

Sequence analysis of the left end of fowl adenovirus genomes

Juan Carlos Corredor · Peter J. Krell ·
É. Nagy

Received: 27 October 2005 / Accepted: 15 November 2005
© Springer Science+Business Media, LLC 2006

Abstract Nucleotide sequence analysis of the left end of the genome of fowl adenoviruses (FAdV) representing species group *C* (FAdV-4 and -10), *D* (FAdV-2) and *E* (FAdV-8) were carried out, and the sequence data was compared to those of FAdV-1 (*FAdV-A*) and FAdV-9 (*FAdV-D*). The viruses were propagated in chicken hepatoma cell line for viral DNA isolation. Restriction endonuclease analysis was performed followed by hybridization with two DNA probes representing the left end of FAdV-9. The identified fragments were sequenced, and the generated data were compared with the GenBank database. Nucleotide sequence homology and amino acid sequence identities were high between members of the same species group, FAdV-2 and -9, and FAdV-4 and -10, whereas different degrees of variations were observed among all FAdVs. Gene arrangement and position of ORFs at the left end of FAdV genomes were largely conserved suggesting similar gene functions. All previously characterized left end ORFs in CELO virus and FAdV-9 were found in all analyzed FAdVs. However, ORF 1C was absent in FAdV-4

and -10, but additional ORFs, most likely corresponding to duplicates of ORF 14, were observed in these viruses.

Keywords Fowl adenovirus (FAdV) · Sequence analysis · Amino acid sequence identity · Gene arrangement · Gene duplication

Introduction

The genus *Aviadenovirus* consists of five fowl adenovirus (FAdV) and one goose adenovirus (GoAdV) species. The FAdV species groups have been defined according to the restriction profiles as described by Zsák and Kisary [1]. Each species group contains one or more serotypes as follows: *FAdV-A* (FAdV-1), *FAdV-B* (FAdV-5), *FAdV-C* (FAdV-4 and -10), *FAdV-D* (FAdV-2, -3, -9 and -11), and *FAdV-E* (FAdV-6, -7, -8a and -8b).

FAdVs have the largest genomes among members of the family *Adenoviridae* [2, 3]. This property plus the low pathogenicity in some of these viruses make FAdVs attractive as recombinant vaccines or gene transfer vectors. Although recombinant vectors based on FAdV-1 (CELO virus), FAdV-9, -8 and -10 have been reported [4–8], very limited information on the molecular biology of these viruses is available. In fact, the complete genome sequence has been reported for only FAdV-1 (CELO virus) and FAdV-9 [2, 3], whereas only partial sequences are available for FAdV-4, -8 and -10 [7, 9].

The left and right ends of the FAdV genomes have no nucleotide sequence homologies to the mastadenoviral early (E) genes 1 (E1), 3 (E3) and 4 (E4),

The sequences reported in this work have been deposited in the Genbank database and assigned the accession numbers DQ208708-DQ208711.

J. C. Corredor · É. Nagy (✉)
Department of Pathobiology, Ontario Veterinary College,
University of Guelph, N1G 2W1 Guelph, Ontario, Canada
e-mail: enagy@ovc.uoguelph.ca

P. J. Krell
Department of Molecular and Cellular Biology, University
of Guelph, N1G 2W1 Guelph, Ontario, Canada

whereas E2, IVA2 and late genes, located in the central part of the genome, are conserved in all members of the family *Adenoviridae* [10]. These observations suggest that novel viral proteins with functional equivalence with the mastadenoviral early genes may be participating in the replication cycle of these viruses. In fact, GAM-1 protein of CELO virus, encoded by the *Gam1* gene at the right end of the genome, has functional equivalence with the mastadenoviral E1 gene products [2, 4, 11, 12], and its importance in viral replication has been reported [13].

Transcriptional and sequence analyses for the genomes of FAdV-1 and -9 have been reported [2, 14–16]. Gene function studies at the left and right ends of the genome of FAdV-1 have been carried out to determine the essentiality of the genes mapping to these locations to be able to use them for cloning of foreign genes [4, 5]. In our laboratory, FAdV-9 has been used as a recombinant virus vector to express the enhanced-green fluorescent protein whose gene was cloned into the tandem repeat region 2 (TR-2), located at the right end of the genome [5], and its growth in vitro and in vivo has been characterized [17]. All these preliminary data suggest the importance of the left and right ends of FAdVs during their replication cycle, vector design, and, probably, pathogenesis.

The nomenclature for naming the ORFs of the FAdV genomes has recently been unified [10]. For example, the left end genomes of the viruses are arranged (from 5' to 3') as follows: ORF 0, ORF 1 (dUTPase homologue), ORF 1A, ORF 1B, ORF 1C, ORF 2 [Rep protein homologue of adeno-associated virus (AAV)], ORF 24 (absent in the genome of CELO virus), ORF 14, ORF 13, and ORF 12 [10]. Previous studies have suggested that ORFs 12 and 13 are duplicates of ORF 2 in the opposite strand, based on identities and amino acid conservation among them [10, 18]. The presence of the Walker motifs, found in the superfamily III helicases, in the Rep protein homologue in AAV-3 and ORFs 2, 12 and 13 of CELO virus and FAdV-9 have been reported [18].

In this study the left end region of FAdVs belonging to species *C*, *D* and *E* was sequenced and analyzed. We did not have any FAdV serotype reliably representing *FAdV-B* species group, and therefore it could not be included in this analysis. The rationale to carry out this study was as follows: (i) unlike FAdV-1 and -9, the nucleotide sequences of the left end of the other FAdVs have not been reported; (ii) the sequence information is important to provide insights about early gene function, evolutionary relationships and genetic features specific for each species group.

Material and methods

Viruses and DNA extraction

The following viruses were used: *FAdV-A* (FAdV-1, CELO virus), *FAdV-C* (FAdV-4, field isolate, and FAdV-10 strain C-2B), *FAdV-D* (FAdV-2, field isolate, and FAdV-9 strain A-2A) and *FAdV-E* (FAdV-8, field isolate). Viruses were propagated in chicken hepatoma cells (CH-SAH cell line) [19]. Cells were infected with 0.5 multiplicity of infection, and when complete cytopathic effect was seen (48–72 h post infection) the cells and supernatants were collected. Virus concentration and partial purification were done as we described previously [5]. The viral pellets were resuspended in TNE buffer (10 mM Tris pH 7.5, 100 mM NaCl, 1 mM EDTA) and DNA was extracted according to Ojkic and Nagy [5]. The DNA was resuspended in TE buffer (10 mM Tris pH 8.0, 1 mM EDTA) and quantitated by spectrophotometry.

Identification of the left ends of FAdVs

The viral DNAs were digested with *Bam*HI, *Eco*RI, *Hind*III and *Not*I and run in 0.8% agarose gels. Nucleic acids were bidirectionally transferred onto nylon membranes (Roche Applied Science) followed by UV crosslinking at 1200 J/cm² (FisherBiotech). The membranes were prehybridized for 2 h followed by overnight hybridization under low (55°C) or high (68°C) stringency conditions. Two probes corresponding to the left end of FAdV-9 were used: one, the entire 1.4 kb *Not*I fragment and the other a 180 bp fragment containing part of the dUTPase gene, and generated through *Sgf*I digestion of the 1.4 kb *Not*I fragment (Fig. 1a). Both probes were labelled with [³²P] dCTP or digoxigenin (DIG) by random priming. Membrane washes were carried out under the same stringency conditions mentioned above. For detection, the membranes were exposed to X-ray films (Kodak Omat XR) or photographed directly for the colorimetric DIG-labelled probes.

The identified fragments were extracted from the gel and cloned into pBluescript (SK-). The end fragments were cloned with blunt end ligation for the terminus and the appropriate restriction site for the other end.

Sequence analysis

The automated sequence of the cloned fragments was initially performed with T3 and T7 primers, and then completed by the primer walking approach. Multiple nucleotide sequence alignments using ClustalW

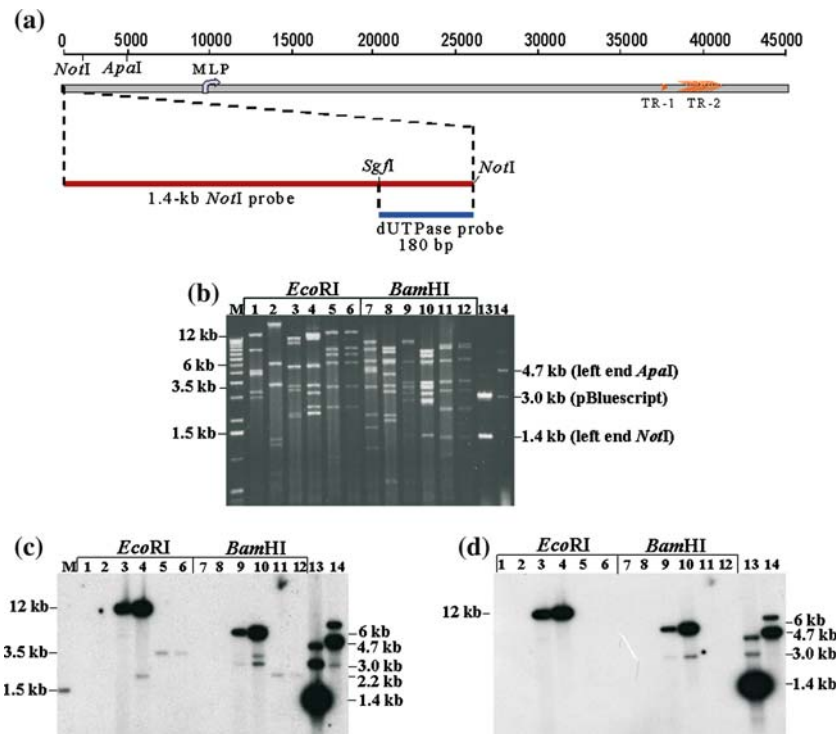


Fig. 1 Identification of the left end regions of FAdV-2, -4 and -8. (a) Location of the 1.4-kb- *NotI* and dUTPase probes in the FAdV-9 genome. (b) *EcoRI* and *BamHI* digestions. Lane M, 1-kb DNA ladder; lanes 1 and 7 FAdV-1; lanes 2 and 8 FAdV-4; lanes 3 and 9 FAdV-9; lanes 4 and 10 FAdV-2; lanes 5–6, and 11–12, FAdV-8 (two field isolates); lanes 13–14, positive controls (1.4-kb *NotI* and 4.7-kb *ApaI* left end fragments of FAdV-9 cloned into pBluescript). (c) Hybridization with the dUTPase probe. Strong hybridization signals are seen for FAdV-2 and -9 (5.8 and 3.8 kb fragments) at high (not shown) and low

stringency conditions. Weak signals for FAdV-1 and -4 (only in overexposed film, not shown), and FAdV-8 (3.5 and 3.8 kb fragments) were observed under low stringency conditions. Cross hybridization is observed with pBluescript (3 kb fragment) and the 1.5 kb band of the 1-kb DNA ladder. (d) Hybridization with the 1.4-kb-*NotI* fragment under low stringency conditions. Positive signals for only FAdV-2 and -9 are observed. The left end of FAdV-10 was subsequently identified by hybridization under high stringency conditions using the identified 3.8 kb *EcoRI* left end of FAdV-4 as a probe (not shown)

(version 1.82) from the European Bioinformatics Institute EMBL-EBI and vector NTI (version 5.5) were carried out to determine the conserved regions and distances among the left end genomes and ORFs of all analyzed FAdVs. The sequence data was blasted with the GeneBank database to determine the nucleotide sequence homologies and amino acid identities and similarities. ORFs encoding polypeptides over 50 amino acids were identified by the ORF finder from NCBI. PSI-BLAST with an inclusion threshold of 0.05 was also used in the analysis. Identities among all FAdV ORF homologues were determined and multiple amino acid sequence alignments were carried out to determine common features among the compared ORFs. These sequences were also compared with the Walker A motif of Rep protein of AAV 3B and papillomavirus E1 helicases, as reported by Washietl and Eisenhaber [18].

Results

Identification of the left end genomes of FAdVs

The DNA banding patterns among the FAdVs corresponded to those reported by Zsák and Kisary [1] (Fig. 1b).

Fragments for all FAdVs were detected with the labelled dUTPase probe in Southern hybridization (Fig. 1c) under low stringency conditions. However, a longer exposure was needed to detect hybridization to FAdV-1 and -4 bands (lanes 1, 2 and 7, 8). Using the 1.4-kb *NotI* probe, signals were observed only for FAdV-2 and -9 even under low stringency conditions (Fig. 1d). Similarly, the 5 kb left end *EcoRI* fragment of FAdV-10 was identified using the 3.5 kb left end *EcoRI* fragment of FAdV-4 as a probe under high stringency conditions (not shown).

Nucleotide sequence homology

Exactly 7455, 7578, 7551 and 7535 bp was sequenced for FAdV-2, -4, -8 and -10, respectively. The nucleotide sequence homologies of these plus the left ends of FAdV-1 and -9 are summarized in Table 1. Global alignments in the first 7455 bp (the smallest fragment sequenced corresponding to FAdV-2) between each species group showed homologies ranging from 47 to 79%, while the homologies were higher (98–99%) within each species group *C* and *D*, respectively. Similarly, the highest homologies from all local alignments generated in the analyzed FAdVs showed values from 54 to 86% among all FAdV groups and 98% between members of the same species group (Table 1).

Multiple nucleotide sequence analysis using ClustalW (version 1.82) were carried out to determine the genetic distances among species groups. Four clusters corresponding to each species group were generated (Fig. 2).

The inverted terminal repeats (ITRs) showed high conservation in all FAdVs, especially within the first 17 nucleotides (Fig. 3). All ITRs of FAdVs started from nucleotides CATCATC as published for those of FAdV-9 and human adenovirus 2 [5, 15]. As previously reported [2], the ITR for FAdV-1 differs from the others by C/G substitutions (GATGAT). The deduced consensus sequence, CATCATC-TATA-TATACC, for the ITRs of FAdVs is shown in Fig. 3.

Amino acid sequence identities

Amino acid identities of the left end ORFs are summarized in Table 2 and the gene arrangement at the left end of all FAdV genomes is depicted in Fig. 4. In this study, the nomenclature used to define each ORF

is that previously reported by us [15, 16], and recently redefined by Davison et al. [10] using the reported genome sequences of FAdV-1 and -9. Therefore, we followed both the latter nomenclature to define the ORFs at the left end of FAdV genomes and their equivalence with the nomenclature we previously gave for FAdV-9 (Fig. 4).

The highest identity values among all FAdVs were observed for the homologues to IVa2 ORF (67–100%) followed by ORF 1 (56–100%) and ORF 2 (44–99%) (Table 2). In addition, identities from 73% and up to 100% were observed in almost all ORFs between members of the same species group: 93–100% for *FAdV-C* (FAdV-4 and -10) and 73–100% for *FAdV D* (FAdV-2 and -9). No significant identities were observed between ORF 1B of FAdV-1 and its homologues in FAdV-2, -4, -9 and -10 and only low identities (34%) for FAdV-8. Similarly, no identities were observed between ORF-1C of FAdV-1 and its homologues in the other FAdVs (Table 2).

Identities with the dUTPase (ORF 1), Rep protein (ORF 2) and IVa2 genes from other viruses and some organisms were found in all FAdVs (Table 3). Of these, the highest identities were between the homologues to dUTPase in FAdV-2 and bovine papular stomatitis virus (63%) and with that of *Arabidopsis thaliana* (59%). Identities from 55 to 59% were observed between the dUTPase of FAdVs and the ORF 007 of the Orf virus (strain D1701). FAdV Rep protein homologue, on the other hand, showed lower identities, between 20 and 36%, when compared to those of parvoviruses. The IVa2 homologues in FAdVs showed identity values from 27 to 43%. The IVa2 homologue in FAdV-8 had the highest identities (40–43%) compared to those of *Duck adenovirus A* and some mastadenoviruses (Table 3).

Table 1 Nucleotide sequence homology (%) among FAdVs

Viruses (Species)	FAdV-1 (A)	FAdV-4 (C)	FAdV-10 (C)	FAdV-2 (D)	FAdV-9 (D)
Global analysis					
FAdV-1 (A)	–				
FAdV-4 (C)	50	–			
FAdV-10 (C)	51	98	–		
FAdV-2 (D)	49	51	51	–	
FAdV-9 (D)	50	50	51	99	–
FAdV-8 (E)	47	50	51	79	79
Local analysis					
FAdV-1 (A)	–				
FAdV-4 (C)	76	–			
FAdV-10 (C)	77	98	–		
FAdV-2 (D)	74	78	54	–	
FAdV-9 (D)	75	78	54	98	–
FAdV-8 (E)	74	86	55	67	67

The highest homology values are in bold

Gene arrangement

Similar gene arrangements were found in all left ends of FAdV genomes (Fig. 4). Homologues to ORFs 0, 1, 1A, 1B, 1C, 2, 24, 14, 13, 12 and IVa2, described previously for FAdV-1 (CELO virus) and FAdV-9 [10], were present in the genomes of all FAdVs except that homologues to ORF 1C could not be detected in FAdV-4 and -10. For FAdV-8, the ORF at nts 1328–1705 had 54 and 77% identities to ORFs 1A and 1B of FAdV-2 and -9, respectively, and therefore this ORF was assigned as ORF 1A/B.

The relative positions of each ORF within the left end regions were generally conserved. The presence of two extra ORFs between ORFs 24 and 13 in FAdV-4 and -10 resulted in a 0.5–1-kb-downstream

Fig. 2 Genetic distances among FAdV species based on nucleotide sequences at the left end genomes. The distance values indicated in the cladogram were calculated using ClustalW

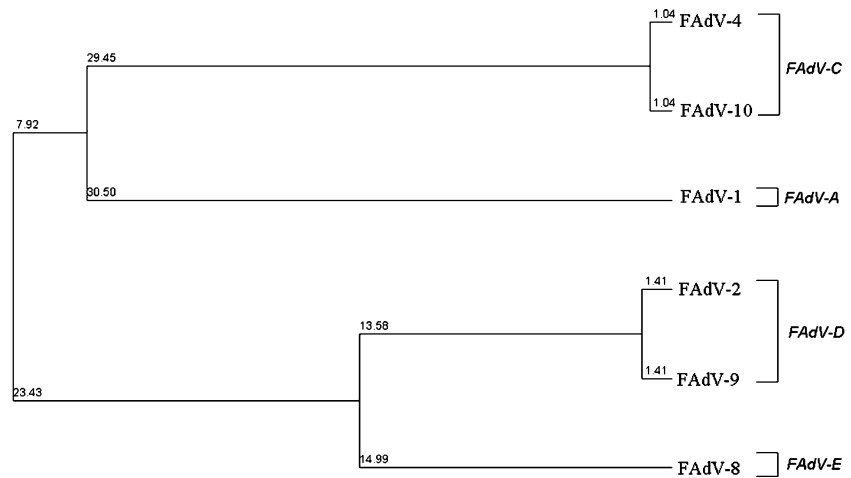


Fig. 3 Inverted terminal repeats (ITRs) of FAdVs. The most conserved, first 17 nucleotides are bolded and underlined. Comparisons between ITRs of FAdV-1 and -9 to those of human adenovirus 2 (HAdV-2) were described by Cao et al. [15]

Species	Serotype	Sequence
<i>Fowl Adenovirus</i>	A FAdV-1	5' <u>GATGAT-GTATAATA-ACC</u> TCAAAAACATA
<i>Fowl Adenovirus</i>	C FAdV-4	5' <u>CATCATCTTATA-TA-ACC</u> GCGTCTTTTGA
	FAdV-10	5' <u>CATCATCTTATA-TA-ACC</u> CGCGTCTTTTG
<i>Fowl Adenovirus</i>	D FAdV-2	5' <u>CATCATC-TATA-TATACC</u> TACATGAATGA
	FAdV-9	5' <u>CATCATC-TATA-TATACC</u> TACATGAATGA
<i>Fowl Adenovirus</i>	E FAdV-8	5' <u>CATCATC-TATA-TATATC</u> TACTTAAAATA
<i>Human Adenovirus</i>	C HAdV-2	5' <u>CATCATC-AATAATATACC</u> TTATTTTCGAT
Consensus sequence		5' <u>CATCATC-TATA-TATACC</u>

displacement of ORFs 14, 13, 12 and IVa2, but their relative orders were retained (Fig. 4). These extra ORFs had identities with ORF 24 and 14 of FAdV-9 and ORF 14 of FAdV-1 (Table 4). To better illustrate the relationship between the assigned ORFs 24 and 14 and these extra ORFs in FAdV-4 and -10, distances represented in a cladogram were determined using ORFs 14 and 24 of FAdV-9 and ORF 14 of FAdV-1 as references. Two major clusters consisting of homologues to ORF 14 and ORF 24 were observed (not shown). The cluster involving the homologues to ORF 14 in FAdV-1, -4 and -9 included these extra ORFs. These extra ORFs were considered ORF 14 duplicates and therefore they were assigned as ORF 14a (nts 3319–3906 in FAdV-4 and nts 3320–3907 in FAdV-10) and ORF 14b (nts 4670–5299 in FAdV-4 and nts 4671–5297 in FAdV-10).

A duplicate of ORF 14, overlapping the assigned ORF 14 in FAdV-8 (nts 3880–4284), was identified (Fig. 4). This ORF had 38 and 44% identity with ORF 14 of FAdV-2 and -9, respectively. No identity with other homologues to ORF 14 in other FAdVs, including its own, was observed (Table 4).

Additional unassigned ORFs were present in FAdV-4 and -8: ORF 2392–2703 of FAdV-4, located

between ORF 2 and ORF 24, and ORF 4413–5219 of FAdV-8, which partially overlapped ORF 13 (Fig. 4).

Relationships among ORFs 2, 24, 14, 13 and 12

No identities and amino acid conservation were clearly observed among ORFs 2, 14 (including its duplicates) and 24 of all FAdVs (data not shown).

Identities between ORFs 14 (including ORF 14a and 14b) and 24 ranged from no significant identity (N.S.I.) to 33%. The highest values (32 and 33%) were found between ORF 14 of FAdV-1 and ORF 24 of FAdV-4 and -10 (Table 4). Identities between ORF 14a and ORF 24 of FAdV-4 and -10 (29–30%) were observed only in these viruses, whereas values between ORF 14b of FAdV-4 and -10 and ORF 24 of all FAdVs, except that of FAdV-8, ranged from 26 to 30%. Identities between homologues to ORF 14 in all FAdVs, except that of FAdV-8, and ORF 14a of FAdV-4 and -10 were 23–31%. Similarly, ORF 14b of FAdV-4 and -10 and ORF 14 of all FAdVs had identities from 23 to 27%. Identities from 96 to 98% were found between ORFs 14a and b of FAdV-4 and their counterparts in FAdV-10 (not shown). Amino acid conservation with the homologues to ORFs 24 and 14,

Table 2 Identities (%) found among ORFs of FAdVs

ORF	Virus (Species)	FAdV-1 (A)	FAdV-4 (C)	FAdV-10 (C)	FAdV-2 (D)	FAdV-9 (D)
ORF 0	FAdV-1 (A)	–	–	–	–	–
	FAdV-4 (C)	60	–	–	–	–
	FAdV-10 (C)	60	96	–	–	–
	FAdV-2 (D)	61	51	51	–	–
	FAdV-9 (D)	61	51	51	98	–
DUTPase ORF 1	FAdV-8 (E)	58	53	53	65	64
	FAdV-1 (A)	–	–	–	–	–
	FAdV-4 (C)	56	–	–	–	–
	FAdV-10 (C)	56	91	–	–	–
	FAdV-2 (D)	69	60	61	–	–
ORF 1A	FAdV-9 (D)	69	60	61	100	–
	FAdV-8 (E)	68	59	60	84	85
	FAdV-1 (A)	–	–	–	–	–
	FAdV-4 (C)	46	–	–	–	–
	FAdV-10 (C)	46	100	–	–	–
ORF 1A/B	FAdV-2 (D)	41	59	59	–	–
	FAdV-9 (D)	41	59	59	73	–
	FAdV-8 (E)	N.S.I.	N.S.I.	N.S.I.	54	54
	FAdV-1 (A)	–	–	–	–	–
	FAdV-4 (C)	N.S.I.	–	–	–	–
ORF 1B	FAdV-10 (C)	N.S.I.	100	–	–	–
	FAdV-2 (D)	N.S.I.	33	33	–	–
	FAdV-9 (D)	N.S.I.	33	33	82	–
	FAdV-8 (E)	34	30	30	77	77
	FAdV-1 (A)	–	–	–	–	–
ORF 1A/B	FAdV-4 (C)	–	–	–	–	–
	FAdV-10 (C)	–	–	–	–	–
	FAdV-2 (D)	N.S.I.	–	–	–	–
	FAdV-9 (D)	N.S.I.	–	–	84	–
	FAdV-8 (E)	N.S.I.	–	–	45	54
ORF 2	FAdV-1 (A)	–	–	–	–	–
	FAdV-4 (C)	51	–	–	–	–
	FAdV-10 (C)	47	97	–	–	–
	FAdV-2 (D)	49	50	47	–	–
	FAdV-9 (D)	48	50	47	99	–
ORF 24	FAdV-8 (E)	47	47	44	78	78
	FAdV-1 (A)	–	–	–	–	–
	FAdV-4 (C)	–	–	–	–	–
	FAdV-10 (C)	–	96	–	–	–
	FAdV-2 (D)	–	38	31	–	–
ORF 14	FAdV-9 (D)	–	38	31	98	–
	FAdV-8 (E)	–	34	30	59	60
	FAdV-1 (A)	–	–	–	–	–
	FAdV-4 (C)	25	–	–	–	–
	FAdV-10 (C)	23	96	–	–	–
ORF 13	FAdV-2 (D)	28	29	28	–	–
	FAdV-9 (D)	28	29	28	99	–
	FAdV-8 (E)	22	28	29	43	44
	FAdV-1 (A)	–	–	–	–	–
	FAdV-4 (C)	54	–	–	–	–
ORF 12	FAdV-10 (C)	54	98	–	–	–
	FAdV-2 (D)	51	44	44	–	–
	FAdV-9 (D)	51	44	45	91	–
	FAdV-8 (E)	42	38	39	65	65
	FAdV-1 (A)	–	–	–	–	–
ORF 13	FAdV-4 (C)	39	–	–	–	–
	FAdV-10 (C)	39	98	–	–	–
	FAdV-2 (D)	41	41	43	–	–
	FAdV-9 (D)	39	40	42	98	–
	FAdV-8 (E)	41	42	42	78	79

Table 2 continued

ORF	Virus (Species)	FAdV-1 (A)	FAdV-4 (C)	FAdV-10 (C)	FAdV-2 (D)	FAdV-9 (D)
IVa2	FAdV-1 (A)	–	–	–	–	–
	FAdV-4 (C)	70	–	–	–	–
	FAdV-10 (C)	70	100	–	–	–
	FAdV-2 (D)	74	72	68	–	–
	FAdV-9 (D)	73	71	67	99	–
	FAdV-8 (E)	78	76	73	94	95

N.S.I. = No significant identity (identities lower than 25% and/or *E* values over 1)

— = ORF not found

Highest identities are in bold

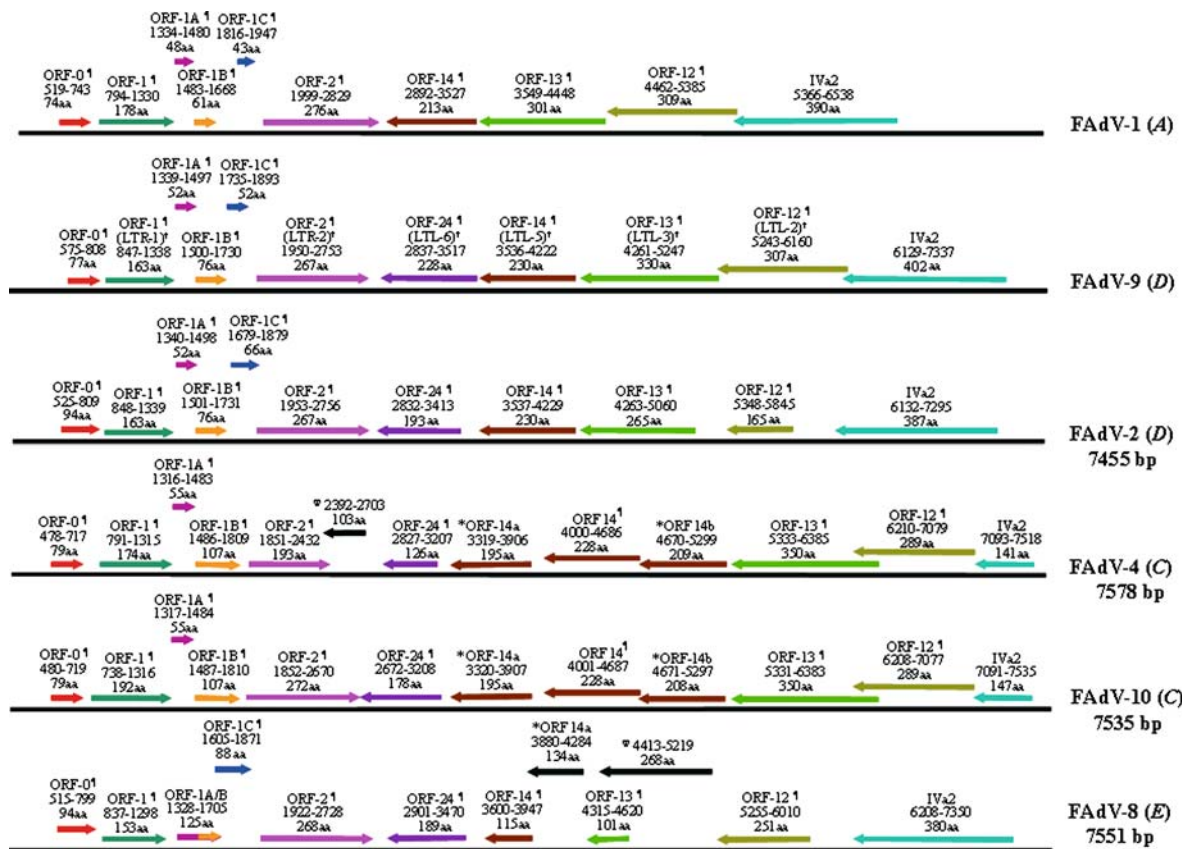


Fig. 4 Putative gene arrangement in FAdV genomes. aa = amino acid, [†]Davison et al. [10]. ORFs 1 (dUTPase), 2 (Rep protein gene homologue), 14, 24, 12 and 13 in FAdV-9 have been

demonstrated to be transcribed in CH-SAH cells. [†]Cao et al. [15]. *ORF 14 duplicates in FAdV-4, -8 and -10. [‡]Additional unidentified ORFs in FAdV-4 and -8

including its duplicates, in the other FAdVs was observed, especially in a stretch of 14 amino acids consisting in the consensus CXCXXPXSFLCQSL (Fig. 5)

Multiple alignments between ORF 2 and ORFs 12 and 13 showed conservation in some amino acid positions (Fig. 6). Alignments between ORF 2 of FAdV-1 and -9 with ORFs 12 and 13 of all FAdVs showed identities corresponding to only ORF 13 of FAdV-2, -4, and -10 (Table 5). To identify the Walker A motif

(GXXXGKT/S), present in the superfamily III ATPase/helicases, data from Washietl and Eisenhaber [18] were utilized. Multiple alignments in all homologues to ORFs 2, 12, 13 and Rep protein of AAV-3 showed some amino acid conservation and alignments with the Walker A motif showed high conservation in the first glycine of this motif. The GKT of this motif were replaced by GRP in ORF 2, CAD in ORF 12, and NAK in ORF 13. Preceding the Walker A motif in Rep

Table 3 Identities (%) of dUTPase, Rep protein and IVa2 protein in other viruses and some organisms

ORF	Source	FAdV-4 (C)	FAdV-10 (C)	FAdV-2 (D)	FAdV-8 (E)
DUTPase	Bovine papular stomatitis virus (ORF 007)	57	56	63	58
	Orf virus strain D1701 (ORF 007)	59	59	58	55
	Bovine adenovirus A (E4 ORFA)	50	51	55	49
	Porcine adenovirus 5 (ORF 2)	57	58	57	52
	Fowlpox virus (ORF FPV038)	53	54	57	59
	Canarypox virus (ATCC VR-111)	54	55	54	54
	Vaccinia virus (Ankara)	54	52	55	52
	<i>Aspergillus fumigatus</i> Af293	57	56	58	51
	<i>Homo sapiens</i>	58	58	55	54
	<i>Mus musculus</i>	54	54	58	56
	<i>Arabidopsis thaliana</i>	55	57	59	55
Rep protein	Adeno-associated virus 8	34	31	29	31
	Avian adeno-associated virus strain DA-1	33	30	28	29
	Bovine adeno-associated virus	31	30	29	31
	Muscovy duck parvovirus	33	31	28	31
	Snake parvovirus 1	24	23	24	26
	Human parvovirus 4 (ORF 1)	27	25	27	25
	Feline panleukopenia virus	36	27	20	24
	Duck adenovirus 1	30	30	38	40
IVa2 protein	Ovine adenovirus 7	29	29	37	42
	Human adenovirus 40	29	29	33	43
	Simian adenovirus 3	27	27	33	41
	Human adenovirus 17	30	30	32	42
	Murine adenovirus 1	29	29	34	40
	Human adenovirus 2	28	28	32	41
	Human adenovirus 11	29	27	32	42

Table 4 Identities (%) among homologues of ORFs 14 and 24 in FAdVs

ORF	FAdVs	FAdV-1 (A)	FAdV-4 (C)	FAdV-4 (C)	FAdV-10 (C)	FAdV-10 (C)	FAdV-2 (D)	FAdV-2 (D)	FAdV-9 (D)	FAdV-9 (D)	FAdV-8 (E)	FAdV-8 (E)
		ORF 14	ORF 14	ORF 24	ORF 14	ORF 24	ORF 14	ORF 24	ORF 14	ORF 24	ORF 14	ORF 24
ORF 14	FAdV-1 (A)	–	25	33	23	32	28	26	28	26	22	N.S.I.
	FAdV-4 (C)	25	–	31	96	31	29	N.S.I.	29	N.S.I.	28	N.S.I.
	FAdV-10 (C)	23	96	31	–	31	28	N.S.I.	28	N.S.I.	29	N.S.I.
	FAdV-2 (D)	28	29	30	28	N.S.I.	–	27	99	27	43	N.S.I.
	FAdV-9 (D)	28	29	30	28	N.S.I.	99	27	–	27	44	N.S.I.
	FAdV-8 (E)	22	28	N.S.I.	29	N.S.I.	43	N.S.I.	44	N.S.I.	–	N.S.I.
	FAdV-8 (E)	22	28	N.S.I.	29	N.S.I.	43	N.S.I.	44	N.S.I.	–	N.S.I.
ORF 14a	FAdV-1 (A)	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
	FAdV-4 (C)	24	29	29	30	29	25	N.S.I.	25	N.S.I.	N.S.I.	N.S.I.
	FAdV-10 (C)	23	25	29	31	30	24	N.S.I.	24	N.S.I.	N.S.I.	N.S.I.
	FAdV-2 (D)	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
	FAdV-9 (D)	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
	FAdV-8 (E)	N.S.I.	N.S.I.	N.S.I.	N.S.I.	N.S.I.	38	N.S.I.	44	N.S.I.	N.S.I.	N.S.I.
	FAdV-8 (E)	N.S.I.	N.S.I.	N.S.I.	N.S.I.	N.S.I.	38	N.S.I.	44	N.S.I.	N.S.I.	N.S.I.
ORF 14b	FAdV-1 (A)	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
	FAdV-4 (C)	23	26	30	27	29	24	27	23	26	27	N.S.I.
	FAdV-10 (C)	25	25	28	26	27	24	27	23	26	27	N.S.I.
	FAdV-2 (D)	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
	FAdV-9 (D)	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
	FAdV-8 (E)	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
	FAdV-8 (E)	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
ORF 24	FAdV-1 (A)	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
	FAdV-4 (C)	33	31	–	31	96	30	38	30	38	N.S.I.	34
	FAdV-10 (C)	32	31	96	31	–	N.S.I.	31	N.S.I.	31	N.S.I.	30
	FAdV-2 (D)	26	N.S.I.	38	N.S.I.	31	27	–	27	98	N.S.I.	59
	FAdV-9 (D)	26	N.S.I.	38	N.S.I.	31	27	98	27	–	N.S.I.	60
	FAdV-8 (E)	N.S.I.	N.S.I.	34	N.S.I.	30	N.S.I.	59	N.S.I.	60	N.S.I.	–

N.S.I. = No significant identity (identities lower than 20% and/or E values over 1)

N.A. = Not applicable due to the absence of these ORFs

Highest identities are in bold

Table 5 Identities (%) between homologues to ORFs 12 and 13 with ORF 2 of FAdV-1 and -9

ORF	Rep protein AAV-3	FAdV-1 ORF2	FAdV-9 ORF2
FAdV-2 ORF 12	17 (0.69)	N.S.I.	N.S.I.
FAdV-4 ORF 12	N.S.I.	N.S.I.	N.S.I.
FAdV-10 ORF 12	N.S.I.	N.S.I.	N.S.I.
FAdV-8 ORF 12	N.S.I.	N.S.I.	N.S.I.
FAdV-2 ORF 13	24 (3×10^{-3})	27 (7×10^{-3})	26 (1.3×10^{-2})
FAdV-4 ORF 13	23 (3.4)	24 (4×10^{-4})	22 (4×10^{-3})
FAdV-10 ORF 13	23 (4.4)	24 (4×10^{-4})	22 (4×10^{-3})
FAdV-8 ORF 13	N.S.I.	N.S.I.	N.S.I.

N.S.I. = No significant identity (values less than 20% and/or *E* values over 1)

E values are in parenthesis

protein of AAV-3, the amino acids NT were conserved in ORFs 2, 12 and 13, with only T/V substitution in ORF 12 of FAdV-2 and -9 (Fig. 6).

Discussion

Nucleotide sequence homology

Sequence analysis at the left end region of FAdVs representing each species group can be very helpful in determining common and unique features, carry out further gene function studies, and conduct genetic manipulations for vector design.

So far, the complete nucleotide sequence of the genomes of only FAdV-1 (CELO virus) and FAdV-9 (strain A-2A) and their transcriptional maps have been reported [2, 3, 14–16]. Partial information on the right end of FAdV-8 and -10 are available, and their implications for early gene function and vector design have been reported [7, 20, 21]. In contrast, no sequence analysis at the left end of those FAdVs has been documented.

Nucleotide sequence homology was 98–99% among members of the same species groups *C* and *D*. These results are consistent with the fact that cross-reaction in virus neutralization assays and cross DNA-hybridization under stringent conditions have been observed for FAdV-4 and -10, suggesting that they could be considered as subtypes of the same serotype [22]. In addition, the restriction patterns with *Bam*HI and *Hind*III of these viruses are similar [1]. In this work, under high stringency conditions the 1.4 kb left end *Not*I and dUTPase probes from FAdV-9 gave strong hybridization signals corresponding to FAdV-2 and -9 only (Fig. 1). In contrast, hybridization under low stringency conditions with dUTPase allowed the detection of the left ends of all FAdVs. Hybridization

signals to FAdV-1 and -4 were barely visible in over-exposed X-ray films when [³²P]-labelled dUTPase probe was used (not shown). Further, no hybridization signals corresponding to these FAdVs were visible when the 1.4 kb *Not*I left end was the probe, probably due to the low nucleotide sequence homology among left end FAdVs upstream of the dUTPase gene. As described, nucleotide sequence homology among all analyzed FAdVs was 47–79%. The 180-bp dUTPase probe corresponding to part of the dUTPase gene has shown to be conserved in all FAdVs (56–100%), and therefore its detection in all FAdVs was possible. Moreover, the left *Eco*R1 fragment of FAdV-10 hybridized to the left *Eco*R1 fragment of the FAdV-4 probe, which further demonstrated their close relationships in terms of nucleotide sequence homologies and cross neutralization in virus assays reported for FAdV-4 and -10 [22].

Multiple nucleotide sequence alignments of the left end genomes of FAdVs generated four clusters corresponding to each species group (Fig. 2). Genetic distances between species *D* and *E* indicated a close relationship between them, and this result is supported by the high identity values among their ORFs (Table 2). For example, homologues to IVa2 in these two groups were 94 and 95%.

Amino acid sequence identities

Homologues to the ORFs in the left end genomes of FAdV-2, -8, and -9 showed generally higher identities to each other than to the other studied viruses. The IVa2 ORF showed the highest identity values among all analyzed FAdVs. However, further nucleotide sequencing may be required to confirm the actual sizes and to obtain more accurate identity values for the homologues to IVa2 in FAdV-4 and -10. This protein seems to be important in viral genome packaging and in late gene transcription from the major late promoter [23–25]. Therefore, IVa2 protein gene, together with E2 and late genes are very important for virus replication and would need to have been maintained in all adenoviruses during evolution [10].

ORFs 1 and 2 have identities with ORFs of known gene functions corresponding to the dUTPase and parvovirus non-structural protein 1 (NS-1), also called Rep protein in AAV. These ORFs may be important for DNA replication of FAdVs whereas the rest of the left end ORFs still have unassigned functions. The roles of these ORFs in virus replication are unknown. In fact, non-sense mutation studies in each ORF did not significantly affect FAdV-1 replication [8], although contradictory results have been reported [4].

Gene arrangement

Similar gene arrangements among the FAdVs suggested similar gene functions. ORF 1C was not observed in FAdV-4 and -10. Instead, homologues to ORF 1B in FAdV-4 and -10 appeared to be larger (107 amino acids) than their homologue counterparts in other FAdVs (61–76 amino acids). Based on the sequence analysis it seems likely that missense mutations abolishing the stop codon of ORF 1B caused this coding region to be larger and additional mutations eliminated ORF 1C in these viruses in all reading frames. Consequently, ORF 1C might not be essential for viral replication. The absence of another ORF, ORF 24 in FAdV-1, has also been reported [10].

Unlike in other FAdVs, ORF 1A and 1B of FAdV-8 appeared to be one continuous coding region represented in one ORF (nts 1328–1705). Therefore, this ORF was assigned as ORF 1A/B. This ORF seem to have originated from frameshift mutations in the 3' end of ORF 1A leading to the junction with ORF 1B. This assumption is based on the fact that homologues to ORF 1A and 1B in all analyzed FAdVs are frameshifted and appear not be continuous nor overlapped. We think it is also possible that ORF 1A and 1B in other FAdVs might have originated from frameshift mutations in an ancestral ORF 1A/B.

We demonstrated the transcription of FAdV-9 ORFs 1, 1A, 1C, 24, 14, 13, 12 and IVa2 in CH-SAH cells [15, 16]. In contrast, the transcription of only ORF 1 (dUTPase) in FAdV-1 has been reported (in nineteen-day-old chicken embryo kidney cells), and its enzymatic activity has been demonstrated [2]. It is not known whether these ORFs in other FAdVs are transcribed in cell culture or in vivo.

The insertion of two additional ORFs between ORF 24 and 13 homologues in FAdV-4 and -10 resulted in approximately 0.5 and 1 kb downstream displacement of ORFs 14, 13 and IVa2, respectively (Fig. 4). When these ORFs were blasted, low identities with ORF 24 and 14 of FAdV-9, and ORF 14 in FAdV-1 virus were observed (Table 4). ORFs at nts 4000–4686 in FAdV-4 and 4001–4687 in FAdV-10 were assigned as ORF 14 because of their relative size (228 aa) when compared to their counterparts in the other FAdVs, except FAdV-8. ORFs at nts 3319–3906 and 4670–5299 in FAdV-4, or 3320–3907 and 4671–5297 in FAdV-10 were considered as duplicates of ORF 14 and therefore we assigned them as ORF 14a and 14b. Despite their low identity values to ORF 14, these ORFs were considered as duplicates of ORF 14 for the following reasons: (i) When blasted into the GenBank database, ORF 14a and 14b showed identities with only ORF 14

of FAdV-1 and -9, whereas no identities to ORF 24 of FAdV-9 were found; (ii) Although ORF 14b had slightly higher identities to ORF 24 in all FAdVs, except that of FAdV-2 and -9, ORF 14a and b were more closely related to ORF 14 of FAdV-1, -4, -9 and -10 than the homologues to ORF 24 when the distances represented in a cladogram were calculated (not shown); (iii) multiple sequence alignments between duplicates of ORF 14 and ORFs 14 and 24 using ClustalW clearly grouped these ORFs with homologues to ORF 14 of all analyzed FAdVs (Fig. 5).

Gene duplication of ORF 14 in FAdV-8 was apparent in ORF 3880–4284 (ORF 14a). Unlike ORF14a and 14b in FAdV-4 and -10, this duplicate did not increase the DNA content nor displaced the downstream ORFs. We think that nonsense mutations may have resulted in a stop codon at position 3880 resulting in the generation of ORF 14a in FAdV-8. This suggestion is based on the fact that ORF 14 of the other FAdVs are all around 230 aa in size, and the sum of ORF 14 (115 aa) and ORF 3880–4284 (135 aa) is just a bit larger at 249 aa. Although the corresponding region from the same clone was sequenced several times sequencing errors might nevertheless have also occurred.

Relationships among ORFs 2, 24, 14, 13 and 12

Detailed analyses for ORF 14 of FAdV-1 and ORFs 14 and 24 of FAdV-9 have revealed that these ORFs are related to each other and have also been considered to belong to superfamily III helicases and to be related to the NS-1 proteins including ORF 2 [18]. In this work, amino acid conservation was evident in these ORFs (Fig. 5), including the duplicates of ORF 14, despite their low identities among them (Table 4).

In this work, we did not detect any identity and amino acid conservation between ORFs 2, Rep protein homologue and papillomavirus E1 helicase and ORFs 14 and 24 using BLAST or PSI-BLAST after several runs (data not shown). In addition, there is no clear amino acid conservation between Papillomavirus E1 helicase and ORFs 14 and 24. However, the preceding amino acid stretch of the Walker A motif of papillomavirus E1 helicase, VSF, partially aligned with the amino acid stretch VLF of the deduced consensus sequence for ORFs 24 and 14 (Fig. 5).

Duplicates of ORF 2 in FAdV-1 and -9 are identified as ORFs 13 and 12 in the opposite strand [10, 18]. Identification of motifs present in the NS-1 proteins of parvoviruses in ORFs 12 and 13 of FAdV-1 and -9 has been reported [18]. This also seems to be the case for these homologues in all analyzed FAdVs, although

more variations in the Walker A motif (GXXXGKT/S) were observed (Fig. 6). Identities between ORF 2 of FAdV-1, -2 and -9 and ORF 13 of all FAdVs except in FAdV-8, were observed suggesting an evolutionary relationship among them. In contrast, no significant identity was found between ORF 2 of FAdV-4, -8 and -10 and ORF 12 in all FAdVs (Table 5), but some conservation was also observed in the amino acids flanking the Walker A motif (Fig. 6). Identities were also seen between the Rep protein of AAV-3 and ORFs 12 (only in that of FAdV-2) and 13 (except that of FAdV-8). However, the high *E* values (above 1) for ORF 13 of FAdV-4 and -10 do not allow us to reliably establish their relationships with the Rep protein of AAV-3.

ORF 2 duplicates seem to be predominant in the left end of FAdVs and their role in virus replication remains intriguing. As discussed above, the transcription of these ORFs have been demonstrated in cell culture [15, 16]. Therefore, these observations raise the following questions: Do all ORF 2 duplicates play important roles during virus replication in cell culture or in vivo? Does the number of ORF 2 duplicates provide advantages or disadvantages for replication in the host? Are duplicates of ORF 14 dispensable for viral replication in FAdV-4 and -10? Are the duplicates functional?

In conclusion, we determined the left end sequences of FAdVs representing members of species group *C*, *D*, and *E* and compared them with the reported sequences of FAdV-1 (*A*) and FAdV-9 (*D*). High nucleotide sequence homologies and high amino acid sequence identities of all ORFs were observed in the members belonging to the same species group. All the studied FAdVs have similar gene arrangement suggesting similar gene functions. Gene duplication in the homologue to ORF 14 in FAdV-4 and -10 resulted in an approximately 1 kb displacement of ORFs 14, 13 and IVa2. The implication of these duplications remains intriguing in terms of evolution and virus replication and further research is needed to define the roles of these ORFs.

Acknowledgements This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Ontario Ministry of Agriculture, Food and Rural Affairs (OMAFRA).

References

1. L. Zsák, J. Kisary, *Intervirology* **22**, 110–114 (1984)
2. S. Chiocca, R. Kurzbauer, G. Schaffne, A. Baker, V. Mautner, M. Cotten, *J. Virol.* **70**, 2939–2949 (1996)
3. D. Ojkic, É. Nagy, *J. Gen. Virol.* **81**, 1833–1837 (2000)
4. A.-I. Michou, H. Lehrmann, M. Saltik, M. Cotten, *J. Virol.* **73**, 1399–1410 (1999)
5. D. Ojkic, É. Nagy, *Virology* **283**, 197–206 (2001)
6. M.A. Johnson, C. Pooley, J.W. Lowenthal, *Dev. Comp. Immunol.* **24**, 343–354 (2000)
7. M. Sheppard, W. Werner, E. Tsatas, R. McCoy, S. Rowse, M. Johnson, *Arch. Virol.* **143**, 915–930 (1998)
8. A. François, N. Etteradossi, B. Delmas, V. Payet, P. Langlois, *J. Virol.* **75**, 5288–5301 (2001)
9. R. McCoy, M. Sheppard, *Virus Genes* **14**, 245–249 (1997)
10. A.J. Davison, M. Benkő, B. Harrach, *J. Gen. Virol.* **84**, 2895–2908 (2003)
11. S. Chiocca, A. Baker, M. Cotten, *J. Virol.* **71**, 3168–3177 (1997)
12. H. Lehrmann, M. Cotten, *J. Virol.* **73**, 6517–6525 (1999)
13. S. Chiocca, V. Kurtev, R. Colombo, R. Boggio, M.T. Scurpi, G. Brosch, C. Seiser, G.F. Draetta, M. Cotten, *Curr. Biol.* **12**, 594–598 (2002)
14. V. Payet, C. Arnauld, J.-P. Picault, A. Jestin, P. Langlois, *J. Virol.* **72**, 9278–9285 (1998)
15. J.X. Cao, P.J. Krell, É. Nagy, *J. Gen. Virol.* **79**, 2507–2516 (1998)
16. D. Ojkic, P.J. Krell, É. Nagy, *Virology* **302**, 274–285 (2002)
17. D. Ojkic, É. Nagy, *Vaccine* **22**, 42–48 (2003)
18. S. Washietl, F. Eisenhaber, *BMC Bioinformatics* **4**, 1471–2105 (2003)
19. H.S. Alexander, P. Huber, J. Cao, P.J. Krell, É. Nagy, *J. Virol. Methods*, **74**, 9–14 (1998)
20. J. Pallister, J.P. Wright, M. Sheppard, *J. Virol.* **70**, 5115–5122 (1996)
21. M.A. Johnson, C. Pooley, J. Ignjatovic, S.G. Tyack, *Vaccine* **21**, 2730–2736 (2003)
22. K. Erny, J. Pallister, M. Sheppard, *Arch. Virol.* **140**, 491–501 (1995)
23. W. Zhan, M.J. Imperiale, *J. Virol.* **74**, 2687–2693 (2000)
24. W. Zhang, J.A. Low, J.B. Christensen, M.J. Imperiale, *J. Virol.* **75**, 10,446–10,454 (2001)
25. C. Tribouley, P. Lutz, A. Staub, C. Keding, *J. Virol.* **68**, 4450–4457 (1994)