JOURNAL OF GENERAL VIROLOGY

RESEARCH ARTICLE

Espínola *et al.*, *Journal of General Virology* 2017;98:2810–2820
DOI 10.1099/jgv.0.000947

MICROBIOLOGY SOCIETY

Computational analysis of a species D human adenovirus provides evidence of a novel virus

Emilio E. Espínola,[1,*] Julio C. Barrios,[1] Graciela Russomando,[1] Santiago Mirazo[2] and Juan Arbiza[2]

Abstract

A human adenovirus (HAdV) species D, was isolated from a hospitalised child with severe lower respiratory infection. It was initially detected in the nasopharyngeal aspirate of the child followed by conventional PCR amplification of the hexon, penton base, and fibre genes. Sanger DNA sequencing and phylogenetic analyses showed characteristics of a recombinant genome not described before. Next Generation Sequencing analysis was performed to reconstruct its complete DNA genome after viral isolation in adenocarcinoma human cell line (A549). A complete genomic sequence of 35.2 kb in length, with a G +C content of 57 % was obtained, related to HAdV-D29 (96 % identity). Imputed serology analysis demonstrated its novel type with a nucleotide sequence identity of 95.3 % (hexon loop 1) and 96 % (hexon loop 2) to HAdV-D9. The penton base gene showed a novel sequence, distantly related to HAdV-D44. The E3 and E4 regions evolved significantly from their ancestors. The fibre gene was almost identical to the knob region of HAdV-D15 but showed an unrelated shaft sequence. In conclusion the genomics of this novel HAdV, designated the HAdV-D83 [P83H9F15] prototype and bearing a new penton base gene, supports the importance of viral evolution to understand modified tissue tropism, enhanced transmission, or altered virulence.

INTRODUCTION

HAdVs cause a wide range of clinical manifestations in humans (from mild to life-threatening diseases), compromising organ systems such as respiratory, gastrointestinal, ocular, and renal, among others, with billions of infected people worldwide [1]. They belong to the *Adenoviridae* family, and are divided into five genera (*Mastadenovirus*, *Aviadenovirus*, *Atadenovirus*, *Siadenovirus*, and *Ichtadenovirus*) [2]. Human adenoviruses belong to the *Mastadenovirus* genera, and are grouped into seven species (HAdV-A to HAdV-G), with over 80 types, according to phylogenomics, serum neutralisation, haemagglutination inhibition and biochemical criteria [3, 4]. Adenoviruses are icosahedral, non-enveloped, double-stranded linear DNA viruses, with a genome of approximately 35 000 bp. Virions consist of a protein shell (capsid) surrounding the genome. The icosahedral shell is composed primarily of 240 capsomeres of hexon trimers (12 per triangular facet of the icosahedron), 12 pentameric penton base capsomeres at each vertex of the icosahedron, and 12 fibres extending from the penton bases, each a trimer of the fibre polypeptide. Loops with hypervariable sequence on the external surface of hexons are important for type-specific immunogenicity and neutralisation [5, 6].

Identification of novel HAdV prototypes was based almost exclusively on neutralisation assays (the first 51 'serotypes'), but nowadays analysis of whole-genome sequence data has evolved as an alternative standard in the identification and characterisation of novel multirecombinant HAdV types (starting with type 52) [7].

HAdV-D is the largest of the seven species, currently with 46 types. The major sites of infections for this species are the eye and gastrointestinal tract, but not the respiratory tract [8]. HAdV-D8, a main etiologic agent of a severe eye infection called epidemic keratoconjunctivitis, was the first to be classified in this species [9]. Most HAdV-Ds were identified during the first two decades of the AIDS epidemic [10], and frequent homologous recombination between types of the same species has been described. Both, recombination between species D types and selection of novel neutralisation epitopes ('immune escape'), contribute

Received 5 August 2017; Accepted 29 September 2017
Author affiliations: [1]Departamento de Biología Molecular y Biotecnología, Instituto de Investigaciones en Ciencias de la Salud, Universidad Nacional de Asunción, Paraguay; [2]Sección Virología, Facultad de Ciencias, Universidad de la República, Montevideo, Uruguay.
***Correspondence:** Emilio E. Espínola, emilioespinola@hotmail.com
Keywords: adenovirus; recombination; DNA sequencing; bioinformatics.
Abbreviations: CPE, cytopathic effect; D-MEM, Dulbecco's modified eagle medium; dN/dS, ratio of non-synonymous to synonymous mutations; HAdV, human adenovirus; HVL1, hypervariable loop 1 (hexon gene); HVL2, hypervariable loop 2 (hexon gene); HVR1, hypervariable region 1 (penton gene); HVR2, hypervariable region 2 (penton gene); NJ, neighbour-joining method; PIV1, parainfluenza virus type 1.
GenBank accession number: KX827426.

to the generation of genome diversity [11, 12]. Except for a few types that cause epidemic keratoconjunctivitis, most of the types of HAdV-D are not associated with severe diseases.

This communication describes a new type of HAdV species D, isolated from nasopharyngeal aspirate of a hospitalised child with severe acute respiratory infection. Whole-genome sequencing and bioinformatics analysis showed multiple recombination events in its phylogeny and the Human Adenovirus Working Group (http://hadvwg.gmu.edu/) has designated it as HAdV-D83.

RESULTS

Molecular typing by conventional PCR

The amplification and sequencing of hexon, penton base, and fibre genes (directly from the clinical sample) were the initial data about the recombinant nature of the HAdV-D detected. A phylogenetic analysis of the penton base gene showed its relatedness to HAdV-D44; the hexon gene (3′ conserved terminal region) and the fibre gene showed its relatedness to HAdV-D15 (data not shown).

Deep next-generation sequencing of whole genome

To reveal the genetic characteristics of the HAdV-D detected, we carried out deep next-generation sequencing of its whole genome. A single contig representing its genome was constructed using both *de novo* and reference assembly. The whole genome was 35 207 bp in length with 57 % G+C content, and a base composition of 28.67 % G, 28.57 % C, 22.49 % A, and 20.27 % T. The genome was predicted to encode 30 ORFs with an organisation similar to other members of the species HAdV-D. Similarity analyses using the HAdV-D83 sequence as a query in LAGAN software with the most closely related prototype sequences of HAdV-D suggested multiple recombination events in its phylogeny (Fig. 1). The complete genomic sequence of HAdV-D83 was deposited in GenBank database (accession number KX827426).

Molecular phylogeny of penton base gene

The penton base gene clustered as a separate branch, and was related to HAdV-D44, Fig. 2(a), with 97.3 % identity (86 % bootstrap value). HVR1 and HVR2 both clustered as a separate branch (not shown). HVR1 showed 92.6 % identity to HAdV-D45, whereas HVR2 showed 95.3 % identity to both HAdV-D32 and HAdV-D44.

Bootscan analysis of the penton base gene, Fig. 2(b), showed distant relationships with other members. No evidence of recombination in flanking nucleotide sequences for the penton base gene was found.

Molecular phylogeny of hexon gene

The hexon gene was most closely related to HAdV-D9, Fig. 3(a), with an overall 96.8 % identity (72 % bootstrap value). Bootscan analysis showed highly similar HVL1 and HVL2 to HAdV-D9, Fig. 3(b).

Molecular phylogeny of fibre gene

The fibre gene was related only to HAdV-D15, Fig. 4(a), with 91.7 % identity (100 % bootstrap value). Bootscan analysis showed similar knob sequence to HAdV-D15, but an unrelated shaft sequence, Fig. 4(b).

Whole-genome of HAdV-D83

Thus, following the convention criteria to classify novel types of adenoviruses (i.e. ‘P’ refers to the penton base, ‘H’ to the hexon HVL1 region, and ‘F’ to the fibre knob) [7], HAdV-D83 should be designated [P83H9F15].

Phylogenetic analysis of the whole-genome sequence of HAdV-D83 with other prototypes of HAdV-D indicated a common ancestor with HAdV-D25, HAdV-D29, HAdV-D58, and HAdV-D63, with the highest sequence identity to HAdV-D29 (96 % identity), and 60 % bootstrap value (Fig. 5). Similar results were obtained using both Simplot (Fig. 6) and bootscan (Fig. 7) analyses.

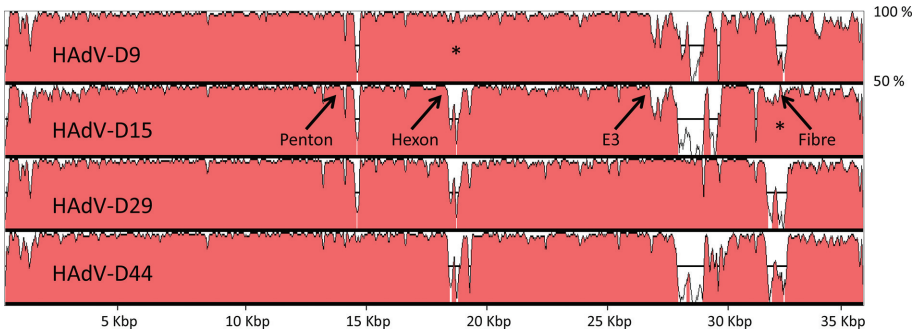


Fig. 1. mVISTA global pairwise sequence alignment of HAdV-D83 compared to representative types from D species (HAdV-D9, HAdV-D15, HAdV-D29, and HAdV-D44). The penton base, hexon, E3, and fibre genes are divergent within species D (indicated by arrows), except for the hexon and fibre regions of HAdV-D9 and HAdV-D15 (*).

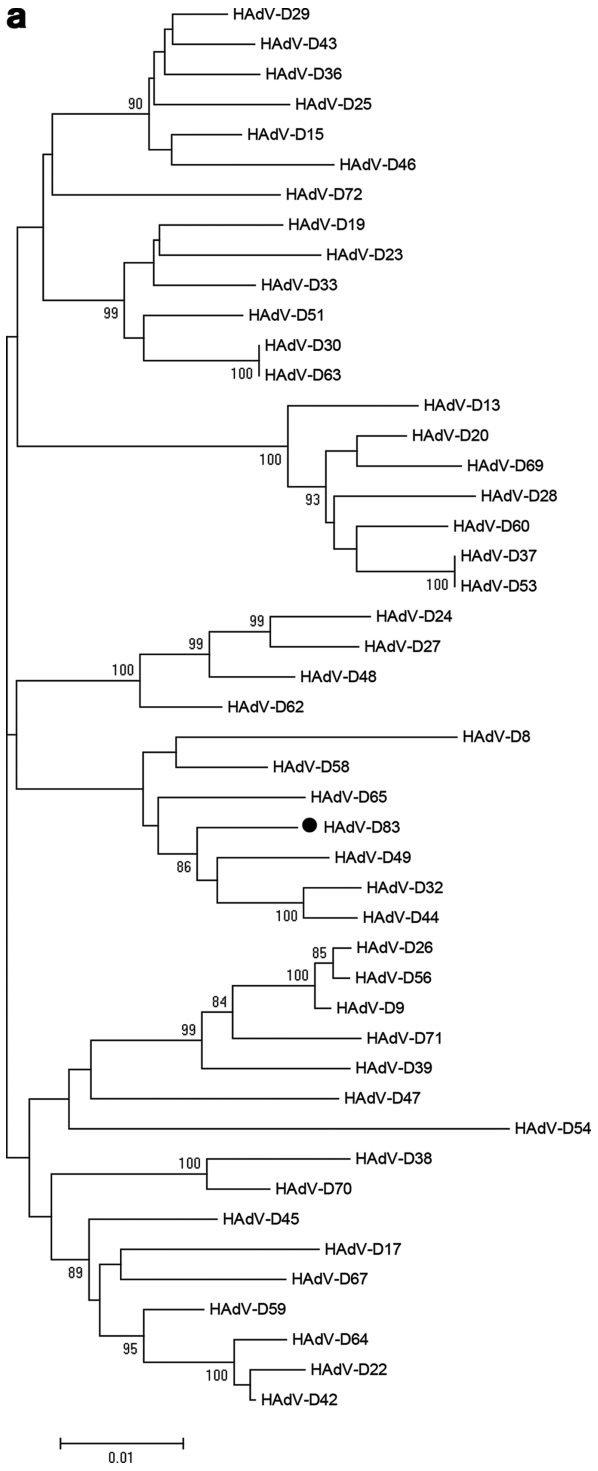
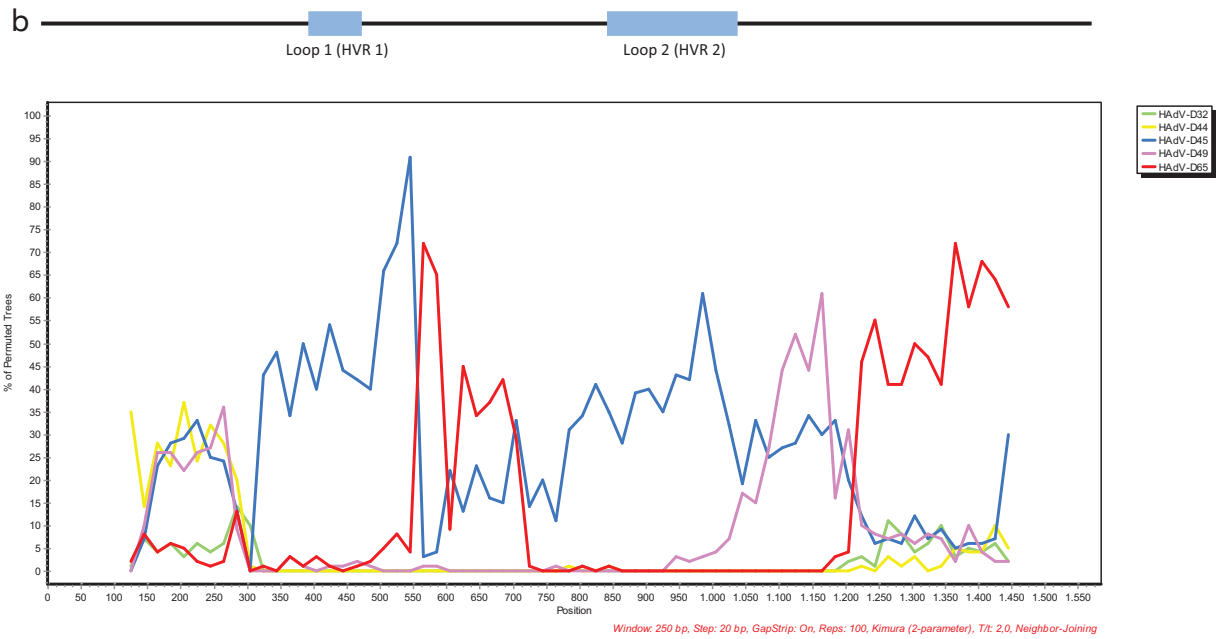


Fig. 2. (a) NJ tree of penton base gene, using 46 HAdV-D prototypes. HAdV-D83 is marked with a filled black circle. Bootstrap values greater than 80 % are shown at branch nodes. Branch distance is indicated by a scale bar at the bottom of the tree. (b) Bootscan analysis of HAdV-D83 penton base gene. A window size of 250 bp and a step size of 20 bp were implemented for analysis using the SimPlot software.

Fig. 2. (cont.)



DISCUSSION

Whole-genome sequencing and bioinformatics analysis showed multiple recombination events in the phylogeny of this new HAdV type, and the Human Adenovirus Working Group (http://hadvwg.gmu.edu/) has designated it as HAdV-D83. The novel HAdV-D83 [P83H9F15] prototype showed key features of a recombinant genome. Its complete genome was determined to be 35 207 bp, coding for 30 putative ORFs. Comparison of HAdV-D83 with other HAdV-D genomes identifies sequence divergence from these in most of its genome regions. The hexon and fibre genes were related to other HAdV types, as a result of a (probably recent) homologous recombination with HAdV-D9 (previously detected in gastrointestinal tract) and HAdV-D15 (previously detected in upper respiratory track and ocular surface), respectively. On the contrary, the penton base gene, E3 and E4 regions showed high diversity (under 97.3 % of identity). In fact, the penton base gene was characterised as a novel HAdV type. For molecular typing by imputed serology analysis, the hexon HVL1 and HVL2 known as the major neutralisation epitope ε and a critical region for the development of adenovirus vaccines [13], is more significant than the clustering of the complete hexon gene [14]. Minimum nucleic acid sequence identity thresholds to the most closely related prototype have been defined for molecular typing with the help of HVL1 (97.6 %) and HVL2 (97.5 %) sequences [14]. Both HVL1 and HVL2 nucleic acid sequence identity data (95.3 and 96 % to the most closely related HAdV-D9) suggested that HAdV-D83 was a new type.

DNA replication and recombination are intimately associated, as suggested by experiments in which recombination was delayed by blocking DNA replication [15]. Besides homologous recombination (proposed as the main mechanism for the evolution of HAdV types), illegitimate recombination has also been proposed as a mechanism, particularly in the hexon hypervariable regions, probably by slippage-misalignment of the HAdV DNA polymerase in repetitive polypurine stretches during single-strand DNA replication [16].

In spite of multiple recombination events in the phylogeny of HAdV-D83, most parts of its genome were novel HAdV sequences with unique phylogenetic relationships. As the hexon contains the major neutralisation epitope ε, positive selection for immune escape can result in rapid evolution of the hexon gene [16]. In fact, HAd-D83 recombined its major neutralisation epitope (HVL1 and HVL2) with HAdV-D9. In general, the ratio dN/dS=0.224 did not indicate positive selection acting on its genome.

HAdV-Ds species are associated to infections of the eye and gastrointestinal tract, but not the respiratory tract [8]. Exceptional cases of respiratory infections, however, have occurred worldwide [17, 18], including fatal outcomes [11, 19]. In some cases, novel recombinant HAdVs can cause unexpected outbreaks. For example, HAdV-D22 has been identified as the source of the novel HAdV-D53, which

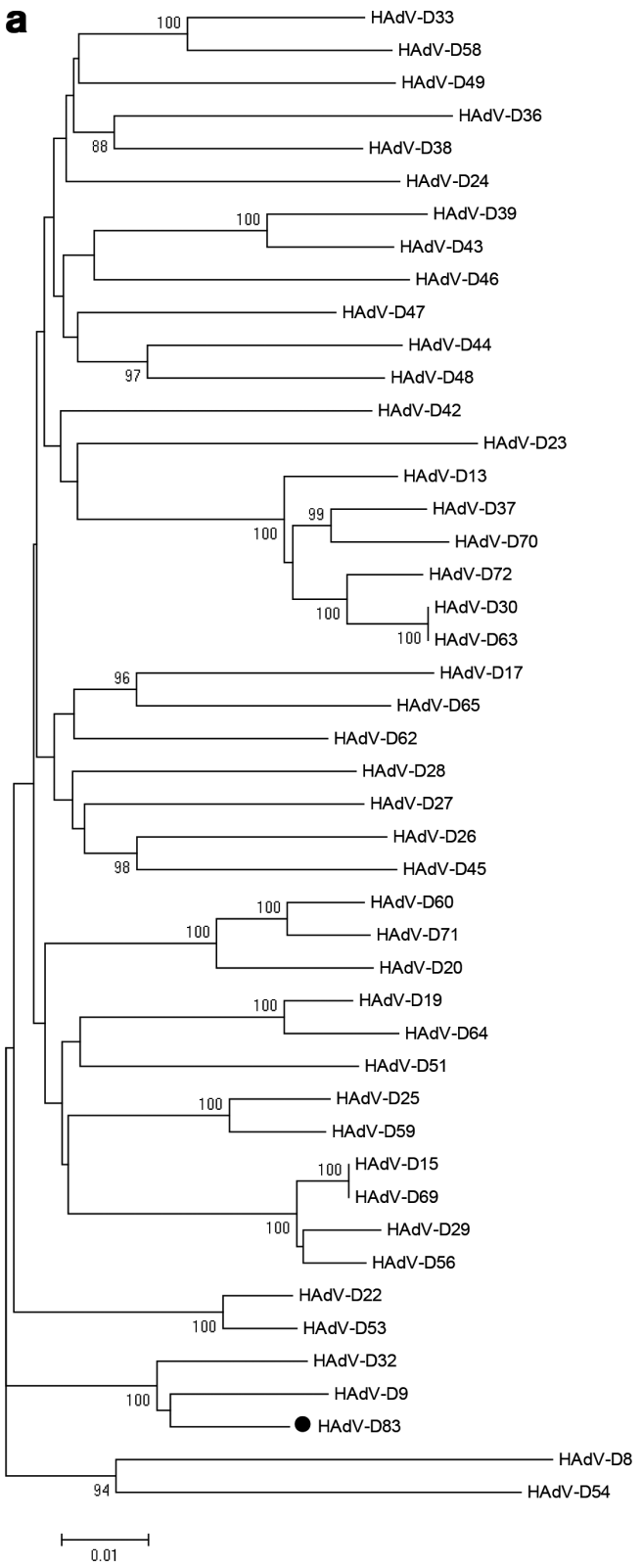
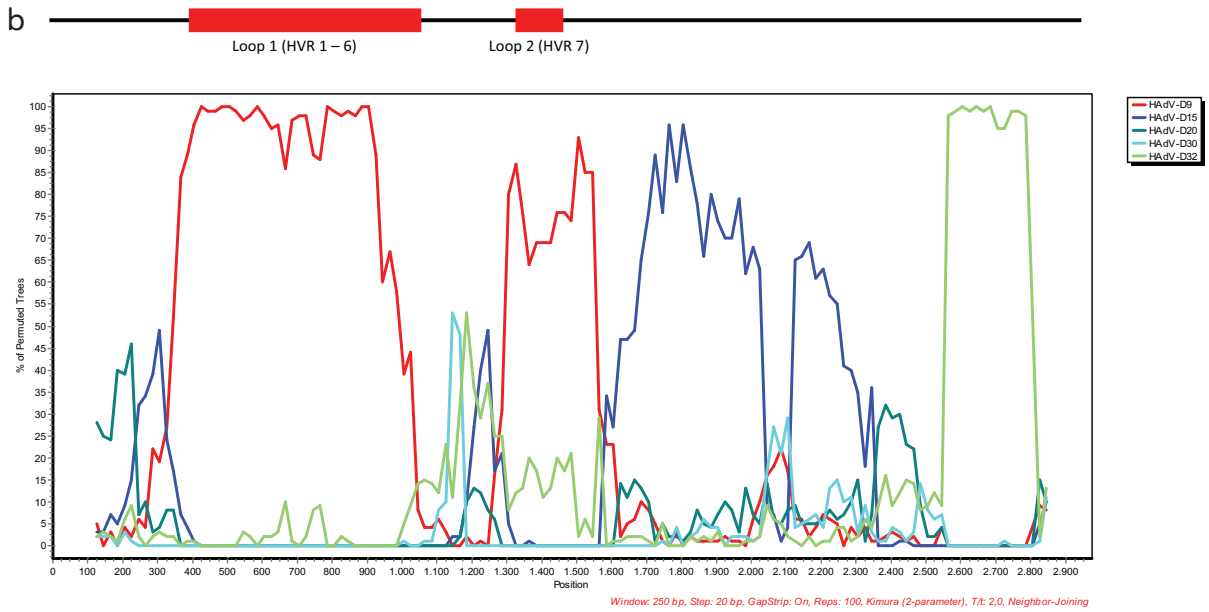


Fig. 3. (a) NJ tree of hexon gene, using 46 HAdV-D prototypes. HAdV-D83 is marked with a filled black circle. Bootstrap values greater than 80 % are shown at branch nodes. Branch distance is indicated by a scale bar at the bottom of the tree. (b) Bootscan analysis of HAdV-D83 hexon gene. A window size of 250 bp and a step size of 20 bp were implemented for analysis using the SimPlot software.

Fig. 3. (cont.)



caused a severe ocular disease outbreak of keratoconjunctivitis in Germany [20]. In our study, we did not find evidence of an outbreak of HAdV-D83, in four years of study in the same hospital.

Since HAdV-D83 was isolated from a child co-infected with PIV1, we cannot conclude that HAdV-D83 infection was the cause of the disease. In human lung cell culture, however, HAdV-D83 produced characteristics CPEs after two viral passages (10 days). The CPEs included cell rounding, detaching from the culture surface and cell lysis.

In conclusion, the new type 83 is a HAdV-D with multiple, consecutive recombination events in its phylogeny. This supports the hypothesis of persistent infections, facilitating co-infections and thus promoting homologous recombination as a key factor for the evolution of HAdV-D types.

METHODS

Clinical case and isolation of HAdV-D83

A 9 months old male child was hospitalised with symptoms of severe acute respiratory infection in July 2010, at the Hospital Pediátrico Niños de Acosta Ñu (San Lorenzo, Paraguay). The patient presented fever (≥38.5 °C), cough, and dyspnea, without cardiac or pulmonary disease; it was positive for rotavirus (using latex agglutination test). A multiplex real-time PCR (FTD Respiratory pathogens 21 plus, Luxembourg) performed in the respiratory sample (nasopharyngeal aspirate) showed a co-infection with PIV1 and HAdV. Amplification of hexon, penton base, and fibre genes were carried out directly from the clinical sample by conventional PCR [21]. After 9 days of hospitalisation the patient left the hospital healthy; bronchiolitis and anaemia were the final diagnosis.

Isolation of HAdV-D83 was performed using 50 µl of the nasopharyngeal aspirate sample, inoculated in human lung carcinoma cells A549 (ATCC CCL-185), maintained in D-MEM (Life Technologies), and collected at 90 % CPE formation. Viral DNA was extracted from cell culture supernatant using an AxyPrep Body Fluid Viral DNA/RNA Miniprep Kit (Axygen Biosciences), following manufacturer's instructions.

Next-generation sequencing and bioinformatics

The HAdV-D83 library was prepared using a GS-FLX Titanium rapid library preparation kit (Roche). Briefly, 500 ng of input DNA was simultaneously fragmented and ligated to adaptors. Adaptor-ligated fragments were amplified for 50 cycles using an emulsion-based clonal PCR kit (Roche), and size-selected using AMPure XP beads (Beckman Coulter). The final library was sequenced on a GS-FLX Titanium platform (454 Life Sciences, Roche). Inverted terminal repeats were sequenced using an ABI 3730XL (Applied Biosystems) through PCR amplification. Quality control included base qualities, GC content, absence of sequence contamination, and sequence comparison and annotation with HAdV-D genome landmarks. The raw sequence dataset was assembled by *de novo* and reference methods using NextGENe software (version 2.4.1; SoftGenetics, LLC.), via

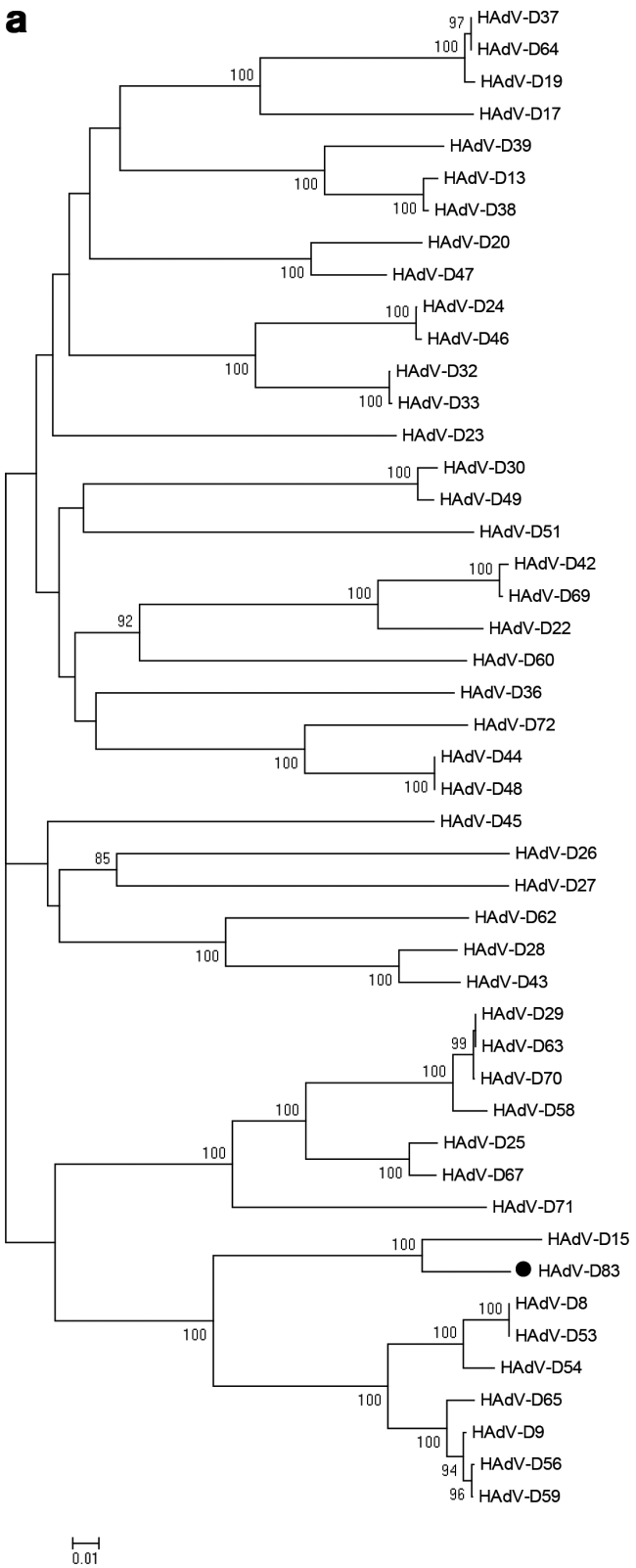


Fig. 4. (a) NJ tree of fibre gene, using 46 HAdV-D prototypes. HAdV-D83 is marked with a filled black circle. Bootstrap values greater than 80 % are shown at branch nodes. Branch distance is indicated by a scale bar at the bottom of the tree. (b) Bootscan analysis of HAdV-D83 fibre gene. A window size of 250 bp and a step size of 20 bp were implemented for analysis using the SimPlot software.

Fig. 4. (cont.)

floton assembly method and default parameters. Minimal sequence coverage was 17- to 19-fold, and accuracy above 99 % (Q20 or better).

Phylogenetics and recombination analysis

Whole-genome and gene alignments were performed using Kalign program [22], which is available online (http://www.ebi.ac.uk/Tools/msa/kalign/), using default parameters. The alignment was visualised and checked for errors using BioEdit v.7.0.5 [23]. A global pairwise alignment was performed using the mVISTA option as part of the LAGAN toolkit (http://lagan.stanford.edu/lagan_web/index.shtml) [24]. Phylogenetic relationships were reconstructed by the NJ method with Kimura's two-parameters as the model of nucleotide substitution, as incorporated in MEGA v5 [25], with 1000 bootstrap replicates.

The ratio dN/dS was estimated using the Nei and Gojobori (1986) method as implemented in the Perl-based SNAP program (http://www.hiv.lanl.gov/content/sequence/SNAP/SNAP.html) [26].

For recombination detection, the multiple alignments were analysed using a bootscan approach implemented in SimPlot version 3.5.1 [27]. For complete genomic analysis, a window size of 1000 bp and a step size of 200 bp were set. Hexon, penton, and fibre genes were analysed by bootscan using different combinations of sequence datasets (both close- and distantly-related HAdV-Ds). The following complete genomic nucleotide sequences representing all available prototypes of the HAdV-D species were used for the analysis (GenBank accession numbers in parentheses): HAdV-D8 (AB448767), HAdV-D9 (AJ854486), HAdV-D13 (JN226747), HAdV-D15 (AB562586), HAdV-D17 (AF108105), HAdV-D19 (JQ326209), HAdV-D20 (JN226749), HAdV-D22 (FJ404771), HAdV-D23 (JN226750), HAdV-D24 (JN226751), HAdV-D25 (JN226752), HAdV-D26 (EF153474), HAdV-D27 (JN226753), HAdV-D28 (FJ824826), HAdV-D29 (JN226754), HAdV-D30 (JN226755), HAdV-D32 (JN226756), HAdV-D33 (JN226758), HAdV-D36 (GQ384080), HAdV-D37 (DQ900900), HAdV-D38 (JN226759), HAdV-D39 (JN226760), HAdV-D42 (JN226761), HAdV-D43 (JN226762), HAdV-D44 (JN226763), HAdV-D45 (JN226764), HAdV-D46 (AY875648), HAdV-D47 (JN226757), HAdV-D48 (EF153473), HAdV-D49 (DQ393829), HAdV-D51 (JN226765), HAdV-D53 (FJ169625), HAdV-D54 (AB333801), HAdV-D56 (HM770721), HAdV-D58 (HQ883276), HAdV-D59 (JF799911), HAdV-D60 (HQ007053), HAdV-D62 (JN162671), HAdV-D63 (JN935766), HAdV-D64 (EF121005), HAdV-D65 (AP012285), HAdV-D67 (AP012302), HAdV-D69 (JN226748), HAdV-D70 (KP641339), HAdV-D71 (KF268207) and HAdV-D72 (KF268335).

Ethical approval

Written informed consent was obtained from parents or guardians prior to study participation. This study was approved by the Ethics Committee at Instituto de Investigaciones en Ciencias de la Salud-Universidad Nacional de Asunción, and Hospital General Pediátrico Niños de Acosta Ñu.

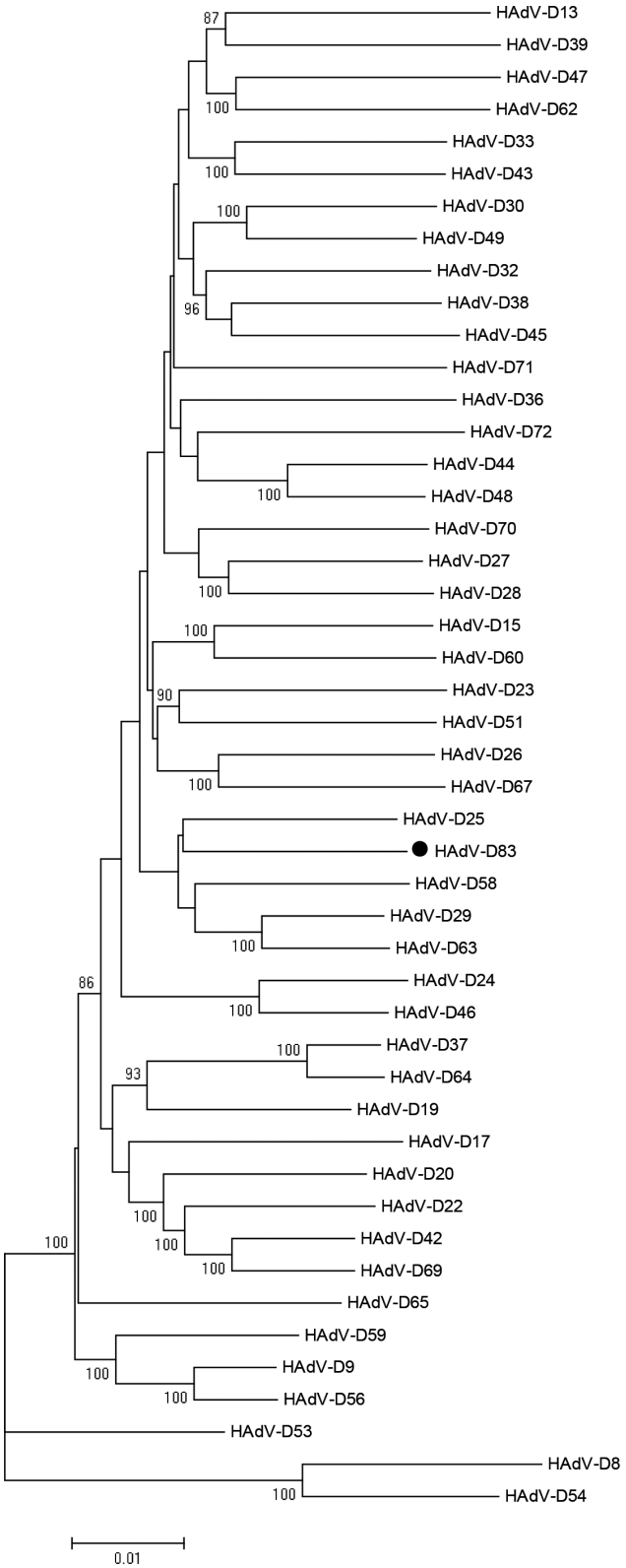


Fig. 5. NJ tree of whole HAdV-D genomes. HAdV-D83 is marked with a filled black circle. Bootstrap values greater than 80 % are shown at branch nodes. Branch distance is indicated by a scale bar at the bottom of the tree.

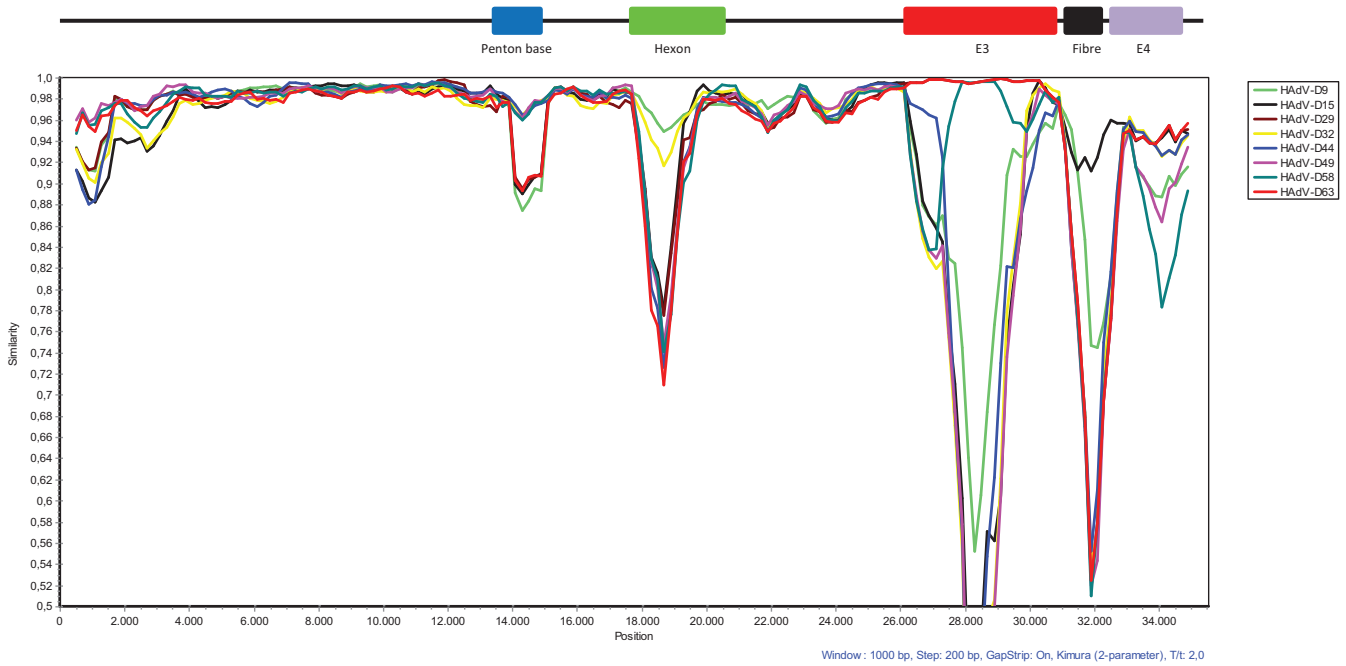


Fig. 6. Simplot analysis of HAdV-D83 whole genome. A window size of 1000 bp and a step size of 200 bp were implemented for analysis using the SimPlot software.

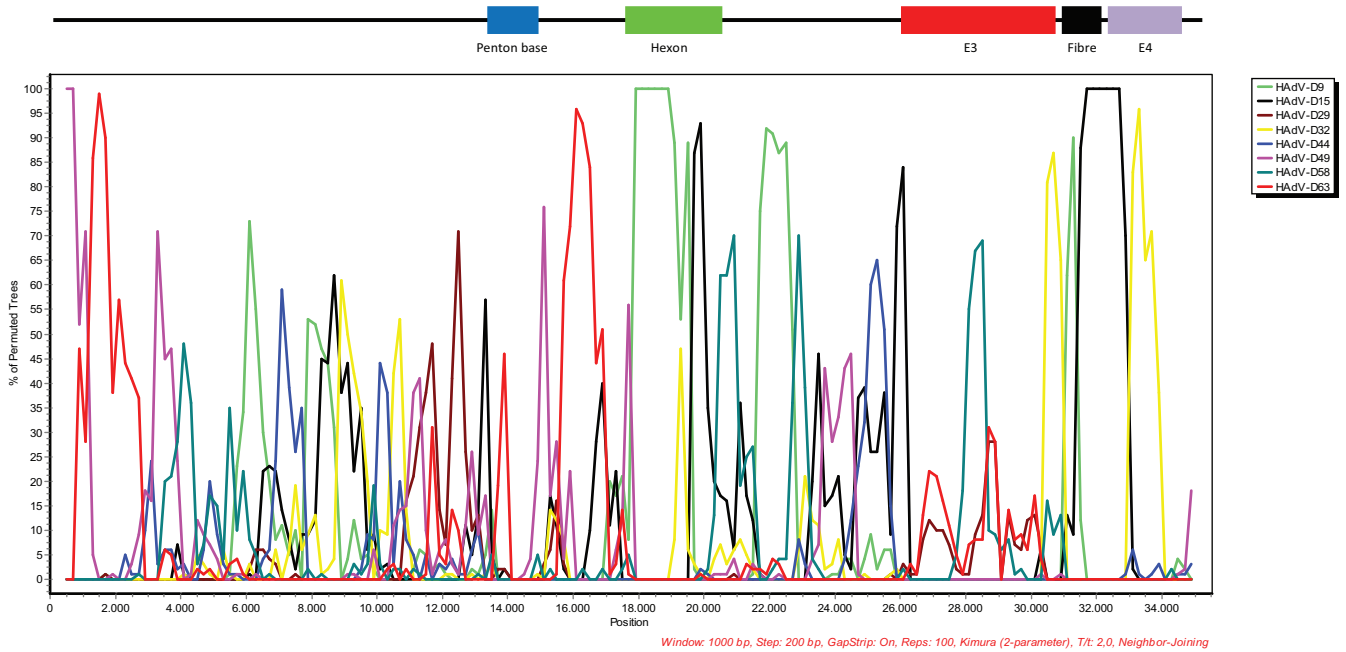


Fig. 7. Bootscan analysis of HAdV-D83 whole genome. A window size of 1000 bp and a step size of 200 bp were implemented for analysis using the SimPlot software.

Funding information
JCB, GR, and EEE received financial support from Universidad Nacional de Asunción (Rectorado-UNA); GR and EEE received funds from PRONII-CONACYT (Paraguay); SM and JA received funds from PEDECIBA and ANII (Uruguay).

Acknowledgements
We would like to thank Dr Wilma Basualdo, Dr Viviana Pavlicich and Dr Carolina Aquino, from Hospital General Pediátrico Niños de Acosta Ñu (Paraguay), for their professional support throughout the study.

Conflicts of interest
The authors declare that there are no conflicts of interest.

References

1. Lynch JP 3rd, Fishbein M, Echavarria M. Adenovirus. *Semin Respir Crit Care Med* 2011;32:494–511.
2. Benko M, Harrach B, Russell WC. Family *Adenoviridae*. In: van Regenmortel HV (editor). *Virus Taxonomy: Seventh Report of the International Committee on Taxonomy of Viruses*. San Diego: Academic Press; 2000. pp. 227–238.
3. Singh G, Robinson CM, Dehghan S, Jones MS, Dyer DW *et al.* Homologous recombination in E3 genes of human adenovirus species D. *J Virol* 2013;87:12481–12488.
4. Matsushima Y, Shimizu H, Kano A, Nakajima E, Ishimaru Y *et al.* Genome sequence of a novel virus of the species human adenovirus d associated with acute gastroenteritis. *Genome Announc* 2013;1:e00068-12.
5. Berk AJ. Adenoviridae. In: Knipe DM, Howley PM, Cohen JI, Griffin DE, Lamb RA *et al.* (editors). *Fields Virology*. Philadelphia: Lippincott Williams & Wilkins; 2013. pp. 1704–1731.
6. Yuan X, Qu Z, Wu X, Wang Y, Liu L *et al.* Molecular modeling and epitopes mapping of human adenovirus type 3 hexon protein. *Vaccine* 2009;27:5103–5110.
7. Seto D, Chodosh J, Brister JR, Jones MS, Members of the Adenovirus Research Community. Using the whole-genome sequence to characterize and name human adenoviruses. *J Virol* 2011;85:5701–5702.
8. Wold WSM, Ison MG. Adenoviruses. In: Knipe DM, Howley PM, Cohen JI, Griffin DE, Lamb RA *et al.* (editors). *Fields Virology*. Philadelphia: Lippincott Williams & Wilkins; 2013. pp. 1732–1767.
9. Jawetz E, Kimura S, Nicholas AN, Thygeson P, Hanna L. New type of APC virus from epidemic keratoconjunctivitis. *Science* 1955;122:1190–1191.
10. De Jong JC, Wermenbol AG, Verweij-Uijterwaal MW, Slaterus KW, Wertheim-van Dillen P *et al.* Adenoviruses from human immunodeficiency virus-infected individuals, including two strains that represent new candidate serotypes Ad50 and Ad51 of species B1 and D, respectively. *J Clin Microbiol* 1999;37:3940–3945.
11. Robinson CM, Singh G, Henquell C, Walsh MP, Peigue-Lafeuille H *et al.* Computational analysis and identification of an emergent human adenovirus pathogen implicated in a respiratory fatality. *Virology* 2011;409:141–147.
12. Robinson CM, Singh G, Lee JY, Dehghan S, Rajaiya J *et al.* Molecular evolution of human adenoviruses. *Sci Rep* 2013;3:1812.
13. Lasaro MO, Ertl HC. New insights on adenovirus as vaccine vectors. *Mol Ther* 2009;17:1333–1339.
14. Madisch I, Harste G, Pommer H, Heim A. Phylogenetic analysis of the main neutralization and hemagglutination determinants of all human adenovirus prototypes as a basis for molecular classification and taxonomy. *J Virol* 2005;79:15265–15276.
15. Young CS, Cachianes G, Munz P, Silverstein S. Replication and recombination in adenovirus-infected cells are temporally and functionally related. *J Virol* 1984;51:571–577.
16. Crawford-Miksza LK, Schnurr DP. Adenovirus serotype evolution is driven by illegitimate recombination in the hypervariable regions of the hexon protein. *Virology* 1996;224:357–367.
17. Barrero PR, Valinotto LE, Tittarelli E, Mistchenko AS. Molecular typing of adenoviruses in pediatric respiratory infections in Buenos Aires, Argentina (1999–2010). *J Clin Virol* 2012;53:145–150.
18. Belsy A, Odalys V, Alexander P, Clara S, Angel G *et al.* Molecular characterization of adenoviral infections in Cuba: report of an unusual association of species D adenoviruses with different clinical syndromes. *Arch Virol* 2009;154:619–627.
19. Henquell C, Boeuf B, Mirand A, Bacher C, Traore O *et al.* Fatal adenovirus infection in a neonate and transmission to health-care workers. *J Clin Virol* 2009;45:345–348.
20. Walsh MP, Chintakuntlawar A, Robinson CM, Madisch I, Harrach B *et al.* Evidence of molecular evolution driven by recombination events influencing tropism in a novel human adenovirus that causes epidemic keratoconjunctivitis. *PLoS One* 2009;4:e5635.
21. Ishiko H, Shimada Y, Konno T, Hayashi A, Ohguchi T *et al.* Novel human adenovirus causing nosocomial epidemic keratoconjunctivitis. *J Clin Microbiol* 2008;46:2002–2008.
22. Lassmann T, Frings O, Sonnhammer EL. Kalign2: high-performance multiple alignment of protein and nucleotide sequences allowing external features. *Nucleic Acids Res* 2009;37:858–865.
23. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp Ser* 1999;41:95–98.
24. Brudno M, Do CB, Cooper GM, Kim MF, Davydov E *et al.* LAGAN and Multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA. *Genome Res* 2003;13:721–731.
25. Tamura K, Peterson D, Peterson N, Stecher G, Nei M *et al.* MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 2011;28:2731–2739.
26. Korber B. *HIV Signature and Sequence Variation Analysis. Computational Analysis of HIV Molecular Sequences*. Dordrecht, Netherlands: Kluwer Academic Publishers; 2000.
27. Lole KS, Bollinger RC, Paranjape RS, Gadkari D, Kulkarni SS *et al.* Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J Virol* 1999;73:152–160.