View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Elsevier - Publisher Connector

Available online at www.sciencedirect.com
SCIENCE DIRECT®
Virology 332 (2005) 114–129
VIROLOGY
www.elsevier.com/locate/yviro

Genomic and bioinformatics analysis of HAdV-7, a human adenovirus of species B1 that causes acute respiratory disease: implications for vector development in human gene therapy

Anjan Purkayastha[a,b,c], Jing Su[a,b,c], Steve Carlisle[d], Clark Tibbetts[b,c], Donald Seto[a,b,c,*]

[a]*Bioinformatics and Computational Biology, School of Computational Sciences, George Mason University, 10900 University Boulevard, MSN 5B3, Manassas, VA 20110, USA*
[b]*HQ USAF Surgeon General Office, Directorate of Modernization (SGR), 5201 Leesburg Pike, Suite 1401, Falls Church, VA 22041, USA*
[c]*Epidemic Outbreak Surveillance (EOS) Consortium, 5201 Leesburg Pike, Suite 1401, Falls Church, VA 22041, USA*
[d]*Commonwealth Biotechnologies, Inc. (CBI), 601 Biotech Drive, Richmond, VA 23235, USA*

Received 20 July 2004; returned to author for revision 24 September 2004; accepted 26 October 2004
Available online 8 December 2004

Abstract

Human adenovirus serotype 7 (HAdV-7) is a reemerging pathogen identified in acute respiratory disease (ARD), particularly in epidemics affecting basic military trainee populations of otherwise healthy young adults. The genome has been sequenced and annotated (GenBank accession no. AY594255). Comparative genomics and bioinformatics analyses of the HAdV-7 genome sequence provide insight into its natural history and phylogenetic relationships. A putative origin of HAdV-7 from a chimpanzee host is observed. This has implications within the current biotechnological interest of using chimpanzee adenoviruses as vectors for human gene therapy and DNA vaccine delivery. Rapid genome sequencing and analyses of this species B1 member provide an example of exploiting accurate low-pass DNA sequencing technology in pathogen characterization and epidemic outbreak surveillance through the identification, validation, and application of unique pathogen genome signatures.
Published by Elsevier Inc.

Keywords: Human adenovirus type 7; HAdV-B1 species; Human gene therapy vector; Acute respiratory disease

Introduction

Adenoviruses infect a very broad spectrum of hosts across the entire span of vertebrates, including fish, frog, and corn snake (Crespo et al., 1998; Davison et al., 2000, 2003; Farkas et al., 2002; Kovacs et al., 2003). The family *Adenoviridae* comprises five genera; within this family are 51 identified serotypes of human adenoviruses (HAdVs) that are included within the genus *Mastadenoviridae*. These HAdVs, in turn, are further divided into six species (formerly subgroups or subgenera) based on differentiating properties that include increasingly important genome differences, as unveiled by whole genome sequencing, as well as restriction enzyme digest patterns and the classical gold standards of serum neutralization and hemagglutination-inhibition tests (Davison et al., 2003). In 1953, the first human adenovirus (HAdV-1) was isolated from adenoid tissue and characterized as a cytopathic agent causing respiratory illness (Rowe et al., 1953). Its complete genome was only recently sequenced and annotated (Lauer et al., 2004). Within a short time, several other HAdVs were soon isolated, characterized, and numbered sequentially. Among these were the serotypes HAdV-4 and HAdV-7, members of species E and B1, respectively. These serotypes are most commonly associated with febrile respiratory disease; in particular, acute respiratory disease (ARD) (Erdman et al., 2002; Hierholzer, 1995).

As of the last global survey, approximately one-fifth of all HAdV infections reported to the World Health Organ-

* Corresponding author. Fax: +1 703 993 8401.
E-mail address: dseto@gmu.edu (D. Seto).

0042-6822/$ - see front matter. Published by Elsevier Inc.
doi:10.1016/j.virol.2004.10.041

ization (WHO) are attributed to HAdV-7 (Erdman et al., 2002; Schmitz et al., 1983). These diseases include respiratory tract illnesses and conjunctivitis. In infants and immunocompromised populations, HAdV-7 can cause outbreaks of severe disease, and, in a few cases, even lead to death (Munoz et al., 1998). Intriguingly, even within young, active, and previously healthy adult populations, epidemic outbreaks and associated mortality have been recorded (Ryan et al., 2001). Three patterns of HAdV-7 respiratory disease epidemics have been recognized: (i) the first type occurs in winter among infants (<2 years old) and may be fatal; (ii) the second occurs in older children and is not usually fatal; and (iii) the third occurs as ARD in military recruits and can have high morbidity rates along with less-common mortality (Ryan et al., 2001; Wadell et al., 1980).

Persistent outbreaks of HAdV-related ARD among basic military trainees led the U.S. Army to develop vaccines against HAdV4 and HAdV7, the two primary agents of adenoviral ARD (Chanock et al., 1966; Gaydos and Gaydos, 1995; Top et al., 1971a, 1971b). Despite their effectiveness, the vaccines were discontinued in 1996. This suspension of the vaccination program resulted in a resurgence of ARD cases to prevaccine era levels (Ryan et al., 2000, 2001, 2002). The reemergence of HAdV as a major causative agent for ARD has prompted the U.S. Department of Defense to restart a HAdV vaccine development program.

The global epidemiology of HAdV-7 genome types has been well documented (Erdman et al., 2002; Kajon and Wadell, 1994; Wadell and Varsanyi, 1978; Wadell et al., 1985). Multiple HAdV-7 genome types have been identified by restriction enzyme analyses (Wadell and Varsanyi, 1978). Global prevalence patterns of these HAdV-7 genome types shift over time (Erdman et al., 2002; Wadell et al., 1985). For example, in the period 1975–1985, the HAdV-7b genome type predominated in North America, Europe, and Australia, while the c, d, and e variants (the 'letter' in the strain designation reflects a restriction enzyme pattern signature) were isolated in other parts of the world (Wadell et al., 1985). In South America, a shift in the prevalence pattern from HAdV-7c to HAdV-7h occurred in 1986; HAdV-7h was shown to cause severe illness and fatalities in children (Kajon and Wadell, 1994). In the 1980s, HAdV-7d displaced HAdV-7b as the predominant genome type (Wadell et al., 1985). In later years, HAdV-7d spread to parts of Asia, causing epidemics in Japan and Korea (Erdman et al., 2002). The HAdV-7h and HAdV-7d2 genome types were recently detected for the first time in North America, signaling a possible shift in the prevalent genome types in the USA (Erdman et al., 2002). The severity and appearance of disease may change with genome types. It is therefore important to understand the genomics and bioinformatics of an ARD-relevant and human disease-relevant B1-species member.

In this report, the complete and annotated genome sequence of prototype HAdV-7p is presented (GenBank Accession no. AY594255). It is important both as a reference genome and a historical footnote, and serves as a valuable and suitable reference for studying the evolution of the various HAdV-7 genome types in the past 50 years. These data provide an additional powerful tool in studying the molecular patho-epidemiology of HAdV-7 and other ARD-causing HAdVs, especially in the development, validation, and deployment of unique pathogen genome signatures and genome-based advanced molecular diagnostics tools.

The phylogeny and evolutionary origins of human adenoviruses have long been a subject of study and debate. This current study also examines the evolutionary relationships between HAdV-7 and a set of recently sequenced simian adenoviruses (Roy et al., 2004). These results have important implications in vector development for human gene therapy and vaccine delivery as the simian adenoviruses are being explored as alternative human gene therapy vectors (Roy et al., 2004).

Results and discussion

Genome sequencing with a sparsely tiled primer array

Given the complete genome sequences of related HAdVs archived either in GenBank (including archived partial genomic and gene sequences) or in the database of ones newly sequenced by the U.S. Air Force, Epidemic Outbreak Surveillance (EOS) Consortium, PCR and DNA sequencing primers were identified and optimized for efficient and low-cost genome sequencing of the HAdV-7 genome based on a sparsely tiled overlap of PCR and sequencing primers. This protocol ensures high quality of data by generating overlapping and complementing sequences for a minimal redundancy of 3-fold coverage per base.

General overview of the HAdV-7p genome sequence

Interestingly, given the past and current importance of the central role of HAdV-7 in ARD and in understanding HAdV patho-epidemiology, there have been no whole genome studies to date. A reference genome of this species B1 pathogen is critical for comparative genomics with current and epidemic field strain isolates. Genomes from other HAdVs have been examined and archived in GenBank: species C (serotypes 1, 2, and 5), species A (12), species D (17), species F (40), and species B2 (11 and 35). Studies such as these are much enhanced given reference genomes for comparative genomics, as well as reference genes for comparative and identity analyses.

The HAdV-7 prototype strain analyzed here is also known as the Gomen strain, isolated as a clinical specimen from a throat washing of a military recruit with pharyngitis

at Fort Ord, CA (Berge et al., 1955). This strain is nearly contemporary with the Greider strain (Rowe et al., 1958), aka HAdV-7a, which has been studied extensively (Crawford-Miksza et al., 1999; Li and Wadell, 1986). There are genome differences between the two strains (manuscript in preparation).

The HAdV-7p genome sequence is 35 306 bp in length and has an overall base composition of 25.36% A, 25.77% C, 25.24% G, and 23.6% T. Its G + C content of 51.03% is within the 50–52% range observed earlier for the HAdV-B species (Shenk, 1996). As observed in the genomes of other members of *Mastadenoviruses*, the HAdV-7 genome is organized into early, intermediate, and late transcription regions. A total of 48 coding sequences were identified in this genome sequence, including seven predicted genes.

Noncoding motifs

Table 1 presents noncoding DNA sequence motifs uncovered in the genome. These motifs include the inverted terminal repeats (ITRs) at either end of the genome (Dan et al., 2001). These are 108 bp long. Sequence alignments show the HAdV-7 ITR to be most similar to those of the B1 species (HAdV-3, 16, and 21). However, the ITRs do not contain the canonical CATCATCAAT motif that is conserved in other HAdVs (Stone et al., 2003). Instead, CTATCTA immediately precedes the "core" motif for DNA replication, TATAATATACC (which differs very slightly from "canonical" AATAATATACC) that binds the preterminal protein–DNA polymerase complex is present at 8–18 bp (Temperley and Hay, 1992). These two deviations from the canonical sequences are also observed in the HAdV-4p genome (Purkayastha et al., in press; Tokunaga et al., 1982), and in genomes annotated from fellow members of the HAdV-B1 subspecies, HAdV-3 and 21 (manuscript in preparation).

In addition to the viral factors, DNA pol and pTP, HAdVs also require a set of host cellular factors for efficient replication. These transcription factor DNA binding motifs are located in the ITR region of HAdV-7 as follows: NFIII/Oct-1 binding site (CATGTAAATGA) at (40–50 bp) and a Sp1 binding site (CTGTGTGG) at (72–79 bp). Additionally, the NFI binding site (TGGAATGGTGCCAA) at 26–39 bp was also identified.

Table 1
HAdV-7 genome noncoding motif annotation

Motif	Function	Position
CTATCT...TAACAT	The inverted terminal repeat	1–108
ATAATATACC	DNApol-pTP binding site	8–18
TGGAATGGTGCCAA	NFI binding site	26–39
CATGTAAATGA	NFIII binding site	40–50
CTGTGTGG	Sp1 recognition site	72–79
TATTTA	TATA box for E1A	481–486
AATAAA	polyA signal for E1A	1494–1499
TATATA	TATA box for E1B	1549–1554
AATGTG	polyA signal for E1B	3405–3410
TAAAGT	TATA box for the pIX gene	3384–3389
AATAAA	polyA signal for pIX gene	3910–3915
AATACA	polyA signal for E2B	(31 374–31 379)c
TGATTGGCTT	Inverted CAAT box for MLP	5823–5832
GCCACGTGAC	Upstream element for MLP	5843–5851
GCCGGGGGGG	MAZ binding site for MLP	5864–5873
TATAAAA	TATA box for MLP	5874–5880
GGGGGCGGACC	MAZ/SP1 binding site for MLP	5881–5891
TCACTGT	Initiator element for MLP	5903–5909
TTGTCAGTTTC	DE1 for MLP	5990–6000
AACGAGGAGGATTTGA	DE2a and DE2b for MLP	6005–6020
AATAAA	polyA signal for L1	13 830–13 835
ATTAAA	polyA signal for L2	17 497–17 502
AATAAA	polyA signal for L3	21 918–21 923
AATAAA	polyA signal for E2A	13 372–13 377
TATAAA	TATA box for E3	27 065–27 070
AATAGA	polyA signal for L4	27 547–27 552
AATAAA	polyA signal for the E3	31 128–31 133
AATATC	polyA signal for L5	34 122–34 127
AATATA	polyA signal for E4	(32 317–32 322)c
TATATATA	TATA box for E4	(35 000–35 005)c
ATAATATACC	DNApol-pTP binding site	(35 289–35 298)c
CTATCT...TAACAT	The inverted terminal repeat	(35 199–35 306)c

DNA sequence motifs are identified and located on the genome of HAdV-7. Their nucleotide signatures and putative functions are indicated. The nucleotide positions of their location are noted in the 5′ to 3′ orientation. Functionality, which is embedded within the complementary strand, is designated by 'c', for example, "(4089–4094)c".

VA RNA

The virus-associated RNA (VA RNA) species are non-protein coding sequences that have been shown to repress the antiviral activity of host interferons (Mathews and Shenk, 1991). For HAdV-7, the VA RNA I (170 bp) and VA RNA II (171 bp) coding sequences were located at 10 423–10 592 and 10 668–10 838 bp. The HAdV-7 VA RNA I is most similar to its counterpart in HAdV-3 with 98% identity. Analysis of VA RNA II returned greater than 90% identical matches with its counterparts in serotypes HAdV-3, 16, and 21, all of which belong to human adenovirus species HAdV-B1. No significant matches were found in the genome sequences of the HAdV-B2 members. This is not surprising, given that HAdV-B2 contains only one VA RNA gene (Kidd et al., 1995). However, in HAdV-C and E species (whose members have two VA RNA genes), no significant regions of identity were found to the HAdV-7 VA RNA II sequence.

Gene coding annotation

Table 2 displays a complete listing of the genes found in HAdV-7p, along with chromosomal locations and putative functions. Genes transcribed in the complementary direction are noted with a “c” designation following the nucleotide locations.

Early genes

E1A

The first transcription unit to be expressed after infection is the E1A. This transcript encodes several E1A proteins that have important roles as transcriptional regulators within the host cell, modulating both viral and cellular gene expression (Flint and Shenk, 1997). These proteins lack sequence-specific DNA-binding activity (Zu et al., 1992). They presumably control gene expression by interacting with cellular elements of the transcription machinery. The alternative splicing of a common RNA precursor, transcribed from a constitutively active promoter, generates multiple E1A products. Three putative E1A proteins of sizes 6.7, 24.6, and 28.3 kDa were identified in the HAdV-7p genome.

E1B

Three coding sequences were identified in the E1B region. The early 20.5-kDa protein has a high BLAST score to the small T antigen that is conserved in other adenoviruses. E1B 54.7-kDa protein has identity to the large T antigen protein, which has been shown to inhibit cellular p53-mediated host defense mechanisms (Yew et al., 1994). This large T antigen protein also plays a role in regulating viral late gene expression. A third coding sequence encodes an 8.9-kDa protein with identity to the 1.26-kb mRNA product that has been identified in the HAdV-C species.

E2

Proteins required for viral DNA replication are encoded in the E2 transcriptional unit. These are divided into two regions, E2A and E2B. HAdV DNA replication system has been shown to require three viral factors: terminal protein precursor, DNA polymerase, and DNA binding protein. A 58.2-kDa DNA binding protein was identified within E2A. In the E2B region, a 133-kDa DNA polymerase and a 75.6-kDa terminal protein precursor were characterized. In addition, three hypothetical proteins of sizes 6.1, 11.3, and 12.6 kDa in the E2B region were uncovered.

E3

The E3 region of human adenoviruses encodes proteins that counter the host immune response (Wold and Gooding, 1991). These proteins are not essential for in vitro growth. The HAdV-7 E3 region was found to encode the 12.1-, 16.1-, 19.3-, 7.7-, 10.3-, 14.9-, and 14.7-kDa proteins. Addition-

Table 2
HAdV-7 genome gene coding annotation

Gene	Product	Location
E1A	6-kDa protein	(577–648, 1250–1351)
E1A	28-kDa protein	(577–1156, 1250–1455)
E1A	32-kDa protein	(577–1063, 1250–1455)
E1B	19-kDa small T antigen	(1603–2139)
E1B	55-kDa protein	(1908–3386)
E1B	hypothetical 1.2kB mRNA protein	(3144–3386)
IX	hexon associated protein IX	(3481–3897)
IVA2	maturation protein IVA2	(3950–5283, 5562–5574)c
E2B	DNA polymerase	(5053–8421)c
Hypo.	hypothetical A-106 protein	(6146–6466)
Hypo.	hypothetical DNA binding protein	(7987–8427)
E2B	12.6-kDa early protein	(8231–8575)c
E2B	terminal protein precursor	(8424–10346)c
Hypo.	hypothetical 14.5-kDa protein	(8550–8951)
E2B	11.3-kDa early protein	(9545–9859)c
Hypo.	hypothetical 9.7-kDa protein	(9759–10034)
VA RNA	VA RNA I	10423–10592
VA RNA	VA RNA II	10668–10838
L1	55-kDa protein	(10869–12026)
E2B	6.1-kDa hypothetical protein	(12047–12223)c
L1	protein IIIa precursor	(12051–13817)
L2	penton protein (protein III)	(13905–15539)
L2	protein VII precursor	(15550–16128)
L2	protein V precursor	(16171–17000)
L3	protein VI precursor	(17554–18306)
L3	hexon protein	(18419–21232)
L3	23-kDa protease	(21269–21898)
E2A	DNA binding protein	(21985–23538)c
L4	100-kDa hexon-assembly associated protein	(23569–26055)
Hypo.	hypothetical protein	(24924–25133)c
L4	22-kDa protein	(25757–26356)
L4	33-kDa protein	(25757–26225, 26275–2663)
L4	protein VIII, hexon-associated protein	(26700–27383)
E3	12.1-kDa glycoprotein	(27383–27703)
E3	16. 1-kDa protein	(27657–28097)
E3	19.3-kDa MHC class I antigen-binding glycoprotein	(28082–28600)
E3	20.6-kDa protein	(28630–29169)
E3	20.6-kDa protein	(29182–29751)
E3	7.7-kDa protein	(29766–29966)
E3	10.3-kDa protein	(30083–30313)
E3	14.9-kDa protein	(30285–30722)
E3	14.7-kDa protein	(30715–31122)
U	U protein	(31141–31305)c
L5	fiber protein	(31320–32297)
E4	Orf 6/7 protein	(32333–32584)c
E4	33.2-kDa protein	(32581–33480)c
E4	13.6-kDa protein	(33383–33751)c
L5	agnoprotein	(33606–34115)
E4	130 aa protein	(34110–34499)c
E4	13.9-kDa protein	(34541–34918)c

Forty-eight coding regions are identified within the HAdV-7 genome sequence. The genes and their products are indicated in the columns (hypothetical and predicted proteins are marked as “Hypo”). Nucleotide positions of the start/stop codons and of the applicable splice sites are noted in the 5′ to 3′ direction. Coding sequences transcribed from the complementary strand are designated by ‘c’, for example, “(30923–31090)c”.

ally, two different 20.6-kDa proteins were contained within this transcript. The 12.1-kDa protein has significant identity to an immunomodulating E3 protein in HAdV-2. A glycoprotein of 16.1-kDa has homologs in other HAdV species. The 19.3-kDa protein is a major histocompatibility class I antigen-binding glycoprotein that prevents the lysis of adeno-infected host cells by cytotoxic T-lymphocytes (Wold and Gooding, 1991). Both 20.6-kDa proteins are similar to the CR1 (conserved region 1)-containing proteins in the E3 region of other HAdVs and SAdVs (Deryckere and Burgert, 1996). The function of this 80 amino acid CR1 domain has not been identified. The 7.7-kDa protein is reported to insert itself into the host cell membrane; its function is yet to be determined (Hong et al., 1988). A HAdV E3 transmembrane protein has identity to the HAdV-7 10.3-kDa protein. This may have a role in downregulating the epidermal growth factor (EGF) receptor (Wold and Gooding, 1991). Both the 14.9-and 14.7-kDa proteins are similar by BLAST scores to adenoviral E3 proteins that are known to protect virus-infected against cells against TNF-induced cytolysis (Horton et al., 1990).

E4

The proteins encoded by the E4 transcription unit have various functions reported in the literature, including viral RNA export and stabilization (Leppard, 1997). For example, the E4 Orf6 protein combines with the E1B 55-kDa protein to inhibit cellular p53 whereas the E4 Orf6/7 has been shown to regulate the cellular transcription factor E2F. The E4 Orf4 controls protein phosphorylation in infected cells. Five putative coding sequences in the HAdV-7 E4 region were identified. These include a 13.9-kDa protein that has identity to the adenoviral E4 protein Orf and contains a dUTPase domain, as well as a 14.4-kDa Orf2 protein (also known as the 130 aa protein), a 13.6-kDa nuclear binding Orf3 protein, a 14.2-kDa Orf4 protein, a 34.7-kDa Orf6 protein, and a 9.4-kDa Orf6/7 protein.

Intermediate genes

IX

As reported in the literature, the intermediate protein IX plays a critical but poorly understood role in controlling DNA packaging (Sargent et al., 2004). It also performs a second function as a transcriptional activator for the major late promoter (MLP) and other viral and cellular promoters in HAdV-5 (Rosa-Calatrava et al., 2001). The physiological role of protein IX as a transcriptional regulator is not clearly understood (San Martin and Burnett, 2003; Sargent et al., 2004). An ORF encoding a 14.1K pIX was identified at 3481–3897 bp. This coding sequence is conserved in *Mastadenviridae*.

IVa2

Protein IVa2, the second intermediate protein, plays a serotype-specific role in packaging viral DNA during adenovirus assembly (Zhang et al., 2001). It also has a role as a transcription factor for the major late genes (Binger and Flint, 1984). An HAdV-7 IVa2 coding sequence was identified at 3950–5574 bp.

Late genes

Adenoviral late genes are transcribed from the MLP. Multiple polyA signals are utilized to produce the various distinct mRNA species. Studies on HAdV-2 have identified essential elements of the MLP (Young, 2003). Based on sequence comparison, all homologs of these regulatory elements in the HAdV-7 MLP were identified. These include the inverted CAAT box (5823–5832 bp); Upstream element (5843–5851bp); TATA box (5874–5880 bp); and MAZ/Sp1 binding sites flanking the TATA box at 5864–5873 and 5881–5891 bp. The Initiator element, which includes the transcription start site for the MLP, is located at 5903–5909 bp. Two downstream elements that recognize the IVa2 protein were identified at 5990–6000 bp (DE1) and 6005–6020 bp (DE2a and DE2b). The late transcription unit encodes the major adenoviral structural proteins and is further subdivided into regions L1–L5, each region being expressed as a distinct mRNA species.

L1

In the HAdV-7 L1 region, the 52-kDa protein (10 869–12 026 bp) and protein IIIa (12 051–13 817 bp) were identified. The 52-kDa protein is thought act as a scaffold for capsid assembly during virus assembly (Hasson et al., 1989). The IIIa protein is found on the outer surface of the virus and seems to have a function in holding the virus facets together (San Martin and Burnett, 2003). In addition, four hypothetical L1 proteins upstream of the 52-kDa coding sequence are contained within this transcript.

L2

Four coding sequences were identified in the L2 transcript. One of these is the penton base protein III coding sequence. Its corresponding gene product is found at 12 virion vertices. The penton protein binds to the host integrins via a conserved Arg–Gly–Asp (RGD) sequence to trigger virus internalization (Wickham et al., 1993). The HAdV-7 penton RGD motif is located at 14 889–14 897 bp. These two surface proteins may contribute to the role of HAdV-7 in infection and disease. The coding sequences for the precursors of proteins VII and V, which are found at the viral core, were located at 15 550–16 128 and 16 171–17 000 bp, respectively.

L3

In the L3 transcript region, three coding sequences were located: minor capsid protein precursor pVI, hexon, and 23-kDa protease. The minor capsid protein is found on the inner capsid surface and probably serves as a structural

intermediate between the capsid and the viral core. In the HAdV-7 genome, the coding sequence for the pVI precursor is located at 17554–18306 bp. The coding sequence for the 105.7-kDa HAdV-7 hexon was located at 18419–21232 bp. The hexon protein is the major structural component of the adenovirus capsid, making up about 63% of the virion mass. The HAdV-7 hexon is a 937 amino acid long protein that shares 95% amino acid identity with its homolog in HAdV-3. This appears to be a determinant for species B1 infection (Erdman et al., 2002; Li and Wadell, 1999). The final coding sequence in the HAdV-7 L3 region encodes a 23-kDa protease and was located at 21269–21898 bp. The 23-kDa protease is required for the cleavage of viral proteins during virus maturation and assembly.

L4

Four coding sequences were identified in the HAdV-7 L4 transcript region. These correspond to the following: 100-kDa protein (23569–26055 bp); 22-kDa protein (25439–25978 bp); 33-kDa protein (25757–26630 bp); and pVIII protein (26700–27383 bp). The 100-kDa nonstructural protein probably has a function in hexon assembly and is required for the efficient translation of the late viral mRNAs (Hayes et al., 1990; Oosterom-Dragon and Ginsberg, 1981). The function of the 22-kDa protein has yet to be identified. The 33-kDa protein is believed to play an indispensible role in virion assembly (Finnen et al., 2001). The pVIII protein is found on the interior surface of the capsid and is thought to serve as a bridge between the capsid and the viral core elements (San Martin and Burnett, 2003). Like its other adenoviral homologs, the HAdV-7 pVIII is rich in proline, arginine, and lysine (~12.7%), and probably has a highly disordered structure.

L5

The L5 region of HAdV-7 encodes the 35.2-kDa fiber protein at 31320–32297 bp. Specifically, the HAdV-7 fiber is 325 amino acids in length. It forms a trimeric fiber assembly that protrudes from the vertices of the icosahedral adenoviral capsid. The N-terminal domain is reported to attach noncovalently to the penton base protein, while the globular C-terminal "knob" domain binds host cell receptor. This fiber is considered to be a major determinant of tissue tropism. In some species, it binds to a cell surface receptor, Coxsackie, and Adenovirus Receptor (CAR) (Howitt et al., 2003). A detailed study of the crystal structure of the HAdV-12 knob domain bound to CAR has revealed the key fiber residues that are required for this binding (Howitt et al., 2003). These include Asp415, Pro417, and Pro418. Another important residue Lys429 is conserved throughout all subtypes except for species F. However, unlike members of the other HAdV species, members of species B apparently do not bind this CAR (Defer et al., 1990). CLUSTALX-based multiple sequence alignment of the fiber sequences from the species A, B, C, and E HAdVs with the HAdV-12 fiber sequence shows that while the Asp415 residue is present in HAdV-2, 4, and 12, it is replaced with an Asn in the HAdV-7 genome (Fig. 1). The residues Pro417 and Pro418 are replaced with Thr and Ala, respectively, in HAdV-7. This lack of conservation of the key binding residues may explain why HAdV-7 and the other HAdV-Bs do not attach to CAR. The L5 region also encodes a putative DNA binding protein at 33606–34115 bp. A similar L5 DNA binding protein coding sequence is found in other HAdV-Bs (unpublished observation).

Miscellaneous proteins

Seven hypothetical proteins were identified across the entire HAdV-7 genome sequence. Four hypothetical coding sequences were identified in the ~5 kb stretch between the MLP initiator element and the L1 52-kDa protein coding sequence: (a) A-106 homolog; (b) DNA binding protein; (c) 14.7-kDa protein; and (d) 9.7-kDa protein. E2B encodes two hypothetical proteins of sizes 11.3 and 12.6 kDa. A 7.9-kDa protein coding sequence lies between the E2 and E4 genes; it has no apparent similarity to any other coding sequence. A short ORF with a high BLAST score against the "U exon" region at position 31141–31305 bp was found, corresponding to the E3 region and reading in the complementary strand. The "U exon" was originally reported in HAdV-40 as a small coding region extending from an initiation codon to a splice donor site and may represent the N-terminal exon of a protein (Davison et al., 1993). However, a downstream exon has yet to be identified and this "U exon" has no known function. The hypothetical A-106 and DNA binding proteins have homologs in HAdV-4, HAdV-1, and HAdV-11. The 14.7-kDa protein has a homolog in HAdV-4, while both the 12.6-kDa and the U exon proteins have homologs in HAdV-1. The presence of these hypothetical proteins strongly suggests that the complete set of proteins encoded by the adenoviral genome sequence has not yet been identified and characterized.

Comparison of the HAdV B1 and B2 subspecies

The HAdV-B clade is further subdivided into subspecies B1 and B2. The B1 subspecies members include HAdV-3, 7, 16, 21, and 50; of these, HAdV-3, 7, and 21 have been implicated in causing respiratory disease. In contrast, members of the B2 subspecies, HAdV-11, 14, 34, and 35, cause urinary tract (UT) infections (with the exception of HAdV-11a and possibly 35). Genome sequences of HAdV-7p and 11p were compared in order to understand better the molecular mechanisms underlying this apparent difference in pathogenicity (Fig. 2). The whole genome sequences of HAdV-7p and 11p were aligned with Pipmaker (http://pipmaker.bx.psu.edu/pipmaker/), giving the result that these two genome sequences have 4768 mismatches and 1032

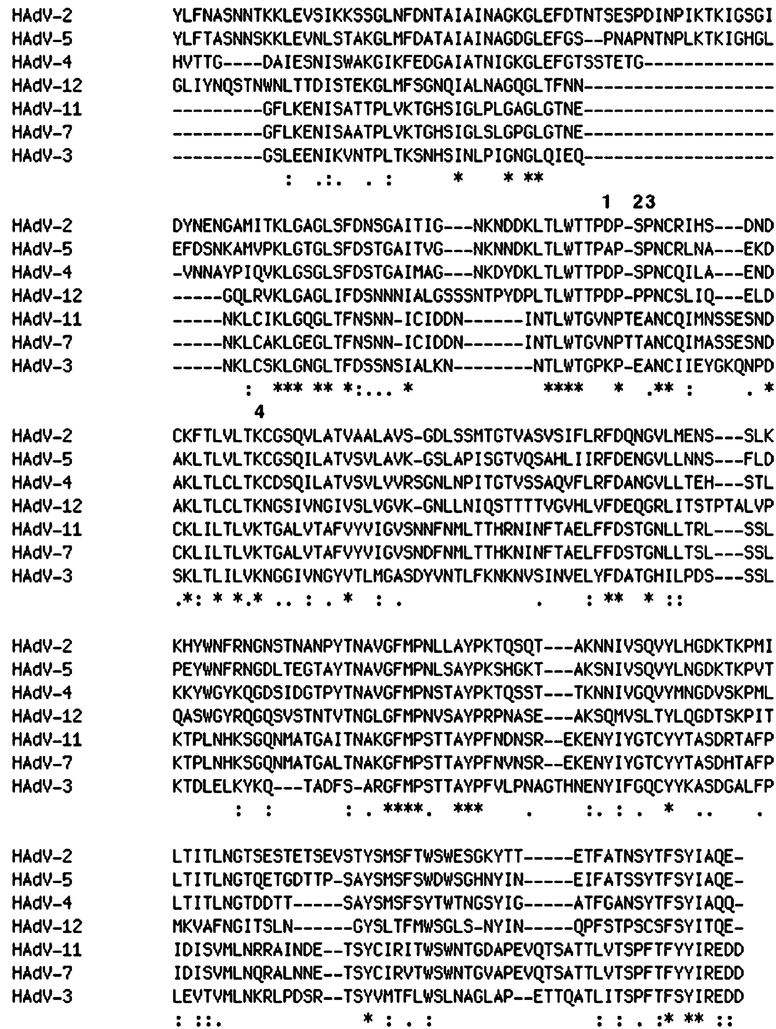

Fig. 1. Multiple sequence alignment (MSA) of fiber proteins of HAdV species A, B, C, and E. The fiber protein interacts with cell receptors, for example, Coxsackie and Adenovirus Receptor (CAR), to determine tropism. Amino acid sequences of the species B (HAdV-3 and 7), C (HAdV-2 and 5), and E (HAdV-4) fibers are aligned with the HAdV-12 (species D) fiber sequence. The three-dimensional structure of the HAdV-12 fiber has been solved, and the key residues involved in CAR binding have been mapped. Some of these key residues are marked by numbers at the top of the alignment to show conservation among the CAR binding species A, C, and E adenoviruses, and relative to HAdV-12: D415 (1), P417 (2), P418 (3), and K429 (4). CLUSTALX is used for the MSA; this software notes amino acid alignments as follows: "*" = conserved amino acid, "." = either size or hydropathy is conserved, and ":" = both size and hydropathy are conserved.

gaps between them, with a matching percentage of 83%. There are two major regions of deletion in the HAdV-11p sequence as compared to HAdV-7p (Fig. 2). The first region comprises three deleted segments of sizes 156, 13, and 44 bp. These three deletions, taken together, eliminate the VA RNA II coding region. The second deletion region is 243 bp in size and deletes the E3 7.7 kDa coding sequence completely. Both the VA RNA II and the E3 7.7-kDa protein play roles in countering the host immune response system. The potential differing responses to HAdV-7 and HAdV-11 from the host immune system may explain, in part, their different pathologies.

The adenovirus fiber binds to host cell receptors and may thus be a major determinant of cell tropism (Howitt et al., 2003). A close examination of the amino acid sequences of the HAdV-7p and 11p fibers revealed fewer differences than may be anticipated in terms of their tropisms. The fibers share an overall identity of 81% (Fig. 3). In the knob region, the part of the fiber that physically interacts with the cell receptor, the two sequences are 94% identical. Of the 11 amino acids differences in the knob region, 7 are substitutions of similar amino acids, conserved in size and hydropathy. Interestingly, it has been reported recently that the HAdV-11 serotype may have arisen as a B2-like virus, possibly HAdV-35, acquiring the fiber coding sequence from HAdV-7 via a homologous recombination event (Stone et al., 2003).

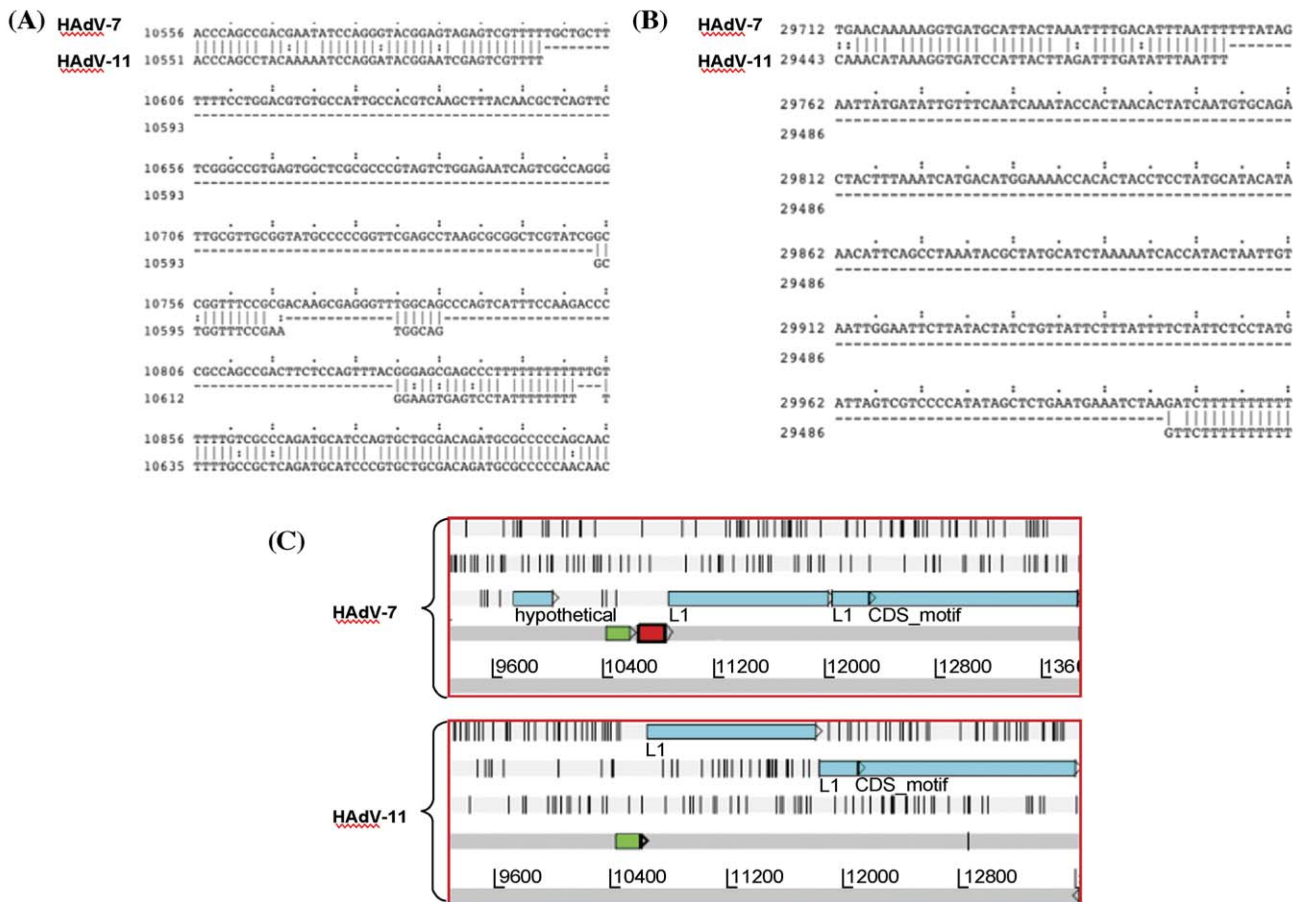

Fig. 2. Selected genome alignments of HAdV-7 and HAdV-11 showing mismatches and gaps. The whole genome sequences of HAdV-7 and 11 were aligned with Pipmaker, highlighting gross genomic differences between subspecies B1 and B2. Displayed are two major deletion regions contained within HAdV-11 as compared to HAdV-7: (A) a deletion region in HAdV-11, comprising 156-, 13-, and 44-bp deleted segments, that eliminates the VA RNA II coding sequence and (B) a 243-bp deletion that removes the E3 7.7-kDa protein coding sequence from HAdV-11. Displayed for each alignment and located between the nucleotide sequence lines for HAdV-7 and HAdV-11 are the identities, differences, and gaps between the two sequences. Key: "|" denotes a perfect match between the two nucleotides; ":" denotes a transition; and "–" denotes a gap. Panel C shows genome maps of a syntenic region from HAdV-7 and HAdV-11. These schematics of the nucleotide sequence are displayed with the three forward frames of translation. Nota bene, in particular, the highlighted VA RNA coding sequences, which are located upstream of the L1 transcription unit, show that VA RNA II is missing from HAdV-11. Color scheme: green = VA RNA I and red = VA RNA II.

Hexon amino acid (aa) sequences of HAdV-7p and 11p were analyzed. The two sequences were 86% identical (Fig. 4). Two regions of variability were identified, marked as VR A (166 aa) and VR B (34 aa). Both variable regions map onto outer loops in the hexon structures and may contribute to immunogenicity (Crawford-Miksza and Schnurr, 1996; Rux et al., 2003). In the rest of the hexon sequences, there were 17 single nucleotide polymorphisms (SNPs). Of these, 12 were of similar amino acid substitutions.

The lack of the VA RNA II and E3 7.7-kDa coding sequences appears to be the most significant difference between the genomes of HAdV-7p and 11p. HAdV-35, another species B2 UT pathogen, lacks these two coding sequences as well (data not shown). As stated earlier, this may account, in part, for the difference in the pathologies due to HAdV-7 and HAdV-11.

Comparison of ARD-relevant serotypes HAdV-7 and HAdV-4

HAdV-7 and HAdV-4 are the most prominent etiologies of the epidemic outbreaks of ARD. This similarity prompted interest in comparisons of their genomes. Whole genome alignments were performed with Pipmaker (http://pipmaker.bx.psu.edu/pipmaker/), showing that these two genome sequences are 77% identical. There are 5350 mismatches and 1275 gaps between the two sequences. A comparison of the gene contents between the two genomes shows that the HAdV-7 genome lacks the 19.4-kDa hypothetical protein, E3 6.3-kDa protein, and 3.5-kDa hypothetical protein, all of which are found in the HAdV-4 genome. On the other hand, the HAdV-4 genome lacks the following coding sequences found in HAdV-7: E2B 6.1-kDa hypothetical protein; E3 7.7-kDa protein; and L5 agnoprotein.

```
HAdV-7   MTKRVRLSDSFNPVYPYEDESTSQHPFINPGFISPNGFTQSPDGVXXXXXXXXXXXXGGS
         MTKRVRLSDSFNPVYPYEDESTSQHPFINPGFISPNGFTQSP+GV            GGS
HAdV-11  MTKRVRLSDSFNPVYPYEDESTSQHPFINPGFISPNGFTQSPNGVLTLKCLTPLTTTGGS

         LQLKVGGGLTIDDTDGFLKENISAATPLVKTGHSIGLSLGPGLGTNENKLCAKLGEGLTF
         LQLKVGGGLT+DDT+GFLKENISA TPLVKTGHSIGL LG GLGTNENKLC KLG+GLTF
         LQLKVGGGLTVDDTNGFLKENISATTPLVKTGHSIGLPLGAGLGTNENKLCIKLGQGLTF
                      F-K
         XXXXXXXXXXXXTLWTGVNPTTANCQIMASSESNDCKLILTLVKTGALVTAFVYVIGVSN
                     TLWTGVNPT ANCQIM SSESNDCKLILTLVKTGALVTAFVYVIGVSN
         NSNNICIDDNINTLWTGVNPTEANCQIMNSSESNDCKLILTLVKTGALVTAFVYVIGVSN
                              ▲      ▲

         DFNMLTTHKNINFTAELFFDXXXXXXXXXXXXXXXXNHKSGQNMATGALTNAKGFMPSTT
         +FNMLTTH+NINFTAELFFD                NHKSGQNMATGA+TNAKGFMPSTT
         NFNMLTTHRNINFTAELFFDSTGNLLTRLSSLKTPLNHKSGQNMATGAITNAKGFMPSTT

         AYPFNVNSREKENYIYGTCYYTASDHTAFPIDISVMLNQRALNNETSYCIRVTWSWNTGV
         AYPFN NSREKENYIYGTCYYTASD TAFPIDISVMLN+RA+N+ETSYCIR+TWSWNTG
         AYPFNDNSREKENYIYGTCYYTASDRTAFPIDISVMLNRRAINDETSYCIRITWSWNTGD
              ▲                   ▲

         APEVQTSATTLVTSPFTFYYIREDD
         APEVQTSATTLVTSPFTFYYIREDD
         APEVQTSATTLVTSPFTFYYIREDD
```

Fig. 3. Alignment of the HAdV-7 and HAdV-11 fiber sequences. The amino acid sequences of the HAdV-7 and HAdV-11 fibers are aligned with BLASTP (using default settings). The shaft-knob junction is marked as "F–K." Amino acid differences in the knob sequences between the two fibers are marked with black triangles. In the BLASTP algorithm, the low-complexity regions in the HAdV-7 sequence (marked as "X") are excluded from the alignment process. A consensus sequence, displayed in between the HAdV-7 and HAdV-11 sequence lines, highlights identical, similar, or different residues between the two sequences. Similar residues, conserved in either size or hydropathy, are marked as "+." Gaps are denoted as "–".

The receptor-binding function of the adenovirus fiber suggests a role in determining cell/tissue tropism, and perhaps, by extension, a role in viral patho-epidemiology. From this perspective, the fiber gene sequences of HAdV-7 and HAdV-4 were compared in greater detail. HAdV-7 and HAdV-4 fiber amino acid sequences have an overall identity of 30%. Fiber knobs, the region that physically interacts with the host cell receptor, had 99 mismatches (33 similar substitutions) and 16 gaps (Fig. 5). It is noteworthy that the fiber knob of HAdV-7 is more similar to the fiber knob of HAdV-11, a species B2 member and a UT pathogen, than to the fiber knob of HAdV-4 that, like HAdV-7, is a respiratory tract pathogen and an ARD agent. The lack of overt correlation between biology and sequence analysis suggests that the fiber may not be a critical determinant of adenovirus patho-epidemiology. A more comprehensive comparative analysis of the HAdV-B1 and HAdV-B2 species points to multiple fiber-swapping recombination events between these two species that have very different pathologies (manuscript in preparation). This further suggests that the fiber may not play a primary role in determining tissue tropism; other, as yet uncharacterized, factors may determine what tissue a serotype/species infects and the type of disease it causes.

Comparison of HAdV-7 and SAdV-21

Previous restriction enzyme digestion patterns have shown that HAdV-4 shares a close phylogenetic relationship with SAdV (Li and Wadell, 1988). Since HAdV-7 is another major pathogen for ARD, the genomes of HAdV-7 and chimpanzee SAdVs were compared for overall identities and relationships. In particular, whole genome alignments of SAdV-21 and HAdV-7, using the BLASTZ algorithm of Pipmaker, show that these genome sequences are 85% identical, with 4332 mismatches and 724 gaps. One major coding difference is that the HAdV-7 genome lacks the CR1-delta1 protein coding sequence, which is present in SAdV-21. SAdV-21, in turn, is missing two homologs of the HAdV-7 coding sequences: E3 7.7-kDa protein and L5 agnoprotein. The fiber amino acid sequences of HAdV-7 and SAdV-21 are 56% identical. An alignment of the fiber knob regions contains 92 mismatches (29 similar substitutions) and 5 gaps (Fig. 6). It appears though that HAdV-7 shares a close phylogenetic relationship with SAdV-21, suggesting a possible origin of HAdV-7 from a zoonotic event.

Comparative genomics: putative evolutionary origins through zoonosis

Again, the genome of HAdV-E species has been shown to have a strong phylogenetic relationship to four simian (chimpanzee) adenoviruses: SAdV-22, 23, 24, and 25 (Purkayastha et al., in press). Data from the above bioinformatics analysis of the HAdV-7p genome suggest a similar close phylogenetic relationship with the outlier from the HAdV-4 study: SAdV-21, another simian (chimpanzee)

Fig. 4. Alignment of the HAdV-7 and HAdV-11 hexon sequences. The amino acid sequences of the HAdV-7 and HAdV-11 hexons are aligned with BLASTP (using default parameter settings). Hexon proteins presumably contribute to the immunogenicity of the virus. The two variable regions are marked as VR A and VR B. A consensus sequence, displayed between the HAdV-7 and HAdV-11 sequences, highlights residues that are identical, similar, or different between the two sequences. Similar residues, conserved in either size or hydropathy, are marked as "+". Gaps are denoted as "–".

adenovirus. To pursue this, genome data from HAdV-3, 7, 11, and 35 were chosen to represent the B species, including both B1 and B2 subspecies; HAdV-4, the sole member, was chosen to represent the E species; and SAdV-21, 22, 23, 24, and 25 were included in the analysis to represent the chimpanzee adenoviruses. As shown in Fig. 7, six representative proteins whose coding sequences are sampled sequentially from across the full linear genome are displayed: E1A 32 kDa, E1B 55 kDa, L1 55 kDa, L2 penton, L3 hexon, L5 fiber, and E4 34 kDa. The amino acid sequences of each protein set were aligned with CLUSTALX (default settings). Phylogenetic trees were drawn using the neighbor-joining method (Saitou and Nei, 1987). This phylogeny study reconfirms a close phylogenetic relationship between the HAdV-Bs and SAdV-21. Therefore, it is likely that the HAdV-B species arose from an interspecies transmission of a SAdV-21-like virus from chimpanzees to humans. Given the diversity among the species B members and the number of HAdV-7 genome types, this transmission is probably not a recent evolutionary event. However, it is also theoretically possible that SAdV-21 has evolved from a HAdV-B-like virus that underwent a human-to-chimpanzee transmission.

The emergence of new viral species from zoonotic events has been well documented in at least two cases (Gao et al., 1999; Hillis et al., 1968; Slattery et al., 1999). It has been noted that two strains of HIV, for example, HIV-1 and HIV-2, are believed to represent cross-species infections to humans from chimpanzee (*Pan troglodytes*) (HIV-1) and sooty mangabey monkey (*Cercocebus torqua-*

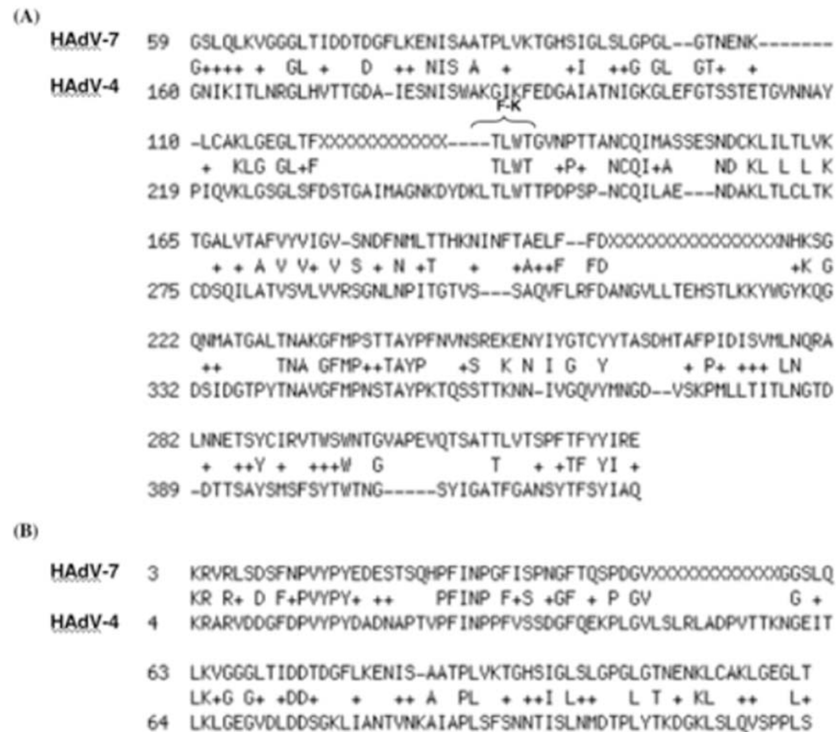

Fig. 5. Alignment of the ARD-relevant HAdV-7 and HAdV-4 fiber sequences. The amino acid sequences of the HAdV-7 and HAdV-4 fibers are aligned with BLASTP (using default parameter settings). There are two regions of alignment: (A) C-terminal knob regions (30% identity), where the shaft-knob junction is marked as "F–K", and (B) N-terminal base region, which interacts with the penton base protein (36% identity). In the BLASTP algorithm, the low-complexity regions in the HAdV-7 sequence (marked as "X") are excluded from the alignment calculation. A consensus sequence, displayed in between the HAdV-7 and HAdV-4 lines, highlights residues that are identical, similar, or different between the two sequences. Similar residues, conserved in either size or hydropathy, are marked as "+". Gaps are denoted as "–".

tus) (HIV-2) (Gao et al., 1999; Hirsch et al., 1989). Similarly, evolutionary studies on human and simian T-cell leukemia/lymphotropic viruses (HTLV/STLV) indicate that HTLV I and II are the result of separate interspecies transfers between simian species and humans (Slattery et al., 1999).

Fig. 6. Alignment of two B1 species, HAdV-7 and SAdV-21, fiber sequences. The amino acid sequences of the HAdV-7 and SAdV-21 fibers are aligned with BLASTP (default settings). The sequences share an overall identity of 56%. The shaft-knob junction is marked as "F–K". A consensus sequence, displayed in between the HAdV-7 and SAdV-21 sequence lines, highlights residues that are identical, similar, or different between the two sequences. Similar residues, conserved in either size or hydropathy, are marked as "+". Gaps are denoted as "–".

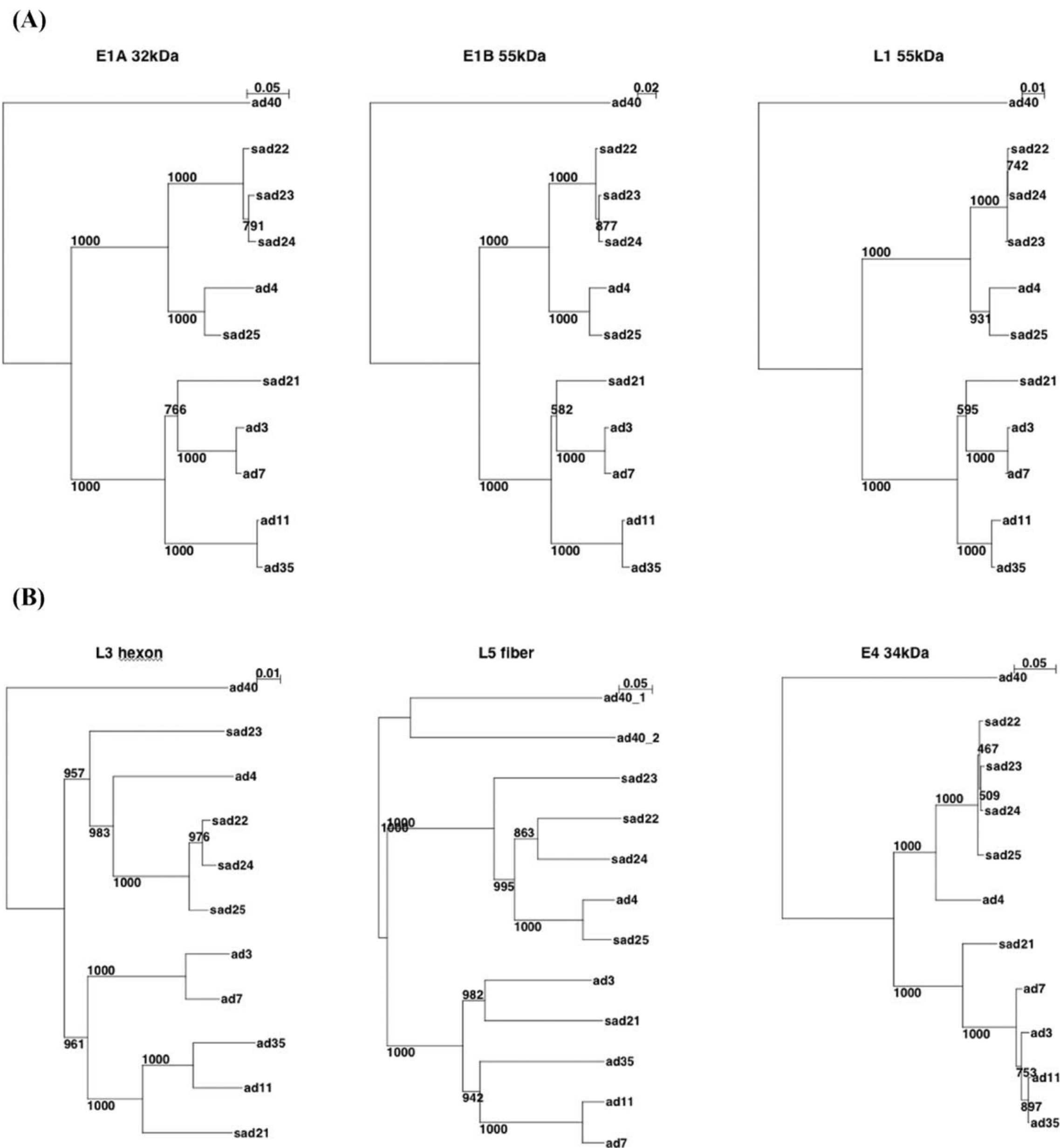

Fig. 7. Phylogenetic analyses of selected HAdV-7 proteins spanning the genome. The amino acid sequences of six proteins from five human adenoviruses and five simian (chimpanzee) adenoviruses are aligned using CLUSTALX using default parameters. The sequences of HAdV serotypes 3, 7, 11, and 35 are used to represent the B species (both B1 and B2 subspecies), while the sequences of serotype 4 are chosen to represent the E species. The sequences of HAdV-40 are used as the "out-group" for reference in all of the trees. These trees are constructed using software incorporating the neighbor-joining method (Saitou and Nei, 1987). The robustness of the trees is measured by bootstrapping (1000 replications). Numbers indicate bootstrap values in support of the adjacent node. (A) Phylogenetic analyses of the E1A 32-kDa, E1B 55-kDa, and L1 55 kDa proteins sequences. (B) Phylogenetic analyses of the L3 hexon, L5 fiber, and E4 34-kDa protein sequences.

Biotechnology implications of HAdV-7 and SAdV genomes: vector development

The evolutionary findings presented here also have far-reaching implications, especially in the growing field of vector construction for DNA vaccine development and human gene therapy. The current HAdV-5-derived vectors can cause problems in human gene therapy protocols, including fatality (Raper et al., 2003). In general, host preexisting immunity to HAdV is enough of a concern to stimulate the development of alternative AdV vectors to which neutralizing antibodies would be rare in the human population. To this end, SAdVs are being considered as alternatives to human adenoviruses in the design of gene and vaccine delivery vectors (Roy et al., 2004). One study reporting lack of neutralizing antibodies against chimpanzee AdV in human serum samples suggests vectors derived from them will be useful as vaccine and gene therapy vectors (Roy et al., 2004). This observation is supported by earlier reports in the literature using serum neutralization, comple-

ment fixation, and hemagglutination-inhibition studies (Basnight et al., 1971; Soike et al., 1969). Intriguingly, one of these reports noted no cross-reaction of the described chimpanzee AdV with antisera of known simian adenovirus types (Basnight et al., 1971).

A conflicting report noted that SAdVs were neutralized by rabbit antisera against the HAdV-4 prototype virion (Li and Wadell, 1988). This observation is supported by others in the literature also using serum neutralization, complement fixation, and hemagglutination-inhibition studies to demonstrate cross-reaction with antisera against HAdV (Hillis et al., 1968; Rowe et al., 1958).

Clearly, there are enough discrepancies in these gold standard clinical assays to warrant further detailed examination into the use of chimpanzee SAdV-based vectors in place of the current HAdV-5-derived vectors for both human gene therapy and DNA vaccine development. The bioinformatics and phylogenetic analyses presented here with HAdV-7, as well as for the genome and bioinformatics analysis for HAdV-4 (Purkayastha et al., in press), suggest that a stronger link may exist between HAdV and SAdV than previously thought. These relationships of the genomes as well as the individual proteins, including the critical determinant outer proteins, coupled with a potential for cross-reaction between preexisting antibodies against HAdV and SAdV-derived vectors with the human host should give pause and concern.

Advanced molecular diagnostics

Analysis of HAdV-7 and related genomes allows the development, validation, and implementation of molecular-based diagnostic assays, particularly through the identification of unique pathogen genome signature sequences. Rapid sequence determinations, coupled with the versatility of these assays, allow nearly "on-the-fly" adjustments of probes and arrays to meet the challenge of new and emerging naturally occurring infectious agents, such as SARS (Ruan et al., 2003), as well as potentially troubling BioThreat agents. The U.S. Air Force Surgeon General Office has supported, as part of the Directorate of Modernization, the Epidemic Outbreaks Surveillance (EOS) program to develop and validate a Respiratory Pathogen Microarray (RPM) assay, based on the Affymetrix technology, in a real-world test bed.

Materials and methods

Cells, virus stocks, and DNA preparation

The American Type Culture Collection (ATCC) is the source for HAdV-7p (ATCC #VR-7). Virus stocks were expanded in A-549 cells (ATCC #CCL-185), a human lung tumor cell line that proved optimal in the genomic study of HAdV-1 (Lauer et al., 2004). Viral and DNA samples were prepared according to protocols reported earlier (Lauer et al., 2004; Le et al., 2001).

DNA sequencing

The HAdV-7p genome was sequenced using a "leveraged" primer walk strategy (Lauer et al., 2004), as primers for PCR amplification and sequencing were available from an earlier HAdV genome sequencing project and from various HAdV-7 gene sequences archived in GenBank. Custom primers were chosen from newly sequenced regions if the sequences had diverged sufficiently from the reference genome to render the original primers useless. Gaps in sequence coverage were closed by amplifying with PCR methodology across the gap using custom and newly identified sequences and also by sequencing the amplicon. Primers were designed either side of the gap using the sequenced regions as reference (with a preference for 19 mers with a 60 °C annealing temperature). PCR was performed using the BIOLASE DNA Polymerase Core Kit (Bioline, Randolph, MA) and 100–200 ng of purified adenoviral DNA as template. Fifty-microliter reactions were used, as per manufacturer's recommendations, and cycled 25 times at 95 °C for 30 s (s), 55 °C for 30 s, 72 °C for 60 s. PCR reactions were carried out in a MJ Research thermocycler (Watertown, MA). Amplicons were purified using the QIAquick® PCR Purification Kit (QIAGEN, Valencia, CA) and sequenced (200 ng per reaction) with the PCR primers using the DYEnamic ET Terminator Cycle Sequencing kit (Amersham Biosciences, Piscataway, NJ) on an ABI Prism 377 DNA Sequencer (Applied Biosystems, Foster City, CA).

Direct sequencing of ITR ends

The 5′ and 3′ ends of the linear HAdV-7 genome were determined by direct sequencing of the repurified genomic DNA. Primers were designed from newly obtained internal sequences (usually internal to the first *Sau* 3A1 recognition site). Template DNA (0.2–1.0 μg/reaction) was repurified by passing through a MicroSpin G-50 column (Amersham Biosciences) and sequenced with the PCR primers using the DYEnamic ET Terminator Cycle Sequencing kit (Amersham Biosciences) on an ABI Prism 377 DNA Sequencer (Applied Biosystems, Inc.).

Sequence assembly and analysis: genome annotation

DNA sequences were trimmed and assembled using Sequencher 4.1.1 (Gene Codes Corporation, Inc.; Ann Arbor, MI). Genome features of the DNA sequence were revealed using the Wisconsin GCG package (SeqWeb v.2).

The genome sequence of HAdV-7 was annotated by parsing the 35 306 bases into 1-kilobase (kb) nonoverlapping segments, as per an algorithm developed, and applied successfully to earlier HAdV-1 genome analyses (Lauer

et al., 2004). Briefly, these segments were queried systematically against the nonredundant NCBI database using the BLASTX program of the BLAST suite sequence-alignment software (Altschul et al., 1990). Default parameters of word size = 3 and expectation = 10, with the BLOSUM62 substitution matrix and with gap penalties of 11 (existence) and 1 (extension), were applied to these analyses. Low-complexity sequences were filtered out of the queries, as per the BLAST algorithm.

For the prediction and identification of putative genes or previously 'not-identified' genes, the web-based software tool GenomeScan was used (Yeh et al., 2001). This tool identified putative exons from the coding sequences, where exon–intron borders were difficult to determine, using exon–intron identification combined with similarity searches to a sequence database (Yeh et al., 2001). These hypothetical proteins were also independently identified using another gene prediction software tool GeneMark (Besemer and Borodovsky, 1999). GeneMark is a fifth-order Hidden Markov Method (HMM)-based gene prediction software. During the course of this annotation, it was noted that while GeneMark had a slightly higher accuracy than GenomeScan, neither gene prediction tool was completely accurate or comprehensive in generating a list of putative genes.

To enhance annotation progress and visualization, Artemis, a genome annotation and editor tool (Berriman and Rutherford, 2003), was used for the annotation process. Both the DNA sequence and the annotation of the HAdV-7p genome are accessible from GenBank (GenBank accession no. AY594255).

Whole genome comparisons

PipMaker (http://bio.cse.psu.edu/pipmaker) was used to perform nucleotide dot-blot identity-based analyses of the whole AdV genomes now available from GenBank, using this HAdV-7p as a reference. These nucleotide comparisons were confirmed using other similar tools, such as Wisconsin GCG package "Compare."

Acknowledgments

We thank Dr. Mary Lou Gibson (Virapur, Inc.; San Diego, CA) for virus growth and DNA preparation. We also thank Jason Seto, Brenda Desjardins, and Charles Caes for their critical comments. Research support was provided through a grant (#DAMD17-03-2-0089) from the U.S. Army Medical Research and Material Command (USAMRMC). Partial support also provided through the Epidemic Outbreak Surveillance Project (EOS), funded through HQ USAF Surgeon General Office (SGR) and the Defense Threat Reduction Agency.

The opinions and assertions contained herein are the private ones of the authors and are not to be construed as official or reflecting the views of the Department of Defense.

During the course of this work, the EOS Consortium included:

Sponsorship: Col. Peter F. Demitry,[1] Lt. Col. Theresa Lynn Difato.[1]

Executive Board and Principal Investigator: Maj Robb K. Rowley,[1] Lt. Col. Eric H. Hanson,[1] Rosana R. Holliday,[2] Clark Tibbetts.[3]

Operational Board and Senior Scientists: Jerry Diao,[2] Curtis White,[4] Elizabeth A. Walter,[5] Russell P. Kruzelock,[6] Jennifer Weller,[7] Donald Seto,[7] David A. Stenger,[8] Maj Brian K. Agan.[9]

Technical Advisors and Collaborating Investigators: CDR Kevin Russell,[10] David Metzgar,[10] Jianguo Wu,[10] Ted Hadfield.[11]

Research and Clinical Staff: Anjan Purkayastha,[12] Jing Su,[12] Baochuan Lin,[8] Dzung Thach,[8] Gary J.Vora,[8] Zheng Wang,[8] Chris Olsen,[2] John Gomez,[4] Jose J. Santiago,[4] Margaret Jesse,[4] Sue A. Worthy,[4] Sue Ditty,[11] John McGraw,[11] Robert Crawford,[11] TSgt Michael Jenkins,[9] Linda Canas.[13]

Operations Support Staff: Cheryl J. James,[2] Kathy Ward,[2] Kenya Grant,[2] Kindra Nix.[4]

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215 (3), 403–410.

Basnight Jr., M., Rogers, N.G., Gibbs Jr., C.J., Gajdusek, D.C., 1971. Characterization of four new adenovirus serotypes isolated from chimpanzee tissue explants. Am. J. Epidemiol. 94 (2), 166–171.

Berge, T.O., England, B., Mauris, C., Shuey, H.E., Lennette, E.H., 1955. Etiology of acute respiratory disease among service personnel at Fort Ord, California. Am. J. Hyg. 62 (3), 283–294.

Berriman, M., Rutherford, K., 2003. Viewing and annotating sequence data with Artemis. Brief. Bioinform. 4 (2), 124–132.

Besemer, J., Borodovsky, M., 1999. Heuristic approach to deriving models for gene finding. Nucleic Acids Res. 27 (19), 3911–3920.

Binger, M.H., Flint, S.J., 1984. Accumulation of early and intermediate mRNA species during subgroup C adenovirus productive infections. Virology 136 (2), 387–403.

Chanock, R.M., Ludwig, W., Heubner, R.J., Cate, T.R., Chu, L.W., 1966. Immunization by selective infection with type 4 adenovirus grown in

---

[1] Department of USAF/SGR.
[2] Department of USAF/SGR (Contractor).
[3] The George Washington University (IPA).
[4] Lackland AFB, TX.
[5] Texas A and M University San Antonio (IPA).
[6] Virginia Tech (IPA).
[7] George Mason University (IPA).
[8] Naval Research Laboratory.
[9] Wilford Hall Medical Center.
[10] Navy Health Research Center.
[11] Armed Forces Institute of Pathology.
[12] George Mason University.
[13] Air Force Institute for Operational Health.

human diploid tissue cultures: I. Safety and lack of oncogenicity and tests for potency in volunteers. JAMA 195 (6), 445–452.

Crawford-Miksza, L., Schnurr, D.P., 1996. Analysis of 15 adenovirus hexon proteins reveals the location and structure of seven hypervariable regions containing serotype-specific residues. J. Virol. 70 (3), 1836–1844.

Crawford-Miksza, L.K., Nang, R.N., Schnurr, D.P., 1999. Strain variation in adenovirus serotypes 4 and 7a causing acute respiratory disease. J. Clin. Microbiol. 37 (4), 1107–1112.

Crespo, R., Shivaprasad, H.L., Droual, R., Chin, R.P., Woolcock, P.R., Carpenter, T.E., 1998. Inclusion body tracheitis associated with avian adenovirus in turkeys. Avian Dis. 42 (3), 589–596.

Dan, A., Elo, P., Harrach, B., Zadori, Z., Benko, M., 2001. Four new inverted terminal repeat sequences from bovine adenoviruses reveal striking differences in the length and content of the ITRs. Virus Genes 22 (2), 175–179.

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 (4), 1308–1316.

Davison, A.J., Wright, K.M., Harrach, B., 2000. DNA sequence of frog adenovirus. J. Gen. Virol. 81 (Pt. 10), 2431–2439.

Davison, A.J., Benko, M., Harrach, B., 2003. Genetic content and evolution of adenoviruses. J. Gen. Virol. 84 (Pt. 11), 2895–2908.

Defer, C., Belin, M.T., Caillet-Boudin, M.L., Boulanger, P., 1990. Human adenovirus–host cell interactions: comparative study with members of subgroups B and C. J. Virol. 64 (8), 3661–3673.

Deryckere, F., Burgert, H.G., 1996. Early region 3 of adenovirus type 19 (subgroup D) encodes an HLA-binding protein distinct from that of subgroups B and C. J. Virol. 70 (5), 2832–2841.

Erdman, D.D., Xu, W., Gerber, S.I., Gray, G.C., Schnurr, D., Kajon, A.E., Anderson, L.J., 2002. Molecular epidemiology of adenovirus type 7 in the United States, 1966–2000. Emerging Infect Dis. 8 (3), 269–277.

Farkas, S.L., Benko, M., Elo, P., Ursu, K., Dan, A., Ahne, W., Harrach, B., 2002. Genomic and phylogenetic analyses of an adenovirus isolated from a corn snake (*Elaphe guttata*) imply a common origin with members of the proposed new genus Atadenovirus. J. Gen. Virol. 83 (Pt. 10), 2403–2410.

Finnen, R.L., Biddle, J.F., Flint, J., 2001. Truncation of the human adenovirus type 5 L4 33-kDa protein: evidence for an essential role of the carboxy-terminus in the viral infectious cycle. Virology 289 (2), 388–399.

Flint, J., Shenk, T., 1997. Viral transactivating proteins. Annu. Rev. Genet. 31, 177–212.

Gao, F., Bailes, E., Robertson, D.L., Chen, Y., Rodenburg, C.M., Michael, S.F., Cummins, L.B., Arthur, L.O., Peeters, M., Shaw, G.M., Sharp, P.M., Hahn, B.H., 1999. Origin of HIV-1 in the chimpanzee Pan troglodytes. Nature 397 (6718), 436–441.

Gaydos, C.A., Gaydos, J.C., 1995. Adenovirus vaccines in the U.S. military. Mil. Med. 160 (6), 300–304.

Hasson, T.B., Soloway, P.D., Ornelles, D.A., Doerfler, W., Shenk, T., 1989. Adenovirus L1 52- and 55-kilodalton proteins are required for assembly of virions. J. Virol. 63 (9), 3612–3621.

Hayes, B.W., Telling, G.C., Myat, M.M., Williams, J.F., Flint, S.J., 1990. The adenovirus L4 100-kilodalton protein is necessary for efficient translation of viral late mRNA species. J. Virol. 64 (6), 2732–2742.

Hierholzer, J., 1995. Adenoviruses. In: L.D. Lennette, E.H., Lennette, E.T. (Eds.), Diagnostic Procedures for Viral, Rickettsial and Chlamydial Infections, 7th ed. American Public Health Association, Washington, pp. 169–188.

Hillis, W.D., Holmes, A.W., Davison, V., 1968. Serologic characterization of adenoviruses isolated from chimpanzees associated with viral hepatitis. Proc. Soc. Exp. Biol. Med. 129 (2), 366–369.

Hirsch, V.M., Olmsted, R.A., Murphey-Corb, M., Purcell, R.H., Johnson, P.R., 1989. An African primate lentivirus (SIVsm) closely related to HIV-2. Nature 339 (6223), 389–392.

Hong, J.S., Mullis, K.G., Engler, J.A., 1988. Characterization of the early region 3 and fiber genes of Ad7. Virology 167 (2), 545–553.

Horton, T.M., Tollefson, A.E., Wold, W.S., Gooding, L.R., 1990. A protein serologically and functionally related to the group C E3 14,700-kilodalton protein is found in multiple adenovirus serotypes. J. Virol. 64 (3), 1250–1255.

Howitt, J., Anderson, C.W., Freimuth, P., 2003. Adenovirus interaction with its cellular receptor CAR. In: Doerfler, W., Böhm, P. (Eds.), Adenoviruses: Model and Vectors in VirusHost Interactions. Springer, Berlin, pp. 331–364.

Kajon, A., Wadell, G., 1994. Genome analysis of South American adenovirus strains of serotype 7 collected over a 7-year period. J. Clin. Microbiol. 32 (9), 2321–2323.

Kidd, A.H., Garwicz, D., Oberg, M., 1995. Human and simian adenoviruses: phylogenetic inferences from analysis of VA RNA genes. Virology 207 (1), 32–45.

Kovacs, G.M., LaPatra, S.E., D'Halluin, J.C., Benko, M., 2003. Phylogenetic analysis of the hexon and protease genes of a fish adenovirus isolated from white sturgeon (*Acipenser transmontanus*) supports the proposal for a new adenovirus genus. Virus Res. 98 (1), 27–34.

Lauer, K.P., Llorente, I., Blair, E., Seto, J., Krasnov, V., Purkayastha, A., Ditty, S.E., Hadfield, T.L., Buck, C., Tibbetts, C., Seto, D., 2004. Natural variation among human adenoviruses: genome sequence and annotation of human adenovirus serotype 1. J. Gen. Virol. 85 (Pt. 9), 2615–2625.

Le, C.T., Gray, G.C., Poddar, S.K., 2001. A modified rapid method of nucleic acid isolation from suspension of matured virus: applied in restriction analysis of DNA from an adenovirus prototype strain and a patient isolate. J. Med. Microbiol. 50 (6), 571–574.

Leppard, K.N., 1997. E4 gene function in adenovirus, adenovirus vector and adeno-associated virus infections. J. Gen. Virol. 78 (Pt. 9), 2131–2138.

Li, Q.G., Wadell, G., 1986. Analysis of 15 different genome types of adenovirus type 7 isolated on five continents. J. Virol. 60 (1), 331–335.

Li, Q.G., Wadell, G., 1988. The degree of genetic variability among adenovirus type 4 strains isolated from man and chimpanzee. Arch. Virol. 101 (1–2), 65–77.

Li, Q., Wadell, G., 1999. Genetic variability of hexon loops 1 and 2 between seven genome types of adenovirus serotype 7. Arch. Virol. 144 (9), 1739–1749.

Mathews, M.B., Shenk, T., 1991. Adenovirus virus-associated RNA and translation control. J. Virol. 65 (11), 5657–5662.

Munoz, F.M., Piedra, P.A., Demmler, G.J., 1998. Disseminated adenovirus disease in immunocompromised and immunocompetent children. Clin. Infect. Dis. 27 (5), 1194–1200.

Oosterom-Dragon, E.A., Ginsberg, H.S., 1981. Characterization of two temperature-sensitive mutants of type 5 adenovirus with mutations in the 100,000-dalton protein gene. J. Virol. 40 (2), 491–500.

Purkayastha, A., Ditty, S.E., Su, J., McGraw, J., Hadfield, T.L., Tibbetts, C., Seto, D., 2004. Genomic and bioinformatics analysis of HAdV-4, a human adenovirus causing acute respiratory disease (ARD). Implications for gene therapy and vaccine vector development. J. Virol. (in press).

Raper, S.E., Chirmule, N., Lee, F.S., Wivel, N.A., Bagg, A., Gao, G.P., Wilson, J.M., Batshaw, M.L., 2003. Fatal systemic inflammatory response syndrome in a ornithine transcarbamylase deficient patient following adenoviral gene transfer. Mol. Genet. Metab. 80 (1–2), 148–158.

Rosa-Calatrava, M., Grave, L., Puvion-Dutilleul, F., Chatton, B., Kedinger, C., 2001. Functional analysis of adenovirus protein IX identifies domains involved in capsid stability, transcriptional activity, and nuclear reorganization. J. Virol. 75 (15), 7131–7141.

Rowe, W.P., Huebner, R.J., Gilmore, L.K., Parrott, R.H., Ward, T.G., 1953. Isolation of a cytopathic agent from human adenoids undergoing spontaneous degradation in tissue culture. Proc. Soc. Exp. Biol. Med. 84, 570–573.

Rowe, W.P., Hartley, J.W., Huebner, R.J., 1958. Serotype composition of the adenovirus group. Proc. Soc. Exp. Biol. Med. 97 (2), 465–470.

Roy, S., Gao, G., Lu, Y., Zhou, X., Lock, M., Calcedo, R., Wilson, J.M., 2004. Characterization of a family of chimpanzee adenoviruses and

development of molecular clones for gene transfer vectors. Hum. Gene Ther. 15 (5), 519–530.

Ruan, Y.J., Wei, C.L., Ee, A.L., Vega, V.B., Thoreau, H., Su, S.T., Chia, J.M., Ng, P., Chiu, K.P., Lim, L., Zhang, T., Peng, C.K., Lin, E.O., Lee, N.M., Yee, S.L., Ng, L.F., Chee, R.E., Stanton, L.W., Long, P.M., Liu, E.T., 2003. Comparative full-length genome sequence analysis of 14 SARS coronavirus isolates and common mutations associated with putative origins of infection. Lancet 361 (9371), 1779–1785.

Rux, J.J., Kuser, P.R., Burnett, R.M., 2003. Structural and phylogenetic analysis of adenovirus hexons by use of high-resolution X-ray crystallographic, molecular modeling, and sequence-based methods. J. Virol. 77 (17), 9553–9566.

Ryan, M., Gray, G., Hawksworth, A., Malasig, M., Hudspeth, M., Poddar, S., 2000. The Naval Health Research Center Respiratory Disease Laboratory. Mil. Med. 165 (7 Suppl. 2), 32–34.

Ryan, M.A.K., Gray, G.C., Malasig, M.D., Binn, L.N., Asher, L.V., Cute, D., Kehl, S.C., Dunn, B.E., Yund, A.J., 2001. Two fatal cases of adenovirus-related illness in previously healthy young adults—Illinois, 2000. Morb. Mortal. Wkly. Rep. 50 (26), 553–555.

Ryan, M.A., Gray, G.C., Smith, B., McKeehan, J.A., Hawksworth, A.W., Malasig, M.D., 2002. Large epidemic of respiratory illness due to adenovirus types 7 and 3 in healthy young adults. Clin. Infect. Dis. 34 (5), 577–582.

Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4 (4), 406–425.

San Martin and Burnett, 2003. Structural studies on adenoviruses. In: Doerfler, W.a.B.P. (Ed.), Adenoviruses: Model and Vectors in Virus–Host Interactions. Sringer, Berlin, pp. 57–94.

Sargent, K.L., Meulenbroek, R.A., Parks, R.J., 2004. Activation of adenoviral gene expression by protein IX is not required for efficient virus replication. J. Virol. 78 (10), 5032–5037.

Schmitz, H., Wigand, R., Heinrich, W., 1983. Worldwide epidemiology of human adenovirus infections. Am. J. Epidemiol. 117 (4), 455–466.

Shenk, T., 1996. Adenoviridae: the viruses and their replication. In: Fields, B., Knipe, D.M., Howley, P.M., et al., (Eds.), Fields Virology, 3rd ed. Lippincott-Raven, Philadelphia, pp. 2111–2148.

Slattery, J.P., Franchini, G., Gessain, A., 1999. Genomic evolution, patterns of global dissemination, and interspecies transmission of human and simian T-cell leukemia/lymphotropic viruses. Genome Res. 9 (6), 525–540.

Soike, K.F., Douglas, J.D., Plager, H., Coulston, F., 1969. Further studies of the viruses of the chimpanzee: II. Viral infection in recently imported chimpanzees captured in Africa. Exp. Mol. Pathol. 11 (3), 333–339.

Stone, D., Furthmann, A., Sandig, V., Lieber, A., 2003. The complete nucleotide sequence, genome organization, and origin of human adenovirus type 11. Virology 309 (1), 152–165.

Temperley, S.M., Hay, R.T., 1992. Recognition of the adenovirus type 2 origin of DNA replication by the virally encoded DNA polymerase and preterminal proteins. EMBO J. 11 (2), 761–768.

Tokunaga, O., Shinagawa, M., Padmanabhan, R., 1982. Physical mapping of the genome and sequence analysis at the inverted terminal repetition of adenovirus type 4 DNA. Gene 18 (3), 329–334.

Top Jr., F.H., Buescher, E.L., Bancroft, W.H., Russell, P.K., 1971a. Immunization with live types 7 and 4 adenovirus vaccines: II. Antibody response and protective effect against acute respiratory disease due to adenovirus type 7. J. Infect. Dis. 124 (2), 155–160.

Top Jr., F.H., Grossman, R.A., Bartelloni, P.J., Segal, H.E., Dudding, B.A., Russell, P.K., Buescher, E.L., 1971b. Immunization with live types 7 and 4 adenovirus vaccines: I. Safety, infectivity, antigenicity, and potency of adenovirus type 7 vaccine in humans. J. Infect. Dis. 124 (2), 148–154.

Wadell, G., Varsanyi, T.M., 1978. Demonstration of three different subtypes of adenovirus type 7 by DNA restriction site mapping. Infect. Immun. 21 (1), 238–246.

Wadell, G., Varsanyi, T.M., Lord, A., Sutton, R.N., 1980. Epidemic outbreaks of adenovirus 7 with special reference to the pathogenicity of adenovirus genome type 7b. Am. J. Epidemiol. 112 (5), 619–628.

Wadell, G., Cooney, M.K., da Costa Linhares, A., de Silva, L., Kennett, M.L., Kono, R., Gui-Fang, R., Lindman, K., Nascimento, J.P., Schoub, B.D., et al., 1985. Molecular epidemiology of adenoviruses: global distribution of adenovirus 7 genome types. J. Clin. Microbiol. 21 (3), 403–408.

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 (2), 309–319.

Wold, W.S., Gooding, L.R., 1991. Region E3 of adenovirus: a cassette of genes involved in host immunosurveillance and virus–cell interactions. Virology 184 (1), 1–8.

Yeh, R.F., Lim, L.P., Burge, C.B., 2001. Computational inference of homologous gene structures in the human genome. Genome Res. 11 (5), 803–816.

Yew, P.R., Liu, X., Berk, A.J., 1994. Adenovirus E1B oncoprotein tethers a transcriptional repression domain to p53. Genes Dev. 8 (2), 190–202.

Young, C.S., 2003. The structure and function of the adenovirus major late promoter. Curr. Top. Microbiol. Immunol. 272, 213–249.

Zhang, W., Low, J.A., Christensen, J.B., Imperiale, M.J., 2001. Role for the adenovirus IVa2 protein in packaging of viral DNA. J. Virol. 75 (21), 10446–10454.

Zu, Y.L., Takamatsu, Y., Zhao, M.J., Maekawa, T., Handa, H., Ishii, S., 1992. Transcriptional regulation by a point mutant of adenovirus-2 E1a product lacking DNA binding activity. J. Biol. Chem. 267 (28), 20181–20187.