Genetic diversity of species Fowl aviadenovirus D and Fowl aviadenovirus E

Ana Marek,1 Győző L. Kaján,2 Carolin Kosiol,3 Mária Benkő,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 serotypes are FAdV-D and -E encompassing eight different FAdV types (Marek et al., 2010b). 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

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).

### Table 1. List of isolates used in this study

<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.

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 2. Percentage sequence identities of complete aviadenovirus genomes

|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
**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 (Grgic et al., 2011; Marek et al., 2010b, 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 (Grgic 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). 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. 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 higher than 70% are indicated. **From GenBank****.
study confirms the grouping of strain SR48 within FAdV-11, based on adequate phylogenetic and genome sequence similarities, which should be considered in future studies (Table 2, Figs 2 and 3). This reassignment is also supported by recently published neutralization assay data in which SR48 was used as reference strain (Steer et al., 2011).

Genes, inherited by all modern AdVs from their common ancestor, are located centrally in the genome and additional, niche-specific genes have accumulated in each lineage, mostly near the genome termini (Davison et al., 2003). In this study, it was shown that the terminal regions of the genome have the most variable sequences in members of avian adenovirus species as well (Figs 1 and S1). However, it is still not clear which genetic features enable a virus to cause specific disease. Recently, genomic conservation and diversity among human adenoviruses (HAdVs) were examined and the penton base, hexon and fibre ORFs and E3 regions were shown to be among the most variable in the HAdV-D genomes (Robinson et al., 2011). As their protein products mediate uptake of the virus into the target cell and/or host immune system recognition of the virus, they may be targets for selective evolutionary pressure. In the present study, hexon, fibre and ORF19 were shown to be among the most variable in the FAdV-D and FAdV-E genomes (Fig. S1). For HAdVs, the areas of greatest intraspecies diversity were different for different species. In this study, the same phenomenon could be observed even between strains belonging to different types within the same species. It would be interesting to further analyse the whole genomes of additional isolates belonging to different avian adenovirus species.

Viruses co-evolving for a long time with their host are thought to be well adapted and not markedly pathogenic. We suggested earlier that viruses in species FAdV-D and FAdV-E have been co-evolving with chickens over a long period (Marek et al., 2014b). However, FAdVs most commonly isolated from IBH cases in chickens belong to FAdV-D and FAdV-E (Kajan et al., 2013; Marek et al., 2010b; Ojkic et al., 2008). Beach et al. (2009) noted genetic differences between virulent and non-virulent turkey haemorrhagic enteritis virus isolates (a member of the genus Siadenovirus) within ORF1, E3 and the fibre protein. However, Grgić et al. (2014) did not note significant differences between fibres of virulent and apathogenic FAdV isolates, which was recently confirmed by Schachner et al. (2016). In order to estimate the influence of viral genetics on pathology, experimental infections with different molecularly manipulated isolates would be necessary.

**CONCLUSION**

The complete genome sequences of FAdV 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 Avian adenovirus into the currently recognized species. The sequenced genomes of FAdV-D and FAdV-E members have a genome organization identical to that of an earlier sequenced FAdV-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 avian adenoviruses is important for taxonomy, diagnostics and pathogenicity studies.

**METHODS**

**Virus isolates.** Eight reference FAdV strains (Kawamura et al., 1964; McFerran et al., 1972) representing different types within the species FAdV-D and FAdV-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 gallus (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 avian adenovirus genomes and for the left and right ends of several additional FAdV 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 Accelerys Gene version 2.5 (Accelerys).

**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 were available from the authors upon request. Sanger sequencing services were provided by LGC 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 avian adenovirus 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 Frank (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: LG1+Γ, hexon: LG+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 sequences 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


