Genetic diversity of species Fowl aviadenovirus D and Fowl aviadenovirus E

Ana Marek,1 Győző L. Kaján,2 Carolin Kosiol,3 Mária Benkö,2 Anna Schachner4 and Michael Hess1,4

1Clinic for Poultry and Fish Medicine, Department for Farm Animals and Veterinary Public Health, Vetmeduni Vienna, Vienna, Austria
2Institute for Veterinary Medical Research, Centre for Agricultural Research, Hungarian Academy of Sciences, Budapest, Hungary
3Institut für Populationsgenetik, Vetmeduni Vienna, Vienna, Austria
4Christian Doppler Laboratory for Innovative Poultry Vaccines, University of Veterinary Medicine, Vienna, Austria

Complete genomes of eight reference strains representing different serotypes within the species Fowl aviadenovirus D (FAdV-D) and Fowl aviadenovirus E (FAdV-E) were sequenced. The sequenced genomes of FAdV-D and FAdV-E members comprise 43,287 to 44,336 bp, and have a gene organization identical to that of an earlier sequenced FAdV-D member (strain A-2A). Highest diversity was noticed in the hexon and fiber genes and ORF19. All genomes sequenced in this study contain one fiber gene. Phylogenetic analyses and G+C content support the division of the genus Aviadenovirus into the currently recognized species. Our data also suggest that strain SR48 should be considered as FAdV-11 instead of FAdV-2 and similarly strain HG as FAdV-8b. The present results complete the list of genome sequences of reference strains representing all serotypes in species FAdV-D and FAdV-E.

INTRODUCTION

Aviadenoviruses infect avian hosts exclusively. Fowl aviadenoviruses (FAdVs) are grouped into five species (Fowl aviadenovirus A to Fowl aviadenovirus E) in the genus Aviadenovirus based on genome organization and phylogeny (Harrach et al., 2011; Harrach & Kajan, 2011). An informal abbreviation of FAdV species, such as for example FAdV-A for Fowl aviadenovirus A, will be used in this paper. FAdVs are widely distributed, and some of them cause inclusion body hepatitis (IBH), hepatitis-hydropericardium syndrome (HHS) and adenoviral gizzard erosions (AGE) in chickens (Hess, 2013). FAdV strains belonging to species FAdV-D and FAdV-E have been isolated mostly from IBH cases, and members of species FAdV-C from HHS outbreaks (Hess et al., 1999; Ojkic et al., 2008; Slavec et al., 2013; Steer et al., 2011; Zadravec et al., 2011). AGE, associated with FAdV-1 infection, have been described in chickens in Japan and Europe (Domanska-Blicharz et al., 2011; Kecskeméti et al., 2012; Manarolla et al., 2009; Marek et al., 2010a; Ono et al., 2001). Before the era of DNA sequencing, serology was the principal means of identifying aviadenovirus types and the 12 serotypes have been grouped into five FAdV species recognized to date as follows: FAdV-A (FAdV-1), FAdV-B (FAdV-5), FAdV-C (FAdV-4 and FAdV-10), FAdV-D (FAdV-2, FAdV-3, FAdV-9 and FAdV-11) and FAdV-E (FAdV-6, FAdV-7, FAdV-8a and FAdV-8b) (Harrach et al., 2011; Hess, 2000). DNA sequencing of the loop 1 (L1) region of the hexon gene is now used frequently for typing FAdVs (Kajan et al., 2013; Marek et al., 2010b; Meulemans et al., 2004; Raue & Hess, 1998).

High-throughput sequencing has become popular in recent years since it permits the rapid and comprehensive analysis of complete aviadenovirus genomes. At least one complete genome sequence is now available for all FAdV species, including: FAdV-A (FAdV-1, also known as CELO virus), FAdV-B (FAdV-5 strain 340), FAdV-C (FAdV-4 strains ON1 and KR5), FAdV-D (FAdV-9 strain A-2A) and FAdV-E (FAdV-8 strain HG) (Chiocca et al., 1996; Grgić et al., 2011; Griffin & Nagy, 2011; Marek et al., 2012, 2013; Ojkic & Nagy, 2000). In addition, the whole genome of numerous non-chicken aviadenoviruses has also been sequenced. These are TAdV-1 (Turkey aviadenovirus B, TAdV-B), GoAdV-4 (Goose aviadenovirus A, GoAdV-A), TAdV-4...
(Turkey aviadenovirus C, TAdV-C), TAdV-5 (Turkey aviadenovirus D, TAdV-D), PiAdV-1 (Pigeon aviadenovirus A, PiAdV-A) and DAdV-2 (Duck aviadenovirus B, DAdV-B) (Kajan et al., 2010, 2012; Marek et al., 2014a, b).

Adenoviruses in general are thought to have co-evolved with a wide range of vertebrate hosts, and thus the genus Aviadenovirus with birds (Harrach, 2014). In this genus, we can indeed observe at least two major clusters containing the AdVs of anseriform birds (DAdV-B and GoAdV-A), and the other the AdVs originating from galliformes, i.e. turkey and fowl adenoviruses (Marek et al., 2014b). The two species that seem to contain the majority of FAdV sero- and genotypes are FAdV-D and -E encompassing eight different FAdV types (Marek et al., 2014b). This might indicate that the viruses in these species have been co-evolving with chickens for a long period. However, the close relatedness and mixed phylogenetic position of the turkey and fowl adenoviruses (Marek et al., 2014b), as well as the high pathogenicity of certain FAdV types, imply that host switches also might have occurred. The increased pathogenicity of a virus is often the consequence of a host switch (Benko & Harrach, 2003; Kohl et al., 2012).

For the correct reconstruction of aviadenovirus evolution, it is important to analyse the whole genomes of additional isolates including, first of all, strains representing not yet examined FAdV types. The main purpose of this study was to obtain the complete genome sequences of reference strains of different types belonging to species FAdV-D and FAdV-E by high-throughput sequencing technology. With the completion of these genome sequences, we expected to gain additional insights into the evolution of the genus Aviadenovirus.

**RESULTS**

**Genome organization**

After filtering for contaminating chicken chromosomal sequence reads, the average coverage for all sequenced genomes was between 250 and 27 000 reads per nucleotide.

**De novo** assembly was optimal when using 1 to 100% of these data (depending on the coverage). Gap closure by PCR and Sanger sequencing resulted in final genome sequences ranging between 43 287 and 44 336 bp with nucleotide composition ranging between 52.8 and 58.0% G+C content (Table 1). The percentage sequence identities to available complete aviadenovirus genome sequences are summarized in Table 2. The intraspecies sequence identities varied between 89.4 and 97.1% for different FAdV-D strains and 92.7 and 97.1% for different FAdV-E strains. The interspecies nucleotide sequence identities varied between 71.2 and 75.4% for FAdV-D and FAdV-E strains. Strain SR48 (FAdV-2) showed higher sequence identity to strain 380 (FAdV-11, 97.1%) than to strain 685 (FAdV-2, 95.8%). Strain HG (FAdV-8) showed higher sequence identity to strain 764 (FAdV-8b, 97.1%) than to strain TR59 (FAdV-8a, 94.1%). Interestingly, strain CR119 (FAdV-6) shared very high sequence identity (97.0%) with strain YR36 (FAdV-7).

All sequenced genomes had a gene organization identical to that of the previously sequenced FAdV-9 (FAdV-D strain A-2A) (Fig. 1).

Global pairwise sequence alignment analyses identified areas of great interspecies diversity. The results for one member of each of the FAdV-D and FAdV-E species (685 and CR119, respectively) are shown in Fig. 1 and for all other FAdV-D and FAdV-E members in Fig. S1 (available in the online Supplementary Material). The genomes of FAdV-D and FAdV-E members display high sequence conservation in the central genomic region (from IVa2 to fibre gene) with all aviadenovirus genomes, and in the terminal genomic regions with each other as well. The terminal regions show lower sequence conservation or none with other aviadenoviruses.

All FAdV-D members sequenced until now have shown high sequence conservation throughout the genome. However, strain 685 has an additional non-coding sequence region near the right genome end in comparison to other sequenced FAdV-D strains (Fig. 1). Strain SR49 shows lower sequence conservation with other sequenced FAdV-D

<table>
<thead>
<tr>
<th>Virus strain</th>
<th>Genome length (bp)</th>
<th>G+C %</th>
<th>GenBank accession no.</th>
<th>Species/type</th>
</tr>
</thead>
<tbody>
<tr>
<td>685</td>
<td>44 336</td>
<td>53.3</td>
<td>KT862805</td>
<td>FAdV-D/FAdV-2</td>
</tr>
<tr>
<td>SR48</td>
<td>43 632</td>
<td>53.3</td>
<td>KT862806</td>
<td>FAdV-D/FAdV-2*</td>
</tr>
<tr>
<td>SR49</td>
<td>43 337</td>
<td>52.8</td>
<td>KT862807</td>
<td>FAdV-D/FAdV-3</td>
</tr>
<tr>
<td>380</td>
<td>43 347</td>
<td>53.3</td>
<td>KT862812</td>
<td>FAdV-D/FAdV-11</td>
</tr>
<tr>
<td>CR119</td>
<td>43 810</td>
<td>57.8</td>
<td>KT862808</td>
<td>FAdV-E/FAdV-6</td>
</tr>
<tr>
<td>YR36</td>
<td>43 525</td>
<td>57.8</td>
<td>KT862809</td>
<td>FAdV-E/FAdV-7</td>
</tr>
<tr>
<td>TR59</td>
<td>43 287</td>
<td>58.0</td>
<td>KT862810</td>
<td>FAdV-E/FAdV-8a</td>
</tr>
<tr>
<td>764</td>
<td>43 666</td>
<td>57.8</td>
<td>KT862811</td>
<td>FAdV-E/FAdV-8b</td>
</tr>
</tbody>
</table>

*Proposed FAdV-11.

A. Marek and others
Table 2. Percentage sequence identities of complete aviadenovirus genomes

<table>
<thead>
<tr>
<th>Species</th>
<th>FAdV-A</th>
<th>FAdV-B</th>
<th>FAdV-C</th>
<th>FAdV-D</th>
<th>FAdV-E</th>
<th>FAdV-G</th>
<th>FAdV-H</th>
</tr>
</thead>
<tbody>
<tr>
<td>SR48</td>
<td>100</td>
<td>93.5</td>
<td>95.9</td>
<td>95.3</td>
<td>95.1</td>
<td>95.7</td>
<td>97.2</td>
</tr>
<tr>
<td>SR49</td>
<td>100</td>
<td>97.8</td>
<td>98.4</td>
<td>97.8</td>
<td>97.5</td>
<td>98.2</td>
<td>97.1</td>
</tr>
<tr>
<td>HBQ12</td>
<td>100</td>
<td>98.9</td>
<td>96.7</td>
<td>96.5</td>
<td>96.8</td>
<td>96.4</td>
<td>96.0</td>
</tr>
<tr>
<td>BJH13</td>
<td>100</td>
<td>97.8</td>
<td>97.3</td>
<td>97.6</td>
<td>97.3</td>
<td>97.5</td>
<td>97.4</td>
</tr>
<tr>
<td>A-2A</td>
<td>100</td>
<td>93.2</td>
<td>95.8</td>
<td>95.4</td>
<td>95.1</td>
<td>95.9</td>
<td>96.2</td>
</tr>
<tr>
<td>CR119</td>
<td>100</td>
<td>97.6</td>
<td>97.4</td>
<td>97.6</td>
<td>97.3</td>
<td>97.4</td>
<td>97.5</td>
</tr>
<tr>
<td>YR36</td>
<td>100</td>
<td>93.9</td>
<td>94.1</td>
<td>93.9</td>
<td>93.8</td>
<td>93.9</td>
<td>93.8</td>
</tr>
<tr>
<td>TR59</td>
<td>100</td>
<td>97.0</td>
<td>96.8</td>
<td>97.0</td>
<td>96.7</td>
<td>97.0</td>
<td>96.8</td>
</tr>
<tr>
<td>764</td>
<td>100</td>
<td>97.1</td>
<td>96.9</td>
<td>97.1</td>
<td>96.8</td>
<td>97.1</td>
<td>96.9</td>
</tr>
<tr>
<td>HG</td>
<td>100</td>
<td>52.4</td>
<td>63.5</td>
<td>57.2</td>
<td>57.2</td>
<td>56.4</td>
<td>57.2</td>
</tr>
</tbody>
</table>

*Proposed FAdV-11.

**DISCUSSION**

The genus *Aviadenovirus* encompasses fowl aviadenoviruses (FAdVs), which were grouped into 12 serotypes (FAdV-1 to -8a and -8b to -11) based on cross-neutralization tests (Hess, 2000). Recently, at least 12 genotypes were revealed by sequence analysis of the hexon loop 1 (L1) region (Marek et al., 2010b). The 12 serotypes constitute five ‘groups’ (now species *Fowl aviadenovirus A* to *Fowl aviadenovirus E*) initially established on the basis of restriction enzyme digest pattern of whole genomes (Zsak & Kisary, 1984). Phylogenetic and sequence analyses of whole genomes supported the division of the genus *Aviadenovirus* into the currently recognized species (Marek et al., 2012, 2013, 2014a, b; Pauly et al., 2015).

Whole-genome sequence identities among members of the various officially accepted aviadenovirus species range from 42.4% [between TAdV-1 (TAdV-B) and GoAdV-4 (GoAdV-A)] to 72.2% [between FAdV-9 (FAdV-D) and FAdV-8b (FAdV-E)] (Marek et al., 2013). In the present study, phylogenetic and sequence analyses confirmed the present division of the genus *Aviadenovirus* into species. The lowest genome sequence identity between the FAdV-D and FAdV-E members and members of different *aviadenovirus* species was 45.9% [between the FAdV-2 (strain 685) and
Fig. 1. Global comparisons of the genome sequences of (a) FAdV-2 (FAdV-D strain 685) and (b) FAdV-6 (FAdV-E strain CR119) to those of other aviadenoviruses. Peaks show regions having >50% sequence identity. At the top, the rightward- and leftward-transcribed strands of the genome are shown in grey with a 2000-nucleotide scale indicated on the latter. The six reading frames are shown in light grey above and below the genome. Protein-encoding regions are depicted as coloured arrows and bars (the ORF prefix omitted). The genes marked by red arrows are conserved in every AdV sequenced to date; green arrows have orthologues in other aviadenoviruses only. Splice sites are indicated by diagonal lines. DBP, DNA-binding protein; ITR, inverted terminal repeat (blue); pTP, terminal protein precursor; *, proposed FAdV-11.
DAdV-2 (DAdV-B)] and 45.0 % between the FAdV-8b (strain HG) and DAdV-2 (DAdV-B), respectively. The highest genome sequence identity was 75.4 % between FAdV-8b (FAdV-E strain 764) and FAdV-3 (FAdV-D strain SR49) (Table 2). Phylogenetic analysis based on the amino acid sequence of the DNA polymerases shows phylogenetic
differences greater than the required 5–15% (Fig. 3b). Therefore, although FAdV-D and FAdV-E are closely related genetically (Grigić et al., 2011; Marek et al., 2012, 2013, 2014a, b), they represent two different aviadenovirus species, which is also supported by differences in the G+C content (Table 1).

To date, the complete genome sequence for a member of FAdV-E was only available for the isolate HG (Grigić et al., 2011). This strain was labelled as FAdV-8 and to date is not assigned to a FAdV type (FAdV-8a or -8b). However, based on partial hexon gene sequences, the clustering of this strain together with FAdV-8b strains was already observed (Marek et al., 2014a, 2014b). This is now supported by the full genome sequence. Originally, typing of FAdV was achieved by cross-neutralization test and the strain SR48 was considered a reference strain of FAdV-2 (McFerran & Connor, 1977). However, partial hexon gene sequences demonstrated the grouping of this strain together with FAdV-11 strains (Marek et al., 2010b; Meulemans et al., 2004). The present

---

**Fig. 2.** Phylogenetic tree based on all available whole-genome sequences of avian AdVs. Genomes of strains 685, SR48, SR49, 380, CR119, YR36, TR59 and 764 (shown in bold) were sequenced in this study whereas the other avian AdV genome sequences have been published previously (Chiocca et al., 1996; Grigić et al., 2011; Griffin & Nagy, 2011; Hess et al., 1997; Kajan et al., 2010, 2012; Kovacs & Benko, 2011; Marek et al., 2012, 2013, 2014a, 2014b; Ojíc & Nagy, 2000; Park et al., 2012; Pitcovski et al., 1998; To et al., 2014; Vera-Hernández et al., 2015; Zhao et al., 2015). Branch lengths are given in number of substitutions per site (see the scale). Bootstrap values are given in percentage for 1000 datasets, and the tree was rooted at the midpoint. *, Proposed FAdV-11; AdV, adenovirus; DAdV, duck AdV; FAdV, fowl AdV; GoAdV, goose AdV; PiAdV, pigeon AdV; PsAdV, psittacine AdV; RAdV, raptor AdV; SPSkAdV, South Polar skua AdV; SkAdV-A, Skua siadenovirus A; TAdV, turkey AdV. Bar, 0.2 nucleotide substitutions per site.
Fig. 3. Phylogenetic trees based on derived amino acid sequences of adenoviral DNA polymerase (a) and hexon (b) sequences. The inset in (b) shows the close-up of species Fowl aviadenovirus D and Fowl aviadenovirus E. Branch lengths are given in number of substitutions per site (see the scale). Bootstrap values are given in percentage for 1000 datasets; only values over 70% are shown.
The complete genome sequences of FAHV reference strains 685, SR48, SR49, 380, CR119, YR36, TR59 and 764 were obtained by Illumina sequencing. Phylogenetic and sequence analyses of the whole genomes support the division of the genus Aviadenovirus into the currently recognized species. The sequenced genomes of FAHV-D and FAHV-E members have a genome organization identical to that of an earlier sequenced FAHV-D member (strain A-2A). The data suggest a common evolutionary origin of strains SR48 and 380, and also of strains HG and 764. Complete genome sequence information on aviadenoviruses is important for taxonomy, diagnostics and pathogenicity studies.

**METHODS**

**Virus isolates.** Eight reference FAHV strains (Kawamura et al., 1964; McFerran et al., 1972) representing different types within the species FAHV-D and FAHV-E (Table 1) were propagated, after plaque purification, on confluent monolayers of chicken embryo liver cells as described previously (Marek et al., 2010b).

**DNA extraction.** Cell culture supernatants were clarified by low-speed centrifugation (10 min at 2000 g) and then ultracentrifuged (3 h at 140,000 g). The pelleted cell-free virions were used for DNA isolation (Marek et al., 2012). The presence of virus DNA in the sample was verified by PCR targeting of the hexon gene (HexA/HexB) (Meulemans et al., 2004).

**Illumina sequencing.** Whole-genome sequencing was performed by using an Illumina system (HiSeq2000, BGI, Hong Kong for 685; GAIIx, Central Service Facility NGS Unit, Vienna, Austria for SR48, SR49, 380, CR119, YR36, TR59 and 764). Paired-end libraries were generated. Multiple virus samples were sequenced in a single lane and sequence reads corresponding to the individual strains were separated by barcoding. Due to propagation of the strains in chicken cells, contamination by chicken genome reads was anticipated. Therefore, all reads were mapped initially against the available genome of Gallus gallus (v. 3.0) and the mitochondrial genome of Gallus sonneratii (AP006746.1), and only the unmapped reads were used for assembly of the viral genomes (Marek et al., 2012).

**De novo assembly.** Excess coverage can hamper de novo assembly. Therefore, we sub-sampled different numbers of reads for different strains (Marek et al., 2013). The whole-genome sequences were then assembled using CLC Genomics Workbench v. 4.0 (CLC bio). By comparison with sequences available for various complete aviadenovirus genomes and for the left and right ends of several additional FAHV genomes (Corredor et al., 2006, 2008), the resulting contigs were manually ordered and orientated (Marek et al., 2012). The contig sequences were aligned using the Accelrys Gene version 2.5 (Accelrys).

**Gap closure using PCR and Sanger sequencing.** In order to close the gaps between contigs by Sanger sequencing, PCR primers were
designed on the basis of the sequences at contig ends. Oligonucleotide primers for amplifying the sequences at one genome end were designed based on obtained sequences from the other genome end because of the symmetric nature of the inverted terminal repeat. Primer sequences are available from the authors upon request. Sanger sequencing services were provided by LGG Genomics (Berlin, Germany). The complete genome sequences for strains 685, SR48, SR49, CR119, YR36, TR59, 764 and 380 were submitted to the GenBank database and assigned to accession numbers KT862805 to KT862812, respectively (Table 1).

Annotation and phylogenetic analyses. FAdV genomes were annotated as previously described (Marek et al., 2014b). Percentage sequence identities of whole avianadivirus genome sequences were calculated using Lasergene software (DNASTAR). Three phylogenetic calculations were performed to assess the correct relationship of the examined strains: based on the complete genome, the amino acid sequence of the viral DNA polymerase, and the amino acid sequence of the hexon, the major capsid protein. The genomes were aligned using PRANK (Löytynoja & Goldman, 2010), while the protein sequences were aligned using MAFFT and the alignments were edited manually using BioEdit (Hall, 1999; Katoh & Toh, 2008). The edited alignment lengths were 106,920 nt, 1020 aa and 896 aa for the complete genome, DNA polymerase and hexon alignments, respectively. The best evolutionary model was GTR+Γ for the tree inference of complete genomes, and it was predicted using ProtTest (Darriba et al., 2011) for the protein sequences (DNA polymerase: LG+I+Γ, hexon: LG+Γ+Γ+F). Phylogenetic analyses were performed using maximum-likelihood methods within the RAxML software package (Stamatakis, 2014). Clade support was assessed using non-parametric bootstrapping with 1000 replicates, and the sequenced strains were compared to all published genome sequences of avian AdVs. Global pairwise alignments to assess sequence identities were performed using mVISTA LAGAN (Brudno et al., 2003).

ACKNOWLEDGEMENTS

The authors thank Irina Prokofieva and Evelyn Berger (both Clinic for Poultry and Fish Medicine, Department for Farm Animals and Veterinary Public Health, Vetmeduni Vienna, Vienna, Austria) for their excellent technical help. The genome analysis work was supported partially by Hungarian Research Fund grant K100163.

REFERENCES


