Infection, Genetics and Evolution 76 (2019) 104035

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Research paper

# Multiple divergent *Human mastadenovirus C* co-circulating in mainland of China

Naiying Mao[a,1], Zhen Zhu[a,1], Pierre Rivailler[a,1], Jianfang Yang[b], Qi Li[c], Guangyue Han[c], Jie Yin[d], Deshan Yu[e], Liwei Sun[f], Hongbo Jiang[f], Zhifei Zhan[g], Xingyu Xiang[g], Hong Mei[h], Xianjun Wang[i], Bo Zhang[i], Pengbo Yu[j], Hong Li[a,k], Zhenqiang Lei[a], Wenbo Xu[a,*]

[a] *WHO WPRO Regional Reference Measles, Rubella Laboratory and Key Laboratory of Medical Virology Ministry of Health, National Institute for Viral Disease Control and Prevention, Chinese Centre for Disease Control and Prevention, Beijing, China*
[b] *Shanxi Provincial Center for Disease Control and Prevention, Taiyuan, China*
[c] *Hebei Provincial Center for Disease Control and Prevention, Shijiazhuang, China*
[d] *Yunnan Provincial Center for Disease Control and Prevention, Kunming, China*
[e] *Gansu Provincial Center for Disease Control and Prevention, Lanzhou, China*
[f] *Changchun Children's Hospital, Changchun, China*
[g] *Hunan Provincial Center for Disease Control and Prevention, Changsha, China*
[h] *Xizang Provincial Center for Disease Control and Prevention, Lasa, China*
[i] *Shandong Provincial Center for Disease Control and Prevention, Jinan, China*
[j] *Shaanxi Provincial Center for Disease Control and Prevention, Xian, China*
[k] *Anhui University of Science and Technology, Huainan, China*

ARTICLE INFO

*Keywords:*
*Human mastadenovirus C*
Molecular evolution
Genotype
Homologous recombination

ABSTRACT

The *human mastadenovirus C* (HAdV-C) cause respiratory infections in children. Homologous recombination was clearly involved in the molecular evolution of HAdV-A, B, and D, but little is known about the molecular evolution of HAdV-C. From 2000 to 2016, 201 HAdV-C strains were collected from nine provinces covering six administrative regions of mainland of China *via* 3 existing surveillance programs, namely the febrile respiratory syndrome surveillance, the acute flaccid paralysis surveillance, and the hand, foot, and mouth disease surveillance system. The genes coding for the capsid protein (penton base, hexon, and fiber) of 201 HAdV-C strains were sequenced and compared with representative sequences publicly available. In addition, the whole genome sequence of 24 representative strains of HAdV-C was generated for further recombination analysis. Phylogenetic analysis of the penton base sequences of HAdV-C revealed six genetic groups (labelled as Px1–6), which showed that the penton base had more variation than previously thought. Based on the penton base, hexon, and fiber gene sequences, 16 new genetic patterns of HAdV-C circulating in mainland of China were identified in this study. Whole genome sequence analysis revealed frequent recombination events among HAdV-C genomes. This study is highly beneficial for case classification, tracking the transmission chain, and further epidemiological exploration of HAdV-C-related severe clinical diseases in the near future. Our data demonstrated that multiple newly divergent HAdV-C co-circulated across mainland China during the research period.

## 1. Introduction

The human adenovirus (HAdV) is a non-enveloped, double-stranded DNA virus belonging to the family of *Adenoviridae* within the genus of *Mastadenovirus* (Benko and Harrach, 2005; Wold et al., 2013). The viral capsid mainly comprises three proteins, hexon, penton base, and fiber (Crawford-Miksza and Schnurr, 1996), whose names are currently used for the HAdV nomenclature proposed by the international HAdV working group (HAdVWG) (Seto et al., 2011). Originally, HAdVs were formally divided into 7 species (A to G) with 52 serotypes based on serum neutralisation and hemagglutination inhibition assays (Benko and Harrach, 2005; Harrach et al., 2011). As of July 2019, 51 additional

* Corresponding author at: WHO WPRO Regional Reference Measles, Rubella Laboratory and Key Laboratory of Medical Virology Ministry of Health, National Institute for Viral Disease Control and Prevention, Chinese Centre for Disease Control and Prevention, No.155, Changbai Road, Changping District, Beijing 102206, China.
*E-mail address:* wenbo_xu1@aliyun.com (W. Xu).
[1] Naiying Mao, Zhen Zhu and Pierre Rivailler contributed equally to this work.

https://doi.org/10.1016/j.meegid.2019.104035
Received 11 April 2019; Received in revised form 4 September 2019; Accepted 5 September 2019
Available online 07 September 2019

HAdV genotypes (HAdV-53 to HAdV-103) have been identified based on genomic and bioinformatic analysis (http://hadvwg.gmu.edu). Most of the new genotypes originate from homologous recombinations of viruses within the same species of HAdV (Ismail et al., 2018). Homologous recombination is an important driving force for molecular evolution of HAdV-A, B, and D (Hage et al., 2017; Seto et al., 2013; Singh et al., 2013; Yang et al., 2009). However, the limited genome sequence data available for HAdV-C prevents any such presumptions (Ismail et al., 2018; Walsh et al., 2011). HAdV now poses a challenge worldwide due to the emergence of novel recombinant viruses (Cao et al., 2014; Cook and Radke, 2017).

HAdV infection is associated with a broad spectrum of clinical diseases including conjunctivitis (species B, D and E), gastroenteritis (species F and G), respiratory infections (species B and C), cystitis (species A and B), and encephalitis (species B and C) (Berk, 2007). HAdV-C is generally prevalent and commonly associated with respiratory tract infection among pediatric patients (Edwards et al., 1985; Scott et al., 2016). Since HAdV-C viruses can shed from the respiratory and gastrointestinal tracts for weeks after infection, viral nucleic acids can be detected from respiratory as well as gastrointestinal specimens of patients and asymptomatic carriers (Garnett et al., 2002). Consequently, HAdV-C derived from feces can also reflect the prevalence of viruses in the population. Since HAdV-C viruses have the ability to establish persistent infection, patients can remain asymptomatic carriers until at least their young adulthood (Garnett et al., 2002). Six types of species C including C1, C2, C5, C6, C57, and C89 have been identified. Among them, C1, C2, C5 and C6 were identified in the USA in 1953, C57 was found in Azerbaijan in 2001 and C89 was identified in 2015 in Germany (Dhingra et al., 2019; Berk, 2007; Walsh et al., 2011). In contrast to other species, penton base gene of species C is highly conserved and not divergent enough to be informative (Ismail et al., 2018; Mao et al., 2017; Rivailler et al., 2019).

No dedicated surveillance for HAdV is established in China. In order to better understand the prevalence and molecular evolution of HAdV-C circulating in mainland of China, HAdV-C strains were obtained through three existing surveillance programs. The respiratory specimens were collected from the febrile respiratory syndrome (FRS) surveillance program, whereas gastrointestinal specimens were obtained through the acute flaccid paralysis (AFP) surveillance program, as well as the hand, foot and mouth disease (HFMD) surveillance program. In the present study, 201 HAdV-C strains were collected from nine provinces, covering all six administrative regions of mainland of China during the period 2000–2016. The penton base, hexon, and fiber gene sequences of 201 strains were determined for genotype identification, and the whole genome sequences (WGSs) of 24 representative strains were generated for recombination analysis.

## 2. Materials and methods

### 2.1. Ethics statement and study design

This study was approved by the second session of Ethics Review Committee of the National Institute for Viral Disease Control and Prevention at China Center for Disease Control and Prevention. All methods were performed in accordance with the relevant guidelines and regulations.

### 2.2. Source of HAdV-C strains

Two-hundred-and-one HAdV-C strains were identified as part of the FRS, AFP, and HFMD surveillance programs in China in the period 2000–2016. Eighty-one strains from the FRS program were isolated from throat swabs, nasopharyngeal secretions, broncho-alveolar lavage fluid, and cerebro-spinal fluid (CSF). One hundred eighteen strains were isolated from stool samples of healthy children and AFP cases. Two strains were isolated from stool samples of HFMD cases. The distribution of collection times and locations is shown in Supplementary Fig. 1, and the 201 samples are listed in Supplementary Table 1.

### 2.3. Virus isolation and sequencing

All 201 strains were obtained after three passages in the human rhabdomyosarcoma (RD) or the laryngeal carcinoma (Hep-2) cell line. The viral DNA was extracted using a QIAamp DNA mini kit (Qiagen, Valencia, CA, USA) following the manufacturer's instructions. Primer pairs were designed and used to amplify the genes of penton base, hexon, and fiber, as well as the whole genome sequence (WGS), based on conserved sequences of prototype HAdV-C types (Supplementary Table 2). Overlapping PCR fragments were amplified, purified, and sequenced according to the previous report (Mao et al., 2017). Sequences were assembled and edited using Sequencher 5.0 (Genecodes Corp, Ann Arbor, MI, USA). In order to obtain high quality data, a minimum of 3-fold coverage for both directions across the genome was generated. In total, the penton base, hexon, and fiber gene sequences of 201 strains and 24 WGSs from 20 genetic groups were generated. All sequences were submitted to GenBank under the following accession numbers: MH322458-MH322473 and MH322475-MH322659 for penton base gene, MH322256-MH322271 and MH322273-MH322457 for hexon gene, and MH322054-MH322069 and MH322071-MH322255 for fiber gene, MK041225-MK041248 for WGS. The WGS of 24 representative strains were also submitted as accession numbers MK041225-MK041248.

### 2.4. Dataset

WGS for 84 of the 103 HAdV genotypes were downloaded from GenBank. The penton base, hexon, and fiber gene sequences of these viral genomes were extracted using Artemis (Rutherford et al., 2000). A total of 2254 HAdV-C sequences were obtained from GenBank (August 2018). Sequences shorter than 1000 nucleotides, patented sequences, sequences derived from modified viruses, sequences of viruses collected in a non-human host, as well as sequences with unknown collection years were excluded. Sequences encoding at least 90% of penton base (> 1545 nt), hexon (> 2617 nt), or fiber (> 1574) open reading frames (ORFs) were selected. The sequences of penton base, hexon, and fiber genes of 201 HAdV-C strains collected in mainland of China were added to the datasets. The final datasets comprised 240 sequences for the penton base gene, 256 sequences for the hexon gene, and 248 sequences for the fiber gene. Each dataset was realigned at the amino acid level, and a nucleotide sequence alignment was re-generated using PAL2NAL tool (Suyama et al., 2006). Alignments were edited in BioEdit (www.mbio.ncsu.edu/BioEdit/bioedit.html). The WGS dataset was established with 24 WGS generated in this study and 25 HAdV-C WGS from the GenBank database.

### 2.5. Phylogenetic analysis

Phylogenetic trees were generated in Mega6 using the neighbor-joining (NJ) method and the maximum composite likelihood nucleotide substitution model (Tamura et al., 2011; Tamura et al., 2013). Maximum likelihood (ML) phylogenetic trees were also generated. Phylogenetic inference was tested by the bootstrap method with 1000 replicates (Saitou and Nei, 1986). Bootstrap values > 70% were shown. Genetic distances were computed in Mega6. Sequences from GenBank were identified by their accession number, country, and year of collection. *De novo* sequences were identified by the name of the province where the viral sample was collected, year of collection, and sample number.

### 2.6. Recombination analysis

Genetic components potentially involved in recombination events were identified using pairwise genetic distances computed on WGSs, as well distinct genomic regions (1–7000, 7001–14,150, 15,867–18,837, 21,745–26,000, 26,001–31,029 and 32,779–end), as previously described (Mao et al., 2017; Rivailler et al., 2019). The numbering was based on the NC_001405 genome. The genomes potentially involved in recombination events were then assessed using BootScan from the SimPlot package, as well as the Recombination Detection Program RDP4 package (Lole et al., 1999; Martin et al., 2015).

## 3. Results

### 3.1. Phylogenetic analysis based on penton base, hexon, and fiber genes

Two-hundred-and-one penton base, hexon, and fiber gene sequences were analysed in the context of 84 representative sequences of seven species of HAdVs. Phylogenetic analysis showed that all 201 viruses belong to species C (Supplementary Fig. 2). No evidence of recombination was found between the 201 HAdV-C and other adenovirus species. The overall genetic pairwise p-distance for HAdV-C hexon is 10.5%, which is in the same range as HAdV-D and HAdV-B hexon genes at 9.1% and 13.5%, respectively (Table 1, Supplementary Table 3). The overall p-distance of HAdV-C fiber was 19.1%, which is comparable to the overall p-distance of HAdV-B and HAdV-D fiber genes at 26.1% and 26.6%, respectively (Table 1). In contrast, the overall p-distance of HAdV-C penton base was 1.2%, whereas the p-distances in the HAdV-D and HAdV-B counterparts were 6.7% and 11.2%, respectively (Table 1).

Phylogenetic trees featuring the 201 *de novo* sequences from China were generated in the context of publicly available sequences, using NJ and ML methods (Supplementary Figs. 3 and 4). For illustration purposes, a subset of 125 sequences was used to generate NJ trees featured in Fig. 1, but most of the major nodes were confirmed by both phylogenetic methods. The NJ tree of fiber sequences featured four major nodes, supported by a 99% bootstrap value corresponding to the four established types, namely F1 (depicted in blue), F2 (green), F5 (purple), and F6 (brown), as shown in Fig. 1. The genetic distance within types was < 1.1%, whereas the smallest distance between types was 22.4% (Table 1) meaning that fiber sequence were well conserved across the types but with high divergence between them.

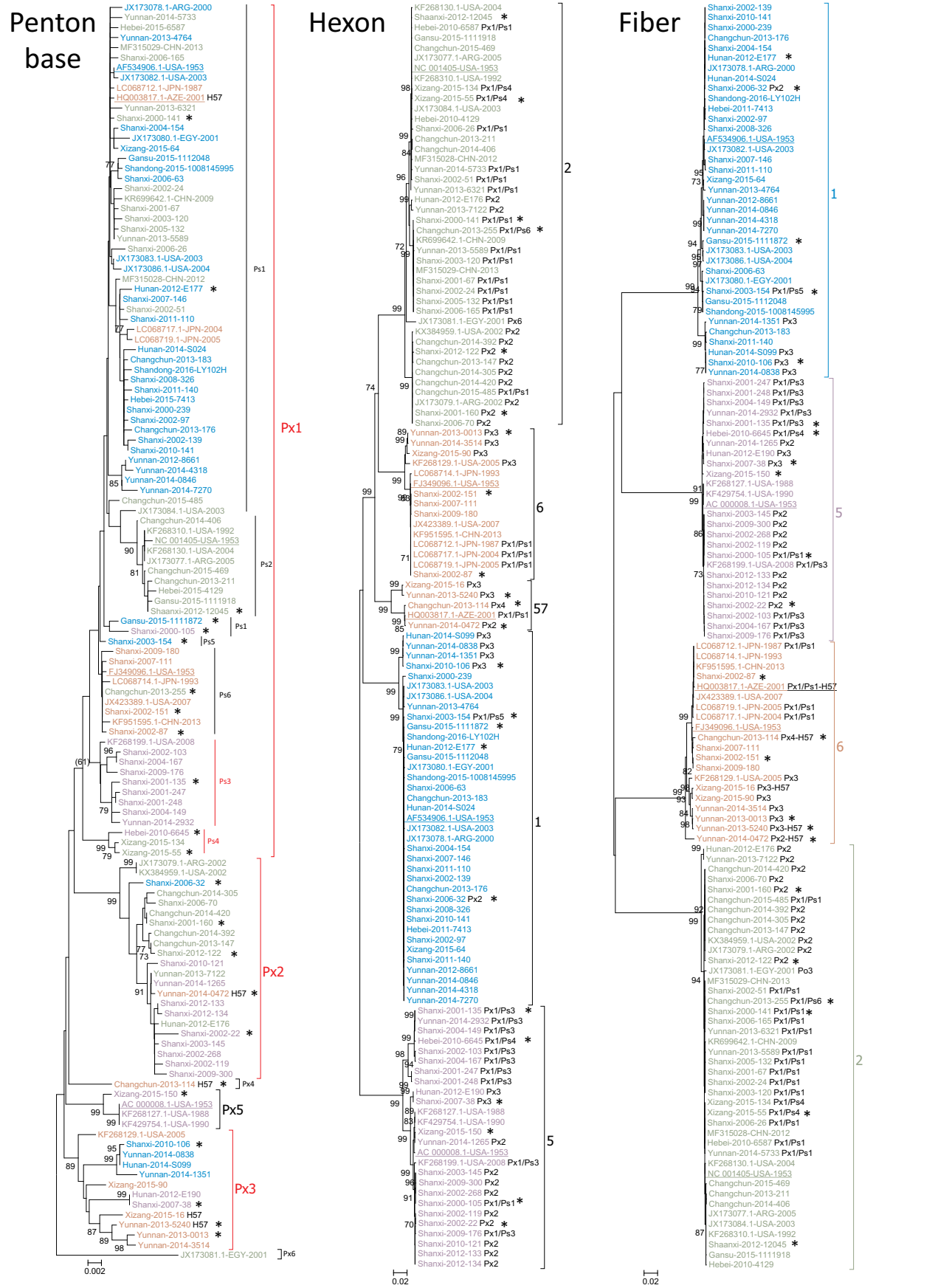
Similarly to the fiber gene, the NJ tree of hexon sequences featured five well-defined genetic clusters, each corresponding to the established hexon type, H1, H2, H5, H6, and H57 (Fig. 1). Except for H57, all viruses had the same type for fiber and hexon but genotype 57 was characterised by F6 and H57. The genetic distance within each type was < 1.5%, with the exception of H57 (2.4%) and H5 (1.6%) (Table 1). In both cases, the divergence was mostly due to a couple of sequences. As more sequences become available, it might be necessary to revisit both types and identify a sub-type within these types. The smallest genetic distance between types was 9.3%, meaning that hexon types were highly divergent between each other (Table 1), and that the five types were completely representative of the genetic diversity of the hexon gene.

The 125 penton base sequences exhibited a significant heterogeneity with a mean genetic distance of 1.2% (Fig. 1, Table 1). Three major nodes, supported by a bootstrap value of 87% or greater, could be identified in the NJ as well as the ML trees (Fig. 1, Supplementary Figs. 3 and 4). Among the 4 previously identified penton-base types (P1, P2, P5 and P6), only P5 was found in a genetic group with a bootstrap value > 70% (99% in the case of P5). Consequently, we arbitrarily decided to develop a new nomenclature and use Px to label the main penton-base genetic groups as well as outlier sequences. The group Px5 contains 4 sequences, including the sequence of type 5 prototype AC_000008 as well as two other USA sequences (KF268127-USA-1988 and KF429754-USA-1990) and one *de novo* sequence from Xizang province of China (Fig. 1). The other 2 main genetic groups were called Px2 and Px3 and the remaining of the sequences was considered as one genetic group called Px1. The group Px2 was supported by a bootstrap value of 99%. It contained 27 unique sequences including two GenBank sequences, namely JX173079-ARG-2002 and KX384959-USA-2002 (Supplementary Fig. 3). The sequences of this group showed a limited divergence among themselves with a p-distance of 0.4% (Table 1). The group Px3 contained 16 sequences including a GenBank sequence KF268129-USA-2005 and 15 sequences from this study (Supplementary Fig. 3). This group was more heterogeneous with a p-distance of 0.9% (Table 1). Finally, the group Px1 was not as well supported, with a bootstrap value of 61% (Fig. 1). Interestingly, this group appeared polyphyletic in the ML tree confirming that this group might not be

**Table 1**
Average genetic distance within and between penton base, hexon, and fiber genetic groups.

| Genetic group [a,b] | Number of analyzed sequences | p-distance within group [c] | Lowest distance between groups [d] |
|---|---|---|---|
| Px1/Ps1 | 53 | 0.4 | 1.6 |
| Px1/Ps2 | 10 | 0.1 | |
| Px1/Ps6 | 9 | 0.1 | |
| Px1/Ps3 | 9 | 0.4 | |
| Px1/Ps4 | 3 | 0.3 | |
| All Px1 sequences | 85 | 0.6 | |
| Px5 | 4 | 0.2 | |
| Px2 | 22 | 0.4 | |
| Px3 | 12 | 0.9 | |
| All penton base sequences | 125 | 1.2 [e] | |
| H1 | 37 | 0.4 | 9.3 |
| H2 | 42 | 1.3 | |
| H5 | 26 | **1.6** | |
| H6 | 15 | 0.7 | |
| H57 | 5 | **2.4** | |
| All hexon sequences | 125 | 10.5 [f] | |
| F1 | 37 | 0.9 | 22.4 |
| F2 | 42 | 0.3 | |
| F5 | 26 | 0.4 | |
| F6 | 20 | 1.1 | |
| All fiber sequences | 125 | 19.1 [g] | |

[a]Average genetic distances were computed using the penton base, hexon, and fiber gene sequences of 125 viral genomes featured in Fig. 1.
[b]Genetic groups previously identified for the penton base genes are highlighted in grey.
[c]Distances within group > 1.5 are in bold font.
[d]Only Px1, Px2, Px3, and P5 genetic groups were considered to compute the distance between penton base genetic groups.
[e]Overall p-distances for penton base species B and D are 11.2% and 6.7%, respectively.
[f]Overall p-distances for hexon species B and D are 13.5% and 9.1%, respectively.
[g]Overall p-distances for fiber species B and D are 26.1% and 26.6%, respectively.




*(caption on next page)*

**Fig. 1.** Neighbor-joining phylogenetic trees of 125 sequences from HAdV-C penton base, hexon, and fiber genes. *De novo* sequences of viruses collected in mainland of China are identified by the province name, collection year, and sample number. The other sequences are identified by their GenBank ID, country of collection, and collection year. Sequences are colour-coded based on the fiber gene type, blue for type 1, green for type 2, purple for type 5, and brown for type 6. The main genetic groups for hexon and fiber genes are shown as a thick bracketed line. The reference strain for each type is underlined. For the penton base tree, four major genetic groups (x1, x2, x3 and x5) are identified by brackets. The subgroups of sequences within x1, Px1/Ps1, Px1/Ps2 and Px1/Ps6 correspond to the previously identified types 1, 2, and 6. New genetic groups, subgroups, and outlier sequences are identified by red labels. Bootstrap values > 70% are indicated. When necessary, bootstrap values < 70% are depicted in parentheses. Penton base and hexon genetic groups are indicated in black after the virus name for genotyping purposes. Genetic groups are listed in Fig. 2. Chinese viruses, for which WGSs have been generated in this study, are depicted by a black star. Trees featuring all sequences generated in this study in the context of previously reported sequences are shown in Supplementary Figs. 3 and 4. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

evolutionary relevant. Because the group Px1 contained sequences of what was previously known as P1, P2 and P6, we felt compelled to consider these subgroups despite not being well supported. These subgroups were labelled Px1/Ps1, Px1/Ps2 and Px1/Ps6 for respectively the previously known types P1, P2 and P6. Two additional subgroups called Px1/Ps3 and Px1/Ps4 were identified. Ps3 contained 14 sequences, including the GenBank sequence KF268199-USA-2008 (Supplementary Fig. 3). Ps4 contained 3 *de novo* sequences and was supported by a bootstrap value of 99% (Fig. 1). Ps4 was extremely well supported in the NJ and the ML trees and was distinct of the Px1 group in the ML tree suggesting that this subgroup could be considered as an independent group (shown with an arrow in Supplementary Fig. 4). However, the limited number of sequences (3) as well as their novelty (all *de novo* sequences) and the discrepancy between ML and NJ suggested that these sequences should be considered with caution. The relevance of this subgroup should be revisited once more related sequences become available. The *de novo* sequence Shanxi-2003-154 did not fit any sequence cluster within group Px1 and was called Px1/Ps5. Two additional sequences, Changchun-2013-114 and JX173081-EGY-2001, were highly divergent compared to all the other penton base sequences and were called Px4 and Px6 respectively (Fig. 1). In order to rule out any potential recombination within the penton base gene, a SimPlot analysis was performed and did not detect any significant recombination within penton-base gene sequences (Supplementary Fig. 5). The lowest genetic distance between the 4 sequence groups (Px1, Px5, Px2, and Px3) was 1.5–1.6%, which meant that these groups were significantly divergent between each other (Table 1, Supplementary Table 3). According to the colour based on the fiber type, most of the genetic groups identified in the penton base phylogenetic tree featured 2, 3, or even 4 colors (Fig. 1), and hence this complex situation provided evidence of recombination between genomes.

In addition, annotating hexon and fiber sequences with their respective penton base genetic group allowed the identification of subgroups within hexon and fiber groups. For example, a subgroup of six sequences in H1, supported by a bootstrap value of 99%, shared the penton base genetic group Px3 (Supplementary Figs. 3 and 4). Similar subgroups can be also identified within H2, H5 and H6.

### 3.2. Genetic patterns among HAdV-C

The combination of 5 types of hexon and 4 types of fiber sequences, as well as 10 genetic groups and subgroups of the penton base sequence resulted in the identification of 20 distinct genetic patterns (Fig. 2). Among these 20 identified genetic patterns, the distribution based on fiber type was relatively equivalent, 4 genetic patterns for F1, 5 genetic patterns for F2 and F6, and 6 genetic patterns for F5. Interestingly, 5 hexon/fiber combinations (H1F1, H2F2, H5F5, H6F6, and H57F6) were found. Based on the analysed data, it seems that hexon and fiber tend to be genetically associated. Among the 201 virus sequences from China reported in this study, one quarter (53 strains) belonged to Px1/Ps1H1F1, represented by the prototype AF534906-USA-1953. Another quarter (49 strains) belonged to Px1/Ps1H2F2, which is likewise represented by the previously reported genome JX173084-USA-2003. Among the 20 genetic patterns, only 4 had been previously reported (C1, C2, C5, and C6) (Fig. 2A). The other 16 new genetic patterns comprised 4 genomes that were previously reported: JX173084-USA-2003 with Px1/Ps1H2F2 (previously reported as P2H2F2), KX384959-USA-2002 with Px2H2F2, KF268199-USA-2008 with Px1/Ps3H5F5 (previously reported as P2H5F5), and KF268129-USA-2005 with Px3H6F6 (previously reported as P6H6F6). Some genetic patterns were found in multiple administrative regions of China (Fig. 2B). For example, Px1/Ps1H1F1 was found in all 6 administrative regions. Px1/Ps1H2F2, Px1/Ps2H2F2, and Px2H2F2 were found in 4 administrative regions, whereas Px3H1F1 and Px1/Ps3H5F5 were found in 3 administrative regions. Six genetic patterns were represented by only one sequence. Twelve genetic patterns concerned viruses only collected in China. Px2H5F5 was the most sampled, mainly in the Shanxi province during the last decade (2002–2014) (Fig. 2B and C). It is worth noting that the Px1/Ps1H57F6 represented by HQ003817 AZE-2001, known as prototype 57, was not detected in China.

### 3.3. Association with clinical symptoms

Among the 201 patients of the present study, 81 patients identified *via* the FRS surveillance program were clinically diagnosed (Table 2). Half of these patients (40) were diagnosed with an upper respiratory tract infection (URI), whereas the other half (39) was diagnosed with a lower respiratory tract infection (LRI). One patient was diagnosed with enteritis [Px1/Ps6H2F2] and another with encephalitis [Px1/Ps1H1F1]. Interestingly, one adenovirus strain was isolated from the cerebro-spinal fluid (CSF) of this patient, demonstrating that this adenovirus infection was associated with encephalitis symptoms. No particular correlation was found between a particular genotype and the distribution of LRI and URI cases. Similarly, no correlation between HAdV-C and AFP or HFMD was found.

### 3.4. Whole genome analysis and recombination characteristics

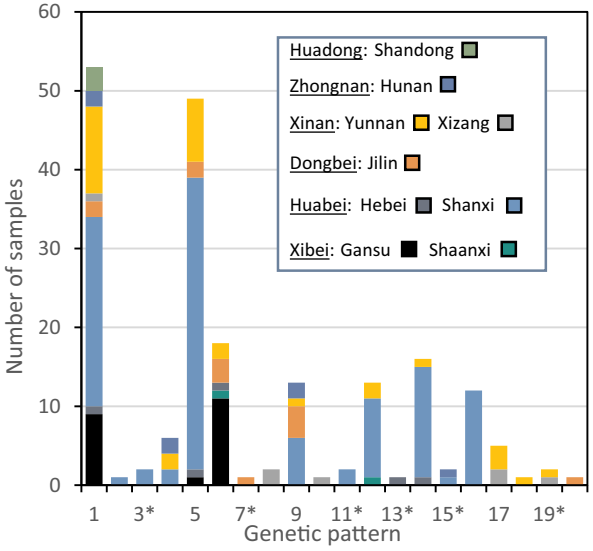
Among the 201 analysed viruses, WGS of 24 strains representing 20 genetic patterns were obtained (Fig. 2A). The average genome length (35,923 nt) was in the range of the previously reported genome length (35,878 +/− 101 nt), and the average GC% content of the genome sequences was 55.27%, similar to the value of previously reported genomes (55.26%). As expected, the 38 genes annotated in NC_001405 were featured in the 24 *de novo* genomes. As the WGSs of 2 strains (Shanxi-2000-105 and Shanxi-2000-106) were identical, only 23 genomes were analysed for potential recombination events and compared to 25 genomes previously reported in GenBank.

A phylogenetic network featured at least 4 distinct clusters corresponding to the known genotypes 1, 2, 5, and 6 (Fig. 3). However, many distinct nodes could be identified, indicating multiple recombination events among these genomes. Furthermore, 7 sequences (Shanxi-2003-154, JX173080–2001-EGY, Shanxi-2010-106, Yunnan-2014-0472, Yunnan-2013-5240, Yunnan-2013-0013, and KF268129–2005-USA) did not fit any main clusters. Phylogenetic trees were built for WGS, as well as for 9 consecutive genomic regions in order to identify potential recombination events (Supplementary Fig. 6). Pairwise genetic p-distances were computed for the corresponding phylogenetic trees (Supplementary Tables 4–14). The lowest pairwise distances obtained from the WGS and the 9 genomic regions were used to identify genetic

A

| Genetic pattern | Penton base | Hexon | Fiber | Collection year | Collection country | Representative virus | Virus chosen for WGS | Known type |
|---|---|---|---|---|---|---|---|---|
| 1 | Px1/Ps1 | H1 | F1 | 1953-2016 | Argentina, China, Egypt, USA | AF534906-USA-1953 | Hunan-2012-E177<br>Gansu-2015-1111872 | C1 |
| **2** | **Px1/Ps5** | **H1** | **F1** | **2003** | **China** | **Shanxi-2003-154** | **Shanxi-2003-154** | |
| **3** | **Px2** | **H1** | **F1** | **2006** | **China** | **Shanxi-2006-32** | **Shanxi-2006-32** | |
| **4** | **Px3** | **H1** | **F1** | **2010-2014** | **China** | **Shanxi-2010-106** | **Shanxi-2010-106** | |
| 5 | Px1/Ps1 | H2 | F2 | 2000-2015 | China, USA | JX173084-USA-2003 | Shanxi-2000-141 | |
| 6 | Px1/Ps2 | H2 | F2 | 1953-2016 | Argentina, China, USA | NC_001405-USA-1953 | Shaanxi-2012-12045 | C2 |
| **7** | **Px1/Ps6** | **H2** | **F2** | **2013** | **China** | **Changchun-2013-255** | **Changchun-2013-255** | |
| **8** | **Px1/Ps4** | **H2** | **F2** | **2015** | **China** | **Xizang-2015-55** | **Xizang-2015-55** | |
| 9 | Px2 | H2 | F2 | 2001-2015 | Argentina, China, USA | KX384959-USA-2002 | Shanxi-2012-122<br>Shanxi-2001-160 | |
| 10 | Px5 | H5 | F5 | 1953-2015 | USA, China | AC 000008-USA-1953 | Xizang-2015-150 | C5 |
| **11** | **Px1/Ps1** | **H5** | **F5** | **2000** | **China** | **Shanxi-2000-105** | **Shanxi-2000-105/106** | |
| 12 | Px1/Ps3 | H5 | F5 | 2001-2014 | China, USA | KF268199-USA-2008 | Shanxi-2001-135 | |
| **13** | **Px1/Ps4** | **H5** | **F5** | **2010** | **China** | **Hebei-2010-6645** | **Hebei-2010-6645** | |
| **14** | **Px2** | **H5** | **F5** | **2002-2014** | **China** | **Shanxi-2002-22** | **Shanxi-2002-22** | |
| **15** | **Px3** | **H5** | **F5** | **2007-2012** | **China** | **Shanxi-2007-38** | **Shanxi-2007-38** | |
| 16 | Px1/Ps6 | H6 | F6 | 1953-2009 | China, USA | FJ349096-USA-1953 | Shanxi-2002-151<br>Shanxi-2002-087 | C6 |
| 17 | Px3 | H6 | F6 | 2005-2015 | China, USA | KF268129-USA-2005 | Yunnan-2013-0013 | |
| **18** | **Px2** | **H57** | **F6** | **2014** | **China** | **Yunnan-2014-0472** | **Yunnan-2014-0472** | |
| **19** | **Px3** | **H57** | **F6** | **2013-2015** | **China** | **Yunnan-2013-5240** | **Yunnan-2013-5240** | |
| **20** | **Px4** | **H57** | **F6** | **2013** | **China** | **Changchun-2013-114** | **Changchun-2013-114** | |

B




C

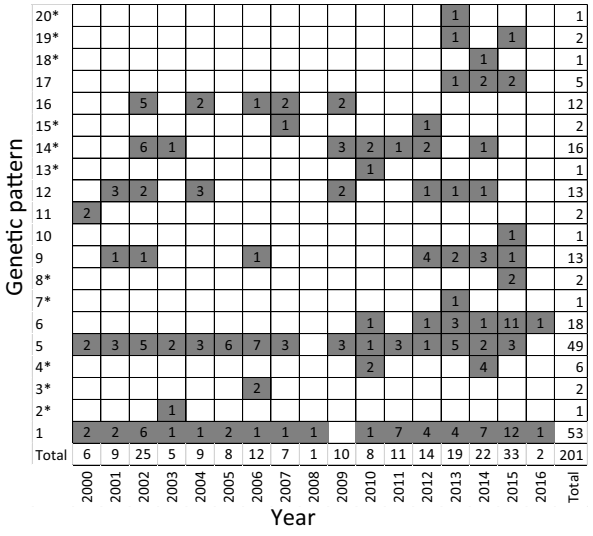



**Fig. 2.** Twenty genetic patterns identified in this study. A. List of the genetic patterns. The sequences of the 201 viruses reported in this study, featured in phylogenetic trees in Supplementary Fig. 3, were used for the genetic analysis. Genetic patterns were identified based on the penton base, hexon, and fiber gene sequences. The genetic patterns are identified by a number, 1–20, for convenience. Collection year, collection country, and representative viruses have been determined based on the dataset provided in Supplementary Fig. 3 and GenBank sequences. The virus collected the earliest was chosen as a representative for each genotype, either from GenBank or from the 201 viruses reported in this study in the cases when no GenBank data was available. Twenty-four viruses chosen for whole genome sequencing are listed. Shanxi-2000-105 and Shanxi-2000-106 genomes are identical. Twelve genetic patterns represented only by viruses collected in China are shown in bold face. Among the 20 genetic patterns identified in this study, four have been previously reported, genotype C1, C2, C5, and C6. The genotype C57 is characterised by a penton base Px1/Ps1, hexon H57 and fiber F6 has not been reported in this study. B. Geographic distribution of the 20 genetic patterns. Each pattern is indicated by its number listed in panel A. Chinese genetic patterns are indicated by a star. Geographic locations are colour coded according to the legend based on province names and administrative regions (underlined). The number of viruses is shown for each genetic patterns and location. C. Temporal distribution of the 20 genetic patterns. The number of viruses is shown for each patterns between 2000 and 2015.

elements potentially involved in a recombination event (Supplementary Table 4). These potential recombination events were further analysed with BootScan and RDP4 recombination analysis. A schematic representation of potential recombination events for the 23 genomes, along with the raw data supporting this analysis is shown in Supplementary Fig. 7. Schemas concerning the 16 newly identified genetic groups are shown in Fig. 4. Whereas most genomes appear to result from the recombination between previously reported genomes, at least one genome, Shanxi-2010-106, might be the result of recombination of genomes that have not been identified yet. Thus, Shanxi-2010-106 is considered an outlier with long branches in all the phylogenetic trees, as well as in the SplitsTree (Huson and Bryant, 2006) (Fig. 3, Supplementary Fig. 6).

**Table 2**
Clinical and genetic information for 81 cases.

| Disease[a,b] | Clinical diagnostic | Sample source | Number of patients | Age (median) | Genetic pattern | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | Pattern[c] | P[a] | H[a] | F[a] |
| LRI | Bronchopneumonia, bronchitis | Throat Swab<br>Nasopharyngeal secretions<br>Bronchoalveolar lavage fluid | 39 | 2M-52 | 1, 5, 6, 9, 15, 17 | Px1/Ps1, Px1/Ps2, Px2, Px3 | all except H57 | all |
| URI | Herpangina, tonsilitis, pharyngitis | Throat Swab | 40 | 6M-31 | 1, 4, 5, 6, 9, 12, 13, 14, 17, 18, 19, 20 | Px1/Ps1, Px1/Ps2, Px1/Ps3, Px1/Ps4, Px2, Px3, Px4 | all | all |
| Enteritis | Enteritis | Throat Swab | 1 | 1.1 | 7 | Px1/Ps6 | H2 | F2 |
| Encephalitis | Encephalitis | CSF | 1 | 4 | 1 | Px1/Ps1 | H1 | F1 |

[a] Clinical and genetic information for all cases are listed in Supplementary Table 13.
[b] LRI: Lower Respiratory tract Infection; URI: Upper Respiratory tract Infection.
[c] Genetic pattern based on Fig. 2A.

## 4. Discussion

The three capsid genes, namely penton base, hexon, and fiber of all 201 HAdV-C strains were sequenced. The genetic types among hexon and fiber could be easily identified because of their high variability, which implies the immune pressure on these two proteins (Walsh et al., 2011). The current analysis of penton base revealed significant genetic diversity among the 201 HAdV-C strains. Previous studies estimated the HAdV-C penton base nucleotide diversity at 0.008, which is much lower than the diversity characterizing the penton base from HAdV-A, B, or D, with 0.117, 0.09, or 0.046, respectively (Ismail et al., 2018). In addition, previous phylogenetic analyses of the penton base featured a few nodes with significantly high bootstrap values, and the genetic variation was considered to be insufficient for genotyping purposes (Mao et al., 2017; Rivailler et al., 2019). The present study shows a very different picture. Even though the overall p-distance of the penton base gene was only 1.2%, two new major genetic groups (Px2 and Px3) as well as two subgroups [Px1/Ps3 and Px1/Ps4] were identified with the support of a high bootstrap value. Based on the analysis result of the penton base gene, all the HAdV-C strains could be divided into four distinct groups (Px1, Px5, Px2 and Px3) with at least 1.5–1.6% genetic distance. The use of Px1 which includes the previous types P1, P2 and P6 actually shows that P1, P2 and P6 were not divergent enough (p distance of 0.6 for the entire group) as others have already reported (Ismail et al., 2018). Distinguishing the types P1, P2 and P6 might not be relevant, and the present study based on 17 consecutive years of data confirmed that the penton base of HAdV-C is more variable than previously assumed (Ismail et al., 2018; Kajan et al., 2018).

Phylogenetic analysis of the penton gene sequences revealed 4 new genetic groups (Px2, Px3, Px1/Ps3 and Px1/Ps4), resulting in 16 new genetic patterns based on penton base, hexon, and fiber gene sequences. Interestingly, among the new genetic patterns, 4 patterns concerned previously reported genomes, namely JX173084, KX384959, KF268199, and KF268129 (Hang et al., 2017). Previous studies reported the sequences KF268129 and JX173079 as outliers from the other penton base sequences (Mao et al., 2017; Rivailler et al., 2019). The present study revealed that these sequences are indeed related to sequences in China and can be clustered within genetic groups Px3 (KF268129) and Px2 (JX173079). Some of the newly identified HAdV-C genetic patterns had been widespread and co-circulated for a long period of time, such as Px1/Ps1H2F2 during 2000–2015, covering 5 of the 9 surveillance provinces and Px2H2F2 during 2001–2015 from 4 the 9 surveillance sites. The nucleotide diversity of penton base gene further encouraged us to conduct homologous recombination and molecular evolution analyses of HAdV-C WGSs.

In the present study, despite the limitations of the geographical and temporal bias of WGS data due to the origin of the viruses from three different surveillance systems, multiple potential recombination events and different recombination patterns were found for each HAdV-C genetic patterns, which demonstrated that the genome of HAdV-C might recombine more frequently than expected. Genomic recombination usually requires co-infection of viruses (Cook and Radke, 2017; Lukashev et al., 2008). HAdV-Cs have the ability to establish persistent infections (Garnett et al., 2002). Consequently, infection with multiple HAdV-C strains can occur leading to recombination of genomic material. Most of the novel genotypes of HAdV usually result from exchanges of the hexon or fiber gene within species that alter tissue tropism and increase virulence (Ismail et al., 2018). In contrast, HAdV-C shows few homologous recombinations between the hexon and fiber genes from the same species. Therefore, HAdV-C genotyping based on neutralisation experiments or capsid genes is not sufficient to clarify recombination phenomena among HAdV-C viruses. WGS sequencing and recombination analysis at the genome level can provide more detailed information on this issue. HAdV-C viruses have the ability to establish persistent infection, and patients can remain asymptomatic carriers until at least their young adulthood (Garnett et al., 2002). In addition, it

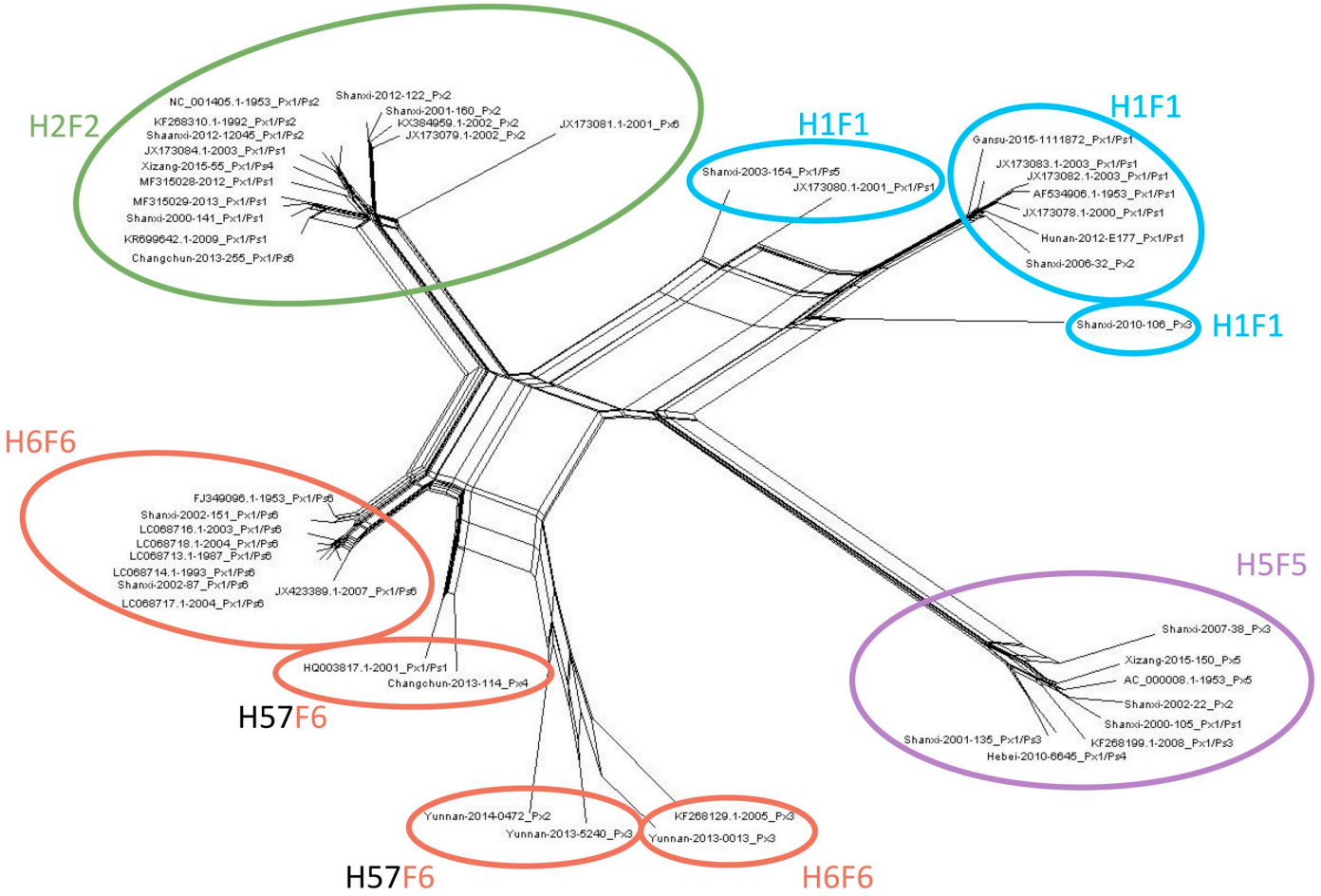



**Fig. 3.** Phylogenetic network built with 48 HAdV-C WGSs. Clusters have been identified based on hexon and fiber types. They are shown with different colors according to Fig. 1, namely fiber type 1 in pastel pink, fiber type 2 in blue, fiber type 5 in pastel yellow, and fiber type 6 in pastel green. The sequences corresponding to the prototype viruses of type 1, 2, 5, 6, and 57 are denoted by a black star. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

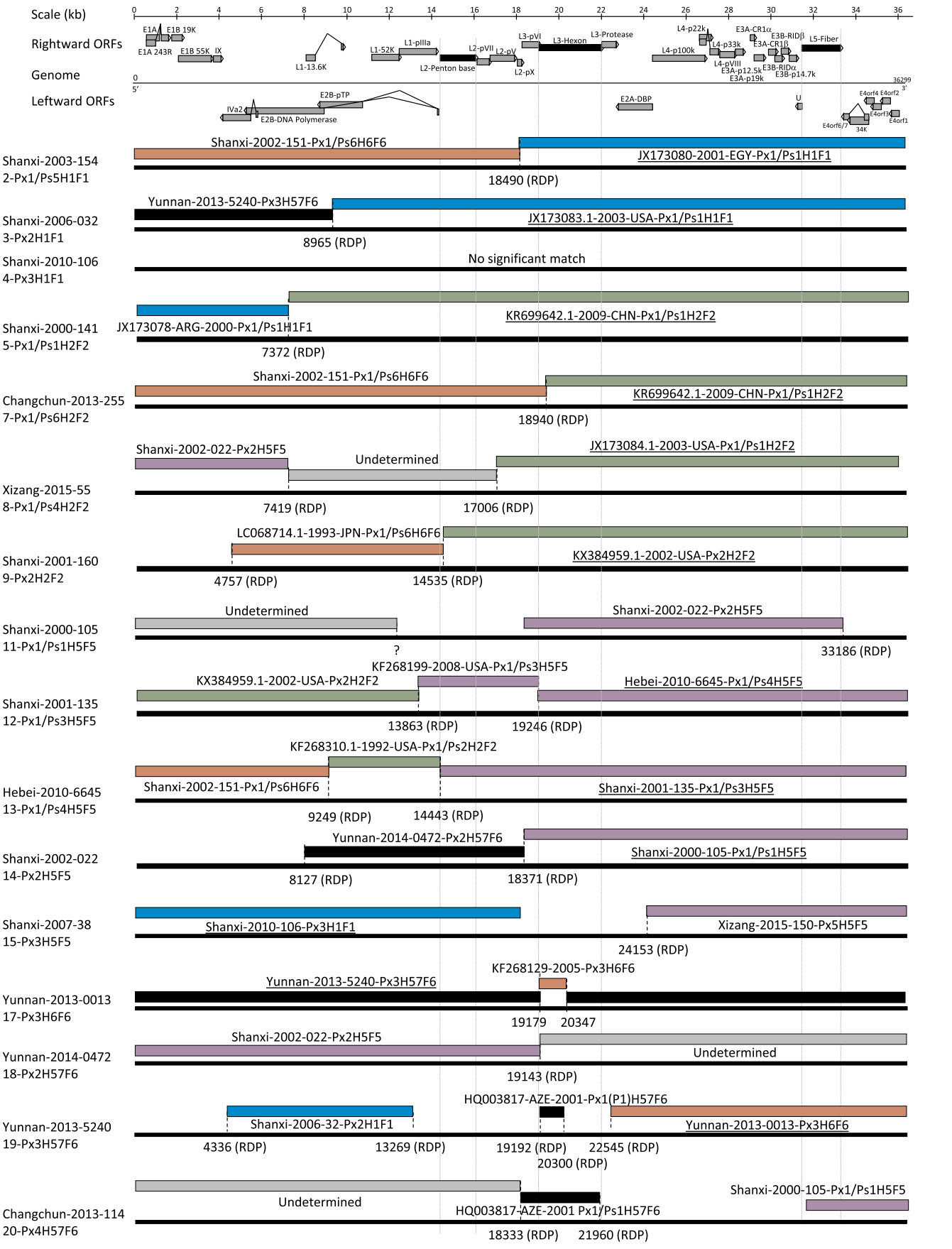
is known that HAdV remain stable in the environment for months, if not years (Rigotto et al., 2011). This viral persistence leads to an uncertainty with regard to the collection dates. Consequently, even though it is possible to predict the genetic elements involved in recombination event, the chronology of the events is more uncertain, as it is difficult to determine which genome is the major parent and which ones are minor parents.

The disease burden of HAdV-C is not clear due to the lack of surveillance systems in China, or even around the world. Among the 201 HAdV-C cases, only 81 patients with an average age of 2.9 years from the FRS system were clinically diagnosed. As expected, almost all patients suffered from upper or lower respiratory tract infections. Although HAdV-C members comprise the most prevalent pathogens among acute respiratory HAdV infections of children, HAdV-C can occasionally cross the blood-brain barrier leading to a central nervous system infection, according to reports from India and Taiwan (Akhil et al., 2016; Huang et al., 2013). In this study, a 4 year old child from Shandong province was diagnosed with encephalitis based on CSF detection results, and confirmed as a HAdV-C genetic pattern Px1/Ps1H1F1 infection. Although the primary infection of HAdV-C mainly occurs in the respiratory tract, the virus can still be excreted in the feces for several months or even years after it is no longer detected in nasopharyngeal swab specimens (Garnett et al., 2002). This could explain the large number of HAdV-Cs (120 strains) isolated from healthy children, AFP, and HFMD cases. Ultimately, this study demonstrated that existing surveillance programs that are not specifically targeted to adenoviruses, could also be useful for monitoring adenovirus evolution.

In conclusion, the present study revealed a significant genetic diversity within the penton base gene of species C, leading to the identification of 16 new genetic patterns and potential genotypes, which would need to be assessed by HAdVWG. WGS analysis showed that homologous recombination was critical for HAdV-C molecular evolution. Though the clinical significance of these new HAdV-C genetic patterns identified in this study remains unknown at present, some of the newly identified HAdV-C may be an increasing public health concern worldwide with regard to acute respiratory diseases such as HAdV-55 (P14H11F14) of HAdV-B. Therefore, the availability of a large set of genomic data based on 17 consecutive years in this study is highly beneficial for case classification, tracking the transmission chain, and further epidemiological exploration of HAdV-C related severe clinical disease in the near future.

## 5. Conclusions

This is the first study that identified 16 new genetic patterns of HAdV-C circulating during 2000–2016 in mainland China and describes molecular evolution of HAdV-C based on divergent sequences of penton base, hexon and fiber genes. Our result reveals that the penton base sequence of HAdV-C cluster in 4 new genetic groups in phylogenetic tree and is more divergent than previously thought. Whole genome sequence analysis of 24 strains HAdV-C representing 20 genetic patterns confirmed frequent recombination events among HAdV-C




*(caption on next page)*

**Fig. 4.** Schematic representation of recombination events within 16 genomes. A genomic map of HAdV-C is shown on the top. The l-strand of the genome is represented by a straight line. Rightward (top) and leftward (bottom) ORFs are represented by grey arrows based on the NC_001405 annotation. ORFs corresponding to the penton base, hexon, and fiber are depicted in black and dotted vertical lines indicate 5′ and 3′ ends of each gene. For illustration purposes, a recombination map is shown for 16 genomes from the 16 new genetic patterns identified in this study. The pattern number is indicated before the genetic pattern definition. For example, genetic pattern 2 is Px1/Ps5H1F1. Analysed genomes are depicted by a thick line. Genomic elements potentially involved in recombination events are shown as a rectangle, coloured based on their fiber type, namely fiber type 1 in pastel pink, fiber type 2 in blue, fiber type 5 in pastel yellow, and fiber type 6 in pastel green. Genomes with hexon H57 are shown in black. Undetermined genomes are shown in grey. Breakpoints are identified based on RPD4 or BootScan output. RDP4 analysis as well BootScan plot and pairwise p-distances are shown in Supplementary Fig. 7, which features a recombination map for all the 23 genomes analysed in this study. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

genomes. The data provide novel insight into the molecular evolution of HAdV-C and demonstrates that homologous recombination was critical for HAdV-C molecular evolution.

## Declaration of competing interest

The authors declare no conflict of interest.

## Acknowledgements

This work was supported by the Key Technologies R&D Program of the National Ministry of Science (2018ZX10713002, 2017ZX10104001-002, and 2018ZX10713001-003), and the National Key Research and Development Program (2017YFC1200303).

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.meegid.2019.104035.

## References

Akhil, C., Suresha, P.G., Sabeena, S., Hindol, M., Arunkumar, G., 2016. Genotyping of human adenoviruses circulating in Southwest India. Virusdisease 27, 266–270.
Benko, M.B., Harrach, W.R., 2005. Family Adenoviridae in Virus Taxonomy. Elsevier Academic Press, San Diego.
Berk, A.J., 2007. Fields virology. In: Knipe, D.M.H., Peter, M. (Eds.), Adenoviridae: The Viruses and Their Replication, 5th edition. Lippincott Williams&Wilkins.
Cao, B., Huang, G.H., Pu, Z.H., Qu, J.X., Yu, X.M., Zhu, Z., Dong, J.P., Gao, Y., Zhang, Y.X., Li, X.H., Liu, J.H., Wang, H., Xu, Q., Li, H., Xu, W., Wang, C., 2014. Emergence of community-acquired adenovirus type 55 as a cause of community-onset pneumonia. Chest 145, 79–86.
Cook, J., Radke, J., 2017. Mechanisms of pathogenesis of emerging adenoviruses. F1000Res 6, 90.
Crawford-Miksza, L., Schnurr, D.P., 1996. Analysis of 15 adenovirus hexon proteins reveals the location and structure of seven hypervariable regions containing serotype-specific residues. J. Virol. 70, 1836–1844.
Dhingra, A., Hage, E., Ganzenmueller, T., Bottcher, S., Hofmann, J., Hamprecht, K., Obermeier, P., Rath, B., Hausmann, F., Dobner, T., Heim, A., 2019. Molecular evolution of human adenovirus (HAdV) species C. Sci. Rep. 9, 1039.
Edwards, K.M., Thompson, J., Paolini, J., Wright, P.F., 1985. Adenovirus infections in young children. Pediatrics 76, 420–424.
Garnett, C.T., Erdman, D., Xu, W., Gooding, L.R., 2002. Prevalence and quantitation of species C adenovirus DNA in human mucosal lymphocytes. J. Virol. 76, 10608–10616.
Hage, E., Dhingra, A., Liebert, U.G., Bergs, S., Ganzenmueller, T., Heim, A., 2017. Three novel, multiple recombinant types of species of *Human mastadenovirus D* (HAdV-D 73, 74 & 75) isolated from diarrhoeal faeces of immunocompromised patients. J. Gen. Virol. 98, 3037–3045.
Hang, J., Vento, T.J., Norby, E.A., Jarman, R.G., Keiser, P.B., Kuschner, R.A., Binn, L.N., 2017. Adenovirus type 4 respiratory infections with a concurrent outbreak of coxsackievirus A21 among United States Army Basic Trainees, a retrospective viral etiology study using next-generation sequencing. J. Med. Virol. 89, 1387–1394.
Harrach, B., Benko, M., Both, G.W., Brown, M., Davison, A.J., Echavarria, M., Heiss, M., Jones, M., Kajon, A., Lehmkuhl, H.D., Mautner, V., Mittal, S., Wadell, G., 2011. Family Adenoviridae in Virus Taxonomy: Classification and Nomenclature of Viruses. pp. 125–141.
Huang, Y.C., Huang, S.L., Chen, S.P., Huang, Y.L., Huang, C.G., Tsao, K.C., Lin, T.Y., 2013. Adenovirus infection associated with central nervous system dysfunction in children. J. Clin. Virol. 57, 300–304.
Huson, D.H., Bryant, D., 2006. Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. 23 (2), 254–267.
Ismail, A.M., Cui, T., Dommaraju, K., Singh, G., Dehghan, S., Seto, J., Shrivastava, S., Fedorova, N.B., Gupta, N., Stockwell, T.B., Madupu, R., Heim, A., Kajon, A.E., Romanowski, E.G., Kowalski, R.P., Malathi, J., Therese, K.L., Madhavan, H.N., Zhang, Q., Ferreyra, L.J., Jones, M.S., Rajaiya, J., Dyer, D.W., Chodosh, J., Seto, D., 2018. Genomic analysis of a large set of currently-and historically-important human adenovirus pathogens. Emerg. Microbes Infect. 7, 10.
Kajan, G.L., Lipiec, A., Bartha, D., Allard, A., Arnberg, N., 2018. A multigene typing system for human adenoviruses reveals a new genotype in a collection of Swedish clinical isolates. PLoS One 13, e0209038.
Lole, K.S., Bollinger, R.C., Paranjape, R.S., Gadkari, D., Kulkarni, S.S., Novak, N.G., Ingersoll, R., Sheppard, H.W., Ray, S.C., 1999. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. J. Virol. 73, 152–160.
Lukashev, A.N., Ivanova, O.E., Eremeeva, T.P., Iggo, R.D., 2008. Evidence of frequent recombination among human adenoviruses. J. Gen. Virol. 89, 380–388.
Mao, N., Zhu, Z., Rivailler, P., Chen, M., Fan, Q., Huang, F., Xu, W., 2017. Whole genomic analysis of two potential recombinant strains within *Human mastadenovirus* species C previously found in Beijing, China. Sci. Rep. 7, 15380.
Martin, D.P., Murrell, B., Golden, M., Khoosal, A., Muhire, B., 2015. RDP4: detection and analysis of recombination patterns in virus genomes. Virus Evol. 1, vev003.
Rigotto, C., Hanley, K., Rochelle, P.A., De Leon, R., Barardi, C.R., Yates, M.V., 2011. Survival of adenovirus types 2 and 41 in surface and ground waters measured by a plaque assay. Environ. Sci. Technol. 45, 4145–4150.
Rivailler, P., Mao, N., Zhu, Z., Xu, W., 2019. Recombination analysis of *Human mastadenovirus C* whole genomes. Sci. Rep. 9, 2182.
Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.A., Barrell, B., 2000. Artemis: sequence visualization and annotation. Bioinformatics 16, 944–945.
Saitou, N., Nei, M., 1986. The number of nucleotides required to determine the branching order of three species, with special reference to the human-chimpanzee-gorilla divergence. J. Mol. Evol. 24, 189–204.
Scott, M.K., Chommanard, C., Lu, X., Appelgate, D., Grenz, L., Schneider, E., Gerber, S.I., Erdman, D.D., Thomas, A., 2016. Human adenovirus associated with severe respiratory infection, Oregon, USA, 2013–2014. Emerg. Infect. Dis. 22, 1044–1051.
Seto, D., Chodosh, J., Brister, J.R., Jones, M.S., Members of the Adenovirus Research, C., 2011. Using the whole-genome sequence to characterize and name human adenoviruses. J. Virol. 85, 5701–5702.
Seto, D., Jones, M.S., Dyer, D.W., Chodosh, J., 2013. Characterizing, typing, and naming human adenovirus type 55 in the era of whole genome data. J. Clin. Virol. 58, 741–742.
Singh, G., Robinson, C.M., Dehghan, S., Jones, M.S., Dyer, D.W., Seto, D., Chodosh, J., 2013. Homologous recombination in E3 genes of human adenovirus species D. J. Virol. 87, 12481–12488.
Suyama, M., Torrents, D., Bork, P., 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res. 34, W609–W612.
Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28, 2731–2739.
Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. 30, 2725–2729.
Walsh, M.P., Seto, J., Liu, E.B., Dehghan, S., Hudson, N.R., Lukashev, A.N., Ivanova, O., Chodosh, J., Dyer, D.W., Jones, M.S., Seto, D., 2011. Computational analysis of two species C human adenoviruses provides evidence of a novel virus. J. Clin. Microbiol. 49, 3482–3490.
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.