




Article

Divergent Evolution of E1A CR3 in Human Adenovirus Species D

Gurdeep Singh ¹, Ashrafali M. Ismail ¹, Jeong Yoon Lee ^{1,†}, Mirja Ramke ¹, Ji Sun Lee ¹, David W. Dyer ², Donald Seto ³, Jaya Rajaiya ^{1,*} and James Chodosh ^{1,*}

¹ Department of Ophthalmology, Howe Laboratory, Massachusetts Eye and Ear Infirmary, Harvard Medical School, Boston, MA 02114, USA; gur.singh@mail.utoronto.ca (G.S.); Mohamed_Ismail@meei.harvard.edu (A.M.I.); jeongyoon.lee@jbnu.ac.kr (J.Y.L.); mirja.ramke@gmx.de (M.R.); ljs02122013@gmail.com (J.S.L.)

² Department of Microbiology and Immunology, University of Oklahoma Health Sciences Center, Oklahoma City, OK 73104, USA; david-dyer@ouhsc.edu

³ Bioinformatics and Computational Biology Program, School of Systems Biology, George Mason University, Manassas, VA 22030, USA; dseto@gmu.edu

* Correspondence: jaya_rajaiya@meei.harvard.edu (J.R.); james_chodosh@meei.harvard.edu (J.C.); Tel.: +1-617-573-6398

† Present address: Molecular Virology Laboratory, Korea Zoonosis Research Institute, Jeonbuk 11, National University, Iksan, Korea.

Received: 16 January 2019; Accepted: 3 February 2019; Published: 8 February 2019



Abstract: Adenovirus E1A is the first viral protein expressed during infection. E1A controls critical aspects of downstream viral gene expression and cell cycle deregulation, and its function is thought to be highly conserved among adenoviruses. Various bioinformatics analyses of E1A from 38 human adenoviruses of species D (HAdV-D), including likelihood clade model partitioning, provided highly significant evidence of divergence of HAdV-Ds into two distinct groups for the conserved region 3 (CR3), present only in the E1A 13S isoform. This variance within E1A 13S of HAdV-Ds was not found in any other human adenovirus (HAdV) species. By protein sequence and structural analysis, the zinc finger motif of E1A CR3, previously shown as critical for transcriptional activation, showed the greatest differences. Subsequent codon usage bias analysis revealed substantial divergence in E1A 13S between the two groups of HAdV-Ds, suggesting that these two sub-groups of HAdV-D evolved under different cellular conditions. Hence, HAdV-D E1A embodies a previously unappreciated evolutionary divergence among HAdVs.

Keywords: adenovirus; E1A; viral evolution

1. Introduction

Human adenoviruses (HAdVs) belong to the genus *Mastadenovirus* within the *Adenoviridae* family, and comprise seven species, HAdV-A to HAdV-G. HAdV-D contains the largest number of viruses presently recognized in GenBank, totaling 61 out of 90 HAdV genotypes. Moreover, HAdV-D is the fastest growing clade with 19 of the 28 most recently emerged genotypes [1–4]. We previously identified homologous recombination as the main evolutionary force driving this rapid expansion [5–9]. Recombination pressure was found to be most prominent in the major structural components of the viral capsid, specifically the penton base, hexon, and fiber proteins, along with the three E3 CR1 open reading frames (ORFs). These regions are also the most hypervariable among all the ORFs in the HAdV-D genome. These insights were possible only through continued efforts to fully characterize newly identified adenoviruses through whole-genome sequencing [5,10]. The relative abundance

of HAdV whole genome sequences now available permits the investigation of novel evolutionary mechanisms adopted by HAdV pathogens during their ontogeny.

Adenoviral E1A is notable for several reasons. Although the protein has not been crystallized, E1A is perhaps the best-characterized adenovirus protein. The *E1A* gene is the first gene expressed during acute infection, and E1A is historically important to our understanding of adenoviral transcriptional activation, the cell cycle, and cell transformation [11–13]. Deletion of E1A renders the virus transcriptionally inactive and incapable of viral replication. Two major E1A transcripts, with sedimentation coefficients of 12S and 13S, are encoded from the E1A ORF by differential splicing [14] and expressed as early as one hour after infection. These two isoforms share three out of four domains, comprising conserved regions (CR) 1, 2, and 4 [15–17], which are conserved among all human and simian adenoviruses (SAdVs). The two E1A isoforms differ by the presence of conserved region 3 (CR3) in the larger E1A 13S and its absence in the 12S transcript. CR3 is critically important to transcriptional activation of other early viral genes [18,19]. However, most prior studies on E1A CR3 were performed only with viruses from HAdV-C (specifically types 2 and 5), in which CR3 is relatively truncated compared to that of other HAdV species: 46 amino acids (aa) in length as compared to, for example, 62 aa in HAdV-D. Beyond one notable study by Ablack and coworkers, comparing the transcriptional activation of E1A CR3 from representative viruses of six HAdV species [20], including HAdV-D9 as the representative for HAdV-D, relatively little is known about the E1A CR3 of those HAdVs that are not within species C.

A major sub-domain described in HAdV-C5 CR3 is the N-terminal zinc finger region, believed to be critical for transcriptional activation. In E1A transactivation, this sub-domain interacts with the mediator complex subunit 23 (MED23) protein and possibly with the TATA-binding protein (TBP), the latter part of the transcription factor II D (TFIID) complex [21–25]. Another important functional CR3 sub-domain is the C-terminal promoter targeting region which binds to TBP-associated factors (TAFs) [26,27] and transcription factors, including OCT-4 and ATF-2 [28,29], whose binding sites are found in the adenoviral early gene promoter regions. CR3 also binds independently to p300 for transcriptional activation [30]. These functions are balanced to some degree by E1A CR1, which interacts with transcriptional co-activators such as the CREB-binding protein (CBP) and p300 to repress their activity in host gene expression [31,32]. CR1 and CR2 both participate in the binding of E1A to the tumor suppressor retinoblastoma protein (pRb) to displace E2F transcription factors, resulting in S-phase entry [33–35]. The C-terminal region of E1A, containing CR4, binds to E1A C-terminal binding protein (CtBP) to repress host transcription through histone modifications [36].

CR1–CR4 are present in the *E1A* ORF of all known HAdVs [16], but it has been shown previously that the impact of E1A on the infected host can be HAdV-species and type-dependent, further suggesting that a detailed examination of E1A in HAdV species other than HAdV-C is warranted. For example, HAdV-A12 E1A is oncogenic in hamsters and down-regulates MHC class I in rat transformed cells, while HAdV-C5 E1A has no similar activity [37–39]. Conservation of putatively critical amino acids in CR regions in multiple species has been studied [17], but the ontogeny of differences between or within species has not. We previously found that proteotyping, a novel bioinformatics tool used to elucidate amino acid signatures, [5,40] was invaluable for understanding intra-species conservation within hypervariable E3 regions of HAdV-D [6], resulting in the identification of unique proteotypes having potentially unique functional features [41]. In the current study, we applied proteotyping along with classical genomic and structural studies to demonstrate two unique subgroupings within the HAdV-D E1A 13S transcript, with evolutionary significance.

2. Materials and Methods

2.1. Sequence Analysis

E1A sequences from HAdVs representing species A–G used in this study were retrieved from NCBI (<https://www.ncbi.nlm.nih.gov/nucleotide/>), and aligned using MAFFT [42] and Clustal W in

MEGA6.06 [43] using default parameters. Phylogenetic trees were generated using the bootstrapped neighbor-joining distance method (500 replicates) with MEGA6.06 and maximum likelihood with PhyML (HKY85 model) [44]. The PhyML generated unrooted tree in Newick format was used to label all HAdV-Ds as a single clade partition (\$1) or two separate clade partitions, group 1 (\$1) and group 2 (\$2), while the remaining HAdVs belonging to the other 6 species (A-C, E-G) were labeled as a single clade partition (\$0 as default). The corresponding labeled trees in Newick format along with *E1A* coding sequences aligned in respective preserved codon position in PHYLIP format were used as input for Clade model C (CmC) divergence selection analysis [45]. PAML 4.8 [46] CmC analysis with settings 'Model = 3 and NSsites = 2' was used to identify divergence selection acting on two clades separating HAdV-Ds into two classes. Using codeml, an alternate model was represented with a phylogenetic tree having each class as separate clades (complex tree) and tested against the tree with all HAdV-Ds in a single clade (simpler tree) as a null model to identify divergence. A likelihood ratio test (LRT) between the models was calculated using the resulting ln-likelihood (ln L) for null and alternate models. The statistical significance was calculated from the LRT value and difference in parameter number using a χ^2 distribution [47].

2.2. Proteotyping

Proteotyping was performed as previously described [5,6]. Amino acid sequences were aligned using ClustalW, and then transferred to MEGA to generate maximum likelihood trees. *E1A* sequences were then rearranged in the multi-sequence alignment in the same order as in the phylogenetic tree. For each amino acid position, the consensus residues were identified and residues matching the consensus at each respective position colored white. Gaps in the alignment were colored black. The remaining amino acids were each assigned a unique color for all occurrences. A unique proteotype was defined as a pattern representing 10% residue divergence from the consensus sequence [5,6].

2.3. Protein Sequence and Structural Analysis

Multi-sequence amino acid alignment of *E1A* CR3 regions representing 38 HAdV-D *E1A* sequences was performed using MAFFT [42]. A protein model for HAdV-D37 and D56 CR3 regions was constructed using QUARK [48] (<http://zhanglab.ccmb.med.umich.edu/QUARK>), an ab initio protein folding and protein structure prediction algorithm, using atomic-level knowledge-based force field and Monte Carlo simulations to generate 3D protein structures in the absence of a template. The generated models were evaluated by Ramachandran plot using RAMPAGE (<http://mordred.bioc.cam.ac.uk/~rapper/rampage.php/>). At least 95% of residues in both models were found in either most-favored or allowed regions. Protein structure analysis and visualization were performed with UCSF Chimera software (<http://www.cgl.ucsf.edu/chimera>). CR3 protein structure backbone dynamics, including flexibility, and ordered and disordered regions, were predicted using DynaMine [49] (<http://dynamine.ibsquare.be/>).

2.4. Codon Usage Bias Analysis

Relative Synonymous Codon Usage (RSCU) values represent the occurrence frequency of a specific codon in a sequence as compared to other synonymous codons coding for the same amino acid under equal codon usage bias. RSCU values were calculated using DnaSP v5 [50] for all individual codons except for three stop codons, methionine, and tryptophan. The RSCU values for 38 HAdV-D sequences (62 amino acids, 59 codons) for *E1A* 13S CR3 sequence were examined by correspondence analysis (CA) in R statistical software version 3.3.2 (<http://www.R-project.org/>) [51], using the CA function in the FactoMineR package [52], and the first two dimensions were plotted.

3. Results

3.1. Segregation of the 13S Isoform of HAdV-D E1A into Two Major Subclades

The evolution of the E1A 13S isoform across 60 HAdVs, representing all seven species (A–G), was analyzed by phylogenomics at the protein level. As shown in Figure 1A, E1A 13S clades segregated as expected according to existing HAdV species delineations. The largest membered HAdV species, B and D, showed further subdivisions with two subclades each, containing roughly equivalent numbers of HAdVs in each subclade. To look more specifically at the CR3 region of E1A 13S, we applied proteotyping to these 60 HAdVs. In this method, a maximum likelihood tree is generated from the putative amino acid sequences of the genomic region of interest, and a unique color is assigned to each amino acid in the resultant amino acid assembly to permit visual discrimination of subgroups. As can be seen from proteotyping (Figure 1B), two distinct HAdV-D CR3 proteotypes emerge and are resolved more clearly than the HAdV-D subclades in Figure 1A. Only one CR3 proteotype was observed for each of the other HAdV species, including HAdV-B (despite it dividing into two subclades at the complete protein level). The CR3 proteotypes for HAdV-A, B, C, E, F, and G were more similar to each other than to HAdV-D. HAdV-C had the smallest E1A CR3 13S region, comprising only 46 amino acids. Compared to HAdV-C5, with the best-studied E1A protein, the CR3 region of HAdV-Ds appears to contain an additional 16 amino acids.

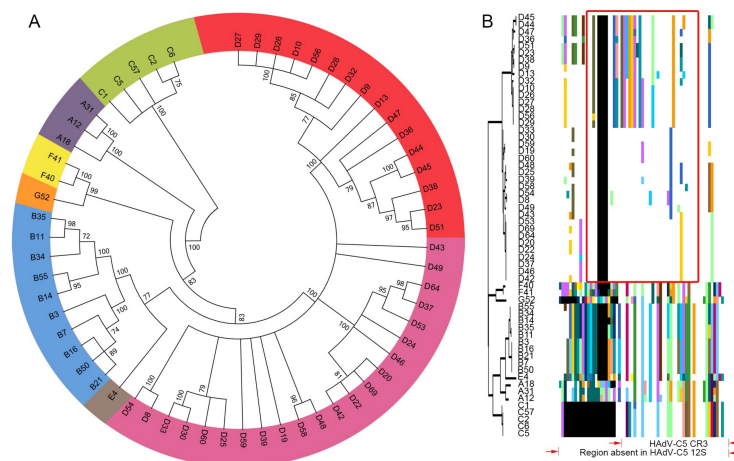


Figure 1. E1A 13S protein coding sequence analysis for human adenovirus (HAdV) species A–G reveals a unique evolutionary pattern for human adenoviruses of species D (HAdV-D). (A) Bootstrap-confirmed (500 replicates) neighbor-joining phylogenetic tree of E1A 13S protein sequence from 60 HAdVs including 38 HAdV-D. Bootstrap values below 70 were collapsed for the condensed tree. Each HAdV species is indicated by a unique color, except HAdV-D species which had two divergent subclades consisting of 16 (red) and 22 (purple) viruses. (B) Proteotyping analysis for the conserved region 3 (CR3) in E1A 13S protein for 60 HAdVs from seven HAdV species. Here, multi-sequence alignment is shown in respect to the maximum-likelihood tree with the consensus sequence hidden as a white background, while non-consensus colored regions represent unique shared protein signatures. The red box outlining the HAdV-Ds represents the most variable region within CR3 for the two HAdV-D groups, arbitrarily named group1 (upper proteotype encompassing HAdV-D45 to D29) and group2 (lower proteotype, encompassing HAdV-D33 to D42), each with highly distinct amino acid signatures. The CR3 of HAdV-Cs, as exemplified by the well-studied HAdV-C5 (red arrows), is the smallest in length among all HAdV species.

Subsequent phylogenetic analysis at the nucleotide level for the unspliced E1A ORF confirmed the clear distinction between these two subgroups within HAdV-D (Figure 2A). By proteotyping, no such division into distinct proteotypes was evident in E1A 12S (Figure 2B), nor was the origin of the difference between the two 13S subgroups apparent beyond the CR3 region (Figure 2C). Due to

differential splicing of the first exon, E1A 12S in HAdV-Ds contains 191 amino acids, while E1A 13S comprises 253 amino acids. Proteotyping of HAdV-D E1A 12S suggested four proteotypes, but these were not distinctly different from one another (Figure 2B). In contrast, proteotyping of 13S showed a remarkable divergence of one proteotype (D29 to D44) from the others, with four proteotypes in total (Figure 2D). The major source of this bifurcation of E1A was the CR3 region, absent in E1A 12S transcript, and present in E1A 13S.

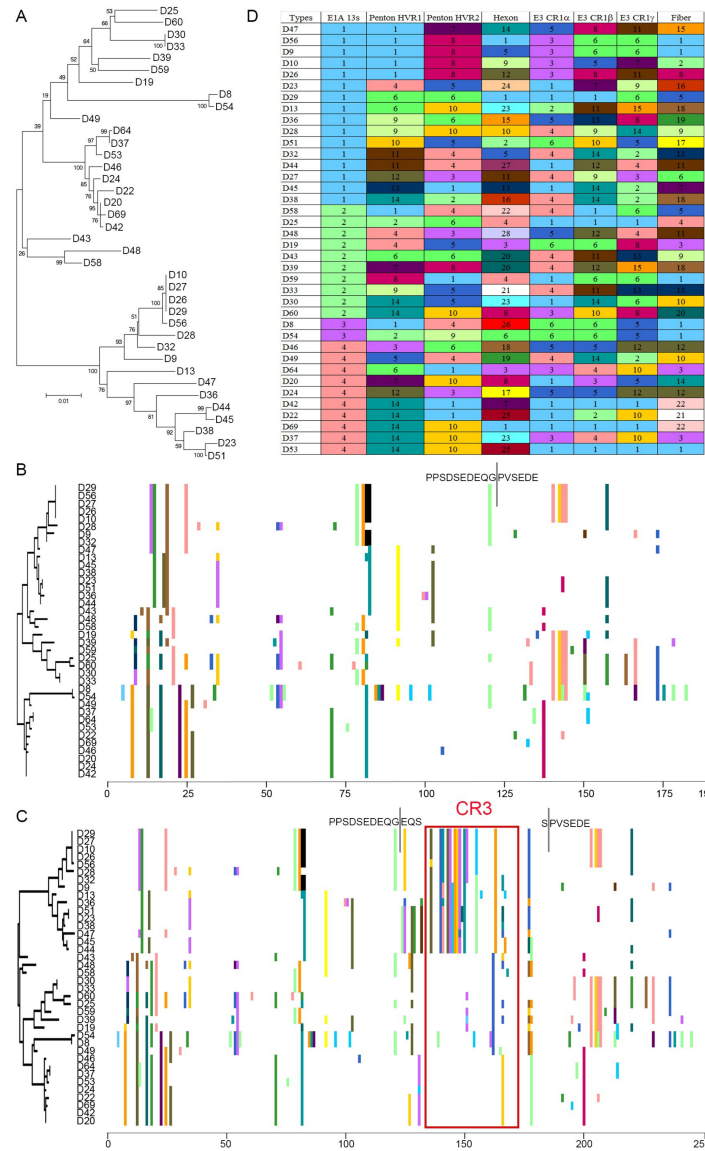


Figure 2. HAdV-D specific analysis of full E1A gene. (A) Bootstrap-confirmed (500 replicates) neighbor-joining phylogenetic tree of unspliced E1A gene nucleotide sequences from 38 HAdV-Ds. Bootstrap values below 80 are indicative of low confidence. Proteotyping analysis was performed for full coding sequences of both E1A isoforms, (B) 12S and (C) 13S, for 38 HAdV-D sequences. No significantly distinct proteotypes were observed for the E1A 12S isoform. The E1A 13S isoform clearly segregated into distinct proteotypes, arbitrarily named group1 (upper) and group2 (lower), due to differences in the CR3 region (as seen by the non-consensus protein signatures). (D) E1A recombination was identified in assortment analysis performed in the genomic position order of putative proteotypes for penton base hypervariable loops (HVL1 and HVL2), hexon, three E3 ORFs (CR1 α , β , and γ) and fiber open reading frames, with respect to E1A 13S proteotypes. For visual representation and understanding recombination at the whole genomic scale, each unique proteotype was assigned a number and color.

3.2. Evidence for Prior Homologous Recombination in HAdV-D E1A

HAdV-D genomes have been shown to contain seven distinguishable hypervariable regions that undergo homologous recombination as a mechanism of viral evolution, leading to the emergence of new pathogens [5,8,9]. These are the penton base hypervariable regions 1 and 2 (separated by ~450 nucleotides), hexon hypervariable regions 1–6 and 7 (closely adjacent), E3 CR1- α , E3 CR1- β , E3 CR1- γ , and fiber genes. Unlike the penton base and hexon genes, which contain hypervariable regions within otherwise highly conserved nucleotide sequence, the three E3 genes and the fiber gene are hypervariable over their entire ORFs. Hypervariable regions can recombine as independent units, or the recombination event initiating at one hypervariable region can extend to include a larger area of the genome [2,6]. As shown by proteotype analysis and subsequent sorting based on E1A 13S (Figure 2D), the data suggest that E1A might be an additional site of homologous recombination in HAdV-D. For example, HAdV-D47, D56, D9, D10, and D26 all share a proteotype including penton base hypervariable region (HVL)1, while the latter four viruses (D56, D9, D10, and D26) also share a proteotype with penton base HVL2. This suggests that the recombination event for HAdV-D47 E1A ended somewhere between penton base HVR1 and HVR2, while for the other four viruses the recombination event concluded after penton base HVR2 but before the hexon hypervariable regions, a pattern observed previously [53].

3.3. Evolutionary Divergence in HAdV-D E1A 13S Zinc Finger-Containing CR3 Region

Our proteotyping and phylogenetic analyses indicated that the two separate clades in HAdV-Ds evolved divergently for the E1A-13S protein. Therefore, we tested divergence within clades and site-specific substitution constraints along entire clades at the codon level through quantifying the diversifying selection pressure acting on each clade in comparison. Clade model C [45] (CmC) in PAML [46] was used to test this hypothesis with whole E1A-13S DNA sequences for 60 HAdVs using nested models. The null model had all the HAdV-Ds labeled as a single partition or one clade (D combined), while the second partition consisted of the rest of the HAdVs belonging to other species as the background clade. Although the CR3 region might be evolving in both clades, this null model assumes no divergence within the D combined clade. The alternate model had three partitions, which included two partitions for the two subclades in HAdV-D and a third as background clade with the rest of the HAdVs belonging to other species. This model assumes the divergence between the clades in HAdV-Ds and hence is a complex version of the simpler null model. There was only one parameter difference between these models. Hence, using one degree of freedom and twice the difference in the ($\ln L$) values between these models, the likelihood ratio test (LRT) produced a $p = 6.7 \times 10^{-16}$. This is consistent with a highly diversifying selection pressure between HAdV-D clades for the full E1A-13S (Table 1). To better understand the role of the CR3 region in this diversifying selection between the two clades, we performed another test by removing the CR3 region from the E1A-13S nucleotide sequence, which accounted for 62 amino acids. Using the same complex and simpler tree partition models as above, CmC was used to identify the divergence between the clades for this shortened sequence without CR3. Interestingly, the difference was no longer significant ($p = 0.108$), demonstrating that the divergent selection within CR3 is the major driver of divergence between the two HAdV-D E1A 13S clades (Table 1).

Table 1. Clade model C (CmC) clade divergence analysis.

Full E1A-13S Nucleotide Sequence							
Model	Np	ln L	κ	Null	LR	df	<i>p</i>
CmC_2partition_Dcombined	123	−7622.25	2.21				
CmC_Alternate_3partition_gro	124	−7589.64	1.98	CmC_Dcombined_2partition	65.22	1	6.71×10^{-16}
62AA CR3 sequence removed from full E1A-13S							
Model	Np	ln L	κ	Null	LR	df	<i>p</i>
CmC_2partition_Dcombined	123	−5826.58	2.06				
CmC_Alternate_3partition_gro	124	−5825.29	2.05	CmC_Dcombined_2partition	2.58	1	0.108

Np = number of parameters; ln L = ln likelihood; κ = ratio of transition to transversion. LR = twice the difference in ln likelihood between complex and simpler nested models. df = degrees of freedom, equal to difference in Np between complex and simpler nested models. *p* = *p*-value.

Further inspection of the HAdV-D 13S CR3 region by multi-sequence alignment revealed the area of greatest divergence between 13S sequences included the zinc finger motif responsible for transcriptional activation (Figure 3A), with an average identity of only 64.7% (55–70% range) between the two HAdV-D clades at the protein level. As previously reported [54], most of the E1A protein was intrinsically disordered except for the region of the CR3 zinc finger motif. The E1A CR3 zinc finger motif has been characterized as a Cys-4 zinc finger, comprised of four cysteine residues [55]. As seen in the multi-sequence alignment, these four cysteine residues, previously described for HAdV-C5, are also conserved in HAdV-D, and found at positions 153, 156, 170, and 173. Many zinc finger motif classes contain multiple histidine residues, and we identified a highly conserved histidine residue at position 159 in all 60 HAdVs. We also found a HAdV-D specific and conserved cysteine residue at position 134. This HAdV-D specific cysteine residue-containing region is absent in HAdV-C5, and its role is yet to be studied. Hence, our multi-sequence alignment suggested that HAdV-Ds have a novel zinc finger motif, with an extra (fifth) cysteine residue. Finally, we noted an additional histidine residue at position 140, present only in the larger of two 13S CR3 subgroups (group2) and of unknown significance.

The structure of E1A has not been resolved. We therefore applied *ab initio* protein modeling to compare the impact of sequence divergence between 13S CR3 subgroups on the corresponding zinc finger fold. Models of the CR3 regions of HAdV-D56, chosen as a representative for group1, and D37, from group2, were studied as examples of each subgroup (Figure 3B,C, respectively), as they share similar tropisms [1]. In the D56 CR3 protein model, four cysteine residues formed a compact zinc finger scaffold in close proximity with the conserved histidine residue at position 159. The secondary structure of the D56 CR3 protein consists of three alpha helices, and the conserved cysteine at position 134 was distant from the core zinc binding domain. In contrast, the D37 CR3 protein model showed an alpha helix with two beta strands. The HAdV-D specific cysteine C134 in D37 was part of the zinc finger scaffold. Only the Ad37 protein model incorporated all five conserved cysteines in its scaffold, while the Ad56 model was able to incorporate only four cysteines, similar to HAdV-C5. Hence, it appears that the predicted D56 E1A zinc finger structure is similar to what was previously predicted for E1A in other HAdV species, having four cysteine residues in the zinc finger, while the D37 E1A proteotype evolved to include a fifth cysteine residue in its zinc finger scaffold structure. The *ab initio* models for HAdV-56 and D37, as shown in Figure 3B,C, respectively, were generated based on the entire 62 aa sequence for the E1A CR3 region. However, the great majority of studies of E1A CR3 have used HAdV-C5. HAdV-Cs are unique in that their E1A CR3 is only 46 aa in length. It could be argued that given the intrinsically disordered character of most of E1A 13S, that the extra 16 aa in other HAdV species do not contribute to the structure of their CR3 zinc finger domain.

To further elucidate the effects of including the entire 62 aa in our models, we next compared *ab initio* models of HAdV-D56 and D37 and assessed their structural overlap using both the canonical 46 and the extended noncanonical 62 aa sequences. As shown in Figure 4, when modeling was restricted to the canonical 46 aa, the predicted structures for D56 and D37 were nearly identical (Figure 4A–C). However, when the full E1A CR3 comprising of 62 aa was modeled, the predicted structures diverged dramatically (Figure 4D–F). Including the entire 62 aa had a greater effect on the predicted structure of D37 than on D56 (Figure 4G,H). It was previously shown that ~57% of the amino acids in E1A CR3 region are required for transcriptional activation based on studies of HAdV-C5 E1A CR3 by conservative mutational analysis [17]. Moreover, the region corresponding to the first half (23 aa out of 46 aa) of the canonical E1A CR3 (of HAdV-C5) [17] in HAdV-Ds is the most divergent region between the two CR3 subgroups with only a 56% average identity. Hence, the amino acid differences between the two CR3 subgroups may be the result of evolutionary divergence in the scaffold structure while maintaining the functionality of the zinc-finger domain.

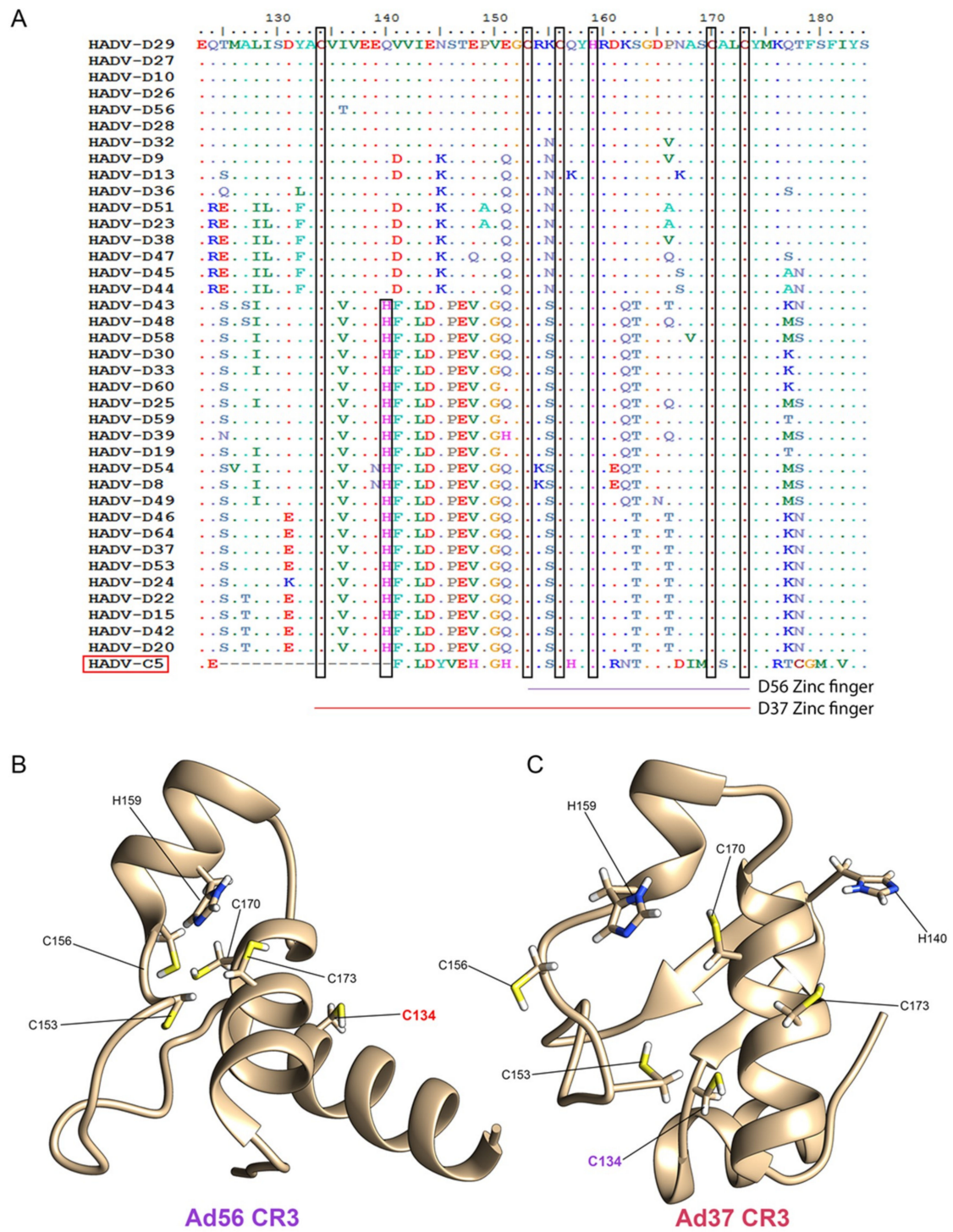


Figure 3. Predicted E1A CR3 zinc-finger scaffold structure in relationship to HAdV-D protein sequence divergence. (A) Multi-sequence alignment of 38 HAdV-Ds and the HAdV-C5 amino acid sequences representing the CR3 region, present only in the 13S E1A isoform. Conserved cysteine and histidine in the CR3 region are shown for both groups. Ab initio protein structure models for the E1A CR3 region from (B) HAdV-D56 and (C) HAdV-D37. Cysteine and histidine residues potentially involved in the zinc finger region are highlighted.

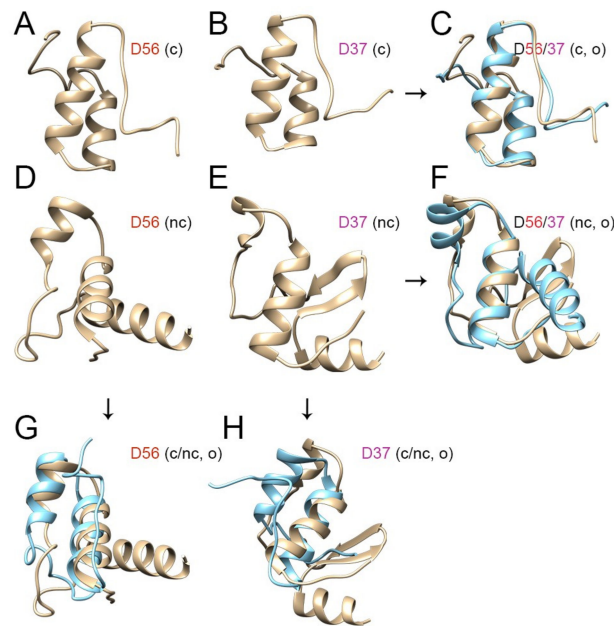


Figure 4. Ab initio protein structure model comparisons for E1A CR3 for HAdV-D56 and D37. The canonical (c) 46 amino acid CR3 regions for (A) D56 and (B) D37 appear very similar (C) when overlapped (o) (D56 in tan color and D37 in cyan). The full-length, noncanonical (nc) 62 amino acid sequences for (D) D56 and (E) D37 are more disparate (F). Overlapped comparison of the 46 and 62 amino acid-based models for D56 (G) shows relatively little difference in the structure of the zinc finger domain for D56, but a significant change in conformation of the putative zinc finger domain for D37 (H).

3.4. Codon Usage Bias Analysis and Divergent Evolution of HAdV-D E1A 13S

DynaMine analysis is a linear regression-based model trained with available NMR data to predict protein dynamics and disorder regions, and is based on amino acid sequences independent of structural modeling. DynaMine was used previously to identify interaction motifs in HAdV-C5 E1A [49]. Our analysis predicted greater structural flexibility in the region 140–160 aa, which flanks the cysteine 134 in HAdV-D37 E1A 13S relative to that of D56 (Figure 5A). Our structural predictions suggest therefore that these two CR3 groups differ principally in the zinc finger motif. We therefore employed codon usage bias analysis to examine the pathway of evolution yielding these two CR3 subgroups. RSCU values representing the frequency of occurrence of synonymous codons in a 59-dimensional codon space for each of the 38 HAdV-Ds were converged to 2-dimensional space. For the HAdV-D E1A 12S isoform, both subgroups had similar RSCU values (Figure 5B), implying that HAdV-D evolved in a similar codon-optimized environment for E1A 12S. However, the same analysis on the CR3 region of the 13S isoform revealed a clear separation between the two CR3 subgroups (Figure 5C). These data suggest that these two E1A CR3 subgroups within HAdV-D evolved in different cellular environments.

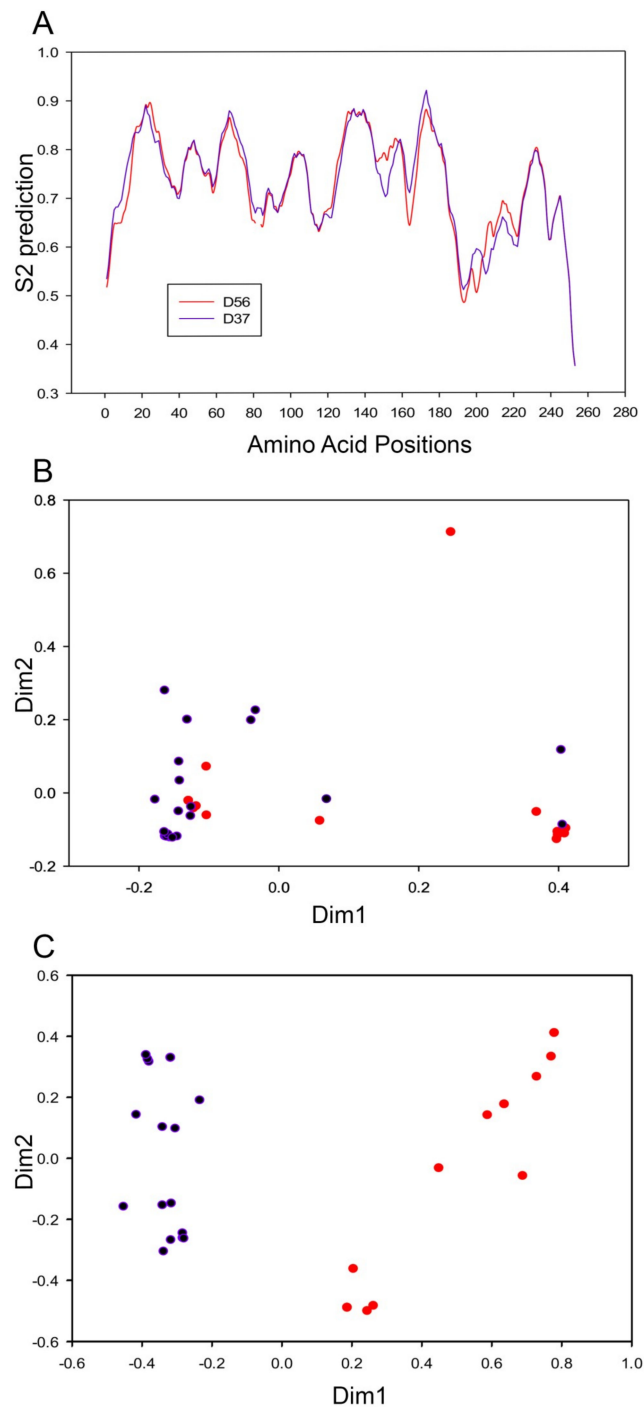


Figure 5. Divergence in structural flexibility and codon usage bias of the E1A protein. (A) DynaMine S2 predictions for E1A 13S protein flexibility and disordered regions for HAdV-D56 and HAdV-D37. Regions with S2 prediction values below 0.7 are flexible, while regions with values above 0.8 are rigid. The CR3 region is the origin for the difference in protein dynamics (at the 140–160 amino acid position) between HAdV-Ds. (B and C) Codon usage bias of 38 HAdV-Ds for 59 codons as summarized by two-dimensional plot performed with correspondence analysis for the E1A 12S isoform (B) and the 13S CR3 region alone (C). Purple color dots represent HAdV-Ds belonging to the one group, while red color dots represent the other. The difference in codon usage bias for the CR3 region indicated its evolution in different cellular environments for each group.

4. Discussion

E1A is one of the most highly studied viral proteins, with pleiotropic influences on viral gene expression, apoptosis, and the cell cycle [13]. The E1A protein consists of four regions thought to be highly conserved among HAdV species. However, in the work presented herein, we demonstrate a unique evolutionary variance in the HAdV-D E1A CR3 region with two highly divergent CR3 subgroups. The three major capsid genes and three of eight ORFs in the E3 transcription unit of HAdV-D were previously shown to contain regions of nucleotide hypervariability with clear segregation into multiple distinct proteotypes, relative to the remainder of the highly conserved HAdV-D genome [5]. HAdV-D is the largest and most rapidly expanding HAdV species [56]. Variation in nucleotide sequences coding for the three major surface capsid proteins, the hexon, penton base, and fiber, and in proteins that directly mediate immune evasion, namely E3 CR1- α , β , and γ [6], suggests successful adaptive evolution to achieve broad infectivity, replication, and viral transmission across multiple host cellular and tissue microenvironments [8,57]. The E1A CR3 region, present in the 13S but not the 12S E1A isoform, is critical to downstream viral gene expression and has a direct impact on viral replication and, therefore, on viral fitness. The *ab initio* modeling, in context with the other *in silico* analyses, suggests that the downstream transcriptional activity of E1A CR3 may differ between HAdV-D subgroups—because of differences in the amino acid backbone—which in turn would impact the motif and folding of the zinc finger scaffold critical to the binding of viral promoters for transcriptional activation. We speculate that the unique zinc-finger fold in the subgroup containing HAdV-D37, with an additional cysteine residue, might generate more robust transcriptional activation. However, whether downstream viral gene expression differs between the two HAdV-D CR3 subgroups remains to be determined by further study. Differences in the evolution (and putative function) between HAdV species have been previously shown to include marked discrepancies in size. For example, the E3 CR1 β gene in HAdV-D is almost twice the length of its homolog in any other HAdV species [58]. Similarly, as noted previously, the HAdV-D E1A CR3 is larger than that reported for HAdV-C5. In the study by Ablack and colleagues comparing transcriptional activation by E1A CR3 across HAdV species, HAdV-D9 E1A CR3 transcriptional activation of a GAL4-DNA-binding domain fusion was extremely low, and well below the comparands from other HAdV species. It would be interesting to perform similar studies comparing HAdV-D9 and other members of the subgroup containing HAdV-D56, with transcriptional activation by members of the subgroup containing HAdV-D37. Further, research into the biology of E1A has contributed to the identification of many important proteins involved in mammalian gene regulation, for example, p300 and CtBP [59]. Therefore, comparing not only transcriptional activity but also the protein interactomes of HAdV-D E1A CR3 subgroups may reveal novel binding partners and transcription factor machinery components whose function is yet to be elucidated.

Our studies of HAdV-D E1A CR3 codon usage bias indicated that both CR3 subgroups came from a common ancestor, but took different evolutionary paths to adapt to different host environments. It is unclear what specific host adaptation pressure(s) may have influenced codon usage between E1A CR3 subgroups. The close relationship between simian and human adenoviruses and prior evidence for interspecies transmission [60,61] and co-evolution [7,62,63] suggest one such possibility. Regardless, our results provide a platform to better understand zinc-finger transcription factor motif evolution in the context of viral pathogenesis, and may contribute to improvements in the design of synthetic zinc finger proteins [64].

In summary, in this study, we present the discovery and delineation of two subgroups of E1A CR3 within HAdV-D species, each with a different evolutionary history and predicted protein structure. Next-generation sequencing, genomics, and bioinformatics have made it possible to understand previously hidden aspects of adenoviral evolution.

Author Contributions: Conceptualization, G.S., J.R., and J.C.; methodology, G.S., A.M.I., J.Y.L., M.R., J.S.L., and J.R.; validation, A.M.I. and J.Y.L.; formal analysis, G.S.; resources, J.R. and J.C.; writing—original draft preparation, G.S.; writing—review and editing, A.M.I., D.W.D., D.S., J.R., and J.C.; supervision, J.R. and J.C.; project administration, J.R. and J.C.; funding acquisition, J.R. and J.C.

Acknowledgments: This work was funded by National Institutes of Health grants EY013124, EY021558, and EY014104, a Senior Scientific Investigator Award grant (to J.C.) from Research to Prevent Blindness, Inc., New York, NY, The Falk Foundation, and the Massachusetts Lions Eye Research Fund.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Robinson, C.M.; Singh, G.; Henquell, C.; Walsh, M.P.; Peigue-Lafeuille, H.; Seto, D.; Jones, M.S.; Dyer, D.W.; Chodosh, J. Computational analysis and identification of an emergent human adenovirus pathogen implicated in a respiratory fatality. *Virology* **2011**, *409*, 141–147. [[CrossRef](#)] [[PubMed](#)]
- Singh, G.; Robinson, C.M.; Dehghan, S.; Schmidt, T.; Seto, D.; Jones, M.S.; Dyer, D.W.; Chodosh, J. Overreliance on the hexon gene, leading to misclassification of human adenoviruses. *J. Virol.* **2012**, *86*, 4693–4695. [[CrossRef](#)]
- Robinson, C.M.; Zhou, X.; Rajaiya, J.; Yousuf, M.A.; Singh, G.; DeSerres, J.J.; Walsh, M.P.; Wong, S.; Seto, D.; Dyer, D.W.; et al. Predicting the next eye pathogen: Analysis of a novel adenovirus. *mBio* **2013**, *4*, e00595-12. [[CrossRef](#)] [[PubMed](#)]
- Singh, G.; Zhou, X.; Lee, J.Y.; Yousuf, M.A.; Ramke, M.; Ismail, A.M.; Lee, J.S.; Robinson, C.M.; Seto, D.; Dyer, D.W.; et al. Recombination of the epsilon determinant and corneal tropism: Human adenovirus species D types 15, 29, 56, and 69. *Virology* **2015**, *485*, 452–459. [[CrossRef](#)] [[PubMed](#)]
- 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.; et al. Molecular evolution of human adenoviruses. *Sci. Rep.* **2013**, *3*, 1812. [[CrossRef](#)] [[PubMed](#)]
- Singh, G.; Robinson, C.M.; Dehghan, S.; Jones, M.S.; Dyer, D.W.; Seto, D.; Chodosh, J. Homologous recombination in E3 genes of human adenovirus species D. *J. Virol.* **2013**, *87*, 12481–12488. [[CrossRef](#)] [[PubMed](#)]
- Dehghan, S.; Seto, J.; Liu, E.B.; Walsh, M.P.; Dyer, D.W.; Chodosh, J.; Seto, D. Computational analysis of four human adenovirus type 4 genomes reveals molecular evolution through two interspecies recombination events. *Virology* **2013**, *443*, 197–207. [[CrossRef](#)] [[PubMed](#)]
- Walsh, M.P.; Chintakuntlawar, A.; Robinson, C.M.; Madisch, I.; Harrach, B.; Hudson, N.R.; Schnurr, D.; Heim, A.; Chodosh, J.; Seto, D.; 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. [[CrossRef](#)]
- Walsh, M.P.; Seto, J.; Jones, M.S.; Chodosh, J.; Xu, W.; Seto, D. Computational analysis identifies human adenovirus type 55 as a re-emergent acute respiratory disease pathogen. *J. Clin. Microbiol.* **2010**, *48*, 991–993. [[CrossRef](#)]
- Seto, D.; Chodosh, J.; Brister, J.R.; Jones, M.S.; Members of the Adenovirus Research, C. Using the whole-genome sequence to characterize and name human adenoviruses. *J. Virol.* **2011**, *85*, 5701–5702. [[CrossRef](#)]
- Nevins, J.R.; Ginsberg, H.S.; Blanchard, J.M.; Wilson, M.C.; Darnell, J.E., Jr. Regulation of the primary expression of the early adenovirus transcription units. *J. Virol.* **1979**, *32*, 727–733. [[PubMed](#)]
- White, E. Regulation of the cell cycle and apoptosis by the oncogenes of adenovirus. *Oncogene* **2001**, *20*, 7836–7846. [[CrossRef](#)] [[PubMed](#)]
- Frisch, S.M.; Mymryk, J.S. Adenovirus-5 E1A: Paradox and paradigm. *Nat. Rev. Mol. Cell Biol.* **2002**, *3*, 441–452. [[CrossRef](#)] [[PubMed](#)]
- Perricaudet, M.; Akusjarvi, G.; Virtanen, A.; Pettersson, U. Structure of two spliced mRNAs from the transforming region of human subgroup C adenoviruses. *Nature* **1979**, *281*, 694–696. [[CrossRef](#)] [[PubMed](#)]
- Kimelman, D.; Miller, J.S.; Porter, D.; Roberts, B.E. E1a regions of the human adenoviruses and of the highly oncogenic simian adenovirus 7 are closely related. *J. Virol.* **1985**, *53*, 399–409. [[PubMed](#)]
- Avvakumov, N.; Wheeler, R.; D'Halluin, J.C.; Mymryk, J.S. Comparative sequence analysis of the largest E1A proteins of human and simian adenoviruses. *J. Virol.* **2002**, *76*, 7968–7975. [[CrossRef](#)]

17. Avvakumov, N.; Kajon, A.E.; Hoeben, R.C.; Mymryk, J.S. Comprehensive sequence analysis of the E1A proteins of human and simian adenoviruses. *Virology* **2004**, *329*, 477–492. [[CrossRef](#)]
18. Berk, A.J.; Lee, F.; Harrison, T.; Williams, J.; Sharp, P.A. Pre-early adenovirus 5 gene product regulates synthesis of early viral messenger RNAs. *Cell* **1979**, *17*, 935–944. [[CrossRef](#)]
19. Jones, N.; Shenk, T. An adenovirus type 5 early gene function regulates expression of other early viral genes. *Proc. Natl. Acad. Sci. USA* **1979**, *76*, 3665–3669. [[CrossRef](#)]
20. Ablack, J.N.; Pelka, P.; Yousef, A.F.; Turnell, A.S.; Grand, R.J.; Mymryk, J.S. Comparison of E1A CR3-dependent transcriptional activation across six different human adenovirus subgroups. *J. Virol.* **2010**, *84*, 12771–12781. [[CrossRef](#)]
21. Boyer, T.G.; Berk, A.J. Functional interaction of adenovirus E1A with holo-TFIID. *Genes Dev.* **1993**, *7*, 1810–1823. [[CrossRef](#)] [[PubMed](#)]
22. Geisberg, J.V.; Lee, W.S.; Berk, A.J.; Ricciardi, R.P. The zinc finger region of the adenovirus E1A transactivating domain complexes with the TATA box binding protein. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 2488–2492. [[CrossRef](#)] [[PubMed](#)]
23. Boyer, T.G.; Martin, M.E.; Lees, E.; Ricciardi, R.P.; Berk, A.J. Mammalian Srb/Mediator complex is targeted by adenovirus E1A protein. *Nature* **1999**, *399*, 276–279. [[CrossRef](#)] [[PubMed](#)]
24. Vijayalingam, S.; Chinnadurai, G. Adenovirus L-E1A activates transcription through mediator complex-dependent recruitment of the super elongation complex. *J. Virol.* **2013**, *87*, 3425–3434. [[CrossRef](#)] [[PubMed](#)]
25. Berk, A.J.; Boyer, T.G.; Kapanidis, A.N.; Ebright, R.H.; Kobayashi, N.N.; Horn, P.J.; Sullivan, S.M.; Koop, R.; Surby, M.A.; Triezenberg, S.J. Mechanisms of viral activators. *Cold Spring Harb. Symp. Quant. Biol.* **1998**, *63*, 243–252. [[CrossRef](#)] [[PubMed](#)]
26. Geisberg, J.V.; Chen, J.L.; Ricciardi, R.P. Subregions of the adenovirus E1A transactivation domain target multiple components of the TFIID complex. *Mol. Cell Biol.* **1995**, *15*, 6283–6290. [[CrossRef](#)] [[PubMed](#)]
27. Mazzarelli, J.M.; Mengus, G.; Davidson, I.; Ricciardi, R.P. The transactivation domain of adenovirus E1A interacts with the C terminus of human TAF(II)135. *J. Virol.* **1997**, *71*, 7978–7983.
28. Scholer, H.R.; Ciesiolka, T.; Gruss, P. A nexus between Oct-4 and E1A: Implications for gene regulation in embryonic stem cells. *Cell* **1991**, *66*, 291–304. [[CrossRef](#)]
29. Liu, F.; Green, M.R. A specific member of the ATF transcription factor family can mediate transcription activation by the adenovirus E1a protein. *Cell* **1990**, *61*, 1217–1224. [[CrossRef](#)]
30. Pelka, P.; Ablack, J.N.; Torchia, J.; Turnell, A.S.; Grand, R.J.; Mymryk, J.S. Transcriptional control by adenovirus E1A conserved region 3 via p300/CBP. *Nucleic Acids Res.* **2009**, *37*, 1095–1106. [[CrossRef](#)]
31. Arany, Z.; Newsome, D.; Oldread, E.; Livingston, D.M.; Eckner, R. A family of transcriptional adaptor proteins targeted by the E1A oncoprotein. *Nature* **1995**, *374*, 81–84. [[CrossRef](#)] [[PubMed](#)]
32. Eckner, R.; Ewen, M.E.; Newsome, D.; Gerdes, M.; DeCaprio, J.A.; Lawrence, J.B.; Livingston, D.M. Molecular cloning and functional analysis of the adenovirus E1A-associated 300-kD protein (p300) reveals a protein with properties of a transcriptional adaptor. *Genes Dev.* **1994**, *8*, 869–884. [[CrossRef](#)] [[PubMed](#)]
33. Liu, X.; Marmorstein, R. Structure of the retinoblastoma protein bound to adenovirus E1A reveals the molecular basis for viral oncoprotein inactivation of a tumor suppressor. *Genes Dev.* **2007**, *21*, 2711–2716. [[CrossRef](#)] [[PubMed](#)]
34. Ferreon, J.C.; Martinez-Yamout, M.A.; Dyson, H.J.; Wright, P.E. Structural basis for subversion of cellular control mechanisms by the adenoviral E1A oncoprotein. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 13260–13265. [[CrossRef](#)] [[PubMed](#)]
35. Dyson, N.; Guida, P.; McCall, C.; Harlow, E. Adenovirus E1A makes two distinct contacts with the retinoblastoma protein. *J. Virol.* **1992**, *66*, 4606–4611. [[PubMed](#)]
36. Berk, A.J. Recent lessons in gene expression, cell cycle control, and cell biology from adenovirus. *Oncogene* **2005**, *24*, 7673–7685. [[CrossRef](#)] [[PubMed](#)]
37. Schrier, P.I.; Bernards, R.; Vaessen, R.T.; Houweling, A.; van der Eb, A.J. Expression of class I major histocompatibility antigens switched off by highly oncogenic adenovirus 12 in transformed rat cells. *Nature* **1983**, *305*, 771–775. [[CrossRef](#)]

38. Bernardis, R.; Schrier, P.I.; Houweling, A.; Bos, J.L.; van der Eb, A.J.; Zijlstra, M.; Melief, C.J. Tumorigenicity of cells transformed by adenovirus type 12 by evasion of T-cell immunity. *Nature* **1983**, *305*, 776–779. [[CrossRef](#)]
39. Gallimore, P.H.; Turnell, A.S. Adenovirus E1A: Remodelling the host cell, a life or death experience. *Oncogene* **2001**, *20*, 7824–7835. [[CrossRef](#)]
40. Obenauer, J.C.; Denson, J.; Mehta, P.K.; Su, X.; Mukatira, S.; Finkelstein, D.B.; Xu, X.; Wang, J.; Ma, J.; Fan, Y.; et al. Large-scale sequence analysis of avian influenza isolates. *Science* **2006**, *311*, 1576–1580. [[CrossRef](#)]
41. Martinez-Martin, N.; Ramani, S.R.; Hackney, J.A.; Tom, I.; Wranik, B.J.; Chan, M.; Wu, J.; Paluch, M.T.; Takeda, K.; Hass, P.E.; et al. The extracellular interactome of the human adenovirus family reveals diverse strategies for immunomodulation. *Nat. Commun.* **2016**, *7*, 11473. [[CrossRef](#)] [[PubMed](#)]
42. Katoh, K.; Misawa, K.; Kuma, K.; Miyata, T. MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **2002**, *30*, 3059–3066. [[CrossRef](#)] [[PubMed](#)]
43. Tamura, K.; Stecher, G.; Peterson, D.; Filipiński, A.; Kumar, S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* **2013**, *30*, 2725–2729. [[CrossRef](#)] [[PubMed](#)]
44. Guindon, S.; Dufayard, J.F.; Lefort, V.; Anisimova, M.; Hordijk, W.; Gascuel, O. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst. Biol.* **2010**, *59*, 307–321. [[CrossRef](#)] [[PubMed](#)]
45. Bielawski, J.P.; Yang, Z. A maximum likelihood method for detecting functional divergence at individual codon sites, with application to gene family evolution. *J. Mol. Evol.* **2004**, *59*, 121–132. [[CrossRef](#)] [[PubMed](#)]
46. Yang, Z. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **2007**, *24*, 1586–1591. [[CrossRef](#)] [[PubMed](#)]
47. Weadick, C.J.; Chang, B.S. An improved likelihood ratio test for detecting site-specific functional divergence among clades of protein-coding genes. *Mol. Biol. Evol.* **2012**, *29*, 1297–1300. [[CrossRef](#)]
48. Xu, D.; Zhang, Y. Ab initio protein structure assembly using continuous structure fragments and optimized knowledge-based force field. *Proteins* **2012**, *80*, 1715–1735. [[CrossRef](#)]
49. Cilia, E.; Pancsa, R.; Tompa, P.; Lenaerts, T.; Vranken, W.F. From protein sequence to dynamics and disorder with DynaMine. *Nat. Commun.* **2013**, *4*, 2741. [[CrossRef](#)]
50. Librado, P.; Rozas, J. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **2009**, *25*, 1451–1452. [[CrossRef](#)]
51. Team, R.C. A Language and Environment for Statistical Computing. 2013. Available online: <https://www.r-project.org/> (accessed on 3 February 2019).
52. Le, S.; Josse, J.; Husson, F. FactoMineR: An R package for multivariate analysis. *J. Stat. Softw.* **2008**, *25*, 1–18. [[CrossRef](#)]
53. Robinson, C.M.; Rajaiya, J.; Walsh, M.P.; Seto, D.; Dyer, D.W.; Jones, M.S.; Chodosh, J. Computational analysis of human adenovirus type 22 provides evidence for recombination among species D human adenoviruses in the penton base gene. *J. Virol.* **2009**, *83*, 8980–8985. [[CrossRef](#)] [[PubMed](#)]
54. Pelka, P.; Ablack, J.N.; Fonseca, G.J.; Yousef, A.F.; Mymryk, J.S. Intrinsic structural disorder in adenovirus E1A: A viral molecular hub linking multiple diverse processes. *J. Virol.* **2008**, *82*, 7252–7263. [[CrossRef](#)] [[PubMed](#)]
55. Culp, J.S.; Webster, L.C.; Friedman, D.J.; Smith, C.L.; Huang, W.J.; Wu, F.Y.; Rosenberg, M.; Ricciardi, R.P. The 289-amino acid E1A protein of adenovirus binds zinc in a region that is important for trans-activation. *Proc. Natl. Acad. Sci. USA* **1988**, *85*, 6450–6454. [[CrossRef](#)] [[PubMed](#)]
56. Kajan, G.L.; Kajan, A.E.; Pinto, A.C.; Bartha, D.; Arnberg, N. The complete genome sequence of human adenovirus 84, a highly recombinant new Human mastadenovirus D type with a unique fiber gene. *Virus Res.* **2017**, *242*, 79–84. [[CrossRef](#)] [[PubMed](#)]
57. Ismail, A.M.; Lee, J.S.; Dyer, D.W.; Seto, D.; Rajaiya, J.; Chodosh, J. Selection Pressure in the Human Adenovirus Fiber Knob Drives Cell Specificity in Epidemic Keratoconjunctivitis. *J. Virol.* **2016**, *90*, 9598–9607. [[CrossRef](#)] [[PubMed](#)]
58. Robinson, C.M.; Seto, D.; Jones, M.S.; Dyer, D.W.; Chodosh, J. Molecular evolution of human species D adenoviruses. *Infect. Genet. Evol.* **2011**, *11*, 1208–1217. [[CrossRef](#)]
59. Turnell, A.S.; Mymryk, J.S. Roles for the coactivators CBP and p300 and the APC/C E3 ubiquitin ligase in E1A-dependent cell transformation. *Br. J. Cancer* **2006**, *95*, 555–560. [[CrossRef](#)]

60. Wevers, D.; Metzger, S.; Babweteera, F.; Bieberbach, M.; Boesch, C.; Cameron, K.; Couacy-Hymann, E.; Cranfield, M.; Gray, M.; Harris, L.A.; et al. Novel adenoviruses in wild primates: A high level of genetic diversity and evidence of zoonotic transmissions. *J. Virol.* **2011**, *85*, 10774–10784. [[CrossRef](#)]
61. Chen, E.C.; Yagi, S.; Kelly, K.R.; Mendoza, S.P.; Tarara, R.P.; Canfield, D.R.; Maninger, N.; Rosenthal, A.; Spinner, A.; Bales, K.L.; et al. Cross-species transmission of a novel adenovirus associated with a fulminant pneumonia outbreak in a new world monkey colony. *PLoS Pathog.* **2011**, *7*, e1002155. [[CrossRef](#)]
62. Dehghan, S.; Seto, J.; Jones, M.S.; Dyer, D.W.; Chodosh, J.; Seto, D. Simian adenovirus type 35 has a recombinant genome comprising human and simian adenovirus sequences, which predicts its potential emergence as a human respiratory pathogen. *Virology* **2013**, *447*, 265–273. [[CrossRef](#)] [[PubMed](#)]
63. Purkayastha, A.; Ditty, S.E.; Su, J.; McGraw, J.; Hadfield, T.L.; Tibbetts, C.; Seto, D. Genomic and bioinformatics analysis of HAdV-4, a human adenovirus causing acute respiratory disease: Implications for gene therapy and vaccine vector development. *J. Virol.* **2005**, *79*, 2559–2572. [[CrossRef](#)] [[PubMed](#)]
64. Gersbach, C.A.; Gaj, T.; Barbas, C.F., 3rd. Synthetic zinc finger proteins: The advent of targeted gene regulation and genome modification technologies. *Acc. Chem. Res.* **2014**, *47*, 2309–2318. [[CrossRef](#)] [[PubMed](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).