs, as expected (Figure 15).

Table 12. Analysis of the genes found in RHTAdV-1

Gene	Identity to TMAdV	Length (relative to TMAdV)
E1B 55K	53%	Partial (362/533)
IX	43%	Shorter (178/249)
IVa2	80%	Shorter (414/457)
pol	67%	Shorter (1151/1177)
pTP	69	Shorter (627/681)
52K	69%	Shorter (390/420)
pIIIa	72%	Shorter (555/609)
III	72%	Shorter (485/546)
pVII	55%	Similar (193/197)
V	52%	Shorter (362/396)
pX	59%	Similar (75/81)
pVI	65%	Shorter (221/294)
hexon	80%	Similar (924/929)
protease	82%	Identical (203/204)
DBP	52%	Shorter (492/510)
100K	67%	Shorter (726/814)
22K	64%	Shorter (164/189)
33K	48% HAdV-4	Similar (205/213)
pVIII	71%	Similar (228/227)
12.5K	52%	Similar (112/114)
NEW	No similarity	-
RID-α	43%	Longer (116/95)
NEW	No similarity	-
14.7K	55%	Similar (122/131)
UXP	30%	Similar (188/198)

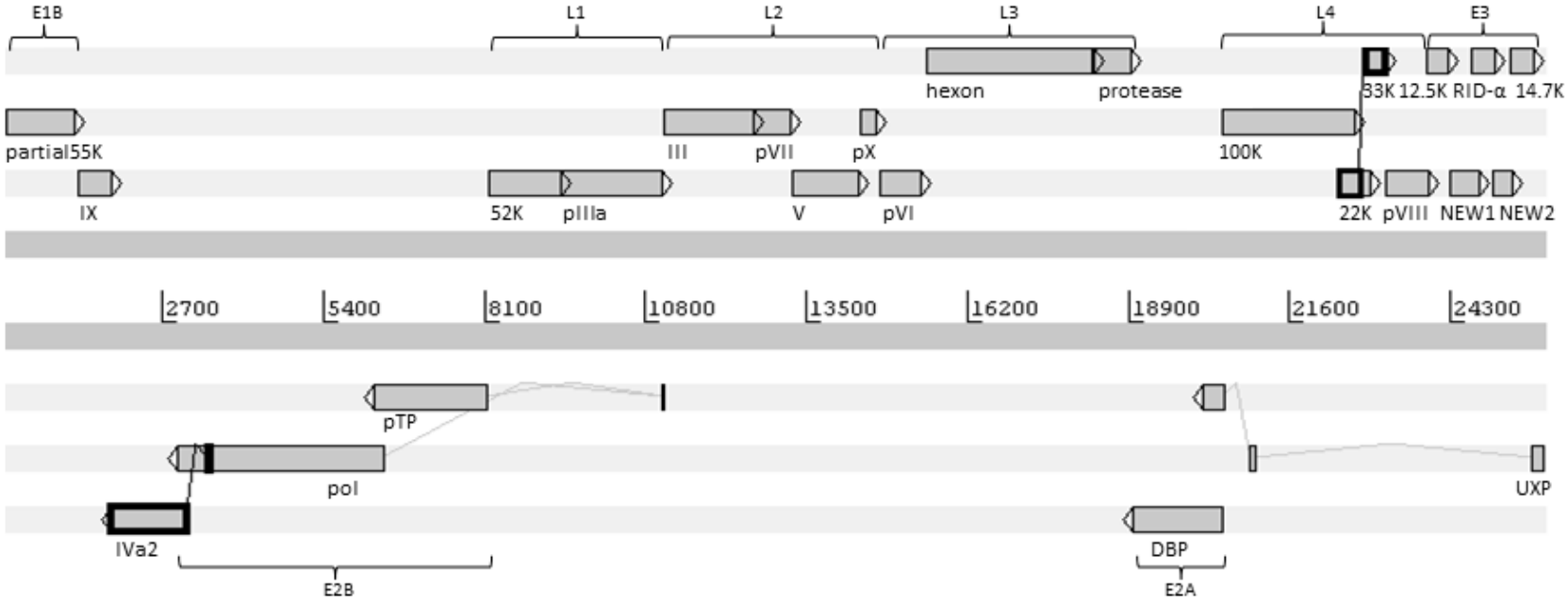


Figure 16. Annotated genome map of red-handed tamarin adenovirus 1. Spliced genes sharing some parts with another ORF are highlighted with thick line.

7. Discussion

7.1 Strains from the ATCC

7.1.1 Production of virions and viral DNA for next-generation sequencing

For production of the studied OWM AdVs for NGS, we decided to use the Vero E6 cells, since it was expected that most of them would propagate on this OWM cell-line. Unfortunately, some of the strains in our possession (SAdV-10 and -13) did not grow on Vero E6, which may have been well the consequence of having not living strains. On the other hand, some strains showed good CPE, but the extracted DNA indicated that there is a parvovirus (adeno-associated virus, AAV) in majority (e.g., in SAdV-4), or there was not any DNA at all (SAdV-9 and -14; Table 5). Consequently, we could not sequence the full genome of these viruses this time (except SAdV-13, sequenced by a colleague by Sanger method). The infection of the cells by three ATCC strains (SAdV-5, -12, and -15) resulted in very high yields of DNA (>1.5 μg/μl), compared to other strains. The reason for this might be the existence of at least two types in each mixture, whereas all the other strains are clean types. The differences in the DNA concentrations gained from each strain are visible in Figure 4, as well as the parvovirus bands. Finding parvovirus DNA in the mentioned cases was not too surprising for us, especially as most of the non-human primate AAVs have been isolated and published as contaminants of AdV preparations including those from ATCC strains (Schmidt *et al*., 2006; reviewed by Balakrishnan & Jayandharan, 2014).

Members of HAdV-G (SAdV-1, -2 and -11) were chosen to be propagated on A549 cells for the possible receptor-binding studies in collaborating laboratory in Umea, Sweden. SAdV-7, another member of HAdV-G species, was propagated on A549 cells earlier by colleagues in Sweden. This laboratory is studying HAdVs mostly, among which is also HAdV-52 from species HAdV-G, and the receptor-binding studies were already performed for this virus (Lenman *et al.*, 2015). This allowed us to compare the results gained for the studied SAdVs to those of HAdV-52. The yields of the HAdV-5 virions were higher than those of the SAdV types (Table 6), as expected when using human cell line A549 for propagation of the virions. SAdV-11, as well as other types, reached good CPE on the cells, but there was no visible band of the AdV after the CsCl gradient. Since we have seen earlier that in some cases parvovirus can be in majority, one would suppose that this might be the case here. However, DNA extracted from the SAdV-11 propagated on Vero E6 cells was very clean (Figure 4), and after the CsCl gradient ultracentrifugation there was no band of AAV visible. Therefore, we think that the concentration of the virions was too low, causing

the lack of the visible AdV band. The concentration of the SAdV-1 and -2 virions was much lower than that of HAdV-5, but still sufficient for the receptor-binding studies.

7.1.2 Partial genome sequencing and ATCC mixtures

Sequencing of four different regions (partial IVa2, *pol*, penton base and hexon genes) of all serotyped, non-sequenced, ATCC OWM AdV strains allowed us to establish their molecular similarities and to assign them into species based on the phylogeny inference and several other biological properties (GC content, host, number of fibre genes, HAG; Pantó *et al.*, 2015). The four studied genes are different by being expressed in different stages of the viral life cycle (representing both early and late genes). The *pol* is important because the species demarcation is based on this protein (Harrach, 2014). The amplified penton base gene fragment proved to be well applicable for detecting different types/variants in SAdV strains earlier supposed to be clean isolates, because the amplified region is a highly variable part. Molecular cloning of amplified penton base gene fragments proved that three of the ATCC strains (SAdV-5, -12 and -15) are mixtures of multiple AdVs. Consequently, even the validity of their original serological comparison with other strain can be questioned. IVa2 nested PCR was found to detect mastadenoviruses very effectively, and for the primate AdVs it proved to be even more sensitive than the *pol* PCR. The amplified partial hexon gene coding the highly conserved N-terminal part is a very popular genome fragment for general AdV PCR (Bányai *et al.*, 2010; Maluquer de Motes *et al.*, 2011; Wevers *et al.*, 2011). As only partial gene sequences were amplified, the validity of the phylogenetic trees based on them is an important question. The bootstrap values mirror the length and conservation status of the amplified fragment: *pol* aa based tree shows values ranging from 89 to 100 for different primate AdV species, and it seems to show correctly also the “time point” of the presumed acquisition of a second fibre gene (Figure 6). From that time-point, all the OWM AdVs and even HAdV members from species HAdV-F and HAdV-G (both of them supposed to have originated from OWM AdVs), have two fibre genes, or even three as seen in SAdV-2 and -17. The only exception is SAdV-18, which probably lost it during a presumed host-switch and adaptation to grivet. Besides these previously non-sequenced SAdVs, three viruses sequenced by other laboratories, SAdV-18 and -20, SAdV strain 23336, and the only sequenced NWM AdV, that from titi monkey, were also assigned to novel species in this study as follows.

Phylogenetic calculations (Figure 6; Pantó *et al.*, 2015) indicate that nine of the studied SAdVs belong either to the species SAdV-A (SAdV-4, -9, 10 and -14) or HAdV-G (SAdV-2, -5 (one of the types in the mixture), -11, -12 (both types) and -15 (both types)), established earlier (Jones *et al.*, 2007; Kovács *et al.*, 2004). AdV genotypes published earlier

indicate that the prevalence of these two groups is high in macaques (Bányai *et al.*, 2010; Lu *et al.*, 2011). Besides these two monkey AdV lineages, other genetic clusters appeared as well. The third most numerous cluster is that of SAdV-5, -8, -49, -50, nine other rhesus macaque isolates (Roy *et al.*, 2012; Roy *et al.*, 2009), and BaAdV-1 (Chiu *et al.*, 2013). This cluster has been proposed to be named species SAdV-B (Roy *et al.*, 2012), and was recently officially accepted by the Executive Committee of ICTV (Podgorski & Harrach, 2015). Presently analyzed SAdV-5 (one type) and -8 are further proved members of it based on our studies, albeit the first members that had been even serotyped by virus neutralization (Rapoza, 1967). Other earlier publications based on partial hexon sequences supported this clade as well (Bányai *et al.*, 2010; Foytich *et al.*, 2014; Maluquer de Motes *et al.*, 2011; Wevers *et al.*, 2011). SAdV-19 proved to be a further member of the previously proposed species SAdV-C (Chiu *et al.*, 2013), which was recently accepted by the Executive Committee of ICTV (Podgorski & Harrach, 2015). This assignment was also confirmed by the GC content (Table 7), and by the uniqueness and uniformity of the host (baboon).

SAdV-13 turned out to be the only member of a separate lineage suggested as a new species, SAdV-D. The exact host species of this AdV type (*Macaca* sp.) is not known, unfortunately. All the constructed phylogenetic trees (Pantó *et al.*, 2015) inferred that SAdV-13 diverged from the other monkey AdVs at an early time. GC content of this virus is also different from all the other viruses (Table 7). The phylogenetically closest AdV is the rhesus monkey isolate 23336, but both the phylogenetic distance and the GC content (46.7% contra 49.9%) differentiate them. Therefore, we propose SAdV-23336 to be the member of a new candidate species named *Simian mastadenovirus H*. As each species must have a type species, we propose to give the next available SAdV type number for this strain. As RhAdV-51 to 53 got their numbers as the last SAdVs with full sequences, we propose to name those three rhesus AdV strains as SAdV-51 to 53 to follow earlier naming system, and give the SAdV type number 54 for the strain 23336 (isolated also from rhesus macaque). The phylogenetic calculations based on both the partial *pol* or full hexon genes show adequate molecular distance to see it confirmed that even SAdV/RhAdV-51 to 53 are adequately distant to get individual type status, and most probably are enough different serologically to represent different serotypes, too.

SAdV-16 is closely related to the proposed species SAdV-B. However, its phylogenetic distance seems to be large enough to propose it to be the first member of a separate species. Different host species, grivet (*Chlorocebus aethiops*), supports the establishment of a new taxon for SAdV-16. However, as host switch is supposed to be relatively common among AdVs (Wevers *et al.*, 2011), it cannot be excluded that this virus has a macaque origin because the two other grivet AdVs (SAdV-17 and -18) are phylogenetically very different from SAdV-16, while very similar to each other. The GC

content of the full genome of SAdV-16 is 57.9%, which is remarkably different from those of the members of SAdV-B (~62%). Based on all data available for this virus, we propose this type to be classified as representative of novel species named SAdV-E.

SAdV-17 and -18 compose the sister taxon of HAdV-F. The phylogenetic distance and host species difference (*Chlorocebus aethiops*) seems to be sufficient to propose a new species, SAdV-F. Also the genome organizations of SAdV-17 and -18 show very important differences compared to HAdV-40 and -41 (the two known members of HAdV-F). SAdV-18 has one fibre gene, SAdV-17 has three fibre genes, while HAdV-40 and -41 have two of it. The HAdV-F members are unique among the primate AdVs, as they do not have RGD motif in the penton base gene, and they also lack the 12.5K gene in their E3 region, while SAdV-17 and -18 have both the RGD motif and the 12.5K gene. The GC content of HAdV-40 and -41 is 51%, while SAdV-17 and -18 have GC rich genomes (58.8-61.4%). We think that the proposal to establish a separate species (SAdV-F) for SAdV-17 and -18 is adequate. Interestingly, both SAdV-17 and -18 are exceptions in the number of fibre genes: while most of the SAdVs from a certain time point in evolution (Figure 6) contain two fibre genes, the SAdV-18 is the sole member with only one fibre gene, whereas the SAdV-17 is one of the two AdVs which proved to have three fibre genes. Nevertheless, other biological properties and phylogeny inference indicate they belong to the same species.

Phylogenetic trees show SAdV-20 as the sister group of SAdV-A, but adequately distant to be a valid separate species (SAdV-G). The only exception is the hexon (Pantó *et al.*, 2015), where SAdV-20 is not separated clearly from SAdV-A members. This may be caused by a recombination event, which occurs often in this gene (Dehghan *et al.*, 2013a; Dehghan *et al.*, 2013b; Walsh *et al.*, 2011). Difference in host species and radically different GC content of the SAdV-20 genome (47.8%) from that of SAdV-A members (54.4-55.8% on the full genome length) confirm our proposal for the new species, SAdV-G.

Comparing the GC percentages of AdV genomes, members of different species proved to have well-distinguishable values (Table 7) and it is a further species demarcation criterion (Harrach, 2014). The limited sequence length of AdVs could lead to controversial inferences. However, data of completed genomes irrefutably make the difference between the proposed species. SAdV-B and -F are GC rich species (58.8-62.9%), SAdV-A, -E, and HAdV-G are moderately GC rich (54.4-57.9%), SAdV-C, -D, -G and HAdV-F have almost balanced GC content (47.8-52.6%), and SAdV-54 strain 23336 (SAdV-H) has the lowest GC content (46.7%; Table 7).

The first classification of monkey AdVs, based on hemagglutination-inhibition test as an early tool of taxon separation (Rapoza, 1967) also helped us in the species demarcation. Every HAG group II member belongs to SAdV-A (Table 7). HAG group III members SAdV-5 and -8 belong to the SAdV-B, however this group includes several members of HAdV-G, too.

This indicates that the HAG probe is not sensitive enough to make always a difference at the species level. Still, it may help to distinguish species from each other, thus this biological property is one of the applicable species demarcation criteria (Harrach, 2014). For example, HAG classification confirmed the uniqueness of SAdV-13 as it is the sole member of the HAG group I. Similarly, SAdV-16 is the sole member of HAG group IV, which confirms its classification as a new taxon.

The tissue tropism of the ATCC strains does not seem to be determined by phylogenetic clustering, i.e., AdVs isolated from internal organs (such as liver and kidney) share common species with apparent enteric SAdVs. It would be interesting to find out if members of different SAdV species have different tissue tropism in macaques, just as it is with HAdVs. E.g., members of HAdV-F, typical enteric AdVs, are found generally in human stool (and are common in waste water), while some HAdV-D members are notorious to infect the cornea and cause epidemic keratoconjunctivitis (Benkő, 2015). An early study reported that SAdVs can cause epidemic conjunctivitis in macaques (Vasileva *et al.*, 1978). By comparing the results with our phylogenetic clustering, it is notable that seemingly only SAdV-A members cause this disease.

The host range of the members of several newly proposed and accepted species, and the previously acknowledged SAdV-A and HAdV-G is mixed (Table 7). AdVs infecting different simian genera usually belong to separate AdV species. However, there are AdV species that contain AdVs of several different monkey species. In contrary, in some cases, AdVs from the same monkey species may belong to different viral species. This is a general feature of SAdV and HAdV species to have mixed host origins, e.g., human, chimpanzee, gorilla and bonobo AdVs were suggested as members of HAdV-B and HAdV-C (Roy *et al.*, 2009). But, crossing the host barrier occurs usually very rarely and mainly only among evolutionary close primate species, most characteristically between apes and humans or among OWMs.

The penton base PCR was the best method in this work to detect different variants in three ATCC deposits. Our conclusion is that SAdV-12 and -15 are mixtures of two types of the HAdV-G species (Figure 7), whilst the SAdV-5 seems to be a mixture of a HAdV-G and a SAdV-B member. The signs of non-homogeneity were observed also in previously conducted cross neutralization experiments (Hull *et al.*, 1958). SAdV-5 (a mixture of HAdV-G and SAdV-B members) showed one way cross neutralization, and SAdV-12 (seemingly a mixture of two HAdV-G members) had two way cross-neutralization with different putative HAdV-G types. This early experiment did not study SAdV-15, so our statement on the mixed type is based on the described penton base sequences, but also on a shot-gun sequencing (Kovács *et al.*, 2004) attempt that revealed heterogeneity of SAdV-15 (unpublished). The end-point dilution assays, used for the separation of at least one type from the mixture, were,

unfortunately, not successful in case of SAdV-12 and -15. The PCR with primers targeting the penton base gene of SAdVs indicated that the obtained types were clean, whereas the NGS proved that they still have a small amount of the 2nd type. Consequently, high similarity of the sequences made it impossible to separate them by bioinformatics programs. We suppose that a very small amount of the 2nd type compared to the huge amount of the prevailing type in each case was not sufficient to be detected by the nested PCR targeting the penton base gene. Therefore, the obtained sequences were clean and indicated that the type is clean. In case of SAdV-5, we managed to obtain a clean sequence of about half of the genome. Interestingly, the sequence proved to be exactly the same as one of the previously published SAdVs, strain A1335 (Roy *et al.*, 2012), belonging to the species SAdV-B. We decided not to sequence the rest of the genome since it is exactly the same as the mentioned one, but anyway it is a very interesting information for us that the ATCC strain described a long time ago (Hoffert *et al.*, 1958) was found again in the rhesus macaques housed in primate facilities (Roy *et al.*, 2012). Furthermore, now we know the full genome sequence of one type of the ATCC strain SAdV-5, whereas the other one, belonging to species HAdV-G, has to be isolated and sequenced. The three ATCC strains proven to be mixtures are currently being analysed by colleagues in the collaborative laboratory in the Netherlands and in Hungary.

Based on comparisons of the phylogenetic analyses and biological properties of OWM AdVs, we established two earlier proposed species SAdV-B and SAdV-C. Furthermore, we proposed the establishment of five new SAdV species, SAdV-D, -E, -F, -G and -H (Figure 17). All of these newly proposed species would contain only OWM AdVs for now. We assume that our proposal of SAdV species will be confirmed in the near future by finding further AdVs and analyzing the phylogeny and full genome characteristics of a large number of SAdVs.

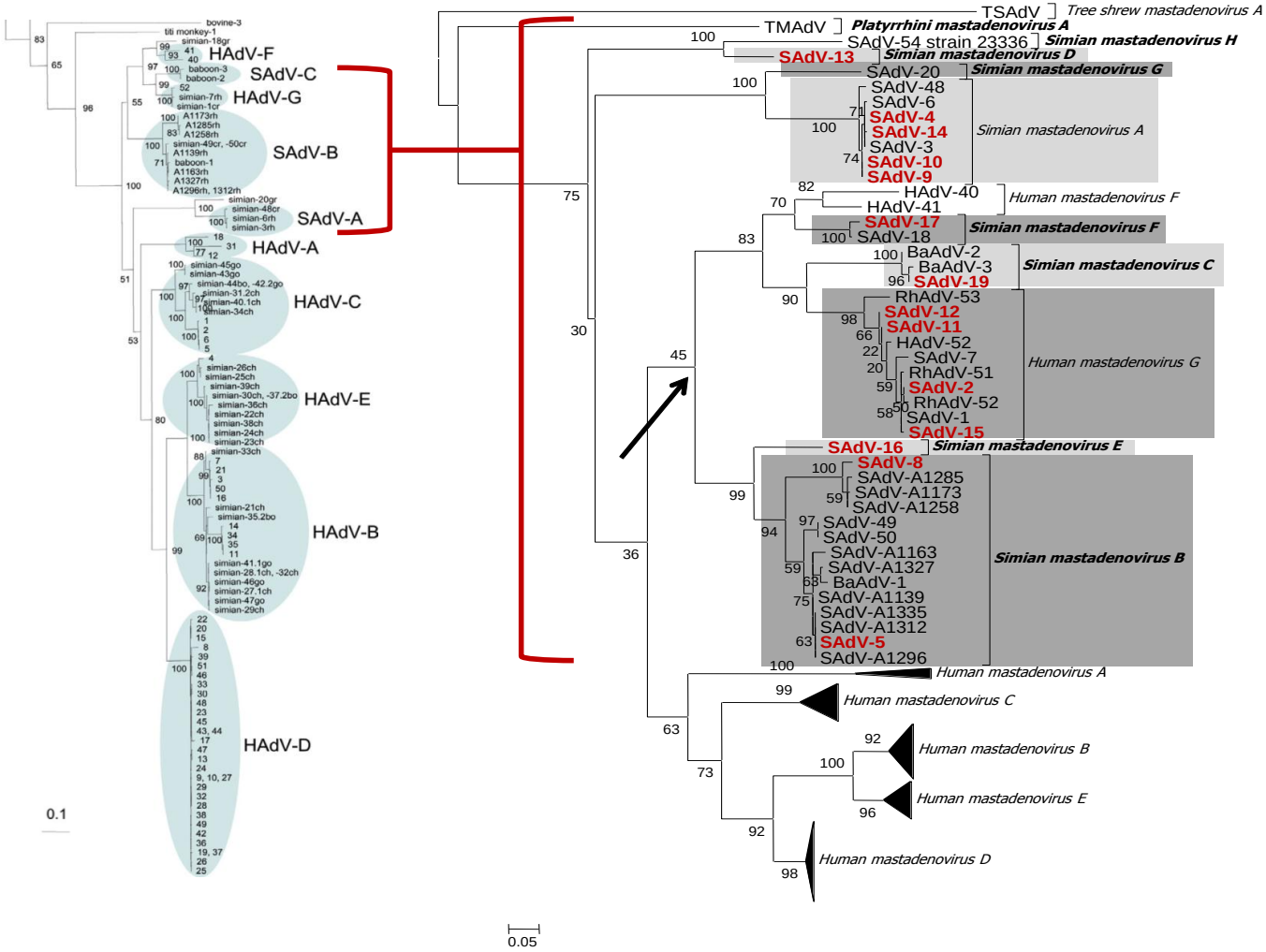


Figure 17. Comparison of partial Figure 1 (containing primate adenoviruses) and Figure 6, with red lines pointing out the difference in number of OWM and NWM species, including HAdV-G which turned out to contain almost exclusively OWM AdVs, before (2 established and 2 proposed) and after this study (4 established and 6 proposed).

7.1.3 Full genome sequencing

During the three and half years, we successfully sequenced the full genomes of six OWM AdVs (SAdV-2, -8, -11, -16, -17 and -19), while SAdV-13 was sequenced by colleague Laura Pantó and the sequencing finished by me, and assigned them to the previously established (HAdV-G, SAdV-A) or newly established/proposed (SAdV-B to H) species. We officially proposed the species SAdV-B and SAdV-C to the ICTV (Podgorski & Harrach, 2015), and they were recently accepted. Species SAdV-D to H for OWM AdVs, and *Platyrrhini mastadenovirus A* for a NWM AdV, were also officially proposed to the ICTV recently, but it will be discussed by the Executive Committee of ICTV only in August 2016 and might be finally accepted in 2017 only. In order to use two different approaches in sequencing methods (traditional and NGS), two of the viruses were sequenced by the traditional method (SAdV-16 and -19), and four (SAdV-2, -8, -11 and -17) by the NGS (Illumina HiSeq2500 system) performed by commercial services (BGI and BaseClear). Consequently, since SAdV-16 and -19 were not sequenced by the NGS, they were not produced in large volumes as there was no need for relatively large amount of DNA, but PCRs were used to amplify the different fragments for sequencing.

The size of OWM 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 six newly sequenced SAdV genomes fall within this range (Table 8), and have 36 or 37 putative genes characteristic for members of the genus *Mastadenovirus* (Figure 9), including the presence of two or three fibre genes (Figure 11), a feature recognised only in members of HAdV species HAdV-G and HAdV-F and in many OWM monkey AdVs in this genus (Alonso-Padilla *et al.*, 2015; Pantó *et al.*, 2015). Besides mastadenoviruses, two fibre genes have also been found in many representatives of different aviadenovirus species (Kaján *et al.*, 2010, 2012; Marek *et al.*, 2014a, 2014b; Zhao *et al.*, 2015) and in only two sequenced atadenoviruses (Pénzes *et al.*, 2014; To *et al.*, 2014). In this study, for the first time, we detected three fibre genes in AdVs (in SAdV-2 from species HAdV-G and SAdV-17 from SAdV-F). This might give great potential to these AdVs for vectorizing purpose, since the fibres of different species can be switched in order to use certain receptors on the host cells. The G+C content, an important species demarcation criterion in AdVs (Harrach, 2014) varies in SAdVs between 47.8% (SAdV-20) and 65.7% (SAdV strain A1139; Roy *et al.*, 2012). Within each of the studied primate AdV species, G+C content varies <3% (Pantó *et al.*, 2015). G+C contents of the six presently studied SAdVs were found to be within this range, and usually corresponded to the G+C content of the species the viruses are proposed to belong to (Table 8).

ITRs of SAdV-19 (127 bp) are longer than those (87 bp) of the other BaAdVs from the species SAdV-C. SAdV-19 also has an overall longer genome, and has been isolated from a different baboon species. It seems that members of the same species may have ITRs of quite different lengths as can be seen also in species HAdV-G (Table 8).

The existence of a protein, coded 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 in the U exon defects virus growth and causes aberrant localisation of the DBP on the viral genome (Tollefson *et al.*, 2007). Amino acid sequence alignments of the UXPs predicted by me in SAdVs and described in HAdVs from HAdV-C (Figure 10) showed that the first and the second exons are relatively well conserved, even in different species of AdVs, whilst the third exon is more variable both in sequence content and length (Table 9). However, the genomic localisation of all the three exons of the studied SAdVs was comparable (Figure 11). Comparison of the third exon of the UXP of the studied SAdVs to other members of the species they are proposed to belong to (data not shown) confirms that in members of the same species the third exon starts at the same upstream position relative to the DBP-encoding region. Furthermore, it also confirms, together with data shown in Figure 11 that the third exon of the UXP is always in the same reading frame relative to DBP. Thus, 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. Anyway, we can expect that defining all the three exons of the UXP in other non-primate AdVs might be more challenging due to increased differences in the sequences of AdV species which are evolutionarily more distant. This especially applies to the third exon of the UXP, which is apparently very variable even within the same species of primate AdVs.

All OWM AdV genomes, studied earlier, have been found to have only one VA-RNA gene, while chimpanzee and most human AdVs (in species HAdV-B to E) have two VA-RNA genes (Kidd *et al.*, 1995; Larsson *et al.*, 1986). In some primate AdVs, VA RNAs have not been studied yet, or they could not be determined probably because of insufficient primer binding due to sequence divergence or due to the possible absence of a VA RNA region in those AdVs (Kidd *et al.*, 1995). The VA-RNA gene of SAdV-16 (strain SA7) was characterised almost 30 years ago, proving that this OWM AdV has only one such gene (Larsson *et al.*, 1986). Our SAdV-16 genome sequence confirmed the presence of this (single) VA-RNA gene, albeit in one of the earlier studies its PCR amplification had failed (Kidd *et al.*, 1995). It has been hypothesised that the failure of the amplification was possibly caused by the use of genetically different strains which have identical designations in the different laboratories (Kidd *et al.*, 1995). The sequence of the VA-RNA genes of SAdV-2, SAdV-11 and SAdV-19 has been published earlier (Kidd *et al.*, 1995), and here we report

their exact position in the genomes. Whilst the sequence length of the VA RNAs of SAdV-16 and -19 is the same as the ones reported earlier (168 in both), that of the SAdV-2 and SAdV-11 is longer (164 nt in both *vs* 93 and 104 nt reported earlier, respectively). 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-2 and -11 (164 in both), SAdV-13 (146 nt), SAdV-16 (168 nt) and SAdV-19 (168 nt). 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, based on sequence comparisons, now we prove for the first time the presence of a VA-RNA gene also in SAdV-8 and SAdV-17. This gene is longer (159 and 169 nt, respectively) than the VA RNAs of most OWM AdVs. SAdV-8 is a member of the species SAdV-B, and SAdV-17 of the proposed species SAdV-F, from which we do not have published information about the VA-RNA gene of any other member. On the other hand, the genome of SAdV-49 (SAdV-B) and SAdV-18 (SAdV-F) has been published, but, by sequence comparison, I determined the position of the VA-RNA gene, which was found to be 157 and 169 nt long, respectively. The length and position of these genes correspond to the VA-RNA genes predicted in the genomes of SAdV-8 and SAdV-17, confirming the existence of this gene in the members of species SAdV-B and -F. Thus we proved that all SAdVs contain at least one VA RNA. Finally, 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).

Part of the SAdV-2 genome (repetition of the ITR and E4 ORF1 in place of the E1B 19K gene), which is different in this type from all the other types, was sequenced by the traditional sequencing as well, in order to confirm the result gained by the NGS. To exclude the possibility that the repetition was caused by a mutation during the propagation in Vero E6 cells in our laboratory, DNA samples from the virus both before and after the propagation were sequenced. The result was in every case exactly the same (with the NGS sequencing and traditional sequencing before and after propagation), therefore we conclude that this virus indeed has this deletion/repetition never seen in other AdVs before. While we excluded a possible sequencing mistake as a reason for this strange genome organization, obviously we cannot exclude that this mutation happened during earlier TC propagation and was not present in the virus while still in the original host. E1B 19K gene is a Bcl-2 homolog that blocks apoptosis induction (Subramanian *et al*., 1995). It is known that deletions of parts or all of the E1B have no effect on recombination (Epstein & Young, 1991) but they do result in failure to protect viral DNA from the DNA endonuclease induced by the E1A gene products and apoptosis, thus limiting the virus replication (reviewed by White, 2001). Consequently, it is quite surprising that the virus would have a natural deletion in this region. On the other hand, several studies showed that the E1B 19K deletion mutant AdVs demonstrate enhanced tumour cell killing and therefore enhanced oncolytic potency (Liu *et al*., 2004;

Sauthoff *et al*., 2000). Accordingly, there seems to be a promising potential of such a virus, which has three fibre genes and a natural deletion of its E1B 19K gene, especially if the goal of the adenoviral vector application is targeted host cell destruction. A further added value may be for some that this is a natural deletion and not a genetic engineered modification.

The hexon-based tree shows different relationship among several simian and human AdVs (Figure 12). This might be the result of recombination. SimPlot analysis of SAdV-19 clearly indicates the probability of a homologous recombination in the hexon gene, and the BootScan analysis suggests that the SAdV-19 hexon gene resulted from a recombination event between SAdV-19 and a yet unknown member of species HAdV-F (perhaps HAdV-40 or a very similar type not known yet; Figure 13). The hexon based phylogenetic tree is not conclusive on the separation of HAdV-F and -G, and SAdV-F (SAdV-F members were not included in the SimPlot analysis because of different number in fibre genes). This also supports the probability of several recombination events among these AdVs. Homologous recombinations most often in the hexon and fibre genes have been described in many primate AdVs (Chiu *et al*., 2013; Dehghan *et al*., 2013a; Dehghan *et al*., 2013b; Walsh *et al*., 2009, 2011). Consequently, a divergent topology on the hexon tree can also be observed in other AdV lineages. For example, HAdV-4 and -16 are clustered together and 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 strain ChAd63 (HAdV-C) is very similar to that of SAdV-36 (HAdV-E), implying an interspecies homologous recombination. SAdV-16 appeared on most trees (Pantó *et al.*, 2015) far enough from the species SAdV-B to be considered as a representative of a new species, SAdV-E. However, the hexon tree shows SAdV-16 within species SAdV-B. A homologous recombination in the hexon of SAdV-16 is therefore very likely. Some other properties of SAdV-16, such as host, G+C content, results of the hemagglutination-inhibition test (Rapoza, 1967) also confirm that it should be considered as a new species, distinct from SAdV-B.

The SAdV-2, -17 and -18 sequences were excluded from the SimPlot analysis based on the full genome sequence because they have three or one fibre gene, respectively. Therefore, their comparison with other SAdVs with two fibre genes would give inconclusive results. Nevertheless, we performed the SimPlot analysis with the fibre-2 sequences separately. The fibre-3 of SAdV-2 and -17 corresponds to the fibre-2 of other SAdVs, therefore it was included in the analysis with fibre-2 genes of other SAdVs. The SimPlot analysis of these fibres did not indicate any recombination event among the so far described AdVs (data not shown). The fibre-1 sequences (including fibre-1 and -2 of both SAdV-2 and -17) were too diverse to be able to align them and perform the SimPlot analysis with all of

them. Nevertheless, we compared the sequences to the ones available in the GenBank and found interesting data regarding the AdVs with three fibre genes: fibre-1 of the SAdV-17 shares 41% identity with HAdV-41 (HAdV-F) fibre-1 gene, whereas the fibre-2 of SAdV-17 shares 48% identity with SAdV-19 (SAdV-C) fibre-1. Therefore the recombination events in this part of the genome are very likely. In case of SAdV-2, fibre-1 shares 55% identity with SAdV-19 fibre-1, whereas fibre-2 seems to be the result of a recombination among different viruses since the first half shares 64% identity with SAdV-1 fibre-1, and the second half only 38% identity with fibre-2 of HAdV-F members. Comparison of the SAdV-19 fibre-1 with available AdV sequences revealed that it shares only 46% identity with that of members of species HAdV-G. The tree based on the fibre-1 aa sequence completely separates SAdV-19 from its proposed species SAdV-C, but also from all the other species we know (Figure 18). With recombination analyses, we were unable to point out any AdV in the known species that could be supposed as the origin of this fibre gene. Apparently the donor AdV (of this fibre gene) is yet to be identified. Nonetheless the evolutionary tree, based on the sequence of the fibre-1, indicates that it might be the most ancient of all the known primate AdVs with two fibre genes.

The organisation of the genomes of all newly and fully sequenced SAdVs was comparable, and very similar to that of the previously sequenced SAdVs. The gene of the UXP homologue was identified in all studied SAdVs based on comparison with the UXP gene of the members of species HAdV-C. However, mRNA studies should be performed to confirm the proposed location and length of the exons. It is clear now that SAdVs cluster into species as well as HAdVs do, and that the future may bring more primate AdV species for Old World monkeys exclusively, as well as for more ancient AdVs, i.e. for those of New World monkeys and prosimians.

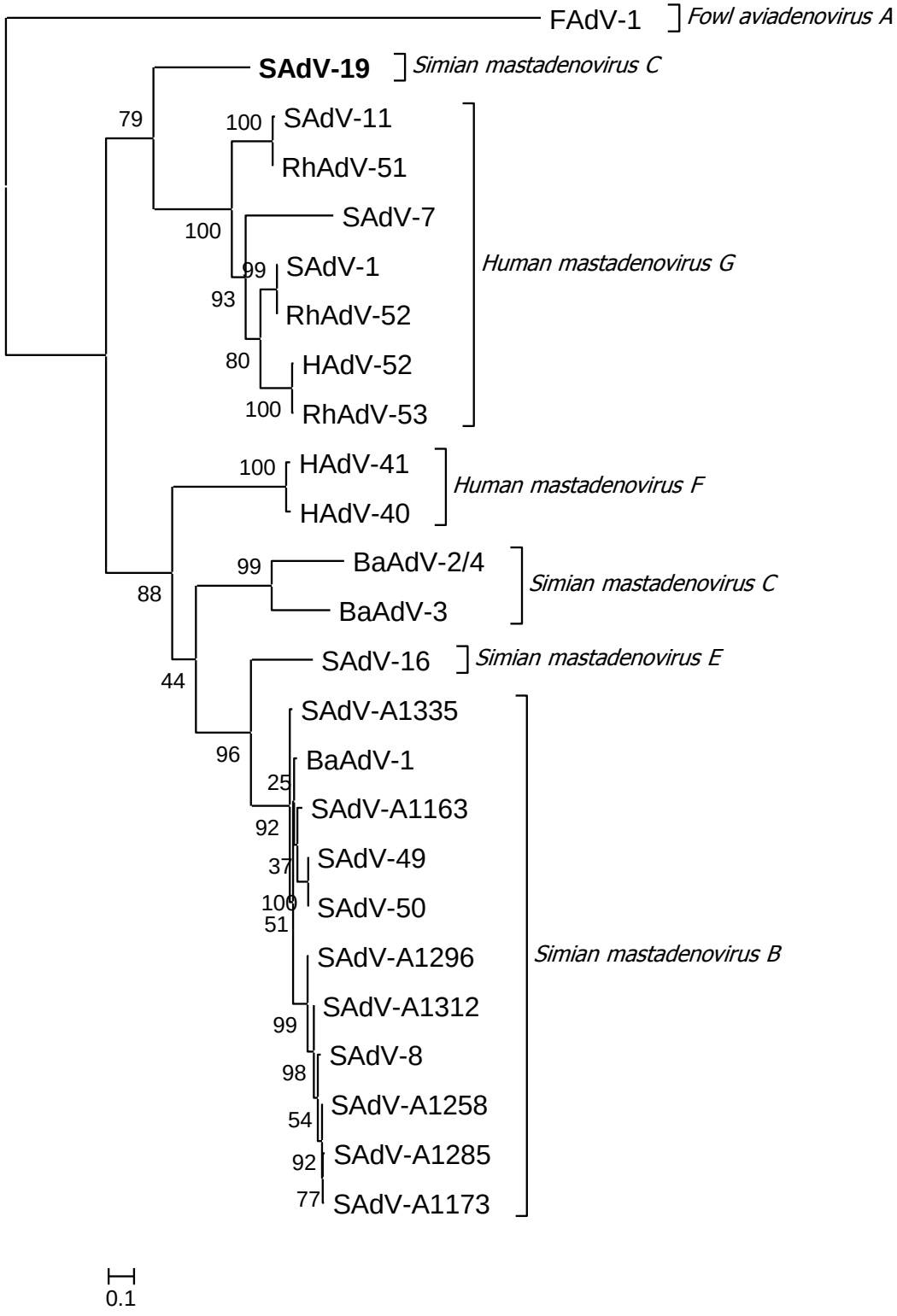


Figure 18. Phylogeny reconstruction (ML) based on full aa sequences from the fibre-1 of primate AdVs that have two fibre genes and fowl AdV-1 (FAdV-1) as an outgroup (LG model with gamma distribution option). Abbreviations: FAdV, fowl AdV; RhAdV, rhesus macaque AdV.

7.1.4 Receptor studies

Majority of HAdVs uses CAR as a primary adhesion receptor (Bergelson *et al.*, 1997; Roelvink *et al.*, 1998; Tomko *et al.*, 1997). Whilst it applies also to the fibre-2 of HAdV-40 and -41 (Roelvink *et al.*, 1998), data about the fibre-1 of the AdVs which have two fibre genes (members of the species HAdV-G and -F, SAdV-B, -C and -E), are quite limited. Most recently data about receptors of HAdV-52 were published by a Swedish laboratory (Lenman *et al.*, 2015), including for the first time data about the receptors of the fibre-1 (as all the earlier studied fibres were similar to fibre-2 proteins). Since we are collaborating with this group, I got a possibility to study also the receptors of the fibre-1 of some SAdVs. In this case we chose to study the members of the species HAdV-G (SAdV-1 and -7 sequenced earlier by Kovács *et al*., 2005 and Roy *et al*., 2011, respectively, and SAdV-11 sequenced by us), and also SAdV-19 (SAdV-C), since its fibre-1 was the most identical (46%) to those of members of HAdV-G. Two HAdVs that contain only one fibre gene but each uses different cellular receptor, CAR (HAdV-5) or sialic acid-containing glycans (HAdV-37), were used as controls. Additional control was the fibre-1 knob of HAdV-52, which binds sialic acid-containing glycans. Consequently, binding efficiency of the HAdV-5 was expected not to be influenced, whereas both HAdV-37 and HAdV-52 were predicted showing reduced binding to cells pre-treated with neuraminidase. Since only one experiment was made, and large deviations were observed, the results should be confirmed with additional experiments in the future. However, from the presented results (Figure 14), we can conclude that the fibre-1 of all the studied viruses uses sialic acid-containing glycans as the receptor on A549 cells, as well as the fibre-1 of HAdV-52 does (Lenman *et al.*, 2015). Since data about the binding of fibre-2 of SAdVs is still missing, an interesting study might be to investigate the binding properties of both fibre-1 and -2 of all the sequenced SAdVs in order to be able to compare them to HAdVs and see what the possibilities are to use them as vectors in gene therapy approaches.

7.2 Novel adenoviruses

7.2.1 Novel adenoviruses detected by PCR

From all the primate groups, we were especially interested in the NWM and prosimian AdVs, therefore every family (except the *Lepilemuridae*) of them was studied. The nested PCR reactions were performed with primer pairs targeting the IVa2 gene of mastadenoviruses, since these primers proved to be the best in our hands for the detection of primate AdVs. Unfortunately, the gained IVa2 sequences obviously could not be compared to the partial

sequences of DNA-dependent DNA polymerase and hexon genes of OWM and NWM AdVs available from other studies (Gál *et al.*, 2013; Hall *et al.*, 2012; Maluquer de Motes *et al.*, 2011; Wevers *et al.*, 2011). Since there were no prosimian AdVs discovered so far, we were especially curious in data we could obtain about their availability, diversity and evolution. Previously we observed that the only fully sequenced NWM AdV, TMAdV, appeared on the phylogenetic tree apart (more ancient) from all the known human and simian AdVs (Chen *et al.*, 2011). This is in correspondence with the taxonomic place of its host (*Callicebus cupreus*), which is clearly more ancient from all the apes and OWMs (Perelman *et al.*, 2011). Samples from five families of prosimians were analyzed in this study, but AdVs were detected in only one of them, *Lemuridae*. We suppose the reason for that is the small number of samples we collected from the other families, giving us lower probability to find the ones harbouring AdVs. Furthermore, novel NWM AdVs were found in both of the studied families of NWMs. In this case, we had mostly the samples again from one family (*Cebidae*), and only a small number from the other family (*Atelidae*). According to the theory on virus−host co-evolution, we expected the newly detected NWM AdVs to group with the TMAdV, and prosimian AdVs to group separately in a more ancient lineage. We also screened some species of OWMs in which AdVs have not been detected yet, and were able to detect AdV in mandrill. Unfortunately, the very short sequence we obtained was not clean enough to be used in the phylogenetic analysis, and should be cloned in future. From the ape samples, we detected AdV in Sumatran orangutan and in siamang. These AdVs were predicted to appear among the previously described ape AdVs. However, siamang AdV, being the first AdV detected in gibbon species, was expected to be more apart from the other ape AdVs.

Although the samples were collected from both captive and wild animals, and from different parts of Europe and Madagascar, we found the same AdV (in some cases only on aa level, and in some cases even on nt level) in several hosts from different areas (i.e., lemur AdV-2, marmoset AdV, tufted capuchin AdV-3, red-handed tamarin AdV-2; Table 11). Furthermore, the same AdV was found even in different species (i.e., lemur AdV-1, lemur AdV-2, marmoset AdV; Table 11). Anyway, all the cases where different species harbor the same AdV are not unexpected since these species are evolutionarily very close (Perelman *et al.*, 2011). Finally, the ability of TMAdV to infect even a human host (Chen *et al.*, 2011) indicates these viruses are sometimes able to cross the species barrier.

7.2.2 Virus isolation attempts

All the samples positive for AdVs were used in the AdV isolation trials on several different cell lines originating from human (A549, HEK293), monkey (Vero E6, MFC), hamster

(CHO-K1) or mouse (3T6, cmt93) tissues. However, despite multiple serial blind passages, isolation was unsuccessful, but the trials are still in progress on the MFC cells. There are several possible reasons for the unsuccessful isolation, but the most probable is the incompatibility of these cell lines with the viruses we are trying to isolate. The main problem for the NWM and prosimian AdVs can be the wrong species origin of the cells, or incompatible cell type. However, the OWM and ape samples were expected to propagate on Vero and/or human cell lines on which many other AdVs were already isolated before, but the problem here could be the very low amount of the virions in the fecal samples in which the AdVs were detected. The only NWM AdV described so far, the TMAdV, was isolated on A549 cell line (Chen *et al.*, 2011). This is very surprising since this cell line has a human origin, evolutionarily very distant from the NWMs. However, a mutation in the E1B 55K gene that results in a putative amino acid change, seen from the 7th passage onwards, is the most probable reason of the adaptation of TMAdV to A549 cells (Yu *et al.*, 2013). The virus with this mutation was also able to infect a variety of additional human and monkey cell lines (Yu *et al.*, 2013). Thus, additional efforts, including different methods, are extremely important for the possible isolation of these viruses, which would give us the possibility to study their properties and a potential for vectorizing.

7.2.3 Phylogeny inference

Three different AdVs were found in four black lemurs from the same area on Madagascar, but phylogenetic calculations indicate they are quite distant from each other (Figure 15): two AdVs (putatively named black lemur AdV-1 and -2) do not cluster with any other prosimian AdV, whereas the third one (putatively named black lemur AdV-3) is closer to other prosimian AdVs, and was detected in two individuals. Similar case was also seen in red lemurs, in which we found three different AdVs: two (lemur AdV-1 and -2) are quite close on the evolutionary tree, whereas one of them (putatively named red lemur AdV) is more apart from the other prosimian AdVs. Remaining prosimian AdVs are clustering together, except the Eastern lesser bamboo lemur AdV, which is not surprising since its host is also apart from other prosimians (Perelman *et al.*, 2011).

In NWM AdVs lineage we can see several clusters (Figure 15), which we compared to the host clusters (Perelman *et al.*, 2011). Red-faced spider monkey, detected by a colleague, was the only one positive for AdVs in the *Atelidae* family, and its AdV is well separated from all the other AdVs originating from the hosts of the other family (*Cebidae*). There is a cluster containing several tamarin and marmoset AdVs, and gray-bellied night monkey AdV. Tamarins and marmosets as hosts are also clustering together, whereas the gray-bellied night monkey is a bit further from them, but still not too far compared to other

NWMs (Perelman *et al.*, 2011). The other cluster contains tufted capuchin, common squirrel monkey and titi monkey AdVs. Tufted capuchins and common squirrel monkeys are evolutionarily very close, therefore their AdVs are logical to cluster together. Furthermore, three tufted capuchin AdVs clustered very close to each other, although the hosts were from different areas (Croatia and Hungary), confirming that they indeed are the real hosts of these AdVs. Common squirrel monkey AdV-2 and -3 are also from different locations in Hungary, but next to each other on the phylogenetic tree. Common squirrel monkey AdV-1 is separated from the other two AdVs, indicating that the common squirrel monkey might not be an original host for this AdV, but the virus may originate from some other species. The position of TMAdV on the tree together with tufted capuchin and common squirrel monkey AdVs is unexpected since its host is evolutionarily quite far from them (Perelman *et al.*, 2011). Anyway, since AdVs usually cause only mild disease or subclinical infections in their hosts, the case of high fatality rate from the TMAdV outbreak (83%; Chen *et al.*, 2011) in titi monkey colony indicates that titi monkey is not the native host for TMAdV. Furthermore, it was already seen that cross-species transmission can result in a pathogenicity usually not seen in the original host, as was hypothesized also about canine AdVs (Jánoska *et al.*, 2011; Kohl *et al.*, 2012; Vidovszky *et al.*, 2015). Another study with TMAdV showed that in common marmosets it did not cause severe infection and animals recovered very fast (Yu *et al.*, 2013), therefore all the data suggests that the origin of TMAdV is in a host from some other family, and not *Pitheciidae*. Another example of an AdV found in NWM but most probably with an origin in some other host is the pygmy marmoset AdV (Gál *et al.*, 2013), which was most closely related to canine and bat AdVs, equine AdV-1 and (very surprisingly) 100% identical in the published partial DNA-dependent DNA polymerase sequence with skunk AdV-1; all these viruses hypothesized as coevolved originally with bats (Kozak *et al.*, 2015; Vidovszky *et al.*, 2015). The infected pygmy marmoset also died because of the respiratory infection most probably caused by this AdV. All these data indicate that AdVs can indeed cross the species barriers among different nonhuman and human primate hosts, but sometimes even between primate and non-primate hosts.

The only newly detected OWM AdV, that from mandrill, was unfortunately not clean type, therefore it could not be included in our phylogenetic analysis. The most probable reason of bad sequence is the existence of at least two AdV types in the sample, both of which detected by our PCR primers. Molecular cloning, which would separate the types, would be the best approach to solve the case. Ape AdVs detected in this study appeared on the phylogenetic tree at the expected place (Figure 15). Sumatran orangutan AdV was placed among the HAdV species, closest to the HAdV-C cluster, to which already many ape AdVs were proposed to belong to (chimpanzee, bonobo and gorilla AdVs). Nevertheless, it does not group closely with any of the known AdVs. We would need longer or even full

genome sequences to decide if this AdV should belong to one of the already established HAdV species, or rather be a member of a new, ape AdV species. Several ape AdVs, previously proposed to belong to the species HAdV-C, appear on the phylogenetic trees well separated from the HAdVs, questioning their proposed membership in this species. Siamang AdV, the first representative of gibbon AdVs, found its place on the IVa2 tree among the OWM AdV species. Nevertheless, it stands alone, as expected, although the position closer to the ape and human AdVs would be more expected. Still, with gibbons being the most ancient apes, it is not surprising that the AdVs they can harbour are close to the AdVs from the OWMs. For now, this AdV would represent a new species, but additional AdVs from gibbons would help in this decision.

7.2.4 Red-handed tamarin adenovirus 1

Although the isolation of the RHTAdV-1 was unsuccessful, we managed to sequence the majority of its genome with the traditional (Sanger) method. Sequence of the ends of the genome is still missing, since these are, as expected, the most difficult to be gained by consensus PCR. The ITR primers used for the amplification of other SAdVs are probably not ideal for this virus, causing the problems with connecting the sequenced region (E1B 55K to UXP gene) with them. However, even the major sequenced part provides us with interesting information about the genes of this AdV: most of the genes are shorter than those of the TMAdV (Table 12), a few are similar, and only one is considerably longer (RID-α in the E3 region). The identities of the genes it shares with the TMAdV are not too high, ranging from 30 to 82%. The genome organization of the sequenced region is similar to that in other mastadenoviruses, with exception of the E3 region (Figure 16). There are two novel genes in the E3 region, but these do not represent additional genes, but rather a “replacement” of some genes seen in other SAdVs. In fact, it seems that one gene is actually missing, since this virus has five genes in the E3 region, instead of the usual six. Three of the genes observed in other primate AdVs (CR1-α, CR1-β and RID-β) are missing from the RHTAdV-1 genome. In the place of the CR1-α and -β there is one novel gene only, and another gene is in place of the RID-β. Since we were able to locate the UXP, we can conclude that there are no other genes downstream from the 14.7K and that we sequenced the full E3 region of this virus. Although the full genome of the TMAdV has been annotated and submitted to the GenBank, the published annotation of the E3 region is not correct, causing us some trouble in comparing the RHTAdV-1 genes with those of TMAdV. It seems that TMAdV also contains several (actually four) novel genes in the E3 region, which are different from our RHTAdV-1 genes. Similarly to the RHTAdV-1, TMAdV also contains the 12.5K, RID-α and 14.7K, but has different genes in the regions of CR1-α, CR1-β and RID-β. Interestingly, in the region of

the CR1-α and -β, TMAdV seems to have three novel genes (whereas RHTAdV-1 has only one), and novel gene in place of RID- β, containing all together seven genes in the E3 region. Therefore, the two NWM AdVs are different in the E3 region. It confirms the existence of variability seen in this region in other mastadenoviruses as well, and we may expect further possible differences in the E3 region of the other, not yet sequenced NWM AdVs and even more in the prosimian AdVs. We may also conclude that gaining the two CR1 homologues and the duplication of the RID-α gene most probably happened after the lineage of the OWMs had evolved as all OWM and ape (and human) AdVs contain CR1-α, CR1-β and RID-β, the homologue of RID-α, (except the deletion mutants SAdV-7 and -18, which seem to lack all E3 genes except 12.5K) (Figure 12). Unfortunately, the RHTAdV-1 sequence of the fibre gene, which can also give valuable information about the virus, is still missing. Phylogeny inference, based on the partial aa sequence of the IVa2 protein, placed the RHTAdV-1 among other NWM AdVs, as expected (Figure 15). However, phylogenetically it is quite distant from the TMAdV, explaining the differences seen in the genome sequence identity.

In this study we detected novel AdVs in apes, OWMs, NWMs and prosimians. NWM and prosimians AdVs clustered into two new lineages of primate AdVs. Majority of these AdVs confirms the theory on the virus−host co-evolution and gives us a broad picture about primate AdVs, representing other lineages beside the previously seen human, ape and OWM AdVs. Additionally, we proposed a novel species for the only fully sequenced NWM AdV (titi monkey AdV) to the ICTV, putatively named *Platyrrhini mastadenovirus A* (Figure 17). As already seen in other AdV lineages, crossing the species barrier is also possible during the AdV evolution, and we can speculate that some of the AdVs we studied here could be a result of such case as well. However, for more reliable conclusions regarding the molecular evolution of these viruses, more extensive genetic studies are planned already in near future.

8. New scientific results

1. Partial genome sequencing of the 14 non-sequenced, serotyped ATCC Old World monkey AdVs enabled their assignment to the previously established (*Human mastadenovirus G* and *Simian mastadenovirus A*) and recently proposed (*Simian mastadenovirus B* and *Simian mastadenovirus C*) or newly proposed (*Simian mastadenovirus D* to *G*) AdV species.

2. Six Old World monkey AdVs (SAdV-2, -8, -11, -16, -17 and -19) fully sequenced and molecularly characterized. For the first time, three fibre genes detected in AdVs.

3. For the first time in SAdVs, all the three exons of the so called U exon protein characterized.

4. An almost fully sequenced New World monkey AdV, which confirms the difference of the E3 region of Old World and New World monkey AdVs.

5. Achievement of the official establishment of two SAdV species, *Simian mastadenovirus B* and *Simian mastadenovirus C* in the ICTV. Proposal to ICTV for six SAdV species, *Simian mastadenovirus D* to *H,* and *Platyrrhini mastadenovirus A*.

6. First receptor binding studies with SAdVs show that fibre-1 knob of members of species *Human mastadenovirus G*, as well as that of SAdV-19, binds the sialic acid-containing glycans.

7. Novel AdVs detected in apes, Old World monkeys, New World monkeys, and for the first time in prosimians support the theory on virus−host co-evolution.

9. References

Abbink, P., Maxfield, L. F., Ng'ang'a, D., Borducchi, E. N., Iampietro, M. J., Bricault, C. A., Teigler, J. E., Blackmore, S., Parenteau, L., Wagh, K., Handley, S. A., Zhao, G., Virgin, H. W., Korber, B. & Barouch, D. H. (2015). Construction and evaluation of novel rhesus monkey adenovirus vaccine vectors. *J Virol* **89**, 1512-1522.

Abrescia, N. G., Cockburn, J. J., Grimes, J. M., Sutton, G. C., Diprose, J. M., Butcher, S. J., Fuller, S. D., San Martín, C., Burnett, R. M., Stuart, D. I., Bamford, D. H. & Bamford, J. K. (2004). Insights into assembly from structural analysis of bacteriophage PRD1. *Nature* **432**, 68-74.

Alonso-Padilla, J., Papp, T., Kaján, G. L., Benkő, M., Havenga, M., Lemckert, A., Harrach, B. & Baker, A. H. (2015). Development of novel adenoviral vectors to overcome challenges observed with HAdV-5 based constructs. *Mol Ther* **24**, 6-16.

Arnberg, N., Edlund, K., Kidd, A. H. & Wadell, G. (2000). Adenovirus type 37 uses sialic acid as a cellular receptor. *J Virol* **74**, 42-48.

Bailey, A. & Mautner, V. (1994). Phylogenetic relationships among adenovirus serotypes. *Virology* **205**, 438-452.

Balakrishnan, B. & Jayandharan, G. R. (2014). Basic biology of adeno-associated virus (AAV) vectors used in gene therapy. *Curr Gene Ther* **14**, 86-100.

Bányai, K., Esona, M. D., Liu, A., Wang, Y., Tu, X. & Jiang, B. (2010). Molecular detection of novel adenoviruses in fecal specimens of captive monkeys with diarrhea in China. *Vet Microbiol* **142**, 416-419.

Benkő, M. (2015). Adenoviruses: Pathogenesis. In Caplan, M., Mitchell, R., Bradshaw, R. & McManus, L. (eds) *Reference Module in Biomedical Sciences*: Elsevier. DOI: 10.1016/B978-0-12-801238-3.02526-5.

Benkő, M. & Harrach, B. (1998). A proposal for a new (third) genus within the family *Adenoviridae*. *Arch Virol* **143**, 829-837.

Benkő, M. & Harrach, B. (2003). Molecular evolution of adenoviruses. *Curr Top Microbiol Immunol* **272**, 3-35.

Benkő, M., Harrach, B. & Kremer, E. J. (2014). Do nonhuman primate or bat adenoviruses pose a risk for human health? *Future Microbiol* **9**, 269-272.

Benkő, M., Harrach, B. & Russell, W. C. (2000). Family *Adenoviridae*. In *Virus Taxonomy, Classification and Nomenclature of Viruses: Seventh Report of the International Committee on Taxonomy of Viruses*, pp. 227-238. Edited by M. H. V. Van Regenmortel, C. M. Fauquet, D. H. L. Bishop, E. B. Carstens, M. K. Estes, S. M. Lemon, J. Maniloff, M. A. Mayo, D. J. McGeoch, C. R. Pringle & R. B. Wickner. San Diego: Academic Press.

Bergelson, J. M., Cunningham, J. A., Droguett, G., Kurt-Jones, E. A., Krithivas, A., Hong, J. S., Horwitz, M. S., Crowell, R. L. & Finberg, R. W. (1997). Isolation of a common receptor for Coxsackie B viruses and adenoviruses 2 and 5. *Science* **275**, 1320-1323.

Berriman, M. & Rutherford, K. (2003). Viewing and annotating sequence data with Artemis. *Brief Bioinform* **4**, 124-132.

Bhat, R. A. & Thimmappaya, B. (1984). Adenovirus mutants with DNA sequence perturbations in the intragenic promoter of VAI RNA gene allow the enhanced transcription of VAII RNA gene in HeLa cells. *Nucleic Acids Res* **12**, 7377-7388.

Capone, S., D'Alise, A. M., Ammendola, V., Colloca, S., Cortese, R., Nicosia, A. & Folgori, A. (2013). Development of chimpanzee adenoviruses as vaccine vectors: challenges and successes emerging from clinical trials. *Expert Rev Vaccines* **12**, 379-393.

Chen, E. C., Yagi, S., Kelly, K. R., Mendoza, S. P., Tarara, R. P., Canfield, D. R., Maninger, N., Rosenthal, A., Spinner, A., Bales, K. L., Schnurr, D. P., Lerche, N. W. & Chiu, C. Y. (2011). Cross-species transmission of a novel adenovirus associated with a fulminant pneumonia outbreak in a new world monkey colony. *PLoS Pathog* **7**, e1002155.

Chiu, C. Y., Yagi, S., Lu, X., Yu, G., Chen, E. C., Liu, M., Dick, E. J., Carey, K. D., Erdman, D. D., Leland, M. M. & Patterson, J. L. (2013). A novel adenovirus species associated with an acute respiratory outbreak in a baboon colony and evidence of coincident human infection. *MBio* **4**, e00084.

Colby, W. W. & Shenk, T. (1981). Adenovirus type 5 virions can be assembled in vivo in the absence of detectable polypeptide IX. *J Virol* **39**, 977-980.

Colloca, S., Barnes, E., Folgori, A., Ammendola, V., Capone, S., Cirillo, A., Siani, L., Naddeo, M., Grazioli, F., Esposito, M., Ambrosio, M., Sparacino, A., Bartiromo, M., Meola, A., Smith, K., Kurioka, A., O'Hara, G., Ewer, K., Anagnostou, N., Bliss, C., Hill, A., Traboni, C., Klenerman, P., Cortese, R. & Nicosia, A. (2012). Vaccine vectors derived from a large collection of simian adenoviruses induce potent cellular immunity across multiple species. *Science Translational Medicine* **4**, 115ra2.

Dán, A., Ruzsics, Z., Russell, W. C., Benkő, M. & Harrach, B. (1998). Analysis of the hexon gene sequence of bovine adenovirus type 4 provides further support for a new adenovirus genus (*Atadenovirus*). *J Gen Virol* **79,** 1453-1460.

Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. (2011). ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* **27**, 1164-1165.

Davison, A. J., Benkő, M. & Harrach, B. (2003). Genetic content and evolution of adenoviruses. *J Gen Virol* **84**, 2895-2908.

Davison, A. J., Telford, E. A., Watson, M. S., McBride, K. & Mautner, V. (1993). The DNA sequence of adenovirus type 40. *J Mol Biol* **234**, 1308-1316.

Dehghan, S., Seto, J., Jones, M. S., Dyer, D. W., Chodosh, J. & Seto, D. (2013a). Simian adenovirus type 35 has a recombinant genome comprising human and simian adenovirus sequences, which predicts its potential emergence as a human respiratory pathogen. *Virology* **447**, 265-273.

Dehghan, S., Seto, J., Liu, E. B., Walsh, M. P., Dyer, D. W., Chodosh, J. & Seto, D. (2013b). Computational analysis of four human adenovirus type 4 genomes reveals molecular evolution through two interspecies recombination events. *Virology* **443**, 197-207.

Diouri, M., Keyvani-Amineh, H., Geoghegan, K. F. & Weber, J. M. (1996). Cleavage efficiency by adenovirus protease is site-dependent. *J Biol Chem* **271**, 32511-32514.

Doszpoly, A., Wellehan, J., Childress, A., Tarján, Z., Kovács, E., Harrach, B. & Benkő, M. (2013). Partial characterization of a new adenovirus lineage discovered in testudinoid turtles. *Infection Genetics and Evolution* **17**, 106-112.

Epstein, L. H. & Young, C. S. H. (1991). Adenovirus homologous recombination does not require expression of the immediate-early E1a gene. *J Virol* **65**, 4475-4479.

Farina, S. F., Gao, G. P., Xiang, Z. Q., Rux, J. J., Burnett, R. M., Alvira, M. R., Marsh, J., Ertl, H. C. & Wilson, J. M. (2001). Replication-defective vector based on a chimpanzee adenovirus. *J Virol* **75**, 11603-11613.

Favier, A. L., Schoehn, G., Jaquinod, M., Harsi, C. & Chroboczek, J. (2002). Structural studies of human enteric adenovirus type 41. *Virology* **293**, 75-85.

Foytich, K. R., Deshazer, G., Esona, M. D., Liu, A., Wang, Y., Tu, X. & Jiang, B. (2014). Identification of new provisional simian adenovirus species from captive monkeys, China. *Emerg Infect Dis* **20**, 1758-1759.

Fuentes-Marins, R., Rodriguez, A. R., Kalter, S. S., Hellman, A. & Crandell, R. A. (1963). Isolation of enteroviruses from the normal baboon (*Papio doguera*). *J Bacteriol* **85**, 1045-1050.

Gelderblom, H. & Maichle-Lauppe, I. (1982). The fibres of fowl adenoviruses. *Arch Virol* **72**, 289-298.

Griffin, B. D. & Nagy, E. (2011). Coding potential and transcript analysis of fowl adenovirus 4: insight into upstream ORFs as common sequence features in adenoviral transcripts. *J Gen Virol* **92**, 1260-1272.

Guindon, S. & Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* **52**, 696-704.

Gál, J., Hornyák, Á., Mándoki, M., Bakonyi, T., Balka, G., Szeredi, L., Marosán, M., Ludányi, T., Forgách, P., Sós, E., Demeter, Z. & Farkas, S. L. (2013). Novel mastadenovirus infection and clinical disease in a pygmy marmoset (*Callithrix* [*Cebuella*] *pygmaea*). *Vet Microbiol* **167**, 695-699.

Hall, N. H., Archer, L. L., Childress, A. L. & Wellehan, J. F. (2012). Identification of a novel adenovirus in a cotton-top tamarin (*Saguinus oedipus*). *J Vet Diagn Invest* **24**, 359-363.

Harrach, B. (2014). Adenoviruses: General features. In Caplan, M., Mitchell, R., Bradshaw, R. & McManus L. (eds) *Reference Module in Biomedical Sciences*: Elsevier. DOI: 10.1016/B978-0-12-801238-3.02523-X

Harrach, B., Benkő, M., Both, G. W., Brown, M., Davison, A. J., Echavarria, M., Hess, M., Jones, M. S., Kajon, A., Lehmkuhl, H. D., Mautner, V., Mittal, S. K. & Wadell, G. (2011). Family *Adenoviridae*. In *Virus Taxonomy, Ninth Report of the International Committee on Taxonomy of Viruses*, pp. 125-141. Edited by A. M. Q. King, M. J. Adams, E. Lefkowitz & E. B. Carstens. Oxford: Elsevier.

Harrach, B. & Podgorski I.I. (2014). Jó vírus – rossz vírus? (Gyógyító vírusok). In: Janda, T. (ed.) *III. ATK Tudományos Nap: A jövőt alakító Európai Uniós agrárkutatási programok az MTA ATK-ban.* 2014.11.13 Martonvásár: MTA ATK, pp. 30-35.

Hoffert, W. R., Bates, M. E. & Cheever, F. S. (1958). Study of enteric viruses of simian origin. *Am J Hyg* **68**, 15-30.

Hull, R. N., Johnson, I. S., Culbertson, C. G., Reimer, C. B. & Wright, H. F. (1965). Oncogenicity of the simian adenoviruses. *Science* **150**, 1044-1046.

Hull, R. N. & Minner, J. R. (1957). New viral agents recovered from tissue cultures of monkey kidney cells. II. Problems of isolation and identification. *Ann N Y Acad Sci* **67**, 413-423.

Hull, R. N., Minner, J. R. & Mascoli, C. C. (1958). New viral agents recovered from tissue cultures of monkey kidney cells. III. Recovery of additional agents both from cultures of monkey tissues and directly from tissues and excreta. *Am J Hyg* **68**, 31-44.

Hull, R. N., Minner, J. R. & Smith, J. W. (1956). New viral agents recovered from tissue cultures of monkey kidney cells. I. Origin and properties of cytopathogenic agents S.V.1, S.V.2, S.V.4, S.V.5, S.V.6, S.V.11, S.V.12 and S.V.15. *Am J Hyg* **63**, 204-215.

Jones, M. S., Harrach, B., Ganac, R. D., Gozum, M. M., Dela Cruz, W. P., Riedel, B., Pan, C., Delwart, E. L. & Schnurr, D. P. (2007). New adenovirus species found in a patient presenting with gastroenteritis. *J Virol* **81**, 5978-5984.

Jánoska, M., Vidovszky, M., Molnár, V., Liptovszky, M., Harrach, B. & Benkő, M. (2011). Novel adenoviruses and herpesviruses detected in bats. *Vet J* **189**, 118-121.

Kaján, G. L., Davison, A. J., Palya, V., Harrach, B. & Benkő, M. (2012). Genome sequence of a waterfowl aviadenovirus, goose adenovirus 4. *J Gen Virol* **93**, 2457-2465.

Kaján, G. L., Stefancsik, R., Ursu, K., Palya, V. & Benkő, M. (2010). The first complete genome sequence of a non-chicken aviadenovirus, proposed to be turkey adenovirus 1. *Virus Res* **153**, 226-233.

Kidd, A. H., Chroboczek, J., Cusack, S. & Ruigrok, R. W. (1993). Adenovirus type 40 virions contain two distinct fibres. *Virology* **192**, 73-84.

Kidd, A. H., Garwicz, D. & Oberg, M. (1995). Human and simian adenoviruses: phylogenetic inferences from analysis of VA RNA genes. *Virology* **207**, 32-45.

Kim, C. S., Sueltenfuss, E. S. & Kalter, S. S. (1967). Isolation and characterization of simian adenoviruses isolated in association with an outbreak of pneumoenteritis in vervet monkeys (*Cercopithecus aethiops*). *J Infect Dis* **117**, 292-300.

Kimelman, D., Miller, J. S., Porter, D. & Roberts, B. E. (1985). E1a regions of the human adenoviruses and of the highly oncogenic simian adenovirus 7 are closely related. *J Virol* **53**, 399-409.

Kiss, I., Matiz, K., Allard, A., Wadell, G. & Benkő, M. (1996). Detection of homologous DNA sequences in animal adenoviruses by polymerase chain reaction. *Acta Vet Hung* **44**, 243-251.

Kohl, C., Vidovszky, M. Z., Mühldorfer, K., Dabrowski, P. W., Radonić, A., Nitsche, A., Wibbelt, G., Kurth, A. & Harrach, B. (2012). Genome analysis of bat adenovirus 2: indications of interspecies transmission. *J Virol* **86**, 1888-1892.

Kovács, E. R. & Benkő, M. (2009). Confirmation of a novel siadenovirus species detected in raptors: partial sequence and phylogenetic analysis. *Virus Res* **140**, 64-70.

Kovács, G. M., Davison, A. J., Zakhartchouk, A. N. & Harrach, B. (2004). Analysis of the first complete genome sequence of an Old World monkey adenovirus reveals a lineage distinct from the six human adenovirus species. *J Gen Virol* **85**, 2799-2807.

Kovács, G. M., Harrach, B., Zakhartchouk, A. N. & Davison, A. J. (2005). Complete genome sequence of simian adenovirus 1: an Old World monkey adenovirus with two fibre genes. *J Gen Virol* **86**, 1681-1686.

Kozak, R. A., Ackford, J. G., Slaine, P., Li, A., Carman, S., Campbell, D., Welch, M. K., Kropinski, A. M. & Nagy, É. (2015). Characterization of a novel adenovirus isolated from a skunk. *Virology* **485**, 16-24.

Larsson, S., Svensson, C. & Akusjärvi, G. (1986). Characterization of a low-molecular-weight virus-associated (VA) RNA encoded by simian adenovirus type 7 which functionally can substitute for adenovirus type 5 VA RNAI. *J Virol* **60**, 635-644.

Lenman, A., Liaci, A. M., Liu, Y., Årdahl, C., Rajan, A., Nilsson, E., Bradford, W., Kaeshammer, L., Jones, M. S., Frängsmyr, L., Feizi, T., Stehle, T. & Arnberg, N. (2015). Human adenovirus 52 uses sialic acid-containing glycoproteins and the coxsackie and adenovirus receptor for binding to target cells. *PLoS Pathog* **11**, e1004657.

Li, Q. G. & Wadell, G. (1988). The degree of genetic variability among adenovirus type 4 strains isolated from man and chimpanzee. *Arch Virol* **101**, 65-77.

Liu, G. Q., Babiss, L. E., Volkert, F. C., Young, C. S. & Ginsberg, H. S. (1985). A thermolabile mutant of adenovirus 5 resulting from a substitution mutation in the protein VIII gene. *J Virol* **53**, 920-925.

Liu, T. C., Hallden, G., Wang, Y., Brooks, G., Francis, J., Lemoine, N. & Kim, D. (2004). An E1B-19 kDa gene deletion mutant adenovirus demonstrates tumor necrosis factor-enhanced cancer selectivity and enhanced oncolytic potency. *Mol Ther* **9**, 786-803.

Lole, K. S., Bollinger, R. C., Paranjape, R. S., Gadkari, D., Kulkarni, S. S., Novak, N. G., Ingersoll, R., Sheppard, H. W. & Ray, S. C. (1999). Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J Virol* **73**, 152-160.

Lopez-Gordo, E., Podgorski, I. I., Downes, N. & Alemany, R. (2014). Circumventing antivector immunity: potential use of nonhuman adenoviral vectors. *Hum Gene Ther* **25**, 285-300.

Lu, J., Wang, Q., Wang, H., Li, G. & Gao, G. (2011). Molecular characterization of adenoviruses in fecal samples of captively bred rhesus macaques in China. *Vet Microbiol* **149**, 461-466.

Ma, H. C. & Hearing, P. (2011). Adenovirus structural protein IIIa is involved in the serotype specificity of viral DNA packaging. *J Virol* **85**, 7849-7855.

Ma, Y. & Mathews, M. B. (1996). Structure, function, and evolution of adenovirus-associated RNA: a phylogenetic approach. *J Virol* **70**, 5083-5099.

Malherbe, H. & Harwin, R. (1963). The cytopathic effects of vervet monkey viruses. *S Afr Med J* **37**, 407-411.

Malouli, D., Howell, G. L., Legasse, A. W., Kahl, C., Axthelm, M. K., Hansen, S. G. & Früh, K. (2014). Full genome sequence analysis of a novel adenovirus of rhesus macaque origin indicates a new simian adenovirus type and species. *Virol Rep* **3-4**, 18-29.

Maluquer de Motes, C., Hundesa, A., Almeida, F. C., Bofill-Mas, S. & Girones, R. (2011). Isolation of a novel monkey adenovirus reveals a new phylogenetic clade in the evolutionary history of simian adenoviruses. *Virol J* **8**, 125.

Mangel, W. F., McGrath, W. J., Toledo, D. L. & Anderson, C. W. (1993). Viral DNA and a viral peptide can act as cofactors of adenovirus virion proteinase activity. *Nature* **361**, 274-275.

Mangel, W. F., Toledo, D. L., Brown, M. T., Martin, J. H. & McGrath, W. J. (1996). Characterization of three components of human adenovirus proteinase activity in vitro. *J Biol Chem* **271**, 536-543.

Marek, A., Ballmann, M. Z., Kosiol, C., Harrach, B., Schlötterer, C. & Hess, M. (2014a). Whole-genome sequences of two turkey 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.

Marek, A., Kaján, G. L., Kosiol, C., Harrach, B., Schlötterer, C. & Hess, M. (2014b). Complete genome sequences of pigeon adenovirus 1 and duck adenovirus 2 extend the number of species within the genus *Aviadenovirus*. *Virology* **462-463**, 107-114.

Marek, A., Kosiol, C., Harrach, B., Kaján, G. L., Schlötterer, C. & Hess, M. (2013). The first whole genome sequence of a *Fowl adenovirus B* strain enables interspecies comparisons within the genus *Aviadenovirus*. *Vet Microbiol* **166**, 250-256.

Mathews, M. B. & Shenk, T. (1991). Adenovirus virus-associated RNA and translation control. *J Virol* **65**, 5657-5662.

Mennechet, F. J., Tran, T. T., Eichholz, K., van de Perre, P. & Kremer, E. J. (2015). Ebola virus vaccine: benefit and risks of adenovirus-based vectors. *Expert Rev Vaccines* **14**, 1471-1478.

Moyer, C. L., Wiethoff, C. M., Maier, O., Smith, J. G. & Nemerow, G. R. (2011). Functional genetic and biophysical analyses of membrane disruption by human adenovirus. *J Virol* **85**, 2631-2641.

Pantó, L., Podgorski, I. I., Jánoska, M., Márkó, O. & Harrach, B. (2015). Taxonomy proposal for Old World monkey adenoviruses: characterisation of several non-human, non-ape primate adenovirus lineages. *Arch Virol* **160**, 3165-3177.

Parks, R. J. (2005). Adenovirus protein IX: a new look at an old protein. *Mol Ther* **11**, 19-25.

Pénzes, J. J., Menéndez-Conejero, R., Condezo, G. N., Ball, I., Papp, T., Doszpoly, A., Paradela, A., Pérez-Berná, A. J., López-Sanz, M., Nguyen, T. H., van Raaij, M. J., Marschang, R. E., Harrach, B., Benkő, M. & San Martín, C. (2014). Molecular characterization of a lizard adenovirus reveals the first atadenovirus with two fibre genes and the first adenovirus with either one short or three long fibres per penton. *J Virol* **88**, 11304-11314.

Pereira, H. G., Valentine, R. C. & Russell, W. C. (1968). Crystallization of an adenovirus protein (the hexon). *Nature* **219**, 946-947.

Perelman, P., Johnson, W. E., Roos, C., Seuánez, H. N., Horvath, J. E., Moreira, M. A., Kessing, B., Pontius, J., Roelke, M., Rumpler, Y., Schneider, M. P., Silva, A., O'Brien, S. J. & Pecon-Slattery, J. (2011). A molecular phylogeny of living primates. *PLoS Genet* **7**, e1001342.

Peruzzi, D., Dharmapuri, S., Cirillo, A., Bruni, B. E., Nicosia, A., Cortese, R., Colloca, S., Ciliberto, G., La Monica, N. & Aurisicchio, L. (2009). A novel chimpanzee serotype-based adenoviral vector as delivery tool for cancer vaccines. *Vaccine* **27**, 1293-1300.

Podgorski I. I. & Harrach, B. (2015). 2 new species in the genus *Mastadenovirus*. Proposal for ICTV. https://talk.ictvonline.org/files/proposals/animal_dna_viruses_and_retroviruses/m/animal_dna_ec_approved/5743.

Podgorski, I. I., Pantó, L., Papp, T., Harrach, B. & Benkő, M. (2016). Genome analysis of four Old World monkey adenoviruses supports the proposed species classification of primate adenoviruses and reveals signs of possible homologous recombination. *J Gen Virol*. Published online: 24/03/2016 (in press). DOI: 10.1099/jgv.0.000465. [Epub ahead of print]

Pring-Akerblom, P., Trijssenaar, F. E. & Adrian, T. (1995). Sequence characterization and comparison of human adenovirus subgenus B and E hexons. *Virology* **212**, 232-236.

Purkayastha, A., Ditty, S. E., Su, J., McGraw, J., Hadfield, T. L., Tibbetts, C. & Seto, D. (2005). Genomic and bioinformatics analysis of HAdV-4, a human adenovirus causing acute respiratory disease: implications for gene therapy and vaccine vector development. *J Virol* **79**, 2559-2572.

Rapoza, N. P. (1967). A classification of simian adenoviruses based on hemagglutination. *Am J Epidemiol* **86**, 736-745.

Roelvink, P. W., Lizonova, A., Lee, J. G., Li, Y., Bergelson, J. M., Finberg, R. W., Brough, D. E., Kovesdi, I. & Wickham, T. J. (1998). The coxsackievirus-adenovirus receptor protein can function as a cellular attachment protein for adenovirus serotypes from subgroups A, C, D, E, and F. *J Virol* **72**, 7909-7915.

Rogers, N. G., Basnight, M., Gibbs, C. J. & Gajdusek, D. C. (1967). Latent viruses in chimpanzees with experimental kuru. *Nature* **216**, 446-449.

Rowe, W. P., Hartley, J. W. & Huebner, R. J. (1956). Additional serotypes of the APC virus group. *Proc Soc Exp Biol Med* **91**, 260-262.

Rowe, W. P., Huebner, R. J., Gilmore, L. K., Parrott, R. H. & Ward, T. G. (1953). Isolation of a cytopathogenic agent from human adenoids undergoing spontaneous degeneration in tissue culture. *Proc Soc Exp Biol Med* **84**, 570-573.

Roy, S., Clawson, D. S., Adam, V. S., Medina, A. & Wilson, J. M. (2011). Construction of gene transfer vectors based on simian adenovirus 7. *J Gen Virol* **92**, 1749-1753.

Roy, S., Gao, G., Clawson, D. S., Vandenberghe, L. H., Farina, S. F. & Wilson, J. M. (2004). Complete nucleotide sequences and genome organization of four chimpanzee adenoviruses. *Virology* **324**, 361-372.

Roy, S., Sandhu, A., Medina, A., Clawson, D. S. & Wilson, J. M. (2012). Adenoviruses in fecal samples from asymptomatic rhesus macaques, United States. *Emerg Infect Dis* **18**, 1081-1088.

Roy, S., Vandenberghe, L. H., Kryazhimskiy, S., Grant, R., Calcedo, R., Yuan, X., Keough, M., Sandhu, A., Wang, Q., Medina-Jaszek, C. A., Plotkin, J. B. & Wilson, J. M. (2009). Isolation and characterization of adenoviruses persistently shed from the gastrointestinal tract of non-human primates. *PLoS Pathog* **5**, e1000503.

San Martín, C. (2012). Latest insights on adenovirus structure and assembly. *Viruses* **4**, 847-877.

San Martín, C., Glasgow, J. N., Borovjagin, A., Beatty, M. S., Kashentseva, E. A., Curiel, D. T., Marabini, R. & Dmitriev, I. P. (2008). Localization of the N-terminus of minor coat protein IIIa in the adenovirus capsid. *J Mol Biol* **383**, 923-934.

Sauthoff, H., Heitner, S., Rom, W. N. & Hay, J. G. (2000). Deletion of the adenoviral E1b-19kD gene enhances tumor cell killing of a replicating adenoviral vector. *Hum Gene Ther* **11**, 379-388.

Schmidt, M., Grot, E., Cervenka P., Wainer, S., Buck, C. & Chiorini, J. A. (2006). Identification and characterization of novel adeno-associated virus isolates in ATCC virus stocks. *J Virol* **80**, 5082-5085.

Schreiner, S., Martinez, R., Groitl, P., Rayne, F., Vaillant, R., Wimmer, P., Bossis, G., Sternsdorf, T., Marcinowski, L., Ruzsics, Z., Dobner, T. & Wodrich, H. (2012). Transcriptional activation of the adenoviral genome is mediated by capsid protein VI. *PLoS Pathog* **8**, e1002549.

Shroyer, E. L., Kelley, S. T., Taylor, P. C., Vanderloo, P. & Lester, T. L. (1979). Three serologic types of adenovirus infections of owl monkeys. *Am J Vet Res* **40**, 532-536.

Singh, M., Shmulevitz, M. & Tikoo, S. K. (2005). A newly identified interaction between IVa2 and pVIII proteins during porcine adenovirus type 3 infection. *Virology* **336**, 60-69.

Strunze, S., Engelke, M. F., Wang, I. H., Puntener, D., Boucke, K., Schleich, S., Way, M., Schoenenberger, P., Burckhardt, C. J. & Greber, U. F. (2011). Kinesin-1-mediated capsid disassembly and disruption of the nuclear pore complex promote virus infection. *Cell Host Microbe* **10**, 210-223.

Subramanian, T., Tarodi, B. & Chinnadurai, G. (1995). Functional similarity between adenovirus E1B 19-kDa protein and proteins encoded by Bcl-2 proto-oncogene and Epstein-Barr virus BHRF1 gene. *Curr Top Microbiol Immunol* **199** (Pt 1), 153-161.

Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* **30**, 2725-2729.

Tarján, Z. L., Pénzes, J. J., Tóth, R. P. & Benkő, M. (2014). First detection of circovirus-like sequences in amphibians and novel putative circoviruses in fishes. *Acta Vet Hung* **62**, 134-144.

Thimmappaya, B., Weinberger, C., Schneider, R. J. & Shenk, T. (1982). Adenovirus VAI RNA is required for efficient translation of viral mRNAs at late times after infection. *Cell* **31**, 543-551.

To, K. K., Tse, H., Chan, W. M., Choi, G. K., Zhang, A. J., Sridhar, S., Wong, S. C., Chan, J. F., Chan, A. S., Woo, P. C., Lau, S. K., Lo, J. Y., Chan, K. H., Cheng, V. C. & Yuen, K. Y. (2014). A novel psittacine adenovirus identified during an outbreak of avian chlamydiosis and human psittacosis: zoonosis associated with virus-bacterium coinfection in birds. *PLoS Negl Trop Dis* **8**, e3318.

Tollefson, A. E., Ying, B., Doronin, K., Sidor, P. D. & Wold, W. S. (2007). Identification of a new human adenovirus protein encoded by a novel late I-strand transcription unit. *J Virol* **81**, 12918-12926.

Tomko, R. P., Xu, R. & Philipson, L. (1997). HCAR and MCAR: the human and mouse cellular receptors for subgroup C adenoviruses and group B coxsackieviruses. *Proc Natl Acad Sci U S A* **94**, 3352-3356.

Vachon, V. K. & Conn, G. L. (2016). Adenovirus VA RNA: An essential pro-viral non-coding RNA. *Virus Res* **212**, 39-52.

Vasileva, V. A., Ivanov, M. T., Rumel, N. B., D'Yachenko, A. G., Kakubava, V. V. & Danelyan, G. A. (1978). Isolation and biological characterization of an adenovirus of rhesus macaques. *Acta Biol Med Ger* **37**, 1281-1287.

Vidovszky, M., Kohl, C., Boldogh, S., Görföl, T., Wibbelt, G., Kurth, A. & Harrach, B. (2015). Random sampling of the Central European bat fauna reveals the existence of numerous hitherto unknown adenoviruses. *Acta Vet Hung* **63**, 508-525.

Wadell, G., Mei, Y. F., Segerman, A., Skog, J. & Lindman, K. (2002). Viral vector for gene therapy. Patent WO 2002053759 A1.

Walsh, M. P., Chintakuntlawar, A., Robinson, C., Madisch, I., Harrach, B., Hudson, N. R., Torres, S., Schnurr, D., Heim, A., Chodosh, J., Seto, D. & Jones, M. S. (2009). Evidence of molecular evolution driven by recombination events influencing tropism and virulence in a novel human adenovirus that causes epidemic keratoconjunctivitis. *PLoS ONE* **4** (6) e5635.

Walsh, M. P., Seto, J., Liu, E. B., Dehghan, S., Hudson, N. R., Lukashev, A. N., Ivanova, O., Chodosh, J., Dyer, D. W., Jones, M. S. & Seto, D. (2011). Computational analysis of two species C human adenoviruses provides evidence of a novel virus. *J Clin Microbiol* **49**, 3482-3490.

Wang, Y., Tu, X., Humphrey, C., McClure, H., Jiang, X., Qin, C., Glass, R. I. & Jiang, B. (2007). Detection of viral agents in fecal specimens of monkeys with diarrhea. *J Med Primatol* **36**, 101-107.

Wellehan, J. F., Johnson, A. J., Harrach, B., Benkő, M., Pessier, A. P., Johnson, C. M., Garner, M. M., Childress, A. & Jacobson, E. R. (2004). Detection and analysis of six lizard adenoviruses by consensus primer PCR provides further evidence of a reptilian origin for the atadenoviruses. *J Virol* **78**, 13366-13369.

Wevers, D., Leendertz, F. H., Scuda, N., Boesch, C., Robbins, M. M., Head, J., Ludwig, C., Kühn, J. & Ehlers, B. (2010). A novel adenovirus of Western lowland gorillas (*Gorilla gorilla gorilla*). *Virol J* **7**, 303.

Wevers, D., Metzger, S., Babweteera, F., Bieberbach, M., Boesch, C., Cameron, K., Couacy-Hymann, E., Cranfield, M., Gray, M., Harris, L. A., Head, J., Jeffery, K., Knauf, S., Lankester, F., Leendertz, S. A., Lonsdorf, E., Mugisha, L., Nitsche, A., Reed, P., Robbins, M., Travis, D. A., Zommers, Z., Leendertz, F. H. & Ehlers, B. (2011). Novel adenoviruses in wild primates: a high level of genetic diversity and evidence of zoonotic transmissions. *J Virol* **85**, 10774-10784.

White, E. (2001). Regulation of the cell cycle and apoptosis by the oncogenes of adenovirus. *Oncogene* **20**, 7836-7846.

Wickham, T. J., Mathias, P., Cheresh, D. A. & Nemerow, G. R. (1993). Integrins alpha v beta 3 and alpha v beta 5 promote adenovirus internalization but not virus attachment. *Cell* **73**, 309-319.

Wodrich, H., Guan, T., Cingolani, G., Von Seggern, D., Nemerow, G. & Gerace, L. (2003). Switch from capsid protein import to adenovirus assembly by cleavage of nuclear transport signals. *EMBO J* **22**, 6245-6255.

Wodrich, H., Henaff, D., Jammart, B., Segura-Morales, C., Seelmeir, S., Coux, O., Ruzsics, Z., Wiethoff, C. M. & Kremer, E. J. (2010). A capsid-encoded PPxY-motif facilitates adenovirus entry. *PLoS Pathog* **6**, e1000808.

Wu, E., Pache, L., Von Seggern, D. J., Mullen, T. M., Mikyas, Y., Stewart, P. L. & Nemerow, G. R. (2003). Flexibility of the adenovirus fibre is required for efficient receptor interaction. *J Virol* **77**, 7225-7235.

Xiang, Z., Gao, G., Reyes-Sandoval, A., Cohen, C. J., Li, Y., Bergelson, J. M., Wilson, J. M. & Ertl, H. C. (2002). Novel, chimpanzee serotype 68-based adenoviral vaccine carrier for induction of antibodies to a transgene product. *J Virol* **76**, 2667-2675.

Yu, G., Yagi, S., Carrion, R., Chen, E. C., Liu, M., Brasky, K. M., Lanford, R. E., Kelly, K. R., Bales, K. L., Schnurr, D. P., Canfield, D. R., Patterson, J. L. & Chiu, C. Y. (2013). Experimental cross-species infection of common marmosets by titi monkey adenovirus. *PLoS One* **8**, e68558.

Zhao, J., Zhong, Q., Zhao, Y., Hu, Y. X. & Zhang, G. Z. (2015). Pathogenicity and complete genome characterization of fowl adenoviruses isolated from chickens associated with inclusion body hepatitis and hydropericardium syndrome in China. *PLoS One* **10**, e0133073.

Zhou, C., Tian, H., Wang, X., Liu, W., Yang, S., Shen, Q., Wang, Y., Ni, B., Chen, S., Fu, X., Fei, R. & Zhang, W. (2014). The genome sequence of a novel simian adenovirus in a chimpanzee reveals a close relationship to human adenoviruses. *Arch Virol* **159**, 1765-1770.

Zöller, M., Mätz-Rensing, K. & Kaup, F. J. (2008). Adenoviral hepatitis in a SIV-infected rhesus monkey

(*Macaca mulatta*). *J Med Primatol* **37**, 184-187.

10. Publications

1. Lopez-Gordo, E., Podgorski, I. I., Downes, N. & Alemany, R. (2014). Circumventing antivector immunity: potential use of nonhuman adenoviral vectors. *Hum Gene Ther* **25**, 285-300. IF: 3.755

2. Pantó, L.*, Podgorski, I. I.*, Jánoska, M., Márkó, O. & Harrach, B. (2015). Taxonomy proposal for Old World monkey adenoviruses: characterisation of several non-human, non-ape primate adenovirus lineages. *Arch Virol* **160**, 3165-3177.

IF: 2.390

*L. Panto and I. I. Podgorski contirbuted equally to this work.

3. Podgorski, I. I., Pantó, L., Papp, T., Harrach, B. & Benkő, M. (2016). Genome analysis of four Old World monkey adenoviruses supports the proposed species classification of primate adenoviruses and reveals signs of possible homologous recombination. *J Gen Virol*. Published online: 24/03/2016 (in press). DOI: 10.1099/jgv.0.000465.

IF: 3.183

11. Acknowledgements

The very first people I would mention here are dr. Balázs Harrach and dr. Mária Benkő, because without their guidance and support I wouldn't be in this life period. I've met many supervisors in the last few years, and learned indubitably that Balázs and Mari are the most encouraging and supporting people whose help was priceless in my early scientific career. I appreciate all the advices I was getting from them for both the private and professional life, making the living and studying in a foreign country much easier.

Next to them stand the people who are under their guidance in the lab, and who helped me with both the work and the fun times in Budapest and abroad. All these people taught me most of the things I know today about the lab work and I am very grateful for that. Special thanks to dr. Tibor Papp who provided great help vitally for the start of the living in Hungary, and also for the help in the projects I've been working on. Thanks to Mónika Ballmann who was my teacher in the first months and helped me to start with my projects. Thanks to Laura Pantó who worked with the simian AdVs prior to my arrival to the lab, and for giving me great basis to continue working on. Thanks to dr. Győző Kaján for his expertise in any problem I had during my stay in Budapest and Sweden. Thanks to Zoltán László Tarján, dr. Andor Doszpoly and dr. Judit Pénzes, my temporary "lab roommates" who were always there to help with my Hungarian language issues. Special thanks to Judit for the fun times we had listening to her concerts. I also thank to dr. Márton Vidovszky and Kinga Böszörményi for making my second part of the stay in Budapest and abroad in very friendly environment, with a lot of jokes and fun times! It was a pleasure to supervise someone as adroit and friendly as Kinga! And thanks to Marci for all the help he provided in preparation of this thesis. I would also like to thank to dr. Edit Eszterbauer for the great help she provided for my private life in Hungary, being an interpreter at any time I needed. And finally, thanks to all the other people of the Institute who made my working days in Hungary easier with just a few nice words or smiles while meeting in the halls.

I also thank to all the people who provided the samples essential for my work, namely dr. Viktor Molnár and dr. Endre Sós from Hungarian zoo, dr. Baptiste Chenet from French zoo, dr. Nadine Mestre-Francés from French University, Maša Ljuština from Croatian zoo, and Katalin Földes for help with the collection of samples. Special thanks to Iris de Winter from Wageningen University for providing the samples from Madagascar. Thanks also to Niklas Arnberg and his group for hosting me in their lab for three weeks, as well as to Eric Kremer and his group for the help during my secondment in Montpellier. And I would also like to thank to the whole ADVance team, who made it possible to work on the project together.

Enormous thanks and gratitude to my whole family, especially to my mom and dad who enabled me to become a person I am. Without their guidance and support I would not be here today. Thanks for always being there for anything I need, for selfless support of every idea and wish I have, and for making my life as fulfilled and happy. The person who I am thankful the most in another way is Luka, who generously and indisputably followed me everywhere, supporting my first steps in the scientific career and being a person I can count on in every situation.

Financial supports were granted by the EU grant ADVance FP7-290002, by the Hungarian Scientific Research Funds NN107632 and the Balassi Institute (Tempus Public Foundation).

12. Appendix

12.1 Tables

Suppl. Table 1. Primers used for sequencing of SAdV-16

Primer name	Primer sequence 5'-3'	Primer name	Primer sequence 5'-3'
ITR-IVa2_ FO	CACTCGGATTCCATCATCA	ITR-IVa2_RE	CCTGGAGCACAACTACGAT
E1A-IX_FO	TTCCGATGCGATGCTCCTGG	E1A-IX_RE	GCATGGACTCTTCTCTCACC
1926_FO	CATCAGCCTGAACCTGTGGAAG	1230_RE	TGAATCAGCTGTTGTGGAGTCG
19K_FO	AATGCCTGAACTTCAGCCTGAG	ITR_RE	CCTACGTCACTTCCGTTCC
19K_FO_II	AAGCGCTCTAGGAGCAATTAGG	2130_RE	CTGCGCCATCAGTTCTCCGTTG
IVa2-*pol*_FO	GAGTAGTTGCCTTCACA	3231_RE	GGCAGGTTGCGCAGGTGATTAG
*pol*_FO	ACGTTCGTTGGTCCAGCAGAGG	55K_RE	CAGGTTGCGCAGGTGATTAG
pTP-III_FO	CGATCTCTGCCATGAACTGCTC	55K_RE_II	CATCCTCTGCATGGCACACTTA
pTP_FO	CTTCCATCTCTTCCACCTCG	IVa2-*pol*_RE	TGGTCCGAGTTCCTCTAC
pTP_FO_II	AGGAGGCGATGCTGCTGATG	*pol*_RE	CCAGCTCATGAACCGCATCTCC
pTP_FO_III	GTAGCCGTAGATGTAGCGCGAC	*pol*_RE_II	CACAGCAGACAGCACCACAG
10620_FO	GAAGAAGAGCTGGAGGCGCAGG	52K_RE	AGTAAGACTCGCGCTCCTCG
pIIIa_FO	CCACCAGCCTGTACTTGATG	pTP-III_RE	GTAGTGGTCGATGATGGCGTTG
III-V_FO	GCAACTATTCCGAGACCATGAC	11940_RE	GATGGTCAGCACCTTCTCGTGG
pVII_FO	ACGAGGTGTACGCGGACGAAGA	pIIIa_RE	ACTGGTACACCTCGCTCTGC
14500_FO	CGTAGAGAAGGACGCCAAGGAC	III-V_RE	TAGCATGAGCGACGAGCGACTG
14730_FO	CTCGAGCCAAGTCAGCAACTAC	III-V_RE_II	CCACCTTCATCTGTTCCAAGAC
III-hex_FO	AGTCATGGTGTCCAGGAAGCAG	16300_RE	GCTACTGTTCCAGGCCTTGACG
hex_FO	GACACCGCGTACTCTTACAAGG	III_RE	GGCTGCACACACAACAGAGAAC
hex_FO_II	GGCTGTATTGTACGCTGAAG	III-pVI_RE	TCTATCGTCTTGGTGACCAG
hex-100K_FO	AACGACGCTGTGCCTAACCAAC	III-hex_RE	CCTTGTAAGAGTACGCGGTGTC
DBP_FO	CTAACCGGAGTGCCTAACAG	hex_RE	GTTGGCGTAGTTGACAGTGTCC
100K_FO	AAGTCATGAGCTGCGCGATGGA	hex_RE_II	TCCATTCGTAGGTGTAGGAG
100K-fibre1_FO	CCAGACAAGCTCATGCAGACCT	hex-100K_RE	TCGCGCACAATAAGGCTCTGAC
100K-fibre2_FO	AAGTGTCTCACTTCGCCTACC	DBP_RE	TGGCCATTATGCATCTGCTG
33K-fibre1_FO	CTCGTGGAGAGCATACAAGAGC	100K-fibre2_RE	ACACGTAGCAGAAGGTGGCAGA
pVIII-E3_FO	CACGCCTTACATGTGGAGCTAC	33K-fibre1_RE	TAGTTGCACGGTAAGACTGTCG
pVIII-U_FO	CATCTGCATCACTGCCGTCAGC	25250_RE	GATACCTCTGCCTCGAATGAGC
pVIII-RID_FO	CGACTCTTGTGTTAAGCGCTGG	pVIII-E3_RE	AGCAACTGTTAGTGGAGCTTCG
CR1-RID_FO	CCTCCAACCTAGCAATTCTG	pVIII-U_RE	TTCTTGCAGCCCATTGGAGGAC
fibre2-ORF1_ FO	CTGCACCGTTAAGGAAGAACTG	pVIII-RID_RE	GCTGCTCCACCTTCTGACGTTG
fibre2-ITR_FO	GGAGTCTAAGCCTTCAAGC	CR1-RID_RE	TACAGCTGGTTGGTCATGAG
34K_FO	GGCGCAGCATCTCATCTGAATC	fibre2-ORF1_RE	GGCATATATAGGCGGTAGTTGG
		34K_RE	CAGTCTTGCAGCCTGATGATAG
		ORF3_RE	TTCGCTGGCTGATATGAACACG

Suppl. Table 2. Primers used for sequencing of SAdV-19

Primer name	Primer sequence 5'-3'	Primer name	Primer sequence 5'-3'
ITR_ FO	GGCCACTCGGATTCCATCAT	ITR-E1A _RE	AACTCTACTCGCCAGCACTC
E1A-IX_FO	GAAGGATCAGATGAGGATGAGC	ITR-IVa2_RE	ATGGCCTACGACGACCTTAC
55K_FO	AATGGATCTCCTAAGGCTGC	E1A-IX_RE	AGCGTAGGTGAGAGTAGCAGAG
IVa2-*pol*_FO	GGTATCATGTCCACTTGAGG	55K_RE	CATGTTGTGCTTAACGCTGG
IVa2-*pol*_FO_II	CTCTGGTCCACGCTAATACA	IVa2-pol_RE	TTACGCAACTTGCTCTCCA
*pol*_FO	TGGCGTAACAACGTAGAAGG	*pol*_RE	AATCTGGTCTTGCAGCACAC
pol-pTP_FO	GGGAGTTGGAAGAGATGGTAGG	*pol*-pTP_RE	CTTTCAGCGCCTGAGAAACTAC
pol-III_FO	GTGCAGGGTAATAAGGTCCACACTGGT	*pol*-pTP_RE_II	TTCTCTCAACTCATGCGACGCA
pol-III_FO_II	CCTCGCTACGAGTGGTTCGA	*pol*-III_RE	CTGATGAAGCTCACGTAGTCCTCCTGTC
pol-III_FO_III	GGTGTTGCAAGGCCAAGTCC	pTP-52K_RE	GCCTTGTGTATCCGATCGTTGC
pTP_FO	CTTGCATCAGTCTCTCCACC	pTP_RE	TCCAAGTCTAAGGCCAGCTC
pTP-52K_FO	CGACCTGAGCGACGTTCAATTC	52K-pIIIa_RE	AGTCATCGGAAGCGTTCA
52K-pIIIa_FO	GGAACTGATGCACAGCTT	III-hex_RE	GTACGAGTAAGCGGTGTCCTCC
III-pVI_FO	AGCAGGCCGTGTACTCGCAACT	pVII-pVI_RE	CTGACCAGTACTACTGTTCCAGGCT
pVII-pVI_FO	CAACCTCCACCGTAGATTCCGTGAT	hex_RE	TGAGCGGAGTTGGCATACAG
III-hex_FO	CTACTGACGTAACATGCGGCTC	hex-protease_RE	GTAACGCAGTGGTCAGGAGTAG
hex_FO	CTCCAATGTGGCCTTGTACC	hex-100K_RE	TTCGTCCTCCACCTCTTCCTCC
hex-protease_FO	CTACTGAGAATGCAGCTGGAAG	hex-pVIII_RE	TCGCTCCGCACCTTATCCTC
protease-DBP_FO	CCGTCATCGGCAACGTATAGAG	100K_RE	AGCGACAGTGGCACTCCAGTAG
DBP_ FO	CGCTACAACTCACGTCTTCC	100K_RE_II	CAGGATGGCCGCCATTATCA
hex-100K_FO	CGTCATCGAAGCCGTCTACCTC	hex-E3_RE	TCGGATGTCGTTCACTTGTGAG
hex-E3_FO	ACCGACACCTACTTCAGCTTGG	fibre1_RE	ATTGGCCGTCCTGTATTACC
ITR-E3_FO	GCCACTGCCATTCGCCTGATT	fibre1-34K_RE	AGATTCAGATGCGCTGCTGC
fibre-ITR_FO	TAGGCAGCGGTCTTAGTATC	E3-ITR_RE	GGCCACTCGGATTCCATCAT
fibre2_FO	GGTCATCGTCCATTGCCTCAAC	ORF_RE	TGTTGGCGCTCAAGACATAC
fibre2_FO_II	TAGCGGAGCAGTGGCCTTAG	ORF_RE_II	ATGCCTCTTCCGTGTCTTCCTC
fibre2_FO_III	ACTGCGGCTTAGTATTGGAG		
34K-ITR_FO	AGCGCATCTGAATCTCACTC		
ORF_FO	TCACCTCGGATACATTGTGC		

Suppl. Table 3. Primers with restriction enzyme recognition sequences (marked bold)

Adenovirus	Forward primer sequence (5'-3') / Reverse primer sequence (5'-3')	Restriction enzymes
SAdV-1	CA**GGTACC**AAGTTTGACAGCAACGGATCC CG**CTGCAG**GTTTATTGTTCTGTAATGTAGC	KpnI, PstI
SAdV-7	CGT**GGATCC**AGATTTAACAGCTCTGGTGCC CG**CTGCAG**GTTTATTGCTCTGCAACATAGC	BamHI, PstI
SAdV-11	CGT**GGATCC**AAGTTTAACAACGATGG CG**CTGCAG**GTTTATTGTTCTGTAATATAGC	BamHI, PstI
SAdV-19	AT**GGTACC**AAGTTTAGTCCCAACGGAA CGC**AAGCTT**GTTATTGTTCTGCAATGT	KpnI, HindIII

Suppl. Table 4. Primers used for sequencing of RHTAdV-1

Primer name	Primer sequence 5'-3'	Primer name	Primer sequence 5'-3'
ITR-55K_FO	CCACTCGGATTCCATCATCA	ITR-55K_RE	ATCTTCACAGACTTCGCGCT
1500_FO	GCGTGCTTCATTGTGAATGTGG	ITR-55K_RE_II	ATCTTCACAGACTTCGCGCT
E1A-55K_FO	TCCTGGTTCTCACGTTCAACTG	3500_RE	CTGTAACGGAATCGGAGCTTCG
55K-IVa2_FO	CATCGTGAACGCGTGCTT	E1A-55K_RE	CTGGCTATGGCATCACTCATGT
55K_IVa2_FO_II	GCCATTGCAAGTGTGTCA	55K-IVa2_RE	TCCAGAACACGAAGCTCT
IVa2-*pol*_FO	TCATCCATAATAATAGCAATAGGCCCT	55K-IVa2_RE_II	TTAAGTCCATAAGTCCAA
IVa2-*pol*_FO_II	AATCAGAGGCTGGATGAC	55K_IVa2_RE_III	GACACACTTGCAATGGCT
*pol*_FO	CTAGGAAGTGGATGAGTTAAAGCAGATG	IVa2-*pol*_RE	GTGACATGCTTAATAGTGTAACAC
pol-pTP_FO	CCACCATTGTGAGCTGTGAGCAT	IVa2-*pol*_RE_II	CACCTTTCTTGAATCTGAAGCTGACGA
pol-pTP_FO_II	GAGCTACGTGACGTTGAC	IVa2-*pol*_RE_III	GTCATCCAGCCTCTGATT
pol-III_FO	GTGCGCAGTAATCAAGAGTG	*pol*-pTP_RE	GTTCATCCAGCATGCAAG
pol-III_FO_II	GGCAGCTAATACAATTTC	*pol*-pTP_RE_II	CTTATCGGCCACCATTGAACGAG
pTP-pIIIa_FO	TCCAATTCTCCATGTCCAGCTC	*pol*-III_RE_II	TTACTCCAGAACCACAAG
pTP-pIIIa_FO_II	CAGGCCATTTTCGTTCC	pTP-pIIIa_RE	CACAAGAACAGCGCGAACAGGAA
pTP-pIIIa_FO_III	TTACAGCAACACCATCCTCGCGT	pTP-pIIIa_RE_II	GTGTGTTCGCCAAGCAGAATCA
pTP-pIIIa_FO_IV	CGAGTTCTAAGTCCTGGT	pTP-pIIIa_RE_III	GCGGATGTGCTTCAGAGTAA
52K_FO	GCTGTATTATCAGGCGTTGGAG	52K_RE	GGACCTGCTTGATACACTTCAC
52K-pIIIa_FO	TAAGTCAGCAACCACACG	52K-pIIIa_RE	TATAGCAGCGGCTCTATG
III-hex_FO	GAGTAGAAGAAGCTGATATAGGAGT	III-hex_RE	CTACCATCATATCCGGCA
III-hex_FO_II	CTTGCCGGATATGATGGT	III-hex_RE_II	ACAATAGTCCTCTGGTG
III-hex_FO_III	TGTCCACTCGCACAACTG	III-hex_RE_III	TTCTGGAGCCACGATATG
III-hex_FO_IV	TGCATATCGTGGCTCCAG	III-hex_RE_IV	CCGCTTTACGACGTTTG
III-hex_FO_V	GACGATCTTCAACCTACC	III-hex_RE_V	GTCATGTTGATGTCGCTG
III-hex_FO_VI	CGGCGGATAGATATACCA	III-hex_RE_VI	GTGCTACTATTCCAGGCT
III-hex_FO_VII	TCGGCACCAACCAGATGA	III-hex_RE_VII	AAGGAATGAACGCCAAGG
III-hex_FO_VIII	AAGTGCGGCTGTGGTAGA	III-hex_RE_VIII	CTCTTTATCAACAGGCACAAATCGT
hex-pVIII_FO	GCCAGGACGCTTCAGAATACCTCAG	hex-pVIII_RE	TTCCGGTTCTAGTCTGAG
hex-pVIII_FO_II	GGTGTTAATGCCGATGGA	hex-pVIII_RE_II	TCGTAAGTGTACGTGCCA
hex-pVIII_FO_III	CCATGTGGAATCAAGCTG	hex-pVIII_RE_III	CGGCCAACTTACAGATGA
hex-pVIII_FO_IV	GGTAACGGACGCTATTGT	hex-pVIII_RE_IV	AAGGCCAGCCAATGAACG
hex-pVIII_FO_V	ACGACAGGTTGCTTACTC	hex-pVIII_RE_V	CGGTGGTTCCTCCTGATT
hex-pVIII_FO_VI	TTGACGTGGTTCGTGTTC	hex-pVIII_RE_VI	ACGAGTCAGTCGCCTATT
hex-pVIII_FO_VII	CCGTTCAAGGTCCATACA	hex-pVIII_RE_VII	TAGCTAGCAAGCACTTCA
hex-pVIII_FO_VIII	AGGAACCACCGCCATTGT	hex-pVIII_RE_VIII	TTACCAATACTACCGCA
hex-pVIII_FO_IX	CCATTAACTTGCCGTTGC	hex-pVIII_RE_IX	GAAGAGCGCAGCACATTG
hex-pVIII_FO_X	CTCCTCCTAGGTTAGAAG	hex-pVIII_RE_X	GATGGTTGAGCAGTTAGC
hex-pVIII_FO_XI	GGAAGATGGCAAGCCAGT	hex-pVIII_RE_XI	GGCCATTCTTCGATGCTG
hex-pVIII_FO_XII	AAGCCTTGCAATCCGGAC	hex-pVIII_RE_XII	GAGTGGTGGTAATGGCAGCTTGTCT
hex-pVIII_FO_XIII	ATGAGCGCCAATCAGAAG	pVIII-ITR_RE	CCACTCGGATTCCATCATCA
hex-pVIII_FO_XIV	ACAAGGCGGTATATGCTG	pVIII_RE	GACGCACTCTAACTCGTTCAAG
pVIII-ITR_FO	AGGCGGCATTGGACAACTT		
pVIII_FO	CTTGAACGAGTTAGAGTGCGTC		
pVIII_FO_II	CCTTGCACAACACTTGTGTTCC		
E3_FO	ATCACGTGGATCTATACGCTGC		

12.2 Accession numbers of GenBank retrieved sequences

HAdV-1 (AF534906), HAdV-2 (ADRCG), HAdV-3 (NC_011203), HAdV-4 (AY487947), HAdV-5 (AC_000008), HAdV-6 (FJ349096), HAdV-7 (AC_000018), HAdV-8 (AB448767), HAdV-9 (AJ854486), HAdV-10 (JN226746), HAdV-11 (AY163756), HAdV-12 (AC_000005), HAdV-13 (JN226747), HAdV-14 (AY803294), HAdV-15 (JN226748), HAdV-16 (AY601636), HAdV-17 (AC_000006), HAdV-18 (GU191019), HAdV-19 (EF121005), HAdV-20 (JN226749), HAdV-21 (AY601633), HAdV-22 (FJ404771), HAdV-23 (JN226750), HAdV-24 (JN226751), HAdV-25 (JN226752), HAdV-26 (EF153474), HAdV-27 (JN226753), HAdV-28 (FJ824826), HAdV-29 (JN226754), HAdV-30 (JN226755), HAdV-31 (AM749299), HAdV-32 (JN226756), HAdV-33 (JN226758), HAdV-34 (AY737797), HAdV-35 (AY128640), HAdV-36 (GQ384080), HAdV-37 (DQ900900), HAdV-38 (JN226759), HAdV-39 (JN226760), HAdV-40 (L19443), HAdV-41 (DQ315364), HAdV-42 (JN226761), HAdV-43 (JN226762), HAdV-44 (JN226763), HAdV-45 (JN226764), HAdV-46 (AY875648), HAdV-47 (JN226757), HAdV-48 (EF153473), HAdV-49 (DQ393829), HAdV-50 (AY737798), HAdV-51 (JN226765), HAdV-52 (DQ923122), SAdV-1 (AY771780), SAdV-3 (AY598782), SAdV-6 (JQ776547), SAdV-7 (DQ792570), SAdV-13 (KP329563), SAdV-18 (CQ982407), SAdV-20 (HQ605912), SAdV-21 (AC_000010), SAdV-22 (AY530876), SAdV-23 (AY530877), SAdV-24 (AY530878), SAdV-25 (AC_000011), SAdV-26 (HB426768), SAdV-27.1 (HC084988), SAdV-27.2 (FJ025928), SAdV-28.1 (HC084950), SAdV-28.2 (FJ025915), SAdV-29 (HC085020), SAdV-30 (HB426704), SAdV-31.1 (HC000816), SAdV-32 (HC085052), SAdV-33 (HC085083), SAdV-34 (HC000847), SAdV-35.1 (HC085115), SAdV-35.2 (FJ025910), SAdV-36 (HC191003), SAdV-37.1 (HB426639), SAdV-37.2 (FJ025919), SAdV-38 (FJ025919), SAdV-39 (HB426607), SAdV-40.1 (HC000785), SAdV-41.1 (HI964271), SAdV-42.1 (HC191035), SAdV-43 (FJ025900), SAdV-44 (FJ025899), SAdV-45 (FJ025901), SAdV-46 (FJ025930), SAdV-47 (FJ025929), SAdV-48 (JQ776547), SAdV-49 (HQ241819), SAdV-50 (HQ241820), SAdV-54 strain 23336 (KM190146), RhAdV-51 (NC_025826), RhAdV-52 (NC_025827), RhAdV-53 (NC_025828), SAdV-ch1 (KF360047), ChAd3 (CS138463), ChAd6 (CS138464), Chseq13 (HH760489), SAdVch36 (CS479277), Chseq62 (HH760538), Chseq63 (HH760539), Chseq65 (HH760541), SAdV-A1139 (JN880448), SAdV-A1163 (JN880449), SAdV-A1173 (JN880450), SAdV-A1258 (JN880451), SAdV-A1285 (JN880452), SAdV-A1296 (JN880453), SAdV-A1312 (JN880454), SAdV-A1327 (JN880455), SAdV-A1335 (JN880456), BaAdV-1 (KC693021), BaAdV-2 (KC693022), BaAdV-3 (KC693023), TMAdV (HQ913600), FAdV-1 (AC_000014), TSAdV (AC_000190).