Virus Genes (2019) 55:769–778
https://doi.org/10.1007/s11262-019-01698-4

ORIGINAL PAPER

Molecular evolution of human adenovirus type 16 through multiple recombination events

Xingui Tian[1] · Hongkai Wu[1] · Rong Zhou[1]

Received: 30 January 2019 / Accepted: 30 July 2019 / Published online: 5 August 2019
© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract

Human mastadenoviruses (HAdVs) are non-enveloped, double-stranded DNA viruses that are comprised of more than 85 types classified within seven species (A–G) based on genomics. All HAdV prototypes and many newly defined type genomes have been completely sequenced and are available. Computational analyses of the prototypes and newly emergent HAdV strains provide insights into the evolutionary history and molecular adaptation of HAdV. Most types of HAdV-B are important pathogens causing severe respiratory infections or urinary tract infections and are well characterized. However, HAdV-16 of the B1 subspecies has rarely been reported and its genome is poorly characterized. In this study, bioinformatics analysis, based on genome sequences obtained in GenBank, suggested that HAdV-16, a prototype HAdV-B species, evolved from multiple intertypic recombination events. HAdV-16 genome contains the hexon loop 1 to loop 2 region from HAdV-E4, the partial hexon conserved region 4 (C4) from the subspecies HAdV-B2, genome region 30,897–33,384 containing the fiber gene from SAdV-35, and other genomic parts from the subspecies HAdV-B1. Moreover, analysis of sequence similarity with HAdV-E4 LI, LII, and SAdV-36 strains demonstrated the recombination events happened rather early. Further, amino acid sequence alignment indicated that the amino acid variations occurred in hypervariable regions (HVRs). Especially, the major difference in HVR7, which contains the critical neutralization epitope of HAdV-E4, between HAdV-16 and HAdV-E4 might explain the low level of cross-neutralization between these strains. Our findings promote better understanding on HAdV evolution, predicting newly emergent HAdV strains, and developing novel HAdV vectors.

Keywords Adenovirus type 16 · Recombination event · Hexon · Adenovirus type 4 · Variation

Introduction

Human mastadenoviruses (HAdVs) are non-enveloped, double-stranded DNA viruses that belong to the *Adenoviridae* family. To date, more than 85 HAdV types, which are classified within seven species (A–G), have been identified and defined using a new paradigm based on genomics [1–4]. HAdVs of species B can be divided into two subspecies, namely subspecies B1 including HAdV-3, HAdV-7, HAdV-16, HAdV-21, and HAdV-50, and subspecies B2 including HAdV-11, HAdV-14, HAdV-34, HAdV-35, and HAdV-55. HAdVs commonly cause respiratory diseases including the common cold and severe pneumonia and also contribute to other diseases such as gastroenteritis, cystitis, conjunctivitis, carditis, and meningoencephalitis. HAdV infections can occur among patients of all ages and susceptible populations include infants, school students, military recruits, and immunocompromised patients [5, 6]. Further, specific types are often associated with particular clinical manifestations.

Edited by Detlev H. Kruger.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11262-019-01698-4) contains supplementary material, which is available to authorized users.

✉ Hongkai Wu
wuhongkai@gmail.com

✉ Rong Zhou
zhourong@gird.cn

[1] State Key Laboratory of Respiratory Disease, National Clinical Research Center for Respiratory Disease, Guangzhou Institute of Respiratory Health, the First Affiliated Hospital of Guangzhou Medical University, Guangzhou Medical University, Guangzhou, China

All HAdV prototype genomes and many new characterized type genomes have been completely sequenced. It is thus possible to investigate the evolutionary history and perform comparative analyses of the prototypes and newly emergent HAdV strains. For example, a re-emergent acute respiratory disease pathogen, named HAdV-55, was revealed to be an intertype recombinant of HAdV-11 and HAdV-14 [7, 8]. Previous computational analysis of HAdV-4, a pathogen that is the only HAdV member of species E, provides insights into its zoonotic origin and molecular adaptation [9]. Its genome encodes a domain of the major capsid protein, hexon, from HAdV-16 that has recombined into the genome chassis of a simian adenovirus. In addition, adenoviruses have been explored as vaccines or gene therapy vectors due to their safety profile and immunogenicity. Recently, some rare HAdV types and non-human adenoviruses have attracted attention as potential vectors to circumvent pre-existing immunity [10, 11]. However, some HAdV prototypes such as HAdV-16 are poorly characterized. Some previous studies indicated an intra-hexon recombination event in the origin of HAdV-16 [12–14]. HAdV-16 had the highest similarity to HAdV-E4 for nucleotide positions 1–1400 of the hexon gene including the neutralization determinant, whereas for other parts of the hexon gene HAdV-16 was more closely related to HAdV-11, which belongs to subspecies HAdV-B2 [12]. However, there is no report on detailed computational analysis of HAdV-16 based on complete genomes.

In this study, bioinformatics analysis based on genome sequences obtained in GenBank demonstrated that HAdV-16 evolved from multiple intertypic recombination events between simian adenovirus (SAdV) SAdV-35, HAdV-E4, subspecies HAdV-B1, and subspecies HAdV-B2, which contains the hexon loop 1 to loop 2 region gene from the HAdV-E4, partial hexon conserved region 4 (C4) gene from the subspecies HAdV-B2, genome region 30,897–33,384 containing the fiber gene from SAdV-35, and other genome parts from the subspecies HAdV-B1. Further amino acid sequence alignment indicated that the amino acid variations in HVRs between HAdV-16 and HAdV-E4 might explain the low-level cross-neutralization between them [15–17].

Materials and methods

Adenovirus genomes

The complete genome sequences of two HAdV-16 strains, the prototype strain ch79, a respiratory tract pathogen [18], and the strain E26, were obtained from GenBank. Sequence and annotation files for the 52 adenovirus genomes used in this study were downloaded from GenBank (Table S1). The GenBank archived genomes without annotation information for coding sequences, SAdV-36 (FJ025917), SAdV-26 (FJ025923), SAdV-35 (FJ025912), SAdV-41 (FJ025913), SAdV-33 (FJ025908), SAdV-27 (FJ025928), and SAdV-46 (FJ025930), were annotated using the software tool Genome Annotation Transfer Utility (GATU) [19] and manually confirmed by comparisons to previously annotated genomes CAdV-Y25 (JN254802) and HAdV-B3-GZ01 (DQ099432).

Phylogenetic analysis of the genomes and genes

ClustalW was used to align all nucleic acid and protein sequences by setting its default score matrices for nucleic acid or protein residues, respectively. The Smart Model Selection was used to select the best substitution model for the resulting alignments of nucleotides or proteins [20]. Then, according to the best substitution models, the maximum likelihood trees were inferred by PhyML [21]. These two processes above were done through the web server PhyML-SMS. Bootstrap analysis with 100 replicates was performed to estimate the robustness of specific tree topologies. The values of no less than 50 were reported. The circular form of the whole genome phylogenetic tree was presented using iTOL [22].

Recombination analysis

Genome and hexon gene similarity scanning analysis was performed for nucleotide sequence recombination assessments by aligning the nucleotides of HAdV-B16-ch79 strain or HAdV-B16-E26 strain (submission sequence) with those of 50 other HAdV strains. Using the nucleotide position of the submission sequence as a reference, scanning window size was set to 200 bp for whole genomes and 100 bp for hexon genes or the genome regions of 30,501–33,600; scanning step size was set to 20 bp for whole genomes and 10 bp for hexon genes or the genome regions of 30,501–33,600. The alignment in each window was replicated 100 times by the bootstrap method. For each replicated alignment, Kimura 2 parameter distance was calculated between the submission sequence and each of the other 50 sequences, prior to which, gaps were removed from the paired sequences. In each window position, the count (close number) of each of the other 50 sequences was the number of times that its distance was closest to the submitted sequence based on 100 repetitions. The heat map of the count was drawn at each location of 50 sequences. The 50 HAdV sequences were divided into nine groups as follows: A, B1, B2, C, D, E-HAd4LI, E-HAd4LII, E-uH, and F. In one replicated alignment of each window position, the close number of the group was counted once when one or more sequences in the corresponding group were closest to the submitted sequence. A similarity line chart was drawn for the close number collected through 100 replicated alignments at each location of nine groups.

At a specific location, a higher close number suggested the sequence of the strain or group was more similar to the submission sequence than that of other strains or groups. The bootstrap and distance computation were performed by corresponding modules in BioPerl. Perl scripts were written to execute the above processes of scanning and calculation. The Heatmaps and line charts were drawn by R package of pheatmap and ggplot2, respectively.

RDP (v4.97) [23] was applied to verify the recombination events on the hexon gene using the two sets of alignments (nucleotides of hexon gene of HAdV-B16-ch79 or HAdV-B16-E26 with those of 50 other HAdVs) separately. Similarly, RDP is also used to verify the recombination events on the genome region of 30,501–33,600 basing the positions of HAdV-B16, by inputting the slices of the two whole genome alignments (HAdV-BV16-ch79 or HAdV-B16-E26 with those of 50 other HAdVs) separately. In RDP, 6 methods, RDP [24], GENECONV [25], MaxChi [26], BootScan [27], SiScan [28], and 3Seq [29] were implemented, setting the window size to 100 and the step size to 10 in methods of BootScan and SiScan, remaining all other settings as their defaults.

Results

Sequence recombination analysis

The similarity plot analysis of the prototype strain, named ch79, of HAdV-B16 (Figure S1A and S1C) and the strain E26 of HAdV-B16 (Figure S1B and S1D) genome sequences compared to 50 other genome sequences suggested multiple recombination events across the genome. The heat maps of HAdV-B16-ch79 (Figure S1A) and HAdV-B16-E26 (Figure S1B) genome sequence similarity plot analyses, as compared to 50 other genome sequences, suggested HAdV-B16-ch79 and HAdV-B16-E26 contained a recombinant genome chassis of the HAdV B1 subspecies, namely, HAdV-3, HAdV-7, HAdV-21, and HAdV-50; moreover, the hexon regions were shared between the HAdV-E species and the HAdV B2 subspecies members HAdV-11, HAdV-34, and HAdV-35. The similarity diagram comparing different groups based on bootstrap analysis (Figure S1C and S1D) also supported this observation. Two HAdV strain genomes, HAdV-B3-Germany and HAdV-68, showed high similarity with HAdV-B16-ch79 (Figure S1A) and HAdV-B16-E26 genomes (Figure S1B) except in the hexon region, suggesting that these two strains might be recombinant viruses originally from HAdV-16.

Further detailed analysis of hexon sequences revealed that HAdV-16 hexon gene contained nucleotides 1–200 and the last part after nucleotides 2200 (ch79 strain) or 2100 (E26 strain), which were most similar to those sequences of its counterpart from the HAdV-B1 subspecies, nucleotides 220–1300, which were most similar to those from the HAdV-E species, and nucleotides 1310–2200 (ch79 strain) or 1310–2100 (E26 strain), which were most similar to those from the HAdV B2 subspecies. Also in the very narrow area after nucleotide 200, there were relatively high similarities between HAdV-16s and HAdV subspecies B2 (Figure S2). These results indicated at least two recombination events that contributed to the current HAdV-16 genomes. In several sub-regions of nucleotides 220–1300, HAdV-16 were less similar to parental sequences suggesting that the recombination events did not happen recently and that random mutations contributed to the divergence. The higher similarity of HAdV-16 hexon nucleotides 220–1300 with those from E-HAdV-4 LI strains, which were isolated 30–50 years ago, compared to that with E-HAdV-4 LII strains, which were isolated in recent years, also suggested that the recombination event in this region did not happen in recent years. It is not clear whether the two recombination events happened at the same time.

Further detailed analysis of genome regions of 30,501–33,600 revealed that HAdV-16 contained nucleotides 30,800–33,500 containing the fiber gene, which were most similar to that from SAdV-35 except for HAdV-B3-Germany and HAdV-B68 (Figure S3). These results indicated the third recombination event in this region of HAdV-16 genome with SAdV-35.

RDP analysis results showed strong evidence of these recombination events on the hexon gene (Table S2). Between the estimated breakpoints of 304 and 1361 on HAdV-16s, the region was recombinant from HAdV-E4s. From the estimated breakpoint of 209–2285 (ch79) or 2123 (E26), the region was recombinant from HAdV subspecies B2 strains. The RDP result also indicated a subsequent recombination event. The ignored nt sites in the detection of this recombination event covered the region that was just right from HAdV-4s (Figure S4/S5). In addition, there was very small recombinant region before the ignored segment. The recombinant signals in this region were much weaker than the signals in the large recombinant region after the ignored segment (Figure S5). RDP estimated more precise breakpoints of the recombination events and the results obtained by its multiple triplet sequences analysis verify the results of the previous one-to-many similarity scanning method. The results of RDP suggested that on hexon gene, HAdV-16 and HAdV subspecies B2 underwent a large fragment recombination, and in the first half of this recombination, recombination with HAdV-4 occurred. This was consistent with the results of the one-to-many similarity scan in which the first half was similar to HAdV-4s and the latter half was similar to HAdV subspecies B2s (Figure S2).

The recombination events on HAdV-16′s genome regions of 30,501–33,600 were also confirmed by RDP analysis

(Table 1). The result showed that among the human strains of subspecies B1, HAdV-16s, HAdV-68, and HAdV-B3-Germany obtained the correspondence segment from the simian strain of SAdV-35.

Phylogenetic analysis

Phylogenetic analysis of the whole genomes indicated that two HAdV-16 strains were within the clade that contained all subspecies B1 members; all subspecies B2 members were contained in another clade (Fig. 1).

According to recombination analysis results, several regions were examined in detail. The phylogenetic tree of IVa2 genes had a similar profile to the whole genome tree (Fig. 2a). Phylogenetic analysis of the penton base genes showed that HAdV-B16-ch79 and HAdV-B16-E26 formed a subclade of the HAdV species B clade with HAdV-3 and HAdV-7 strains, but that HAdV-21 and HAdV-50 of subspecies B1 were contained in another subclade with strains of subspecies B2. Surprisingly, HAdV-E4-vaccine formed a separate clade with HAdV-36 of species D (Fig. 2b). Phylogenetic analysis of the hexon genes showed that HAdV-B16-ch79 and HAdV-B16-E26 formed a separate subclade which was apart from all other types of HAdV-B1 subspecies and clustered into a large clade containing all of HAdV-E4s. HAdV-D/C/A/F were also contained in this clade but with relatively long branch lengths (Fig. 2c). Phylogenetic analysis of the fiber genes showed that HAdV-B16-ch79 and HAdV-B16-E26 formed a subclade with HAdV-B3-Germany, HAdV-68, and SAdV-35, on a branch apart from HAdV-B3-GB and HAdV-B3-GZ01. Further alignment of amino acid sequences showed HAdV-16 fibers were most similar to that of SAdV-35 with a similarity of 97.45%. Amino acid sequences of HAdV-16 fibers had similarities of 51.40–63.17% with those from other HAdV-B types (data not shown). This result supported the recombination event of HAdV-16 fiber from SAdV-35. HAdV-21 and HAdV-50 of subspecies B1 formed a subclade with HAdV-35 and HAdV-34 of subspecies B2, whereas HAdV-7 of subspecies B1 formed a subclade with HAdV-11, HAdV-14, and HAdV-55 of subspecies B2 (Fig. 2d).

Hexon analysis

The hexon protein is the predominant target of serotype-specific neutralizing antibodies [30–36], and the serotype-specific neutralization epitopes on hexon are located mainly on the tower region (loop 1 and loop 2), which consists of seven hypervariable regions (HVRs) [33, 37–42]. Similarity plot analysis of hexon genes (Figure S2) suggested high similarity between HAdV-16 and HAdV-E in the V1 to V3 region, which contains all seven HVRs, and between HAdV-16 and HAdV subspecies B2 in the first

Table 1 Recombination events on the genome regions of 30,501–33,600 of HAdV-B16 from RDP analysis

Event	Begin[a]	End[a]	Recombinant sequence	Minor parental sequences	Major parental sequences	*p* Value in detection methods[b]					
						R	G	B	M	S	T
1	30,897	33,490	HAdV-B68 HAdV-B3-Germany HAdV-B16-ch79	HAdV-B7-vaccine HAdV-B7-Gomen HAdV-B21-Germany HAdV-B21-NHRC63218 HAdV-B50	SAdV-35	7.11E−55	8.74E−45	1.67E−52	1.56E−15	2.06E−20	2.14E−51
2	30,792	33,384	HAdV-B68 HAdV-B3-Germany HAdV-B16-E26	HAdV-B7-vaccine HAdV-B7-Gomen HAdV-B21-Germany HAdV-B21-NHRC63218 HAdV-B50	SAdV-35	2.54E−57	4.72E−32	1.57E−54	6.81E−16	4.01E−15	5.48E−54

[a]The begin and end nt positions of estimated breakpoints on the recombinant genomes of HAdV-B16-ch79 and HAdV-B16-E26

[b]*R* RDP, *G* GENECONV, *B* BootScan, *M* MaxChi, *S* SiScan, *T* 3Seq

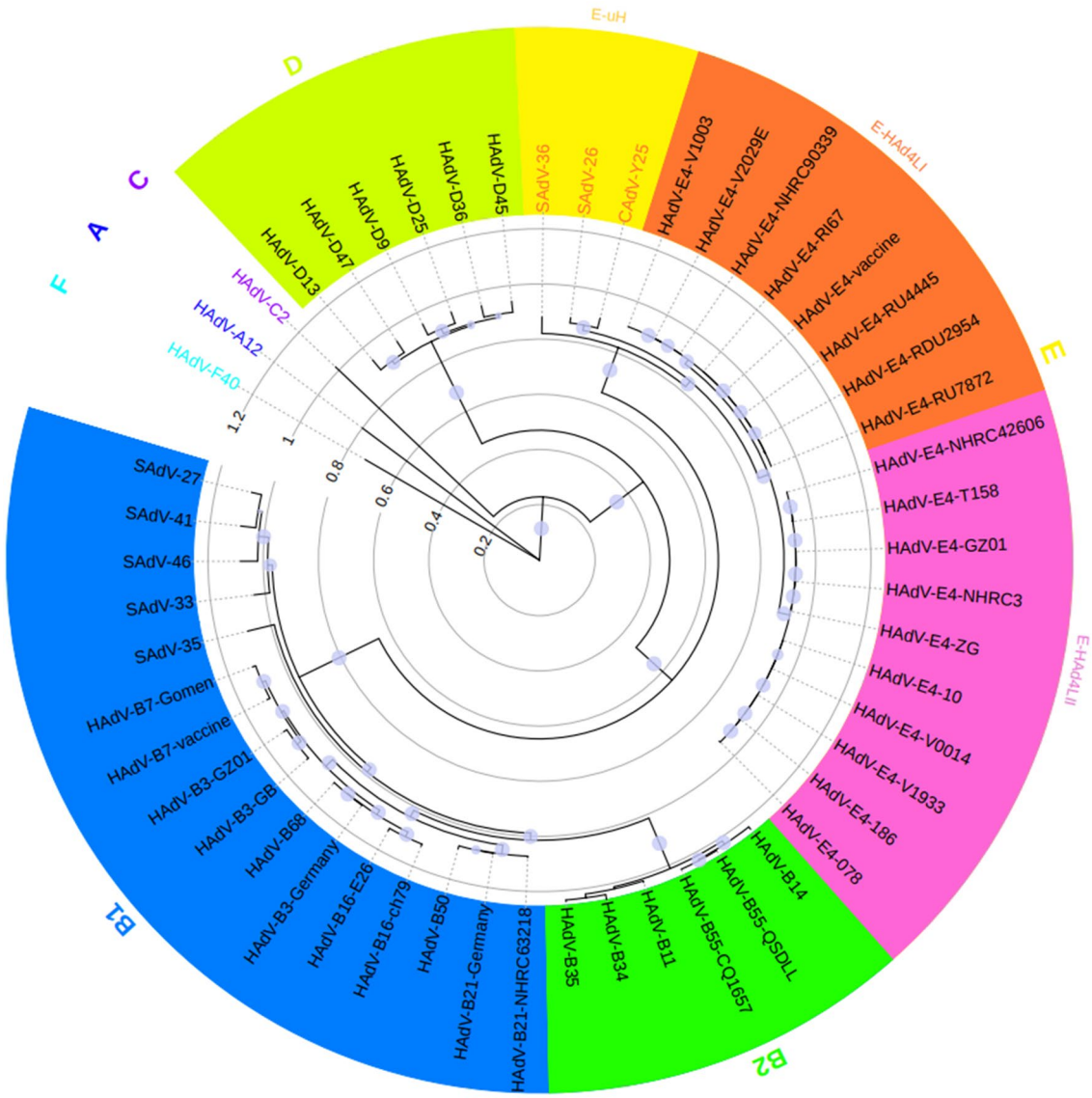

Fig. 1 Phylogenetic trees of 52 whole genomes of adenovirus strains. The phylogenetic tree was conducted by using the maximum likelihood method with the substitution model GTR + G + I (GTR matrix; G, gamma shape parameter: fixed; I, proportion of invariable sites: fixed). The size of gray dots on the branch points represents the consistent quantity derived from bootstrap analysis (a range from 54 to 100)

half of the C4 region. Therefore, the two regions (amino acid 121–456 and 443–695) were analyzed separately. Based on the phylogenetic tree for the region sequences (amino acid 121–456; Fig. 3a), HAdV-B16-ch79 and HAdV-B16-E26 formed a subclade within the HAdV-E clade, apart from the subclade of HAdV-E4 LI strains and the subclade of HAdV-E4 LII strains. Surprisingly, CAdV-Y25 strain formed a clade with HAdV-9 of species D. For the region sequences (amino acid 443–695; Fig. 3b), HAdV-B16-ch79 and HAdV-B16-E26 formed a branch within the HAdV subspecies B2 subclade, apart from the subclades of HAdV subspecies B1 strains. These analyses confirmed the results of the similarity plot analysis.

For a more detailed analysis, the hexon amino acid sequences were compared using multi-sequence alignment (Figure S6). Seven hypervariable regions (HVR1–7) were identified. The arrows show the amino acid substitutions of HAdV-16 and HAdV-E strains, which were all within the HVRs, except for one substitution. Amino acid variations in HVRs between HAdV-16 and HAdV-E4 might explain the low level of cross-neutralization between them [15–17]. The major difference between HAdV-16 and

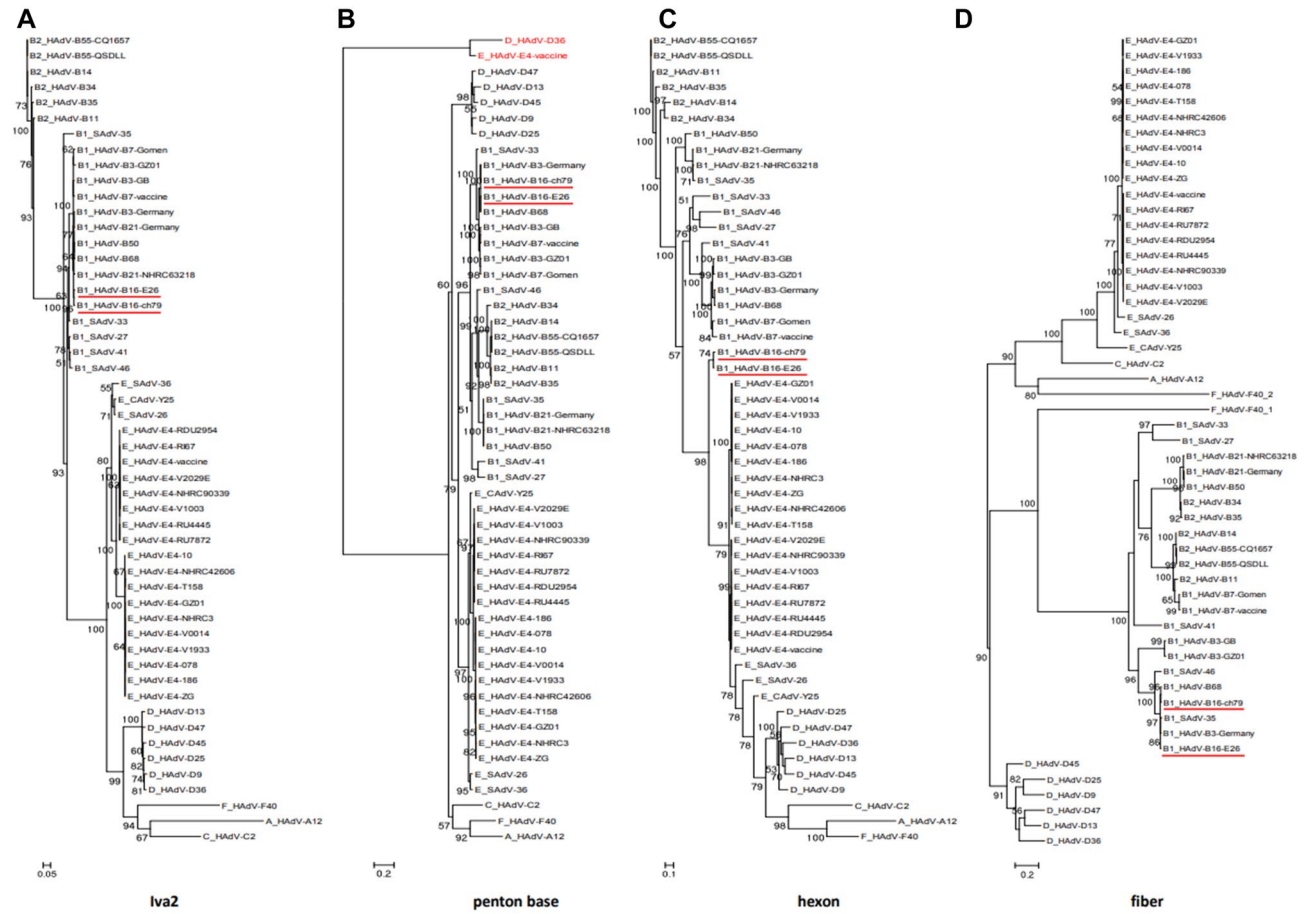

Fig. 2 Phylogenetic trees of four adenoviral genes of structural proteins, IVa2 (**a**), penton base (**b**), hexon (**c**) and fiber (**d**). The phylogenetic trees were constructed based on nucleotide sequences by using the maximum likelihood method with the substitution models: GTR + G + I for Iva2, hexon and fiber, GTR + G for penton base (GTR matrix; G, gamma shape parameter: fixed; I, proportion of invariable sites: fixed). The entry names began with the species (A–F) derived from whole genome phylogenetic analysis of the corresponding strains. Two strains of HAdV-B16 were marked with red underlines. Strains grouped into a clade inconsistent with their species were printed in red. Two fiber genes in HAdV-F40 were named "F_HAdV-F40_1" and "F_HAdV-F40_2" (**d**)

HAdV-E hexons was confined to HVR7 (loop 2). The residue AVAGTSGTQ in HAdV-16 was substituted with the residue AGSEK of HAdV-E4, which was conserved in HAdV-E4 strains. Previously, we identified this amino acid residue (AGSEK) in HVR7 as critical for the neutralization epitope of HAdV-E4 [33]. It could thus be inferred that this residue (AVAGTSGTQ) might be critical for HAdV-16 neutralization. It was also noted that this residue in HAdV-16 was more similar to that in SAdV-36 HVR7 than that in HAdV-E4.

Discussion

The one-to-many similarity scanning method in this study could provide the similarity profiles between one submission sequence and dozens of other sequences which were classified to several groups (Figure S1/S2/S3/). But, there were two limitations that need to be issued: First, although the values on heatmaps or line charts were computed basing on distances of nucleotides, they were the numbers of "times with the closest distance in the bootstrap." So, they were the count values and different from the distance values originally calculated by the model. For example, in some regions of the alignments, when the distances of all sequences to the submitted sequence were not very close, there were still some sequences that could obtain very high-count values. This method could sensitively detect the possible recombination signals, but the calculated values did not reflect the true distances. Second, this method did not contain any phylogenetic information. In general, under this method, the identification of sequence recombination requires the synthesis of the results of various other analytical methods.

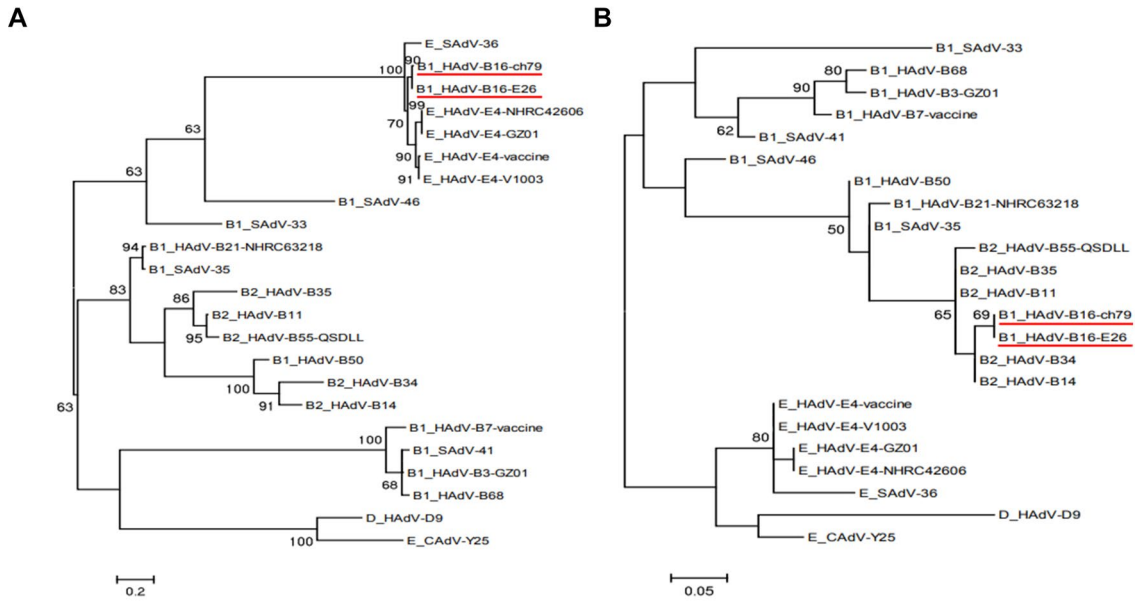

Fig. 3 Phylogenetic trees of two regions of the adenoviral hexon gene. The phylogenetic trees were constructed based on amino acid sequences by using the maximum likelihood method with the substitution models: JTT + G + I + F for (**a**) amino acid 121–456, CpREV + G + I + F for (**b**) amino acid 443–695 (JTT matrix; CpREV matrix; G, gamma shape parameter: fixed; I, proportion of invariable sites: fixed; F, equilibrium frequencies: empirical). Two strains of HAdV-B16 were marked with red underlines

Openly accessible and available databases provide large amounts of high-resolution virus genome data, which are useful for understanding infectious diseases and pathogens. All HAdV prototypes and many newly characterized genome types are available in the GenBank database. High-throughput sequencing and fast computational analysis methods will help to understand the molecular mechanism of evolution of old and novel viruses and to predict emerging pathogens.

Reported here, the HAdV-16 genome provides an example of a reanalysis of HAdV prototype genomes. As shown in Figure S1, using all 51 HAdV prototype genomes as references, the relationship between each HAdV-16 genome region with those of other HAdV prototype genomes could be easily revealed. Here, all prototypes of HAdV-B and typic strains of two HAdV-E subgroups [43, 44] were included. With this method, evidence is provided that HAdV-16 contains a recombinant subspecies B1 genome chassis and a hexon composite of E4 and subspecies B2. Several early studies reported a high degree of homology between HAdV-16 and HAdV-E4 hexons [9, 13, 14]. However, this is the first report that the HAdV-16 hexon 1310–2200 nucleic acid sequence is most similar to that from HAdV subspecies B2, as compared to that with other strains, and that HAdV-16 might have evolved from two recombination events. In the hexon gene, which contained recombination events, a phylogenetic tree showed that HAdV-16 s formed a clade with HAdV-E4s, and that this clade formed a higher-order clade with SAdV-V36, SAdV-26, CAdV-Y25, and HAdV-D strains, in that order (Fig. 2c). In the phylogenetic tree based on hexon region sequences (amino acid 121–456), HAdV-16s formed a clade with HAdV-E4 s, and then formed a higher-order clade with SAdV-36 and SAdV-46 (Fig. 3a). These two higher-order clades did not include HAdV-B. HAdV-68 strain Arg827/04 was an intertypic recombinant adenovirus with a serotype 3-like hexon gene and a serotype 16-like fiber isolated in Argentina from an infant admitted to the hospital with acute respiratory disease [45]. In this study, it was found that HAdV-68 and another strain HAdV-B3-Germany showed high similarity with HAdV-B16-ch79 (Figure S1A) and HAdV-B16-E26 genomes (Figure S1B) except in the hexon region, suggesting that these two strains might be recombinant viruses originally from HAdV-16. The results of the detailed similarity plot analysis of genome regions of 30,501–33,600 (Figure S3), RDP analysis (Table 1), and phylogenetic analysis of the fiber genes (Fig. 2d) suggested the third recombination event of HAdV-B16-ch79 and HAdV-B16-E26 from SAdV-35.

A previous report holds that HAdV-E4 evolved from recombination of HAdV-16 and a simian adenovirus SAdV-E26 [9]. Based on our result, most of the genome regions of HAdV-16 were very similar to HAdV-B comparing with other groups, and HAdV-16 clustered into HAdV-B clade in the phylogenetic tree of whole genome; this suggests that it is more likely that HAdV-16 evolved from HAdV-B with recombination of HAdV-E4, and the direction is not from HAdV-16 to HAdV-E4. Moreover, especially considering

the fact that HAdV-16 is a human adenovirus, there are few opportunities of co-infection with a simian adenovirus. In contrast, HAdV-E4 and most types of HAdV-B are important human respiratory pathogens which may more likely co-infect human beings to produce novel recombinant adenovirus.

Although SAdV-36 was less similar to HAdV-16 s than HAdV-E4 s based on nucleotide distance measuring (Figure S2A and S2B), it did cluster into a single branch closest to the clade of HAdV-16s and HAdV-E4s (Fig. 2c); further, the amino acid alignment of seven HVRs revealed the high similarity of HAdV-16 and SAdV-36. Therefore, HAdV-16 might have evolved from an ancestor that is common to HAdV-E and SAdV-36. It was also found that in many regions, HAdV-16 had less similarities with all of the other HAdV prototypes (Figure S1), indicating that more mutations still occur in these regions compared to those in counterparts of similar strains. This finding also supported the fact that HAdV-16 evolved from a relatively ancient ancestor. The evolutionary origin of any novel HAdV genomes might also be quickly identified in this manner.

The phylogenetic trees based on penton base genes, hexon genes, and fiber genes had different profiles compared to those of the whole genome tree, in which HAdV-21 and HAdV-50 of subspecies B1 were found to be contained in subclades with subspecies B2 strains (Fig. 2). Based on the phylogenetic tree of fiber genes, HAdV-7 of subspecies B1 formed a subclade with HAdV-11, HAdV-14, and HAdV-55 of subspecies B2. These results indicate the possible complexity of HAdV evolution, which might be related to recombination. Serum neutralization tests (SNs) have been used to differentiate and type HAdVs over the last century. However, SN depends on the neutralization epitopes that are mainly located in seven HVRs of hexon. Sometimes a small change in critical amino acid residues might contribute to a dramatic reversal of SN [33]. In this study, HAdV-16 was found to form a clade with HAdV-E4 based on the hexon region, which was also reported in some previous studies [9, 12–14]. However, SN data in early studies did not agree with the high similarity based on the hexon loop1 sequence between HAdV-16 and HAdV-E4. These studies reported that HAdV-16 and HAdV-E4 were neutralized by all heterologous sera tested with 4- to 32-fold lower heterologous titers compared to homologous titers [15–17]. Although the V1 to V3 region (loop1) of the HAdV-16 hexon gene was highly similar to that of HAdV-E4, a major variation in HVR7 (loop 2) was found between them. Our recent research identified this residue (AGSEK) in HVR7 as the critical neutralization epitope of HAdV-E4 [33]. This might explain the low-level cross-reaction between HAdV-16 and HAdV-E4. Further, the residue VAGTSGTQ in HVR7 might be critical for the neutralization of HAdV-16 which should be further confirmed. It would thus be interesting to detect the antisera cross-neutralization between HAdV-16 and SAdV-36. Moreover, there are some other epitopes of lesser importance that contribute to detectable cross-neutralization.

To the best of our knowledge, other than HAdV-E4, HAdV-16 is the only strain that evolved from two different HAdV species. The structural complexity and long-term adaptive evolution of adenoviruses may make interspecific recombination difficult. In our previous experiment, a chimeric adenovirus Ad3/H4 could not be rescued by replacing the HAdV-3 hexon gene with that of HAdV-E4. This result indicates an incompatibility with the hexon of other species, which could result in defective packaging [36]. This incompatibility agrees with a study on the HAdV-5 vector [46]. The potential incompatibility between various species of HAdV may influence Ad evolution. Based on the findings on HAdV-16 hexon, it might be possible to obtain a chimeric adenovirus Ad3/V4 by replacing the HAdV-3 hexon V1 to V3 region with the corresponding region of HAdV-E4. It is unclear whether this chimeric hexon would affect the stability and proliferation of HAdV-16 due to a lack of experimental data. However, these findings on HAdV-16 indicate that new Ad strains could still emerge from recombination between different species, thus highlighting the importance of continuous monitoring.

In conclusion, this study provides evidence of HAdV-16 molecular evolution through multiple recombination events, based on computational analysis. In addition, the variations in HVR7 between HAdV-16 and HAdV-E4 might have contributed to the low level of cross-neutralization. Our findings will enhance our understanding of HAdV evolution, monitoring newly emergent HAdV strains, and developing novel HAdV vaccines and vectors.

Acknowledgements This study was supported by grants from the National Key Research and Development Program of China (2018YFC1200100), National Natural Science Foundation of China (NSFC 31570163), and the Youth Project of State Key Laboratory of Respiratory Disease (SKLRD-QN-201713). We sincerely thank the reviewers for their comments, which have greatly improved the quality of this study.

Author contributions XT conceived and wrote the paper, HW performed most computational analyses and revised the paper, and RZ provided advice and supervised the work. All authors read and approved the final version of the manuscript.

Compliance with ethical standards

Conflicts of interest The authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. Espinola EE, Barrios JC, Russomando G, Mirazo S, Arbiza J (2017) Computational analysis of a species D human adenovirus provides evidence of a novel virus. J Gen Virol 98(11):2810–2820
2. Seto D, Chodosh J, Brister JR, Jones MS, Members of the Adenovirus Research C (2011) Using the whole-genome sequence to characterize and name human adenoviruses. J Virol 85:5701–5702
3. Yoshitomi H, Sera N, Gonzalez G, Hanaoka N, Fujimoto T (2017) First isolation of a new type of human adenovirus (genotype 79), species Human mastadenovirus B (B2) from sewage water in Japan. J Med Virol 89:1192–1200
4. Hashimoto S, Gonzalez G, Harada S, Oosako H, Hanaoka N, Hinokuma R, Fujimoto T (2018) Recombinant type Human mastadenovirus D85 associated with epidemic keratoconjunctivitis since 2015 in Japan. J Med Virol 90:881–889
5. Lenaerts L, De Clercq E, Naesens L (2008) Clinical features and treatment of adenovirus infections. Rev Med Virol 18:357–374
6. Sandkovsky U, Vargas L, Florescu DF (2014) Adenovirus: current epidemiology and emerging approaches to prevention and treatment. Curr Infect Dis Rep 16:416
7. Walsh MP, Seto J, Jones MS, Chodosh J, Xu W, Seto D (2010) Computational analysis identifies human adenovirus type 55 as a re-emergent acute respiratory disease pathogen. J Clin Microbiol 48:991–993
8. Yang Z, Zhu Z, Tang L, Wang L, Tan X, Yu P, Zhang Y, Tian X, Wang J, Zhang Y, Li D, Xu W (2009) Genomic analyses of recombinant adenovirus type 11a in China. J Clin Microbiol 47:3082–3090
9. Dehghan S, Seto J, Liu EB, Walsh MP, Dyer DW, Chodosh J, Seto D (2013) Computational analysis of four human adenovirus type 4 genomes reveals molecular evolution through two interspecies recombination events. Virology 443:197–207
10. Iampietro MJ, Larocca RA, Provine NM, Abbink P, Kang ZH, Bricault CA, Barouch DH (2018) Immunogenicity and cross-reactivity of rhesus adenoviral vectors. J Virol 92:e00159-18
11. Michael NL (2012) Rare serotype adenoviral vectors for HIV vaccine development. J Clin Invest 122:25–27
12. Madisch I, Harste G, Pommer H, Heim A (2005) Phylogenetic analysis of the main neutralization and hemagglutination determinants of all human adenovirus prototypes as a basis for molecular classification and taxonomy. J Virol 79:15265–15276
13. Pring-Akerblom P, Trijssenaar FE, Adrian T (1995) Sequence characterization and comparison of human adenovirus subgenus B and E hexons. Virology 212:232–236
14. Ebner K, Pinsker W, Lion T (2005) Comparative sequence analysis of the hexon gene in the entire spectrum of human adenovirus serotypes: phylogenetic, taxonomic, and clinical implications. J Virol 79:12635–12642
15. Rafajko RR (1964) Production and standardization of adenovirus types 1 to 18 reference antisera. Am J Hyg 79:310–319
16. Wigand R (1987) Pitfalls in the identification of adenoviruses. J Virol Methods 16:161–169
17. Hierholzer JC, Stone YO, Broderson JR (1991) Antigenic relationships among the 47 human adenoviruses determined in reference horse antisera. Arch Virol 121:179–197
18. Lin B, Wang Z, Vora GJ, Thornton JA, Schnur JM, Thach DC, Blaney KM, Ligler AG, Malanoski AP, Santiago J, Walter EA, Agan BK, Metzgar D, Seto D, Daum LT, Kruzelock R, Rowley RK, Hanson EH, Tibbetts C, Stenger DA (2006) Broad-spectrum respiratory tract pathogen identification using resequencing DNA microarrays. Genome Res 16:527–535
19. Tcherepanov V, Ehlers A, Upton C (2006) Genome Annotation Transfer Utility (GATU): rapid annotation of viral genomes using a closely related reference genome. BMC Genomics 7:150
20. Lefort V, Longueville JE, Gascuel O (2017) SMS: smart Model Selection in PhyML. Mol Biol Evol 34:2422–2424
21. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol 59:307–321
22. Letunic I, Bork P (2016) Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res 44:W242–W245
23. Martin DP, Murrell B, Golden M, Khoosal A, Muhire B (2015) RDP4: detection and analysis of recombination patterns in virus genomes. Virus Evol 1:vev003
24. Martin D, Rybicki E (2000) RDP: detection of recombination amongst aligned sequences. Bioinformatics 16:562–563
25. Padidam M, Sawyer S, Fauquet CM (1999) Possible emergence of new geminiviruses by frequent recombination. Virology 265:218–225
26. Smith JM (1992) Analyzing the mosaic structure of genes. J Mol Evol 34:126–129
27. Martin DP, Posada D, Crandall KA, Williamson C (2005) A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints. AIDS Res Hum Retrovir 21:98–102
28. Gibbs MJ, Armstrong JS, Gibbs AJ (2000) Sister-scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. Bioinformatics 16:573–582
29. Boni MF, Posada D, Feldman MW (2007) An exact nonparametric method for inferring mosaic structure in sequence triplets. Genetics 176:1035–1047
30. Gall J, Kass-Eisler A, Leinwand L, Falck-Pedersen E (1996) Adenovirus type 5 and 7 capsid chimera: fiber replacement alters receptor tropism without affecting primary immune neutralization epitopes. J Virol 70:2116–2123
31. Roy S, Clawson DS, Calcedo R, Lebherz C, Sanmiguel J, Wu D, Wilson JM (2005) Use of chimeric adenoviral vectors to assess capsid neutralization determinants. Virology 333:207–214
32. Sumida SM, Truitt DM, Lemckert AA, Vogels R, Custers JH, Addo MM, Lockman S, Peter T, Peyerl FW, Kishko MG, Jackson SS, Gorgone DA, Lifton MA, Essex M, Walker BD, Goudsmit J, Havenga MJ, Barouch DH (2005) Neutralizing antibodies to adenovirus serotype 5 vaccine vectors are directed primarily against the adenovirus hexon protein. J Immunol 174:7179–7185
33. Tian X, Qiu H, Zhou Z, Wang S, Fan Y, Li X, Chu R, Li H, Zhou R, Wang H (2018) Identification of a critical and conformational neutralizing epitope in human adenovirus type 4 hexon. J Virol 92:e01643-17
34. Wu H, Dmitriev I, Kashentseva E, Seki T, Wang M, Curiel DT (2002) Construction and characterization of adenovirus serotype 5 packaged by serotype 3 hexon. J Virol 76:12775–12782
35. Yu B, Dong J, Wang C, Zhan Y, Zhang H, Wu J, Kong W, Yu X (2013) Characteristics of neutralizing antibodies to adenovirus capsid proteins in human and animal sera. Virology 437:118–123
36. Tian X, Su X, Li H, Li X, Zhou Z, Liu W, Zhou R (2011) Construction and characterization of human adenovirus serotype 3 packaged by serotype 7 hexon. Virus Res 160:214–220
37. Bradley RR, Maxfield LF, Lynch DM, Iampietro MJ, Borducchi EN, Barouch DH (2012) Adenovirus serotype 5-specific neutralizing antibodies target multiple hexon hypervariable regions. J Virol 86:1267–1272
38. Pichla-Gollon SL, Drinker M, Zhou X, Xue F, Rux JJ, Gao GP, Wilson JM, Ertl HC, Burnett RM, Bergelson JM (2007) Structure-based identification of a major neutralizing site in an adenovirus hexon. J Virol 81:1680–1689
39. Qiu H, Li X, Tian X, Zhou Z, Xing K, Li H, Tang N, Liu W, Bai P, Zhou R (2012) Serotype-specific neutralizing antibody epitopes

of human adenovirus type 3 (HAdV-3) and HAdV-7 reside in multiple hexon hypervariable regions. J Virol 86:7964–7975
40. Rux JJ, Kuser PR, Burnett RM (2003) Structural and phylogenetic analysis of adenovirus hexons by use of high-resolution x-ray crystallographic, molecular modeling, and sequence-based methods. J Virol 77:9553–9566
41. Tian X, Ma Q, Jiang Z, Huang J, Liu Q, Lu X, Luo Q, Zhou R (2015) Identification and application of neutralizing epitopes of human adenovirus type 55 hexon protein. Viruses 7:5632–5642
42. Yuan X, Qu Z, Wu X, Wang Y, Liu L, Wei F, Gao H, Shang L, Zhang H, Cui H, Zhao Y, Wu N, Tang Y, Qin L (2009) Molecular modeling and epitopes mapping of human adenovirus type 3 hexon protein. Vaccine 27:5103–5110
43. Li QG, Wadell G (1988) The degree of genetic variability among adenovirus type 4 strains isolated from man and chimpanzee. Arch Virol 101:65–77
44. Kajon AE, Lamson DM, Bair CR, Lu X, Landry ML, Menegus M, Erdman DD, St. George K (2018) Adenovirus type 4 respiratory infections among civilian adults, Northeastern United States, 2011-2015(1). Emerg Infect Dis 24:201–209
45. Kajon AE, Dickson LM, Murtagh P, Viale D, Carballal G, Echavarria M (2010) Molecular characterization of an adenovirus 3-16 intertypic recombinant isolated in Argentina from an infant hospitalized with acute respiratory infection. J Clin Microbiol 48:1494–1496
46. Youil R, Toner TJ, Su Q, Chen M, Tang A, Bett AJ, Casimiro D (2002) Hexon gene switch strategy for the generation of chimeric recombinant adenovirus. Hum Gene Ther 13:311–320

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.