High genetic stability of co-circulating human adenovirus type 31 lineages over 59 years

Jasper Göttig, Claas Baier, Victoria Panagiota, Britta Maecker-Kolhoff, Akshay Dhingra, and Albert Heim

Abstract

Type 31 of human adenovirus species A (HAdV-A31) is a significant pathogen primarily associated with diarrhoea in children but also with life-threatening disseminated disease in allogeneic haematopoietic stem cell transplant (HSCT) recipients. Nosocomial outbreaks of HAdV-A31 have been frequently described. However, the evolution of HAdV-A31 has not been studied in detail. The evolution of other HAdV types is driven either by intertypic recombination, where different types exchange genome regions, or by immune escape selection of neutralisation determinants. Complete genomic HAdV-A31 sequences from sixty diagnostic specimens of the past 18 years (2003–21) were generated, including fourteen specimens of a presumed outbreak on two HSCT wards. Additionally, twenty-three complete genomes from GenBank were added to our phylogenetic analysis as well as in silico generated and previously published restriction fragment polymorphism (RFLP) data. Phylogenetic analysis of eighty-three genomes indicated that HAdV-A31 evolved slowly with six lineages co-circulating. The two major lineages were lineage 1, which included the prototype from 1962 and nine recent isolates, and lineage 2, which split into four sublineages and included most isolates from 2003 to 2021. The average nucleotide identity within lineages was high (99.8 per cent) and identity between lineages was 98.7 and 99.2 per cent. RFLP data allowed the construction of a lower-resolution phylogeny with two additional putative lineages. Surprisingly, regions of higher diversity separating lineages were found in gene regions coding for non-structural and minor capsid proteins. Intertypic recombinations were not observed, but the phylogeny of lineage 3 was compatible with an interlineage recombination event in the fibre gene. Applying the phylogenetic analysis to the presumed nosocomial outbreak excluded two suspected transmission events and separated it into two different, simultaneous outbreaks caused by different sublineages of lineage 2. However, due to the high nucleotide identity within HAdV-A31 lineages, the proof of infection chains remains debatable. This in-depth study on the molecular phylogeny of HAdV-A31 highlights the high genetic stability of co-circulating HAdV-A31 lineages over almost six decades. It also supports the epidemiological hypothesis that HAdV-A31 circulates as an etiological agent of a childhood disease infecting immunologically naive patients without strong positive selection of immune escape variants and recombinants.

Key words: Adenovirus; molecular epidemiology; molecular evolution; complete genomic sequencing; nosocomial; genomics; DNA virus

Introduction

Viruses from the family of Adenoviridae are non-enveloped, icosahedral, and common throughout the vertebrates (Russell and Benkó 1999). Adenoviridae are separated into the five genera Atadenovirus, Aviadenovirus, Diadenovirus, Ichtadenovirus, and Mastadenovirus (Benkó, Harrach, and Russell 2006; Benkó et al. 2002; Davison, Benkó, and Harrach 2003; Kovács et al. 2003). The human adenoviruses (HAdV), first isolated in 1953 from human adenoidal tissue (Rowe et al. 1953; Hillemann and Werner 1954), belong to the Mastadenovirus genus. Their linear dsDNA genome is ~35 kb in length and encodes about 30–40 proteins (Davison, Benkó, and Harrach 2003). HAdVs are further separated into seven species (A–G) subdivided into 104 types, with only four of these as members of species HAdV-A. Types were initially distinguished by serotyping the neutralisation loops of the hexon capsid protein (HAdV types 1–51) and later by sequencing all three major capsid proteins penton, hexon, and fibre (types 52–105) (Aoki et al. 2011, Seto et al. 2011).

Recombination between types of the same HAdV species is an important mechanism for the evolution of human adenoviruses resulting in multiple recombinants of species HAdV-B, -C, and -D (Robinson et al. 2013). Zoonotic adenovirus types can also...
recombine with HAdV (Jacobs et al. 2004). In contrast, only a single recombinant type (HAdV-A61) has been described in species HAdV-A (Matsushima et al. 2011). In addition to recombination, new types may evolve via diversifying selection (immune escape) of the neutralisation determinant (Robinson et al. 2013).

HAdV-A31 was isolated from the stool of Bristol children and serumotyped in 1962. HAdV-A31 is associated with gastroenteritis in children, sometimes in small outbreaks. HAdV-A31 circulates worldwide and has a high seroprevalence already in children (about 15 per cent), but the latest systematic study on this topic dates from 1982 (D’Ambrosio et al. 1982; Mennechet et al. 2019). HAdV-A31 frequently causes nosocomial outbreaks in paediatric oncology wards (Leruez-Ville et al. 2006; Swartling et al. 2015; Houldcroft et al. 2018; Myers et al. 2021). Moreover, it was recognised as a significant transplant pathogen in the early 2000s, causing disseminated disease in haematopoietic stem cell transplanted (HSCT) children in sporadic and outbreak-associated cases (Munoz, Piedra, and Demmler 1998; Walls, Shankar, and Shingadia 2003). HAdV-A31 was found in 22 per cent of sporadic HAdV infections of paediatric HSCT recipients (Mynarek et al. 2014), the second most frequently detected HAdV type after HAdV-C2 in this patient group at high risk of life-threatening adenovirus infections.

For tracing of infection chains, restriction fragment length polymorphism (RFLP) analysis was used initially, and several subtypes (genome types) of HAdV-A31 were identified (Th. et al. 1986; Adrian, Wigand, and Richter 1987; Adrian and Wigand 1989). Moreover, this technique was applied to HAdV-A31 isolates from immunocompromised patients in search of highly virulent genome types (Johansson et al. 1991). A complete genomic sequence (CGS) of the HAdV-A31 prototype and partial sequences of seven clinical isolates from immunocompromised patients were published in 2009 (Hofmayer et al. 2009). In addition to the RGD (Arginine, Glycine, Aspartic Acid) motif in the penton base, a second RGD motif was found in pIX, unique compared to all other HAdV types. Partial sequencing of genomes (hexon and E3 region) was also applied to a prolonged outbreak of adenovirus A31 in allogeneic stem cell transplant recipients, and results suggested nosocomial transmission (Swartling et al. 2015). Subsequently, a series of nosocomial outbreaks of HAdV-A31 at a single medical centre was analysed by complete genomic sequencing. Surprisingly, identical sequences were also found in patients from different outbreaks. However, results were not entirely proving transmission as the genetic diversity of the endemically circulating HAdV-A31 remained obscure (Houldcroft et al. 2018; Myers et al. 2021). Prior to 2021, only three HAdV-A31 CGSs were available in public databases, and thus, the sequence diversity of circulating HAdV-A31 strains was unknown.

Therefore, we sequenced the genomes and analysed the molecular phylogeny of 60 HAdV-A31 clinical isolates from 2003 to 2021, including fourteen isolates from two recent concomitant nosocomial outbreaks. Moreover, twenty recently published complete genomic HAdV-A31 sequences from the UK nosocomial outbreak and three CGSs from GenBank were included in the phylogenetic analysis.

Materials and methods

HAdV-A31 isolates and CGSs

HAdV-A31-positive samples (stool or cell culture supernatant from A549 cultures used for virus isolation) originated from the collection of the German national adenovirus reference laboratory and were sequenced as described below. Furthermore, all twenty-three available complete genomic HAdV-A31 sequences from GenBank were included in the phylogenetic analysis. All sequences generated for or used in this study are listed in Supplementary Table 1.

Ethical statement

The project was reviewed by the ethics committee at the university clinic (ethical vote 10,100 BOK_2021, Hannover Medical School). Additional anonymised, archival, HAdV-A31 containing samples were provided by the German National Reference Laboratory for Adenovirus.

High-throughput sequencing and de novo assembly

DNA was extracted from 400μl HAdV-A31 positive stool or cell culture supernatant (depending on the availability and virus load) using a Qiagen Blood Kit on a QIAcute. Library preparation was performed using the NEBNext Ultra II FS DNA Library Prep Kit for Illumina according to the manufacturer’s protocol. Final libraries were inspected on an Agilent Bioanalyzer, normalised, multiplexed, and sequenced on an Illumina MiSeq using a 600v3 Reagent Kit to generate 2 × 300 bp paired-end reads with an average of 1.25 million reads per sample.

De novo assembly was performed as previously described. Briefly, human reads were removed, and reads were trimmed with fastsp and assembled with SPAdes, which usually resulted in a single high-coverage contig constituting the entire HAdV-A31 genome. Finally, genomes were polished using Pilon (Walker et al. 2014), and genome termini were manually examined and corrected using a mapping of the reads against the HAdV-A31 reference genome. CGSs were additionally inferred using Geneious Prime 2020.1.2. Finally, genomes were deposited in GenBank (Accession numbers MZ983552–MZ983610 & OM1027527, see Supplementary Table 1).

Phylogenetic analysis

Multiple sequence alignment of the complete sequences was carried out using MAFFT v7.450 (Katoh and Standley 2013). Phylogenetic trees were constructed with the HAdV-A61 reference genome as the outgroup using RAxML v8 under the GTR GAMMA model with 100 rapid bootstrapping replicates and search for the best-scoring maximum likelihood (ML) tree (Stamatakis 2014). For comparison, the CGS phylogeny was additionally inferred using MrBayes 3.2.6 (GTR & invgamma model with 500,000 Markov Chain Monte Carlo (MCMC) steps and 50,000 burn-in steps) and Geneious Tree Builder (neighbour-joining with default parameters and 1,000 bootstrap replicates) (Ronquist et al. 2012). SimPlot 3.5.1 was used to generate similarity plots (SimPlots) and to perform BootScan recombination analyses using default parameters, a window size of 1,000 bp (BootScan) or 1,500 bp (SimPlot), and a step size of 200 bp (BootScan) or 300 bp (SimPlot) (Lole et al. 1999). TreeTime with default parameters was used to attempt inference of a time-calibrated ML phylogeny (Sagulenko, Puller, and Neher 2018). In silico RFLP analysis was performed in Geneious Prime 2020.1.2 (Biomatters) with the nine restriction enzymes used in the previous RFLP genome typing work (BarnHI, BclI, BglII, BstEII, EcoRI, HindIII, KpnI, Smal, and XhoI) (Johansson et al. 1991). CGSs were digested with each enzyme separately, calculated fragments were rounded to the nearest 100 bp length, and fragments shorter than 400 bp were discarded to match the data with the fragment patterns from Fig. 1 of Johansson et al. (1991). A 112-bit binary
string representing the presence or absence of all 112 occurring fragments from all enzymes was compiled for all CGSs as well as the nine described genome types from Table 1 of Johansson et al. (1991). A phylogenetic tree of those binary strings was created using FreeTree with the Nei & Li UPGMA distance model and 100 bootstrap replicates (Pavlicek, Hrdá, and Flegr 1999). All phylogenetic trees were visualised and annotated in R using ggplot2 and ggtree (Wickham 2016, Yu 2020).

Analysis for positive selection

branch-site unrestricted statistical test for episodic diversification (BUSTED) and mixed effects model of evolution (MEME) as implemented on datamonkey.org were utilised to identify genes and sites under positive or diversifying selection (Weaver et al. 2018). All genes displaying high diversity in the SimPlot were analysed in BUSTED, which tests for gene-wide, non-site-specific selection (Murrell et al. 2015). Genes exhibiting significant evidence for selection in BUSTED were further analysed using MEME, which identifies individual sites evolving under positive or diversifying selection (Murrell et al. 2012).

Results

HAdV-A31 phylogeny

Phylogenetic analysis of eighty-three complete genomic HAdV-A31 sequences exhibited six distinct lineages containing multiple identical or barely divergent (99.75–100 per cent identity) sequences (Fig. 1).

Lineage 1 included the prototype from 1962 and isolates from as late as 2019. Lineage 1 had a 99.79 per cent average nucleotide identity and even 99.84 per cent identity between the 1962 prototype and the last available isolate from 2019. Bootstrapping of the sub-cluster 1b was not unequivocal with a value of 58 per cent in the ML tree, and future data may change the clustering of these two 1b isolates. These clustered with lineage 2 when phylogeny was inferred with the neighbour-joining method but also with lineage 1 using Bayesian methods (Supplementary Figs 1 and 2). Lineage 2, which had four sub-clusters (2a–2d), included the majority of the genomic sequences (62 of 83), which originated from 2007 to 2021 and exhibited a 99.85 per cent average nucleotide identity. The average identity between lineages 1 and 2 was high (99.62 per cent), whereas other lineages were more divergent to lineage 1 (98.72–99.21 per cent). Lineages 3 and 4 were represented only by a single sequence and lineage 6 only by two divergent to lineage 1 (98.72–99.21 per cent). Lineages 3 and 4 were multiple identical or barely divergent (99.75–100 per cent identity) sequences from 2007 to 2021 and exhibited a 99.85 per cent average nucleotide identity. The majority of the genomic sequences (62 of 83), which originated from 2007 to 2021 and exhibited a 99.85 per cent average nucleotide identity. The average identity between lineages 1 and 2 was high (99.62 per cent), whereas other lineages were more divergent to lineage 1 (98.72–99.21 per cent). Lineages 3 and 4 were represented only by a single sequence and lineage 6 only by two identical sequences. Overall, HAdV-A31 isolates covering 59 years were highly conserved, and co-circulation of the closely related lineages 1 and 2 over a period of more than 10 years was evident. Intratypic evolution was so slow that constructing time-calibrated phylogenies based on either the CGSs or just the region of highest diversity (spanning DNA pol, pTP, p52K, and pIIIa) was unsuccessful (Supplementary Figs 3 and 4). Furthermore, we detected no SNPs between sequences obtained from cell culture isolated virus and direct sequencing from stool samples of the same patient.

In silico RLFP analysis

Previous work attempted to resolve the strain diversity of HAdV-A31 isolates from the 1960s to 1980s using RFLP of nine common restriction enzymes. Thirteen genome types (D1–D13) and their corresponding enzyme codes were described (Johansson et al. 1991). We performed in silico RFLP analysis of the present eighty-three CGSs with the same restriction enzyme set and generated the corresponding 112-bit fragment fingerprint to create a combined RFLP phylogeny with the previously published genome types (Fig. 2). As a positive control, the enzyme code of the prototype sequence (genome type D1) was reproduced by our in silico analysis. With a few exceptions, the overall structure of the RFLP-based phylogenetic tree was concordant with the CGS-based phylogeny. However, clustering of sublineage 1b was with lineage 2, and clustering of lineages 2d and 3 was with lineage 1. Genome type D7 (isolated for the first time in 1976 in Minnesota) had an identical RFLP fingerprint to five completely sequenced sublineage 1a isolates from 2005 to 2019. Genome type D8 (isolated in Massachusetts, 1982) also clustered within sublineage 1a.

The most abundant lineage in our analysis, lineage 2, was clustered with genome types D2, D9, D10, and D11. Genome type D2 was most closely related to the isolates of sublineages 2a, 2b, and 2c from 2003 to 2021 and was isolated as early as 1965 in Illinois. Thus, lineage 2 can be traced back almost as far as lineage 1, which includes the prototype from 1962. Genome types D12 and D13 did not cluster with any of the lineages sequenced in the present study and likely represent additional lineages. D13 was sampled in Tokyo in 1985 and is, compared to all other HAdV-A31 genome types, most closely related to HAdV-A61, a hexon-recombinant HAdV-A31 isolated in Kyoto, Japan, in 2004 (Matsushima et al. 2011). Genome types D3–D6, sampled in Germany around 40 years ago, do not appear in the dendrogram, as the full fragment fingerprint could not be inferred from the literature (Adrian and Wigand 1989). Their partially described patterns did not match up with any CGS.

Evolution of genome regions

Hotspots of evolution separating the lineages were found in the polymerase/pTP, p52K/pIIIa, protease/DBP, E3, and the fibre gene region (Fig. 3). Positive selection of immune escape variants should be expected in the fibre knob and hexon region. However, only lineages 3 and 5 had evolved significantly in the fibre compared to the prototype (Fig. 4), and the amino acid substitutions were not accumulated in the ε determinant of the fibre knob. Only lineages 5 and 6 had evolved in the hexon gene compared to the prototype sequence. These two lineages clustered together in the hexon tree (Fig. 4), but only one nonsynonymous mutation (of three) was located in the neutralisation epitope ε of lineages 5 and 6. Sublineage 1b sequences also exhibited a single amino acid substitution in the ε determinant, but all other lineages had an ε determinant amino acid sequence identical to the prototype. Moreover, selection analysis of the hexon and fibre alignment using BUSTED revealed the absence of positive selection (e.g. immune escape). The penton gene was highly conserved (Fig. 4). Lineage 6, phylogenetically most distant from the prototype sequence, had only two amino acid sequence substitutions in the penton, and even HAdV-A61 had only a single amino acid substitution compared to the HAdV-A31 prototype sequence. This amino acid substitution was weakly indicative of gene-wide positive selection in BUSTED, but selection could not be confirmed by the site-specific MEME analysis (Supplementary Table 2).

Potential hotspots of evolution were in gene regions coding for non-structural and minor capsid proteins. Lineages 4–6 had the lowest sequence identity (98.29–98.45 per cent) compared to the prototype sequence in the polymerase/pTP gene region of transcription unit E2, whereas in the E3 region, lineages 3, 5, and 6 were most divergent from the prototype sequence (98.00–98.24 per cent identity). A poly(T) stretch of variable length (longest in lineage 5) contributed significantly to the sequence divergence between the late (L3) and E2B promoter gene region. Although the genes of the L1 proteins p52K and pIIIa of lineages 4–6 had up to fifty-eight SNPs in the corresponding enzyme codes were described (Johansson et al. 1991). We performed in silico RFLP analysis of the present eighty-three CGSs with the same restriction enzyme set and generated the corresponding 112-bit fragment fingerprint to create a combined RFLP phylogeny with the previously published genome types...
compared to the prototype sequence, these resulted only in a few amino acid substitutions (five in lineages 4 and 5 and four in lineage 6). Only the amino acid substitutions in the protease were weakly indicative of gene-wide positive selection in BUSTED, but selection could again not be confirmed by the site-specific MEME analysis (Supplementary Table 2).

Intertypic and interlineage recombination

Intertypic recombination with the classical serotypes 12 and 18 of species HAdV-A was not observed in the phylogeny of HAdV-A31 lineages. However, lineages 5 and 6 were more closely related to the genomic backbone of HAdV-A61 than the other lineages (Fig. 1). HAdV-A61 is a recombinant genotype with a hexon loop.
Figure 2. RFLP fingerprint phylogeny. Phylogenetic tree based on the 112-bit in silico RFLP fingerprints from the eighty-three HAdV-A31 CGSs and the nine genome types described in (Johansson et al., 1991). The HAdV-A31 and -A61 prototype sequences are marked with a rhombus at the tip point, and the genome types are labelled bold. Sequences with identical RFLP fingerprints from the same lineage were collapsed for clarity with the sampling period and number of CGS in parentheses. The lineage colouring from Figure 1 was retained, and bootstrap support values were binned into three categories (≥85 per cent, 50–84 per cent, and <50 per cent), indicated by filled boxes at the node points.

Figure 3. Similarity Plot (SimPlot) of HAdV-A31 lineages. SimPlot of the nucleotide identity between the HAdV-A31 prototype sequence and the consensus sequences of the six identified lineages along the whole genome.
Figure 4. Phylogenetic trees of the penton, hexon, and fibre genes of all eighty-three HAdV-A31 CGSs. The HAdV-A31 and -A61 prototype sequences are marked with a rhombus at the tip point. Bootstrap support values were binned into three categories (≥95 per cent, 75–94 per cent, and <75 per cent) indicated by filled boxes at the node points. The lineage colouring from Figures 1 and 2 was retained, and all three trees are visualised at the same scale. The sequences obtained from the outbreak on the HSCT wards are additionally labelled with the patient identifier in brackets; see Figure 6 for details. HAdV-A61 was omitted from the hexon phylogeny due to the low genetic identity to the HAdV-A31 sequences (∼87 per cent).

1 and 2 (ε determinant) sequence of HAdV-A12 and a genomic backbone of an ancestral HAdV-A31 strain, which was presumably also a precursor of lineages 5 and 6. Phylogeny of lineage 3 is compatible with an ancient interlineage recombination of a precursor of lineages 1 and 2 with lineage 5 in the fibre region (Figs 4 and 5).

No other interlineage recombinations were observed in the phylogeny of circulating HAdV-A31 strains. Thus, recombination was probably not a significant driver of the evolution of HAdV-A31 lineages.

An outbreak of HAdV-A31 on two HSCT wards

Between December 2020 and May 2021, a clustering of HAdV-A31 infections on two HSCT wards was reported: nine cases (A1–A9) on ward A with exclusively adult patients and five cases on ward P with exclusively paediatric patients (Fig. 6). Simultaneously, two additional cases on a haematological ward for adults (ward E) and a general ward for children (ward N) of the same hospital were observed. All four wards were not epidemiologically linked. All patients from ward P were routinely screened for HAdV, whereas routine screening was established on ward A after detection of the second case.

All nine patients on ward A were infected with a single sublineage 2a strain (0 SNPs), whereas among the five ward P cases, a co-circulation of three strains (sublineages 2a and 2c and lineage 6) was observed. Therefore, multiple introductions of HAdV-A31 must have occurred. Three of the five cases (P1, P3, and P4, infected with sublineage 2c; ≤2 SNPs) were epidemiologically and phylogenetically linked, indicating transmission events. Patient P2 had a unique strain that clustered with lineage 6. Interestingly, patient P5 clustered phylogenetically with sublineage 2a of the adult HSCT patients from ward A (A1–A9). Patient P5 was cared for by the paediatric HSCT team, and ward P is located in another building than ward A. This missing epidemiological link clearly supports the assumption that the sublineage 2a strain was introduced (at least) twice into the hospital. A third introduction of the same strain can be suspected in patient E1, who was on the regular (non-HSCT) haematology and oncology ward E, which is also located in a different building than ward A and ward P. A fourth introduction of the same strain cannot be wholly excluded because the last three patients (A7, A8, and A9) infected on ward A did not have an immediate temporal overlap with first six patients (A1–A6), and transmission by unsampled patients can also be excluded due to routine stool screening for HAdV. However, patient A7 occupied the same room as patient A1, and patient A8 followed patient A7 in exactly this room. Before A7 and A8 were admitted to this room, intensified cleaning with virucidal disinfectants was performed. However, an environmental persistence of HAdV-A31
triggering transmission cannot be excluded entirely. In parallel to these cases, HAdV-A31 was also isolated from a paediatric patient (N1) on ward N, epidemiologically unrelated to the other wards and presenting with a unique strain from lineage 5.

**Virus load kinetics and clinical course in outbreak patients**

All fourteen patients suffered from diarrhoea with a very rapid increase of stool virus loads, followed by an increase of blood virus loads with high peaks (>1 × 10^5 c/ml) as observed in patients A1, A7, P1, and P4 (Fig. 7). All other patients had peak virus loads in blood <5 × 10^4 c/ml or were even negative for HAdV-DNAemia (patients A3, A4, A5, and P2). All but two patients (A1 and P1) cleared the infection, usually after discharge from the hospital (Fig. 7). For example, P4 shed HAdV-A31 with faeces for 8 months. Patients A1 and P1 died on Day 49 and Day 87 after detection of the HAdV-A31 infection, respectively. Criteria for a disseminated infection were fulfilled, high HAdV loads in
blood in both patients \((6 \times 10^6\) and \(5 \times 10^5\) c/ml, respectively) and detection of HAdV in other body fluids. Patient A1 was also positive for HAdV-A31 in urine \((4 \times 10^8\) c/ml) and eye swabs. In P1, a bronchoalveolar lavage and cerebrospinal fluid were also positive for HAdV-DNA, and HAdV-DNA was also detected in a colon biopsy with \(1.3 \times 10^3\) c/cell. However, these patients probably did not succumb to the infection as virus loads in peripheral blood had subsequently decreased to or below the limit of quantification \((1,000\) c/ml) (Fig. 7).

### Discussion

In this study, we sequenced 60 HAdV-A31 genomes from samples collected during the past 18 years in Germany to elucidate the type’s evolution and diversity. In addition, twenty-three publicly available HAdV-A31 genomes from other regions (the prototype of 1962 and isolates from 2005 to 2020 from UK, Tunisia, and the USA) were included in the phylogenetic analysis. Overall, six slowly evolving lineages of HAdV-A31 were identified. Historic genome typing data generated by RFLP analysis (Johansson et al. 1991) were considered to search for additional lineages. Genome types D12 (New York, 1981) and D13 (Tokyo, 1985–6) could represent two additional lineages; however, bootstrap values were ambiguous due to the low-resolution phylogenetic data provided by RFLP codes. Moreover, the overall RFLP tree structure was not entirely concordant with the CGS tree. For example, the recombinant lineage 3 clustered with lineage 1. This may also be explained by the low-resolution phylogenetic data generated by RFLP, which are insufficient to resolve a recombination event. Although the overall structure of lineages 1 and 2 seemed to agree with the CGS tree, sublineage 2d clustered with lineage 1 and sublineage 1b vice versa with lineage 2. Therefore, designating genome types D8–D11 as potential sublineages of lineage 2 was dubious.

Nevertheless, the co-circulation of the most abundant lineages 1 and 2, which was demonstrated as early as 2003 by CGS data, could be traced back to at least 1965. This is supported by the overall structure of the RFLP tree, which includes genome type D2 detected as early as 1965 in its lineage 2 and the prototype from 1962 in lineage 1. Interestingly, genome type D2 from 1965 had as many as 68 of 70 co-migrating restriction fragments with two sublineage 2c isolates from 2007 and 2019. Similar modes of evolution—high genetic stability of long-term co-circulating lineages—have also been reported for other types (2, 3, 4, and 55) across multiple HAdV species (B, C, and E) (Dhingra et al. 2019; Gonzalez et al. 2019; Hang et al. 2020; Duan et al. 2021a, 2021b). Separation of lineages was inferred to have occurred over 600 years ago in the case of HAdV-E4, but this was not achieved for all the other types described above and HAdV-A31 (Gonzalez et al. 2019). Furthermore, the slow evolution of HAdV species C was recently demonstrated by the recovery of ancient HAdV-C1 and C2 genomes from 31,600-year-old milk teeth (Nielsen et al. 2021).

Similar to the evolution of HAdV-C1, -C2, -D8, and -B55, positive selection of potential immune escape mutants in the \(\varepsilon\) determinant has not been observed in the HAdV-A31 lineage evolution (Hage et al. 2017b; Dhingra et al. 2019; Hang et al. 2020). This is in contrast to multiple amino acid substitutions in the \(\varepsilon\) determinant described in the evolution of HAdV-B3, E4, and B7 strains (Gonzalez et al. 2019; Duan et al. 2021a, 2021b). These types tend to circulate with different, and often stable, genome types at different geographically separated army installations infecting at least partially immune recruits, which may exert higher selective pressure for immune escape variants (Blasiole et al. 2004; Kajon et al. 2007; Metzgar et al. 2007; Coleman et al. 2021). In contrast, HAdV-A31 predominately circulates in immune naive children almost like a childhood disease and occasionally in nosocomial outbreaks in highly immunosuppressed adult patients (Schmitz, Wigand, and Heinrich 1983; Adrian and Wigand 1989; Swartling et al. 2015; Shieh 2021). Therefore, the endemic circulation of HAdV-A31 probably does not require immune escape. A previously described positive selection of E3 CR1 beta mutants could not be confirmed in a more extensive set of sequences in the present
study (Hofmayer et al. 2009), although the nucleic acid sequence diversity between lineages was slightly higher than average in the E3 region (Fig. 3). Significant sequence variability in the E3 region, which codes for gene products interfering with the host’s immune response, was also observed between the two lineages of HAdV-E4 and HAdV-D types (Ginsberg et al. 1989; Lichtenstein et al. 2004; Windheim, Hilgendorf, and Burgert 2004; Ismail et al. 2018; Gonzalez et al. 2019). Higher sequence diversity between lineages of HAdV-A31 was also observed in the polymerase/pTP region and L1 proteins p52K and pI1α, possibly due to relaxed purifying selection pressure on those regions (Hughes and Hughes 2007; Wertheim and Kosakovský Pond 2011). Only lineage 5 (and the recombinant lineage 3) of HAdV-A31 had evolved in the fibre (but surprisingly not in its immunogenic γ determinant).

Although lineage 3 was an intratypic recombinant, no intraspecies recombinations (with other types of species HAdV-A) were found in any HAdV-A31 lineage. So far, the only intraspecies A recombination was observed in the phylogeny of HAdV-A61, which has a genomic backbone most closely related to HAdV-A12 with 84% nucleic sequence identity (Matsushima et al. 2011). Recombination is a substantial driver of species HAdV-D-type evolution but has also been observed in the evolution of HAdV-B and HAdV-C types (Walsh et al. 2009; Robinson et al. 2011, 2013; Dhingra et al. 2019; Ismail et al. 2019; Rivaille et al. 2019). Recombination requires the co-circulation of multiple types of the same HAdV species resulting in co-infection of a host cell with at least two types, or superinfection of a latently infected cell with another type, as described for HAdV species C and D (Kho et al. 1995; Ivanova et al. 2012; Kosulin et al. 2016; Hage et al. 2017a; Kosulin 2019). This prerequisite seems hardly fulfilled in the evolution of HAdV-A types as our collection of diagnostic specimens of the last 20 years almost exclusively comprises HAdV-A31 isolates (eighty-four HAdV-A31, three HAdV-A18, and one HAdV-A61). However, in the evolution of HAdV-A31 lineages, a recombination of the most frequently circulating lineages 1 and 2 would be most probable but was not observed in our data set, whereas the phylogeny of lineage 3 indicated recombination of lineage 5 with an ancestor of lineages 1 and 2.

HAdV-A31 is infamous for causing prolonged and hard to control outbreaks on HSCT wards, which may include environmental intermediates or transmission by unsampled patients (Leruez-Ville et al. 2006; Swartling et al. 2015; Houldcroft et al. 2018; Myers et al. 2021). However, multiple introductions to the hospital should also be considered because a presumed outbreak on two different wards of the same hospital was, surprisingly, caused by three distinct HAdV-A31 strains. Detailed analysis in this study revealed at least four independent introductions of these three strains to two HSCT wards. In parallel, two other HAdV-A31 introductions were observed on other wards of the same hospital during the same period.

In a recent work on the healthcare-associated transmission of HAdV-A31, ≤3 SNPs were observed in epidemiologically linked monophyletic clusters and ≤2 SNPs within a single host, which is in accordance with our findings (Myers et al. 2021). However, in our study, sequence identical isolates (≤2 SNPs) were observed in lineage 2a even over a period of 7 years and a distance of 550 km, of course without any epidemiological link (see MZ983553—Hanover 2008 and MZ983609—Ulm 2015 in Fig. 1). Therefore, detection of ≤3 SNPs or even sequence identity does not ultimately confirm a suspected but questionable healthcare-associated HAdV-A31 transmission event. This was recently suspected, when almost identical HAdV-A31 strains (≤3 SNPs) were identified on a single HSCT ward separated by an interval of 1 and 4 years (Myers et al. 2021). We also observed an interval, albeit much shorter (1–2 days), between sequence identical cases A1, A7, and A8 occupying the same room. Although the room was disinfected thoroughly, environmental persistence and transmission of HAdV cannot be excluded but neither can multiple introductions to the ward due to the low sequence diversity of HAdV-A31 strains circulating in the community. The latter was demonstrated by our study that compared fourteen CGSs of a presumed outbreak with forty-six CGSs from Germany originating from an 18-year period and twenty-three complete genomic HAdV-A31 sequences from the USA, UK, and Tunisia. This substantial sequence collection highlighted the genetic stability of HAdV-A31 over many years. Previously, the significance of sequence identity for proving questionable infection chains may have been overestimated.

From a technical point of view, different sequencing approaches were used in the present study and previous studies, e.g. direct sequencing from stool samples with or without bait enrichment or sequencing of cell culture isolates (Depledge et al. 2011; Houldcroft et al. 2018; Myers et al. 2021). These methods yielded almost identical CGSs with ≤2 SNPs (see, e.g. MZ983563—Hanover 2012 and MW686780—UK 2011 in Fig. 1), which were all included in the phylogenetic analysis of the present study.

In conclusion, this study demonstrated the high genetic stability of HAdV-A31 and the co-circulation of its stable lineages over at least 59 years. Therefore, complete genomic sequencing is the ultimate tool to exclude presumed transmission events, e.g. in a healthcare setting. However, using sequence identity to prove transmission chains can be misleading without sufficient sampling of circulating HAdV-A31 strains in the community.

Data availability

HAdV-A31 complete genomic sequences generated for this study are available at GenBank accession numbers MZ983552—MZ983610 & OM372572 (see Supplementary Table 1). Accession numbers for pre-existing GenBank sequences used in this study are also listed in Supplementary Table 1.

Supplementary data

Supplementary data are available at Virus Evolution online.

Acknowledgements

We thank Hans Nitschko, Helga Mairhofer, Anna-Maria Eis-Hubinger, Ulrike Reber, and Reimar Johne for providing us with additional HAdV-A31 material and isolates. We also thank Jenny Witthuhn, Sandra Flucht, and Lars Steinbrück for technical assistance.

Funding

This study was Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—Projektnummer 158989968 - SFB 900 (project Z1) and under Germany’s Excellence Strategy EXC 2155 ‘RESIST’—Project ID 390874280. J.G. was supported by the graduate programme ‘Infection Biology’ of the Hanover Biomedical Research School (HBRS) and the Center for Infection Biology (ZIB).
Conflict of interest: None declared.

Author contributions
J.G. conducted the sequencing, analysed and interpreted the data, and drafted the manuscript. C.B. provided and analysed contact tracing data and drafted the manuscript. B.M.-K. and V.P. provided clinical data and revised the manuscript. A.D. contributed to sequencing and genome assembly. A.H. designed and supervised the study, interpreted data, and reviewed the manuscript. All authors contributed to the article and approved the submitted version.

References
Adrian, T., and Wigand, R. (1989) ‘Genome Type Analysis of Adenovirus 31, a Potential Causative Agent of Infants’ Enteritis’, Archives of Virology, 105: 81–7.
Adrian, T., Wigand, R., and Richter, J. (1987) ‘Gastroenteritis in Infants, Associated with a Genome Type of Adenovirus 31 and with Combined Rotavirus and Adenovirus 31 Infection’, European Journal of Pediatrics, 146: 38–40.
Aoki, K. et al. (2011) ‘Towards an Integrated Human Adenovirus Designation System that Utilizes Molecular and Serological Data and Serves Both Clinical and Fundamental Virology’, Journal of Virology, 85: 5703–4.
Benkó, M. et al. (2002) ‘First Molecular Evidence for the Existence of Distinct Fish and Snake Adenoviruses’, Journal of Virology, 76: 10056–9.
Benkó, M., Harrach, B., and Russell, W. C. (2000) ‘Family Adenoviridae’. In: van Regenmortel, M. H. V. et al. (eds) Virus Taxonomy: Seventh Report of the International Committee on Taxonomy of Viruses, pp. 227–38. San Diego: Academic Press.
Blasiole, D. A. et al. (2004) ‘Molecular Analysis of Adenovirus Isolates from Vaccinated and Unvaccinated Young Adults’, Journal of Clinical Microbiology, 42: 1686–93.
Coleman, K. K. et al. (2021) ‘Six Decades of Human Adenovirus Type 4 Infections Reviewed: Increasing Infections among Civilians Are a Matter of Concern’, Clinical Infectious Diseases, 73: 740–6.
D’ambrosio, E. et al. (1982) ‘Neutralizing Antibodies against 33 Human Adenoviruses in Normal Children in Rome’, Journal of Hygiene, 89: 155–61.
Davison, A. J., Benkó, M., and Harrach, B. (2003) ‘Genetic Content and Evolution of Adenoviruses’, Journal of General Virology, 84: 2895–908.
Deepledge, D. P. et al. (2011) ‘Specific Capture and Whole-Genome Sequencing of Viruses from Clinical Samples’, PLoS One, 6(11): e27805.
Dhingra, A. et al. (2019) ‘Molecular Evolution of Human Adenovirus (HAdV) Species C’, Scientific Reports, 9: 1039.
Duan, Y. et al. (2021a) ‘Genetic Analysis of Human Adenovirus Type 7 Strains Circulating in Different Parts of China’, Virologica Sinica, 36: 382–92.
Duan, Y. et al. (2021b) ‘Molecular Characteristics of Human Adenovirus Type 3 Circulating in Parts of China during 2014–2018’, Frontiers in Microbiology, 12: 688661.
Ginsberg, H. S. et al. (1989) ‘Role of Early Region 3 (E3) in Pathogenesis of Adenovirus Disease’, Proceedings of the National Academy of Sciences, 86: 3823–7.
Gonzalez, G. et al. (2019) ‘Genomic Characterization of Human Adenovirus Type 4 Strains Isolated Worldwide since 1953 Identifies Two Separable Phylogroups Evolving at Different Rates from Their Most Recent Common Ancestor’, Virology, 538: 11–23.
Hage, E. et al. (2017a) ‘Three Novel, Multiple Recombinant Types of Species of Human Mastadenovirus D (HAdV-D 73, 74 & 75) Isolated from Diarrhoeal Faeces of Immunocompromised Patients’, Journal of General Virology, 98: 3037–45.
Hage, E. et al. (2017b) ‘Molecular Phylogeny of a Novel Human Adenovirus Type 8 Strain Causing a Prolonged, Multi-state Keratoconjunctivitis Epidemic in Germany’, Scientific Reports, 7: 4–12.
Hang, J. et al. (2020) ‘Human Adenovirus Type 55 Distribution, Regional Persistence, and Genetic Variability’, Emerging Infectious Diseases, 26: 1497–505.
Hillermeier, M. R., and Werner, J. H. (1954) ‘Recovery of New Agent from Patients with Acute Respiratory Illness’, Experimental Biology and Medicine, 85: 183–8.
Hofmayer, S. et al. (2009) ‘Unique Sequence Features of the Human Adenovirus 31 Complete Genomic Sequence Are Conserved in Clinical Isolates’, BMC Genomics, 10: 557.
Houldcroft, C. J. et al. (2018) ‘Use of Whole-Genome Sequencing of Adenovirus in Immunocompromised Pediatric Patients to Identify Nosocomial Transmission and Mixed-Genotype Infection’, The Journal of Infectious Diseases, 218: 1261–71.
Hughes, A. L., and Hughes, M. A. K. (2007) ‘More Effective Purifying Selection on RNA Viruses than in DNA Viruses’, Gene, 404: 117–25.
Ismail, A. M. et al. (2018) ‘Adenoviromics: Mining the Human Adenovirus Species D Genome’, Frontiers in Microbiology, 9: 2178.
Ismail, A. M. et al. (2019) ‘Genomic Foundations of Evolution and Ocular Pathogenesis in Human Adenovirus Species D’, FEBS Letters, 593: 3583–608.
Ivanova, O. E. et al. (2012) ‘Adenovirus Isolation Rates in Acute Flaccid Paralysis Patients’, Journal of Medical Virology, 84: 75–80.
Jacobs, S. C. et al. (2004) ‘Characterization and Manipulation of the Human Adenovirus 4 Genome’, Journal of General Virology, 85: 3361–6.
Johansson, M. E. et al. (1991) ‘Genome Analysis of Adenovirus Type 31 Strains from Immunocompromised and Immunocompetent Patients’, Journal of Infectious Diseases, 163: 293–9.
Kajon, A. E. et al. (2007) ‘Molecular Epidemiology of Adenovirus Type 4 Infections in US Military Recruits in the Postvaccination Era (1997–2003)’, The Journal of Infectious Diseases, 196: 67–75.
Kato, K., and Standley, D. M. (2013) ‘MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability’, Molecular Biology and Evolution, 30: 772–80.
Khoo, S. H. et al. (1995) ‘Adenovirus Infections in Human Immunodeficiency Virus-positive Patients: Clinical Features and Molecular Epidemiology’, Journal of Infectious Diseases, 172: 629–37.
Kosulin, K. (2019) ‘Intestinal HAdV Infection: Tissue Specificity, Persistence, and Implications for Antiviral Therapy’, Viruses, 11: E804.
Kosulin, K. et al. (2016) ‘Persistence and Reactivation of Human Adenoviruses in the Gastrointestinal Tract’, Clinical Microbiology and Infection, 22: 381.e1–381.e8.
Kovács, G. M. et al. (2003) ‘Phylogenetic Analysis of the Hexon and Protease Genes of a Fish Adenovirus Isolated from White Sturgeon (Acipenser transmontanus) Supports the Proposal for a New Adenovirus Genus’, Virus Research, 98: 27–34.
Leruez-Ville, M. et al. (2006) ‘Description of an Adenovirus A31 Outbreak in a Paediatric Haematology Unit’, Bone Marrow Transplantation, 38: 23–8.
Lichtenstein, D. L. et al. (2004) ‘Functions and Mechanisms of Action of the Adenovirus E3 Proteins’, International Reviews of Immunology, 23: 75–111.
Lole, K. S. et al. (1999) ‘Full-length Human Immunodeficiency Virus Type 1 Genomes from Subtype C-infected Serocconverters in India,'
with Evidence of Intersubtype Recombination’, Journal of Virology, 73: 152–60.
Matsushima, Y. et al. (2011) ‘Genomic Characterization of a Novel Human Adenovirus Type 31 Recombinant in the Hexon Gene’, Journal of General Virology, 92: 2770–5.
Mennechet, F. J. D. et al. (2019) ‘A Review of 65 Years of Human Adenovirus Seroprevalence’, Expert Review of Vaccines, 18: 597–613.
Metzgar, D. et al. (2007) ‘Abrupt Emergence of Diverse Species B Adenoviruses at US Military Recruit Training Centers’, The Journal of Infectious Diseases, 196: 1465–73.
Munoz, F. M., Piedra, P. A., and Demmler, G. J. (1998) ‘Disseminated Adenovirus Disease in Immunocompromised and Immunocompetent Children’, Clinical Infectious Diseases, 27: 1194–200.
Murrell, B. et al. (2012) ‘Detecting Individual Sites Subject to Episodic Diversifying Selection’, Molecular Biology and Evolution, 32: 1365–71.
Murrell, B. et al. (2015) ‘Gene-Wide Identification of Episodic Selection’, Molecular Biology and Evolution, 32: 1365–71.
Nielsen, S. H. et al. (2021) ‘31,600-Year-old Human Virus Genomes Support a Pleistocene Origin for Common Childhood Infections’, bioRxiv, 2021: 06.28.450199.
Pavlícek, A., Hrdá, S., and Flegr, J. (1999) ‘Free-Tree—Freeware Program for Construction of Phylogenetic Trees on the Basis of Distance Data and Bootstrap/Jackknife Analysis of the Tree Robustness. Application in the RAPD Analysis of Genus Frenkelia’, Folia Biologica, 45: 97–9.
Rivailler, P. et al. (2019) ‘Recombination Analysis of Human Mastadenovirus C Whole Genomes’, Scientific Reports, 9: 2182.
Robinson, C. M. et al. (2011) ‘Molecular Evolution of Human Species D Adenoviruses’, Infection, Genetics and Evolution, 11: 1208–17.
Robinson, C. M. et al. (2013) ‘Molecular Evolution of Human Adenoviruses’, Scientific Reports, 3: 1812.
Ronquist, F. et al. (2012) ‘MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice across a Large Model Space’, Systematic Biology, 61: 539–42.
Rowe, W. P. et al. (1953) ‘Isolation of a Cytopathogenic Agent from Human Adenoids Undergoing Spontaneous Degeneration in Tissue Culture’, Experimental Biology and Medicine, 84: 570–3.
Russell, W. C., and Benkö, M. (1999) ‘Adenoviruses (Adenoviridae): Animal Viruses’. In: Granoff, A., and Webster, R. G. (eds) Encyclopedia of Virology, 2nd edn. pp. 14–21. Oxford: Elsevier.
Sagulenko, P., Fuller, V., and Neher, R. A. (2018) ‘TreeTime: Maximum-Likelihood Phylodynamic Analysis’, Virus Evolution, 4: vex042.
Schmitz, H., Wigand, R., and Heinrich, W. (1983) ‘Worldwide Epidemiology of Human Adenovirus Infections’, American Journal of Epidemiology, 117: 455–66.
Seto, D. et al. (2011) ‘Using the Whole-Genome Sequence to Characterize and Name Human Adenoviruses’, Journal of Virology, 85: 5701–2.
Shieh, W.-J. (2021) ‘Human Adenovirus Infections in Pediatric Population – An Update on Clinico-Pathologic Correlation’, Biomedical Journal, 45: 38–49.
Stamatakis, A. (2014) ‘RAxML Version 8: A Tool for Phylogenetic Analysis and Post-analysis of Large Phylogenies’, Bioinformatics, 30: 1312–3.
Swartling, L. et al. (2015) ‘Prolonged Outbreak of Adenovirus A31 in Allogeneic Stem Cell Transplant Recipients’, Transplant Infectious Disease, 17: 785–94.
Th., A. et al. (1986) ‘DNA Restriction Analysis of Adenovirus Prototypes 1 to 41’, Archives of Virology, 91: 277–90.
Walker, B. J. et al. (2014) ‘Pilon: An Integrated Tool for Comprehensive Mutational Variant Detection and Genome Assembly Improvement’, PLoS One, 9: e112963.
Walls, T., Shankar, A., and Shingadia, D. (2003) ‘Adenovirus: An Increasingly Important Pathogen in Paediatric Bone Marrow Transplant Patients’, The Lancet Infectious Diseases, 3: 79–86.
Walsh, M. P. et al. (2009) ‘Evidence of Molecular Evolution Driven by Recombination Events Influencing Tropism in a Novel Human Adenovirus that Causes Epidemic Keratoconjunctivitis’, PLoS One, 4: e6535.
Weaver, S. et al. (2018) ‘Datamonkey 2.0: A Modern Web Application for Characterizing Selective and Other Evolutionary Processes’, Molecular Biology and Evolution, 35: 773–7.
Wertheim, J. O., and Kosakovsky Pond, S. L. (2011) ‘Purifying Selection Can Obscure the Ancient Age of Viral Lineages’, Molecular Biology and Evolution, 28: 3355–65.
Wickham, H. (2016) ggplot2: Elegant Graphics for Data Analysis. New York: Springer-Verlag.
Windheim, M., Hilgendorf, A., and Burgert, H. G. (2004) ‘Immune Evasion by Adenovirus E3 Proteins: Exploitation of Intracellular Trafficking Pathways’, Current Topics in Microbiology and Immunology, 273: 29–85.
Yu, G. (2020) ‘Using ggtree to Visualize Data on Tree-Like Structures’, Current Protocols in Bioinformatics, 69: e96.