The complete genome sequence of human adenovirus 84, a highly recombinant new *Human mastadenovirus D* type with a unique fiber gene

Győző L. Kaján[a,b], Adriana E. Kajon[c], Alexis Castillo Pinto[d], Dániel Bartha[b], and Niklas Arnberg[a]
[a] Division of Virology, Department of Clinical Microbiology, Umeå University, SE-90185 Umeå, Sweden
[b] Institute for Veterinary Medical Research, Centre for Agricultural Research, Hungarian Academy of Sciences, Hungária krt. 21, H-1143 Budapest, Hungary
[c] Infectious Disease Program, Lovelace Respiratory Research Institute, 2425 Ridgecrest Drive SE, Albuquerque, NM 87108, United States
[d] Instituto Conmemorativo Gorgas de Estudios de la Salud, Calle 36 Este, Panamá, Panama

Corresponding author: Győző L. Kaján, kajan.gyozo@agrar.mta.hu

Abstract

A novel human adenovirus was isolated from a pediatric case of acute respiratory disease in Panama City, Panama in 2011. The clinical isolate was initially identified as an intertypic recombinant based on hexon and fiber gene sequencing. Based on the analysis of its complete genome sequence, the novel complex recombinant *Human mastadenovirus D* (HAdV-D) strain was classified into a new HAdV type: HAdV-84, and it was designated Adenovirus D human/PAN/P309886/2011/84[P43H17F84]. HAdV-D types possess usually an ocular or gastrointestinal tropism, and respiratory association is scarcely reported. The virus has a novel fiber type, most closely related to, but still clearly distant from that of HAdV-36. The predicted fiber is hypothesised to bind sialic acid with lower affinity compared to HAdV-37. Bioinformatic analysis of the complete genomic sequence of HAdV-84 revealed multiple homologous recombination events and provided deeper insight into HAdV evolution.

Keywords

human adenovirus, complete genome, HAdV-84, homologous recombination, HAdV-D, acute respiratory disease

Human adenoviruses (HAdVs) belong to genus *Mastadenovirus* and are classified into seven species: *Human mastadenovirus A-G* (Harrach et al., 2011). Within each of these species, virus strains were originally classified as serotypes using serum neutralisation (Harrach et al., 2011), but currently bioinformatic methods are used for typing and also for the description of new genotypes both for human and non-human adenoviruses (Kaján, 2016; Madisch et al., 2006; Seto et al., 2011; Seto et al., 2013). Candidate novel HAdV genotypes are currently evaluated, approved and assigned names by the Human Adenovirus Working Group (hadvwg.gmu.edu) to reduce conflicts and confusion with designation.

Human adenoviruses are causative agents of respiratory, gastrointestinal, urinary and ocular diseases, but adenovirus-based vectors are also being developed for cancer treatment, prevention of infectious diseases, and/or to correct genetic disorders (Alonso-Padilla et al., 2016; Baden et al., 2016; Majhen et al., 2014; Zhang and Seto, 2015).

As part of our effort seeking to identify new potential vector candidates, we obtained the complete genomic sequence for a respiratory isolate originally identified as a unique intertypic recombinant of species HAdV-D strains. Strain P309886 originated from a convenience collection of pediatric respiratory isolates from the Gorgas Institute in Panama. It was isolated from a nasopharyngeal swab of a nine month old female hospitalized for acute lower respiratory infection in Panama in 2011. The clinical isolate was easily propagated in A549 human alveolar epithelial cells. Intracellular viral DNA was purified from infected cells following the extraction protocol developed by Kajon and Erdman (2007). Amplification and sequencing of the hexon hypervariable regions 1-7 and of the fiber gene identified the strain as a candidate new genotype with a HAdV-17-like hexon and a novel fiber gene most closely related to that of HAdV-36. Highly pure genomic DNA was further characterized by Ion Torrent next generation sequencing at the Uppsala Genome Center of the National Genomics Infrastructure, SciLifeLab (Uppsala, Sweden). The resulting 412,336 reads were normalised to a 60-times coverage using BBNorm from the BBTools suite. The normalised reads were *de novo* assembled using Mira version 4.9.5_2 (Chevreux et al., 1999), and the original sequence reads were mapped to the resulting consensus sequence using the Geneious mapper on the highest sensitivity with five iterations in Geneious 9.1.8 (Kearse et al., 2012). The new consensus sequence was annotated based on HAdV-D reference strain genome annotations using the Annotate & Predict function of Geneious. The annotation was manually checked and edited. Apparent reading frame shifts were identified in the pVI and the IVa2 open reading frames, which were resolved by additional Sanger sequencing (pVI) or by a thorough manual analysis of the involved reads and edition of the consensus sequence (IVa2).

The complete genome sequence of strain P309886 was found to be 35,257 bp long with a G+C-content of 57.3%. After mapping the sequence reads to the *de novo* assembly, the final read coverage minimum was 85, the mean was 2830.6, the read coverage's standard deviation was 1080.0, and the average of base quality sums was 2789.2. A typical HAdV-D genome layout was observed with 37 protein coding sequences and two virus-associated RNAs (Figure 1A). The genome sequence of P309886 was submitted to the NCBI Nucleotide database with accession number MF416150.

Based on the preliminary identification of the novel virus as a member of species *Human mastadenovirus D*, all known HAdV types belonging to this species were included into subsequent, comparative analyses. Phylogenetic analyses were based on multiple sequence alignments conducted using MAFFT (Katoh and Standley, 2013). The MAFFT E-INS-i multiple sequence alignment method was used for the complete genome (alignment A) and the E3 genome region (alignment B) nucleotide sequences, and for the complete hexon (alignment C), hexon loop 1 (determined based on the results of Yuan et al. (2009), alignment D) and penton base (alignment E) amino acid (aa) sequences. The MAFFT G-INS-i multiple sequence alignment method was used for the complete DNA polymerase (alignment F), fiber (alignment G) and fiber knob (alignment H) aa sequences. Recombination events were analysed using SimPlot 3.5.1 (Lole et al., 1999) based on alignment A and its subslices corresponding to the penton base, hexon and fiber genes and the E3 genomic region. Phylogenetic calculations were performed using RAxML 8.2.10 (Stamatakis, 2014) based on Gblocks 0.91b edited alignments A-H (Talavera and Castresana, 2007). Evolutionary model selections for A and B nucleotide alignments were performed using MEGA 6.06 (Tamura et al., 2013), and for the protein alignments using RAxML. The robustness of the trees was determined by a non-parametric bootstrap calculation using 1024 repeats. (To accelerate phylogenetic analyses, the calculations were run parallel on 32 processor cores. RAxML always calculates integer multiples of the used cores, and uses the smallest value above the desired one as the number of replicates. In this case, 1000 replicates were targeted, so 1024 [32*32] was conducted.) Phylogenetic trees were visualized using MEGA 6.06 (Tamura et al., 2013), trees were rooted on midpoint and bootstrap values were applied in percentage and indicated if they reach 75%. After all eight phylogenetic calculations, the pair-wise distances (branch lengths) were also calculated for the different trees to assess the closest related HAdV types to strain P908886. Sequence identity values to these closest related HAdV types were determined from the unedited sequence alignments using Geneious.

The DNA polymerase protein of strain P309886 shares 99.5% aa identity with that of HAdV-58, which further supports classification into the family *Adenoviridae*, genus *Mastadenovirus*, species *Human mastadenovirus D* according to the species demarcation criteria of the International Committee on Taxonomy of Viruses (Harrach et al., 2011). Further typing was based on the analysis of homologies with the genes encoding major capsid proteins of adenoviruses (hexon, penton base and fiber) (Seto et al., 2011). The analysis suggested P309886 to be a multiple recombinant HAdV-D with penton base protein being most similar to that of HAdV-43 (aa identity: 98.3%), the hexon being most similar to that of HAdV-17 (99.4%), and the fiber (84.1%) and its knob domain (75.7%) being distantly related to the corresponding sequences of HAdV-36 (summarised in Table 1 and in Figure 1B). Such low levels of sequence identity for the fiber suggested the designation of the virus with a new HAdV type ordinal: HAdV-84. Accordingly the strain was designated as Adenovirus D human/PAN/P309886/2011/84[P43H17F84].

The viral genome possesses a typical HAdV-D backbone. Because the conserved, less variable coding sequences share a high percentage of identity with numerous HAdV-D types (Table 2), it is impossible to identify which type exactly served as the ancestral backbone. Further evaluation of our phylogenetic and recombination analysis data (Table 1 and Figure 1) revealed that in addition to the already mentioned types, several additional types appear to have contributed to the final composition of the HAdV-84 genome, through multiple recombination events. Recombination is one of the major driving forces in HAdV evolution (Gonzalez et al., 2015; Robinson et al., 2013). The penton base's variable loop is most similar to that of HAdV-8, while the remaining penton base backbone is more similar to HAdV-43. The nucleotide sequence encoding membrane glycoprotein E3 CR1-gamma can be divided to two stretches based on the similarity to other HAdV-D types. Its 5' end (genome interval nt 28,874–29,450), coding for the N-terminus of the protein, is most similar to HAdV-48, while the 3' end (nt 29,451–29,752) is most similar to HAdV-62. SimPlot analysis (Figure 1B) also suggests that CR1-beta proteins of HAdV-27, -28, -70 and -84 (the analysed strain) are relatively closely related to each other, and these proteins are distantly related to their homologues found in other HAdV-D adenoviruses. Most probably, these genome regions originate from a common ancestor, and several recombination events placed them in the four distinct HAdV types. Singh et al. (2013) identified the E3 region of the genome as a hot-spot for recombination events, and grouped the HAdV-27/28 CR1-beta proteins into one proteotype. The corresponding proteins of HAdV-70 and -84 can also be classified into the same proteotype 9 cluster.

Phylogenetic and SimPlot analyses also suggest that in addition to recombination events, accumulation of mutations also separated coding sequences in the HAdV-84 genome from their homologues in other HAdV types (Figures 1 and 2). This is most evident in the case of the fiber protein, where the level of separation is very high, resulting in a unique and very distantly related fiber. A bioinformatic analysis was performed to assess this fiber's potential receptor-binding capabilities. Desmoglein 2- (DSG2), CD46-, sialic acid- and coxsackie and adenovirus receptor- (CAR) binding was analysed by comparing the HAdV-84 fiber sequence to those of HAdV-3 (NCBI Protein: YP_002213796) (Wang et al., 2013), HAdV-11 (AAN62521) (Persson et al., 2007), HAdV-19 (1UXB_A) and HAdV-37 (ABK59080) (Burmeister et al., 2004; Nilsson et al., 2011; Seiradake et al., 2006), respectively. Alignments with the above mentioned fiber knobs' aa sequences were conducted using the MAFFT G-INS-I method. The fiber knob structure was modelled, too, based on fiber knob structures of HAdV-11, -19 and -37 using SWISS-MODEL's alignment mode (Biasini et al., 2014). Intermolecular hydrogen bonds were compared in Swiss PdbViewer (Guex et al., 2009) and conserved/non-conserved receptor-binding residues were analysed. The results are summarised in Table 3. Based on the low number of conserved or predicted binding residues, HAdV-84 is hypothesised not to bind to DSG2, CD46 or CAR. However, all amino acids providing a direct contact to simple sialic acid and to more complex sialic acid-containing – GD1a-like glycan motifs are conserved in HAdV-84 or only replaced by an aa belonging to the same aa class. E.g., Pro317 is replaced by an isoleucine at the corresponding alignment position, both belonging to the nonpolar, hydrophobic class. Furthermore, most of these conserved amino acids were also predicted to form a hydrogen bond with the receptor. The electrostatic surface potential of the HAdV-84 fiber knob and that of HAdV-37 (sialic acid binder, 1UXA) and HAdV-5 (CAR-binder, 1KNB) (Xia et al., 1994) were also compared using the Adaptive Poisson-Boltzmann Solver plug-in of PyMol 1.8.4.0 (Baker et al., 2001; Dolinsky et al., 2007; Schrödinger, 2015). Whereas there are areas of positive electrostatic surface potential, the HAdV-84 fiber knob is less positive both in general and in the sialic acid-binding central pocket, too, compared to the overall positively charged HAdV-37 knob surface (Figure 3). The predicted isoelectric points of the fiber knobs are closer for HAdV-37 and -84 (9.59 and 8.41, respectively; 6.78 for HAdV-5) than the electrostatic surface potentials, still a less positive charge was observed here, too. Since charge-dependent interactions contribute to the fiber knob-sialic acid interaction (Arnberg et al., 2002), we hypothesised that HAdV-84 binds sialic acid with lower affinity compared to HAdV-37. However, these predictions should be handled cautiously, as even crystallographic structure might contradict measured affinity (Seiradake et al., 2006), and here only modelled structures were analysed. Sialic acid is used mainly by ocular adenoviruses, and most respiratory adenoviruses use other receptors such as CD46, DSG2 and/or CAR (Arnberg, 2012). These results might indicate that specific types of sialic acid-binding adenoviruses (Nilsson et al., 2011) may also cause respiratory infections.

Cases of HAdV-D respiratory infection are only reported from neonate patients. Similar to HAdV-84, HAdV-56, another member of species HAdV-D, was isolated from neonatal cases of pneumonia, although this type was also detected in cases of keratoconjunctivitis (Kaneko et al., 2011; Moallem et al., 2016; Robinson et al., 2011). This indicates that HAdV-56 and -84 have the ability to infect the respiratory tract of neonates and young children.

No serum neutralisation was performed to characterize HAdV-84 but, based on its genomic characteristics, it is likely that it will behave as a serologically intermediate strain like many HAdV-D strains described in the past as representing examples of intertypic recombination (Eiz et al., 1995; Wigand and Adrian, 1989).

This study was made possible by funding from FP7 Marie Curie Actions via the ADVEC consortium (grant agreement no.: 324325) and by intramural funding from the Lovelace Respiratory Research Institute. Phylogenetic analysis were performed on resources provided by SNIC through Uppsala Multidisciplinary Centre for Advanced Computational Science (UPPMAX) under Project b2014163.

References

Alonso-Padilla, J., Papp, T., Kaján, G.L., Benkő, M., Havenga, M., Lemckert, A., Harrach, B., Baker, A.H., 2016. Development of Novel Adenoviral Vectors to Overcome Challenges Observed With HAdV-5-based Constructs. Mol Ther 24(1), 6-16.

Arnberg, N., 2012. Adenovirus receptors: implications for targeting of viral vectors. Trends Pharmacol Sci 33(8), 442-448.

Arnberg, N., Kidd, A.H., Edlund, K., Nilsson, J., Pring-Akerblom, P., Wadell, G., 2002. Adenovirus type 37 binds to cell surface sialic acid through a charge-dependent interaction. Virology 302(1), 33-43.

Baden, L.R., Karita, E., Mutua, G., Bekker, L.G., Gray, G., Page-Shipp, L., Walsh, S.R., Nyombayire, J., Anzala, O., Roux, S., Laher, F., Innes, C., Seaman, M.S., Cohen, Y.Z., Peter, L., Frahm, N., McElrath, M.J., Hayes, P., Swann, E., Grunenberg, N., Grazia-Pau, M., Weijtens, M., Sadoff, J., Dally, L., Lombardo, A., Gilmour, J., Cox, J., Dolin, R., Fast, P., Barouch, D.H., Laufer, D.S., Group, B.-I.-H.S., 2016. Assessment of the Safety and Immunogenicity of 2 Novel Vaccine Platforms for HIV-1 Prevention: A Randomized Trial. Ann Intern Med 164(5), 313-322.

Baker, N.A., Sept, D., Joseph, S., Holst, M.J., McCammon, J.A., 2001. Electrostatics of nanosystems: application to microtubules and the ribosome. Proc Natl Acad Sci U S A 98(18), 10037-10041.

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Gallo Cassarino, T., Bertoni, M., Bordoli, L., Schwede, T., 2014. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res 42(Web Server issue), W252-258.

Burmeister, W.P., Guilligay, D., Cusack, S., Wadell, G., Arnberg, N., 2004. Crystal structure of species D adenovirus fiber knobs and their sialic acid binding sites. J Virol 78(14), 7727-7736.

Chevreux, B., Wetter, T., Suhai, S., 1999. Genome sequence assembly using trace signals and additional sequence information. German conference on bioinformatics. Vol. 99, 45-56.

Dolinsky, T.J., Czodrowski, P., Li, H., Nielsen, J.E., Jensen, J.H., Klebe, G., Baker, N.A., 2007. PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. Nucleic Acids Res 35(Web Server issue), W522-525.

Eiz, B., Adrian, T., Pring-Akerblom, P., 1995. Immunological adenovirus variant strains of subgenus D: comparison of the hexon and fiber sequences. Virology 213(2), 313-320.

Gonzalez, G., Koyanagi, K.O., Aoki, K., Watanabe, H., 2015. Interregional Coevolution Analysis Revealing Functional and Structural Interrelatedness between Different Genomic Regions in Human Mastadenovirus D. J Virol 89(12), 6209-6217.

Guex, N., Peitsch, M.C., Schwede, T., 2009. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: a historical perspective. Electrophoresis 30 Suppl 1, S162-173.

Harrach, B., Benkő, M., Both, G.W., Brown, M., Davison, A.J., Echavarría, M., Hess, M., Jones, M.S., Kajon, A., Lehmkuhl, H.D., Mautner, V., Mittal, S., Wadell, G., 2011. Family Adenoviridae. In: King, A.M.Q., Lefkowitz, E., Adams, M.J., Carstens, E.B. (Eds.), Virus Taxonomy: IXth Report of the International Committee on Taxonomy of Viruses. Elsevier, San Diego, pp. 125-141.

Kajon, A.E., Erdman, D.D., 2007. Assessment of genetic variability among subspecies B1 human adenoviruses for molecular epidemiology studies. Methods Mol Med 131, 335-355.

Kaján, G.L., 2016. Poultry Adenoviruses. In: Liu, D. (Ed.), Molecular detection of animal viral pathogens. CRC Press, Boca Raton, pp. 735-746.

Kaneko, H., Aoki, K., Ohno, S., Ishiko, H., Fujimoto, T., Kikuchi, M., Harada, S., Gonzalez, G., Koyanagi, K.O., Watanabe, H., Suzutani, T., 2011. Complete genome analysis of a novel intertypic recombinant human adenovirus causing epidemic keratoconjunctivitis in Japan. J Clin Microbiol 49(2), 484-490.

Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30(4), 772-780.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., Drummond, A., 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28(12), 1647-1649.

Lole, K.S., Bollinger, R.C., Paranjape, R.S., Gadkari, D., Kulkarni, S.S., Novak, N.G., Ingersoll, R., Sheppard, H.W., Ray, S.C., 1999. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. J Virol 73(1), 152-160.

Madisch, I., Wölfel, R., Harste, G., Pommer, H., Heim, A., 2006. Molecular identification of adenovirus sequences: a rapid scheme for early typing of human adenoviruses in diagnostic samples of immunocompetent and immunodeficient patients. J Med Virol 78(9), 1210-1217.

Majhen, D., Calderon, H., Chandra, N., Fajardo, C.A., Rajan, A., Alemany, R., Custers, J., 2014. Adenovirus-based vaccines for fighting infectious diseases and cancer: progress in the field. Hum Gene Ther 25(4), 301-317.

Moallem, M., Song, E., Jaggi, P., Conces, M.R., Kajon, A.E., Sánchez, P.J., 2016. Adenovirus and "Culture-Negative Sepsis" in a Preterm Neonate. AJP Rep 6(4), e417-e420.

Nilsson, E.C., Storm, R.J., Bauer, J., Johansson, S.M., Lookene, A., Ångström, J., Hedenström, M., Eriksson, T.L., Frängsmyr, L., Rinaldi, S., Willison, H.J., Pedrosa Domellöf, F., Stehle, T., Arnberg, N., 2011. The GD1a glycan is a cellular receptor for adenoviruses causing epidemic keratoconjunctivitis. Nat Med 17(1), 105-109.

Persson, B.D., Reiter, D.M., Marttila, M., Mei, Y.F., Casasnovas, J.M., Arnberg, N., Stehle, T., 2007. Adenovirus type 11 binding alters the conformation of its receptor CD46. Nat Struct Mol Biol 14(2), 164-166.

Robinson, C.M., Singh, G., Henquell, C., Walsh, M.P., Peigue-Lafeuille, H., Seto, D., Jones, M.S., Dyer, D.W., Chodosh, J., 2011. Computational analysis and identification of an emergent human adenovirus pathogen implicated in a respiratory fatality. Virology 409(2), 141-147.

Robinson, C.M., Singh, G., Lee, J.Y., Dehghan, S., Rajaiya, J., Liu, E.B., Yousuf, M.A., Betensky, R.A., Jones, M.S., Dyer, D.W., Seto, D., Chodosh, J., 2013. Molecular evolution of human adenoviruses. Sci Rep 3, 1812.

Schrödinger, 2015. The PyMOL Molecular Graphics System, 1.8.4.0 ed.

Seiradake, E., Lortat-Jacob, H., Billet, O., Kremer, E.J., Cusack, S., 2006. Structural and mutational analysis of human Ad37 and canine adenovirus 2 fiber heads in complex with the D1 domain of coxsackie and adenovirus receptor. J Biol Chem 281(44), 33704-33716.

Seto, D., Chodosh, J., Brister, J.R., Jones, M.S., 2011. Using the whole-genome sequence to characterize and name human adenoviruses. J Virol 85(11), 5701-5702.

Seto, D., Jones, M.S., Dyer, D.W., Chodosh, J., 2013. Characterizing, typing, and naming human adenovirus type 55 in the era of whole genome data. J Clin Virol 58(4), 741-742.

Singh, G., Robinson, C.M., Dehghan, S., Jones, M.S., Dyer, D.W., Seto, D., Chodosh, J., 2013. Homologous recombination in E3 genes of human adenovirus species D. J Virol 87(22), 12481-12488.

Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30(9), 1312-1313.

Talavera, G., Castresana, J., 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst Biol 56(4), 564-577.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol 30(12), 2725-2729.

Wang, H., Yumul, R., Cao, H., Ran, L., Fan, X., Richter, M., Epstein, F., Gralow, J., Zubieta, C., Fender, P., Lieber, A., 2013. Structural and functional studies on the interaction of adenovirus fiber knobs and desmoglein 2. J Virol 87(21), 11346-11362.

Wigand, R., Adrian, T., 1989. Intermediate adenovirus strains of subgenus D occur in extensive variety. Med Microbiol Immunol 178(1), 37-44.

Xia, D., Henry, L.J., Gerard, R.D., Deisenhofer, J., 1994. Crystal structure of the receptor-binding domain of adenovirus type 5 fiber protein at 1.7 A resolution. Structure 2(12), 1259-1270.

Yuan, X., Qu, Z., Wu, X., Wang, Y., Liu, L., Wei, F., Gao, H., Shang, L., Zhang, H., Cui, H., Zhao, Y., Wu, N., Tang, Y., Qin, L., 2009. Molecular modeling and epitopes mapping of human adenovirus type 3 hexon protein. Vaccine 27(37), 5103-5110.

Zhang, Q., Seto, D., 2015. Chimpanzee Adenovirus Vector Ebola Vaccine--Preliminary Report. N Engl J Med 373(8), 775-776.

Figure legends

Figure 1. Genomic layout and recombination analysis of human adenovirus 84.
A. The genomic layout, the SimPlot analysis and the BootScan analysis of the complete genome. Green arrows in the genome map represent protein coding sequences, red arrows represent virus associated RNAs and brown arrows represent the inverted terminal repeats. The colour code of SimPlot/BootScan analyses is available in part B. B. SimPlot and BootScan analyses of the penton base, hexon and fiber encoding genes and that of the E3 genomic region. Human adenovirus types showing the highest sequence identity in some characteristic domains or coding sequences of the E3 region are denoted by the letter H and the type number (e.g. H8 – human adenovirus 8), while the designation “NEW” expresses that human adenovirus 84 has a novel fiber type

Figure 2. Phylogenetic analysis of human adenovirus 84 (strain P309886) and all *Human mastadenovirus D* types available.
The complete genome and the E3 region analysis was based on nucleotide sequences, all other analyses were based on derived amino acid sequences.

Figure 3. Electrostatic surface potential of human adenovirus 37 (A), 84 (B) and 5 (C) fiber knob trimers viewed from the distal end of the fiber protein. Blue represents +1 kT/e electrostatic potential, red represents -1 kT/e electrostatic potential. On (A) and (B) the bound sialic acid is represented as a stick model. (A) and (C) are based on crystal structures 1UXA and 1KNB, respectively, whereas (B) is modelled based on 1UXA.

Tables

Table 1. Comparative phylogenetic analysis of human adenovirus 84 (strain P309886).

Analysed stretch	**Most similar HAdV type**					**Closest BLAST hit (06/09/2017) (BlastN for genome, BlastP for proteins)**			
	Type No	**Reference strain**	**Compared**		**Identity %**	**HAdV type**	**Accession number**	**Identity %**	**Query cover**
			Strain's name	**Genome's Acc No**					
Complete genome	27	BP-4	n.a.	JN226753	94.2%	P38H32F27	KF268325	94.6%	100%
DNA polymerase	58	human/ARG/ AdCor-96-487/ 1996/58 [P58H58F29]	human/ARG/ AdCor-96-487/ 1996/58 [P58H58F29]	HQ883276	99.5%	65	BAL41712	99.3%	92.9%
Penton base	43	n.a.	n.a.	JN226762	98.3%	43	AFK92731	98.3%	100%
Hexon	17	Ch22	D17	HQ910407	99.4%	P29H17F30	AGT77650	99.5%	100%
Hexon loop 1	17	Ch22	D17	HQ910407	97.9%	17	AFH03083	98.9%	91.3%
Fiber	36	275	n.a.	GQ384080	84.1%	36	ACY04488	84.1%	100%
Fiber knob	36	275	n.a.	GQ384080	75.7%	36	CAH18761	75.7%	100%

Acc No: accession number in the NCBI Nucleotide database; BLAST: basic local alignment search tool; HAdV: human adenovirus; Identity %: Pairwise identity percentage (nucleic acid identity for genomes and amino acid identity for proteins); n.a.: not available

Table 2. Closest BlastP hits (06/09/2017) of three conserved, less variable protein coding sequences originating from human adenovirus 84 (strain P309886).

Coding sequences	**BlastP hits**	
	Identity %	**HAdV types**
DNA polymerase	99.2-99.3 %	15, 23, 29, 48, 49, 65
100K	98.4-98.8 %	13, 29, 47, P9H46F39
core protein V	99.1-99.4 %	39, 45, 71, 83 and three additional not yet typed HAdV-D strains

Table 3. Potential receptor-binding capabilities of the human adenovirus 84 fiber

Compared HAdV type	Recep-tor	PDB ID	PMID	Binding sites for compared HAdV types		Corresponding site in HAdV-84		Additional predicted binding sites in HAdV-84	Ratio[a] (HAdV-84/ compared type)
						Conserved/ same group/ different aa/ missing (gap)	Predicted binding site		
3	DSG2	n.a.	23946456	direct[b]	Asn186	conserved	n.a.	n.a.	25%[c]
					Val189	different aa			
					Ser190	conserved			
					Asp261	missing			
					Phe265	missing			
					Leu277	missing			
					Leu296	different aa			
					Glu299	missing			
11	CD46	2O39	17220899	direct	Arg280(3)	missing	no	Tyr291/B Asn335/C	9%
					Ala281	missing	no		
					Ile282	missing	no		
					Asn283	different aa	no		
					Asp284(3)	different aa	no		
					Gln305	same group	Asn334/C		
					Thr306	conserved	no		
19	sialic acid	1UXB	15220447	direct	Tyr312	conserved	Tyr315/A(2)BC	Asn313/C	89%
					Pro317	same group	Ile320/ABC		
					Lys345	conserved	no		
37	sialic acid	1UXA	15220447	direct	Tyr312	conserved	Tyr315/A(2)BC(2)	Asn313/C	100%
					Pro317	same group	Ile320/ABC		
					Lys345	conserved	no		
	GD1a	3N0I	21151139	direct	Tyr312/A	conserved	Tyr315/A(2)	-	117%
					Pro317/A	same group	Ile320/A		
					Lys345/A	conserved	no		
					Tyr312/B	conserved	Tyr315/B(2)		
					Pro317/B	same group	Ile320/B		
					Lys345/B	conserved	Lys348/B		
				water medi-ated	Ser344/B	same group	no	Asn313/AB(2) Tyr321/AB Ser312/BC Thr349/B Trp346/C	220%
					Lys345/B	conserved	Lys348/B		
					Thr310/C	same group	Asn313/C		
					Ser344/C	same group	no		
					Ser344/C	same group	no		
	CAR	2J12	16923808	direct	Asp191(2)	conserved	no	Asn269/ABC Asp301/ABC	40%
					Ser193	conserved	Ser190/ABC		
					Gln200	same group	no		
					Lys202(2)	conserved	Lys199/ABC		
					Lys205	conserved	no		
					His231	different aa	no		
					Tyr266(2)	conserved	Tyr261/A(2)B(2)C(2)		
					Ser277(2)	different aa	no		
					Ser299	different aa	no		
					Glu351(2)	different aa	no		

The amino acid residues are represented by their three letter abbreviation and their ordinal in the fiber protein's amino acid sequence. If available, the protein chain designation (A/B/C) is given, too. The number of hydrogen bonds is given in parenthesis if more than one. E.g.: Asp191(2).

[a] Ratio of binding sites predicted for HAdV-84 and observed in the compared human adenovirus types

[b] Not binding residues, but critical residues in desmoglein 2 binding are listed here.

[c] Ratio of conserved residues in HAdV-84 and critical residues for binding in HAdV-3.

Abbreviations: aa: amino acid; CAR: coxsackievirus and adenovirus receptor; CD46: cluster of differentiation 46; DSG2: desmoglein 2; GD1a: GD1a ganglioside; HAdV: human adenovirus; n.a.: not available; PDB ID: The Protein Data Bank accession number of the compared fiber knob structure; PMID: PubMed accession number of corresponding publication

Figures

Figure 1.

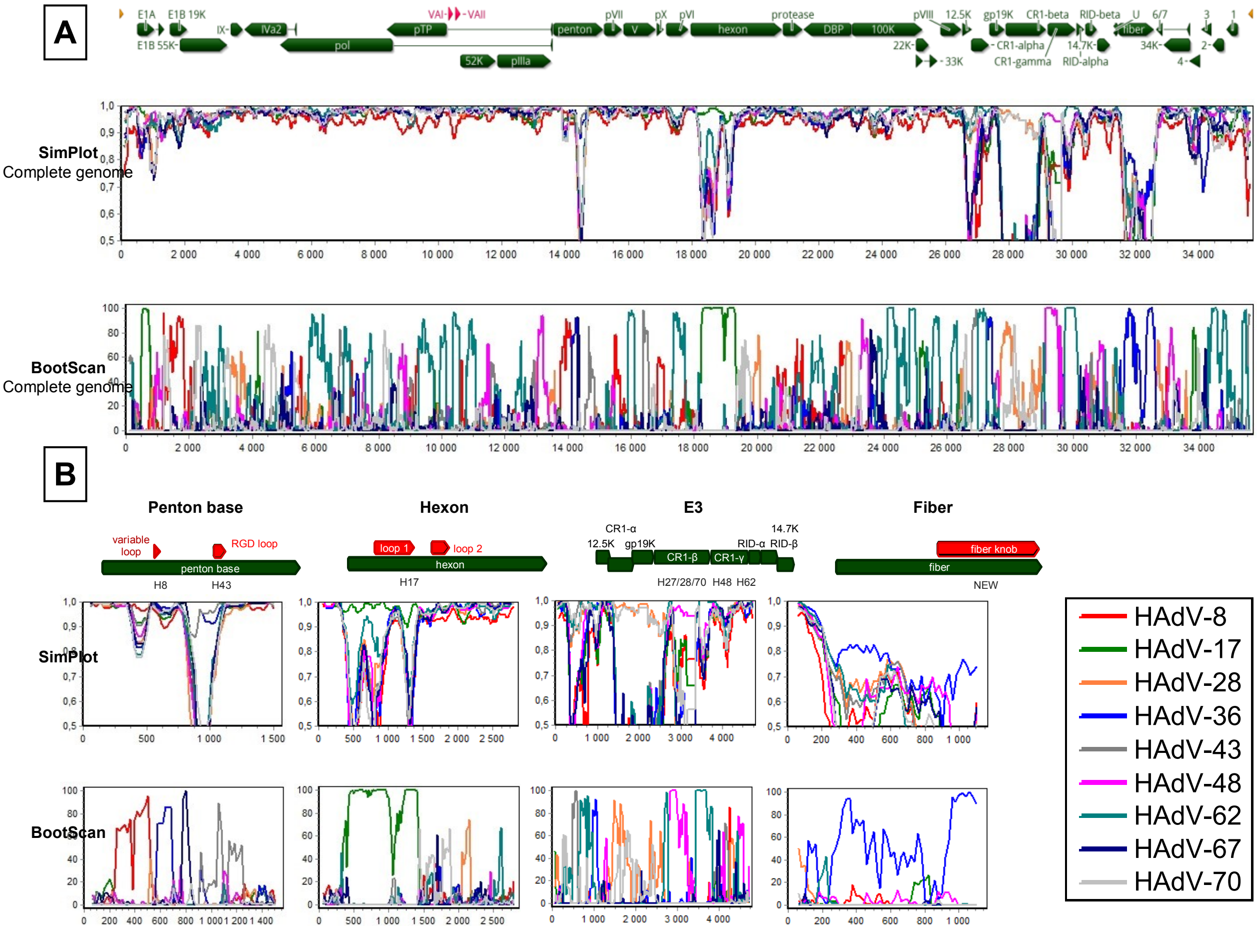
A
E1A
E1B 19K
E1B 55K
IX
IVa2
pol
VAI
VAII
pTP
52K
pIIIa
penton
pVII
V
pX
pVI
hexon
protease
DBP
100K
22K
33K
pVIII
12.5K
gp19K
CR1-alpha
CR1-gamma
CR1-beta
RID-alpha
RID-beta
14.7K
U
fiber
6/7
34K
3
2
1
4
SimPlot
Complete genome
BootScan
Complete genome
B
Penton base
Hexon
E3
Fiber
variable loop
RGD loop
penton base
H8
H43
loop 1
loop 2
hexon
H17
CR1-α
12.5K
gp19K
CR1-β
CR1-γ
RID-α
RID-β
14.7K
H27/28/70
H48
H62
fiber knob
fiber
NEW
SimPlot
BootScan
HAdV-8
HAdV-17
HAdV-28
HAdV-36
HAdV-43
HAdV-48
HAdV-62
HAdV-67
HAdV-70

Figure 2.

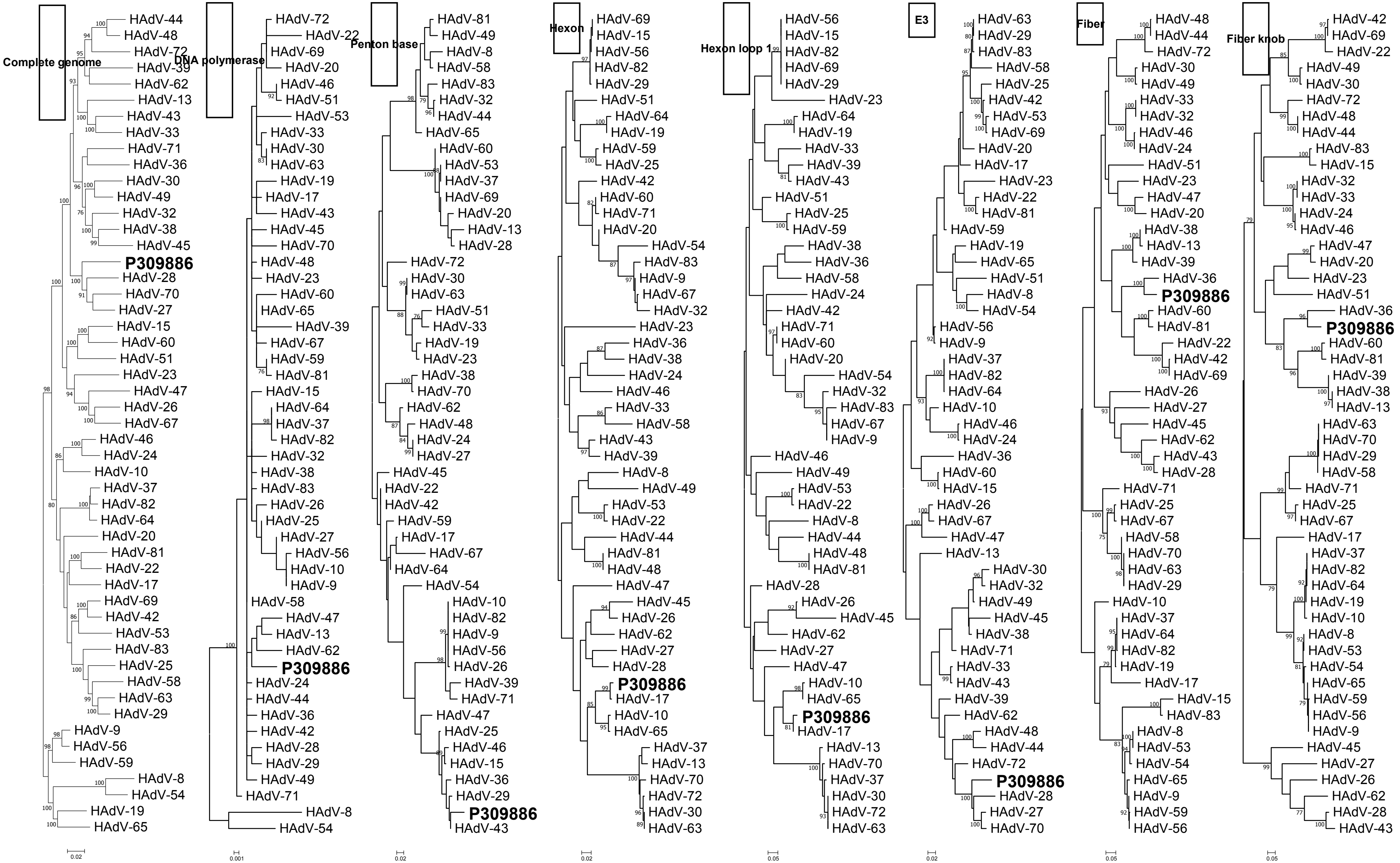
Complete genome
HAdV-44
HAdV-48
HAdV-72
HAdV-39
HAdV-62
HAdV-13
HAdV-43
HAdV-33
HAdV-71
HAdV-36
HAdV-30
HAdV-49
HAdV-32
HAdV-38
HAdV-45
P309886
HAdV-28
HAdV-70
HAdV-27
HAdV-15
HAdV-60
HAdV-51
HAdV-23
HAdV-47
HAdV-26
HAdV-67
HAdV-46
HAdV-24
HAdV-10
HAdV-37
HAdV-82
HAdV-64
HAdV-20
HAdV-81
HAdV-22
HAdV-17
HAdV-69
HAdV-42
HAdV-53
HAdV-83
HAdV-25
HAdV-58
HAdV-63
HAdV-29
HAdV-9
HAdV-56
HAdV-59
HAdV-8
HAdV-54
HAdV-19
HAdV-65
0.02
DNA polymerase
HAdV-72
HAdV-22
HAdV-69
HAdV-20
HAdV-46
HAdV-51
HAdV-53
HAdV-33
HAdV-30
HAdV-63
HAdV-19
HAdV-17
HAdV-43
HAdV-45
HAdV-70
HAdV-48
HAdV-23
HAdV-60
HAdV-65
HAdV-39
HAdV-67
HAdV-59
HAdV-81
HAdV-15
HAdV-64
HAdV-37
HAdV-82
HAdV-32
HAdV-38
HAdV-83
HAdV-26
HAdV-25
HAdV-27
HAdV-56
HAdV-10
HAdV-9
HAdV-58
HAdV-47
HAdV-13
HAdV-62
P309886
HAdV-24
HAdV-44
HAdV-36
HAdV-42
HAdV-28
HAdV-29
HAdV-49
HAdV-71
HAdV-8
HAdV-54
0.001
Penton base
HAdV-81
HAdV-49
HAdV-8
HAdV-58
HAdV-83
HAdV-32
HAdV-44
HAdV-65
HAdV-60
HAdV-53
HAdV-37
HAdV-69
HAdV-20
HAdV-13
HAdV-28
HAdV-72
HAdV-30
HAdV-63
HAdV-51
HAdV-33
HAdV-19
HAdV-23
HAdV-38
HAdV-70
HAdV-62
HAdV-48
HAdV-24
HAdV-27
HAdV-45
HAdV-22
HAdV-42
HAdV-59
HAdV-17
HAdV-67
HAdV-64
HAdV-54
HAdV-10
HAdV-82
HAdV-9
HAdV-56
HAdV-26
HAdV-39
HAdV-71
HAdV-47
HAdV-25
HAdV-46
HAdV-15
HAdV-36
HAdV-29
P309886
HAdV-43
0.02
Hexon
HAdV-69
HAdV-15
HAdV-56
HAdV-82
HAdV-29
HAdV-51
HAdV-64
HAdV-19
HAdV-59
HAdV-25
HAdV-42
HAdV-60
HAdV-71
HAdV-20
HAdV-54
HAdV-83
HAdV-9
HAdV-67
HAdV-32
HAdV-23
HAdV-36
HAdV-38
HAdV-24
HAdV-46
HAdV-33
HAdV-58
HAdV-43
HAdV-39
HAdV-8
HAdV-49
HAdV-53
HAdV-22
HAdV-44
HAdV-81
HAdV-48
HAdV-47
HAdV-45
HAdV-26
HAdV-62
HAdV-27
HAdV-28
P309886
HAdV-17
HAdV-10
HAdV-65
HAdV-37
HAdV-13
HAdV-70
HAdV-72
HAdV-30
HAdV-63
0.02
Hexon loop 1
HAdV-56
HAdV-15
HAdV-82
HAdV-69
HAdV-29
HAdV-23
HAdV-64
HAdV-19
HAdV-33
HAdV-39
HAdV-43
HAdV-51
HAdV-25
HAdV-59
HAdV-38
HAdV-36
HAdV-58
HAdV-24
HAdV-42
HAdV-71
HAdV-60
HAdV-20
HAdV-54
HAdV-32
HAdV-83
HAdV-67
HAdV-9
HAdV-46
HAdV-49
HAdV-53
HAdV-22
HAdV-8
HAdV-44
HAdV-48
HAdV-81
HAdV-28
HAdV-26
HAdV-45
HAdV-62
HAdV-27
HAdV-47
HAdV-10
HAdV-65
P309886
HAdV-17
HAdV-13
HAdV-70
HAdV-37
HAdV-30
HAdV-72
HAdV-63
0.05
E3
HAdV-63
HAdV-29
HAdV-83
HAdV-58
HAdV-25
HAdV-42
HAdV-53
HAdV-69
HAdV-20
HAdV-17
HAdV-23
HAdV-22
HAdV-81
HAdV-59
HAdV-19
HAdV-65
HAdV-51
HAdV-8
HAdV-54
HAdV-56
HAdV-9
HAdV-37
HAdV-82
HAdV-64
HAdV-10
HAdV-46
HAdV-24
HAdV-36
HAdV-60
HAdV-15
HAdV-26
HAdV-67
HAdV-47
HAdV-13
HAdV-30
HAdV-32
HAdV-49
HAdV-45
HAdV-38
HAdV-71
HAdV-33
HAdV-43
HAdV-39
HAdV-62
HAdV-48
HAdV-44
HAdV-72
P309886
HAdV-28
HAdV-27
HAdV-70
0.02
Fiber
HAdV-48
HAdV-44
HAdV-72
HAdV-30
HAdV-49
HAdV-33
HAdV-32
HAdV-46
HAdV-24
HAdV-51
HAdV-23
HAdV-47
HAdV-20
HAdV-38
HAdV-13
HAdV-39
HAdV-36
P309886
HAdV-60
HAdV-81
HAdV-22
HAdV-42
HAdV-69
HAdV-26
HAdV-27
HAdV-45
HAdV-62
HAdV-43
HAdV-28
HAdV-71
HAdV-25
HAdV-67
HAdV-58
HAdV-70
HAdV-63
HAdV-29
HAdV-10
HAdV-37
HAdV-64
HAdV-82
HAdV-19
HAdV-17
HAdV-15
HAdV-83
HAdV-8
HAdV-53
HAdV-54
HAdV-65
HAdV-9
HAdV-59
HAdV-56
0.05
Fiber knob
HAdV-42
HAdV-69
HAdV-22
HAdV-49
HAdV-30
HAdV-72
HAdV-48
HAdV-44
HAdV-83
HAdV-15
HAdV-32
HAdV-33
HAdV-24
HAdV-46
HAdV-47
HAdV-20
HAdV-23
HAdV-51
HAdV-36
P309886
HAdV-60
HAdV-81
HAdV-39
HAdV-38
HAdV-13
HAdV-63
HAdV-70
HAdV-29
HAdV-58
HAdV-71
HAdV-25
HAdV-67
HAdV-17
HAdV-37
HAdV-82
HAdV-64
HAdV-19
HAdV-10
HAdV-8
HAdV-53
HAdV-54
HAdV-65
HAdV-59
HAdV-56
HAdV-9
HAdV-45
HAdV-27
HAdV-26
HAdV-62
HAdV-28
HAdV-43
0.05

Figure 3.