



OPEN Hybrid sequencing for detailed genetic characterization of human adenoviruses

Bin Fang^{1,4}, Juan Lai^{2,4}, Yongfeng Liu^{2,4}, Lin-lin Liu¹, Xiao Yu¹, Xiang Li¹, Haili Ke², Xianfeng Zhang¹✉ & Xin Zhang^{2,3}✉

Human adenoviruses (HAdVs) are highly contagious and have significant clinical implications in the pediatric population. In the present study, we employed a combination of long-read sequencing and short-read sequencing to accurately reconstruct 32 genomes of HAdVs. The phylogenetic analyses based on the whole genome and genes revealed distinct sub-clusters within HAdV-B and -E. For HAdV-C, the phylogenetic trees constructed from *hexon*, *fiber*, and E3 gene sequences consistently matched the whole-genome phylogeny, reflecting the high sequence diversity in these regions. Notably, in regions with high sequence diversity, we observed a higher number of recombination breakpoints and lower GC content. Additionally, the E4 gene region of HAdV-C exhibited a Ka/Ks ratio > 1, indicating that positive selection may be driving the fixation of advantageous mutations. These genetic characterization analyses are crucial for enhancing future surveillance of HAdVs, facilitating a more strategic and proactive approach to monitoring their evolution, diversity, and epidemiological trends.

Keywords HAdV, ONT, FASTASeq 300, Evolution, Co-infection, Recombination

Human adenoviruses (HAdVs) are non-enveloped viruses with linear, double-stranded DNA genomes, and belong to the genus Mastadenovirus, family *Adenoviridae*. The DNA genome is ~ 35 Kb and encodes 30–40 proteins¹. The outer shell of the icosahedral-shaped capsid is built by three main proteins: hexon, penton, and fiber, contributing greatly to the early stages of HAdVs infection². Notably, these three major capsid genes are well-recognized as hot spots for homologous recombination, a key driver in the molecular evolution of HAdVs and the emergence of novel pathogenic strains^{3,4}. HAdVs are classified into seven groups (HAdV-A to HAdV-G), with further classification into 116 genotypes recognized by the Human Adenovirus Working Group (<http://hadv.wg.gmu.edu/>), in March 2024. Initially, genotypes were classified based on serum neutralization assays targeting the hexon protein, which identified serotypes 1 to 51. Subsequent classification has incorporated genetic characteristics, including phylogenetic and recombination analysis of the three major capsid proteins, and more recently, viral whole-genome sequencing (WGS) data, which has led to the identification of genotypes 52–111^{5,6}.

The progress in WGS and bioinformatics has markedly enhanced our comprehension of the HAdV genomic landscape. Beyond the taxonomic classification based on phylogenetic sequence alignment, the genomic profiling of HAdVs has also unveiled the pivotal role of recombination dynamics in their evolution. Homologous recombination, encompassing both intra- and interspecies events, is the principal driving force of HAdV evolution, significantly contributing to the emergence of novel and pathogenic strains with a propensity for broader tissue tropisms and heightened virulence⁷. A compelling case in point is HAdV-B55, a recombinant strain that arose from the intertypic recombination between HAdV-B11, typically linked to renal infections, and HAdV-B14, recognized for causing respiratory diseases. HAdV-B55 has adeptly sidestepped herd immunity against HAdV-B14 by incorporating the *hexon* gene from HAdV-B11, while retaining the respiratory pathogenicity characteristic of HAdV-B14. This dual capability renders it a formidable respiratory threat to both civilian and military populations^{8–10}. Notably, among the 61 novel pathogenic genotypes identified post HAdV-52, nearly all are recombinants¹¹. The significance of WGS is further highlighted by its capacity to elucidate the complexities of mixed and superinfection. It provides valuable insights into the genetic architecture of pathogens,

¹Hubei Provincial Center for Disease Control and Prevention, Wuhan 430079, China. ²GeneMind Biosciences Company Limited, Shenzhen 518001, China. ³State Key Laboratory of Primate Biomedical Research, Institute of Primate Translational Medicine, Kunming University of Science and Technology, Room 1405, Administration Building, No. 727 Jingming South Road, Chenggong District, Kunming 650500, Yunnan, China. ⁴Bin Fang, Juan Lai and Yongfeng Liu contributed equally to this work. ✉email: hbczczhangxf@163.com; 603986119@qq.com

enabling the identification of multiple strains or species co-existing within an infected host, and thus helps the study of recombination and genetic diversity within viral populations^{12,13}.

WGS also offers an exceptional vantage point for exploring the HAdV genetic diversity, a critical determinant in their evolutionary trajectory, pathogenicity and immune escape^{14,15}. Particularly, the heterogeneity within the major capsid genes, which are pivotal for engaging neutralizing antibodies and host cellular receptors, can confer the virus with the capacity to evade host immune responses and manifest distinct tissue affinities¹⁶. The major neutralizing epitopes, located within the hypervariable regions of hexon and characterized by high sequence diversity, are frequently engaged by the host's immune defenses and are a focus of ongoing research to develop effective vaccines and therapeutics^{17,18}. The penton and fiber proteins play important roles in host cell binding and internalization, respectively. Fiber exhibits a high affinity for the cell surface receptors, including coxsackievirus-and-adenovirus receptor (CAR), CD46, CD86, and sialic acid¹⁹, while penton base aids in viral internalization by engaging $\alpha\beta 3$ or $\alpha\beta 5$ integrins through an RGD motif. In addition, the early-transcribed regions E1A, E1B, E2, E3, and E4 encode proteins that play pivotal roles in activating transcription, facilitating viral replication, and modulating the cellular environment to promote viral production^{20–22}. Variations within these gene regions can significantly influence the virus's infectivity. The genetic diversity of HAdV genes may arise from a variety of sources, including genetic mutations, recombination events, and selective pressures from the host's immune system. Understanding the underpinnings of genetic diversity is essential for devising innovative therapeutic approaches and crafting vaccines utilizing adenovirus vectors^{23,24}.

Long-read sequencing technology adeptly navigates the obstacles that short-read sequencing often faces, such as the fragmented nature of sequences, the complexity of repetitive regions, and the heterogeneity of strains, thereby enabling the acquisition of high-quality viral genomes. Despite the advantages, the higher error rate associated with long-read sequencing can impede achieving the precision needed for base-level resolution and reliable identification of genomic variations. Consequently, hybrid sequencing has emerged as an efficacious strategy for *de novo* genome assembly. In this approach, the long reads generated by Oxford Nanopore Technologies (ONT) serve as the structural framework for assembly, adeptly resolving repetitive regions and correctly ordering contigs. Meanwhile, the short reads derived from the high-throughput sequencing platforms (e.g. GeneMind FASTASeq 300) are instrumental in refining the assembly. They enhance the sequence accuracy through polishing and clarifying any lingering ambiguities^{25–27}. This study aims to integrate ONT and FASTASeq 300 (FA) sequencing data to reconstruct highly accurate consensus sequences of various HAdV isolates and to shed light on the molecular epidemiology and genetic diversity of these viruses, contributing to a deeper understanding of their biology and evolution.

Results

Patient characteristics

A total of 26 patients with acute respiratory tract infection symptoms were enrolled in this study. Some information for three patients was missing. The gender distribution among the other twenty-three HAdV-positive patients was predominantly male, with sixteen individuals (69.57%) identified as male and seven (30.43%) as female, yielding a male-to-female ratio of 2.3:1. The age of the patients spanned from one to eleven years, with a median age of 3.78 years. The mean Ct (cycle threshold) value for the 26 samples analyzed was 11.52, varying from 7.41 to 19.75, as detailed in Supplementary Table S1. These Ct values are indicative of the relative quantity of viral nucleic acid present in the samples, with lower values suggesting higher viral loads.

Assembly and genotyping results

We collected 26 HAdV-positive samples. After meticulous removal of host DNA contamination and exclusion of low-quality reads, we generated 781,691,800 short reads with 150 bp in length, from the FA platform. The abundance of sequencing reads ensured the accuracy of sequencing outcomes by enabling the construction of a consensus sequence from multiple reads, effectively countering the effects of random errors. Additionally, we obtained a total of 1.40 million long reads (ranging from 500 to 48,000 nt) from the ONT platform (Supplementary Table S2). To evaluate the quality and completeness of our assemblies (Table 1), CheckV was applied to scrutinize the consensus sequences that were initially constructed using Nanopore long reads and subsequently polished by FA short reads. Every assembled sequence met 'High-quality' (i.e. > 90% completion) or 'complete' standard, suggesting an assembly of near-perfect fidelity suitable for subsequent genomic characterization. Notably, high-quality contigs were also identified in all short-read assemblies when utilizing solely FA reads, except S28, which cannot be assembled into a complete contig exceeding 30 Kb. Furthermore, just one contig longer than 30 Kb was found for each sample in the assembled results of FA reads. In contrast, it's intriguing to observe that eight out of 26 samples yielded multiple contigs longer than 30 Kb after de-redundancy, according to the assembled data amalgamated from two sequencing technologies. The presence of multiple adenovirus assembly sequences larger than 30 Kb within the same sample suggested the possibility of co-infection. Except for S9, all samples' genotypes, as determined by the closest nucleotide sequence alignment to a curated reference list, were highly congruent between the two assembly options. The pairwise synteny analysis also demonstrated a substantial degree of sequence homology across the whole genomes between two assembly methods, with the exception of S9_contig 1 (Supplementary Figure S1). Genotyping S9 by integrating long-read and short-read assembly discovered C2 and C5 contigs (> 35 Kb) simultaneously, while the short-read assembly alone found C2 infection. This potential co-infection needs further investigation to be interpreted and confirmed. In conclusion, our findings indicated that the hybrid assembly approach, which integrates both long and short reads, is capable of yielding complete and high-quality HAdVs genomes. Additionally, this approach holds promise for identifying co-infections and uncovering potential recombination events.

Samples	No. of contigs	Contigs	Contig length	GC (%)	Average Depth		CheckV		Completeness (%)	Contamination (%)	References/Group	aligned base/ref (%)	Identity (%)
					FA	ONT	Quality	ONT					
S1	1	S1	35,225	50.92	32,226	1946	Complete	100	0	MW748655.1/B3	94.822	94.356	
S16	1	S16	35,225	50.93	68,658	7948	Complete	100	0	MW748645.1/B3	94.886	94.967	
S21	2	S21_contig1	35,089	50.91	103,588	13,613	High-quality	99.77	0	MW748655.1/B3	94.537	94.437	
		S21_contig2	35,236	50.91	103,887	13,567	High-quality	100	0	MW748655.1/B3	94.879	94.384	
S25	1	S25	35,222	50.92	499	96	High-quality	100	0	MW748655.1/B3	94.899	94.441	
S33	1	S33	35,228	50.92	120,382	9011	High-quality	100	0	MW748655.1/B3	94.899	94.425	
S41	2	S41_contig1	35,236	50.9	115,864	16,550	High-quality	100	0	MW748655.1/B3	94.716	94.222	
		S41_contig2	35,200	50.93	115,434	16,557	High-quality	100	0	MW748655.1/B3	94.517	94.119	
S49	1	S49	35,234	50.91	77,995	4982	High-quality	100	0	MW748649.1/B3	94.985	94.508	
S5	1	S5	35,232	50.92	76,923	20,338	High-quality	100	0	MW748649.1/B3	94.985	94.514	
S59	1	S59	35,231	50.92	68,086	6837	Complete	100	0	MW748655.1/B3	94.896	94.414	
S63	1	S63	34,693	48.81	34,312	5058	High-quality	98.68	0	MK123980.1/B55	94.961	94.889	
S71	2	S71_contig1	34,709	48.81	63,662	5133	High-quality	98.51	0	KJ883521.1/B55	95.036	95.148	
		S71_contig2	34,711	48.8	63,660	5132	High-quality	98.74	0	KJ883521.1/B55	94.955	95.062	
S11	2	S11_contig1	35,157	51.11	97,137	17,029	High-quality	99.95	0	MT941568.1/B7	94.682	94.601	
		S11_contig2	35,205	51.1	97,137	17,029	High-quality	100	0	MT950360.1/B7	94.908	94.657	
S55	1	S55	35,891	55.29	76,626	27,184	High-quality	100	0	MH183293.1/C1	94.745	94.89	
S6	1	S6	35,956	55.45	76,472	25,649	Complete	100	0	MN513344.1/C1	93.953	94.032	
S60	1	S60	35,943	55.5	61,418	5160	Complete	100	0	MN513344.1/C1	94.517	94.63	
S15	1	S15	35,900	55.21	80,679	8370	High-quality	100	0	MK896858.1/C2	86.679	86.769	
S28	1	S28	35,903	55.27	128,668	38,407	High-quality	100	0	MF315029.1/C2	98.381	98.532	
S58	1	S58	35,816	55.27	71,291	28,302	High-quality	100	0	MF315029.1/C2	94.213	94.586	
S62	1	S62	35,891	55.25	60,525	9737	Complete	100	0	MK883607.1/C2	98.665	98.203	
S9	2	S9_contig1	35,848	55.22	44,868	14,334	High-quality	100	0	MH121094.1/C5	87.397	87.517	
		S9_contig2	35,880	55.23	64,959	14,454	High-quality	100	0	MH121086.1/C2	97.247	96.887	
S51	1	S51	35,309	55.32	68,123	20,039	High-quality	100	0	KF429754.1/C5	90.365	91.906	
S64	1	S64	35,906	55.15	71,874	14,432	High-quality	100	0	MF681662.1/C5	88.957	88.707	
S48	1	S48	35,886	55.22	80,426	6790	High-quality	100	0	KF268199.1/C5	88.986	89.09	
S3	1	S3	35,920	56.29	1733	141	High-quality	100	0	KF006344.1/E4	97.92	98.029	
S4	2	S4_contig1	35,346	56.48	34,973	4791	High-quality	100	0	KF006344.1/E4	96.385	98.059	
		S4_contig2	35,841	56.33	35,232	4745	High-quality	100	0	KF006344.1/E4	98.045	98.371	
S43	1	S43	35,925	56.28	15,639	9298	High-quality	100	0	KF006344.1/E4	97.925	98.021	

Table 1. Assembly statistics and quality assurance data. #No. of contigs: number of contigs. #Ref: Reference. #aligned base/ref (%): (aligned base/reference base)*100. #Identity (%): (aligned base/contigs base)*100.

Phylogenetic and diversity analysis of whole genome sequences

The phylogenetic tree (32 *de novo* assembly, 7 references) exhibited three well-defined clusters, HAdV-B, C, and E, as robustly supported by bootstrap analysis (Fig. 1A). Further delineation within the HAdV-B cluster revealed three sub-clusters, namely B3, B7, and B55, which were aligned with the existing reference sequences with strong bootstrap value support. HAdV-C also displayed three distinct sub-clusters (C1, C2, and C5), closely interwoven with their respective reference strains. Noteworthy, both FA and ONT sequencing platforms consistently achieved 100% coverage across a spectrum of sequencing depths. However, a minor deviation was observed in four samples, which show a slightly lower coverage rate of 99% at specific sequencing depths. To substantiate the findings from phylogenetic analysis, a heatmap was conducted based on the ANI values of different HAdV samples (Fig. 1B). The results reaffirmed the presence of three clusters (B, C, and E) and seven sub-cluster (B3, B7, B55, C1, C2, C5, and E4) identified in this study. It was noteworthy that S9-contig1 and S9-contig2 have been classified into distinct phylogenetic branches, suggesting a divergence in their evolutionary trajectories. In terms of genetic diversity, the various HAdV species exhibited unique profiles (Fig. 1C). HAdV-Cs were distinguished by a pronounced dichotomy between high nucleotide sequence conservation and stereotypical focal variations in regions including the *hexon* gene, *fiber* gene, and *E3* transcription unit. In contrast to HAdV-Cs, HAdV-Bs showed a relatively higher degree of genomic diversity, with an additional peak of high variability within the *penton* gene region, in addition to the *hexon*, *fiber*, and *E3* regions. HAdV-Es, on the other hand, exhibited the lowest genetic diversity across the entire genomes compared to HAdV species B and C. This reduced diversity

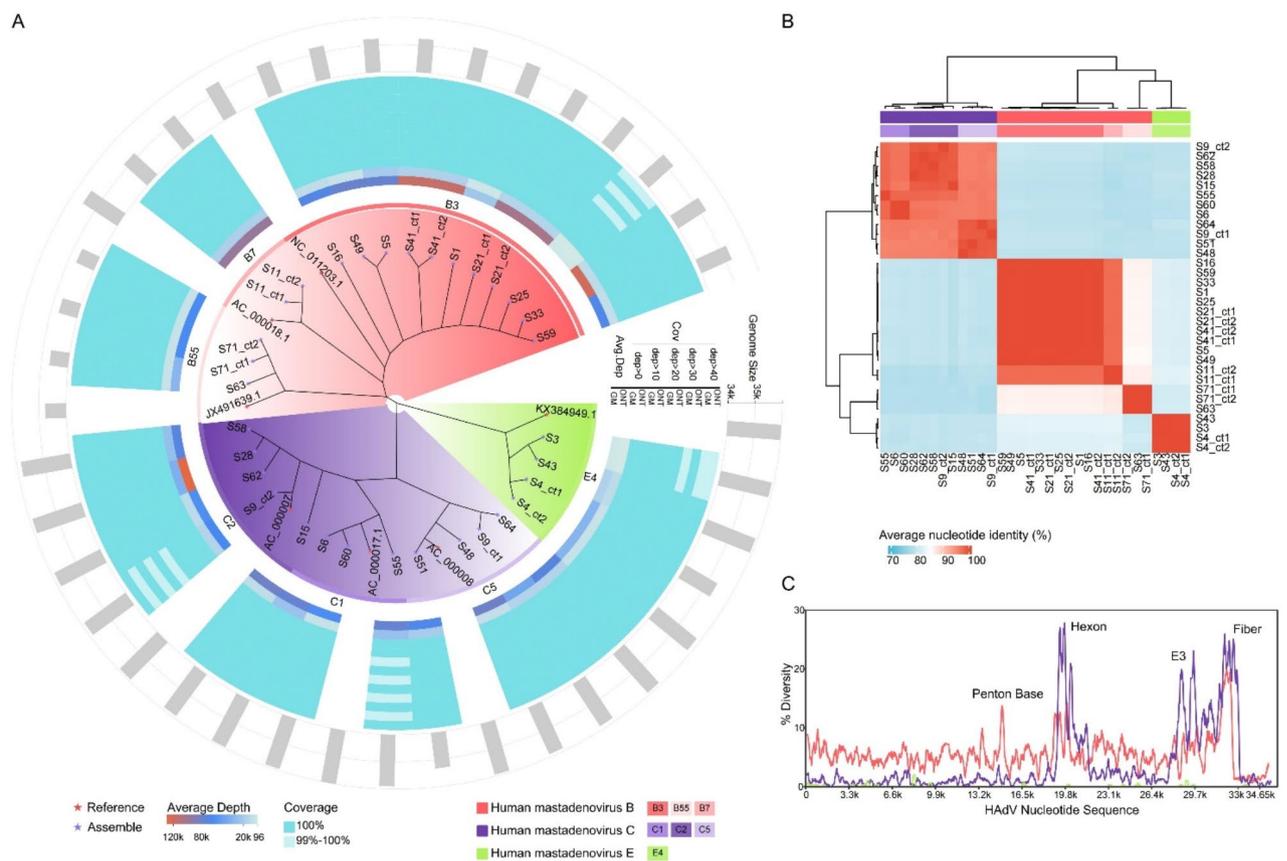


Fig. 1. Phylogenetic and diversity analysis of whole genome sequences. **(A)** The phylogeny of HAdV strains was constructed by using the maximum likelihood method and the Kimura two-parameter model in MEGA (11.0) with 1000 replicates. The phylogeny tree plot was visualized via tvBOT. HAdVs were clustered together to form three distinct clades, and each clade with high bootstrap support (> 95). The inner to outer rings display the average sequencing depth, the coverage at different sequencing depths (cyan blocks), and the genome size (the outermost grey bars). The height of the bar corresponds to the size of the genome, which was typically around 35 kb. The coverage across various depths was fundamentally 100%. The violet and red star symbols represent *de novo* assembled genomes and reference sequences in this study, respectively. The color column indicates different HAdV genotypes. **(B)** Heat map of the ANI for HAdV strains calculated by FastANI (v1.33). Higher values (red) correspond with higher genetic homology and lower phylogenetic distance between HAdV strains. **(C)** Nucleotide diversity plots displayed the mean percentage of nucleotide differences per site across the genome. The values were calculated by DnaSP (v6.12.03). The colors light coral, while purple, and green represented HAdV-B, HAdV-C, and HAdV-E species, respectively. ct, contig; Avg., average; Cov, coverage; dep, depth.

could potentially be attributed to the limited samples of HAdV-E identified in the study and their close genetic relationship.

Co-infection inference

As previously mentioned, genotyping based on reference alignment revealed a mixed infection in sample S9. This finding was further substantiated by the phylogenetic dendrogram, where, S9-contig1 was found to be clustered within a monophyletic cluster specific to HAdV-C5, while S9-contig2 exhibited a pronounced evolutionary relationship with HAdV-C2 strains (Fig. 1A and B). Figure 2A illustrated the outcomes of an *in silico* restriction enzyme analysis, which corroborated the genetic similarity between S9-contig1 and the reference HAdV-C5 virus, as well as between S9-contig2 and the reference HAdV-C2 virus, when digested by the endonuclease *Bal* I, *Xba* I and *Sac* I. However, upon closer examination, minor variations in the restriction profiles of S9-contig1 were observed in comparison to HAdV-C5 under the endonucleases *Dra* I and *Bam*H I, indicating some level of genetic divergence. ANI analysis delineated two distinct clades nested with S9-contig1 and S9-contig2 (Fig. 2B). Additionally, sequence alignment revealed high pairwise genome identities, with remarkably high values of 99.69% between S9-contig2 and HAdV-C2, and 98.68% between S9-contig1 and HAdV-C5. Furthermore, a considerable degree of molecular similarities was observed across the *hexon*, *penton*, and *fiber* genes, albeit with occasional point mutations, as depicted in Fig. 2C and D. In conclusion, the coexistence of these nearly full-length contigs, each affiliated with distinct genotypes, highlighted the superiority of *de novo* assembly combining ONT long reads and FA short reads. This approach confirmed the intricate nature of HAdV infection in sample S9 and also highlighted the importance of meticulous genotyping in uncovering the underlying evolutionary dynamics and the complexity of viral infections.

Phylogenetic and diversity analysis of viral genes

The clustering of *hexon* and *fiber* gene sequences consistently mirrored the clustering observed in complete genomic sequences, albeit with some divergence in the *penton* gene clustering pattern, particularly among HAdV-C species (Fig. 3A-C). Notably, four HAdV-C5 strains (S64, S48, S51, S9_contig1) were found to be phylogenetically interspersed within the five HAdV-C2 strains (S15, S28, S58, S62, S9_contig2) regarding the *penton* gene region. Similarly, there were no clear subclades of HAdV-C observed in the *E1A*, *E1B*, *DNA pol*, and *E4* gene regions, while the *E3* sequences clustered in accordance with the complete genomic sequences (Fig. 3D-H). Further, we employed ANI analysis to investigate gene-level sequence similarity. The analysis revealed a high degree of genetic conservation within the *E1A*, *E1B*, *DNA pol*, *penton*, and *E4* gene regions of HAdV-C, whereas the *hexon*, *fiber*, and *E3* gene regions displayed a lower percentage of nucleotide identity (Fig. 3I). Correspondingly, the variability observed within the *hexon*, *fiber*, and *E3* gene regions indicated a higher level

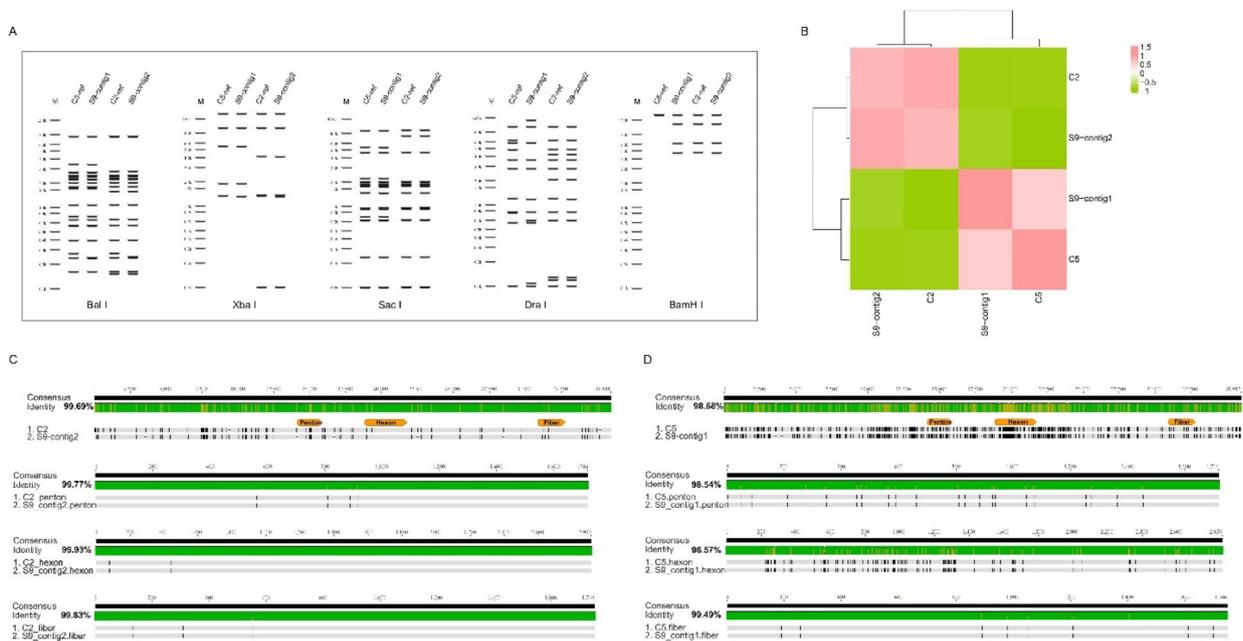


Fig. 2. Investigation of S9 co-infection. (A) *In silico* restriction enzyme analysis of S9 contigs and closely related genotypes, C2 and C5. (B) ANI analysis. (C, D) Nucleotide sequence alignment of the complete genome sequences and major capsid genes *hexon*, *penton* base, and *fiber*. Pairwise identities were marked on the figure. Vertical black lines indicate disagreements compared to the reference and dashes indicate sequence gaps. These analyses were conducted in Geneious Prime.

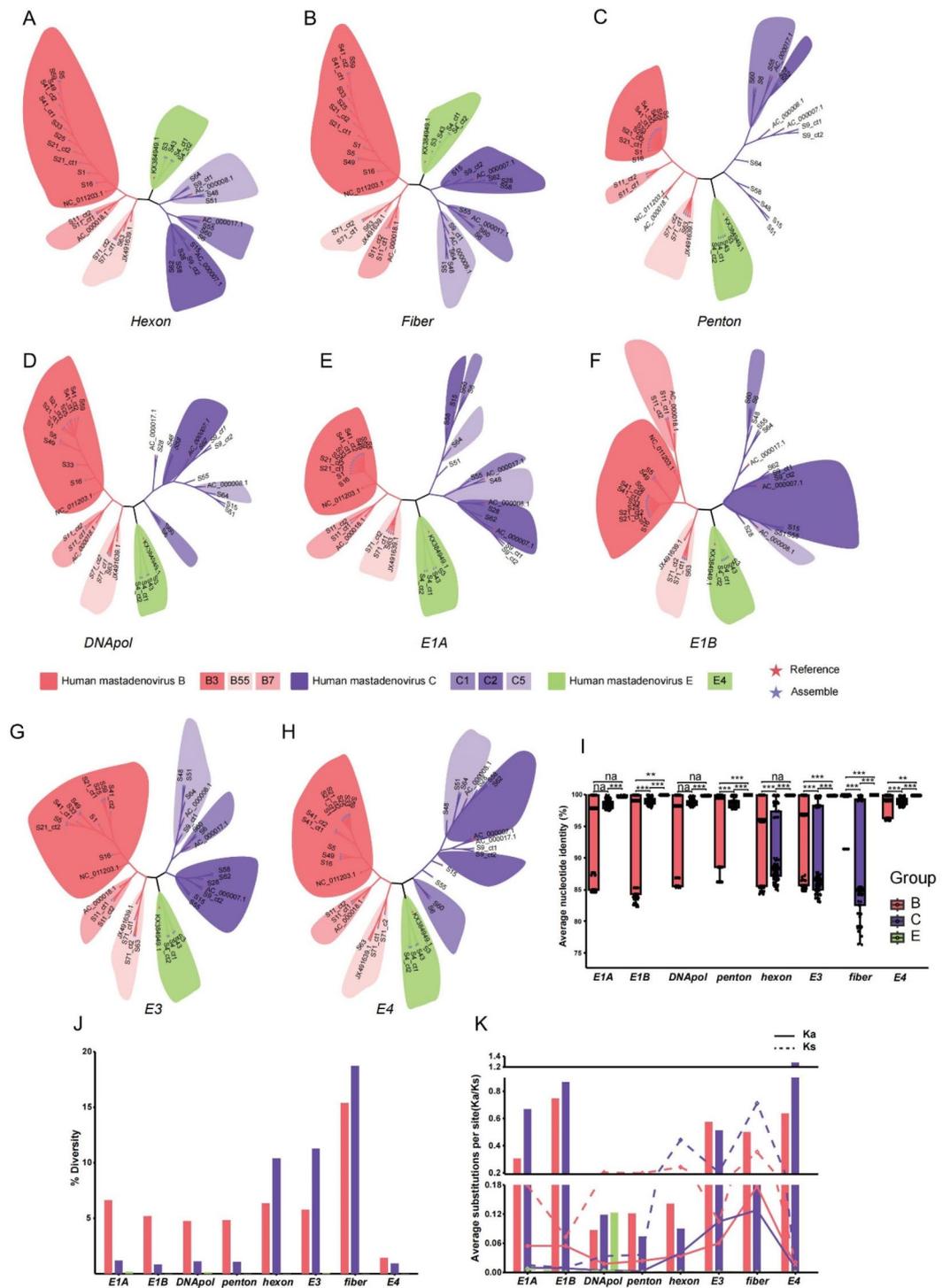


Fig. 3. Analysis of the phylogeny, genetic divergence, and diversity of 8 HAdVs gene regions. **(A–H)** Phylogenetic trees of the *hexon*, *fiber*, *penton*, *DNA pol*, *E1A*, *E1B*, *E3*, and *E4* genes, respectively. The phylogenetic tree was calculated by using the maximum-likelihood method and the Kimura two-parameter model in MEGA (11.0). The statistical robustness of branches was estimated by 1,000 bootstraps. The phylogenetic tree plot was visualized via tvBOT. **(I)** The ANI of different gene regions were calculated by FastANI. **(J)** Mean nucleotide diversity representing the average number of nucleotide differences per site. **(K)** The synonymous (K_s) and non-synonymous (K_a) mutations of 8 HAdVs gene regions calculated using DnaSP (v6.12.03) software. K_a and K_s changes were represented in solid and dashed lines, respectively. The bars represent the ratio (K_a/K_s) for each gene.

than other genes within the HAdV-C genome (Fig. 3J). Generally, HAdV-B genes exhibited greater diversity, particularly in the *hexon*, *fiber*, and *E3* gene regions. Furthermore, the synonymous (Ks) and nonsynonymous (Ka) substitution rates, recognized indicators of selective pressure, were calculated for each of the eight genes of HAdV-B, C, and E species (Fig. 3K). As anticipated, the majority of HAdV-B and HAdV-C genes exhibited a Ka/Ks ratio smaller than 1, implying that these genes were predominantly under purifying selection. On the other hand, the Ka/Ks ratio of the *E4* gene region of HAdV-C displayed a Ka/Ks ratio exceeding 1, indicating the action of positive selection. Given the pivotal role of the *E4* transcription unit in modulating cellular functions and in facilitating viral DNA replication and RNA processing, we hypothesize that the observed positive selection in this region may enhance the virus's fitness. This selective advantage may lead to the emergence of viral variants better adapted to the host environment, thus augmenting their capacity to evade host defenses, optimize replication strategies, and manipulate cellular machinery for their benefit. It is important to note that this hypothesis necessitates further empirical investigation to validate its implications fully.

Recombination analysis

The results of reticulate networks and pair-wise homoplasy index (*Phi*) test indicated the recombination promise of HAdV-B and HAdV-C strains (Fig. 4A and B). Compared with HAdV-B, HAdV-C exhibited a higher degree of reticulations, indicative of recombination, which was corroborated by the phi test with significant support ($p = 0$). Besides, the topologies of these recombination networks delineated three distinct phylogenetic branches within the HAdV-B and HAdV-C sequences, consistent with the evolutionary relationships (Fig. 1A). Then, RDP4 software was employed to identify the recombination events and to localize potential recombination breakpoints. A total of six recombination events within the HAdV-B strains and thirty within HAdV-C strains were identified, respectively, each supported by more than three algorithms (Supplementary Tables S3 and S4). The breakpoint distribution, ascertained using a sliding window strategy, further validated the recombination events of HAdV-B and HAdV-C. Moreover, this analysis also concurred with the observation that HAdV-C experienced a higher number of recombination events (Fig. 4C and D). The breakpoints were found to be dispersed throughout the entire genome, with two notable hotspot regions corresponding to the *hexon* gene and a region proximal to the *fiber* gene, especially for HAdV-C. This widespread distribution underscores the possibility of dynamic nature of recombination in shaping the genetic diversity of these HAdV strains. It is widely acknowledged that co-infection with at least two strains of virus is the prerequisite for recombination, which is one of the key mechanisms for genetic diversification²⁸. To delve deeper into the recombination events in the S9 sample, which is indicative of a potential mixed infection, we conducted an exhaustive recombination analysis using both BootScan and RDP4 software. Our findings highlighted a remarkable genomic similarity, with S9-contig2 exhibiting nearly 100% sequence identity to the C2 reference strains (NC_001405.1). In contrast, S9-contig1 seemed to be a recombinant sequence, potentially involving the C5 reference strain (AC_000008.1) and S9-contig2 (Fig. 4E). Subsequent in-depth analysis with BootScan corroborated this hypothesis, pinpointing the C5 reference strain and S9-contig2 as the parental strains contributing to the genetic composition of S9-contig1 (Fig. 4F).

Comparative genomics analysis

Pairwise synteny analysis revealed a high degree of sequence similarity among strains within the HAdV-B species (Fig. 5A, Supplementary Figure S2A), as well as HAdV-C species (Fig. 5B, Supplementary Figure S2B). It has been reported that GC-rich motifs are associated with increased genomic stability and may contribute to relative resistance to homologous recombination²⁹. To validate this hypothesis, a gene-by-gene analysis for GC content was performed on the genomes of HAdV-B and C. This analysis, in conjunction with the prior assessment of gene diversity in Fig. 3J, disclosed that the genomic regions more prone to homologous recombination exhibited a marked decrease in GC content. This decrease was particularly evident within the *hexon*, *E3*, *fiber*, and *E4* gene regions. The findings suggested a correlation between GC content and the propensity for recombination events, providing insight into the genomic architecture that influences the evolutionary dynamics.

Discussion

The global impact of disease caused by HAdVs has indeed propelled research into pathogen detection, typing, and analysis. Undoubtedly, the advent of high-resolution genomic data coupled with sophisticated bioinformatics tools has accelerated the studies of viral molecular evolution and genetic characterization, and revolutionized the conventional serology-based typing method^{16,9,30}. By leveraging the complementary strengths of both long- and short-read sequencing, our integrated approach generated accurate consensus sequences with high accuracy, long read, and cost-effectiveness. Thirty-two HAdV genome sequences from 26 samples were obtained in this study, which can serve as a valuable reference cohort to facilitate high-resolution analyses of HAdV genomic characteristics and contribute to a deeper understanding of their epidemiology and evolution.

In the present retrospective study, HAdV-B3 emerged as the predominant genotype, representing 34.38% (11/32) of the typed HAdV infections, followed by HAdV-C2 at 15.63%, HAdV-C5 at 12.5%, and HAdV-E4 at 12.5%. These predominant isolates are commonly implicated in respiratory HAdV infection, both in China and globally^{31,32}. The diversity of HAdV genotypes in the limited samples highlights the intricate dynamics of adenovirus transmission and the urgency for a robust surveillance system. Notably, during the present study period, an outbreak of acute respiratory infections in children caused by HAdV-B7 was also reported in Hubei Province³³. This finding emphasizes the concurrent circulation of HAdV-B3 alongside other adenovirus types, such as HAdV-B7, within regions experiencing HAdV epidemics. This co-circulation of multiple HAdV genotypes, generally, may facilitate genetic recombination events, potentially leading to the emergence of novel strains with enhanced virulence or transmissibility. Furthermore, the hypervariable regions of globally circulating HAdV-3 strains have exhibited a pronounced degree of heterogeneity and have undergone continuous mutations

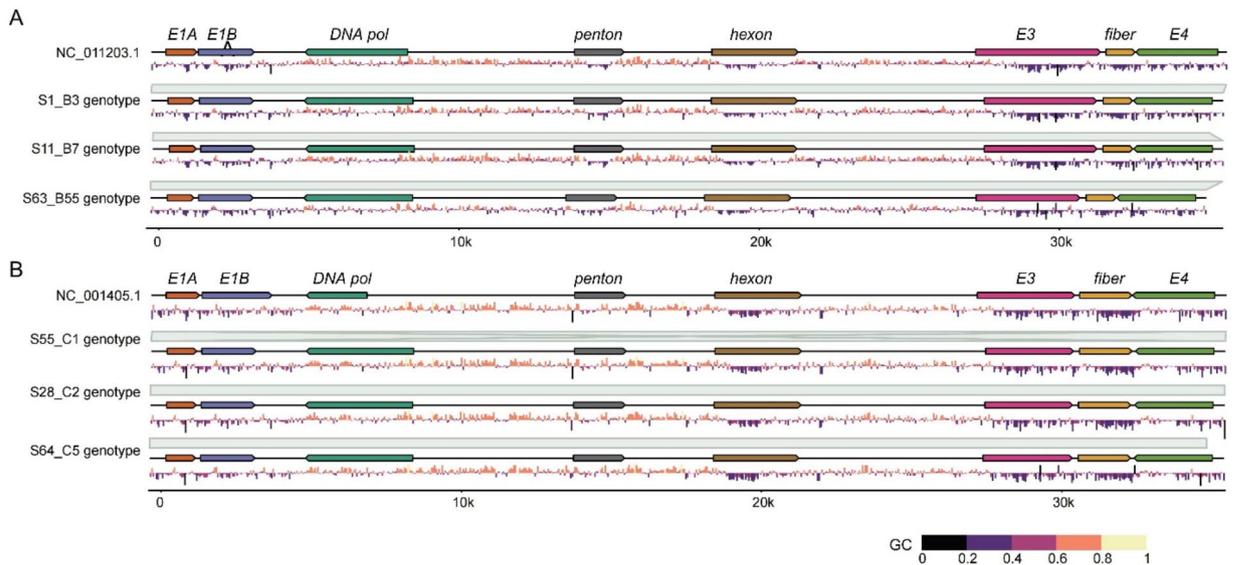


Fig. 5. Genomic comparison and characterization of HAdV genomes. **(A)** Representation of HAdV-B. **(B)** Representation of HAdV-C. The genes of interest were distinctly annotated within the schematic genomic diagrams, presented at the summit of the illustration for each strain. Syntenic regions were identified between each pair and were connected by gray-colored blocks, facilitating visual comparison. The peak plot depicted the distribution of GC content across the genome. In the GC content plot, a value of 50% represents the midline, with GC contents exceeding 50% positioned above the horizontal reference line, and those below 50% situated below this line.

since their original isolation, which may account for the increased prevalence of HAdV-3 respiratory infections worldwide³³. HAdV-C DNA has been reported to persist in a latent state in lymphoid cells following primary infection^{34,35}. This prolonged latency, coupled with the high prevalence of infections with HAdV-C types, may increase the probability of superinfection, thereby promoting the evolution of novel HAdV-C types by intertypic recombination. Intriguingly, the mixed infection inferred in this study for sample S9 was specifically associated with HAdV-C genotypes, highlighting the complexity of HAdV infections and the importance of detailed genotyping in understanding their epidemiological patterns and evolutionary potential.

In the gene-level phylogenetic analysis for HAdV-C, we observed multiple sequence stretches that lacked clear clustering within the *penton*, *DNA pol*, *E1A*, *E1B*, and *E4* gene regions, and correspondingly these genes exhibited higher nucleotide identity and lower sequence diversity than the *hexon*, *fiber* and *E3* sequences (Fig. 3). As key structural components, three major capsid genes, *penton*, *hexon*, and *fiber* genes are commonly utilized for viral molecular typing and recombination identification^{6,11}. In the present study, the genetic types of *hexon* and *fiber* were readily identifiable due to their high variability, while the *penton* base showed less distinct typing branches, suggesting that its genetic variation may be insufficient for precise genotyping. This observation is supported by previous studies, which reported a nucleotide diversity of 0.008 for the HAdV-C *penton* base, significantly lower than those of HAdV-B (0.102) or HAdV-E (0.02471)³⁶. Our analysis of the *penton* base sequence diversity also confirmed its relatively lower genetic variability (Fig. 3). Moreover, the high genetic conservation observed in the *E1* region, *DNA pol*, and *E4* can be attributed to functional constraints, as these genes play essential roles in the viral life cycle, including virus-host interactions, viral DNA replication, cell-cycle modulation, host's immune response interference, etc. Interestingly, the *E3* gene region of HAdV-C showed clear clustering, as the *E3* sequences of different HAdV-C prototypes were highly divergent, consistent with a previous report³⁷. The phylogenetic consistency between the *hexon* and *fiber* genes with the *E3* gene regions suggested the co-evolution potential among these gene regions in HAdV-C species. Furthermore, the analysis of recombination breakpoint distribution in HAdV-C revealed a distinct peak within the *E3* gene region, indicating that the *E3* region is a hotspot for recombination events in HAdV-C, which likely contributed significantly to the observed high sequence diversity in this region (Fig. 3). Considering the immunomodulatory function of *E3* region, its genetic diversity presented a unique opportunity to explore the evolutionary adaptations of HAdVs in response to selective pressures from host immune systems.

The genes of HAdVs were predominantly subjected to purifying selection ($Ka/Ks < 1$), which is instrumental in preserving the long-term stability of biological function by eliminating deleterious mutations³⁸. However, our study detected a striking deviation of the *E4* gene region of HAdV-C, where a Ka/Ks ratio > 1 was observed. This finding suggested the influence of positive selection, potentially conferring a selective advantage by promoting the fixation of beneficial mutations within the *E4* gene. Nonetheless, the specific role of positive selection on the *E4* gene has not been thoroughly investigated in the current literature, highlighting the need for additional experimental validation to support this hypothesis. The *E4* transcription unit, comprising seven open reading frames (ORFs), encodes proteins that modulate cellular functions and facilitate viral DNA replication and

transcription²². Notably, *E4* ORF3 has been implicated in attenuating the innate immune response of cells infected with HAdV³⁹. Therefore, the positive selection action on the *E4* region may be associated with the refinement of immune evasion strategies or the adaptation to novel host environments. Future studies should focus on elucidating the functional implications of these nonsynonymous substitutions within the *E4* region. Such studies are expected to yield novel insights into the molecular mechanisms that drive viral adaptation and pathogenesis, furthering our understanding of HAdV evolution in the context of host-virus interactions.

This study had several inherent limitations. Firstly, the limited sample size constrained the scope of our conclusions. The 32 strains identified as belonging to HAdV-B, -C, or -E may not fully represent the broader diversity of adenoviruses in circulation, which could impact the generalizability of our findings. Consequently, we recommend that future studies should incorporate a larger and more diverse sample cohort to bolster the robustness of the results and to offer a more comprehensive understanding of adenovirus genomic characterization. Secondly, this study focused on investigation of HAdVs' genetic characterization utilizing sequencing methodologies and bioinformatics analysis, however, it did not encompass *in vitro* experiments. Especially for the inferred co-infection sample S9, ideal confirmation would involve isolating HAdV strains of genotypes C2 and C5 from this clinical specimen. However, regrettably, we do not possess additional clinical samples to facilitate further validation of this inference. Although the limited quantities of DNA samples hampered the ability to perform *in vitro* experimental validation, the results of this study provide valuable insights into the genomic characteristics of adenoviruses.

In conclusion, this study assembled 32 HAdV genomes by integrating reads from the ONT and FA platforms. Following the acquisition of reliable genotyping outcomes, we conducted phylogenetic, diversity, and recombination analyses, achieving a comprehensive and systematic understanding of the genetic characterization of HAdV-B, -C, and -E. This work highlighted the merits of integrating long- and short-read sequencing technologies, enabling the detection of potential co-infection cases at the genomic level. Additionally, the high sequence diversity and potential recombination events observed, particularly within the *hexon*, *fiber*, and *E3* gene regions, underscored the necessity of HAdV epidemiological surveillance in China. Lastly, the sequencing dataset generated in this study has enriched the public repository with a trove of high-quality reference sequences, serving as a valuable resource for the analysis and understanding of future emerging viruses. The findings in this study provide insights in developing effective diagnostic tools, therapeutic strategies, and vaccines against HAdVs.

Materials and methods

Ethics statement and human sample collection

This study was approved by the research ethics board of the Hubei Provincial Center for Disease Control and Prevention. The need for informed consent from participants was waived as the retrospective samples were collected from the influenza surveillance system of Hubei Province, and all the data in the surveillance system were anonymized. Nasopharyngeal swab specimens from pediatric patients with symptoms of respiratory infection were collected from the influenza surveillance system of Xianning, Hubei Province, from 2018 to 2019. All methods were performed in accordance with relevant guidelines and regulations. Respiratory virus screening was performed using a commercial D³ Ultra DFA Respiratory Virus Screening and ID Kit (Diagnostic Hybrids, USA). 26 HAdV-positive samples were collected and stored at -80°C .

HAdV cultivation and genomic DNA extraction

Followed by three freeze-thaw cycles, 200 μL HAdV-positive samples were then filtered through 0.22 syringe filters into cell plates containing human laryngeal carcinoma cells HEp-2 for virus isolation and propagation. The cells were grown in Dulbecco's Modified Eagle Medium supplemented with 100 IU/mL penicillin, 100 $\mu\text{g}/\text{mL}$ streptomycin, and 2% (v/v) fetal bovine serum until 90% cytopathic effects were observed. The viral genomic DNA was extracted using the EZ1 Virus Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. The extracted DNA was stored at -80°C for later library construction.

Library preparation and sequencing

ONT MinION Mk1C

AMPure[®] XP beads (Beckman Coulter, USA) were used for viral DNA purification. DNA concentration and absorbance were measured with the Qubit 1X dsDNA HS Assay Kit equipped with a Qubit[™] 4 Fluorometer (Thermo Fisher Scientific, USA). The library was prepared using the Nanopore Rapid Barcoding Kit SQK-RBK004 (ONT). Specifically, we utilized the Fragmentation Mix RB provided in the kit to perform fragmentation of the DNA and concurrently attach a pair of unique barcodes to the ends of DNA fragments. This was followed by the attachment of sequencing adapters. Subsequently, the constructed libraries were sequenced using the FLOW-MIN106D R9.4.1 flow cells on the MinION Mk1C platform (ONT).

GeneMind FA

The viral DNA was purified and quantified following the aforementioned methods, hereafter, subjected to Nextera XT DNA Library Preparation Kit (Illumina, San Diego, CA, USA) for library construction, according to the manufacturer's instructions. Briefly, the DNA samples were fragmented by sonication to a size of 300 bp, then DNA fragments were end-polished, A-tailed, and ligated with the full-length adaptor for sequencing, followed by PCR amplification. Then, the PCR products were purified and quantified, and the libraries were analyzed for size distribution. Final libraries were sequenced on FA to generate 2×150 bp paired-end reads.

Raw data processing

Firstly, the sequencing data of ONT was filtered using Porechop 0.2.4 (default parameters) to remove adapters. Thereafter, minimap2 (2.24-r1122, -x map-ont) was used to map the sequencing reads to the human reference genome (GCF_000001405.40_GRCh38.p14) to discard host reads, followed by alignment to adenovirus reference genomes to obtain the adenovirus specific reads. Finally, NanoFilt (2.8.0-q 10 -l 500 --maxlength 50000 --headcrop 50) was invoked to further screen the specific reads to get high quality, long-read adenovirus reads (Qscore ≥ 10 , length: 500 ~ 50,000 bp). The sequencing data of FA was aligned to the human reference genome (GCF_000001405.40_GRCh38.p14) using Bowtie2 (version 2.4.5, --end-to-end) to remove host reads. SOAPnuke (2.1.6, -n 0.001 -l 20 -q 0.2 --polyX 50 --ada_trim adaMR=0.25) was then applied to filter low-quality reads as follows: (1) reads with ≥ 60 bp with a quality score of < 20 ; (2) reads with ≥ 1 N base; and (3) reads with an overlap of ≥ 8 bp with adapters.

Genome assembly and quality assessment

Genome assembly combining ONT long-read with FA short-read reads

Using Proovread (v2.14.1) with the default parameters, the trimmed FA reads were utilized to polish the processed ONT reads. The cleaned ONT reads from different samples were then individually assembled with the Canu assembler (v2.2; genomeSize = 45k) under the data's actual maximum sequencing depth. To obtain the final consensus sequences, Pilon (v1.24, --fix snps, indels) used the FA short reads and default parameters to perform error correction on the assembled contigs. To obtain the final sequences for later genomic characterization, we extracted the sequences longer than 34 Kb⁴⁰ and subjected them to CD-HIT (v4.8.1, -c 0.9 -aS 0.8 -d 0) for de-redundancy. The quality of assembled sequences was assessed using CheckV (v1.0.1, end_to_end), an automated pipeline for estimating the quality and completeness of viral genomes.

Genome assembly using FA short-read reads only

The trimmed FA paired-end reads were assembled using the MEGAHIT (v1.2.9) with parameters as follows: minimum k-mer size, 21 bp; maximum k-mer size, 141 bp; k-mer step size, 10 bp; minimum contig length, 34 Kb to obtain the consensus sequence. Similarly, CheckV (v1.0.1) was invoked to evaluate the assembly quality with default parameters.

Genotyping and annotation

The assembled sequences were aligned to a curated reference list of complete adenovirus genome sequences using Minimap2 (v2.26, -x asm5) to identify the most closely related HAdV types. The reference dataset consisted of 1,165 human mastadenovirus genomes covering all A–G species from the NCBI database. Annotation of each element and open-reading-frame (ORF) was conducted according to HAdV reference genome using a genome annotation program, VIGOR (v3.0, default parameters).

Phylogenetic analysis

This study enrolled a total of 39 whole-genome sequences for phylogenetic analysis, including seven reference sequences from the NCBI database and 32 newly assembled genomes. In addition, the sequences of *E1A*, *E1B*, *DNA polymerase*, *penton*, *hexon*, *E3*, *fiber*, and *E4* gene regions were also selected for phylogenetic analysis. Multiple sequence alignments were conducted using ClustalW (v2.1, default parameters) for the *E1A*, *E1B*, *E3*, and *E4* gene regions and whole-genome sequences, and MUSCLE (v3.8.31, default parameters) for the *DNA polymerase*, *penton*, *hexon*, and *fiber* genes. Automated alignment trimming was performed using trimAl (v1.4.1, default parameters) software. Phylogenetic trees were inferred using MEGA (v11.0) with the maximum likelihood method and the Kimura two-parameter model, validated by 1,000 bootstrap replicates for robustness. The phylogenetic trees were visualized with tvBOT.

Nucleotide identity and diversity analysis

The nucleotide identity of the assembled draft genomes and genes was calculated using FastANI (v1.33), employing an alignment-free approach to compute the average nucleotide identity (ANI). FastANI was configured with 500 bp and 200 bp fragment settings for whole-genome and gene sequences, respectively. Whole-genome nucleotide diversity (π) was calculated using the DNA polymorphism option in the DNA Sequence Polymorphism (DnaSP v6.12.03) software⁴¹, with a window size of 200 nt and a step size of 20 nt. The gene nucleotide diversity (π) was calculated after excluding gapped sites. Additionally, nonsynonymous substitutions per nonsynonymous site (Ka) and synonymous substitutions per synonymous site (Ks) were determined using the synonymous and nonsynonymous substitutions option in DnaSP v6.12.03.

Recombination analysis

Potential genomic recombination events were identified based on the sequence alignment results from RDP4 (v 4.95)⁴², SimPlot (v3.5.1)³⁶ and SplitsTree6 (v 6.3.27, <https://github.com/husonlab/splitstree6/releases/tag/6.3.27>). Recombination network trees were constructed using SplitsTree6, and the pair-wise homoplasy (*Phi*) test was applied to assess the statistical significance of the recombination networks. Within RDP4, six distinct programs—RDP, GENECONV, 3Seq, Maximum Chi Square (MaxChi), SiScan, and BootScan—were utilized with default settings to detect recombination breakpoints. Only breakpoints identified by at least three programs with a Bonferroni-corrected P value < 0.05 were reported. Duplicates and flagged breakpoints were excluded. The distribution of recombination breakpoints along the genome was illustrated in a dedicated plot.

Comparative genomics analysis

Genome annotation, encompassing eight core genes of interest, was schematically represented using the ggenes R package (version 0.5.1). Additionally, a comparative genomics plot, depicting the positions of the eight core genes, syntenic regions, and GC content across the genomes, was visualized with the gggenomes R package (version 1.00).

Data availability

The authors confirm that the data supporting the findings of this study are available within the article. The datasets of FASTASeq 300 and ONT analyzed during the current study are available from China National GeneBank DataBase with the accession number CNP0005871 and CNP0005888, respectively.

Received: 19 August 2024; Accepted: 5 November 2024

Published online: 27 November 2024

References

1. Davison, A. J., Benkő, M. & Harrach, B. Genetic content and evolution of adenoviruses. *J. Gen. Virol.* **84**, 2895–2908 (2003).
2. Medina-Kauwe, L. K. Development of adenovirus capsid proteins for targeted therapeutic delivery. *Therapeutic Delivery*. **4**, 267–277 (2013).
3. Ismail, A. M. et al. Genomic analysis of a large set of currently—and historically—important human adenovirus pathogens. *Emerg. Microbes Infections*. **7**, 1–22 (2018).
4. Walsh, M. P. et al. Evidence of molecular evolution driven by recombination events influencing tropism in a novel human adenovirus that causes epidemic keratoconjunctivitis. *PLoS One*. **4**, e5635 (2009).
5. Aoki, K. et al. Toward an integrated human adenovirus designation system that utilizes molecular and serological data and serves both clinical and fundamental virology. *J. Virol.* **85**, 5703 (2011).
6. Seto, D., Chodosh, J., Brister, J. R., Jones, M. S. & Community, A. R. Using the whole-genome sequence to characterize and name human adenoviruses. *J. Virol.* **85**, 5701 (2011).
7. Robinson, C. M. et al. Molecular evolution of human adenoviruses. *Sci. Rep.* **3**, 1812 (2013).
8. Walsh, M. P. et al. Computational analysis identifies human adenovirus type 55 as a re-emergent acute respiratory disease pathogen. *J. Clin. Microbiol.* **48**, 991–993 (2010).
9. Seto, D., Jones, M. S., Dyer, D. W. & Chodosh, J. Characterizing, typing, and naming human adenovirus type 55 in the era of whole genome data. *J. Clin. Virol.: Off. Publ. Pan Am. Soc. Clin. Virol.* **58**, 741–742 (2013).
10. Zhang, Q., Seto, D., Cao, B., Zhao, S. & Wan, C. Genome sequence of human adenovirus type 55, a re-emergent acute respiratory disease pathogen in China. *Am. Soc. Microbiol.* **86**, 12441–12442 (2012).
11. Wu, X. et al. Molecular typing and rapid identification of human adenoviruses associated with respiratory diseases using universal PCR and sequencing primers for the three major capsid genes: Penton base, hexon, and fiber. *Front. Microbiol.* **13**, 911694 (2022).
12. Myers, C. E. et al. Using whole genome sequences to investigate adenovirus outbreaks in a hematopoietic stem cell transplant unit. *Front. Microbiol.* **12**, 667790 (2021).
13. Ghedin, E. et al. Mixed infection and the genesis of influenza virus diversity. *J. Virol.* **83**, 8832–8841 (2009).
14. Oliveira, E. R. & Bouvier, M. Immune evasion by adenoviruses: A window into host–virus adaptation. *FEBS Lett.* **593**, 3496–3503 (2019).
15. Sanjuan, R. & Domingo-Calap, P. Genetic diversity and evolution of viral populations. *Encyclopedia Virol.*, 53 (2021).
16. Matsushima, Y. et al. Construction of new primer sets for corresponding to genetic evolution of human adenoviruses in major capsid genes through frequent recombination. *Jpn. J. Infect. Dis.* **67**, 495–502 (2014).
17. Qiu, H. et al. 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 (2012).
18. Bradley, R. R. et al. Adenovirus serotype 5-specific neutralizing antibodies target multiple hexon hypervariable regions. *J. Virol.* **86**, 1267–1272 (2012).
19. Zhang, Y. & Bergelson, J. M. Adenovirus receptors. *J. Virol.* **79**, 12125–12131 (2005).
20. Flint, J. Organization of the adenoviral genome. *Adenoviruses: Basic. Biology Gene Therapy*, 17–30 (1999).
21. Horwitz, M. S. Function of adenovirus E3 proteins and their interactions with immunoregulatory cell proteins. *J. Gene Medicine: cross-disciplinary J. Res. Sci. gene Transf. its Clin. Appl.* **6**, 172–S183 (2004).
22. Weitzman, M. D. Functions of the adenovirus E4 proteins and their impact on viral vectors. *Front. Biosci.* **10**, e17 (2005).
23. Liu, L. et al. Genetic diversity and molecular evolution of human adenovirus serotype 41 strains circulating in Beijing, China, during 2010–2019. *Infect. Genet. Evol.* **95**, 105056 (2021).
24. Weklak, D. et al. Genetic and chemical capsid modifications of adenovirus vectors to modulate vector–host interactions. *Viruses*. **13**, 1300 (2021).
25. Zhou, L. et al. Long-read sequencing unveils high-resolution HPV integration and its oncogenic progression in cervical cancer. *Nat. Commun.* **13**, 2563 (2022).
26. Boldogkői, Z., Moldován, N., Balázs, Z., Snyder, M. & Tombác, D. Long-read sequencing—a powerful tool in viral transcriptome research. *Trends Microbiol.* **27**, 578–592 (2019).
27. Oehler, J. B., Wright, H., Stark, Z., Mallett, A. J. & Schmitz, U. The application of long-read sequencing in clinical settings. *Hum. Genomics*. **17**, 73 (2023).
28. Trémeaux, P. et al. SARS-CoV-2 co-infections and recombinations identified by long-read single-molecule real-time sequencing. *Microbiol. Spectr.* **11**, e00493–e00423 (2023).
29. Gruss, A., Moretto, V., Ehrlich, S., Duwat, P. & Dabert, P. GC-rich DNA sequences block homologous recombination in vitro. *J. Biol. Chem.* **266**, 6667–6669 (1991).
30. Lin, B. et al. Broad-spectrum respiratory tract pathogen identification using resequencing DNA microarrays. *Genome Res.* **16**, 527–535 (2006).
31. Esposito, S. et al. Epidemiology and clinical characteristics of respiratory infections due to adenovirus in children living in Milan, Italy, during 2013 and 2014. *PLoS One*. **11**, e0152375 (2016).
32. Yu, J., Zhao, S. & Rao, H. Molecular characterization of human respiratory adenoviruses infection in Xining City, China in 2018. *Viol. Sin.* **36**, 545–549 (2021).
33. Haque, E., Banik, U., Monwar, T., Anthony, L. & Adhikary, A. K. Worldwide increased prevalence of human adenovirus type 3 (HAdV-3) respiratory infections is well correlated with heterogeneous hypervariable regions (HVRs) of hexon. *PLoS One*. **13**, e0194516 (2018).
34. Garnett, C., Erdman, D., Xu, W. & Gooding, L. R. Prevalence and quantitation of species C adenovirus DNA in human mucosal lymphocytes. *J. Virol.* **76**, 10608–10616 (2002).

35. Kosulin, K. et al. Persistence and reactivation of human adenoviruses in the gastrointestinal tract. *Clin. Microbiol. Infect.* **22**, 381e381–381e388 (2016).
36. Fang, B. et al. Genetic characterization of human adenoviruses in patients using metagenomic next-generation sequencing in Hubei, China, from 2018 to 2019. *Front. Microbiol.* **14**, 1153728 (2023).
37. Dhingra, A. et al. Molecular evolution of human adenovirus (HAdV) species C. *Sci. Rep.* **9**, 1039 (2019).
38. Cvijović, I., Good, B. H. & Desai, M. M. The effect of strong purifying selection on genetic diversity. *Genetics*. **209**, 1235–1278 (2018).
39. Ullman, A. J., Reich, N. C. & Hearing, P. Adenovirus E4 ORF3 protein inhibits the interferon-mediated antiviral response. *J. Virol.* **81**, 4744–4752 (2007).
40. Houldcroft, C. J. et al. Use of whole-genome sequencing of adenovirus in immunocompromised pediatric patients to identify nosocomial transmission and mixed-genotype infection. *J. Infect. Dis.* **218**, 1261–1271 (2018).
41. Rozas, J. et al. DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol. Biol. Evol.* **34**, 3299–3302 (2017).
42. Rivaller, P., Mao, N., Zhu, Z. & Xu, W. Recombination analysis of human mastadenovirus C whole genomes. *Sci. Rep.* **9**, 2182 (2019).

Acknowledgements

We thank all the participants who contributed in the study. We want to thank the members of the GeneMind team who contributed to the development of the FASTASeq 300.

Author contributions

B.F, J.L and Y.F.L: conceptualization, investigation, writing & editing. L.L.L, X.Y and X.L: sample collection, DNA extract & library construction. H.L.K: sequencing. X.F.Z and X.Z: directing the project, bioinformatics analysis. All authors reviewed the manuscript.

Funding

This study was supported by the National Key Research and Development Program of China 2023YFC3403001 and 2021YFC2600205.

Declarations

Competing interests

The authors declare no competing interests.

Ethics declarations

Approval for this study was obtained from the research ethics board of the Hubei Provincial Center for Disease Control and Prevention. The need for informed consent from participants was waived as the retrospective samples were collected from the influenza surveillance system of Hubei Province.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-78960-9>.

Correspondence and requests for materials should be addressed to X.Z. or X.Z.

Reprints and permissions information is available at www.nature.com/reprints.

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

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024