Screening for and comparative molecular characterization of adenoviruses in South American and European wildlife to reveal geographic and evolutionary differences

Project closing report (NN107632)

We aimed at the recognition and characterization of novel adenoviruses (AdVs) in animals of South America and Europe. We planned to compare them to each other and to other AdVs detected in wild vertebrates. The reports on wild animals are unfortunately still very sparse even in the better studied regions. But we need data on the AdVs of exotic and diverse vertebrate lineages to possibly understand the evolutionary past of the whole virus family (considering both a general coevolution and occasional host switches).

In the first year, we shared our methods with the Brazilian collaborators for the non-invasive sampling of animals, the extraction of the DNA from such samples, the pan-adenovirus PCR and DNA sequencing of the amplicons, and the comparative bioinformatics analyses of the gained sequences. A large number of wild bird samples could be collected mainly in the Amazon basin (in the Manaus Reserve), and 4 novel atadenoviruses, 1 siadenovirus and 3 aviadenoviruses had been identified. The 4 atadenoviruses were from blue-and-yellow macaw (*Ara ararauna*, Psittaciformes), keel-billed toucan (*Ramphastos sulfuratus*, Piciformes), house wren (*Troglodytes aedon*, Passeriformes) and from a species of owl order Pulsatrix (*Pulsatrix* sp., Strigiformes). The siadenovirus originated from white-throated toucan (*Ramphastus tucanus*, Piciformes). Finally, the 3 aviadenoviruses identified by sequencing were from great kiskadee (*Pitangus sulphuratus*), a passerine bird, slaty-breasted wood rail (*Aramides saracura*, Gruiformes) and from a *Tigrisoma* sp. (a genus of heron, order Pelecaniformes).

However, in the second year, the Brazilian source supply started to decrease. Thus we turned also to zoo samples. While theoretically the wild living animals may have different viruses than the captive ones, we and other scientists see and suppose everyday host switches to be very rare. Indeed, this approach helped identifying several AdVs occurring in animals with South American origin. From zoo birds, we found a novel siadenovirus (burrowing parrot or Patagonian conure (*Cyanoliseus patagonus*) and a novel atadenovirus (from monk parakeet (*Myiopsitta monachus*). These studies confirm the earlier results that Passeriformes and Psittaciformes host atadenoviruses in larger numbers, but members of some other avian orders may be infected by atadenoviruses as well. The zoo approach to collect samples from South American species was extended to German, Austrian and further Hungarian zoos. It was needed also because e.g. the few Xenarthra species in the Budapest zoo were found to be negative, but we still could hope to find AdVs in samples from other zoos. Unfortunately, the Xenarthra samples proved to be negative both from the resampled Hungarian species and those from foreign zoos.

We started also several new collaborations to collect original Brazilian samples from wild life (resulting mainly samples from wild mammals) and examined bats from Mexico, too. From crab-eating fox (*Cerdocyon thous*), also known as the forest fox, a medium-sized canid endemic to the central part of South America, we amplified fowl adenovirus 8b. As this was a fecal sample and aviadenoviruses never infect mammals, this virus must originate from a prey animal. This assumption seemed to be further confirmed by finding also other viruses in the feces of the same animal, namely a circovirus with 97% nucleic acid identity to porcine circovirus 2. Also another crab-eating fox feces from this collection was found to contain a circovirus with a 95% nucleic acid identity with equine-associated cyclovirus 1. So it seems one fox ate pig, the other horse meat or remaining. These cases seem to prove that these Brazilian carnivores most probably steal and eat garbage of human populations.

From a white headed marmoset a novel simian mastadenovirus was amplified, which has only 78% identity with the most similar AdV, namely porcine AdV-5. Thus this virus is novel and may be the own, co-speciated AdV of this New World monkey (NWM) (or originate from some unknown mammals being a prey).

Another Brazilian white headed marmoset contained a novel atadenovirus. Atadenoviruses are present in sqamate reptiles, birds and ruminants. The original hosts of atadenoviruses supposed to be the scaled reptiles (Squamata). The atadenoviruses found in birds or ruminants because of host switch have a low percent of G+C content probably due to the accommodation process, while the atadenoviruses of reptiles have a balanced, ~50% G+C content. As the DNA sequence of this novel AdV has a balanced G+C content, this AdV is supposed to originate from a scaled reptile, i.e., from a reptile prey of this marmoset.

A common mouse opossum (a marsupial) turned out to have an AdV in its feces, which is 100% identical with the turkey AdV-5, a typical aviadenovirus. Aviadenoviruses are present only in birds, so this AdV is supposed to be the virus of prey (a bird or egg or garbage). These cases highlight the need for constant precautions in interpreting molecular findings.

Earlier, only a single frog AdV was known, a siadenovirus, thus we put considerable efforts to screen amphibian samples especially from the species living in South America. Finally, we could find a second frog AdV, actually even three variants of it (differing in 3 amino acids in the PCR amplified very short partial polymerase fragment) in green and black poison dart frog (*Dendrobates auratus*), which are native to the northwestern parts of South America and to Central America, and Golfoducean poison dart frog (*Phyllobates vittatus*), which lives in Costa Rica. To our big surprise, these viruses turned out to be atadenoviruses. Thus it is clear, that these novel AdVs could not have coevolved with these hosts but are more likely switched to the frogs from a yet unknown host, most probably a lizard or snake. These emerged strains seem to be very pathogenic for frogs (mass mortalities were seen in some pet shops) and capable of crossing the species barrier between these two frog species. We successfully sequenced longer parts of the viral genomes between the genes of the IVa2 and the protease, the conserved genome part in AdVs (made possible by designing consensus primers based on alignments). However, all other amphibian samples studied (including several frogs closely related to the positive species) were negative. After so many screening of frog samples, we have to suppose that there is no real frog AdV cospeciated with this or other amphibian hosts.

From rodents, we could amplify AdV gene fragments from several species with South American origin. In fecal sample of a degu or brush-tailed rat (*Octodon degus*), we detected a novel AdV. The degu is endemic to the "matorral" ecoregion of central Chile. We obtained both a DNA polymerase and a hexon gene fragment. We detected novel AdVs also in capybara (*Hydrochoerus hydrochaeris*), a large rodent native to South America, in guinea pig (*Cavia porcellus*), which originated in the Andes, and in a common agouti, which are native to Middle America, northern and central South America. Parallel, we amplified and partially characterized novel AdVs from Eurasian rodents: wood mouse (*Apodemus sylvaticus*), stripped field mouse (*Apodemus agrarius*), European hamster (*Cricetus cricetus*), golden hamster (*Mesocricetus auratus*), Djungarian hamster or Siberian dwarf hamster (*Phodopus sungorus*), common vole (*Microtus arvalis*), acatia rat (*Thallomys paedulcus*) and Mongolian gerbil (*Meriones unguiculatus*). We amplified and sequenced a partial gene also from the AdV earlier isolated from deer mouse (*Peromyscus maniculatus*), native to North America. According to our results, the relatedness of viruses is primarily influenced by the evolutionary distance of their hosts and not by the (present) geographical distribution of the hosts.

New World monkeys are primates found exclusively in Central and South America and in some parts of Mexico in the wild. On the evolutionary level, they are very distant from the

Old World monkeys (OWMs), apes and man, so we hypothesized that they may have characteristically different AdVs. Very few such AdVs were described, and so far only one, a titi monkey AdV, has been fully sequenced. We have detected several novel AdVs in NWMs by PCR. The OWM and NWM AdVs proved to be well separated on the phylogenetic tree, and even more the novel prosimian AdVs identified by us. However, gaining longer genome fragments from the NWM AdVs proved to be a challenge except in one case. The sequencing of the AdV found in a red-handed tamarin yielded a long genome sequence from the IVa2 gene to the U exon. It means that also the entire E3 region was determined, which is the most variable part of primate AdV genomes. While all OWM AdVs share practically the same organization in their E3 region, the new genome sequence from the NWM AdV showed three unique E3 genes (in place of the first three OWM E3 genes; the latter ones occurring also in the ape and human AdVs). The three novel genes are different also from the titi monkey AdV E3 genes in this genome region (which are also unique). This was a surprising finding that even the first two NWM AdVs sequenced are so divergent. For comparison, we sequenced the full genomes of all serotyped OWM AdVs that had not been sequenced: simian adenovirus 2 (SAdV-2), -4, -5, -8, -9, -10, -11, -12, -13, -14, -15, -16, -17, -19 either by classical Sanger sequencing or by next generation (Illumina) sequencing after propagating them on Vero cells. Thus we have a good base for comparison with the NWM AdVs. While there were characteristic differences in number of fiber genes (these helps in establishing species demarcations), the genomes showed a practically identical gene organization. Even the E3 regions showed very similar organization in OWM AdVs with only occasional gene deletions or fusions. Based on the phylogenetic distances, number of fiber genes, G+C content, host, we grouped them and proposed to establish officially six new species for them (*Simian mastadenovirus B* to *G*). The International Committee on Taxonomy of Viruses (ICTV) is supposed to accept this proposal in the first weeks of March (by a final electronic balloting).

Beside Xenarthra (occurring only in South America), Afrotheria comprises a sister taxa at the base of the placental mammal radiation. While we could not get sample from the Brazilian manatee, the only Afrotheria species living there, we succeeded in identifying the very first Afrotheria AdV ever. It is from a short-eared elephant shrew (*Macroscelides proboscideus*), which is a relative of the elephants in spite of looking like a small rodent. We could amplify four different gene fragments from this AdV (IVa2, DNA polymerase, hexon and pVIII; and we also connected the IVa2 and DNA polymerase gene fragments). However, the phylogenetic tree showed a surprising placement for this Afrotheria. It was supposed to be the most ancient branch of the mastadenoviruses, but did not situate so deep. Thus we made special effort to get a second elephant shrew sample from another zoo. We could get sample from Vienna, and we indeed could amplify again an AdV. It turned out to be an identical virus thus confirming our original finding. We have no explanation yet why it does not reflect the more ancient place of Afrotheria. Perhaps this virus got to this new host by host switching from a more modern unknown mammal.

In the last year we got a request to help to diagnose a case: king penguin chicks were dying in Southern Chile. We could detect and partially characterize a siadenovirus in the sample of these South American/Antarctic birds. The found virus is quite similar to but still a distinct type from those AdVs identified earlier in gentoo and chinstrap penguins. As siadenoviruses are generally more pathogenic than aviadenoviruses, the own viruses of birds, this virus may be the reason of the dying, however, not all the samples contained it, thus a screening for further pathogens is still going on.

Altogether, we identified a relatively large number of novel AdVs, and concluded that the present geographical differences are not crucial for the differences among the AdVs, e.g. in the case of birds. Coevolution may be more important in causing the seen phylogenetic

differences as exampled well by OWM vs NWM AdVs. On the other hand, the very deep evolutionary differentiations cannot be always well detected.

To get deeper knowledge about proteins of the AdVs, we expressed the fiber gene of several interesting AdVs by bacterial expression system and, with the help of Spanish collaborators, produced adequate amount of pure fiber protein, crystallized it and established the three dimensional structure by X-ray crystallography. We could establish the 3D structure of a siadenovirus the first time (both of the very pathogenic turkey haemorrhagic enteritis virus, i.e., turkey AdV-3, and its non-pathogenic variant, a vaccine strain with some minor differences in the fiber) and of an atadenovirus (the bovine AdV-4 recognized as the first type of the novel lineage of BAdVs and described by Hungarian scientists, Bartha and coworkers). We expressed also the fiber knob of goose AdV-4 (sequenced earlier by us) and its structure was solved, too. The establishing of fiber structures helps to understand the cellular receptor usage of the studied AdVs. In case of TAdV-3, *in vitro*, sialyllactose was identified as a ligand by glycan microarray analysis, nuclear magnetic resonance spectroscopy, and crystallography. We propose that, *in vivo*, the TAdV-3 fiber may bind a sialic acid-containing cell surface component.

As we found several atadenoviruses during this work, we were interested also in the function of a protein that occurs only in atadenoviruses and mastadenoviruses: the 34K protein. Furthermore, it occurs always in duplicates in the atadenoviruses what we proved by sequencing several reptilian atadenovirus genomes (both from snakes and lizards). We expressed several atadenovirus and mastadenovirus 34K genes and studied their function with the help of Canadian scientists. We could report that all adenoviruses known to encode such proteins maintain the potential to form an E3 ubiquitin ligase complex to ubiquitinate specific substrates, leading to their degradation by the proteasome. This decides a long debate regarding the LH3 protein of atadenovirus, a homologue of the mastadenovirus E1B 55K, if is a structural protein or part of such cell manipulating protein complex. In atadenoviruses it is both a structural protein (cementing the virion structure instead of the missing protein IX) AND part of the ligase complex. (In mastadenoviruses, it has only the later function.) We could accurately predict Cullin usage for the 34K of mastadenoviruses and all but one atadenovirus. Interestingly, in nonhuman primate adenoviruses, we found a clear segregation of Cullin binding, with Cul5 utilized by viruses infecting great apes and Cul2 by Old/New World monkey viruses, suggesting that a switch from Cul2 to Cul5 binding occurred during the period when great apes diverged from monkeys.

While we could amplify so many novel AdVs and some of them have been fully sequenced, we had to face repeatedly the old dilemma, if the existence of a virus can be accepted if only molecular data are available. After a prolonged discussion and a special meeting devoted to this problem finally, the Executive Committee of ICTV (I am being one of the members of the EC, as chair of the Animal DNA Viruses and Retroviruses Subcommittee) officially decided last year that even these viruses can get official taxonomical classification. This will clarify not only the problems of our research group but also of those hundreds of virologists who are faced with thousands of novel viruses characterized only on molecular level (e.g. found by metagenomics) but not yet isolated. We just cannot ignore them anymore. A consensus statement was published about this approach and decision by us in the Nature Reviews Microbiology (a 24,727 IF journal).