Journal of General Virology (2010), **91**, 2601–2609 **DOI** 10.1099/vir.0.020032-0

Prevalence and genetic diversity of adeno-associated viruses in bats from China

Yan Li,[1] Xingyi Ge,[1] Chung-Chau Hon,[2] Huajun Zhang,[1] Peng Zhou,[1] Yunzhi Zhang,[1,3] Yi Wu,[4] Lin-Fa Wang[4] and Zhengli Shi[1]

Correspondence
Zhengli Shi
zlshi@wh.iov.cn

[1]State Key Laboratory of Virology, Wuhan Institute of Virology, Chinese Academy of Sciences (CAS), Wuhan, PR China

[2]Unité Biologie Cellulaire du Parasitisme, INSERM U389, France

[3]Yunnan Institute of Endemic Diseases Control and Prevention, Dali, PR China

[4]CSIRO Livestock Industries, Australian Animal Health Laboratory, Geelong, Australia

Bats are increasingly being recognized as important natural reservoirs of different viruses. Adeno-associated viruses (AAVs) are widely distributed in primates and their distribution in bats is unknown. In this study, a total of 370 faecal swab samples from 19 bat species were collected from various provinces of China and examined for the presence of AAVs. The mean prevalence rate was 22.4 % (83 positives out of 370 samples), ranging from 10 to 38.9 % among different bat species. The genome sequence spanning the entire *rep–cap* ORFs was determined from one chosen AAV-positive sample (designated BtAAV-YNM). Phylogenetic analysis of the entire *rep–cap* ORF coding sequences suggested that BtAAV-YNM is relatively distant to known primate AAVs, but phylogenetically closer to porcine AAV strain Po3. Further analysis of the partial *cap* ORF sequences of bat AAV samples (n=49) revealed a remarkably large genetic diversity, with an average pairwise nucleotide identity of only 84.3 %. Co-presence of multiple distinctive genotypes of bat AAV within an individual sample was also observed. These results demonstrated that diverse AAVs might be widely distributed in bat populations.

Received 10 January 2010
Accepted 16 June 2010

INTRODUCTION

Adeno-associated viruses (AAVs) were discovered in 1965 as a contaminant of adenovirus preparations (Atchison *et al.*, 1965). Since productive infection of AAV was initially found to happen only with co-infection of an adenovirus or a herpesvirus (Bauer & Monreal, 1986; Goncalves, 2005), AAVs are therefore ascribed to a separate genus, named *Dependovirus*, within the family *Parvoviridae* (Muzyczka & Berns, 2001). However, it was later discovered that AAVs do not always require a helper co-infection for a productive infection and cellular genotoxic stress can also activate the efficient replication of AAVs (Schlehofer *et al.*, 1983; Yakobson *et al.*, 1987; Yalkinoglu *et al.*, 1988). Although AAV may elicit mild immune responses, the virus is not currently known to cause disease in humans. AAV can infect both dividing and non-dividing cells and may incorporate its genome into that of the host cell. These features make AAV an attractive candidate for usage as viral vectors for gene therapy (Grieger & Samulski, 2005).

The non-enveloped virion of AAV is approximately 25 nm in diameter and encapsidates a ssDNA genome of about 4.7 kb that contains two large ORFs, named *rep* and *cap*, flanked by inverted terminal repeats (ITRs). The ITRs are the only *cis*-acting elements required for genome replication and packaging. The *rep* ORF is composed of four overlapping genes encoding replication proteins required for viral life cycle, while the *cap* ORF contains three overlapping genes encoding the viral capsid proteins that are assembled into icosahedral virion shells (Wu *et al.*, 2006).

At present, 13 primate AAVs designated AAV1–AAV13 have been identified from human or non-human primates (Atchison *et al.*, 1965; Bantel-Schaal & zur Hausen, 1984; Chiorini *et al.*, 1997, 1999; Gao *et al.*, 2002, 2003, 2004; Melnick *et al.*, 1965; Mori *et al.*, 2004; Schmidt *et al.*, 2008; Schnepp *et al.*, 2005, 2009). In addition to primates, a number of phylogenetically distinct AAVs have also been identified from other animals such as cow (Schmidt *et al.*, 2004), goat (Arbetman *et al.*, 2005), pig (Bello *et al.*, 2009), chicken (Bossis & Chiorini, 2003; Hess *et al.*, 1995), waterfowl (Brown *et al.*, 1995) and snake (Farkas *et al.*, 2004).

The GenBank/EMBL/DDBJ accession numbers for sequences reported in this paper are GU226879–GU226916, GU226971, HQ142870–HQ142879.

Supplementary figures are available with the online version of this paper.

Bats are one of the most diverse groups of mammals on earth (Altringham *et al.*, 1996) and are natural reservoirs of a number of emerging viruses, including henipaviruses (Eaton *et al.*, 2005), severe acute respiratory syndrome coronavirus (SARS-CoV) (Lau *et al.*, 2005; Li *et al.*, 2005), Ebola virus (Leroy *et al.*, 2005) and Marburg virus (Towner *et al.*, 2007). Although DNA viruses have been discovered in bats, the majority of viruses identified in bats are RNA viruses (Wong *et al.*, 2007). Recently, a novel bats adenovirus was isolated in a primary cell culture derived from kidney tissues of a fruit bat, *Pteropus dasymallus yayeyamae* (Maeda *et al.*, 2008). In addition, PCR screening has revealed the presence of seven genetically different gammaherpesviruses and one novel betaherpesvirus in seven different European bat species in the family *Vespertilionidae* (Wibbelt *et al.*, 2007). Independently, another study using PCR analysis has identified a novel gammaherpesvirus in a Serotine bat (*Eptesicus serotinus*) in Hungary during an investigation of a sick bat (Molnar *et al.*, 2008).

During a recent surveillance study in different regions of China, we have discovered a number of adenoviruses, which were found in a wide host range of bat species with remarkable genetic diversity (Li *et al.*, 2010). Here, we report the discovery and genetic diversity of AAVs among different bat species in China. Although our sampling size and location were not exhaustive, the data gathered in this study was sufficient to demonstrate the high prevalence and large genetic diversity of AAVs in the bat populations in China.

RESULTS

Identification and prevalence of AAVs in bats

From 2007 to 2008, a total of 370 faecal swab samples were collected from 19 bat species in five provinces in China and screened for the presence of AAV using a PCR primer pair targeting the conserved region of known AAV genomes (expected amplicon size, 447 bp). Eighty-three samples were found to be positive (83/370, 22.4 %) and they are distributed among 10 insectivorous bat species with a prevalence rate ranging from 10 to 38.9 % (Table 1). Possibly

Table 1. Prevalence of AAVs among 19 bat species in five provinces of China

Bat species	Sampling location					
	Tianjin	Hubei	Guangdong	Yunnan	Hainan	All locations
	No. positives/no. tested (%)					
Rousettus leschenaulti	–	–	–	–	0/1	0/1
Eptesicus serotinus	–	0/1	–	–	–	0/1
Rhinolophus pusillus	–	–	–	–	0/1	0/1
Rhinolophus affinis	–	–	3/11 (27.3)	2/4 (50.0)	9/45 (20.0)	14/60 (23.3)
Rhinolophus luctus	–	–	–	–	0/1	0/1
Rhinolophus sinicus	–	7/28 (25.0)	4/11 (36.4)	–	1/7 (14.3)	12/46 (26.1)
Rhinolophus pearsoni	–	0/1	–	1/9 (11.1)	–	1/10 (10.0)
Rhinolophus macrotis	–	0/3	–	1/1 (100.0)	–	1/4 (25.0)
Hipposideros armiger	–	1/8 (12.5)	–	4/16 (0.25)	0/12	5/36 (13.9)
Hipposideros larvatus	–	–	3/13 (23.1)	–	–	3/13 (23.1)
Hipposideros pomona	–	–	0/9	–	0/18	0/27
Myotis daubentoni	–	–	–	7/18 (38.9)	–	7/18 (38.9)
Myotis horsfieldii	–	–	–	–	0/8	0/8
Myotis ricketti	2/36 (5.56)	5/21 (23.8)	–	21/33 (63.6)	–	28/90 (31.1)
Myotis davidii	–	0/2	–	–	–	0/2
Myotis chinensis	–	0/3	–	–	–	0/3
Myotis spp.	–	–	–	0/13	–	0/13
Scotophilus kuhlii	–	–	9/26 (34.6)	–	–	9/26 (34.6)
Minipopterus schreibersii	–	–	–	3/10 (30.0)	–	3/10 (30.0)
All species	2/36 (5.56)	13/67 (19.4)	19/70 (27.1)	39/104(37.5)	10/93(10.7)	83/370(22.4)

due to the limited sampling size ($n<10$), no AAV was detected in *Rousettus leschenaultia*, *E. serotinus*, *Rhinolophus pusillus*, *Rhinolophus luctus*, *Myotis davidii* or *Myotis chinensis*. It is also noteworthy that all 27 samples collected from *Hipposideros pomona* were negative, while the other two closely related *Hipposideros* species, *Hipposideros armiger* and *Hipposideros larvatus* have prevalence rates of 13.9 ($n=36$) and 23.1 % ($n=13$), respectively. These results suggest that bat AAVs are widely distributed in different bat populations in China.

Genome sequence analysis of a representative bat AAV collected from *Myotis ricketti*

The genome sequence (spanning the entire *rep–cap* ORFs) of one bat AAV-positive sample collected from *M. ricketti* in Yunnan (YN) province was characterized. This bat AAV strain was designated BtAAV-YNM (GenBank accession no. GU226971). Its putative ORFs of *rep* and *cap* are predicted to be 1839 and 2172 nt in length, respectively (Fig. 1). The intergenic region of BtAAV-YNM is 17 nt in length, which appears to be slightly closer to the major mammalian AAVs (Fig. 1). The deduced Rep and Cap proteins of BtAAV-YNM share an average of 52.7 % (ranging from 51.2 to 53.5 %) and 60.2 % (ranging from 56.5 to 63.2 %) pairwise amino acid identity, respectively, to their homologues of mammalian AAVs, including AAV1–13 (except for AAV9 whose full-length ORFs are unavailable), bovine AAV and goat AAV (Supplementary Fig. S1, available in JGV Online). A plot of similarity (i.e. Simplot) across the full genomes of AAVs suggests that BtAAV-YNM is not likely to originate from any recent recombination events of known AAVs (Supplementary Fig. S2, available in JGV Online). These results suggest the genome of BtAAV-YNM is considerably divergent from the reference genomes of currently known AAVs.

Phylogenetic origin of bat AAVs

We constructed a large-scale maximum-likelihood (ML) phylogeny for all relevant sequences from members of the genus *Dependovirus* obtained from GenBank ($n=297$; both full-length and partial genomes) and all bat AAV sequences derived from this study ($n=49$; will be described later). This large-scale ML phylogeny suggests bat AAV is a novel AAV member of the genus *Dependovirus* and is phylogenetically closer to mammalian AAVs than to avian, waterfowl or snake AAVs (Supplementary Fig. S3, available in JGV Online). To further investigate the phylogenetic position of BtAAV-YNM, we constructed a refined Bayesian phylogeny for 33 carefully chosen AAV sequences based on the large-scale ML phylogeny, which should represent the principal genetic diversity of members of the genus *Dependovirus* (Fig. 2). In this Bayesian phylogeny, the AAVs can be classified into 14 clades. BtAAV-YNM shares a well-supported monophyl with porcine AAV strain Po3, and this monophyl shares a most recent common ancestor with the majority of primate AAVs (AAV1–13 except for AAV5). These results suggest that BtAAV-YNM is closely related to porcine AAV strain Po3 and distantly related to primate AAVs (except for AAV5). Our data also imply that BtAAV-YNM is relatively divergent from AAV5 and AAVs from goat, bovine, rodents, chicken, waterfowl and snake, based on both the phylogeny topology and nucleotide similarity (Fig. 2).

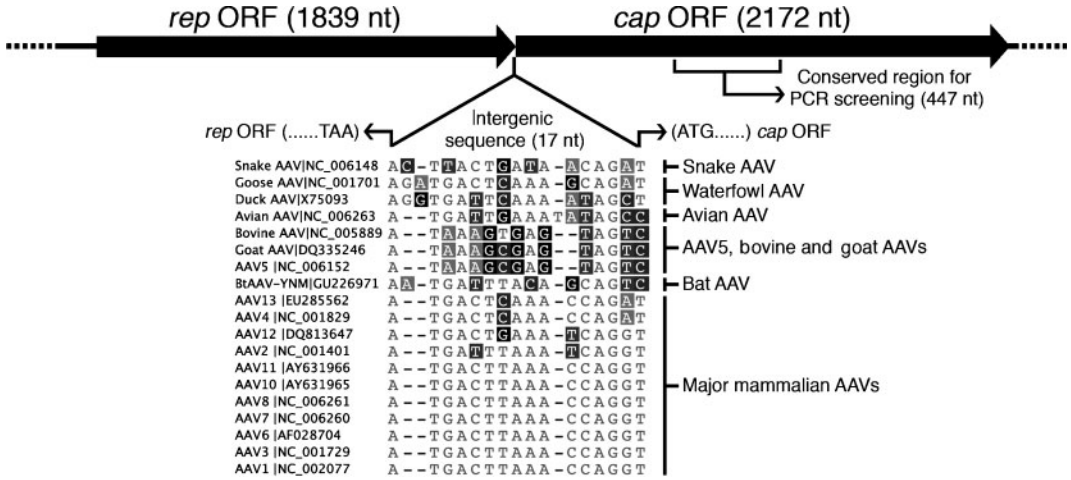
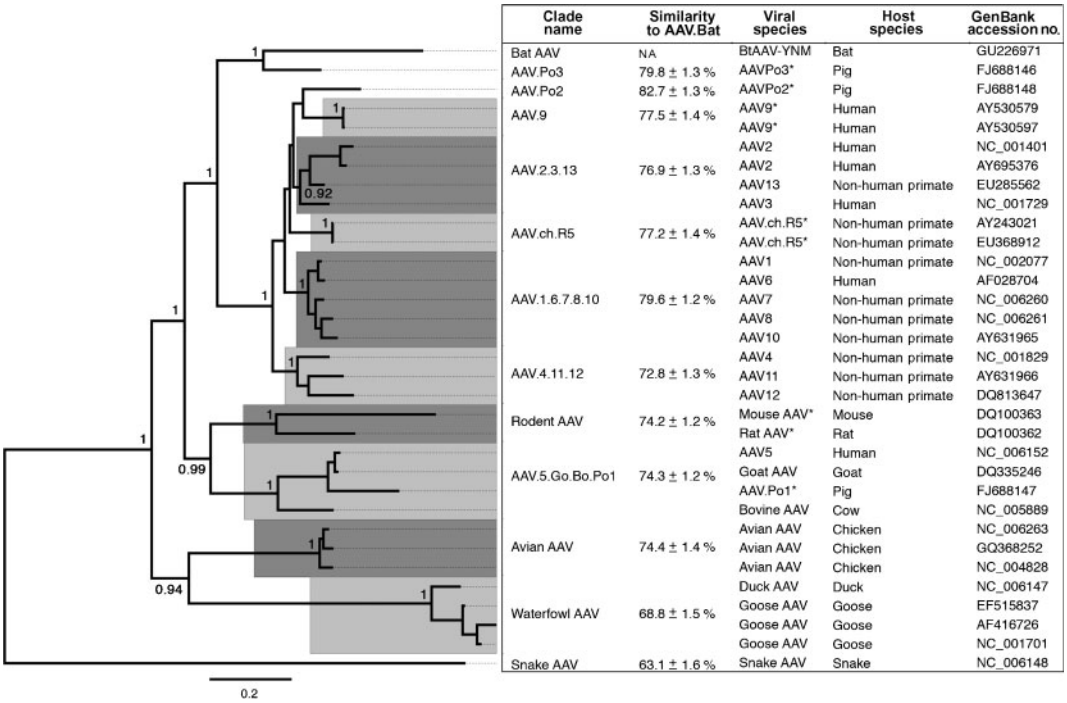


Fig. 1. The genome of BtAAV-YNM. A schematic diagram showing the relative scale of *rep* and *cap* ORFs, intergenic sequence, as well as the partial 5′ and 3′ UTR sequences of BtAAV-YNM. The region chosen for PCR screening is indicated. The intergenic sequences of BtAAV-YNM are aligned with those of other AAVs. The residues that are different from the 25 % consensus are highlighted in the alignment. The AAV species are shown on the left of the alignment and a rough classification of the AAV species is shown on right of the alignment.



Clade name	Similarity to AAV.Bat	Viral species	Host species	GenBank accession no.
Bat AAV	NA	BtAAV-YNM	Bat	GU226971
AAV.Po3	79.8 ± 1.3 %	AAVPo3*	Pig	FJ688146
AAV.Po2	82.7 ± 1.3 %	AAVPo2*	Pig	FJ688148
AAV.9	77.5 ± 1.4 %	AAV9*	Human	AY530579
		AAV9*	Human	AY530597
		AAV2	Human	NC_001401
AAV.2.3.13	76.9 ± 1.3 %	AAV2	Human	AY695376
		AAV13	Non-human primate	EU285562
		AAV3	Human	NC_001729
AAV.ch.R5	77.2 ± 1.4 %	AAV.ch.R5*	Non-human primate	AY243021
		AAV.ch.R5*	Non-human primate	EU368912
		AAV1	Non-human primate	NC_002077
		AAV6	Human	AF028704
AAV.1.6.7.8.10	79.6 ± 1.2 %	AAV7	Non-human primate	NC_006260
		AAV8	Non-human primate	NC_006261
		AAV10	Non-human primate	AY631965
		AAV4	Non-human primate	NC_001829
AAV.4.11.12	72.8 ± 1.3 %	AAV11	Non-human primate	AY631966
		AAV12	Non-human primate	DQ813647
Rodent AAV	74.2 ± 1.2 %	Mouse AAV*	Mouse	DQ100363
		Rat AAV*	Rat	DQ100362
		AAV5	Human	NC_006152
AAV.5.Go.Bo.Po1	74.3 ± 1.2 %	Goat AAV	Goat	DQ335246
		AAV.Po1*	Pig	FJ688147
		Bovine AAV	Cow	NC_005889
		Avian AAV	Chicken	NC_006263
Avian AAV	74.4 ± 1.4 %	Avian AAV	Chicken	GQ368252
		Avian AAV	Chicken	NC_004828
		Duck AAV	Duck	NC_006147
Waterfowl AAV	68.8 ± 1.5 %	Goose AAV	Goose	EF515837
		Goose AAV	Goose	AF416726
		Goose AAV	Goose	NC_001701
Snake AAV	63.1 ± 1.6 %	Snake AAV	Snake	NC_006148

Fig. 2. The phylogenetic origin of BtAAV-YNM. The phylogeny shown on the left is a Bayesian phylogeny constructed by using BEAST, based on the concatenated codon alignment of full-length ORFs of *rep* and *cap*. The numbers at selected nodes represent the Bayesian posterior probability support. The phylogeny is rooted using snake AAV. Bar, 0.2 nt substitutions per site. The shadings on the phylogeny correspond to the clades tabulated and shown on the right. The AAV species highlighted with asterisks are AAVs whose full-length ORF sequences are not available, and the gaps were treated as missing data in the phylogeny construction. The second column of the right panel represents the mean ± SEM pairwise nucleotide similarities (1000 bootstrap replicates) of the clade members to BtAAV-YNM, which were calculated based on the alignment of the *cap* region (818 nt) that is common to all taxa by using MEGA v4.1 (Kumar *et al.*, 2004). NA, Not applicable.

Genetic diversity of bat AAVs

To investigate the genetic diversity of bat AAVs, partial *cap* ORF sequences were obtained from direct sequencing of the PCR products (n=49; selected from the positive samples of the PCR screening assay). The amplified region covers aa 249–397 of *cap* ORF of AAV2 (GenBank accession no. NC_001401), containing two previously identified hypervariable regions (Gao *et al.*, 2003). Careful inspection of the sequencing chromatograms suggested sequence heterogeneity in the amplicons of nine samples (i.e. overlapping multicolour peaks), implying the co-presence of multiple distinct genotypes within an individual sample of bat AAV. These nine samples were collected from four different bat species, including *Rhinolophus sinicus*, *Rhinolophus affinis*, *Myotis daubentoni* and *M. ricketti*. To further investigate the existence of multiple AAV genotype in these samples, representative PCR products from each of the four different bat species (n=4; the other five samples were discarded) were cloned and 15 clones from each sample were sequenced. The majority consensus sequences of the highly similar cloned sequences were regarded as the representative sequence of the distinctive AAV genotypes in the individual samples (refer to Methods). Three of the four samples contain two distinctive genotypes, and one of the four samples contains three distinctive genotypes. These genotypes can be classified into topologically distinct sublineages (Fig. 3b; will be discussed later). These representative consensus sequences (n=9), along with the other directly sequenced bat AAV samples (n=40) and other mammalian AAV sequences (n=93; selected from the large-scale ML phylogeny in Supplementary Fig. S3), were used to construct an ML phylogeny (Fig. 3a), aiming to investigate the genetic diversity of bat AAVs with respect to mammalian AAVs.

In the ML phylogeny in Fig. 3(a), all bat AAVs fall into a monophyletic lineage that is closely related to AAVPo3, agreeing with the Bayesian phylogeny in Fig. 2. Moreover, based on visual inspections of the topology and internal branch lengths of the lineages, the diversity of AAV.Bat lineage appears to be comparable to that of the major mammalian AAV lineage. In fact, the average pairwise nucleotide identity among bat AAVs (84.3 %) is comparable to that among the major mammalian AAVs (86.2 %) (Fig. 3a). These results suggest that the bat AAVs sampled in this study are considerably diverse. To further investi-

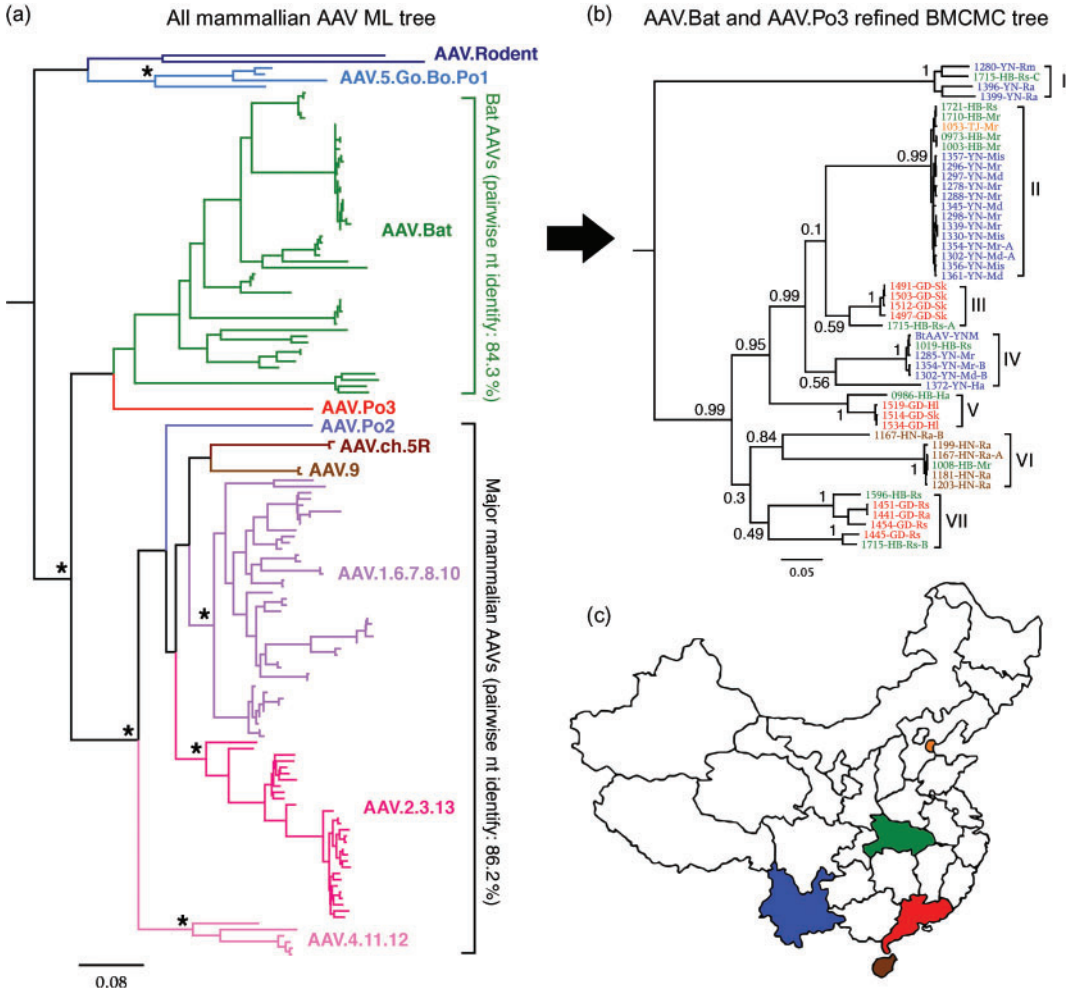


Fig. 3. Genetic diversity of bat AAVs. (a) An ML phylogeny constructed using PHYML, based on the concatenated codon alignment of the full-length and partial ORFs of *rep* and *cap*. The asterisks at selected nodes indicate their proportion of bootstrap support (100 replicates) is greater than 70 %. The phylogeny was rooted using avian AAVs. Bar, 0.08 nt substitutions per site. Bat AAV and major mammalian AAV lineage are highlighted using brackets and the intra-lineage pairwise nucleotide similarity based on the *cap* region (414 nt) are shown. (b) A Bayesian phylogeny constructed by using BEAST, based on the codon alignment of a conserved *cap* region as indicated in Fig. 1. The phylogeny is rooted using porcine AAV strain Po3. Bar, 0.05 nt substitutions per site. The numbers at selected nodes represent the Bayesian posterior probability support. The taxa names are coloured according to their geographical origins as indicated in (c). The taxa names are labelled as (sample number)–(geographical origin)–(host bat species). Abbreviations of geographical origins and host bat species: TJ, Tianjin; HB, Hubei; GD, Guangdong; YN, Yunnan; HN, Hainan; Ra, *Rhinolophus affinis*; Rs, *Rhinolophus sinicus*; Rm, *Rhinolophus macrotis*; Ha, *Hipposideros armiger*; Hl, *Hipposideros larvatus*; Md, *Myotis daubentoni*; Mr, *Myotis ricketti*; Sk, *Scotophilus kuhlii*; Mis, *Minipopterus schreibersii*. For individual samples with multiple distinct genotypes, an additional letter is appended after the species abbreviation. (c) A map of China and the five provinces from where samples were taken.

gate the population structure of bat AAVs, a refined Bayesian phylogeny was constructed for all bat AAV sequences (n=49; Fig. 3b). Based on the tree topology, bat AAVs can be classified into seven sublineages (designated sublineage I–VII). First and the foremost, host species specificity was not observed among these sublineages, suggesting that these bat AAV genotypes could be transmitted between different bat species without too many restrictions. Second, a semi-restricted geographical distribution pattern was observed among these bat AAV sublineages. The bat AAVs collected from province Hubei (HB) appear in all seven sublineages, which are relatively more diverse than bat AAVs collected from other provinces. If the bat AAVs collected from province HB were not considered, a considerable pattern of geographical restriction can be observed, i.e. sublineage I and VI are specific to YN; sublineage III, V and VII are specific to Guangdong (GD); sublineage VI is specific to Hainan

(HN). Such a semi-restricted pattern of geographical distributions will be further investigated in future surveillance studies. Lastly, sublineage I appears to be remarkably divergent from all other sublineages, and full genome characterization of the representatives selected from all sublineages should help to delineate the phylogenetic relationship between these bat AAV sublineages.

DISCUSSION

Here, we report the discovery of AAVs from 10 insectivorous bat species, including *R. affinis*, *R. sinicus*, *Rhinolophus pearsoni*, *Rhinolophus macrotis*, *H. armiger*, *H. larvatus*, *M. daubentoni*, *M. ricketti*, *Minipopterus schreibersii* and *Scotophilus kuhlii* distributed in five provinces in China. The average prevalence rate was 22.4 %, indicating that AAVs are considerably prevalent among different bat species. Although no AAV sequences were detected in the other nine species sampled in this study, it is believed that the failure of detection is most probably due to the limited sampling size (with the exception of *H. pomona* in HN and GD). We speculate that AAVs are prevalent in other bat species as well, which argues for a much wider surveillance to cover more bat species from more geographical locations. The bat AAVs discovered in this study also display remarkable genetic diversity, which seems to be independent of host species specificity.

AAV genomes are widely distributed throughout tissues of human and non-human primates as proviral DNAs (Gao *et al.*, 2002, 2003, 2004; Mori *et al.*, 2004). Previous reports have suggested that primate AAVs have undergone substantial evolution and homologous recombination during natural infections (Gao *et al.*, 2003, 2004; Takeuchi *et al.*, 2008). Nonetheless, Simplot analysis of the genome of BtAAV-YMN obtained in this study suggested that bat AAVs are unlikely to originate from recent recombination between known AAVs.

As its name suggests, AAV was initially thought to be closely associated with adenoviruses. In our study, the same set of 370 bat samples was also subjected to PCR detection of bat adenoviruses. Although we have detected the presence of bat adenoviruses in several bat species, *M. ricketti* and *S. kuhlii* (Li *et al.*, 2010), the co-presence of bat adenovirus and AAV were found in only 19.3 % of the samples. Due to the fact that other viruses such as herpesvirus and vaccinia virus might also help AAV to replicate (Schlehofer *et al.*, 1986; Wibbelt *et al.*, 2007), it is worthwhile extending future surveillance to include these viruses as well. It is now known that AAV replication is not strictly helper-dependent and genotoxic stress can also activate AAV genome replication (Yakobson *et al.*, 1987; Yalkinoglu *et al.*, 1988). Future studies will enable us to determine whether the replication of bat AAVs is strictly helper-dependent. In this study, we consistently detected AAVs in faecal swabs of various bat species, indicating that AAV viral particles are probably shed from tissues of the digestive tracts in bats. However, we have no data to show whether these virus genomes are integrated in the host genome and how they were replicated and released from tissues.

AAVs are being evaluated as vectors for human gene therapy. Following the establishment of the first infectious clone of AAV2 in 1982 (Samulski *et al.*, 1982), AAV2 vectors have rapidly gained popularity in gene therapy applications. Up to now, several AAV serotypes and over 100 AAV novel isolates have been acquired from adenovirus stocks or from animal tissues (Atchison *et al.*, 1965; Bantel-Schaal & zur Hausen, 1984; Gao *et al.*, 2002; Hoggan *et al.*, 1966; Mori *et al.*, 2004; Rutledge *et al.*, 1998; Schmidt *et al.*, 2006). Utilization of alternative AAV serotypes not only helps to lower the vector load due to their potentially higher transduction efficiency, but also helps to escape from the pre-existing neutralizing antibodies generated as a result of humoral immune responses against natural infection or prior treatment with AAV-based vectors. In addition, AAV serotypes and variants can serve as templates for design of tissue targeted capsid constructs that will serve to expand and complement the current range of AAV vectors (Wu *et al.*, 2006). Here, we have shown that the Cap protein of BtAAV-YNM is relatively divergent from other mammalian AAVs (ranged from 56.5 to 63.2 % aa identity), which implies that bat AAV Cap protein might have a low cross-serological reactivity to existing mammalian AAVs and could possibly be used as alternative vectors for gene therapy.

In summary, the discovery of novel AAVs in bats and their remarkable genetic diversity provide new insights into the ecology and evolution of AAVs and reinforce the role of bats as a reservoir of viruses that sometimes might pose zoonotic threats to public health. The discoveries of diverse bat DNA viruses mimic the trend of bat RNA virus discoveries and highlight the importance of further surveillance of bat.

METHODS

Sample collection. Bat samples were collected during May 2007 to May 2008, using Methods described previously (Li *et al.*, 2005). Faecal swab samples were collected in viral transport medium and transported to the laboratory in liquid nitrogen. Samples were stored at –80 °C until analysed.

Viral nucleic acid extraction and PCR detection. Viral DNA was extracted from bat faecal swabs with QIAamp DNA Blood Mini kits (Qiagen) following the manufacturer's instructions. DNA was eluted in 60 µl elution buffer and stored at −20 °C. One pair of primers (forward: 5′-CTGCCCACCTACAACAACCA-3′ and reverse: 5′-GTAYTCBAGGCAGTAGAAGG-3′), which targets conserved regions of the *cap* ORF, was used for PCR screening for AAV. The amplified region corresponds to aa 249–397 of *cap* ORF of AAV2, which contains two previously identified hypervariable regions, i.e. HVR3 and 4 (Gao *et al.*, 2003). The PCR amplification was carried out in a 25 µl reaction mix containing 2 µl extracted DNA, 2.5 µl PCR buffer, 20 pmol of each primer, 0.2 mM dNTP and 0.5 U *Taq* DNA polymerase (Promega). After an initial incubation at 94 °C for 5 min,

30 cycles of amplification were carried out consisting of denaturation at 94 °C for 30 s, annealing at 50 °C for 30 s, extension at 72 °C for 30 s and a final extension at 72 °C for 10 min. Standard precautions were taken to avoid PCR contamination, and a negative control was included in each PCR assay.

Sequencing of BtAAV-YNM genome region spanning the entire *rep–cap* ORFs. Bat AAV identified from bat species *M. ricketti* in the YN province (designated BtAAV-YNM) was selected for further full-length genome sequencing. PCR was performed with random and conserved primers (sequences available upon request) and PCR products were gel purified and cloned into pGEM-T Easy Vector (Promega). At least three independent clones were subjected to sequencing to obtain a consensus sequence for each of the amplified regions. The sequence of the *cap* ORF 3′ end of the genome was determined using a 3′ RACE kit (Roche Diagnostics) with nested primers, which are complementary to the known sequence of the genome. The sequence of *rep* ORF 5′ end was determined by conducting the 3′ RACE of the complementary strand. All these sequences were assembled into a contig and the presented sequence is the majority consensus of this contig. The obtained sequence covered the start and stop codons of both *rep* and *cap* ORFs, as well as the partial 5′ and 3′ UTR of 199 and 53 nt, respectively. No palindrome sequence is detected in the 199 nt of 5′ UTR, using inverted repeat finder (http://mobyle.pasteur.fr/cgi-bin/portal.py?form=palindrome).

Sequencing of bat AAV partial genome regions. After the PCR screening, the amplicons of expected size (447 nt) were gel purified using E.Z.N.A Gel Extraction kit (Omega Bio-Tek) and sequenced with both forward and reverse primers. Forty-nine PCR products were successfully sequenced. The sequencing chromatograms were inspected carefully for overlapping multicolour peaks, which is an indicator of sequence heterogeneity in the amplicons. Of the nine samples that showed sequence heterogeneity, four samples were selected for further analyses and the other five samples were discarded. The PCR products of these four samples were cloned into pGEM-T Easy Vector (Promega) and 15 clones of each sample were sequenced. The potential *in vitro* recombinant sequences (i.e. PCR artefacts) were screened and discarded using Recombination Detection Program v2.0 (Martin *et al.*, 2005). Highly similar sequences (>98 % nt identity) derived from the same sample were grouped and their majority consensus sequences were taken as the representatives of the distinctive genotypes in these samples. These sequences are deposited in GenBank (accession nos GU226879–GU226950).

Datasets and phylogenetic analyses. All available sequences of the members of the genus *Dependovirus* were obtained from GenBank (n=643; 28th February, 2010). Sequences shorter than 150 nt and artificial sequences (e.g. patented vector sequences) were discarded, generating a dataset of 297 sequences including both full-length and partial genomes. The codons of *rep* and *cap* ORFs of these sequences were extracted and aligned with all bat AAV sequences obtained in this study, based on their amino acid sequences, using MUSCLE v3.6 (Edgar, 2004). The alignment was manually corrected and the ambiguously aligned codons were manually removed. To construct large-scale phylogenies with more than 100 taxa (i.e. Supplementary Fig. S1, n=341; and Fig. 3a, n=141), a heuristic ML method implemented in PHYML v3.0 was used (Guindon & Gascuel, 2003). These ML phylogenies were estimated under the HKY85 nucleotide substitution model, with four categories of gamma rate heterogeneity and a proportion of invariant sites. The robustness of these ML topologies was assessed using 100 bootstrap replicates. To construct refined phylogenies with less than 50 taxa (i.e. Fig. 2, n=33; Fig. 3b, n=49), a Bayesian Markov Chain Monte Carlo (MCMC) method implemented in BEAST v1.5.3 was used (Drummond & Rambaut, 2007). These Bayesian phylogenies were estimated under SRD06 codon substitution model (Shapiro *et al.*, 2006) with the assumption of a lognormally relaxed molecular clock (Drummond *et al.*, 2006) and a non-parametric coalescent model (Drummond *et al.*, 2005). The mean substitution rate prior was fixed at 1.0. The MCMC chains were run for 10 million replicates and samples were taken at an interval of 1000 replicates. The MCMC burn-ins were determined visually in TRACER of the BEAST program and the effective sample size of all parameters were greater than 200. The final Bayesian phylogenies were summarized from these MCMC samples and the final branch lengths were expressed as the mean number of nucleotide substitutions per site.

ACKNOWLEDGEMENTS

We thank Dr April Davis (Wadsworth Center, New York State Department of Health) for revising the manuscript. This work was jointly funded by the State Key Program for Basic Research Grant (2011CB504700), National Natural Science Foundation of China (30970137) and Guangdong Natural Science Foundation (8151009101000005).

REFERENCES

Altringham, J. D., McOwat, T. & Hammond, L. (1996). *The Evolution and Diversity of Bats. Bats: Biology and Behaviour*. London, UK: Oxford University Press.

Arbetman, A. E., Lochrie, M., Zhou, S., Wellman, J., Scallan, C., Doroudchi, M. M., Randlev, B., Patarroyo-White, S., Liu, T. & other authors (2005). Novel caprine adeno-associated virus (AAV) capsid (AAV-Go.1) is closely related to the primate AAV-5 and has unique tropism and neutralization properties. *J Virol* **79**, 15238–15245.

Atchison, R. W., Casto, B. C. & Hammon, W. M. (1965). Adenovirus-associated defective virus particles. *Science* **149**, 754–756.

Bantel-Schaal, U. & zur Hausen, H. (1984). Characterization of the DNA of a defective human parvovirus isolated from a genital site. *Virology* **134**, 52–63.

Bauer, H. J. & Monreal, G. (1986). Herpesviruses provide helper functions for avian adeno-associated parvovirus. *J Gen Virol* **67**, 181–185.

Bello, A., Tran, K., Chand, A., Doria, M., Allocca, M., Hildinger, M., Beniac, D., Kranendonk, C., Auricchio, A. & other authors (2009). Isolation and evaluation of novel adeno-associated virus sequences from porcine tissues. *Gene Ther* **16**, 1320–1328.

Bossis, I. & Chiorini, J. A. (2003). Cloning of an avian adeno-associated virus (AAAV) and generation of recombinant AAAV particles. *J Virol* **77**, 6799–6810.

Brown, K. E., Green, S. W. & Young, N. S. (1995). Goose parvovirus – an autonomous member of the *Dependovirus* genus? *Virology* **210**, 283–291.

Chiorini, J. A., Yang, L., Liu, Y., Safer, B. & Kotin, R. M. (1997). Cloning of adeno-associated virus type 4 (AAV4) and generation of recombinant AAV4 particles. *J Virol* **71**, 6823–6833.

Chiorini, J. A., Kim, F., Yang, L. & Kotin, R. M. (1999). Cloning and characterization of adeno-associated virus type 5. *J Virol* **73**, 1309–1319.

Drummond, A. J. & Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* **7**, 214.

Drummond, A. J., Rambaut, A., Shapiro, B. & Pybus, O. G. (2005). Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol Biol Evol* **22**, 1185–1192.

Drummond, A. J., Ho, S. Y., Phillips, M. J. & Rambaut, A. (2006). Relaxed phylogenetics and dating with confidence. *PLoS Biol* **4**, e88.

Eaton, B. T., Broder, C. C. & Wang, L. F. (2005). Hendra and Nipah viruses: pathogenesis and therapeutics. *Curr Mol Med* **5**, 805–816.

Edgar, R. C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**, 113.

Farkas, S. L., Zadori, Z., Benko, M., Essbauer, S., Harrach, B. & Tijssen, P. (2004). A parvovirus isolated from royal python (*Python regius*) is a member of the genus *Dependovirus*. *J Gen Virol* **85**, 555–561.

Gao, G. P., Alvira, M. R., Wang, L., Calcedo, R., Johnston, J. & Wilson, J. M. (2002). Novel adeno-associated viruses from rhesus monkeys as vectors for human gene therapy. *Proc Natl Acad Sci U S A* **99**, 11854–11859.

Gao, G., Alvira, M. R., Somanathan, S., Lu, Y., Vandenberghe, L. H., Rux, J. J., Calcedo, R., Sanmiguel, J., Abbas, Z. & Wilson, J. M. (2003). Adeno-associated viruses undergo substantial evolution in primates during natural infections. *Proc Natl Acad Sci U S A* **100**, 6081–6086.

Gao, G., Vandenberghe, L. H., Alvira, M. R., Lu, Y., Calcedo, R., Zhou, X. & Wilson, J. M. (2004). Clades of adeno-associated viruses are widely disseminated in human tissues. *J Virol* **78**, 6381–6388.

Goncalves, M. A. (2005). Adeno-associated virus: from defective virus to effective vector. *Virol J* **2**, 43.

Grieger, J. C. & Samulski, R. J. (2005). Adeno-associated virus as a gene therapy vector: vector development, production and clinical applications. *Adv Biochem Eng Biotechnol* **99**, 119–145.

Guindon, S. & Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* **52**, 696–704.

Hess, M., Paul, G., Kling, S. & Monreal, G. (1995). Molecular characterization of two strains of the avian adeno-associated virus (AAAV). *Arch Virol* **140**, 591–598.

Hoggan, M. D., Blacklow, N. R. & Rowe, W. P. (1966). Studies of small DNA viruses found in various adenovirus preparations: physical, biological, and immunological characteristics. *Proc Natl Acad Sci U S A* **55**, 1467–1474.

Kumar, S., Tamura, K. & Nei, M. (2004). MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief Bioinform* **5**, 150–163.

Lau, S. K., Woo, P. C., Li, K. S., Huang, Y., Tsoi, H. W., Wong, B. H., Wong, S. S., Leung, S. Y., Chan, K. H. & Yuen, K. Y. (2005). Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats. *Proc Natl Acad Sci U S A* **102**, 14040–14045.

Leroy, E. M., Kumulungui, B., Pourrut, X., Rouquet, P., Hassanin, A., Yaba, P., Delicat, A., Paweska, J. T., Gonzalez, J. P. & Swanepoel, R. (2005). Fruit bats as reservoirs of Ebola virus. *Nature* **438**, 575–576.

Li, W., Shi, Z., Yu, M., Ren, W., Smith, C., Epstein, J. H., Wang, H., Crameri, G., Hu, Z. & other authors (2005). Bats are natural reservoirs of SARS-like coronaviruses. *Science* **310**, 676–679.

Li, Y., Ge, X., Zhang, H., Zhou, P., Zhu, Y., Zhang, Y., Yuan, J., Wang, L. F. & Shi, Z. (2010). Host range, prevalence, and genetic diversity of adenoviruses in bats. *J Virol* **84**, 3889–3897.

Maeda, K., Hondo, E., Terakawa, J., Kiso, Y., Nakaichi, N., Endoh, D., Sakai, K., Morikawa, S. & Mizutani, T. (2008). Isolation of novel adenovirus from fruit bat (*Pteropus dasymallus yayeyamae*). *Emerg Infect Dis* **14**, 347–349.

Martin, D. P., Williamson, C. & Posada, D. (2005). RDP2: recombination detection and analysis from sequence alignments. *Bioinformatics* **21**, 260–262.

Melnick, J. L., Mayor, H. D., Smith, K. & Rapp, F. (1965). Association of 20-millimicron particles with adenoviruses. *J Bacteriol* **90**, 271–274.

Molnar, V., Janoska, M., Harrach, B., Glavits, R., Palmai, N., Rigo, D., Sos, E. & Liptovszky, M. (2008). Detection of a novel bat gammaherpesvirus in Hungary. *Acta Vet Hung* **56**, 529–538.

Mori, S., Wang, L., Takeuchi, T. & Kanda, T. (2004). Two novel adeno-associated viruses from cynomolgus monkey: pseudotyping characterization of capsid protein. *Virology* **330**, 375–383.

Muzyczka, N. & Berns, K. I. (2001). *Parvoviridae*: the viruses and their replication. In *Fields Virology*, 4th edn, pp. 2327–2360. Edited by D. M. Knipe & P. M. Howley. Philadelphia, PA: Lippincott, Williams & Wilkins.

Rutledge, E. A., Halbert, C. L. & Russell, D. W. (1998). Infectious clones and vectors derived from adeno-associated virus (AAV) serotypes other than AAV type 2. *J Virol* **72**, 309–319.

Samulski, R. J., Berns, K. I., Tan, M. & Muzyczka, N. (1982). Cloning of adeno-associated virus into pBR322: rescue of intact virus from the recombinant plasmid in human cells. *Proc Natl Acad Sci U S A* **79**, 2077–2081.

Schlehofer, J. R., Heilbronn, R., Georg-Fries, B. & zur Hausen, H. (1983). Inhibition of initiator-induced SV40 gene amplification in SV40-transformed Chinese hamster cells by infection with a defective parvovirus. *Int J Cancer* **32**, 591–595.

Schlehofer, J. R., Ehrbar, M. & zur Hausen, H. (1986). Vaccinia virus, herpes simplex virus, and carcinogens induce DNA amplification in a human cell line and support replication of a helpervirus dependent parvovirus. *Virology* **152**, 110–117.

Schmidt, M., Katano, H., Bossis, I. & Chiorini, J. A. (2004). Cloning and characterization of a bovine adeno-associated virus. *J Virol* **78**, 6509–6516.

Schmidt, M., Grot, E., Cervenka, P., Wainer, S., Buck, C. & Chiorini, J. A. (2006). Identification and characterization of novel adeno-associated virus isolates in ATCC virus stocks. *J Virol* **80**, 5082–5085.

Schmidt, M., Voutetakis, A., Afione, S., Zheng, C., Mandikian, D. & Chiorini, J. A. (2008). Adeno-associated virus type 12 (AAV12): a novel AAV serotype with sialic acid- and heparan sulfate proteoglycan-independent transduction activity. *J Virol* **82**, 1399–1406.

Schnepp, B. C., Jensen, R. L., Chen, C. L., Johnson, P. R. & Clark, K. R. (2005). Characterization of adeno-associated virus genomes isolated from human tissues. *J Virol* **79**, 14793–14803.

Schnepp, B. C., Jensen, R. L., Clark, K. R. & Johnson, P. R. (2009). Infectious molecular clones of adeno-associated virus isolated directly from human tissues. *J Virol* **83**, 1456–1464.

Shapiro, B., Rambaut, A. & Drummond, A. J. (2006). Choosing appropriate substitution models for the phylogenetic analysis of protein-coding sequences. *Mol Biol Evol* **23**, 7–9.

Takeuchi, Y., Myers, R. & Danos, O. (2008). Recombination and population mosaic of a multifunctional viral gene, adeno-associated virus cap. *PLoS One* **3**, e1634.

Towner, J. S., Pourrut, X., Albarino, C. G., Nkogue, C. N., Bird, B. H., Grard, G., Ksiazek, T. G., Gonzalez, J. P., Nichol, S. T. & Leroy, E. M. (2007). Marburg virus infection detected in a common African bat. *PLoS ONE* **2**, e764.

Wibbelt, G., Kurth, A., Yasmum, N., Bannert, M., Nagel, S., Nitsche, A. & Ehlers, B. (2007). Discovery of herpesviruses in bats. *J Gen Virol* **88**, 2651–2655.

Wong, S., Lau, S., Woo, P. & Yuen, K. Y. (2007). Bats as a continuing source of emerging infections in humans. *Rev Med Virol* **17**, 67–91.

Wu, Z., Asokan, A. & Samulski, R. J. (2006). Adeno-associated virus serotypes: vector toolkit for human gene therapy. *Mol Ther* **14**, 316–327.

Yakobson, B., Koch, T. & Winocour, E. (1987). Replication of adeno-associated virus in synchronized cells without the addition of a helper virus. *J Virol* **61**, 972–981.

Yalkinoglu, A. O., Heilbronn, R., Burkle, A., Schlehofer, J. R. & zur Hausen, H. (1988). DNA amplification of adeno-associated virus as a response to cellular genotoxic stress. *Cancer Res* **48**, 3123–3129.