Coding potential and transcript analysis of fowl adenovirus 4: insight into upstream ORFs as common sequence features in adenoviral transcripts

Bryan D. Griffin and Éva Nagy

Department of Pathobiology, Ontario Veterinary College, University of Guelph, 50 Stone Road East, Guelph, ON N1G 2W1, Canada

Recombinant fowl adenoviruses (FAdVs) have been successfully used as veterinary vaccine vectors. However, insufficient definitions of the protein-coding and non-coding regions and an incomplete understanding of virus–host interactions limit the progress of next-generation vectors. FAdVs are known to cause several diseases of poultry. Certain isolates of species FAdV-C are the aetiological agent of inclusion body hepatitis/hydropericardium syndrome (IBH/HPS). In this study, we report the complete 45667 bp genome sequence of FAdV-4 of species FAdV-C. Assessment of the protein-coding potential of FAdV-4 was carried out with the Bio-Dictionary-based Gene Finder together with an evaluation of sequence conservation among species FAdV-A and FAdV-D. On this basis, 46 potentially protein-coding ORFs were identified. Of these, 33 and 13 ORFs were assigned high and low protein-coding potential, respectively. Homologues of the ancestral adenoviral genes were, with few exceptions, assigned high protein-coding potential. ORFs that were unique to the FAdVs were differentiated into high and low protein-coding potential groups. Notable putative genes with high protein-coding capacity included the previously unreported fiber 1, hypothetical 10.3K and hypothetical 10.5K genes. Transcript analysis revealed that several of the small ORFs less than 300 nt in length that were assigned low coding potential contributed to upstream ORFs (uORFs) in important mRNAs, including the ORF22 mRNA. Subsequent analysis of the previously reported transcripts of FAdV-1, FAdV-9, human adenovirus 2 and bovine adenovirus 3 identified widespread uORFs in AdV mRNAs that have the potential to act as important translational regulatory elements.

INTRODUCTION

Adenoviruses (AdVs) of the genus Mastadenovirus have been and are being developed as anti-cancer agents (Huebner et al., 1956; Cody & Douglas, 2009; Yamamoto & Curiel, 2010) and vaccine vectors (Lasaro & Ertl, 2009). The problem of pre-existing immunity against human adenovirus type 5 (HAdV-5), exemplified in the STEP human immunodeficiency virus (HIV) trial that employed recombinant HAdV-5 (Buchbinder et al., 2008; McElrath et al., 2008), has generated new interest in the development of less characterized AdV serotypes and non-human AdVs as both oncolytic (Cody & Douglas, 2009; Gallo et al., 2005; Shashkova et al., 2005) and vaccine vectors (Barouch, 2008; Lasaro & Ertl, 2009; Sharma et al., 2010). The development of AdV-based veterinary vaccine vectors is ongoing (Bangari & Mittal, 2006; Ferreira et al., 2005). Fowl adenoviruses (FAdVs) of the genus Aviadenovirus that includes species FAdV-A to FAdV-E (Adair & Fitzgerald, 2008; Benkő et al., 2005) are being developed as vaccine vectors. Some FAdVs cause disease, including a subset of FAdV-4 of species FAdV-C that are the aetiological agents of the economically important inclusion body hepatitis/hydropericardium syndrome (IBH/HPS) (Adair & Fitzgerald, 2008). The first generation of FAdV-based vaccine vectors have proven effective at eliciting an antibody response against a delivered transgene (Corredor & Nagy, 2010; Ojkic & Nagy, 2003), and in chickens have conferred protective immunity against infectious bursal disease virus (IBDV) (Francois et al., 2004; Sheppard et al., 1998) and infectious bronchitis virus (Johnson et al., 2003). Analysis of the complete genomes of FAdV-1, the chicken embryo lethal orphan (CELO) virus (Chiocca et al., 1996) and FAdV-9 (Ojkic & Nagy, 2000) (species FAdV-A and FAdV-D, respectively), and the terminal genomic regions of FAdV-2, -4, -10 and -8 (Corredor et al., 2006, 2008) has shown that
the FAdVs share a common genome organization. The central genomic region consists primarily of ancestral adenoviral genes, dubbed the genus-common genes, that are highly conserved in all adenovirus genera. The terminal genomic regions consist of novel genes, dubbed the genus-specific genes, that show no homology to the corresponding regions of the mastadenoviruses and are expected to function in interactions with the host cell (Berk, 2007; Davison et al., 2003a). In the most studied FAdV, FAdV-1, protein functions have been attributed to only three of the predicted 19 genus-specific genes, ORF1 (Weiss et al., 1997), ORF22 (Lehrmann & Cotten, 1999) and Gallus anti-mort-1 (GAM-1) (Chiocca, 2007; Glotzer et al., 2000; Lehrmann & Cotten, 1999).

In the present study, we report the complete genome sequence of a non-pathogenic isolate of FAdV-4 and a combined analysis of protein-coding potential, employing the BD-based Gene Identification tool (IBM) together with evaluation of sequence conservation among FAdV species. We further explore the possibility that a subset of the previously annotated ORFs within the terminal genomic regions of the FAdVs constitute upstream ORFs (uORFs), cis elements in mRNAs that are involved in regulating translation, typically reducing the translation initiation efficiency at the downstream ORF start codon and thereby reducing protein expression (Calvo et al., 2009; Meijer & Thomas, 2002; Morris & Geballe, 2000). Identification of the most probable protein-coding and non-coding ORFs and those ORFs that contribute to uORFs provides an additional basis for future functional studies and will facilitate the rational design of next-generation FAdV-based veterinary vaccine vectors. Further, this work may assist in the identification of virulence factors in FAdVs, including pathogenic IBH/HPS-associated strains of FAdV-4.

RESULTS

General properties of the FAdV-4 genome

The genome of the non-pathogenic FAdV-4 is 45667 bp in length with an overall base composition of 23.3 % A, 27.7 % C, 26.9 % G and 22.1 % T and a G+C content of 54.6 mol%. The FAdV-4 genome was larger than that of FAdV-9 (45063 bp) (Ojkic & Nagy, 2000) and FAdV-1 (43804 bp) (Chiocca et al., 1996), making it the largest AdV genome reported. Employing the inclusion criteria that an ORF must contain a methionine start codon, correspond in size to a peptide \( \geq 50 \) aa, and not be overlapped by greater than 60% of its length by a larger ORF, we identified a total of 46 potential protein-coding ORFs distributed on both strands (57% on the sense strand and 43% on the antisense strand) (Fig. 1). As with other avian adenoviruses, the FAdV-4 genome was organized into a central genomic region (nt 7100–33 000) that consisted primarily of ORFs that were homologous to the genus-common genes and left and right terminal genomic regions (nt 1–7100 and nt 33000–45667, respectively) that consisted of ORFs homologous to the genus-specific genes. Of the 46 ORFs that were identified in the FAdV-4 genome, 18 represented genus-common genes and 28 represented genus-specific genes. The annotation of the predicted protein-coding regions is listed in Table 1.

Phylogeny

Phylogenetic analysis of the predicted amino acid sequences of the hexon gene was performed using the neighbour-joining method of Molecular Evolutionary Genetics Analysis (MEGA) 4.0 (Tamura et al., 2007). The FAdV-4 ON1 isolate clustered as expected with species FAdV-C, close to an IBH/ HPS-associated strain of FAdV-4, supported by high bootstrap values at all nodes (\( \geq 98% \)) (Fig. 2).

A global pairwise alignment of the genome sequence of FAdV-4 with those of FAdV-1, FAdV-9, and human adenovirus 2 (HAdV-2) was carried out using mVISTA Limited Area Global Alignment of Nucleotides (LAGAN) software (Brudno et al., 2003) (Fig. 3). The central region of the FAdV-4 genome displayed high sequence conservation (Fig. 3, shaded regions) with FAdV-1 and FAdV-9 and moderate sequence conservation with HAdV-2 (Fig. 3, white regions). The terminal regions of the FAdV-4 genome showed low sequence conservation with FAdV-1 and FAdV-9 and no sequence conservation with HAdV-2. The specific ORFs that showed regions with greater than 50% sequence conservation with FAdV-1 and/or FAdV-9 are also noted in Table 1.

Protein-coding potential analysis

Assessment of the protein-coding potential of the 46 identified ORFs was carried out by processing the genomic DNA sequence with the BD-based Gene Identification tool (IBM), an application of the Bio-Dictionary-based Gene Finder (BDGF) algorithm (Shibuya & Rigoutsos, 2002) for which there is precedent for eukaryotic viral gene prediction (Murphy et al., 2003; Rigoutsos et al., 2003). We employed a combined analysis that further took into account (i) the existence of orthologous ORFs in other FAdVs, (ii) experimental detection of orthologous transcripts (Cao et al., 1998; Ojkic et al., 2002; Payet et al., 1998), and (iii) explicitly demonstrated orthologous protein expression, or implied protein expression through a point mutation resulting in an altered viral growth phenotype in FAdV-1 (François et al., 2001; Lehrmann & Cotten, 1999). Of the 21 ORFs located within the central genomic region, 18 were homologues of genus-common genes, and, with the exception of U exon and pVII, were assigned high protein-coding potential by BDGF (Fig. 1 and Table 1). Based on the known functions of the genus-common genes in the mastadenoviruses, these 18 ORFs can be expected to encode proteins that are indispensable (Berk, 2007; Davison et al., 2003a). The most notable features observed within the central genomic region were the two predicted fiber genes, fiber 1 and fiber 2, and the
three hypothetical genes with predicted molecular masses of 11.7, 10.3 and 10.5 kDa, respectively (11.7K, 10.3K and 10.5K) (Fig. 1). The two fiber genes were located adjacent to one another between U exon and ORF22. JPred secondary structure prediction of the FAdV-4 fiber 1 and fiber 2 head domains showed a similar β-strand arrangement to FAdV-1 fiber 1 and fiber 2, indicating that both fiber genes are likely to encode functional proteins (Fig. 4). Three hypothetical ORFs, non-homologous to the genus-common genes, were located between DNA-binding protein (DBP) and 100K; 11.7K was assigned low protein-coding potential and 10.3K and 10.5K were assigned high protein-coding potential by BDGF (Fig. 1 and Table 1).

Of the 25 ORFs located within the left and right terminal genomic regions that were homologous to genus-specific genes, 15 and 10 ORFs were assigned high and low protein-coding potential, respectively (Fig. 1 and Table 1). Among the 15 ORFs assigned high protein-coding potential were ORF1, ORF22 and GAM-1, whose orthologues in FAdV-1 are the only FAdV genus-specific genes with functionally characterized proteins (Chiocca, 2007; Glotzer et al., 2000; Lehrmann & Cotten, 1999; Weiss et al., 1997); ORF2, ORF12, ORF13, ORF19A and ORF4, which have cellular and/or viral homologues within the families Parvoviridae, Herpesviridae or Poxviridae (Chiocca et al., 1996; Corredor et al., 2006, 2008; Ojkic & Nagy, 2000; Washietl & Eisenhaber, 2003); and ORF16, which shows conservation of a cellular ADP-ribosyltransferase domain (Corredor et al., 2006; Washietl & Eisenhaber, 2003) (Table 1). The six remaining genus-specific gene orthologues that were assigned high protein-coding potential in our analysis, namely, ORF0, ORF1B, ORF24, ORF20B, ORF43 and ORF17, were without identifiable functional
Table 1. Protein-coding potential analysis of FAdV-4 ORFs

The protein-coding potential for each ORF is indicated in the leftmost column, with dark and light grey fields representing high and low protein-coding potential, respectively. In all other columns, dark grey fields indicate characteristics that suggest an ORF is protein-coding and light grey fields designate characteristics that suggest an ORF is non-coding. Normalized BDGF scores are listed for each ORF unless the score was ≤ 0 or the ORF was excluded by the BDGF algorithm (denoted with a dash). Instances where BDGF designated a start codon other than that defining the largest potential ORF are indicated by a superscript letter. The occurrence of an orthologous ORF in FAdV-1 and/or FAdV-9 is indicated as positive (+) if reports have demonstrated protein expression or that a point mutation or deletion has resulted in an altered phenotype (François et al., 2001; Lehrmann & Cotten, 1999). NR, No reports were identified.

<table>
<thead>
<tr>
<th>FAdV-4 ORF</th>
<th>ORF position</th>
<th>Strand</th>
<th>Length (aa)</th>
<th>Kozak-AUG</th>
<th>BDGF score (normalized)</th>
<th>mVista LAGAN</th>
<th>Transcription</th>
<th>Expression</th>
<th>ORF1BL</th>
<th>ORF1A</th>
<th>ATG Stop</th>
<th>Protein superfamily, function of homologue, putative function</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORF0</td>
<td>478–717</td>
<td>+</td>
<td>79</td>
<td>–</td>
<td>11.1</td>
<td>+ + NR NR</td>
<td>+ + NR NR</td>
<td>+ + NR NR</td>
<td>+ NR NR</td>
<td>+ NR NR</td>
<td>+</td>
<td>uORF on ORF1 mRNA (FAdV-1, -9)</td>
</tr>
<tr>
<td>ORF1</td>
<td>790–1314</td>
<td>+</td>
<td>174</td>
<td>+</td>
<td>2.3</td>
<td>+ + + + +</td>
<td>+ + +</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
<td>dUTPase superfamily</td>
</tr>
<tr>
<td>ORF1B</td>
<td>1485–1808</td>
<td>+</td>
<td>107</td>
<td>–</td>
<td>12.8</td>
<td>- - NR NR</td>
<td>+ + + + + +</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
<td>Parvo_NS1 superfamily</td>
</tr>
<tr>
<td>ORF2</td>
<td>1850–2665</td>
<td>+</td>
<td>271</td>
<td>-</td>
<td>11.5</td>
<td>+ + + + + +</td>
<td>+ + + + + +</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
<td>Related to ORF14</td>
</tr>
<tr>
<td>ORF2A</td>
<td>3203–2667</td>
<td>-</td>
<td>178</td>
<td>0</td>
<td>0.8</td>
<td>- + NR NR</td>
<td>+ + + + + +</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
<td>Related to ORF14</td>
</tr>
<tr>
<td>ORF2Aa</td>
<td>3905–3735</td>
<td>-</td>
<td>176</td>
<td>0</td>
<td>0.2</td>
<td>- + NR + N</td>
<td>+ N + + N +</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
<td>Related to ORF14</td>
</tr>
<tr>
<td>ORF2B</td>
<td>4653–3997</td>
<td>+</td>
<td>228</td>
<td>0</td>
<td>0.7</td>
<td>- + NR + N</td>
<td>+ + + + + +</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
<td>Related to ORF14</td>
</tr>
<tr>
<td>ORF2B4</td>
<td>5396–4667</td>
<td>-</td>
<td>209</td>
<td>0</td>
<td>0.3</td>
<td>- + NR + N</td>
<td>+ + + + + +</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
<td>Related to ORF14</td>
</tr>
<tr>
<td>ORF3</td>
<td>6130–5330</td>
<td>+</td>
<td>266</td>
<td>1</td>
<td>1.5</td>
<td>+ + + + + +</td>
<td>+ + + + + +</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
<td>Related to ORF14</td>
</tr>
<tr>
<td>ORF12</td>
<td>7076–6207</td>
<td>+</td>
<td>289</td>
<td>2</td>
<td>2.2</td>
<td>+ + + + + +</td>
<td>+ + + + + +</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
<td>Related to ORF24</td>
</tr>
<tr>
<td>ORF12a</td>
<td>8274–7090</td>
<td>+</td>
<td>394</td>
<td>-</td>
<td>3.5</td>
<td>+ + + + + +</td>
<td>+ + + + + +</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
<td>Related to ORF24</td>
</tr>
<tr>
<td>ORF13</td>
<td>12046–1252</td>
<td>+</td>
<td>47</td>
<td>-</td>
<td>0.7</td>
<td>+ + + + + +</td>
<td>+ + + + + +</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
<td>Related to ORF24</td>
</tr>
<tr>
<td>ORF13a</td>
<td>13892–12021</td>
<td>-</td>
<td>602</td>
<td>5</td>
<td>5.3</td>
<td>+ + + + + +</td>
<td>+ + + + + +</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
<td>Related to ORF24</td>
</tr>
<tr>
<td>ORF14</td>
<td>13959–15158</td>
<td>+</td>
<td>399</td>
<td>2.6</td>
<td>+ + + + + +</td>
<td>+ + + + + +</td>
<td>+ + + + + +</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
<td>Related to ORF24</td>
</tr>
<tr>
<td>ORF14A</td>
<td>15145–16917</td>
<td>+</td>
<td>590</td>
<td>4.6</td>
<td>+ + + + + +</td>
<td>+ + + + + +</td>
<td>+ + + + + +</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
<td>Related to ORF24</td>
</tr>
<tr>
<td>ORF14B</td>
<td>3203–2667</td>
<td>-</td>
<td>178</td>
<td>-</td>
<td>1.5</td>
<td>+ + + + + +</td>
<td>+ + + + + +</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
<td>Related to ORF24</td>
</tr>
<tr>
<td>ORF15</td>
<td>35087–34746</td>
<td>-</td>
<td>113</td>
<td>-</td>
<td>1.3</td>
<td>- + NR + N</td>
<td>+ + + + + +</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
<td>Related to ORF20 (FAdV-1 and -9)</td>
</tr>
<tr>
<td>ORF15a</td>
<td>34050–33784</td>
<td>-</td>
<td>88</td>
<td>0</td>
<td>0.8</td>
<td>+ * NR NR</td>
<td>+ N + + N +</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
<td>Related to ORF20 (FAdV-1 and -9)</td>
</tr>
<tr>
<td>ORF16</td>
<td>39105–39263</td>
<td>+</td>
<td>52</td>
<td>0.1</td>
<td>+ + NR NR + + NR NR</td>
<td>+ + + + + +</td>
<td>+ + + + + +</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
<td>Related to ORF20 (FAdV-1 and -9)</td>
</tr>
<tr>
<td>ORF16a</td>
<td>3905–39263</td>
<td>+</td>
<td>52</td>
<td>0.1</td>
<td>+ + NR NR + + NR NR</td>
<td>+ + + + + +</td>
<td>+ + + + + +</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
<td>Related to ORF20 (FAdV-1 and -9)</td>
</tr>
<tr>
<td>ORF17</td>
<td>38124–38381</td>
<td>+</td>
<td>41</td>
<td>1.4</td>
<td>+ + + + + +</td>
<td>+ + + + + +</td>
<td>+ + + + + +</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
<td>Related to ORF20 (FAdV-1 and -9)</td>
</tr>
<tr>
<td>ORF17a</td>
<td>41156–40680</td>
<td>-</td>
<td>158</td>
<td>0.9</td>
<td>+ + NR + + NR + + NR</td>
<td>+ + + + + +</td>
<td>+ + + + + +</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
<td>Related to ORF20 (FAdV-1 and -9)</td>
</tr>
<tr>
<td>ORF18</td>
<td>41566–41149</td>
<td>+</td>
<td>135</td>
<td>1.8</td>
<td>+ + + + + +</td>
<td>+ + + + + +</td>
<td>+ + + + + +</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
<td>Related to ORF20 (FAdV-1 and -9)</td>
</tr>
<tr>
<td>ORF18a</td>
<td>42216–44678</td>
<td>+</td>
<td>820</td>
<td>1.3</td>
<td>+ + + + + +</td>
<td>+ + + + + +</td>
<td>+ + + + + +</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
<td>Related to ORF20 (FAdV-1 and -9)</td>
</tr>
<tr>
<td>ORF19</td>
<td>4472–45308</td>
<td>+</td>
<td>178</td>
<td>1.8</td>
<td>+ + + + + +</td>
<td>+ + + + + +</td>
<td>+ + + + + +</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
<td>Related to ORF20 (FAdV-1 and -9)</td>
</tr>
<tr>
<td>ORF19a</td>
<td>4472–45308</td>
<td>+</td>
<td>178</td>
<td>1.8</td>
<td>+ + + + + +</td>
<td>+ + + + + +</td>
<td>+ + + + + +</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
<td>+ + + + +</td>
<td>Related to ORF20 (FAdV-1 and -9)</td>
</tr>
</tbody>
</table>

a, 17160–18566; b, 23921–25243; c, 26607–28940; d, 35330–36400; e, 38124–38381; f, 41149–41541.

*FAdV-1 and FAdV-9 ORF20A do not begin with a start codon.
domains or cellular or viral homologues (Table 1). Of the 15 ORFs assigned high protein-coding potential by BDGF, only three, ORF43, ORF19A and ORF4, were not conserved between FaDV-4 and FaDV-1 and/or FaDV-9 in the BLASTp and LAGAN analyses (Table 1 and Fig. 3), and all 15 ORFs were conserved between FaDV-4 CA, FaDV-10 C-2B (Corredor et al., 2006, 2008) and FaDV-4 ON1 (unpublished results). Ten of the 15 ORFs had orthologues in FaDV-1 and/or FaDV-9 with reported mRNA expression, including ORF1, ORF1B, ORF2, ORF13, ORF12, ORF22, ORF20B, GAM-1, ORF16 and ORF19A (Table 1). Lastly, eight of the 15 ORFs had orthologues in FaDV-1 that either encode proteins or for which indirect evidence of protein expression had been shown through mutational analysis or deletion resulting in an altered phenotype, namely ORF1, ORF2, ORF24, ORF12, ORF22, GAM-1, ORF17 and ORF16.

---

**Fig. 2.** Phylogenetic analysis of FaDV-4 based on the predicted amino acid sequences of the hexon. The phylogenetic tree was constructed using the neighbour-joining, p-distance method in MEGA version 4.0 (Tamura et al., 2007). The Gonnet protein weight matrix was used to generate the CLUSTAL W alignments. Percentage bootstrap confidence levels as determined for 1000 pseudoreplicates are shown at the relevant internal nodes. Genetic distance in substitutions per nucleotide is indicated by the scale in the lower left of the panel. White sturgeon adenovirus (WSAdV-1) was included as an outgroup. FaDV, Falcon adenovirus; EDS, egg drop syndrome virus; HEV, hemorrhagic enteritis virus; IBH/HPS, strain from India.

---

**Fig. 3.** Global pairwise sequence alignment of FaDV-4 with FaDV-9, FaDV-1 and HAdV-2, using mVISTA LAGAN. The height of the plot on the y-axis indicates the percentage nucleotide identity of selected AdV genomes with the FaDV-4 genome. The x-axis indicates the nucleotide position on the FaDV-4 genome. Regions of high sequence conservation are shaded. The FaDV-4 genome is shown to scale above the alignment. ORFs are depicted as arrows with dark grey and light grey representing high and low protein-coding potential, respectively. Chevrons depict the boundaries of the central genomic region.
The 10 ORFs that were assigned low protein-coding potential consisted of ORF14A, ORF14, ORF14B, ORF20A, ORF20, ORF19, ORF27, ORF28, ORF29 and ORF30. Of these, only ORF19 had an orthologue with a functional domain or cellular or viral homologue (Corredor et al., 2008; Washietl & Eisenhaber, 2003). Three of the 10 ORFs, ORF20A, ORF20 and ORF29, were highly conserved between FAdV-4 and FAdV-1 and/or FAdV-9 in the BLASTP and LAGAN analyses (Table 1 and Fig. 3), and four ORFs, ORF14A, ORF14, ORF14B and ORF19, were highly conserved between FAdV-4 and FAdV-1 and/or FAdV-9 in the BLASTP analysis (Table 1). Only two of the ten ORFs, ORF20 and ORF19, had orthologues in FAdV-1 and/or FAdV-9 with reported mRNA expression, and only ORF14 had an orthologue for which there was indirect evidence of protein expression (Table 1). A flowchart of the analysis is depicted in Supplementary Fig. S1 (available in JGV Online).

uORFs in the 5′-UTR region of FAdV-4 mRNAs

To explain the occurrence of several ORFs that were not predicted to have high protein-coding potential, but were highly conserved among FAdVs, we hypothesized that a subset of the low coding potential ORFs that corresponded to a peptide of 100 aa in length (ORF20A, ORF20, ORF27, ORF28, ORF29 and ORF30) contributed to uORFs, defined by a start codon in the 5′ UTR that is out-of-frame with the downstream coding sequence. RT-PCR was carried out on RNA extracted from FAdV-4-infected cells with ORF22 gene-specific primers (see Fig. 5b, upper panel).
for primer locations and Supplementary Table S1, available in JGV Online, for primer sequences). ORF22 mRNA was detected at 6, 12 and 18 h post-infection (p.i.) (Fig. 5a). Sequencing of the PCR product revealed that ORF20A and ORF20 both contributed to large uORFs in the ORF22 mRNA (Fig. 5b, upper panel). The ORF22 mRNA contained a long 5' UTR, greater than 718 nt in length. The ORF20 uORF was 277 nt (91 aa) in length with several upstream start codons (uAUGs), and the ORF20A uORF, located immediately upstream (3 nt) of the ORF22 AUG, was 267 nt in length (91 aa) with no intervening uAUGs (Fig. 5b, upper panel). We reconstituted the ORF22 mRNA transcript sequences of FAdV-1 and FAdV-9, in silico, based on the experimentally determined splice donor and acceptor sites (Ojkic et al., 2002; Payet et al., 1998) and used GeneQuest (Lasergene v8.0; DNASTAR) to examine the transcripts for the presence of uORFs. While splicing did not occur within the sequenced region of the FAdV-4 ORF22 5' UTR, analysis of the FAdV-1 and FAdV-9 ORF22 5' UTRs showed interspecies variation in splice donor and acceptor sites that directed the removal of different portions of ORF20 and ORF20A (Fig. 5b, lower panels). All three of the examined FAdV serotypes (FAdV-1, -4 and -9) had a long ORF22 5' UTR that contained a minimum of four uAUGs and at least one large uORF. Additional RT-PCR analysis revealed that ORF29 was an uORF in the 5' UTR of the FAdV-4 GAM-1 mRNA and that the TR-1 repeat region, nested entirely within ORF29, conferred with each repeat unit an additional uAUG to ORF29 (unpublished results).

Codon usage analysis with DAMBE revealed that 11 of the 14 ORFs with the largest cumulative difference in codon usage from Gallus gallus were ≤100 aa in length, eight of which were low coding potential ORFs, namely ORF29, pVII, hypothetical 11.7 kDa, ORF30, ORF27, U exon, ORF28 and ORF20A (Table 1, coloured light grey).

**DISCUSSION**

First-generation recombinant FAdV vectors have been successfully developed as veterinary vaccines; however, the
limited understanding of the protein-coding regions and virus–host interactions restrict their progression into next-generation vectors. The complete 45,667 bp genome sequence of a non-pathogenic FAdV-4 was determined and 46 putative genes were identified. We evaluated the protein-coding potential of FAdV-4 and discriminated each of the putative genes into one of two classes, representing either low or high protein-coding potential. Through RT-PCR analysis, we determined that several ORFs that were assigned low protein-coding potential in our analysis occurred as large uORFs, greater than 153 nt (50 aa) in length, ORF20A and ORF20 in the ORF22 mRNA (Table 1 and Fig. 5) and ORF29 in the GAM-1 mRNA (Table 1; unpublished results).

Accurate determination of FAdV protein-coding regions, given the limited availability of experimental data, poses a challenge. The high G + C content of the FAdVs (Chiocca et al., 1996; Ojkic & Nagy, 2000) correlates with a reduced number of randomly occurring A/T-rich stop codons and may confer a greater prevalence of small non-coding ORFs (Silke, 1997). Further, in every FAdV species that has been examined, extended genomic regions have been identified that consist solely of small ORFs corresponding to proteins less than 100 aa in length (Chiocca et al., 1996; Corredor et al., 2006, 2008; Davison et al., 2003a; Ojkic & Nagy, 2000). Additionally, HAdV-2 E3 CR1-alpha (E3 6.7K) and E3 RID-alpha (E3 10.4K) are 61 and 91 aa in length, respectively (Berk, 2007). The conventional FAdV annotation criterion that defines the minimal ORF size as corresponding to 50 aa in length (Chiocca et al., 1996; Ojkic & Nagy, 2000) is sensible; however, it increases the likelihood that spurious ORFs are retained within genome annotations. In our study, BDGF discriminated between small ORFs of high (ORF0, 10.3K and 10.5K) and low conservation among closely related species has proven to be very effective (Clamp et al., 2002). The overlapping uORF, ORF29, has the potential to confer regulation of FAdV-4 GAM-1 expression in this manner. Ribosome shunting (Morley & Coldwell, 2008) has been described for several viral transcripts, including the human adenovirus late mRNAs (Yueh & Schneider, 1996). It was shown that uORFs are necessary structures for ribosome shunting on the cauliflower mosaic virus (CaMV) 35S RNA (Pooggin et al., 2000) and on the prototype foamy virus (PFV) (pre)genomic RNA (pgRNA) (Schepetilnikov et al., 2009). Lastly, uORFs could enable IRES-mediated translation during cellular stress in a manner similar to cat-1 Arg/Lys transporter (Fernandez et al., 2005). The common theme of the aforementioned mechanisms is that they confer selective gene expression under specific cellular conditions, which could be relevant
The absence or occurrence of one or more upstream ORFs (uORFs) within the 5'-UTR of each mRNA is indicated. Transcript sequences were reconstituted in silico based on the experimentally determined splice donor and acceptor sites that are reported for FAdV-1 (Payet et al., 1998), FAdV-9 (Ojkic et al., 2002), HAdV-2 (Broker, 1984; Perricaudet et al., 1979; Virtanen et al., 1984; Wold et al., 1984) and BAdV-3 (Idamakanti et al., 1999; Reddy et al., 1998, 1999; Zheng et al., 1999). Boxes with two shades of grey indicate that transcripts encoded within the 5'-UTR of these mRNAs contain uORFs in 5'-UTR, those coded by this gene have been detected with and without a uORF(s). A detailed list of the FAdV uORFs is provided as Supplementary Table S2 (available in JGV Online).

### Table 2. uORFs in HAdV-2, BAdV-3, FAdV-1 and FAdV-9 mRNAs

To determine whether uORFs were common features in FAdV transcripts, we reconstituted the mRNA transcript sequences of the avian adenoviruses, FAdV-1 and FAdV-9, in silico, based on the experimentally determined splice donor and acceptor sites (Ojkic et al., 2002; Payet et al., 1998). Subsequent examination using GeneQuest (Lasergene v8.0; DNASTAR) identified uORFs in a number of FAdV transcripts with the potential to act as translational regulatory elements (Table 2). In several instances, annotated ORFs from FAdV-1 and/or FAdV-9 were present, in full-length, upstream of a second full-length ORF in potential bicistronic transcripts, or contributed to large uORFs, greater than 153 nt (50 aa) in length. The potential bicistronic mRNAs included FAdV-1 ORF12 (FAdV-1 ORF13 mRNA) and ORF16 (FAdV-1 DBP mRNA). Previously annotated ORFs that contributed to large uORFs included ORF0 (FAdV-1 and FAdV-9 ORF1 mRNA), ORF20 (FAdV-1 ORF22 mRNA), and pVII [FAdV-1 pX (μ) mRNA]. Our analysis also revealed two previously unannotated FAdV-9 ORFs and one FAdV-1 ORF that resulted from the fusion of two or more exons and defined a large uORF in their respective mRNAs, ORF41008–41287, 41378–41385 (FAdV-9 ORF11 mRNA), ORF33998–34018, 32741–32986 (FAdV-9 ORF22 mRNA), and ORF39291–39305, 24716–24780, 23455–23667, 23289–23340 (FAdV-1 DBP mRNA). Additionally the unannotated FAdV-1 ORF, ORF23198–23566, was identified as a large uORF in the 100K mRNA. Small uORFs less than 153 nt (50 aa) in length were identified in several transcripts encoding both genus-common and genus-specific genes, including FAdV-1 ORF13, ORF22, ORF19, DBP, 100K and pX, and FAdV-9 ORF13, DNA pol, pTP, ORF22 and ORF11. FAdV-4 ORF0 and ORF16 were also among the 14 ORFs with the largest cumulative difference in codon usage from Gallus gallus (Table 1). A detailed list of uORFs in the FAdVs with their nucleotide coordinates is provided as Supplementary Table S2 (available in JGV Online).

<table>
<thead>
<tr>
<th>Gene</th>
<th>Sense</th>
<th>Species</th>
<th>Gene</th>
<th>Sense</th>
<th>Species</th>
</tr>
</thead>
<tbody>
<tr>
<td>E1A</td>
<td>+</td>
<td>HAdV-C</td>
<td>ORF0</td>
<td>+</td>
<td>FAdV-A</td>
</tr>
<tr>
<td>E1B 19K</td>
<td>+</td>
<td>HAdV-B</td>
<td>ORF1</td>
<td>+</td>
<td>FAdV-D</td>
</tr>
<tr>
<td>E1B 55K</td>
<td>+</td>
<td></td>
<td>ORF1A</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>IX</td>
<td>-</td>
<td></td>
<td>ORF1B</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>IA2</td>
<td>-</td>
<td></td>
<td>ORF1C</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Pol</td>
<td>-</td>
<td></td>
<td>ORF2</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>gTP</td>
<td>+</td>
<td></td>
<td>ORF24</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>52K</td>
<td>+</td>
<td></td>
<td>ORF14</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>pIIa</td>
<td>+</td>
<td></td>
<td>ORF13</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>+</td>
<td></td>
<td>ORF12</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>pVII</td>
<td>+</td>
<td></td>
<td>ORF1</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>V</td>
<td>+</td>
<td></td>
<td>ORF2</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>pX</td>
<td>-</td>
<td></td>
<td>ORF1</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>pVI</td>
<td>+</td>
<td></td>
<td>ORF10</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Hexon</td>
<td>-</td>
<td></td>
<td>ORF20</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Protease</td>
<td>-</td>
<td></td>
<td>pVII</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>pV</td>
<td>+</td>
<td></td>
<td>DBP</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>pVII</td>
<td>+</td>
<td></td>
<td>100K</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>92K</td>
<td>+</td>
<td></td>
<td>33K</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>E3 12.5K</td>
<td>+</td>
<td></td>
<td>22K</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>E3 CR1-a (6.7K)</td>
<td>+</td>
<td></td>
<td>pVIII</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>E3 gp 19K</td>
<td>+</td>
<td></td>
<td>U exon</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>11.6 kD protein</td>
<td>-</td>
<td></td>
<td>Fiber 1</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>E3 RID-a (10.4K)</td>
<td>+</td>
<td></td>
<td>Fiber 2</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>E3 RID-β (14.5K)</td>
<td>+</td>
<td></td>
<td>pVII</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>E3 14.7K</td>
<td>+</td>
<td></td>
<td>E4 ORF67</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>U exon</td>
<td>-</td>
<td></td>
<td>E4 ORF5</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Fiber</td>
<td>-</td>
<td></td>
<td>E4 ORF6</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>E4 ORF67</td>
<td>+</td>
<td></td>
<td>E4 ORF4</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>E4 ORF4</td>
<td>+</td>
<td></td>
<td>E4 ORF3</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>E4 ORF3</td>
<td>+</td>
<td></td>
<td>E4 ORF2</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>E4 ORF2</td>
<td>+</td>
<td></td>
<td>E4 ORF1</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>BAdV E4 ORF1</td>
<td>-</td>
<td></td>
<td>BAdV E4 ORF1</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>BAdV E4 ORF2</td>
<td>-</td>
<td></td>
<td>BAdV E4 ORF2</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>BAdV E4 ORF3</td>
<td>-</td>
<td></td>
<td>BAdV E4 ORF3</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>BAdV E4 ORF4</td>
<td>-</td>
<td></td>
<td>BAdV E4 ORF4</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>BAdV E4 ORF5</td>
<td>-</td>
<td></td>
<td>BAdV E4 ORF5</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

Gene not present in genome
Transcript has not been characterized
mRNA does not contain uORF(s) in 5'-UTR
mRNA contains uORF(s) in 5'-UTR
during the cellular stress that is known to occur during FAdV (Glotzer et al., 2000) and HAdV (Berk, 2007; O’Shea et al., 2005) infection.

### Similar analysis of previously reported transcript sequences for the mastadenoviruses, HAdV-2 and bovine adenovirus 3 (BAdV-3) revealed uORFs in a number of early gene transcripts (Table 2). In both HAdV-2 and BAdV-3, the sole transcript detected encoding E1B-55K is partially overlapped by E1B-19K in a potential bicistronic mRNA (Broker, 1984; Reddy et al., 1999; Zheng et al., 1999). Further, the most abundant transcript encoding HAdV-2 E3-14.5K, also the sole detected transcript encoding E3-10.4K, is a potential bicistronic mRNA, encoding both subunits of the receptor internalization and degradation (RID) complex, 10.4K and 14.5K, and also contains a 195 nt (64 aa), E3-10.4K-overlapping uORF, named 7.5K (Broker, 1984; Wilson-Rawls et al., 1990). Further, both of the detected transcripts encoding E3-19K also encode the E3-19K-overlapping, E3-6.7K, in potential bicistronic
mRNAs and contain a 102 nt (33 aa) uORF, named 3.6K, that is present in all the E3 mRNAs (Broker, 1984; Wilson-Rawls et al., 1990). Protein expression from 3.6K and the larger 7.5K has never been demonstrated, whereas protein expression from 6.7K has been shown (Broker, 1984; Wilson-Rawls et al., 1990). There is, therefore, precedent for both non-coding and protein-coding uORFs in HAdV-2. Within the BAdV-3 E3 transcription unit, the E3-121R contains a 78 nt (25 aa) uORF (Idamakanti et al., 1999). Small uORFs less than 153 nt (50 aa) occur in several HAdV-2 and BAdV-3 transcripts, including HAdV-2 E4-34K and E4-ORF6/7 (Virtanen et al., 1984) and BAdV-3 E1A (Reddy et al., 1999) and E3-284R (Idamakanti et al., 1999).

The identification of several large uORFs in FAdV-1, -9 and -4 mRNAs suggests that exhaustive transcriptome analysis, including 5′ and 3′ RACE analysis, would be informative. Without consideration of the 5′-UTR sequences, a point mutation or deletion designed to knock out a putative ORF could inadvertently disrupt a uORF on another mRNA, resulting in a change of phenotype attributed to the wrong genome region. For example, the lack of observable CPE caused by a point mutation in FAdV-1 ORF16 that was attributed to loss of ORF16 protein function (François et al., 2001) may have resulted from disruption of the ORF16 uORF in the DBP 5′-UTR. Reliable interpretation of genome deletions is important for vaccine vector design and functional studies. Further, it remains a possibility that uORF-dependent translational regulation of the as-yet-undetermined virulence factors of FAdVs plays a role in FAdV pathogenesis, including FAdV-4-associated IBH/HPS.

**ACKNOWLEDGEMENTS**

We thank Dr Davor Ojkic of the Animal Health Laboratory, University of Guelph for generously providing the virus isolate. We thank Dr Isidore Rigoutsos of the IBM Bioinformatics and Pattern Discovery Group for consultation and for insightful correspondences. We also thank Drs Helena Grgić and Juan Carlos Corredor, and Mrs Sheila Watson for invaluable technical assistance and Dr Peter J. Krell for critical reading of the manuscript. 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**


3389–3402.


731.

cancer virotherapy.

analysis of terminal regions of the fowl adenovirus type 8 genome.

the fowl adenovirus 9 genome is suitable for foreign gene insertion/

1270


Clamp, M., Fry, B., Kamal, M., Xie, X. H., Cuff, J., Lin, M. F., Kellis, M.,


Chiocca, S., Kurzbauer, R., Schaffner, G., Baker, A., Mautner, V. &


Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. Proc Natl Acad Sci U S A 106, 7507–7512.


