

# Emergence and characterization of a putative novel human adenovirus recombinant HAdV-C104 causing pneumonia in Southern China

Tianxing Ji,<sup>1,\*</sup> † Ling Li,<sup>2,†</sup> Wenrui Li,<sup>3,†</sup> Xuehua Zheng,<sup>4,†</sup> Xianmiao Ye,<sup>4</sup> Hongliang Chen,<sup>5</sup> Qiang Zhou,<sup>1</sup> Hongyun Jia,<sup>1</sup> Bo Chen,<sup>1</sup> Zhen Lin,<sup>1</sup> Haoyu Chen,<sup>1</sup> Shiwen Huang,<sup>1</sup> Donald Seto,<sup>6</sup> Ling Chen,<sup>4,7,\*</sup> and Liqiang Feng<sup>7,\*</sup>

<sup>1</sup>Clinical Laboratory Medicine Department, The Second Affiliated Hospital of Guangzhou Medical University, Guangzhou 510260, P. R. China, <sup>2</sup>Guangzhou KingMed Center for Clinical Laboratory, Guangzhou 510330, P. R. China, <sup>3</sup>Dongguan Children's Hospital, Dongguan 523326, P. R. China, <sup>4</sup>Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, Guangzhou 510530, P. R. China, <sup>5</sup>Clinical Laboratory Medicine Department, Deqing County People's Hospital, Zhaoqing 526600, P. R. China, <sup>6</sup>Bioinformatics and Computational Biology Program, School of Systems Biology, George Mason University, Manassas, VA 20110, USA and <sup>7</sup>State Key Laboratory of Respiratory Diseases, The First Affiliated Hospital of Guangzhou Medical University, Guangzhou 510120, P. R. China

\*Corresponding authors: E-mails: jitianxing7021@163.com (T.J.); feng\_liqiang@gibh.ac.cn (L.F.); chen\_ling@gibh.ac.cn (L.C.)

†These authors contributed equally to this work.

## Abstract

In 2017, a survey of the molecular epidemiology of human adenovirus (HAdV) infections in Southern China based on hexon and fiber genotype demonstrated that the most prevalent genotypes of HAdV were HAdV-3 ( $n = 62$ ), HAdV-2 ( $n = 21$ ), and HAdV-7 ( $n = 16$ ). In addition, two patients were co-infected with two genotypes of HAdV. Interestingly, a novel human adenovirus C recombinant genotype strain was isolated from one of the pneumonia patients in this survey. Phylogenetic, recombination, and proteotyping analysis showed that this novel pathogen originated from the recombination of parental viruses harboring the HAdV-1 penton and hexon gene, and the HAdV-2 fiber gene. It was named 'P1H1F2' and was assigned as HAdV-C104 based on the nomenclature protocol of using three major capsid proteins for characterization. Subsequent *in vitro* experiments demonstrated that HAdV-C104 had comparable proliferation capacity to HAdV-1, HAdV-2, and another recombination genotype P1H2F2. In addition, the HAdV-C104 infected patient was diagnosed with pneumonia and recovered after antiviral therapy. This report strengthens the hypothesis of recombination as a major pathway for the molecular evolution of HAdV-C species.

**Key words:** human adenovirus; molecular epidemic; molecular evolution.

## 1. Introduction

Human adenoviruses (HAdVs) are non-enveloped, double-stranded DNA viruses of the *Adenoviridae*. More than 100 genotypes of HAdVs, classified into seven species (A–G), have been identified and are characterized according to their physicochemical, biological, and genetic characteristics (<http://hadv.wg.gmu.edu/>). HAdVs are common pathogens responsible for a broad spectrum of human diseases such as acute respiratory disease, conjunctivitis, gastroenteritis, urinary infection, and meningoencephalitis ([Ismail et al. 2018a](#)). Several HAdVs, including HAdV-B (HAdV-3, -7, -55), HAdV-C (HAdV-1, -2, -5, -6), and HAdV-E (HAdV-4), have been recognized as primary pathogens associated with respiratory infections in pediatric patients ([Khanal et al. 2018](#)). Among these HAdVs, HAdV-C viruses are the second most common HAdV pertaining to human respiratory diseases causing more than half of the adenovirus infections in immunocompromised patients. HAdV-C viruses are also clinically more significant, in terms of disease severity, than HAdV-B and HAdV-E viruses in hematopoietic stem cell transplant recipients ([Wang et al. 2016](#); [Khanal et al. 2018](#); [Dhingra et al. 2019](#); [Yang et al. 2019](#)).

Evidence from the HAdV-B and HAdV-D species demonstrates that homologous recombination and immune selection are the major driving forces for the molecular evolution of HAdVs ([Robinson et al. 2013](#); [Ismail et al. 2018a](#)). Homologous recombination involving the three major capsid proteins (hexon, penton base, and fiber) can lead to a high diversity of HAdV-B and HAdV-D genotypes, with additional immune selection pressure presumably contributing to the adaptability of the diversity of these two genotypes ([Ismail et al. 2018a,b](#)). In regard to the HAdV-C species, recombination or escape mutations have been observed in ‘low diversity’ regions, including the penton base or the early gene regions, but not in the ‘high diversity’ hexon and fiber regions ([Dhingra et al. 2019](#); [Mao et al. 2019](#)). Compared with the HAdV-B and HAdV-D species, few novel genotypes within the HAdV-C species have been reported ([Mao et al. 2019](#)).

In this study, we performed a molecular epidemiological survey of HAdVs in Southern China, comprising the Guangdong, Guangxi, and Hunan province. A novel species C HAdV strain HAdV-C104 harboring the penton base and hexon genes from HAdV-1 with the fiber gene from HAdV-2 was isolated. Phylogenetic and genome recombination analyses were conducted on this novel recombinant HAdV strain. In addition, we investigated and compared the replication dynamics of this novel recombinant with the other members of the HAdV-C species. The clinical symptoms and treatment history of the novel isolate-infected patient were also retrospectively analyzed.

## 2. Methods

### 2.1 Ethics statement and human sample collection

This study was approved by the Ethics Committees of Guangzhou KingMed Clinical Laboratory Center and Deqing County People’s Hospital. Written informed consent was obtained from each volunteer or their parents. Nasopharyngeal swab specimens from the Guangdong, Guangxi, and Hunan province were sent to Guangzhou KingMed Diagnostics Group Co., Ltd for respiratory virus detection using a commercial respiratory virus immunofluorescence detection kit (Diagnostic Hybrids, USA). HAdV-positive samples were collected and stored at a temperature of  $-80^{\circ}\text{C}$ . Further, the clinical records of

the novel recombination HAdV-infected patient were retrospectively reviewed. The clinical information obtained for HAdV-C104 (MH558113) infected patient included age, diagnosis, drugs used, and prognosis.

### 2.2 HAdV culture and genomic DNA extraction

HAdV-positive samples were freeze-thawed three times, then filtered through 0.22 syringe filters into cell plates containing human lung carcinoma cells A549. The cells were grown in DMEM supplemented with 2 per cent fetal bovine serum. After incubation for 4–8 days, the cells were subjected to viral genomic DNA extraction using the phenol–chloroform method until obvious cytopathic effects were evident.

### 2.3 Molecular characterization of HAdVs

Molecular typing of the HAdVs was carried out by sequencing the hexon and fiber gene fragments using corresponding primers devised by [La Rosa et al. \(2011\)](#). The genomic fragments were amplified by TaKaRa Ex Taq<sup>®</sup> (Takara, Japan). The amplification products were subjected to bi-directional sequencing using an ABI Prism BigDye Terminator Cycle Sequencing Ready Reaction kit version 3.1 (ThermoScientific, USA). Forward and reverse sequences were assembled into a single consensus sequence using Vector NTI. The assembled sequences were subjected to BLAST analysis to identify the most closely related genotype. Finally, the molecular types of HAdV-positive isolates were determined based on the hexon and fiber genotype.

### 2.4 Complete genome sequence and annotation

Firstly, the complete genomes of HAdV isolates were sequenced using a Sanger chemistry-based primer walking method. In brief, 3,000 bp genomic fragments were amplified using a high-fidelity Prime STAR<sup>®</sup> Max DNA Polymerase kit (Takara, Japan) with an overlap of 100–200 bp between adjacent fragments. The corresponding PCR primers were designed according to the conserved sequences among species C HAdVs ([Supplementary Table S1](#)). Subsequently, PCR fragments were subjected to bi-directional sequencing using an ABI Prism BigDye Terminator Cycle Sequencing Ready Reaction kit version 3.1 (ThermoScientific). For the gaps and ambiguous sequences, the corresponding regions were amplified by PCR using new primers designed according to the known sequences, and re-sequencing was then performed. Annotation of each element and open-reading-frame was conducted according to a previously described strategy ([Wang et al. 2016](#)). The complete genomic sequences were deposited in the NCBI database.

### 2.5 Next-generation sequencing of HAdV-C104

To further ascertain the complete genome sequences of HAdV-C104, next-generation sequencing was performed. In brief, HAdV-C104 was propagated in 293 cells, and purified using the CsCl gradient method. Subsequently, the genomic DNA was extracted using the phenol–chloroform method. The purified DNA was sheared by sonication. Afterward, the sequencing libraries were prepared directly using ND602-0102VAHTS PCR-Free DNA Library Prep Kit (Illumina, USA). After quality control, the libraries were sequenced on Illumina NovaSeq 6000 platform ( $2 \times 150$  bp paired-ends run) (Illumina). The obtained sequence data were quality-controlled and *de novo* assembled using Metaviral SPAdes assembler ([Dmitry 2020](#)).

## 2.6 Genome analysis

The sequences were aligned with Clustal W. Phylogenetic trees were constructed by maximum likelihood implemented in MEGA7.0 using the best-fit nucleotide substitution model. The reliability of all phylogenetic trees was evaluated by 1,000 bootstraps. Recombination analysis was performed using seven methods (RDP, GENECONV, BOOTSCAN, MaxChi, CHIMAERA, SiSCAN, and 3SEQ) implemented in the RDP4 suite (Martin et al. 2015). The probability of a putative recombination event was corrected using the Bonferroni procedure with a cutoff of  $P < 0.01$ . To avoid false identification, only events supported by at least five of the seven methods were considered to be recombinants. In addition, the recombination events were further ascertained using Simplot software under the following parameters: window size of 500 bp, step size of 100 bp, gap stripping, kimura (2-parameter). The following HAdV genome sequences were used in genomic analysis: HAdV-1 (MH121116), HAdV-2 (JX173077.1), HAdV-5 (AC\_000008.1), HAdV-6 (FJ349096.1), P1H1F2 (MH558113), HAdV-57 (HQ003817.1), HAdV-89 (MH121097), HAdV-1 (MK357714), P1H2F2 (MK357715), HAdV-2 (MN088492), and HAdV-7 (AC\_000018.1).

## 2.7 Proteotyping analysis

Sequence identity was calculated using EMBOSS Needle ([https://www.ebi.ac.uk/Tools/psa/emboss\\_needle/](https://www.ebi.ac.uk/Tools/psa/emboss_needle/)) and a 10 per cent sequence divergence threshold was used to identify unique proteotypes (Singh et al. 2013).

## 2.8 Plaque formation assay

A 100 per cent confluent A549 cells were infected with HAdVs at an MOI of 0.1 for 2 h in a cell incubator. The cells were then washed with PBS and overlaid with 3 ml of 1.2 per cent agarose mixed 1:1 with  $2\times$  MEM containing 2 per cent FBS, 100 IU/ml penicillin, and 100  $\mu$ g/ml streptomycin. The cells were then incubated at 37 °C in air atmospheric conditions with 5 per cent CO<sub>2</sub>. Five days after the infection, the cells were fixed, permeabilized, and incubated with 500  $\mu$ l anti-HAdV-2 serum (1:200 dilution) and 500  $\mu$ l HRP-labeled goat anti-human IgG antibody (1:2,000 dilution, Beyotime Biotechnology, China). Finally, the cells were stained with 500  $\mu$ l AEC peroxidase substrate for 20 min at room temperature. The strain was decanted, and the plates were rinsed three times with water. The size of the plaque was determined using Nano Measurer 1.2.

## 2.9 Statistical analysis

All the statistical analyses were performed using GraphPad Prism version 5. Differences between groups were calculated using the Mann–Whitney *U* tests. *P*-values  $< 0.05$  were considered statistically significant.

## 3. Results

### 3.1 Molecular epidemiology of HAdVs

Between 1 January 2017 and 30 December 2017, a total of 13,467 nasopharyngeal swab samples were sent to Guangzhou KingMed Clinical Laboratory Center for respiratory virus detection. Among them, 123 specimens were identified as HAdV positive. The genotypes of 123 specimens HAdV isolates were successfully determined based on the nucleotide sequences of hexon and fiber genes (Fig. 1). The total HAdV positive rate was 1.25 per cent in 2017, with the highest infection rate observed in

June (Fig. 2A). The majority of the infected children were younger than 4 years old (Fig. 2B). The most common genotype was HAdV-3 ( $n = 62$ ), followed by HAdV-2 ( $n = 21$ ), HAdV-7 ( $n = 16$ ), HAdV-1 ( $n = 14$ ), HAdV-5 ( $n = 6$ ), and HAdV-4 ( $n = 1$ ) (Fig. 2C; Supplementary Fig. S1). Three specimens were further characterized by virus culture and subsequent plaque purification as their genotypes were considered discordant based on the sequences of the hexon and fiber genes (Fig. 2C; Supplementary Fig. S1). Molecular typing of single viral plaque suggested that S89 patient was co-infected with HAdV-3 and HAdV-7 co-infection, S10 was co-infected with HAdV-1 and HAdV-5, S88 patient was infected with a novel recombinant genotype as the hexon sequence from the single viral plaque indicated HAdV-1 whereas the fiber sequence from the same plaque indicated HAdV-2 (Fig. 2C; Supplementary Fig. S1).

### 3.2 Phylogenetic and genomic analyses of the novel HAdV-C isolate

The complete genome of HAdV isolate from S88 patient and other three HAdV-C species strain isolated in a previous study (Ji et al. 2020) were sequenced and deposited in the NCBI Genbank (Accession numbers: MH558113, MK357714, MK357715, and MN088492, respectively). Moreover, the complete genome sequence of MH558113 was further verified using next-generation sequencing. MH558113 isolate was designated as HAdV-C104 genotype (<http://hadvhwg.gmu.edu>), whose genes were annotated (Supplementary Table S2). Next, we performed phylogenetic analysis of these HAdV-C isolates obtained in this study along with other representative HAdV-C viruses. Phylogenetic analysis based on the complete genome showed that HAdV-C104 formed a distinct cluster (Fig. 3A). Phylogenetic analyses of the penton base and hexon gene sequences showed that HAdV-C104 clustered with HAdV-1 while the HAdV-C104 fiber gene clustered with HAdV-2 (Fig. 3B–D). Therefore, HAdV-C104 was typed as a novel recombinant genotype and was named P1H1F2 (<http://hadvhwg.gmu.edu>) according to the current nomenclature rules based on the three major capsid gene identities (Seto et al. 2011).

Additionally, phylogenetic analysis based on the complete genome showed that MK357715 formed a cluster with HAdV-89, with its penton base gene sequence clustering with that of HAdV-1, and the hexon and fiber gene sequences clustering with those of HAdV-2 (Fig. 3). Therefore, this isolate was denoted as P1H2F2. The MK357714 isolate was typed as HAdV-1 because its penton base, hexon, and fiber gene sequences were all similar to those of HAdV-1 strains (Fig. 3). Accordingly, the MN088492 isolate belonged to HAdV-2 because its penton base, hexon, and fiber gene were all phylogenetically close to those of the HAdV-2 prototype strain (Fig. 3).

Further, the nucleotide sequence diversity of the recombination hotspot region was analyzed using the EMBOSS needle (Ismail et al. 2018b). Our results suggested that the penton base and hexon nucleotide sequences of HAdV-C104 were similar to that of HAdV-1, but the E3 and fiber sequences were similar to HAdV-2 (Table 1).

Next, we analyzed the molecular phylogeny of the early genes of HAdV-C104. The E1A and E1B of HAdV-C104 were similar to the HAdV-1 strain, MH121116. However, the E1A gene of HAdV-6 (FJ349096.1) was closed to HAdV-57 (HQ003817.1) and P1H2F2 (MK357715). HAdV-1 (MK357714) was similar to HAdV-2 (MN088492). The E1B gene of HAdV-2 (MF044052.1) clustered with HAdV-5 (AC000008.1), E1B gene of HAV-89 (MH121097), HAdV-6 (FJ349096.1), and P1H2F2 (MK357715) formed a cluster

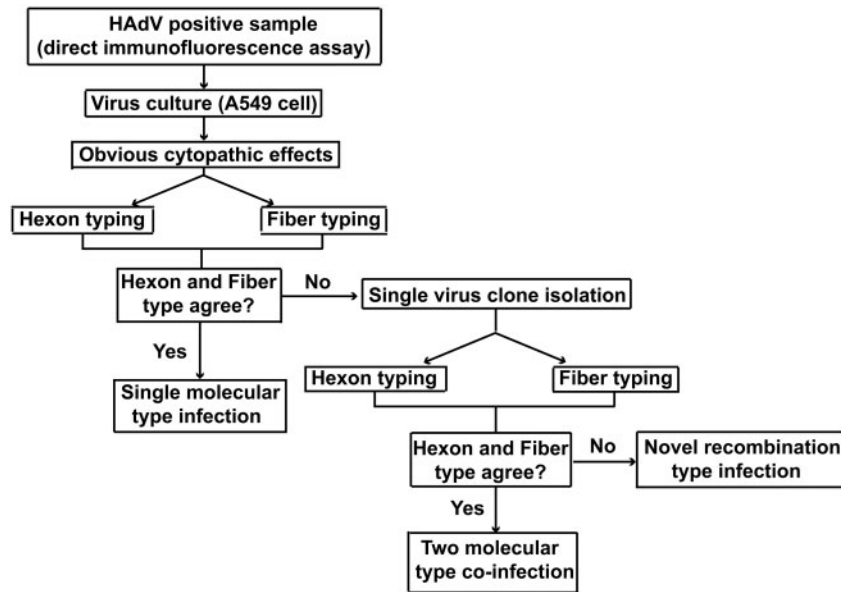


Figure 1. Schematic illustration of the determination of human adenovirus genotype in this study.

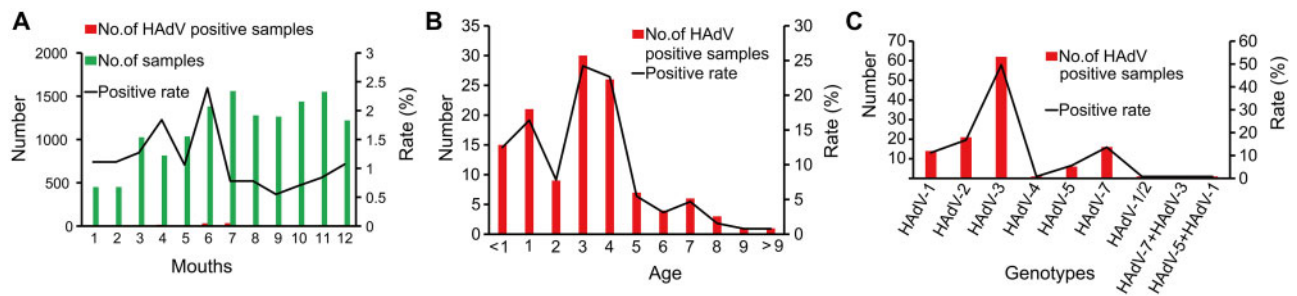


Figure 2. Epidemic characteristics of HAdVs infection in Southern China. (A) Proportion of HAdV detected in nasopharyngeal swabs, by months. (B) Proportion of HAdV detected in nasopharyngeal swabs, by age. (C) Proportion of HAdV infection by different molecular types.

(Fig. 4A and B). These data indicated the presence of recombination hotspots in the E1A and E1B regions of HAdV-C species. The E3 genes of HAdV-C104 clustered with prototypes HAdV-2, HAdV-6, and HAdV-89, and were distant from prototypes other HAdV-C strain including HAdV-1, HAdV-5, and HAdV-57 (Fig. 4C). However, there were no obvious sequence differences in the E4 genes of the analyzed HAdV-C genotypes (Fig. 4D), implying that the E4 gene, of all HAdV-C species, had the same ancestor.

### 3.3 Recombination analysis

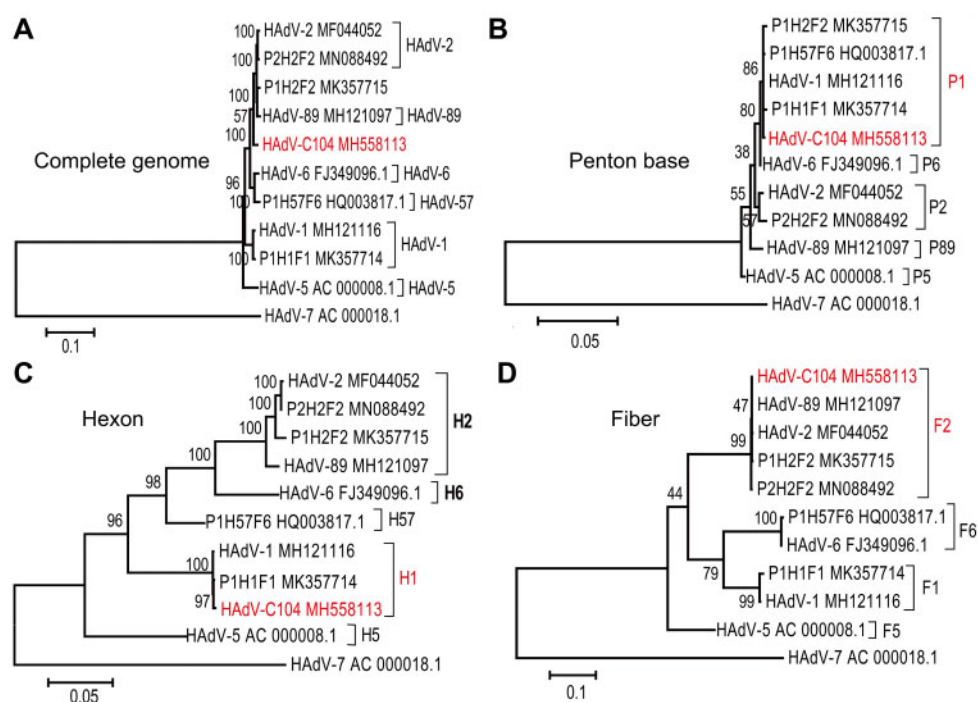
To further explore the recombination findings for HAdV-C104, genome recombination analysis was carried out using RDP4 (Martin et al. 2015) and the Simplot software tools. Three recombination events involving HAdV-1 and HAdV-2 were identified with a high level of confidence as recombinants, according to all seven methods implemented in the RDP package (Table 2). Simplot analysis consolidated the recombination events of HAdV-1 and HAdV-2 (Fig. 5A and B). Accordingly, phylogenetic tree analyses demonstrated that the major region of 1–8,699 bp and 10,052–22,302 bp in the HAdV-C104 genome originated from HAdV-1, while the minor region of 8,700–10,051 and 22,304–36,127 bp in the HAdV-C104 genome was derived from HAdV-2 (Fig. 6A–D).

### 3.4 Proteotyping analysis

The amino acid sequence diversity of the recombination hotspot region was analyzed using the EMBOSS needle. Based on a published study (Singh et al. 2013), proteotyping analysis further supported the assertion that HAdV-C104 was a novel recombination adenovirus originating from parental HAdV-1 and HAdV-2 viruses, with penton base and hexon of HAdV-1, fiber, and E3 gene encoding proteins of HAdV-2 (Table 3). These data and results further improve our understanding on the evolution of HAdV-C species in which the major driving force is recombination; consistent with previous reports on novel pathogens of HAdV-B and HAdV-D species (Ismail et al. 2018a).

### 3.5 Comparative analysis of the proliferation of the novel HAdV-C isolate with other closely related HAdVs

Since the recombination of the capsid protein genes could alter the tropism and pathogenicity of the resultant virus (Wang et al. 2016; Ismail et al. 2018a,b), we explored whether the recombinant isolate HAdV-C104 had different cyto-pathogenicity than the parental isolates HAdV-1 and HAdV-2, and the recently reported recombinant isolate P1H2F2 (Wang et al. 2016). This analysis was performed using a plaque formation assay. The results showed that the sizes of plaques formed by HAdV-C104 were comparable with those formed by HAdV-1 (MK357714),



**Figure 3.** Phylogenetic relationship between HAdV-C104 and other human adenovirus C species based on complete genome (A), penton base (B), hexon (C), and fiber (D). Maximum likelihood trees were constructed using the best fit nucleotide substitution model, GTR + G + I (complete genome), K2 + G (penton), GTR + G (hexon), HKY + G (fiber) in MEGA7.0. HAdV-7 (AC\_000018.1) were donated as outgroup. The novel types are indicated in red.

**Table 1.** Percent identity of nucleotide acid between HAdV-C104 (MH558113) penton base, hexon, fiber, E3, and counterparts from other human adenovirus C species.

Genotypes	HAdV-C104 (MH558113)			
	Penton (%)	Hexon (%)	E3 (%)	Fiber (%)
HAdV-1	99.9	99.8	78.5	72.7
HAdV-2	98.8	85.4	100	99.8
HAdV-5	98.0	84.0	77.3	73.8
HAdV-6	99.7	84.2	97.2	69.0
HAdV-57	99.9	87.9	95.3	68.7
HAdV-89	97.9	85.3	99.7	99.7

HAdV-2 (MN088492), and P1H2F2 (MK357715) (Fig. 7A and B). This suggests that HAdV-C104 might have similar virulence as these other genotypes in cultured cells.

### 3.6 Clinical characteristics of the novel HAdV-C isolate-infected patient

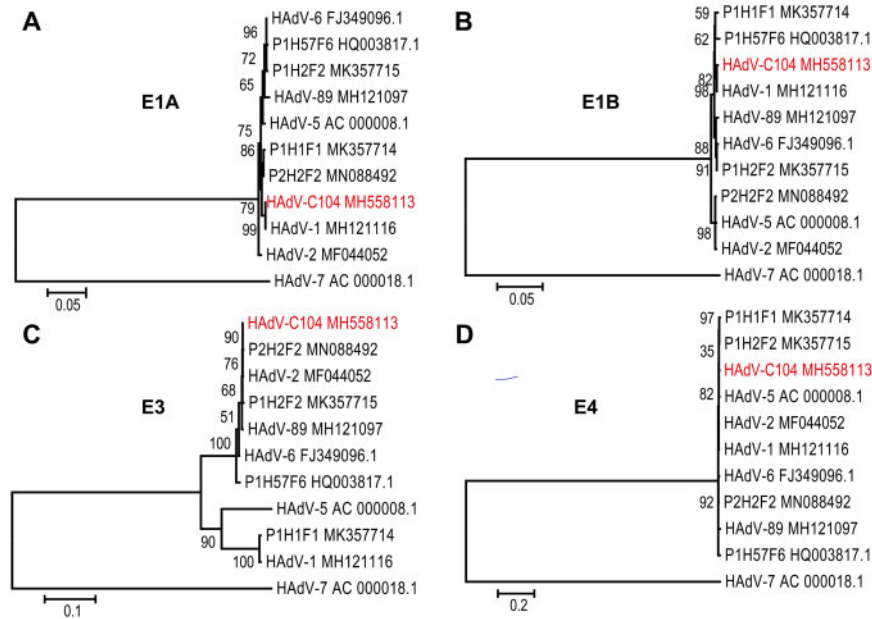
It has been reported that the recombination isolate P1H2F2 can cause severe acute respiratory infection in children (Wang et al. 2016). Therefore, we retrospectively reviewed the clinical characteristics of a patient infected with HAdV-C104. This HAdV-C104 infected patient was a girl of 2 years 8 months old. She was referred to the People's Hospital of Deqing County due to purulent nasal discharge and high fever. Physical and serological examination showed high white blood cell count ( $15.36 \times 10^9/l$ ), low neutrophil granulocyte rate (46.7%), high body temperature ( $40.2^\circ C$ ), fatigue, pharyngeal congestion, 'rough sounding' breathing, and a heart rate of 126 beats/min. Chest radiographic findings revealed increased bilateral lung markings along with

small patches of blurred shadows near the hilus of the lung; suggesting the presence of pneumonia (Supplementary Fig. S2). The patient received azithromycin lactobionate injection after *M. pneumoniae* antibodies IgG and IgM were reported as positive. However, the patient continued to experience intermittent fevers 2 days after the treatment. Subsequently, the patient was diagnosed with HAdV infection and received a recombination interferon  $\alpha$ -2b injection. After a course of antiviral treatment (4 days), the patient recovered and was discharged.

## 4. Discussion

Here, the molecular epidemiological survey of HAdVs revealed that the prominent molecular types of HAdVs were HAdV-3, HAdV-2, and HAdV-7 in the Guangdong, Guangxi, and Hunan province of Southern China, in 2017. Interestingly, two patients co-infected with two different molecular types were identified and another was found to be infected with a novel recombinant HAdV-C104. Phylogenetic, recombination, and proteotyping analysis demonstrated that the HAdV-C104 strain originated from the recombination of HAdV-1 and HAdV-2, and was named P1H1F2 based on adenovirus nomenclature rules. *In vitro* experiment demonstrated that HAdV-C104 exhibited comparable proliferation dynamics to the parental HAdV isolates, including HAdV-1, HAdV-2, and P1H2F2. Retrospective analyses indicated that the HAdV-C104 infected patient recovered after antiviral therapy.

Our results indicated that the HAdV positive rate in Southern China in 2017 was 1.25 per cent, lower than in other regions (Yao et al. 2019; Zhao et al. 2020). This difference may be related to the use of different detection methods. PCR detection has higher sensitivity than indirect immunofluorescence detection. However, more healthy controls had low levels of HAdV DNA in nasopharyngeal aspirate (NA) (Schjelderup Nilsen et al.



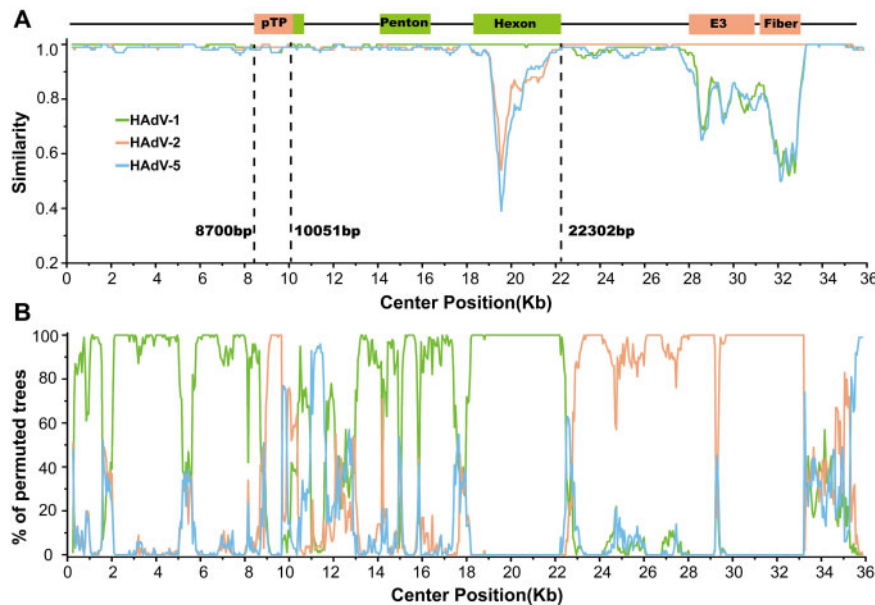
**Figure 4.** Phylogenetic relationship between HAAdV-C104 and other human adenovirus C species based on E1A (A), E1B (B), E3 (C), and E4 (D). The nucleotide sequences were aligned using Clustal W in MEGA7.0. Maximum likelihood trees were constructed using the best fit nucleotide substitution model K2 (E1A), HKY + G (E1B), HKY + G (E3), HKY + I (E4) in MEGA7.0. HAAdV-7 (AC\_000018.1) were donated as outgroup. The novel types are indicated in red.

**Table 2.** The recombination events identified by RDP4.

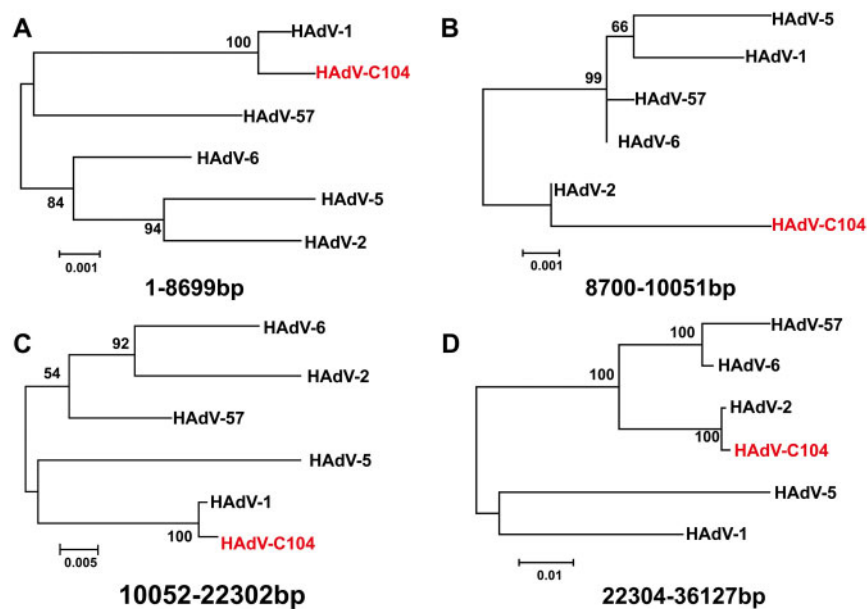
Breakpoint positions					Detection methods						
Begin	End	Recombinant sequence(s)	Minor Parental Sequence(s)	Major Parental Sequence(s)	RDP	GENECONV	Bootscan	Maxchi	Chimaera	SiSscan	3Seq
1 <sup>a</sup>	7789	HAAdV-C104 <sup>b</sup>	HAAdV-1	HAAdV-2	9.19E-75	2.67E-65	1.79E-53	2.46E-19	1.93E-07	5.45E-18	3.33E-15
12788	22280	HAAdV-C104	HAAdV-1	HAAdV-2	1.75E-185	2.88E-189	7.33E-184	8.75E-57	3.09E-02	1.96E-62	3.33E-15
33026	36127 <sup>a</sup>	HAAdV-C104 <sup>b</sup>	Unknown	HAAdV-2	7.05E-11	3.99E-04	1.18E-14	2.66E-02	3.59E-04	NS	1.64E-04

<sup>a</sup>The actual breakpoint position is undetermined (it was most likely overprinted by a subsequent).

<sup>b</sup>The recombinant sequence may have been misidentified (one of the identified parents might be the recombinant). Minor parent: parent contributing the smaller fraction of sequence. Major parent: parent contributing the larger fraction of sequence. Unknown: only one parent and a recombinant need be in the alignment for a recombination event to be detectable. The sequence listed as unknown was used to infer the existence of a missing parental sequence. NS: no significant P-value was recorded for this recombination event using this method.



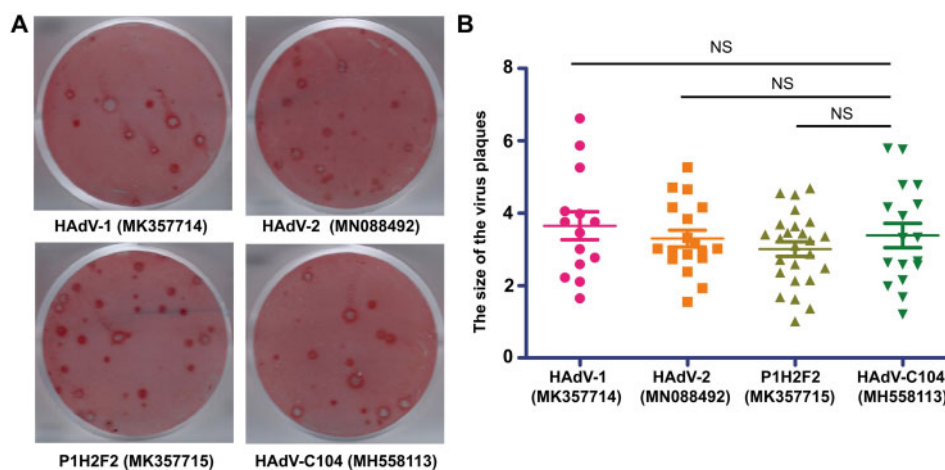
**Figure 5.** Recombination analysis between HAAdV-C104 and other human adenovirus C species using Simplot (A) and Bootscan (B).



**Figure 6.** Phylogenies of the recombination regions including 1 – 86,99, 8,700 – 10,051, 10,052 – 22,302, 22,304 – 36,127 bp. Maximum likelihood trees were constructed using the best fit nucleotide substitution model HKY + G + I (1 – 8,699 bp), T92 (8,700 – 10,051 bp), GTR + G + I (10,052 – 22,302 bp), GTR + G (22,304 – 36,127 bp) in MEGA7.0.

**Table 3.** Percent identity of amino acid between HAdV-C104 (MH558113) penton base, hexon, fiber and counterparts from other human adenovirus C species.

Genotypes	HAdV-C104 (MH558113)									
	Penton (%)	Hexon (%)	12.5 (%)	CR1- $\alpha$ (%)	19K (%)	CR1- $\beta$ (%)	RID- $\alpha$ (%)	RID- $\beta$ (%)	14.7K (%)	Fiber (%)
HAdV-1	100.0	100.0	98.1	78.1	93.1	52.9	93.4	84.3	80.5	72.4
HAdV-2	99.1	90.9	100	98.4	100.0	100.0	100.0	100.0	100.0	99.7
HAdV-5	98.3	86.8	88.8	81.0	81.9	66.3	91.2	78.4	78.1	69.3
HAdV-6	99.8	84.1	71.0	100.0	100.0	98.0	98.9	96.9	100.0	67.7
HAdV-57	99.8	91.1	98.1	98.4	97.5	98.0	98.9	96.9	100.0	66.8
HAdV-89	98.6	94.7	99.1	100	100.0	99.0	100.0	99.2	100.0	99.8



**Figure 7.** (A) Plaque formation of HAdV-C104, HAdV-1, HAdV-2, and P1H2F2. (B) Distribution of the plaque sizes of HAdV-C104, HAdV-1, HAdV-2, and P1H2F2.

2019). Hence, the detection of HAdV DNA per se by PCR was not associated with respiratory tract infections, only HAdV culture positive in NA and high levels of HAdV DNA  $>10^6$  copies/ml were strongly associated with RTI in hospitalized children

(Schjelderup Nilsen et al. 2019). In this present study, almost all indirect immunofluorescence detected adenovirus positive samples produced cytopathic effects. It is plausible that viral culture and the indirect immunofluorescence detection method

could be useful as a human adenovirus pathogenic infection diagnostic test (Schjelderup Nilsen et al. 2019). The top three circulating genotypes in Southern China in 2017 were HAdV-3, HAdV-2, and HAdV-7, which is quite different from those observed in Beijing (year, 2017) (Yao et al. 2019), Xining (year, 2017) (Yu et al. 2019), and Guangzhou (year, 2012–3) (Chen et al. 2016), but similar to those observed in the Hebei province (year, 2017) (Zhao et al. 2020). These results suggest that different regions have different molecular epidemics, with the prominent genotype varying with time. Therefore, it is necessary to continuously monitor molecular epidemics.

Based on the HAdV nomenclature rules proposed by the international HAdV working group, the recombination of high diversity regions significantly diversified the genotypes of HAdV-B and D species (Seto et al. 2011; Ismail et al. 2018a). However, the HAdV-C species only contains six genotypes, including HAdV-1, HAdV-2, HAdV-5, HAdV-6, HAdV-57, and HAdV-89 (Dhingra et al. 2019). Among them, HAdV-57 and HAdV-89 have a unique hexon and penton base, respectively (Walsh et al. 2011; Dhingra et al. 2019). Recently, data have shown that multiple recombination events in the HAdV-C penton base resulted in several novel recombinant isolates, and one of these novel recombinant isolate infections might be related to serious respiratory disease (Wang et al. 2016; Mao et al. 2019). However, these recombinant isolates were not defined as novel genotypes as the penton base gene sequence of the HAdV-C species exhibited significantly lower heterogeneity than the hexon and fiber gene sequence, and those recombinant events did not significantly change the biological properties of the isolate (Mao et al. 2019). Here, an emerging novel HAdV-C isolate was involved in the recombination of HAdV-1 hexon, and HAdV-2 fiber and E3; which are high diversity regions of the HAdV-C species. The two major capsid proteins (hexon and fiber) play key roles in the biological properties of HAdVs. Therefore, the recombinant isolate GD2467 was a novel genotype according to the HAdV nomenclature rules (Seto et al. 2011) and was formally designated as HAdV-C104. Just recently, a novel recombination HAdV-C isolate possessing a genomic backbone of type HAdV-C89 and a unique insertion of HAdV-C1 in the hexon sequence was isolated (Tahmasebi et al. 2020). In addition, recent reports have shown that recombination and mutation events of another high diversity region E3 gene resulted in a novel HAdV-C isolate (Zhang et al. 2019; Ji et al. 2020). The above data provides the evidence to support HAdV-C species evolution through recombination, with recombination regions including high (hexon, fiber, and E3) and low (penton, E1, and E4) diversity region; similar to other reports of isolates of species HAdV-B and HAdV-D (Singh et al. 2013; Ismail et al. 2018a; Dhingra et al. 2019; Mao et al. 2019).

Although, HAdV-C104 exhibited similar virulence as HAdV-1, HAdV-2, and P1H2F2, and the HAdV-C104-infected patient recovered after appropriate antiviral therapy, any putative novel recombinant adenovirus isolate could present additional unknown risks, including severe outbreaks in dense populations. This, in turn, can result in severe morbidity and mortality, particularly in immunosuppressed patients who are already susceptible to C adenoviruses species (Dhingra et al. 2019; Mao et al. 2019). Therefore, it is necessary to continuously monitor the novel recombinant HAdV-C104 isolate epidemic.

In conclusion, we report a recombinant novel adenoviral human pathogen that supports HAdV-C species evolution through recombination. Therefore, complete genome sequencing and genome recombination analysis are recommended for thorough identification of adenoviruses in the surveillance and

characterization of adenoviral pathogens in the general community. We suggest that the novel recombinant HAdV-C104 epidemic should be continuously monitored.

## Supplementary data

Supplementary data are available at *Virus Evolution* online.

## Funding

This work was supported by the National Natural Science Foundation of China (31470892), National Key Research and Development Project (2016YFC1200900), National Science and Technology Major Project (2017ZX10204401003, 2018ZX10101003005), Guangzhou Health Care and Cooperative Innovation Major Project (201803040004), Guangzhou and the grant of CAS Youth Innovation Promotion Association (2014328), Guangzhou Health Science and technology project (20201A011078), Guangdong Provincial Department of Education Youth innovative talents Project (2017KQNCX168) and Guangzhou Science and Technology Project (201904010214).

## Authors' contributions

T.J. did the methodology, data collection, and analysis; and wrote the original draft; L.L. and W.L. were collected the samples and analyzed; X.Y. did the methodology; X.Z. has contributed in statistical analysis; H.C., Q.Z., H.J., B.C., Z.L., Y.H. contributed for sample collection, molecular typing; D.S. revised the paper; L.C. was participated in conceptualization, funding acquisition, and supervision; L.F. was participated in conceptualization, funding acquisition, and supervision, methodology, writing the original draft, writing review and editing. All the authors participated in the paper preparation and approved the submission of the manuscript.

**Conflict of interest:** None declared.

## References

- Chen, Y. et al. (2016) 'Molecular Identification and Epidemiological Features of Human Adenoviruses Associated with Acute Respiratory Infections in Hospitalized Children in Southern China, 2012-2013', *PLoS One*, 11: e0155412.
- Dhingra, A. et al. (2019) 'Molecular Evolution of Human Adenovirus (HAdV) Species C', *Scientific Reports*, 9: 1039.
- Dmitry, A. (2020) 'Metaviral SPAdes: Assembly of Viruses from Metagenomic Data', *Bioinformatics*, 36: 4126–9.
- Ismail, A. M. et al. (2018a) 'Genomic Analysis of a Large Set of Currently-and Historically-Important Human Adenovirus Pathogens', *Emerging Microbes & Infections*, 7: 10.
- et al. (2018b) 'Adenoviromics: Mining the Human Adenovirus Species D Genome', *Frontiers in Microbiology*, 9: 3005.
- Ji, T. et al. (2020) 'Molecular Typing and Genomic Characteristic of Human Adenoviruses in Datong', *Journal of Medical Virology*, 92: 3111–8.
- Khanal, S. et al. (2018) 'The Repertoire of Adenovirus in Human Disease: The Innocuous to the Deadly', *Biomedicine*, 6: 30.
- La Rosa, G. et al. (2011) 'Molecular Characterization of Adenovirus from Clinical Samples through Analysis of the Hexon and Fiber Genes', *The Journal of General Virology*, 92: 412–20.



- Mao, N. et al. (2019) 'Multiple Divergent Human Mastadenovirus C co-Circulating in Mainland of China', *Infection, Genetics and Evolution : Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases*, 76: 104035.
- Martin, D. P. et al. (2015) 'RDP4: Detection and Analysis of Recombination Patterns in Virus Genomes', *Virus Evolution*, 1: vev003.
- Robinson, C. M. et al. (2013) 'Molecular Evolution of Human Adenoviruses', *Scientific Reports*, 3: 1812.
- Schjelderup Nilsen, H. J. et al. (2019) 'Human Adenovirus in Nasopharyngeal and Blood Samples from Children with and without Respiratory Tract Infections', *Journal of Clinical Virology*, 111: 19–23.
- Seto, D. et al.; Members of the Adenovirus Research Community. (2011) 'Using the Whole-Genome Sequence to Characterize and Name Human Adenoviruses', *Journal of Virology*, 85: 5701–2.
- Singh, G. et al. (2013) 'Homologous Recombination in E3 Genes of Human Adenovirus Species D', *Journal of Virology*, 87: 12481–8.
- Tahmasebi, R. et al. (2020) 'Genomic Analyses of Potential Novel Recombinant Human Adenovirus C in Brazil', *Viruses*, 12: 508.
- Walsh, M. P. et al. (2011) 'Computational Analysis of Two Species C Human Adenoviruses Provides Evidence of a Novel Virus', *Journal of Clinical Microbiology*, 49: 3482–90.
- Wang, Y. et al. (2016) 'Phylogenetic Evidence for Intratypic Recombinant Events in a Novel Human Adenovirus C That Causes Severe Acute Respiratory Infection in Children', *Scientific Reports*, 6: 23014.
- Yang, J. et al. (2019) 'Human Adenovirus Species C Recombinant Virus Continuously Circulated in China', *Scientific Reports*, 9: 9781.
- Yao, L. H. et al. (2019) 'Human Adenovirus among Hospitalized Children with Respiratory Tract Infections in Beijing, China, 2017-2018', *Virology Journal*, 16: 78.
- Yu, J. et al. (2019) 'Molecular Characterization of Human Respiratory Adenovirus Infection in Children from November 2016 to October 2017 in Xining City, China', *Biomedical and Environmental Sciences: BES*, 32: 38–41.
- Zhang, W. et al. (2019) 'Genome Analysis of a Novel Recombinant Human Adenovirus Type 1 in China', *Scientific Reports*, 9: 4298.
- Zhao, M. C. et al. (2020) 'Molecular and Clinical Characterization of Human Adenovirus Associated with Acute Respiratory Tract Infection in Hospitalized Children', *Journal of Clinical Virology*, 123: 104254.