Multi-species co-circulation of adenoviruses identified by next generation sequencing during an outbreak in coastal Kenya in 2023

Authors
Arnold W. Lambisia¹, Martin Mutunga¹, Esther N. Katama¹, Charles N. Agoti¹,², Charlotte J. Houldcroft ³

Affiliations
1. KEMRI-Wellcome Trust Research Programme, Epidemiology and Demography Department, Kilifi, Kenya
2. Pwani University, School of Health and Human Sciences, Kilifi, Kenya
3. University of Cambridge, Department of Genetics, Downing Street, Cambridge, CB2 3EH, United Kingdom

*Corresponding author:
Contact details: P.O. Box 230, Kilifi-80108, Kenya,
Email: alambisia@kemri-wellcome.org

Key words: Human adenovirus, ONT, enteric, Kenya, RespiCov

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
Abstract (247 words)

Background

Although seven human adenovirus (HAdV) species are known to exist, only F (types 40 and 41) and G, are identified as diarrhoeal disease agents. The role of other HAdV species in diarrhoeal disease remains unclear and data of their prevalence is limited. We describe HAdV species and types in hospitalised children with diarrhoea in coastal Kenya.

Methods

329 stool samples collected between June 2022 and August 2023 from children aged <13-years were screened for HAdV using quantitative polymerase chain reaction (qPCR). Positive HAdV cases were genotyped by adenovirus primers from the RespiCoV panel by amplification, next generation sequencing followed by phylogenetic analysis.

Results

65 samples (20%) tested HAdV positive from which five HAdV species were identified. Other than HAdV F, other species included A, B, C and D; these were detected as either mono-detections or coinfections. Six HAdV F identified by NGS had been missed by our qPCR typing method. This appeared to be as a result of a 133-nucleotide deletion in the long fiber protein which abrogated a primer and probe binding site. Based on VESIKARI scores grading of diarrheal disease severity, 93% of the HAdV cases presented with severe disease. One child with an HAdV F infection died.

Conclusion

Our study shows the enormous diversity and clinical characteristics of HAdV species in children with diarrhoea in coastal Kenya. These data offers an opportunity to improve current diagnostic assays, increase knowledge of HAdV in Africa for control of outbreaks in the future.
Background (2584 words)

Human adenoviruses (HAdV) are non-enveloped, double-stranded DNA viruses that belong to the mastadenovirus genus. To date, seven different species, A to G, and 114 different types of HAdV have been described by the human adenovirus working group (http://hadvwg.gmu.edu/). These HAdVs are associated with a variety of diseases presentation including gastroenteritis (F and G), respiratory tract infections (A, B, C and E), and keratoconjunctivitis (D).

HAdV types F40 and F41 are a common etiology of mild to severe diarrhoea among children under the age of five years. However, other HAdV species such as A, B, C, D and E have been detected among diarrhoea cases although their contribution to diarrhoea disease is unclear. In China, HAdV B3 has been associated with diarrhoea (adjusted odds ratio = 9.205, p < 0.001). In Canada, HAdV F40/41, detections had the highest attributable fraction (96%; 95% confidence interval (C.I), 92.3 to 97.7%) to diarrhoea symptoms compared to species A, B, C and E, but HAdV C1, C2, C5 and C6 were also attributed to 52% (95% C.I, 12 to 73%) of the symptoms. A previous study in Kenya reported predominance of HAdV species D and F in urban and rural settings respectively, among cases with diarrhoea; but other types including B3, B21, C2, C5 and C6 were also detected among diarrheal cases.

The KEMRI-Wellcome Trust Research Programme (KWTRP) has been conducting a prospective hospital-based rotavirus surveillance study at Kilifi County Hospital paediatric ward in coastal Kenya. The prevalence of adenoviruses of any species among paediatric diarrhoea cases in coastal Kenya has been reported to be 15.9% (95% C.I 12.8 to 19.5). However, HAdV F only accounts for approximately half of the HAdV detections (7.3%, 95% C.I 5.2-10.1) with the rest of the HAdVs untyped.

This study aimed to genotype HAdV positive samples detected between June 2022 and August 2023 to determine the circulating non-F HAdVs and any HAdV-Fs that may have been missed by real-time PCR screening. HAdV genotyping is typically done by amplifying a region of the hexon gene, followed by Sanger sequencing. Here we used adenovirus primers from the RespiCoV panel, applying these primers to stool-derived nucleic acid extracts for the first time, and sequencing the amplicons on the Oxford Nanopore Technologies (ONT) platform.
Methods

Study site and population.

The target population was children below the age of 13 years admitted to Kilifi County Hospital (KCH) who presented with diarrhoea as one of their illness symptoms i.e three or more loose stools in a 24-hour period 16.

Laboratory Methods

Total Nucleic Acid (TNA) Extraction and Screening

TNA was extracted from 0.2 g (or 200 µl if liquid) of stool samples using QIAamp® Fast DNA Stool Mini kit (Qiagen, UK) as previously described. Pan-HAdV (forward primer: 5’-GCCCACTGGTCTTTACATGCACATC-3'; probe: ‘FAM-TCGGAGTACCTGAGCCGGTGCA-MGBNFQ’; and Reverse primer: 5’-GCCACGTTGGGTTTCTAAACTT-3’) and HAdV-F (forward primer: 5’-CAGCTTTTGGACTAGT-MGBNFQ’; and Reverse primer: 5’-ACTGGATAGCTAGCGGC-3’) primers and probes, and TaqMan Fast Virus 1-Step Master Mix were used for screening as previously described 17. The thermocycling conditions were 95°C for 20 seconds and 35 cycles of 94°C for 15 seconds and 60°C for 30 seconds.

DNA amplification

The primers used in amplification were adopted from the RespiCov panel 15. Briefly, 14 adenovirus primers were pooled into one tube and resuspended in nuclease free water to generate a 10µM working concentration. TNA from HAdV positive samples were amplified using the Q5® Hot Start HighFidelity 2X Master Mix (NEB) kit. The master mix was prepared as follows: Q5® Hot Start High-Fidelity 2X Master Mix (6.25 µl), H2O (3 µl), HAdV Primer pool (2 µl), and DNA (1.25 µl). The reaction was then incubated on a thermocycler using the following conditions: 98°C for 30 seconds followed by 35 cycles of 98°C for 15 seconds, 65°C for 30 seconds and 72°C for 20 seconds and a final extension of 72°C for 5 minutes.

Library Preparation and Oxford Nanopore Technologies (ONT) Sequencing

Library preparation was performed using the SQK-LSK114 ligation kit with SQK-NBD114.96 barcoding kit. Briefly, the amplicons were end-repaired, barcoded, and pooled into one tube,
and adapters ligated to the library and the final library sequenced using the FLOW-MIN106D R9.4.1 flow cell on the GridION platform (ONT) for one hour.

**Long Fiber amplification and Illumina sequencing**

Primers that could amplify the long fiber protein were obtained by picking forward (HAdV-F41_1kb_jh_85_LEFT:ACACTACAMTCCCCTTGACATCC) and reverse primers (HAdV-F41_1kb_jh_87_RIGHT:AAGAAAATGAGCAGCAGCGGATG) from the whole genome sequencing HAdV primers designed elsewhere ([Quick F41 WGS primers](#)). The mastermix reaction was prepared as described in the above section and incubated on a thermocycler using the following conditions: 98°C for 30 seconds followed by 35 cycles of 98°C for 15 seconds and 65°C for 5 minutes. The amplicons generated were used for library preparation using an Illumina library preparation kit as recommended by the manufacturer. Briefly, the amplicons were tagmented, indexed, and amplified. The libraries were then normalized, pooled, and sequenced as paired-end reads (2*150 bp).

**Data analysis**

**Genotyping**

The FASTQ reads from the GridION were trimmed using porechop v.0.2.4 and mapped to the HAdV reference genomes (DQ923122.2, NC_001460.1, NC_001454.1, NC_001405.1, AC_000006.1, AC_000018.1, AC_000008.1, NC_012959.1) using minimap2 v.2.24-r1122 ([https://github.com/lh3/minimap2](https://github.com/lh3/minimap2)). Variant calling and consensus sequence generation was done using ivar v.1.3.1 ([https://github.com/andersen-lab/ivar](https://github.com/andersen-lab/ivar)) with a minimum read depth of 20. Taxonomic classification was done using BLASTN ([https://blast.ncbi.nlm.nih.gov/](https://blast.ncbi.nlm.nih.gov/)).

The generated consensus genomes were aligned with contemporaneous global HAdV sequences on GenBank using MAFFT ([https://mafft.cbrc.jp/alignment/software/](https://mafft.cbrc.jp/alignment/software/)) and maximum likelihood trees generated using iqtree 2 ([http://www.iqtree.org/](http://www.iqtree.org/)). The trees were then annotated and visualized using ggtree ([https://guangchuangyu.github.io/software/ggtree](https://guangchuangyu.github.io/software/ggtree)).
Long fiber primer check validation

Short-read paired fastq data obtained from the Illumina MiSeq platform was trimmed using fastp with a phred score of q30. The cleaned reads were then mapped to the HAdV-F reference genome using bwa (https://github.com/lh3/bwa). The primers were then trimmed from the BAM files and consensus genomes generated using ivar.

For quality check, the cleaned reads were also assembled using a denovo approach using MetaSPAdes v3.13 (http://cab.spbu.ru/software/spades). The generated contigs were compared with consensus genomes from the reference guide approach.

Real-time PCR primer and probe sequences were then aligned to the generated consensus long fiber sequences to check for differences in binding sites using Geneious Prime® 2023.2.1 (https://www.geneious.com).

Disease severity

The Vesikari Clinical Severity Scoring System Manual was used to estimate disease severity as previously described 18. The following parameters were used: maximum number of stools and vomiting per day, duration of diarrhoea and vomiting episodes in days, temperature, dehydration status and treatment. The Vesikari grading categories were mild, moderate, and severe for scores of <7, 7-10 and ≥ 11 respectively.
Results

HAdV Epidemiology

Between 3rd June 2022 and 28th August 2023, 329 children with diarrhoea as one of their illness symptoms were consented and gave stool samples for enteric viruses screening. A total of 65 (20%) cases had a HAdV detection in their sample when using a pan-adenovirus real-time PCR assay. Forty-three samples were successfully sequenced and genotyped using adenovirus RespiCoV primers to determine the circulating HAdV species and types in stool.

Single or multiple HAdV types were detected in the samples (Figure 1). Single HAdV type detections were as follows: F (n=13), C (n=12), D (n=3), B (n=2) and A (n=2). Codetection of HAdV samples was also observed in 10 samples: C and F (n=4), A and C (n=3), A and D (n=1), F and D (n=1), and C, F and D (n=1) (Figure 1B).
**Figure 1:** Temporal distribution of human adenoviruses in coastal Kenya between June 2022 and August 2023. A) Temporal plot showing distribution of F, mixed F (F and non-F coinfections) and non-F. The primary y axis shows the proportions of adenovirus species while the secondary y axis shows the total number of diarrhoea monthly cases. B) Temporal plot showing the distribution of HAdV species between June 2022 and August 2023.

The HAdV species were further characterized into types (Figure 2). Both HAdV F40 (n=7) and F41 (n=13) were detected in HAdV species F-positive cases. Within HAdV species A, A31 (n=4), A18 (n=2) and A12 (n=1) were identified. Only HAdV B7 was identified in species B. HAdV type C1 (n=5), C2 (n=5), C5 (n=5) and C89 (n=5) were identified within species HAdV C. Within HAdV D, D54 was identified in two samples while the rest of the sequences could not be genotyped further phylogenetically (Figure 2).
Adenovirus outbreak in May 2023

In May 2023, there was a noticeable increase in adenovirus detections among children presenting with diarrhoea over the study period. From the 13 HAdV cases in May 2023, 69% of the cases were genotyped as non-HAdV F (D (n=3), B (n=2), C (n=2), A and C (n=1) and A (n=1)) and only 4 samples were genotyped as HAdV F. Notably, 38% of the 13 HAdV cases had coinfections with norovirus GII (n=4) and astrovirus (n=1).

Demographic and clinical characteristics of HAdV cases

The majority of the HAdV cases were male (60.5%) and between the age of 12 to 59 months (65.1%). Coinfections with rotavirus A, norovirus GII, sapovirus and astrovirus were identified across the different HAdV species. Majority of the cases (93%) presented with severe disease, including all the non-F cases (Table 1). Only one fatality was identified in the HAdV F cases and none in the non-F cases (Table 1).
Table 1: Demographic and clinical characteristics of HAdV genotyped cases in coastal Kenya

<table>
<thead>
<tr>
<th></th>
<th>HAdV F (n=13)</th>
<th>HAdV Codetection F and non-F (n=7)</th>
<th>HAdV non-F (n=23)</th>
<th>Total (n=43)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>7 (53.8%)</td>
<td>2 (28.6%)</td>
<td>6 (26.1%)</td>
<td>15 (34.9%)</td>
</tr>
<tr>
<td>Male</td>
<td>6 (46.2%)</td>
<td>4 (57.1%)</td>
<td>16 (69.6%)</td>
<td>26 (60.5%)</td>
</tr>
<tr>
<td>No data</td>
<td>0 (0.0%)</td>
<td>1 (14.3%)</td>
<td>1 (4.3%)</td>
<td>2 (4.7%)</td>
</tr>
<tr>
<td><strong>Age (Months)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median (interquartile range)</td>
<td>13 (7–21)</td>
<td>11.5 (11–28.5)</td>
<td>15.5 (10.2–20)</td>
<td>14 (10–21)</td>
</tr>
<tr>
<td><strong>Age group (months)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;12</td>
<td>1 (7.7%)</td>
<td>2 (28.6%)</td>
<td>3 (13.0%)</td>
<td>6 (14.0%)</td>
</tr>
<tr>
<td>12 - 23</td>
<td>5 (38.5%)</td>
<td>1 (14.3%)</td>
<td>9 (39.1%)</td>
<td>15 (34.9%)</td>
</tr>
<tr>
<td>24 - 59</td>
<td>5 (38.5%)</td>
<td>2 (28.6%)</td>
<td>6 (26.1%)</td>
<td>13 (30.2%)</td>
</tr>
<tr>
<td>≥60</td>
<td>2 (15.4%)</td>
<td>2 (28.6%)</td>
<td>5 (21.7%)</td>
<td>9 (20.9%)</td>
</tr>
<tr>
<td><strong>Coinfections</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rotavirus A</td>
<td>1 (7.7%)</td>
<td>2 (28.6%)</td>
<td>3 (13.0%)</td>
<td>6 (14.0%)</td>
</tr>
<tr>
<td>Norovirus GII</td>
<td>0 (0.0%)</td>
<td>1 (14.3%)</td>
<td>4 (17.4%)</td>
<td>5 (11.6%)</td>
</tr>
<tr>
<td>Sapovirus</td>
<td>1 (7.7%)</td>
<td>0 (0.0%)</td>
<td>2 (8.7%)</td>
<td>3 (7.0%)</td>
</tr>
<tr>
<td>Astrovirus</td>
<td>0 (0.0%)</td>
<td>1 (14.3%)</td>
<td>0 (0.0%)</td>
<td>1 (2.3%)</td>
</tr>
<tr>
<td><strong>Disease severity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moderate</td>
<td>2 (15.4%)</td>
<td>1 (14.3%)</td>
<td>0 (0.0%)</td>
<td>3 (7.0%)</td>
</tr>
<tr>
<td>Severe</td>
<td>11 (84.6%)</td>
<td>6 (85.7%)</td>
<td>23 (100.0%)</td>
<td>40 (93.0%)</td>
</tr>
<tr>
<td><strong>Outcome</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alive</td>
<td>12 (92.3%)</td>
<td>6 (85.7%)</td>
<td>22 (95.7%)</td>
<td>40 (93.0%)</td>
</tr>
<tr>
<td>Dead</td>
<td>1 (7.7%)</td>
<td>0 (0.0%)</td>
<td>0 (0.0%)</td>
<td>1 (2.3%)</td>
</tr>
<tr>
<td>No data</td>
<td>0 (0.0%)</td>
<td>1 (14.3%)</td>
<td>1 (4.3%)</td>
<td>2 (4.7%)</td>
</tr>
</tbody>
</table>

Failure of HAdV-F real-time PCR Assay

The genotyping results detected six additional HAdV-F positives that were missed by real-time PCR. Successful sequencing of the long fiber protein where the real-time PCR primers bind showed that the sequences had a 133 base deletion that spanned the forward primer and probe binding regions leading to failure in the HAdV-F real-time PCR assay (Figure 3).
Figure 3: An alignment of HAdV F sequences from missed and detected HAdV F samples mapped to the primers and probe sequences. Dots show consensus and Ns show gaps in the primer and probe binding sites.

Ct Value distribution

HAdV F cases had significantly lower Ct values compared to non-F cases (p = 0.002) (Figure 4). There was no significant difference in Ct value among HAdV F cases and cases that had codetections of HAdV-F and non-F. Eight non-F cases had a Ct value of less than 25 (high viral load) but six of these samples also had coinfections with norovirus GII, rotavirus A and sapovirus.

Figure 4: Comparison of HAdV viral load (inverse of cycle threshold value) among F, mixed F (coinfection of F and non-F) and non-F cases.
Discussion

The study findings show that multiple HAdV species and types were in circulation between June 2022 and August 2023 in coastal Kenya and that our current real-time PCR primers for detecting HAdV F may be missing positive cases due to a 133-nucleotide deletion in the long fiber protein which abolishes a primer and probe binding site in some circulating variants. This is the first application of the adenovirus part of the RespiCov method to stool samples for genotyping enteric viral infections. Previous studies from China, Tunisia, Kenya and Brazil have reported detection of non-F HAdVs similar to our findings 3,7,9,10,12,19,20. Non-F types such as B3, C1, C2, C5 and C6 have been associated with increased risk of diarrhoea but direct causation is not yet clear 3,10. In this study, we did not detect HAdV B3 but detected the HAdV types C1 and C5; although we cannot conclude that they were the main cause of diarrhoea in these individuals. The RespiCov method used could not clearly genotype all HAdV species D phylogenetically. This is likely due to the relatively short ~300 base pair hexon region that is amplified and the highly recombinatorial nature of species D HAdVs within the hexon locus 21.

Within HAdV F, there is a high divergence in the strains that have been reported to be circulating in Kenya and other regions across the globe 17,22–24. If this divergence occurs in primer and probe binding sites, nucleic acid amplification methods used for detection may be affected. The genotyping results in this study showed six HAdV F cases were missed by the F-specific real-time PCR assay. Amplification of the long fiber protein revealed a 133 base deletion that impacted the HAdV-F real-time PCR assay, and it is advisable that similar primers used in previous works 13,17 may be missing some positives and need redesigning, a phenomenon previously seen in SARS-CoV-2 25, HIV 26 and Chlamydia trachomatis 27.

HAdV F samples had a significantly higher viral load (lower Ct value) compared to non-F samples, similar to previous studies 10,19. HAdV species F is highly associated with diarrhoea when the Ct value is below 22.7 and cases and controls can be discriminated at Ct 30.5 suggesting that the HAdV F detected in this study may be clearly associated with diarrhoea 9,28. Interestingly, the HAdV-F detections with Ct value above 30 had a rotavirus A coinfection suggesting that they could be the secondary cause of diarrhoea. The non-F HAdVs detected had a lower viral load and were detected with other enteric viruses including rotavirus A, norovirus GII and sapovirus suggesting that they may be associated with carriage due to
prolonged shedding or gut contamination from respiratory infections rather than diarrhoea

This study had some limitations. First, there was no clinical data on the respiratory symptoms of the patients to help interpret the non-enteric HAdVs detected in stool samples which are usually associated with respiratory illnesses. Secondly, only four common enteric viruses were screened for, and these were frequently detected as coinfections in non-F cases. Screening for additional causes such as bacteria, parasites and helminths may have helped elucidate whether other enteric pathogens contribute to the detection of non-F cases.

In conclusion, there is a high diversity of HAdV types found among diarrhoea cases in coastal Kenya and this may reflect the situation in Africa where there is limited data. Timely and accurate genotyping of these HAdV cases is key for troubleshooting failure of molecular detection assays, estimation of diarrhoea prevalence associated with HAdV-F and implementation of interventions to reduce the burden of HAdV associated diarrhoea.
Declarations

Ethics approval and consent to participate.

The research protocol for the study was approved at Kenya Medical Research Institute (KEMRI), by the Scientific and Ethics Review Unit (SSC#2861) in Nairobi, Kenya.

Acknowledgement

We are grateful to the study participants who provided samples and members of the pathogen epidemiology and omics group at KEMRI-Wellcome Trust Programme who did sample collection and laboratory processing. This manuscript was written with the permission of Director KEMRI CGMRC.

Funding.

This study was funded in part by the Cambridge-Africa ALBORADA Research Fund to Drs Agoti and Houldcroft. This research was funded in part by the Wellcome Trust [226002/Z/22/Z]. For the purpose of Open Access, the author has applied a CC-BY public copyright license to any author accepted manuscript version arising from this submission.

Competing interests

The authors declare no conflict of interest.
Data availability

The datasets used and/or analyzed during the current study are available from the KWTRP Research repository via https://doi.org/10.7910/DVN/XCHBND. The HAdV sequences were deposited on GenBank and can be accessed using the accession numbers PP318651-PP318703.

Authors' contributions

CAN and CJH sourced the study funding. CNA, CJH and AWL designed the study laboratory assay. AWL and MM did the laboratory experiments. EK and AWL managed the study data and did the data analysis. AWL, CAN and CJH wrote the first manuscript draft. All authors read, revised, and approved the final manuscript.
References


19. do Nascimento LG, Fialho AM, de Andrade J da SR, de Assis RMS, Fumian TM. Human


22. Götting J, Cordes AK, Steinbrück L, Heim A. Molecular Phylogeny of human adenovirus type 41 lineages. bioRxiv [Internet]. 2022;2022.05.30.493978. Available from: https://www.biorxiv.org/content/10.1101/2022.05.30.493978v1%0Ahhttps://www.biorxiv.org/content/10.1101/2022.05.30.493978v1.abstract


