SIADENOVIRUSES:
A COMPARATIVE MOLECULAR AND PHYLOGENETIC STUDY

Brief Version of the Ph.D. Thesis

Endre Kovács

Supervisor: Prof. Dr. Mária Benkő, DSc

Veterinary Medical Research Institute
Hungarian Academy of Sciences
Budapest

Biological Studies Doctorate School
Faculty of Science, Eötvös Loránd University
Head: Prof. Dr. Anna Erdei, DSc

Doctorate Program: Theoretical and Evolutionary Biology
Head: Prof. Dr. Eörs Szathmáry, DSc

2011

1 INTRODUCTION

Adenoviruses are double-stranded DNA viruses whose characteristic genome organization allowed for the elaboration of a clear-cut natural classification within the family. The central region of the genome is conserved in all adenoviruses and contains genes that are useful for phylogeny reconstructions. The flanking regions, the left and right ends of the genome, contain genes that are greatly varied in orientation and number and are homologous only within each particular genus. The analysis of these ORFs and genes usually in concert with data provided by phylogenetic trees ensures a good basis for the genus-level affiliation of any adenovirus. At an early stage, the basis for classification used to be the host origin, only the two genera *Mastadenovirus* and *Aviadenovirus* existed for adenoviruses isolated from mammals and birds, respectively. Based on molecular taxonomy, the family *Adenoviridae* is currently divided into five distinct genera. Apart from the two mentioned above, the host origin of a third genus that contains the only known fish adenovirus, *Ichtadenovirus*, is clear. In two other genera (*Atadenovirus, Siadenovirus*), viruses of the representatives of different vertebrate classes occur in a mixed manner. According to an early hypothesis, the five genera correspond to the adenoviral lineages that had codiverged with the five major classes (fish, amphibians, reptiles, birds, mammals). However, the targeted analysis of adenoviruses of lower vertebrates supported this hypothesis only partly. The first representatives of the genus *Atadenovirus* were found in ruminants, but viruses isolated from birds and marsupials were classified here as well. All adenoviruses found so far in various lizards and snakes proved to be atadenoviruses, therefore this lineage is currently considered a clade that had codiverged with squamate reptiles (Squamata).

At the outset of my Ph.D. work, the genus *Siadenovirus* contained only two viruses, and it should be noted that the genomes of both were sequenced completely. One of these, turkey adenovirus 3 (TAdV-3), which causes hemorrhagic enteritis in turkeys and marble spleen disease in chickens, was reclassified from aviadenoviruses to this genus. The other one is the formerly unclassified frog adenovirus 1 (FrAdV-1), which is the only adenovirus isolated from amphibians. The new genus was established for these two viruses because of their special genome organization and close genetic relationship. The name of this taxon refers to the sialidase-like gene present only in these two types of adenovirus. According to the early hypothesis, siadenoviruses used to be considered a lineage that had codiverged with amphibians.

2 OBJECTIVES

PCR has revolutionized virological techniques in so far as direct detection and determination of genome sequences of non-isolated viruses have become possible. By means of a consensus, nested PCR suitable to detect probably all members of the *Adenoviridae* family, numerous novel adenoviruses were found in samples of various animals in the last couple of years. The highly degenerate primers of the PCR target an adenoviral enzyme with the most conserved amino acid motifs, namely the gene of the DNA dependent DNA polymerase. Although the length of the fragment that can be aquired from the gene is only 300 nucleotides (nt), it is still appropriate for preliminary phylogenetic calculations and the genus-level classification of newly detected viruses.

In this manner was the existence of a seemingly new siadenovirus proven, which was detected in the internal organs of dead raptorial birds. Since the same virus occurred in several host species, it was officially named raptor adenovirus 1 (RAdV-1). On the basis of the short polymerase sequence it was determined to be a pathogen presumably representing a new virus species that was significantly distinct from TAdV-3 in spite of the similar pathological effects. Tissue culture isolation of the virus was unsuccessful, so we decided to make an attempt at the further characterization of the virus by means of PCR and sequencing done directly on the pathological material. In addition, the examination of further novel siadenoviruses found in the course of screening PCR investigations was also planned, as far as the conditions allowed. The nucleotide sequence of a significant genome fragment was determined from a new siadenovirus found in the sample of a Great tit (*Parus major*). This virus was temporarily named great tit adenovirus 1 (GTAdV-1). In case of the successful execution of the planned work, the genome organization considered characteristic to the members of the *Siadenovirus* genus was expected to be confirmed. Furthermore, the rise of the number of known siadenoviruses would be expected to increase the reliability of the phylogenetic calculations. The ultimate goal of the study was the clarification of the host origin of the lineage represented by siadenoviruses.

3 MATERIALS AND METHODS

3.1. Sample origin. Disease and sporadic deaths were experienced among exotic raptorial birds in two British aviaries. The following birds were involved: Harris hawk (*Parabuteo unicinctus*), Bengal eagle owl (*Bubo bengalensis*), Verreaux's eagle owl (*B. lacteus*). The extracted nucleic acid from the internal organs of these three birds constituted the basis of the present work. The other virus examined was detected in an extracted nucleic acid sample, which was formerly found negative to avian influenza virus in a nation-wide survey. Regrettably, by the time the new adenovirus was confirmed in the Great tit sample, the carcass of the bird was no longer available. No more than 40 µl of the nucleic acid remained following the survey. In order to preserve the sample containing the interesting virus, the nucleic acid solution was subjected to a non-specific amplification by means of the REPLI-g® Mini Kit (Qiagen).

3.2. Primers and PCR. At the beginning, degenerate, consensus primers applicable to the whole family were used, followed by non-degenerate, genus-specific and type-specific primers. Three types of DNA polymerases were used based on the expected size of the PCR products. These included RedTaq® DNA Polymerase (Sigma-Aldrich), Phusion™ High-Fidelity DNA Polymerase (Finnzymes), and FailSafe™ PCR System (Epicentre). Each reaction was optimized according to the enzyme and primers used, and to the size of the DNA fragment to be acquired.

3.3. Molecular cloning. Specific DNA fragments exceeding 0.8 kb in size were cloned either into pBluescript® II KS(+/-) (Stratagene) or pGEM®-T Easy (Promega) vectors, depending on the DNA polymerase used in the PCR.

3.4. Poly-A tailing. The exact determination of the inverted terminal repeat sequences (ITR) was performed by means of the 5'/3' RACE Kit (Roche) using a modified protocol.

3.5. Sequencing. Sequencing was carried out with the BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems). DNA fragments were sequenced on both strands directly with the PCR primers, then the strategy of primer walking was applied.

3.6. Sequence analysis. Raw sequence data were viewed with the BioEdit Sequence Alignment Editor v5.0.7 and/or the free software FinchTV v1.4. The identity of each fragment was verified by BlastX searches. The sequence files were processed and assembled by the Staden package using the programs Pregap4 v1.5 and Gap4 v4.10. The genome was annotated by Artemis Release 10. Splice donor and acceptor sites were located manually in the vicinity of regions corresponding to the splice sites in TAdV-3 and FrAdV-1. SimPlot was used to demonstrate the distance across the complete siadenoviral genomes. The assembled genome sequence was translated in all the six reading frames by means of the JavaScript DNA Translator v1.1.

3.10. Phylogenetic analysis. Multiple amino acid alignments were made either by MultAlin v5.4.1 or by ClustalX v2.0. Maximum likelihood analysis was carried out through the Mobyle portal using PhyML. Bayesian inference was computed by MrBayes v3.1.2.

4 RESULTS

4.1. The complete sequence of the first non-isolated adenovirus was determined. The RAdV-1 genome represents the first non-isolated adenovirus and the third completely sequenced siadenovirus. The genome is 26,284 nt in length, its G+C content is 38.5%, the ITR is 29 nt long. Out of the 25 ORFs found in the genome, 18 are conserved in all adenoviruses, and 6 are specific to the *Siadenovirus* genus. One further ORF (ORF9) is presumed to be active.

Three ORFs are present in the hypothetical E1 transcription unit. In the leftmost position is the putative sialidase. In its encoded protein, three repeats of the aspartic acid (D) box motif of bacterial sialidases (S/T-X-D-[X]-G-X-T-W/F) are discernible. ORF4, originally described in the genome of TAdV-3, and the so-called hydrophobic protein are both present as opposed to the genome of FrAdV-1, where no homolog of ORF4 could be identified.

Between the genes of pVIII and fiber, a homolog of the E3 gene of TAdV-3, as well as the U exon were identified.

Close to the right end of the genome, three ORFs were found. ORF7 and 8 have their counterparts in TAdV-3 and FrAdV-1. However, the encoded product of ORF9, apparently specific to RAdV-1, consists of 146 amino acids, but no known protein is homologous to this product. The complete genome sequence of RAdV-1 is retrievable from GenBank under the accession no. EU715130.

4.2. Approximately half of the genome sequence of another non-isolated adenovirus, GTAdV-1, was determined. A total of 13,628 nt was acquired from the central, conserved genome part of GTAdV-1, the G+C content of which is 37.5%. This genome fragment contains 8 full (DNA polymerase, pTP, 52K, pIIIa, III, pVII, pX, pVI) and 2 partial (IVa2 and hexon). Due to the low concentration of viral DNA in the sample, out of about 50 attempts only 6 PCRs yielded specific fragments. The partial genome sequence of GTAdV-1 was deposited in GenBank under the accession no. FJ849795.

4.4. The two newly characterized members of the Siadenovirus genus occupy a common branch with TAdV-3 in the phylogenetic tree, separate from FrAdV-1. The siadenovirus clade represents a distinct, relatively ancient offshoot in the adenovirus family tree. Phylogenetic calculations imply that RAdV-1 and GTAdV-1 share a recent common ancestor with TAdV-3. FrAdV-1 is consistenty placed on a separate branch with another recently described, novel siadenovirus, which caused mass mortality among Sulawesi tortoises (*Indotestudo forstenii*).

4.5. The increasing number of siadenoviruses found in birds along with a high A+T content of the DNA in each member of the genus challenges the hypothesis of siadenoviruses having originated from amphibian hosts.

5 DISCUSSION

5.1. Current state of Siadenovirus.

TAdV-3 for years had been considered an avian AdV out of the ordinary. When its genome sequence became available, however, the differences from the prototype avian AdV, fowl AdV-1, were greater than expected. The genome of the latter was significantly longer (~45 kb), whereas that of TAdV-3 was as much shorter (~26 kb), than the genome size characteristic to mastadenoviruses (~36 kb) considered universal at an early stage. A couple of years later the genome sequence of FrAdV-1 showed that TAdV-3 was not a unique, very special virus, but the two constituted a distinct group, phylogenetically separate from aviadenoviruses. In the meantime, another completely sequenced, special avian adenovirus, duck adenovirus 1, turned out to possess a fourth type of genome organization, which on the

basis of phylogenetic calculations belonged to a fourth clade, the newly established *Atadenovirus* genus.

The new genome sequence, that of RAdV-1, entirely consolidated the genome features characteristic to siadenoviruses. The genomes of these AdVs probably represent the shortest, minimal AdV genome sufficient for propagation.

In the last couple of years, animal-related adenovirus detection—by PCR and DNA sequencing—seems to have been gathering pace, and with that numerous novel adenoviruses, as well as several novel siadenoviruses have been reported. While, despite targeted attempts, no new adenoviruses were found in amphibians, siadenoviruses were detected in the representatives of several bird species. Besides RAdV-1 and GTAdV-1, short sequences of the genes DNA polymerase and hexon from the siadenoviruses of Plum-headed parakeet (*Psittacula cyanocephala*) and Budgerigar (*Melopsittacus undulatus*) were reported. Surprisingly, the gene sequence of a new siadenovirus was acquired from Sulawesi tortoises that had suffered a mass mortality following an illegal import. The virus was described as the etiological agent of the disease. This is the first reptilian adenovirus that occurred in reptiles outside of the order Squamata. Since only short gene sequences are available from these newly detected siadenoviruses, the analysis of RAdV-1 and GTAdV-1 represents a significant step in the confirmation of the characteristic genome organization of siadenoviruses.

5.2. Comparative genome analysis. With an approximate 26 kb, their genome is significantly shorter than that of a typical aviadenovirus. With respect to the genes encoded by the E2A, E2B, and L transcription units, representatives of the two genera show little difference. Differences between the two genome types are most manifest in the gene composition of the flanking regions. Especially the number of ORFs in place of the mastadenoviral transcriptional units E1 and E4 is particularly high in aviadenoviruses with 9 and 11, respectively, whereas it is extremely low (2–3) in the corresponding region of siadenoviruses. Aviadenoviruses (atadenoviruses likewise) completely lack the conventional E3 unit, but siadenoviruses consistently feature a single ORF of unknown function in this region between pVIII and fiber. The genes of two structural proteins (V and IX) present only in mastadenoviruses are expectedly absent from these two siadenoviruses.

ORF4 is a short putative gene coding for ~100 aa, located immediately downstream of sialidase. This ORF is absent from FrAdV-1. However, another shorter ORF, the putative gene of a hydrophobic protein, was detected in all three siadenoviruses a couple of nt downstream in a different reading frame, largely in overlap with ORF4. Similarities between

the putative hydrophobic protein of the three siadenoviruses are rather superficial. Accordingly, the "E1" region of FrAdV-1 is constituted from 2 possible ORFs, whereas that of TAdV-3 and RAdV-1 contains three ORFs. Further experiments are required in order to find out whether these ORFs, along with ORF9 newly found in RAdV-1, are functional as genes. It should be noted that the *in vitro* propagation in monolayered cell cultures of both isolated siadenoviruses is cumbersome. FrAdV-1, isolated from Leopard frog (*Rana pipiens*), was propagated in a cell line (TH-1) derived from the heart tissue of the Common box turtle (*Terrapene carolina*). TAdV-3 replicates only in a turkey-derived lymphoblastoid cell line transformed with the Marek's disease virus. At any rate, in the course of a preliminary study, mRNAs transcribed from the sialidase gene during the early phase of infection were detected in cells infected by TAdV-3.

A possible explanation for the existence of sialidases in siadenoviruses can be theorized as a gene capture event through horizontal gene transfer before the last common ancestor of siadenoviruses started to split, therefore they can be considered xenologs of the bacterial sialidases. Although the function of the adenoviral sialidase is unknown, it may be relevant to note that certain human adenovirus types such as HAdV-8, 19 and 37, causative agents of epidemic keratoconjunctivitis, use sialic acid residues instead of the coxsackie-adenovirus receptor as a port of entry.

For the grouping of adenoviruses, the nucleotide composition of the genome, that is the proportion the bases G+C and A+T, has been used for quite some time. In members of the three genera (*Mast-, Avi-* and *Ichtadenovirus*) whose host origin is considered to be clarified, the proportion is usually balanced (45–55%) or slightly biased in favour of G+C. Contrarily, a strikingly high (>60%) A+T content was discernible in the initially characterized atadenoviruses (isolated from ruminants, birds and a marsupial). The name of the genus testifies of the early recognition of this fact. However, adenoviral sequences from lizards and snakes possessed a surprisingly balanced base composition, albeit phylogenetic reconstructions clearly suggested their inclusion in a common clade with atadenoviruses. In accordance with an early hypothesis, increase in the genomic A+T content may be linked to host switches. In other virus families, certain amount of empirical evidence is available in favour of this hypothesis. The A+T content in the sequence of every siadenovirus so far known is above 60%.

5.3. Phylogeny. Currently, five distinct clades (the number of which is likely to rise in the future), all representing separate genera—*Mastadenovirus, Aviadenovirus, Atadenovirus,*

Siadenovirus and *Ichtadenovirus*—, are discernible within the family *Adenoviridae*. In phylogenetic reconstructions, the genus *Siadenovirus* is consistently placed near the base of the tree. It was originally thought of as the lineage that had cospeciated with amphibians, but the results of the investigation aimed at the verification of this hypothesis seem to act as a refutation. In spite of targeted attempts, no other frog AdV has yet been found. On the other hand, the number of siadenoviruses identified in birds has increased significantly from one to at least four, thus predominating in number. This fact together with the most recently demonstrated siadenovirus in Sulawesi tortoises implies that amphibians may not represent the original hosts of siadenoviruses but perhaps another reptilian order different from *Squamata*. Although birds are considered a separate class, but in actual fact, reptiles cannot be regarded monophyletic without the inclusion of birds. Adenoviruses of birds (*Aviadenovirus*) constitute a distinct clade, so it would not be surprising if the rest of the reptilian orders had their own characteristic adenovirus group. Most recently in the United States and in Hungary at almost the same time, adenoviral sequences from various sliders and box turtles were acquired, which proved to be distinct from those in the other five genera. The at least three new types of turtle adenoviruses form a sixth clade (and probably a new genus).

Thus, according to current knowledge, at least four AdV lineages are present in reptiles: atadenoviruses in lizards and snakes (Squamata), aviadenoviruses in birds, siadenoviruses and members of a new group in various turtles. On the other hand, birds harbour representatives of three adenoviral genera. It appears that in case of certain adenoviruses, host switches are not uncommon between hosts species of close evolutionary relationship. Notably, adenoviral infection in worm lizards (Amphisbaenia) and in tuatara (Rhynchocephalia) has not been reported yet. Adenoviral particles observed by light and electron microscopy have been reported in crocodiles, but no sequence data are available. Thus this ancient reptilian lineage cannot be excluded as the original hosts of siadenoviruses or another AdV genus yet to be discovered.

6 PUBLICATIONS

6.1. Publications relevant to the thesis

Kovács, E.R., Benkő, M. (2011). Complete sequence of raptor adenovirus 1 confirms the characteristic genome organization of siadenoviruses. *Infect. Genet. Evol.* 11, 1058–1065. (IF: 3.223)

Kovács, E.R., Jánoska, M., Dán, Á., Harrach, B., Benkő, M. (2010). Recognition and partial genome characterization by non-specific DNA amplification and PCR of a new siadenovirus species in a sample originating from *Parus major*, a great tit. *J. Virol. Methods* 163, 262–268. (IF: 2.077)

Kovács, E.R., Benkő, M. (2009). Confirmation of a novel siadenovirus species detected in raptors: Partial sequence and phylogenetic analysis. *Virus Res.* 140, 64–70. (IF: 2.429)

6.3. Congress abstracts

Kovács, E.R., Harrach, B., Benkő, M. Complete sequence and genetic features of raptor adenovirus 1: a novel, non-isolated species in the genus *Siadenovirus*. *ESVV 8th International Congress of Veterinary Virology*, Budapest, Hungary, 23–26 August, 2009.

Kovács, E.R., Harrach, B., Benkő, M. Genome analysis of raptor adenovirus 1: a novel, non-isolated type, first member of a proposed new species in the genus *Siadenovirus*. *9th International Adenovirus Meeting*, Dobogókő, Hungary, 26–30 April, 2009.

Kovács, E.R., Benkő, M. Two novel siadenoviruses: molecular and phylogenetic analysis. *Adenoviruses: Basic Biology to Gene Therapy*, Zadar, Croatia, 23–24 September, 2008.

Kovács, E.R., Benkő, M. Two novel siadenoviruses: molecular and phylogenetic analysis. *14th International Congress of Virology*, Istanbul, Turkey, 10–15 August, 2008.

Kovács, E.R., Zsivanovits, P., Benkő, M. Genome sampling of a novel siadenovirus suspected to cause fatalities among raptors. *8th International Adenovirus Meeting*, Zürich, Switzerland, 30 August–02 September, 2006.

Kovács, E.R. Biodiversity of animal adenoviruses. *DNA Tumor Viruses 2006 Meeting*, Salk Institute for Biological Studies, La Jolla, CA, USA, 11–16 July, 2006.