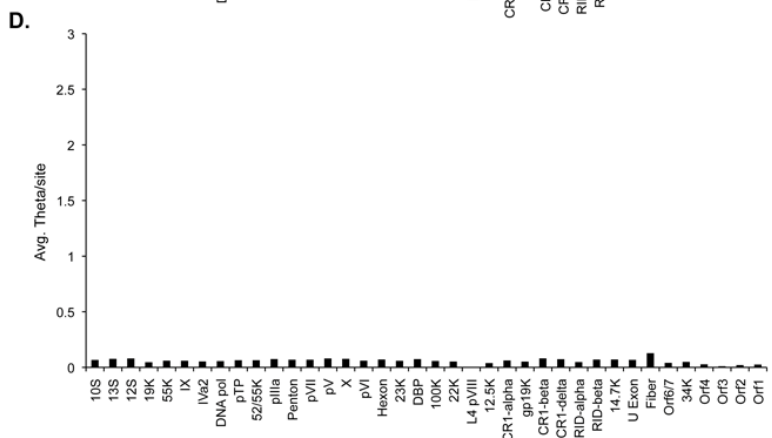
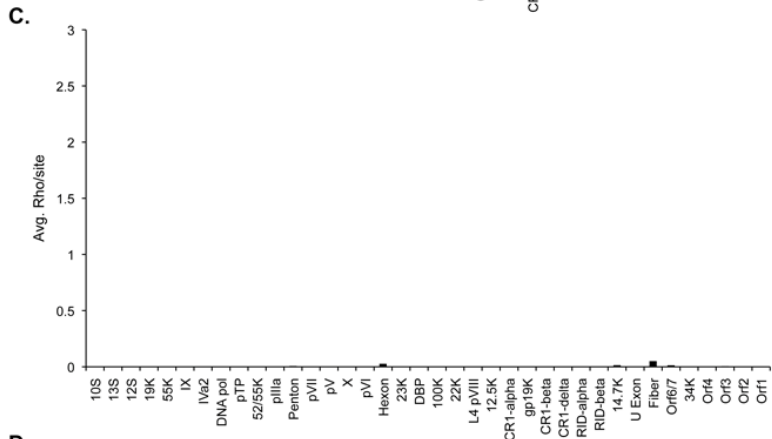
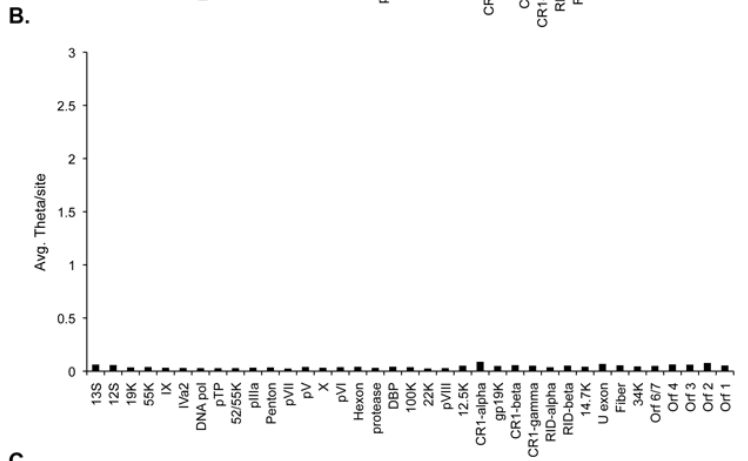
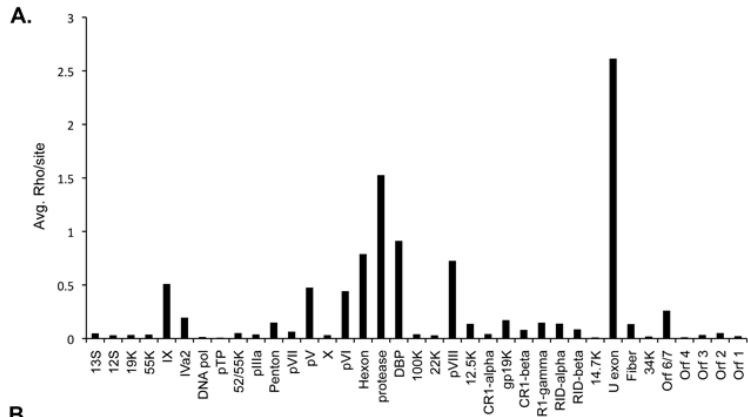


Supplementary information for:

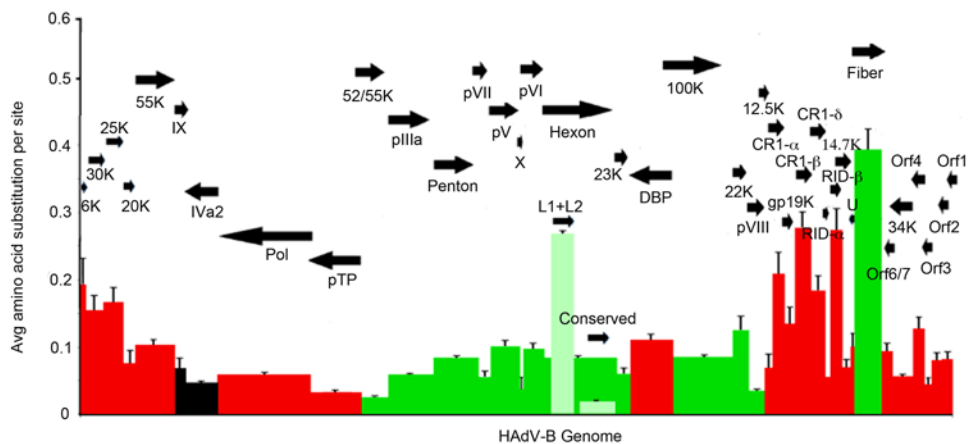
**Molecular evolution of human adenoviruses**

Christopher M. Robinson, Gurdeep Singh, Jeong Yoon Lee, Shoaleh Dehghan, Jaya Rajaiya, Elizabeth B. Liu, Mohammad A Yousuf, Rebecca A. Betensky, Morris S. Jones, David W. Dyer, Donald Seto, James Chodosh

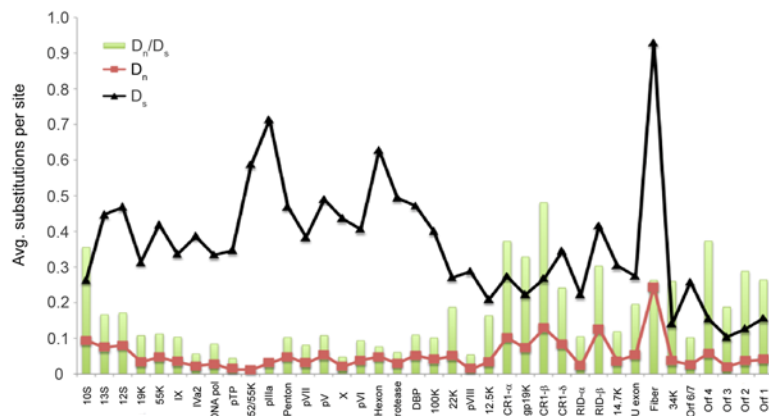


**Supplemental Fig. 1.** Recombination and mutation analysis. (A) Average rho (recombination) rate per site for each HAdV-D gene. (B) Average theta (mutation) rate per site for each HAdV-D gene. (C) Average rho rate per site for each HAdV-B gene. (D) Average theta rate per site for each HAdV-B gene.

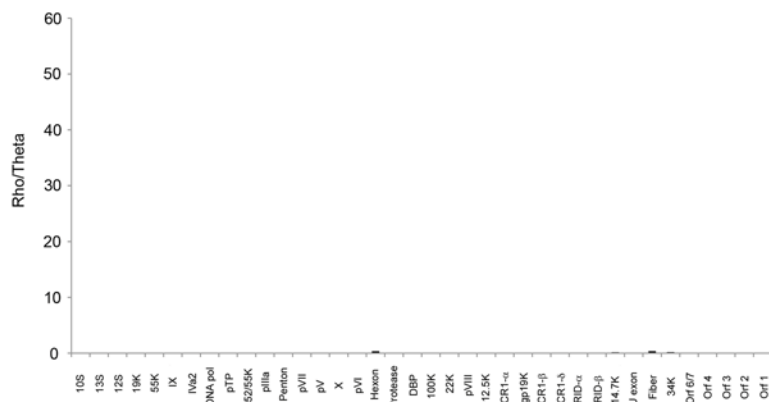
A.



B.

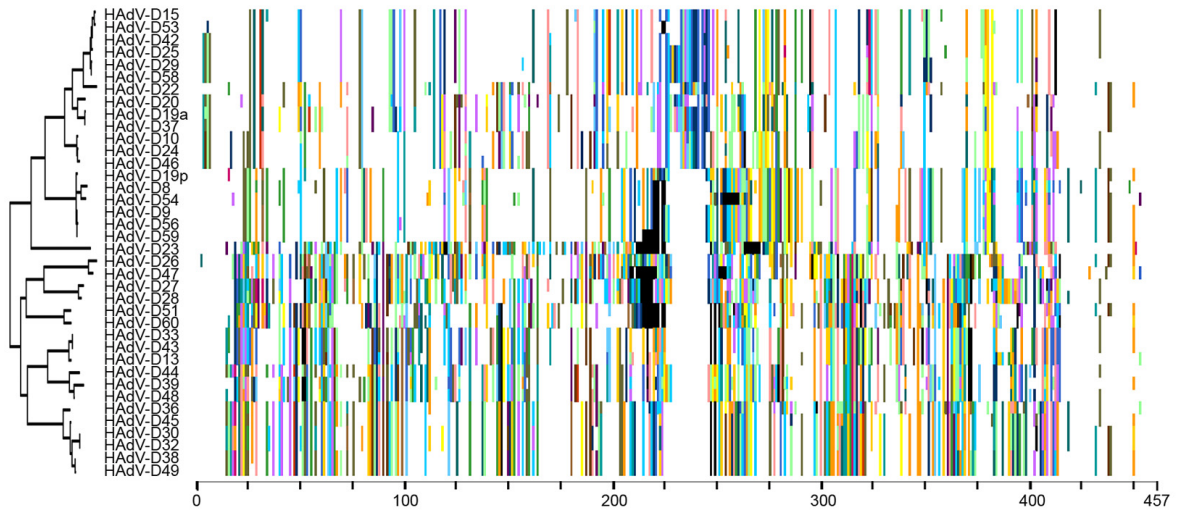


C.

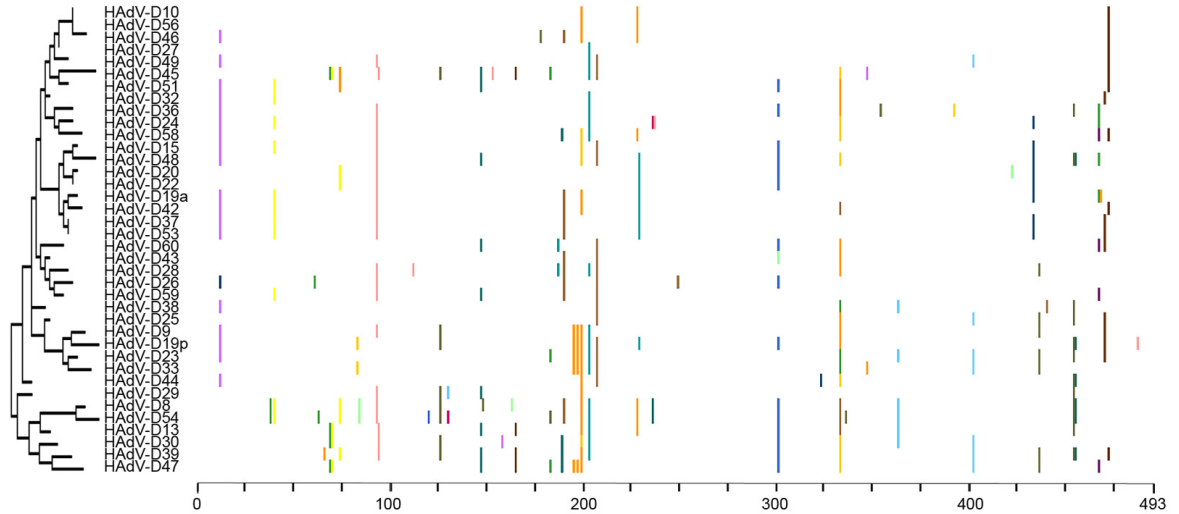


**Supplemental Fig. 2.** HAdV-B Evolution. (A) Amino acid diversity calculated in MEGA 4.02, measuring the average amino acid substitution for each HAdV-B protein. Each bar in the graph corresponds to a protein as represented by arrows. Red = early. Light green is hypervariable region of hexon. Dark green=late genes. Black= intermediate genes. (B) Analysis of synonymous and non-synonymous mutations across the HAdV-B genome calculated using MEGA software. Synonymous ( $D_s$ ) and non-synonymous ( $D_n$ ) changes are represented in black and red lines, respectively. Green bars represent the ratio ( $D_n/D_s$ ) for each gene. (C) Analysis of the rho (recombination) and theta (mutation) ratio as determined by RDP3 software for each gene in the HAdV-B genome.

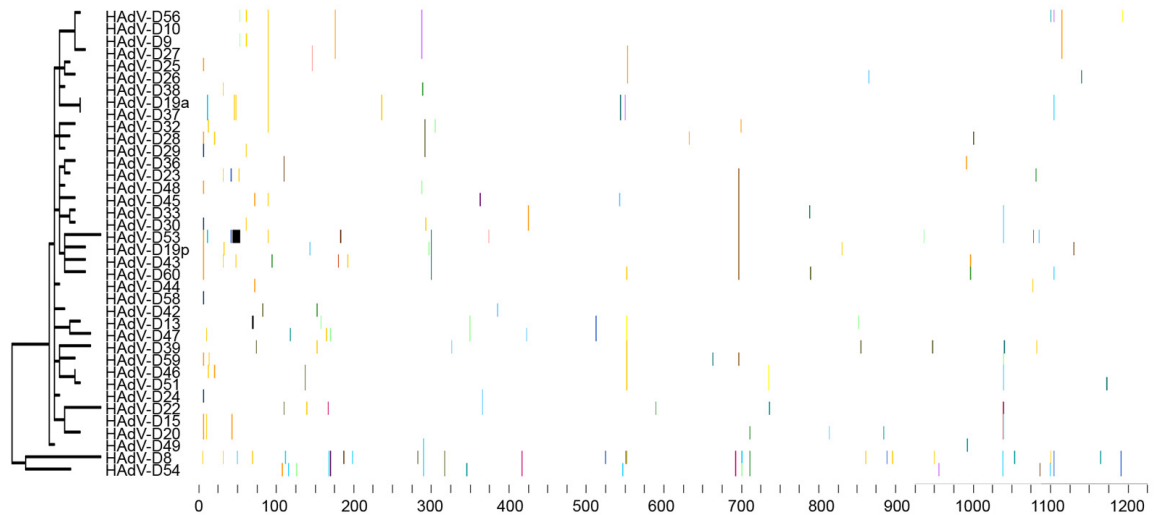
### A. E3 CR1- $\beta$



### B. DNA Binding Protein



### C. DNA Polymerase



**Supplemental Fig 3.** Proteotyping assignments. Neighbor-joining phylogenetic trees are shown on the left for each protein. The amino acid signatures are shown to the right. Each amino acid that was variable from consensus sequence was assigned a color. White regions represent consensus amino acids at each position. (A) HAdV-D CR1- $\beta$ , (B) HAdV-D DNA binding protein, (C) HAdV-D DNA polymerase.

Supplemental Table 1. Human adenoviruses sequenced in this study.

Type	Genbank Accession #	Country	Year	454 coverage	N50	% GC
HAdV-D10	JN226746	USA	1956	168	8,511	57.1
HAdV-D13	JN226747	USA	1955	*	*	57.3
HAdV-D15	JN226748	Saudi Arabia	1955	102	34,403	57.0
HAdV-D20	JN226749	Saudi Arabia	1956	103	27,633	56.6
HAdV-D23	JN226750	Saudi Arabia	1956	189	12,916	56.9
HAdV-D24	JN226751	Arabia	1956	149	34,831	56.7
HAdV-D25	JN226752	USA	1956	72	12,149	57.0
HAdV-D27	JN226753	USA	1958	53	7,815	56.9
HAdV-D29	JN226754	USA	1959	173	28,280	57.2
HAdV-D30	JN226755	USA	1959	107	15,396	57.1
HAdV-D32	JN226756	USA	1965	*	*	56.9
HAdV-D33	JN226758	USA	1965	100	34,416	57.0
HAdV-D38	JN226759	Netherlands	1984	94	34,448	56.9
HAdV-D39	JN226760	El Salvador	1977	*	*	57.3
HAdV-D42	JN226761	France	1982	91	34,315	56.9
HAdV-D43	JN226762	USA	1982	128	23,826	57.1
HAdV-D44	JN226763	USA	1982	100	34,327	57.1
HAdV-D45	JN226764	USA	1982	121	34,390	57.0
HAdV-D47	JN226757	USA	1982	128	32,897	57.3
HAdV-D51	JN226765	Amsterdam	1989	186	34,380	57.1

\* Sequenced on a ABI 3730XL DNA sequencer to over 8 fold coverage



Supplemental Table 2. Whole Genome Rho/Theta Analysis

HAdV species	rho/gene	theta/gene	rho/theta	Total number of sites (excluding gaps)
HAdV-A	665	5575	0.119282511	33530
HAdV-B	6.9	3381.725	0.002040379	33602
HAdV-C	18.2	1357.762	0.013404411	35484
HAdV-D	86.9	1738.634	0.049981767	32826
HAdV-E	22.3	2706.393	0.008239749	33654
HAdV-F	N/A	N/A	N/A	N/A
HAdV-G	35.1	2292	0.015314136	29395

N/A: Unable to analyze due to number of HAdV types

Supplemental Table 3. Shared/Unshared Proteotypes

Types	Penton (HVL1)	Penton (HVL2)	Hexon	Fiber	CR1- $\beta$
HAdV-D8					
HAdV-D9					
HAdV-D10					
HAdV-D13					
HAdV-D15					
HAdV-D19a					
HAdV-D19p					
HAdV-D20					
HAdV-D22					
HAdV-D23					
HAdV-D24					
HAdV-D25					
HAdV-D26					
HAdV-D27					
HAdV-D28					
HAdV-D29					
HAdV-D30					
HAdV-D32					
HAdV-D33					
HAdV-D36					
HAdV-D37					
HAdV-D38					
HAdV-D39					
HAdV-D42					
HAdV-D43					
HAdV-D44					
HAdV-D45					
HAdV-D46					
HAdV-D47					
HAdV-D48					
HAdV-D49					
HAdV-D51					
HAdV-D53					
HAdV-D54					
HAdV-D56					
HAdV-D58					
HAdV-D59					
HAdV-D60					
Total no. of proteotypes	14	10	28	22	14
No. unshared	5	3	20	12	4
No. shared	9	7	8	10	10
Recombinants	$\geq 24$	$\geq 28$	$\geq 10$	$\geq 16$	$\geq 24$

Gray color denotes shared proteotypes; white denotes unshared (unique) proteotypes.

Supplemental Table 4. Location of homologous GC/AT transition zones

	Median Position	Sequence	Proportion
Penton base	13801-13830	CCGAGGCCAGCACGC/ AGACGATAAATTTTG	38/38
	14341-14370	CCCGCCCTACTGGAT/ ACCAAAAAATATCTG	14/38
	14304-14348	CCCGCCCTCCTTGAT/ GTGCCCAAGTACTTG/ GAAAGCAAGAAGAAA	24/38
	15031-15075	CCGCGCGTGCTTTCC/ AGTCGCACCTTCTAA/ AAAAATGTCTATTCT	38/38
Hexon	18179-18223	CCCCAAGGGCGCCCC/ CAATCCCAGTCAGTG/ GACTACCAAAGAAAA	38/38
	19426-19470	GCTACGTGCCCTTCC/ ACATCCAAGTGCCCC/ AAAAGTTCTTTGCCA	38/38
	21193-21237	CCGCGCCGCCATCGA/ ACACGCCACCGCTTT/ TGATAAAATGAAACA	38/38
CR1-β	27379-27423	TGCCCCCTACCAAGG/ AGAACATGGTGGGTT/ TTTCTTTGGCTTTTG	38/38
	27486-27530	GCAAGCCCAGGTACG/ GAAATGAAGAGAAGG/ AAAAATTGCTATAAA	38/38
	28711-28755	GGCAATCGTGGTGGG/ TGTGATTGCGGGCTT/ CATAACTCTGATCAT	38/38
	28774-28815	CTGCTGCCGCAAGCG/ TCCCAGGGCATA/ CAATCATATGGTAGA	38/38
Fiber	30748-30777	CGCGACCTCCTCCCC/ ATGAACTGATGTTGA	38/38
	30955-30999	CCCCGGCGGGCGGCG/ AACTTCCTCCACACC/ TTGAAAGGGATGTCA	38/38
	32131-32175	GGCGGGCGGCGAACT/ TCCTCCACACCTTGA/ AAGGGATGTCAAATT	38/38
	32221-32250	CAGCACGGGTGGCCT/ TAAATAGGGGAATGT	38/38