

COMPUTATIONAL ANALYSIS OF HUMAN ADENOVIRUS EVOLUTION AND
DEVELOPMENT OF BIOINFORMATICS TOOLS

by

Elizabeth Liu
A Dissertation
Submitted to the
Graduate Faculty
of
George Mason University
in Partial Fulfillment of
The Requirements for the Degree
of
Doctor of Philosophy
Bioinformatics and Computational Biology

Committee:

_____	Dr. Donald Seto, Dissertation Director
_____	Dr. Jason Kinser, Committee Member
_____	Dr. Dmitri Klimov, Committee Member
_____	Dr. Andrea Weeks, Committee Member
_____	Dr. James D. Willett, Director, School of Systems Biology
_____	Dr. Donna M. Fox, Associate Dean, Office of Student Affairs & Special Programs, College of Science
_____	Dr. Peggy Agouris, Dean, College of Science
Date: _____	Fall Semester 2015 George Mason University Fairfax, VA

Computational Analysis of Human Adenovirus Evolution and Development of
Bioinformatics Tools

A dissertation submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy at George Mason University

By

Elizabeth B. Liu
Master of Science
George Mason University, 2006
Bachelor of Science
Virginia Polytechnic Institute & State University, 2005

Director: Donald Seto, Professor
School of System Biology

Fall Semester 2015
George Mason University
Fairfax, VA

Copyright 2015 Elizabeth B. Liu
All Rights Reserved

DEDICATION

This is dedicated to my loving parents Pingping Fan & Weiguo Ge and to my grandparents QinghuaYang & Rensheng Fan, who have supported me through all the years of pursuing my degrees.

ACKNOWLEDGEMENTS

I would like to thank the many friends, relatives, and supporters who have made this happen. My loving parents and grandparents supported me throughout the years. Dr. Donald Seto, and the members of my committee Dr. Jason Kinser, Dr. Dmitri Klimov and Dr. Andrea Weeks for their invaluable help and input. I would also like to thank members of Dr. Seto's lab (Shoaleh Dehghan, Jason Seto and Michael Walsh) for their valuable support and help.

For the HAdV16 analysis, I would like to thank Dr. Clark Tibbetts, James M. Clark and Dr. Anjan Purkayastha for initial discussions; and Drs. David Metzgar, Morris Jones and James Chodosh for continuing discussions. During the collection of this genome data set and preliminary analyses, DS (2002-2004), AP (2003-2005), JMC (2004) and CT (2001-2005) were affiliated with the HQ USAF Surgeon General Office, Directorate of Modernization (SGR) and the Epidemic Outbreak Surveillance (EOS) Program, 5201 Leesburg Pike, Suite 1401, Falls Church, VA 22041. Portions of this work were funded, during these time periods, by a grant from the U.S. Army Medical Research and Material Command (USAMRMC) (DAMD17-03-2-0089) and additional support was through the EOS Project, funded by HQ USAF Surgeon General Office, Directorate of Modernization (SGR) and the Defense Threat Reduction Agency (DTRA). Therefore, it is noted that the opinions and assertions contained herein are the private ones of the authors and are not to be construed as official or reflecting the views of the U.S. Department of Defense. I would also like to thank George Mason University's Provost for their Ph.D. thesis completion grant.

TABLE OF CONTENTS

	Page
List of Tables.....	vii
List of Figures.....	viii
List of Appendices.....	ix
Abstract.....	x
Chapter 1 - Introduction.....	1
Chapter 2 - Computational analysis of Human Adenovirus Type 16.....	6
Abstract.....	6
Introduction.....	6
Material & methods	9
Results.....	9
Genome sequence analysis	9
Comparative genome analysis	11
Hexon recombination.....	14
Phylogeny analysis.....	17
Discussion	21
Recombination, molecular evolution and new serotype	21
Conclusion	24
Chapter 3 - Computational analysis of Human Adenovirus Type 21	26
Abstract.....	26
Introduction.....	26
Material & methods	27
Results.....	28
Genome sequence analysis	28
Comparative whole genome analysis.....	28
Hexon recombination.....	31
Phylogeny analysis.....	35
Discussion	38
Conclusion	38
Chapter 4 - Human Adenovirus Type 58.....	40
Introduction.....	40
Chapter 5 - Human Adenovirus Type 59.....	52
Introduction.....	52
Chapter 6 - Bioinformatics Tools Development to Enhance Viral Genome Analysis ..	62
Introduction.....	62
Discussion	64

DrawBar	64
GeneMap	68
Human Adenovirus Working Group Website.....	73
Chapter 7- Future Directions	75
Appendices.....	77
References	92

\

LIST OF TABLES

Table	Page
Table 1 – Taxonomy of Human Adenovirus	4
Table 2 – Percent Identity of HAdV-B16	13
Table 3 – Percent Identity of HAdV-B21	31

LIST OF FIGURES

Figure	Page
Figure 1. Structure of Human Adenovirus.....	3
Figure 2. HAdV-B16 whole genome mapping	10
Figure 3. zPicture analysis of HAdV-B16	12
Figure 4. Whole genome recombination analysis HAdV-B16	15
Figure 5. Hexon recombination analysis of HAdV-B16	16
Figure 6. Whole genome phylogenetic analysis of HAdV-B16	18
Figure 7. Hexon genome phylogenetic analysis of HAdV-B16	19
Figure 8. Hexon halves phylogenetic analysis of HAdV-B16.....	20
Figure 9. Fiber knobs phylogenetic analysis of HAdV-B16.....	21
Figure 10. zPicture analysis of HAdV-B21	30
Figure 11. Whole genome recombination analysis HAdV-B21	33
Figure 12. Hexon recombination analysis of HAdV-B21	34
Figure 13. Whole genome phylogenetic analysis of HAdV-B21	35
Figure 14. Hexon genome phylogenetic analysis of HAdV-B21	36
Figure 15. Hexon halves phylogenetic analysis of HAdV-B21.....	37
Figure 16. Fiber knobs phylogenetic analysis of HAdV-B21.....	38
Figure 17. DrawBar input screen	66
Figure 18. DrawBar result screen	67
Figure 19. GeneMap input screen	70
Figure 20. GeneMap result screen 1	71
Figure 21. GeneMap result screen 2	72
Figure 22. HAdV working group website serotyping tool	74

LIST OF APPENDICES

Appendix	Page
Appendix 1. DrawBar source code	77
Appendix 2. GeneMap source code	82

ABSTRACT

COMPUTATIONAL ANALYSIS OF HUMAN ADENOVIRUS EVOLUTION AND DEVELOPMENT OF BIOINFORMATICS TOOLS

Elizabeth Liu, Ph.D.

George Mason University, 2015

Dissertation Director: Dr. Donald Seto

Human adenoviruses (HAdVs) may be highly contagious and may be human pathogens that can cause a wide range of illnesses including respiratory, gastrointestinal and ocular infections. Individuals with immune deficiency are especially prone to such infections, leading to fatalities. Even though adenoviruses continue to cause concerns due to notable mortality and morbidity in human populations, their existence may also provide a possible benefit for patients with a broader range of illnesses as well. In recent studies, for example, adenoviruses have been used in gene therapy and vaccine vector development. Recently, bioinformatics and genomics are both high-resolution approaches and resources available for studying the adenovirus for such purposes. One observation is that genome recombination is a driving force in the molecular evolution of human adenoviruses. This has implications for their use as vectors, particular across host species. Computational analysis of this event can provide a better understanding of the role recombination plays in the human adenovirus evolution and pathology, which may later

provide for a rational design of vaccines and for gene delivery vector development.

Custom developed bioinformatics tools will also help to facilitate the process of data mining and analysis, and its presentation. In the course of this project, the genomes of three respiratory and gastrointestinal pathogens, HAdV-B16, B21, D58 and D59, were analyzed using bioinformatics tools to understand their origins and evolution as pathogens, in particular, changes in their genomes. To facilitate this, several tools were developed to assist this genome analysis. The analysis of both HAdV-B16 and HAdV-B21, archived 1950s prototypes, provided examples of “then novel and emergent” HAdVs arising as the result of genome recombination events with simian adenoviruses, across host species. Recently emergent HAdV-D58 and HAdV-D59 are novel pathogens that are characterized by genome sequencing and analysis. Their results have shown also that recombination plays an important role in their molecular evolution.

CHAPTER 1 - INTRODUCTION

Introduction

The first two human adenoviruses (HAdVs) were isolated around the same time in 1953: 1) from a child's adenoid tissue as a non-specific human respiratory infectious agent and 2) from an U.S. Army recruit that presented with respiratory disease (Rowe et al., 1953; Hilleman et al., 1954). Since the first isolation of the virus, numerous members of the *Adenoviridae* family have been identified and characterized. It is likely all vertebrates are affected by adenoviruses (AdVs), including but not limited to fish, frogs, snakes, birds, canines, and primates, for example, chimpanzee and human. Human adenoviruses are grouped under the genus Mastadenovirus (mammalian) (Fenner et al., 1993). Within this, there are seven species (A-G) of HAdVs recognized, with more than 70 types that are identified based on immunochemistry, originally, and homologies of nucleic acid sequences and hexon and fiber protein sequences, as well as biological and genomic properties recently (Lion et al., 2014). The different types are associated with different tissue tropism, such as the upper and lower respiratory tracts, urinary and digestive tracts and eyes, allowing for characteristic diseases in these tissues and organs.

AdVs are non-enveloped icosahedral viruses comprising a nucleocapsid and a linear double strand DNA with a genome size average of 30 kb that encodes about 30

proteins (Rowe et al., 1953). The icosahedral capsid contains 12 vertices and seven surface proteins (**Figure 1**). It has a unique spike-like fiber protein, associated with each penton base of the capsid, which is the cell recognition domain, enabling the attachment of the virus to the host cell. The major outer proteins of the capsid are 240 hexon (protein I), 12 penton base at the vertices (protein II), and 12 protruding trimeric fibers (protein IV) attached to each of the penton base. These outer capsid proteins define individual virus types, and include the recognition site for host's immune system interaction. They are also very useful for serotyping e.g., antibody assay, which was the main method for the classification of AdVs. Recently, computational analyses of whole genome sequence and the individual protein sequences have proved to be a faster, more complete and more efficient process for typing AdVs (Jones et al., 2007).

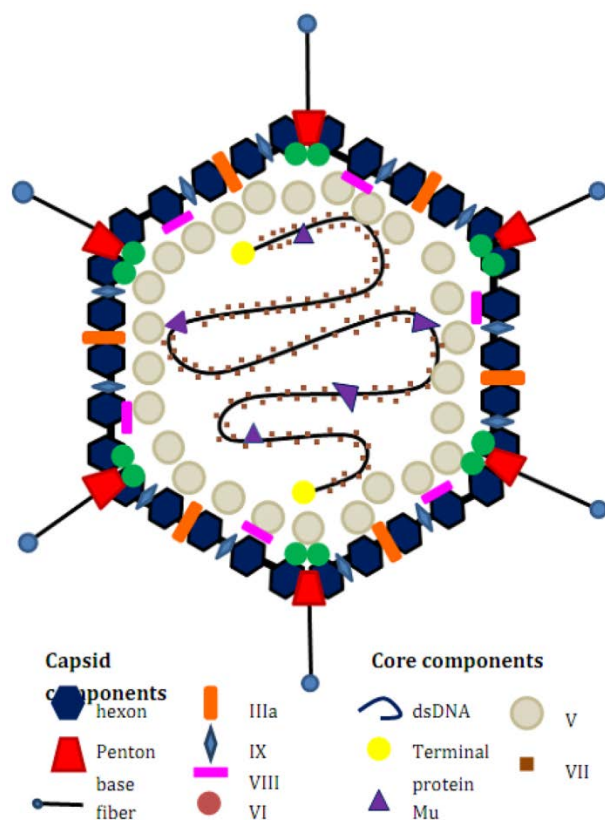


Figure 1 - Structure of a human adenovirus.

The hexon proteins provide structural stability while penton base and fiber proteins are responsible for host recognition and virus penetration. Adapted from Wayne, M.M.Y.; Sing, C.W. Anti-Viral Drugs for Human Adenoviruses. *Pharmaceuticals* **2010**, *3*, 3343-3354.

Known diseases caused by HAdVs include but are not limited to, gastroenteritis, acute febrile pharyngitis, pharyngoconjunctival fever, acute respiratory disease, pneumonia, keratoconjunctivitis, pertussis like syndrome, acute hemorrhagic cystitis, meningoencephalitis and hepatitis (Jones et al., 2007). The HAdV-A species has also been shown to initiate sarcoma development in certain rodents (Ogawa K. 1989). Respiratory diseases are mainly due to species HAdV-B, E, and C. Ocular diseases are caused by HAdV-B, E, and D species. Gastroenteritis is due to the HAdV-F serotypes 40

and 41, although recent studies have shown HAdV-D serotype causes also gastroenteritis (Liu et al., 2011; Liu et al., 2012). HAdV-E only has one serotype, 4; it is primarily responsible for acute respiratory disease but may cause ocular disease at times. HAdV-4 is also one of the two HAdVs that has a vaccine developed against it (Jones et al., 2007), indicating its importance as a respiratory pathogen.

The classification of AdVs is complex. As mentioned earlier, HAdVs are classified under the genus Mastadenovirus. Currently there are more than 70 accepted HAdV types (unpublished observation) based on genomics (HAdV-1 to HAdV-70), which are categorized in seven species (HAdV-A to G) as shown in Table 1.

Table 1 - Taxonomy of Human Adenoviruses.

Using genomics and bioinformatics approaches, all of the recognized types of human adenoviruses are parsed into species. These conform to original observation based on serology, sequence comparisons and biological attributes.

Species	Types
A	12, 18, 31, 61
B	3, 7, 11, 14, 16, 21, 34, 35, 50, 55, 66, 68
-B1	3, 7, 16, 21, 50, 66, 68
-B2	11, 14, 34, 35, 55
C	1, 2, 5, 6, 57
D	8, 9, 10, 13, 15, 17, 19, 20, 22, 23, 24, 25, 26, 27, 28, 29, 30, 32, 33, 36, 37, 38, 39, 42, 43, 44, 45, 46, 47, 48, 49, 51, 53, 54, 56, 58, 59, 60, 62, 63, 64, 65, 67, 69, 70
E	4
F	40, 41
G	52

In the past, HAdV serotyping and species classification were defined by reactivity of the outer coat proteins to discriminating antibodies. Additional biological properties were used as well, e.g., oncogenic potential and hemagglutination properties (Lion et al.,

2014). However, these are lacking since they only probe a small view of the virus, for example the antibody epitope DNA sequence for the hexon protein is approximately 2.6% of the entire genome. Another drawback is that they are also expensive and time consuming to perform. Recent studies using DNA sequencing and bioinformatics methods, including phylogenetic analysis and amino acid sequence analysis, have provided a suitable alternative that is cost-effective, quantitative and much less time-consuming. Thus, bioinformatics has become a preferred and reliable method for demonstrating how those viruses are related through molecular evolution by using primary sequence data (Seto et al., 2009).

Currently, bioinformatics tools are lacking for mining viral genome data. Several computational tools were developed to meet this need and were applied to the study of four human adenovirus pathogens. The main objective of this thesis is to examine the genomes and determine the mechanisms of the molecular evolution of three respiratory and one gastrointestinal human pathogens, HAdV-16, 21, 58 and 59. This is accomplished by studying the changes in their genomes in order to understand the genesis of emergent and novel human viral pathogens.

CHAPTER 2 – COMPUTATIONAL ANALYSIS OF HUMAN ADENOVIRUS Type 16

Abstract

Molecular evolution of human adenoviruses (HAdVs) is driven by recombination. The computational analysis of HAdV-B16, a subspecies B1 member, provides evidence for recombination between subspecies B2 genomes within the second half portion of the hexon gene, previously unreported, and HAdV-E4, species E, within the first half portion, an interspecies recombination event previously not known. As HAdV-B16 is a candidate human gene transfer vector and HAdV-E4 is an important human pathogen, understanding the role recombination plays in the evolution and pathoepidemiology of HAdV has applications in the rational design of vaccines and for gene delivery vector development.

Introduction

Human adenoviruses (HAdVs) have been characterized using available assays since the 1950s (Hilleman et al., 1954; Rowe et al., 1953). As pathogens, they are of interest because they may occur in highly contagious outbreaks infecting 100s (Binn et al., 2007; Engelmann et al., 2006; Ishiko et al., 2008) and can also cause a wide range of

diseases, including respiratory, ocular, gastrointestinal and metabolic (Echavarria et al., 2009). There are 70 different types partitioned into six (A-F) species based on biology, immunochemistry and recombinant DNA methodologies, with genomics recently providing additional prototypes for defining a new species G (Echavarria et al., 2009; Ishiko et al., 2008; Jones et al., 2007; Walsh et al., 2009).

Genomics and bioinformatics are high-resolution approaches and resources that are available for studying HAdVs at the primary nucleotide sequence level. Computational analysis of the genome data is providing insights into the molecular evolution of HAdVs, with the three novel “types” (HAdV-G52, D53 and B55) characterized and christened using sequence analysis (Jones et al., 2007; Walsh et al., 2009a; Walsh Seto et al., 2009) rather than serological characterization. Two of these are the results of genome recombination (Walsh et al., 2009a; Walsh et al., 2009b), providing support for the hypothesis that recombination is a major driving force for novel prototypes (Crawford-Miksza et al., 1996). This was also recently suggested by the limited analyses of sixteen species C field isolates by serology (Lukashev et al., 2008). A re-examination of archived prototypes by genomics provides additional support that novel HAdV arise as the result of recombination. An example is a penton base recombination characterized in HAdV-D22 (Robinson et al., 2009).

This project describes two partial hexon recombinations uncovered during the computational analysis of HAdV-B16. One is as an acceptor and the other is as a donor. The donated sequence was from HAdV-E4 (Dehghan et al., 2013), representing an interspecies event. Previous hypotheses and limited laboratory data indicated

interspecies recombination does not occur for the HAdV genomes (Lukashev et al., 2008; Wadell et al., 1980; Williams et al., 1975).

Species HAdV-B is subdivided into two subspecies; using molecular biological techniques and taking into account their biology as well as proteome differences (Echavarria et al., 2009; Wadell et al., 1984; Wadell et al., 1980). Members of B1 (HAdV-B3, B7, B16, B21 and B50) are human acute respiratory disease (ARD) pathogens, with two exceptions, HAdV-B50 (De Jong et al., 1999) and HAdV-B16. HAdV-B16 was isolated originally from conjunctival scrapings in 1955 and recognized as a new serotype (Bell et al., 1959; Hierholzer et al., 1991, Pereira et al., 1963). It was therefore associated with ocular disease (Bell et al., 1959; Feng et al., 1959) originally, but subsequently recognized as a respiratory pathogen as well, causing pharyngoconjunctival fever (Echavarria et al., 2009), pneumonia (Morgan et al., 1984) and other respiratory disease (Metzgar et al., 2005). HAdV-B16 appears to be either an underreported or uncommonly encountered HAdV (D'Ambrosio et al., 1982; Metzgar et al., 2005; Morgan et al., 1984) as there are few mentions in the literature. Members of subspecies B2, with the exception of HAdV-B14 (Louie et al., 2008; Van Der Veen et al., 1957) and HAdV-B55 (formerly "HAdV-B11a") (Walsh et al., 2010; Yang et al., 2009; Zhu et al., 2009), are not associated with respiratory disease (Echavarria et al., 2009).

Aside from its pathogenicity, HAdV-B16 is of interest as a vector candidate for gene delivery in gene therapy protocols (Skog et al., 2007). It is reported to infect human low-passage brain tumor cells as well as cancer stem cells, unlike HAdV-C5, giving it an important advantage.

Materials and Methods

HAdV-B16 is archived at the American Type Culture Collection (ATCC; Manassas, VA) as “VR-17”, strain ch. 79. This was obtained and processed using a protocol described for similar HAdV sequencing projects (Purkayastha et al., 2005; Seto et al., 2009), with virus growth in A-549 cells and DNA purification outsourced to Virapur, LLC. (San Diego, CA). Commonwealth Biotechnologies, Inc. (Richmond, VA) sequenced the genome, using the Sanger chemistry with the DYEnamic ET Terminator Cycle Sequencing kit (Amersham Biosciences; Piscataway, NJ); sequence ladders were resolved on an ABI Prism 377 DNA Sequencer (Applied Biosystems; Foster City, CA); and assembled it using DNA Sequencer (GeneCodes, Inc.; Ann Arbor, MI). A minimum of three-fold sequences, covering both directions, with overall five-fold coverage, was supplemented with PCR-driven amplification and re-sequencing of areas that were found to be ambiguous upon sequence assembly and genome annotation.

Results

Genome sequence analysis

HAdV-B16 has a genome of 35,522 nucleotides with a GC content of 51%, consistent with subspecies B1 (51%) and differing from species B2 (49%) and other HAdV species; GC% is a species-defining criterion. Its genome is approximately 94% identical to the HAdV-B1 members and 82% to HAdV-B2 members, with lower

identities to other serotypes and species (the next highest is HAdV-E4 at 72%). It contains two VA-RNAs, as well as other genome and proteome attributes, that fit it into subspecies B1 rather than B2 (Wadell et al., 1980; Wadell et al., 1984). Mapping of HAdV-B16 proteins is shown in **Figure 2**.

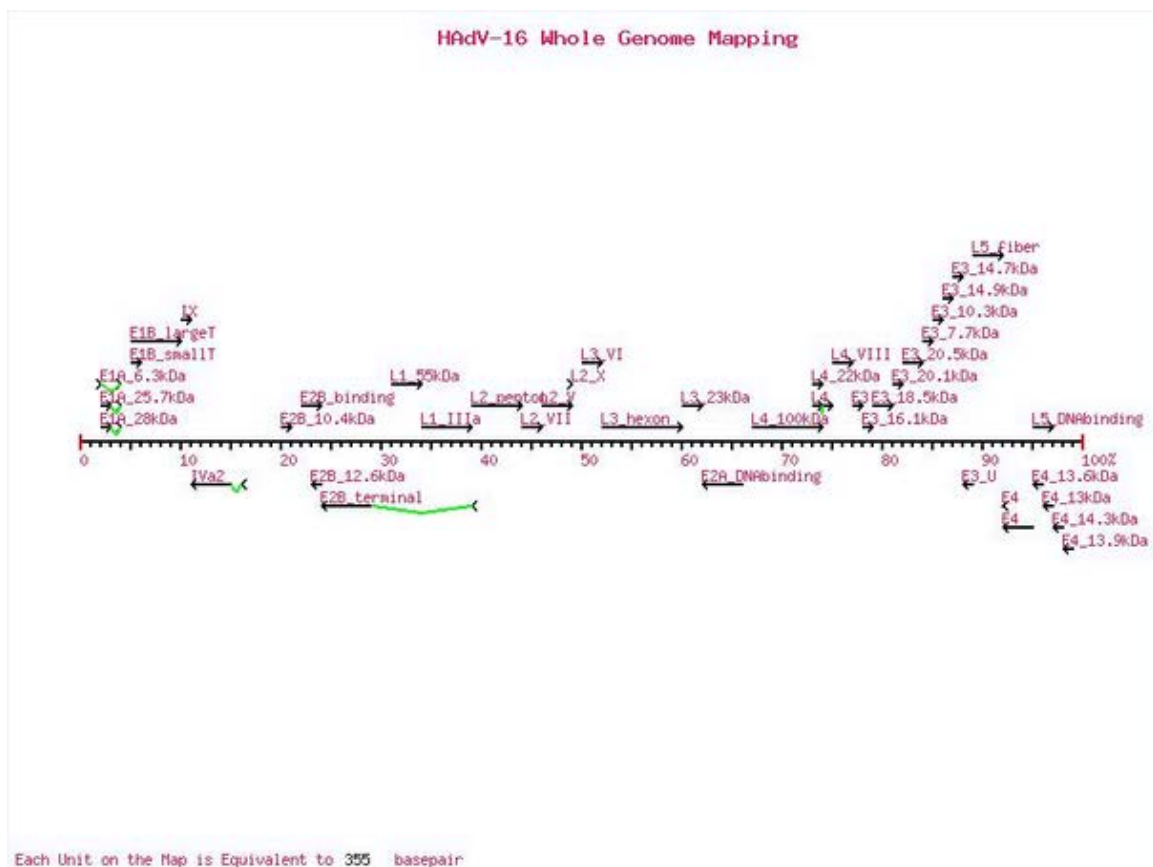


Figure 2 - HAdV-B16 whole genome mapping.
Whole genome protein mapping of HAdV-B16 with GEMEMap software on binf.gmu.edu/eliu1/genemap/

Comparative genome analysis

Initial comparative sequence analysis was performed using zPicture (<http://zpicture.dcode.org/>), a blastz algorithm based dynamic alignment visualization tool, to align the HAdV-B16 hexon genome against each of the eleven sequences from the respiratory pathogens of species E, subspecies of B1 and B2 members. The zPicture genome results showed sequence divergence across the genomes for the B1 genomes, including within the hexon sequence. However, the B2 genomes showed sequence conservation at the distal portion of the hexon gene, with similarity levels higher than the B1 genomes (93% vs. 88%). Interestingly, HAdV-E4 showed a low similarity at the distal portion, 81%, but a much higher level 97%, at the proximal end. The zPicture sequence comparisons of the hexon sequences are shown in **Figure 3**.

Further analysis of HAdV-B16 included using the EMBOSS NEEDLE pairwise sequence alignment tool (EMBL-EBI 2015) to calculate percent identity to each of the eleven genomes to provide a more detailed relationship between each of the major protein genes. The findings are consistent with zPicture results. The hexon sequences showed that the proximal HAdV-B16 sequence that is highly similar (94.6%) to its HAdV-E4 counterpart, comprising approximately 900 nucleotides that represent 31.9% of the HAdV-B16 hexon gene and 2.53% of the whole genome. This is in line with similar hexon recombinants reported in the literature for HAdV-D53 and B55 (Walsh et al., 2009; Walsh et al., 2010).

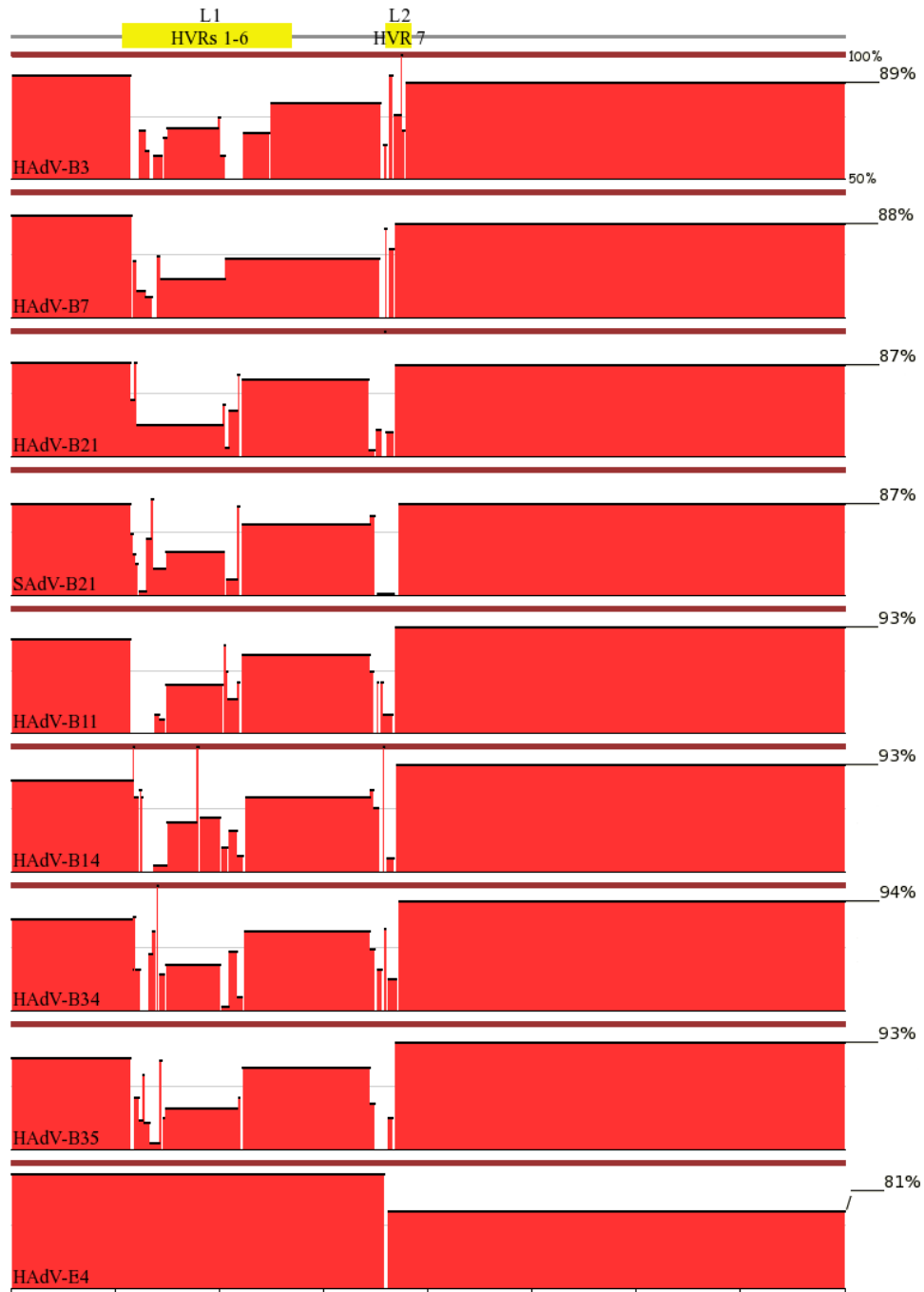


Figure 3 - zPicture Analysis of HAdV-B16.

The HAdV-B16 hexon gene was aligned, using zPicture, against subspecies B1 (HAdV-B3, B7, B21, B50 and SAdV-B21), subspecies B2 (HAdV-B11, B14, B34, B35, and B55), and HAdV-E4. The x-axis ranges from nucleotide 1 to 2800. Numbers along the y-axis represent the percent identity from 50% to 100%. Hypervariable regions L1 and L2 are indicated at the top of the alignments, for reference and are approximate locations. GenBank accession numbers: HAdV-B16 (AY601636), HAdV-B3 (AY599836), HAdV-B7 (AY594255), HAdV-B21 (AY601633), HAdV-B50 (AY737798), SAdV-B21 (AC_000010), HAdV-B11 (AC_000015), HAdV-B14 (AY803294), HAdV-B34 (AY737797), HAdV-B35 (AC_000019), HAdV-B55 (FJ643676) and HAdV-E4 (AY594253).

In contrast, the distal sequences of HAdV-B16 showed 92.5% identity with HAdV-E4, where the zPicture alignments showed that the distal portion of the HAdV-B16 hexon as having high similarities to the subspecies B2 sequences, for example percent identity results showed 98% identity with HAdV-B11, HAdV-B34 and HAdV-B35; 97% with rest of the HAdV-B species, see **Figure 3**. This region represents approximately 34.0% of the hexon and 2.70% of the genome (**Table 2**). Also noted is that HAdV-B16 has the highest percent identity of 95% with HAdV-B3 and B7 in the penton base gene, while the rest of percent identities of the reference genomes are in the low to mid 80%. The fiberknob region of HAdV B-16 also shows the highest percent identity to Simian Adenovirus (SAdV) 35.1 and 35.2 of 97.4% and 97.9% respectively.

Table 2 - Percent Identity of HAdV-B16.

Percent identities of the nucleotide coding sequences of selected HAdV-B16 coding regions to homologous sequences from viruses in species HAdV-B, HAdV-E and SAdV-B.

Species	Types	Penton base	Hexon	Hexon Proximal	Hexon Distal	Fiber knob
Human	HAdV-B3	95.0	86.8	75.4	97.4	61.8
Human	HAdV-B7	95.0	86.8	75.2	97.4	52.3
Human	HAdV-B21	84.8	85.2	72.4	97.2	52.2
Simian	SAdV-B21	84.0	85.0	71.8	97.6	79.4
Human	HAdV-B50	84.7	86.1	74.2	97.2	51.9
Human	HAdV-B11	84.2	85.5	72.1	98.0	51.7
Human	HAdV-B14	85.1	85.4	72.2	97.8	50.9
Human	HAdV-B34	85.9	85.2	71.6	98.0	51.9
Human	HAdV-B35	84.1	85.3	71.8	98.0	51.9
Simian	SAdV-B35.1	84.5	84.7	71.7	97.0	97.4
Simian	SAdV-B35.2	84.5	84.7	71.7	97.0	97.9
Human	HAdV-B55	85.4	85.4	72.0	97.8	50.6
Human	HAdV-E4	80.2	93.5	94.6	92.5	29.7
Simian	SAdV-27.1	84.9	85.2	74.1	96.0	55.7

Hexon recombination

To explore the zPicture and percent identity results in greater detail, and to examine its origins, the HAdV-B16 genome was examined using software to detect sequence recombination, SimPlot 3.5.1 (<http://sray.med.som.jhmi.edu/SCSoftware/SimPlot>). SimPlot calculates and plots the percent identity of the query sequence to a panel of reference sequences in a sliding window, which move across the alignment of steps (Lole et al., 1999). This was approached initially by using whole genome and hexon sequences from all of the sequenced HAdVs, and then narrowing the eventual query set to the B1, B2 and E genomes. Bootscan analysis, an option of SimPlot, revealed a high degree of similarity of the proximal portion of the HAdV-B16 hexon with HAdV-E4 (**Figure 4A**). SimPlot analysis of the distal portion of the HAdV-B16 hexon indicated conservation to the B2 species collectively (**Figure 4B**). Iterations of recombination analyses using each of the B2 genomes separately showed each contributed equally to the proximal recombination as shown with the Bootscan analyses (**Figure 5A**). Together, these provide evidence for a recombination event with an ancestral species B2 genome at the distal portion of hexon (**Figure 5B**), unlike the two other HAdV recombinations described recently involving the proximal portions of the hexon gene (Walsh et al., 2009; Walsh et al., 2010). Bootscan analysis is an option of SimPlot, with adjustable window, step and repeat sizes; it repeatedly generates bootscan phylogenetic trees using random halves of the sequence within a given window. The reference sequence that has been clade with the query sequence the most number of times will be represented on the top of the plot.

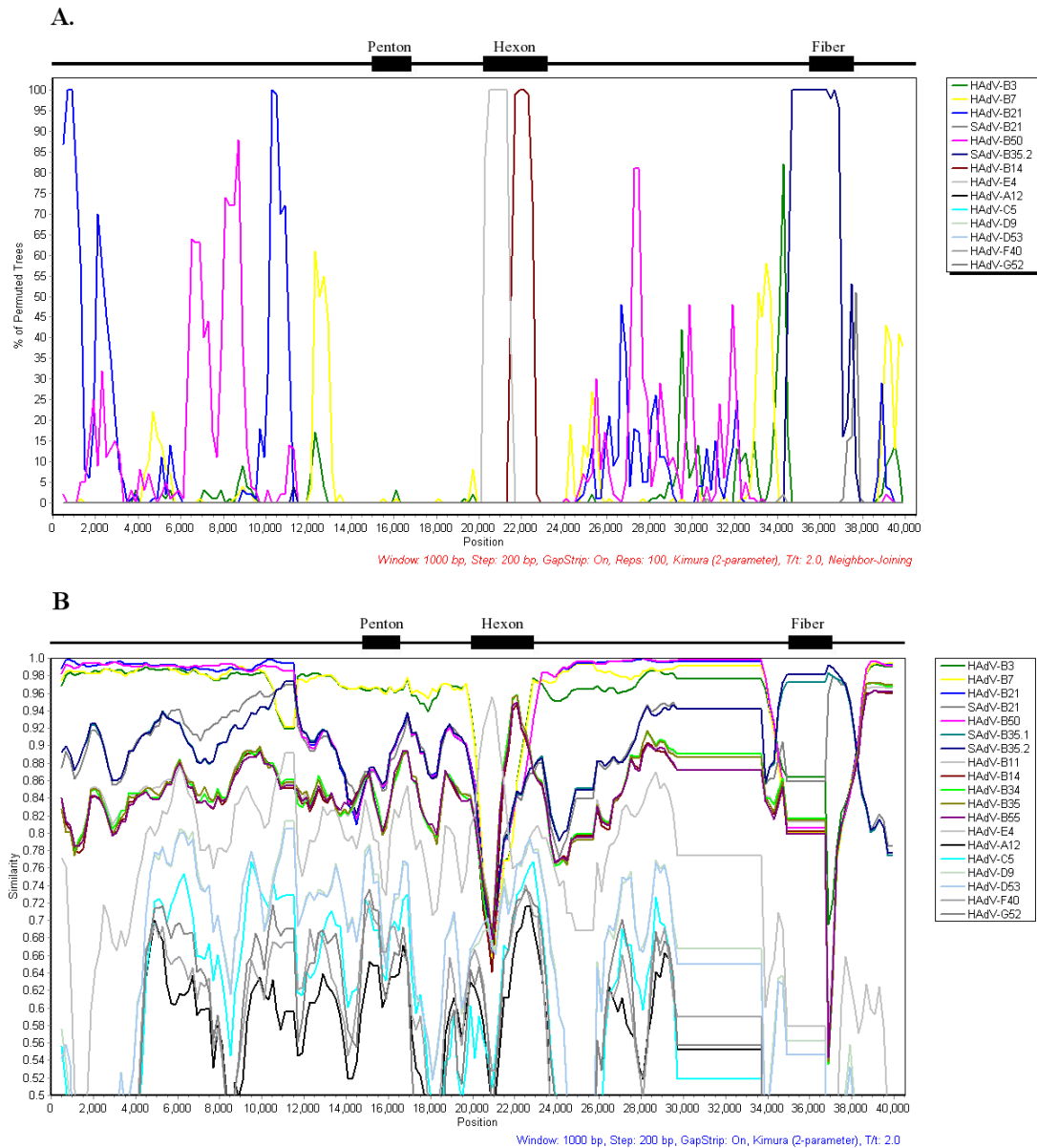


Figure 4 - Whole genome recombination analysis of HAdV-16.

A) Whole genome Bootscan analysis of HAdV-B16 with representative HAdV genomes, showing sequence similarity to HAdV-E4 at the hexon sequence, with all four subspecies B2 genomes contributing equally and diluting out the similarity. (window size 1000bp, step size 200bp, repeat 100) **B)** Whole genome SimPlot analysis of HAdV-B16 with representative HAdV genomes. (window size 1000bp, step size 200bp, repeat 100).

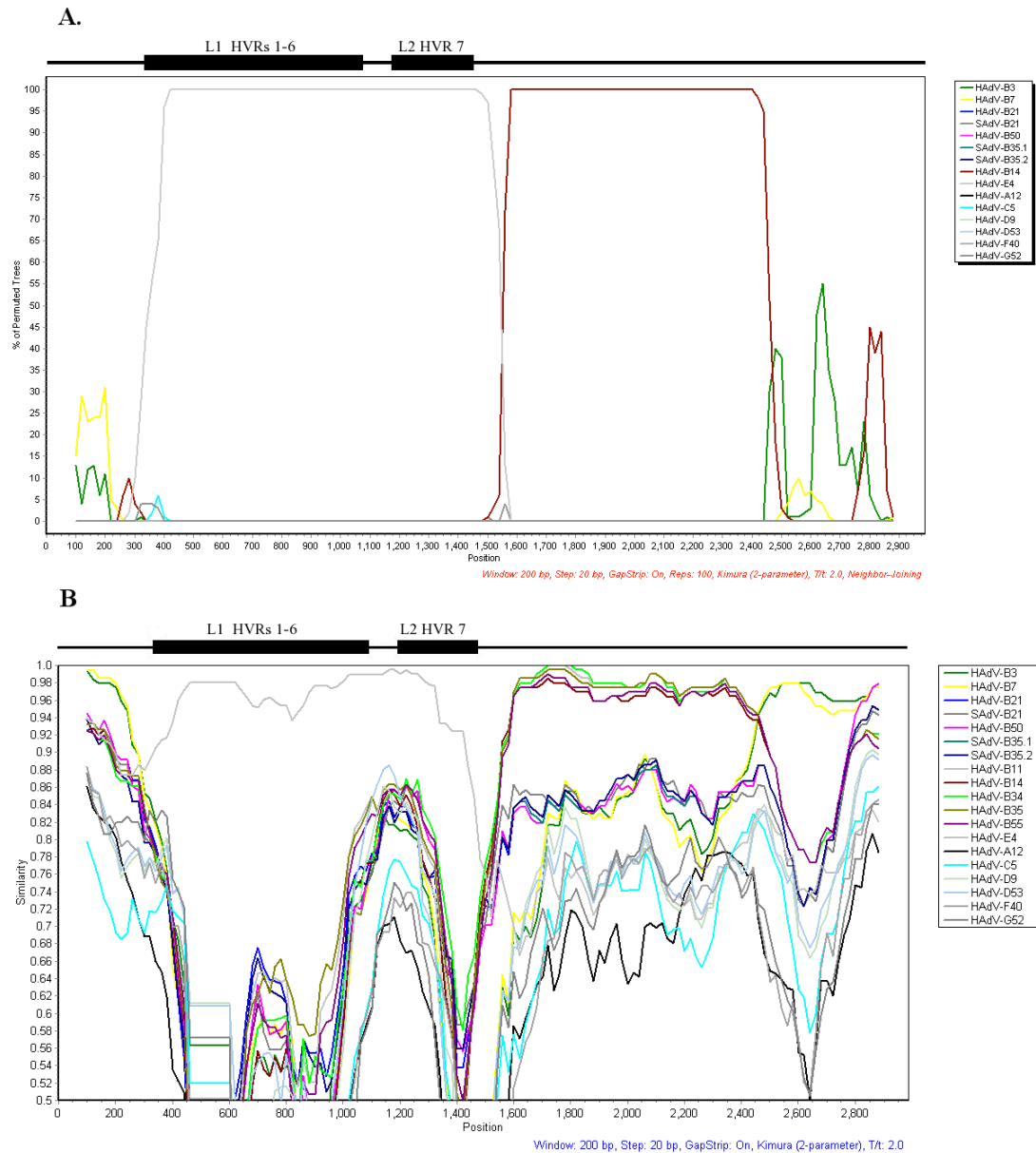


Figure 5 - Whole genome recombination analysis of HAdV-16.

A) Hexon Bootscan analysis of HAdV-B16 with representative HAdVs. This is a composite, using HAdV-B14 as a representative of the subspecies B2. Additional iterations with each B2 shows the identical pattern, and inclusion of all B2 members “competed” out the high similarity. (window size 200bp, step size 20bp) **B)** Hexon SimPlot analysis of HAdV-B16 with representative HAdVs, showing the subspecies B2 sequence contributions. (window size 200bp, step size 20bp) GenBank accession numbers, in addition to ones noted earlier, are as follows: HAdV-A12 AC_000005, HAdV-C5 AC_000008, HAdV-D9 AJ854486, HAdV-D53 FJ169625, HAdV-F40 NC_001454, HAdV-G52 DQ923122.

The second recombination event, involving the proximal portion of the hexon, suggests HAdV-B16 contributed to the evolution of HAdV-E4 (Dehghan et al., 2013), or vice versa. This is hypothesized as a host adaptation event as HAdV-E4 is the only human HAdV of species E with the rest being chimpanzee adenoviruses (SAdV-E22 to E25). The virus was reported as originating from a zoonotic event (Purkayastha et al., 2005).

Given the whole genome recombination analysis, without clear similarities to other genomes, it is likely the HAdV-B16 genome is an “ancient” sequence that has accumulated enough nucleotide changes to show divergence from other genomes as shown in the recombination analysis. The distal B2 recombination is likely a “recent” event, as it retains a high level of identity to the B2 sequences, e.g., without subsequent accumulated nucleotide changes. The “donation” of the proximal sequence to HAdV-E4 is likely recent as well, from the HAdV-E4 genome perspective (Dehghan et al., 2013).

Phylogeny analysis

Molecular Evolutionary Genetics Analysis (MEGA) 4.0.2 (Tamura et al., 2007) was used for phylogenetic analysis using bootstrap-confirmed neighbor-joining trees of the HAdV genomes, hexon gene and its parts, penton base gene and fiber gene, allowing a detailed examination of HAdV evolution and providing an additional view of these recombination events (data not shown). **Figure 6** displays a portion comprehensive whole genome phylogeny analysis tree. Shown in this phylogenetic snapshot are the B1 and B2 genomes forming a subclade together, as species B, but branching separately as

subspecies. As expected, HAdV-B16 subclades with species B1 members. HAdV-E4 is also shown in a subclade separated from species B and the other HAdV species. This clade includes other members of species E, the chimpanzee adenoviruses (Purkayastha et al., 2005; Purkayastha et al., 2005) (data not shown).

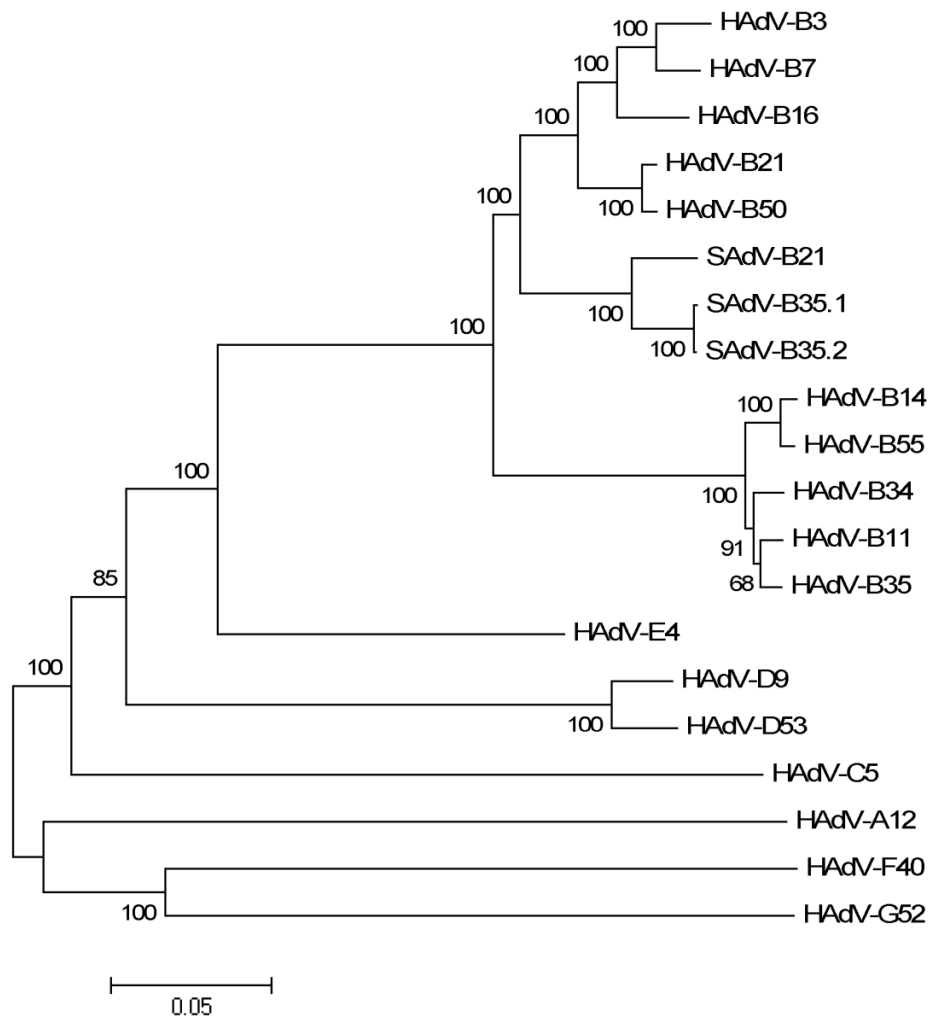


Figure 6 - Whole genome phylogenetic analysis of HAdV-B16.

The phylogenetic tree was constructed from aligned sequences using MEGA, via the neighbor-joining methods and a bootstrap test of phylogeny. Bootstrap values shown at the branching points indicate the percentages of 1000 replications produce the clade. A Bootstrap value of 70 and above is considered to be robust.

Phylogeny analysis of the hexon genes (**Figure 7**) shows HAdV-B16 in the same clade as, but branching away from, HAdV-B3 and B7. It has a similarity to the B2 subclade and the chimpanzee SAdV-B21 and HAdV-B21 subclade. Dividing the hexon sequence into proximal and distal subsequences, (**Figure 8A and 8B**), defined in the zPicture analysis, the proximal portion subclades with HAdV-E4. The distal portion of hexon branches with B2 subclade. Both hexon and fiber phylogeny trees presented the same results as the whole genome tree, the HAdV-B16 genes subclade with B1 members and away from both species E and subspecies B2 (data not shown). These results are consistent with and support the findings of the zPicture and the Bootscan analyses.

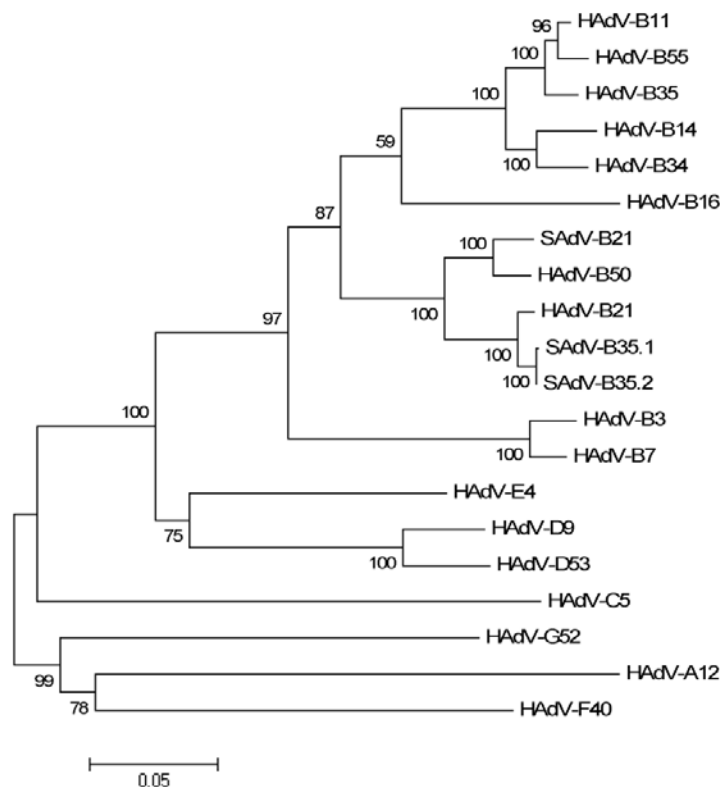


Figure 7 - Hexon phylogenetic analysis of HAdV-B16. Bootstrap neighbor-joining trees hexon gene sequence relationships, with the species B members and representatives of the other HAdV species for reference.

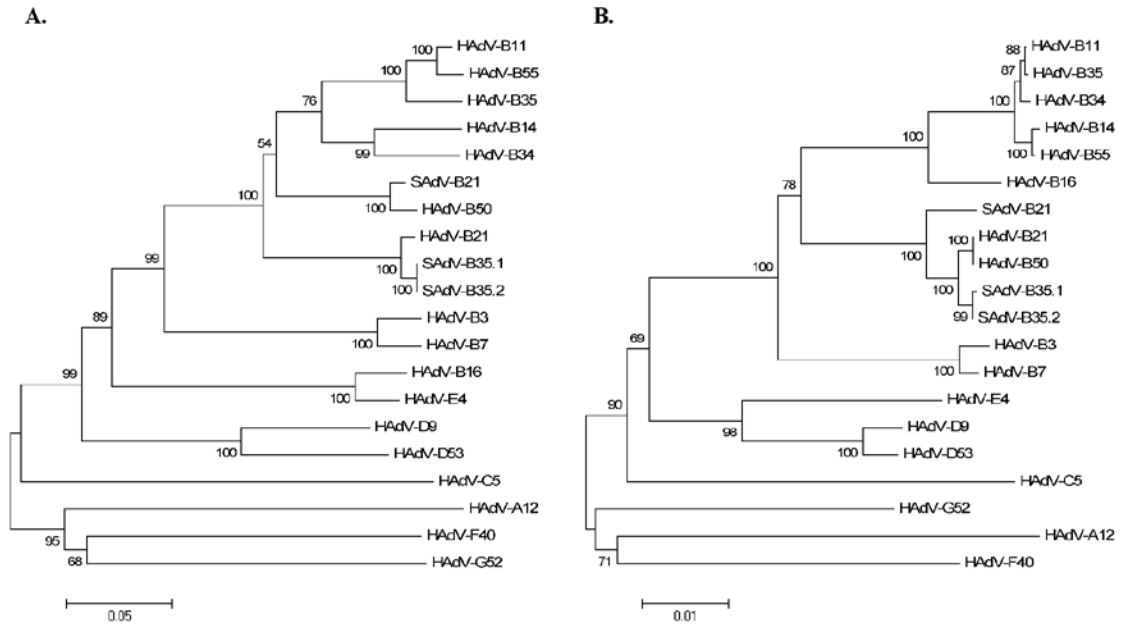


Figure 8 - Hexon halves phylogenetic analysis of HAdV-B16.

A) Proximal portion of Hexon sequence phylogenetic relationships. **B)** Distal portion of Hexon sequence phylogenetic relationships. Phylogenetic trees were generated from aligned sequences using MEGA, via the neighbor-joining method and a bootstrap test of phylogeny.

The phylogenetic analysis of HAdV-B16 fiber portion especially the fiber knob region appears to be grouped closest to the simian adenoviruses, SAdV-B35.1, SAdV-B35.2 and then to SAdV-B21, away from rest of the HAdV-Bs, the closest human adenovirus is HAdV-B3, while rest of the HAdV-Bs were branched separately (**Figure 9**).

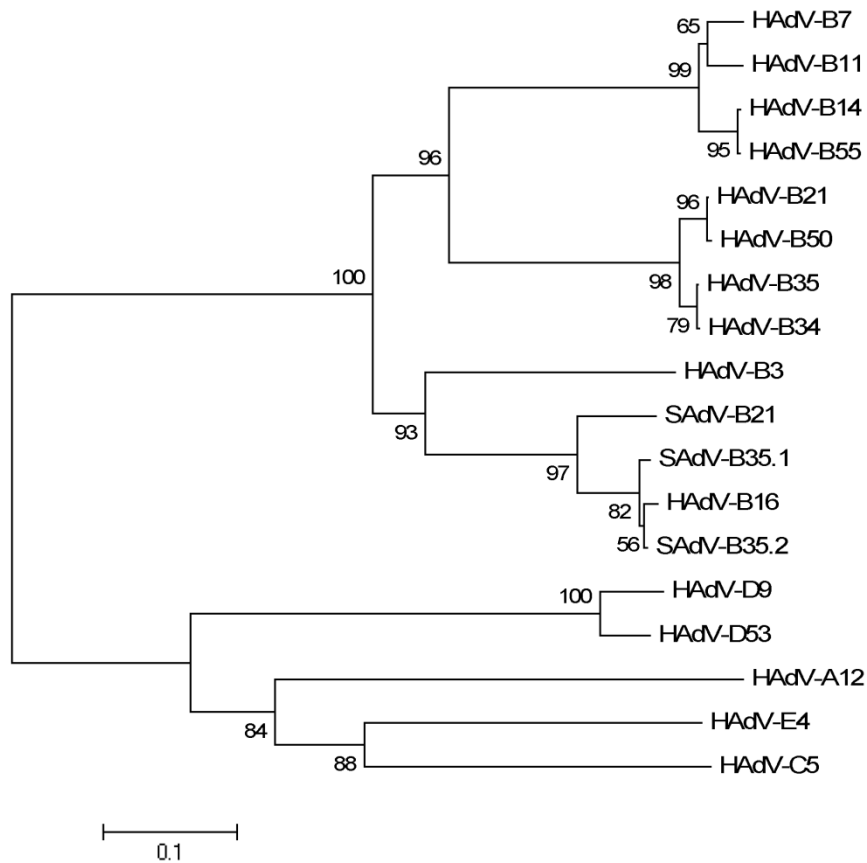


Figure 9 - Fiber knob phylogenetic analysis of HAdV-B16.

Phylogenetic tree was constructed from aligned sequences using MEGA, via the neighbor-joining method and a bootstrap test of phylogeny

Discussion

Recombination, molecular evolution and new serotype

As double-stranded DNA viruses, HAdV genomes are relatively stable with minor nucleotide changes, such as base substitution and insertion/deletions. This was documented in genome sequence comparisons between prototype and vaccine strains of HAdV-E4 and B7 (Purkayastha et al., 2005), as well as with a recent HAdV-B7 field

strain and its prototype (Seto et al., 2010). A similar result was found for longer time-spans, e.g., across minimum of fifty years of circulation, for five HAdV-B3 genomes (Mahadevan et al., 2010), and forty-two years, shown for several strains of HAdV-B7, which were assayed for antigenic differences, presumably reflecting strain variations, by serum neutralization tests and hexon sequencing (Crawford-Miksza et al., 1999); this contrasted with the HAdV-E4 genomes noted in the same study

HAdV genomes also undergo relatively large-scale changes, as recombination, which is noted as antigenic shifts. As the hexon is a target for neutralizing antisera (Toogood et al., 1992), it would be expected that changes in this epitope would result in altered serological response. In the past, this signified a novel HAdV serotype. A striking example is demonstrated recently for the newly recognized HAdV-B55 (misnamed as “HAdV-B11, QS” and “HAdV-B11a” (Zhu et al., 2009)). This is a recently reanalyzed HAdV identified as the recent re-emergent pathogen responsible for a highly contagious ARD outbreak (Walsh et al., 2010; Yang et al., 2009; Zhu et al., 2009) in which a recombination of the proximal portion of the hexon results in a change in serum neutralization patterns. The genome “chassis” of HAdV-B55 is predominantly HAdV-B14 (97.4%), with the partial hexon sequence of HAdV-B11 (2.6%) allowing it to serotype as HAdV-B11, a renal and urinary tract virus (Li et al., 1999; Numazaki et al., 1968). The genome recombination resolves the cell tropism riddle of “B55” having the cell tropic characteristics of B14 while serotyping as “B11” (Li et al., 1991; Mei et al., 1998). The original naming of HAdV-B55 as “HAdV-B11a” by serology was also

inconsistent with its biology and pathology as a respiratory pathogen, which archetype HAdV-11 was not.

In this study, HAdV-B16 is revealed to present a unique recombination event. To date, the only two hexon recombinants characterized in genomic detail involve a transfer of the proximal sequence (Walsh et al., 2009, Walsh et al., 2010), in contrast to this HAdV-B16 recombination. The distal portion of an ancestral B2 HAdV genome is shuffled, resulting in a “then-new” HAdV that warranted the recognition of a new serotype in the 1950s (Bell et al., 1959; Hierholzer et al., 1991; Pereira et al., 1963). This study also shows that HAdV-B16 contributed a proximal portion of its hexon gene to HAdV-E4, an interspecies event (the second such noted), perhaps allowing that virus to adapt and optimize to a human host following a zoonotic infection from chimpanzees to humans (Purkayastha et al., 2005; Purkayastha et al., 2005). In addition, this genome analysis explains earlier reports, and resolves a riddle, that of the “bilateral cross-neutralization observed between Ad16 of subgenus B and Ad4 of subgenus E” reported in the literature twenty-five years ago (Hierholzer et al., 1991).

HAdV genome recombination has been reported in the literature (Williams et al., 1975). Using molecular typing and/or serological techniques to characterize new variants and more virulent strains, such as HAdV-B7h (Kajon et al., 1996), as well as “intertypic” or “intermediate” strains (Bourisnell et al., 1981; Engelmann et al., 2006; Hierholzer et al., 1988; Hierholzer et al., 1976; Ishiko et al., 2008), recombinants have been suggested. Recombination was noted as a frequent event during the analyses of sixteen species C field isolates (Lukashev et al., 2008) and noted in a survey of novel AIDS-associated

HAdVs (Crawford-Miksa et al., 1996). Recombinant HAdV genomes have also been generated *in vitro* (Boursnell et al., 1981; Mautner et al., 1984). Given all these observations, recombination was hypothesized as a driving force for the molecular evolution of new HAdV serotypes (Crawford-Miksa et al., 1996). Recent reports using genomics have reconfirmed this, e.g., identified, characterized and christened HAdV-D53 and HAdV-B55 as novel recombinant HAdV, with partial proximal hexon transfers (Walsh et al., 2009; Walsh et al., 2010). Re-analysis of HAdV-22 by genomics and computational methods shown this is also a recombinant, albeit at the penton base gene (Robinson et al., 2009). Genomics has extended and refined these observations, for example, disproving a hypothesis with limited molecular typing data for sixteen species C field isolates that interspecies recombination does not occur for “available complete genome sequences of AdB, AdC and AdD species” (Lukashev et al., 2008).

It should be noted a recombination does not automatically define nor necessitate the recognition of a new “type”. An example of this is described for two recently isolated and characterized field strains of HAdV-E4 (Dehghan et al., 2013).

Conclusion

Genomic and bioinformatics comparisons of the HAdV-B16 genome to other HAdV genomes identified two exclusive and partial hexon gene recombination events, one as an interspecies donator to HAdV-E4 and the other as an acceptor of an ancestral subspecies B2 sequence. The former event explains the observed serological cross-

reaction, noted in the literature, with HAdV-E4. This is a unique snapshot of the molecular evolution of HAdV, and represents an exception to the previous hypotheses and observations that HAdV genomes did not shuffle sequences across species.

CHAPTER 3 – COMPUTATIONAL ANALYSIS OF HUMAN ADENOVIRUS TYPE 21

Abstract

One driving force of human adenoviruses (HAdVs) molecular evolution is genome recombination resulting in emergent and novel pathogens in the past and the present. HAdV-B21 is a human respiratory pathogen that is a recombinant containing a large genomic sequence, including the major capsid proteins penton base and hexon, that has near sequence identity with genomic sequences identified from both chimpanzee (SAdV-B35.1) and bonobo (SAdV-B35.2) AdVs. Understanding the role that recombination plays in adenovirus evolution and pathoepidemiology is important in vaccine development, along with their long-term effectiveness and in the development of gene therapy vectors, using SAdVs as an alternative to viruses with pre-existing immunity in humans. Genome recombination provides the realization that non-human simian species are reservoirs for potentially highly contagious and deadly human adenoviral pathogens through zoonosis.

Introduction

This study of the detailed analysis of HAdV-B21 is a genomic examination of an archived prototype, *circa* 1950s. The purpose of this genomic analysis is very similar to the analysis of HAdV-B16 discussed in Chapter 2, which is to support additionally that novel HAdVs may arise from genome recombination. HAdV-B21 also belongs to subspecies B1, whose members are usually responsible for respiratory diseases. There has been a report that shows an increase in the incidence of fatal adenovirus infections (Rowe et al., 1953). A number of those severe disease cases have been linked to HAdV-B21 (Lahm et al., 2010). The exact etiology for this unexpected high mortality remains unknown; the referenced case reports a patient with severe pneumonia resulting in hemophagocytic lymphohistocytosis (HLH) with acute respiratory distress syndrome and rapid progressive multi-organ dysfunction syndrome. It was proposed that an association between HAdV-B21 and HLH may, at least in part, explain the recent observed increase in incidence of fatal adenoviral infection (Lahm et al., 2010). To explore whether genome recombination is involved in the molecular evolution of HAdV-B21 in general, this study reexamines the original isolate using genomic and bioinformatics.

Material & Method

HAdV-B21 (AV-1645) was purchased from the American Type Culture Collection (ATCC; Manassas, VA). This virus was processed using protocols described for similar HAdV sequencing projects (Lauer et al., 2004; Purkayastha et al., 2005a; Seto

et al., 2009) by Virapur, LLC. (San Diego, CA); these include growth in A-549 cells and subsequent DNA purification.

Genome sequencing was outsourced to Commonwealth Biotechnologies, Inc. (Richmond, VA), applying the Sanger chemistry with the DYEnamic ET Terminator Cycle Sequencing kit (Amersham Biosciences; Piscataway, NJ); ladders were resolved on an ABI Prism 377 DNA Sequencer (Applied Biosystems; Foster City, CA); and assembled using DNA Sequencher (GeneCodes, Inc.; Ann Arbor, MI). Across the genome, an average of five-fold sequencing and a minimum of three-fold coverage, and both directions. Re-sequencing of areas that were found to be questionable upon sequence assembly and genome annotation was PCR-driven amplification and sequencing. Quality control included genome annotation, with comparisons to earlier types 1, 4, and 7 genome data, including the prototype and vaccine (Lauer et al., 2004; Purkayastha et al., 2005a; Purkayastha et al., 2005b). Annotation was performed using the Genome Annotation Transfer Utility (GATU) software tool (Tcherepanov et al., 2006), and recorded and visualized using Artemis, a genome viewer (<http://www.sanger.ac.uk/resources/software/artemis/>) (Rutherford et al., 2003).

Whole genome sequences used in this analysis are listed here along with their accession number: HAdV-B21 (AY601633), HAdV-B16 (AY601636), HAdV-B3 (AY599836), HAdV-B7 (AY594255), HAdV-B50 (AY737798), SAdV-B21 (AC_000010), HAdV-B11 (AC_000015), HAdV-B14 (AY803294), HAdV-B34 (AY737797), HAdV-B35 (AC_000019), SAdV-B35.1 (FJ025912), SAdV-B35.2 (FJ025910), HAdV-B55 (FJ643676) and HAdV-E4 (AY594253).

Results

Genome sequence analysis

HAdV-B21 has a genome size of 35,382 nucleotides with a GC content of 51%, consistent with subspecies B1 (51%) and differing from B2 (49%) and other HAdV species; GC% is a species criterion. Its genome is approximately 95% and 83% identical to the HAdV-B1 and HAdV-B2 members respectively, with much lower identities to other species and serotypes, next highest is HAdV-E4 at 74%. The virus fits into subspecies B1 instead of B2 as it contains two VA-RNAs, as well as other genome and proteome attributes as noted by earlier reports (Wadell et al., 1980; Wadell et al., 1984).

Comparative whole genome analysis

Comparative whole genome sequence analysis was performed using zPicture. This dynamic alignment tool aligns the HAdV-B21 genome against genomes of the B1 and B2 members, all HAdVs, along with SAdV-B21, SAdV-B35.1 and SAdV-B35.2, which are from non-human simian hosts. The zPicture whole genome results showed sequence divergence across the genomes for the B1 genomes and the SAdV genomes (**Figure 10**). The HAdV-B21 genome is almost identical to HAdV-B50 genome, except for the divergence in the proximal hexon region. Rather, the proximal hexon region shows high sequence conservation with SAdV-B35.1 and SAdV-B35.2 instead.

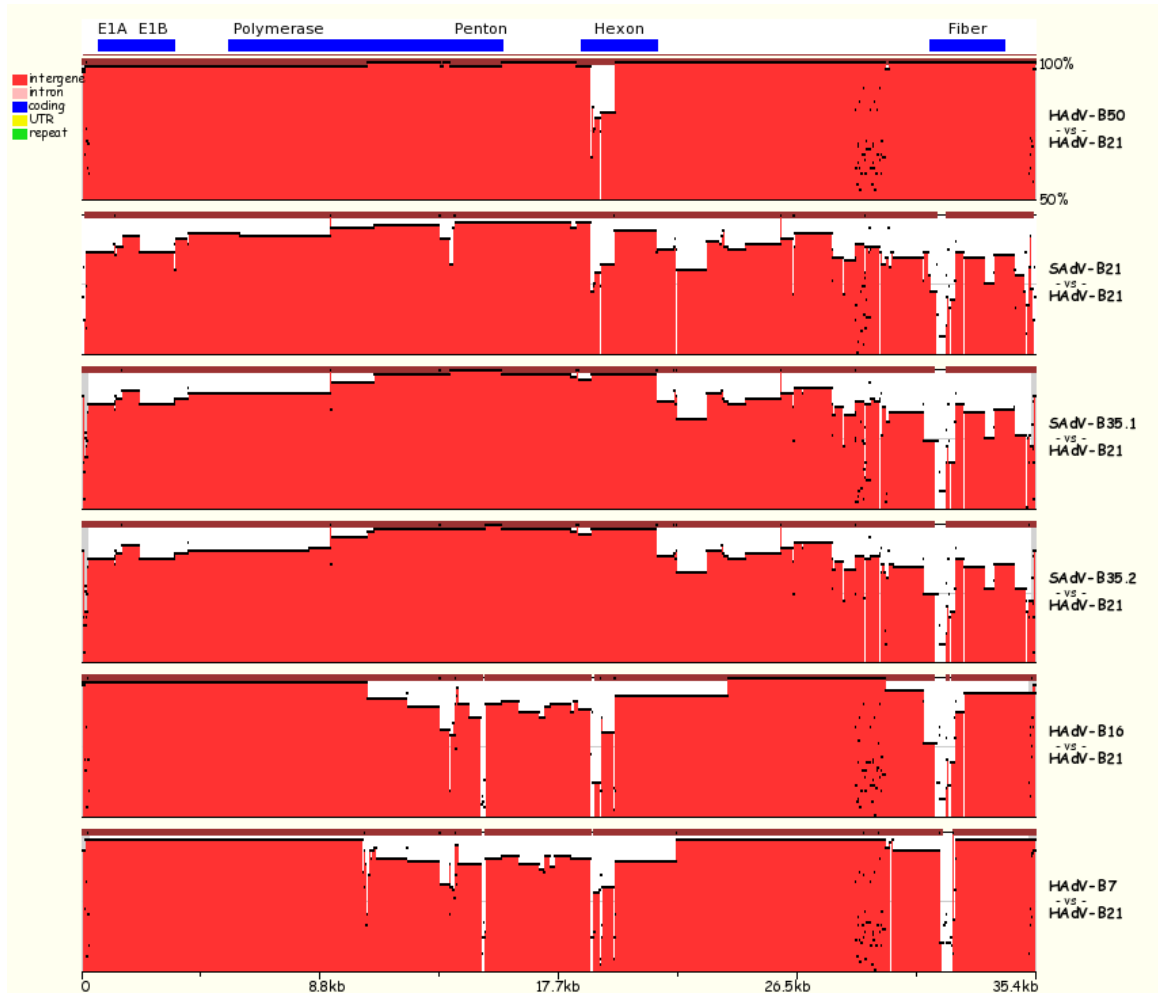


Figure 10 - zPicture Analysis of HAdV-B21.

The HAdV-B16 hexon gene was aligned, using zPicture, against HAdV-B50, B16, B7 and SAdV-B21, B35.1 and B35.2. The x-axis ranges from nucleotide 1 to 36000. Numbers along the y-axis represent the percent identity from 50% to 100%. Protein regions are indicated at the top of the alignments, for reference and are approximate locations.

Each surface protein of HAdV-B21 was further analyzed by calculating its percent identity to the rest of the reference genomes mentioned above. These results are shown in **Table 3**. HAdV-B21 penton base region has a percent identity of 99.3 to HAdV-B50, SAdV-B35.1 and SAdV-B35.2. The hexon region has the highest percent identity of 98.0 and 98.1 to SAdV-B35.1 and B35.2 respectively. Upon closer

examination, the proximal portion of HAdV-B21 only has 84.7% identity of HAdV-B50 and 99.3% identity to SAdV-B35.1 and B35.2. This region comprises approximately 900 nucleotides that represent 31.9% of the HAdV-B21 hexon and 2.5% of the total genome. In contrast the distal hexon region is 100% identical to HAdV-B50, and 99.8% identity to SAdV-B35.1 and B35.2. The fiber knob region is well-conserved with HAdV-B50; it has a percent identity of 99.7% and diverges away from the SAdVs numbers.

Table 3 - Percent identity of HAdV-B21.

Percent identities of the nucleotide coding sequences of selected HAdV-B21 coding regions to homologous sequences from viruses in species HAdV-B, HAdV-E and SAdV-B.

Species	Types	Penton base	Hexon	Hexon Proximal	Hexon Distal	Fiber knob
Human	HAdV-B3	85.8	85.5	74.9	95.6	57.0
Human	HAdV-B7	85.6	85.9	75.9	95.4	59.7
Human	HAdV-B16	84.8	85.2	72.4	97.2	52.2
Simian	SAdV-B21	97.7	91.3	83.1	99.0	56.7
Human	HAdV-B50	99.3	92.6	84.7	100.0	99.7
Human	HAdV-B11	92.2	93.5	89.1	97.6	59.4
Human	HAdV-B14	92.0	90.2	82.5	97.4	60.0
Human	HAdV-B34	92.5	90.3	82.4	97.6	95.2
Human	HAdV-B35	92.3	91.3	84.6	97.6	94.8
Simian	SAdV-B35.1	99.3	98.0	96.1	99.8	53.4
Simian	SAdV-B35.2	99.3	98.1	96.3	99.8	53.1
Human	HAdV-B55	92.2	92.4	87.1	97.4	60.0
Human	HAdV-E4	83.9	83.3	72.9	92.9	26.0

Hexon recombination

To further explore the zPicture and percent identity results in greater detail, and to examine HAdV-B21's origins, a sequence recombination detection software SimPlot was used to calculate and plots the percent identities of the HAdV-B21 to a panel of reference

sequences in a sliding window that moves across in steps of the alignment in order to detect recombination (**Figure 11A and 11B**). A closer examination of the hexon region is done by SimPlot and Bootscan analyses, they displayed a high degree similarity of the proximal half of HAdV-B21 with SAdV-B35.1 and B35.2; distal half of the HAdV-B16 indicated conservation to HAdV-B50 (**Figure 12A and 12B**). Due to the near-identical percent similarity between SAdV-B35.1 and B35.2 that will compete against each other in Bootscan graph and obscuring the true result, SAdV-B35.1 was excluded and SAdV-B35.2 was used in this figure to represent the recombination event.

Given the whole genome recombination analysis and the hexon recombination analysis, it is likely that HAdV-B21 is an ancient genome sequence that has accumulated nucleotides changes across species and shown divergence from other genomes. The proximal SAdV-B35 recombination contains a high level of identity to HAdV-B16 which can be a recent event. The zoonotic transfer is consistent with the ones observed in previous studies, with the direction of the zoonosis undetermined (Dehghan et al., 2013).

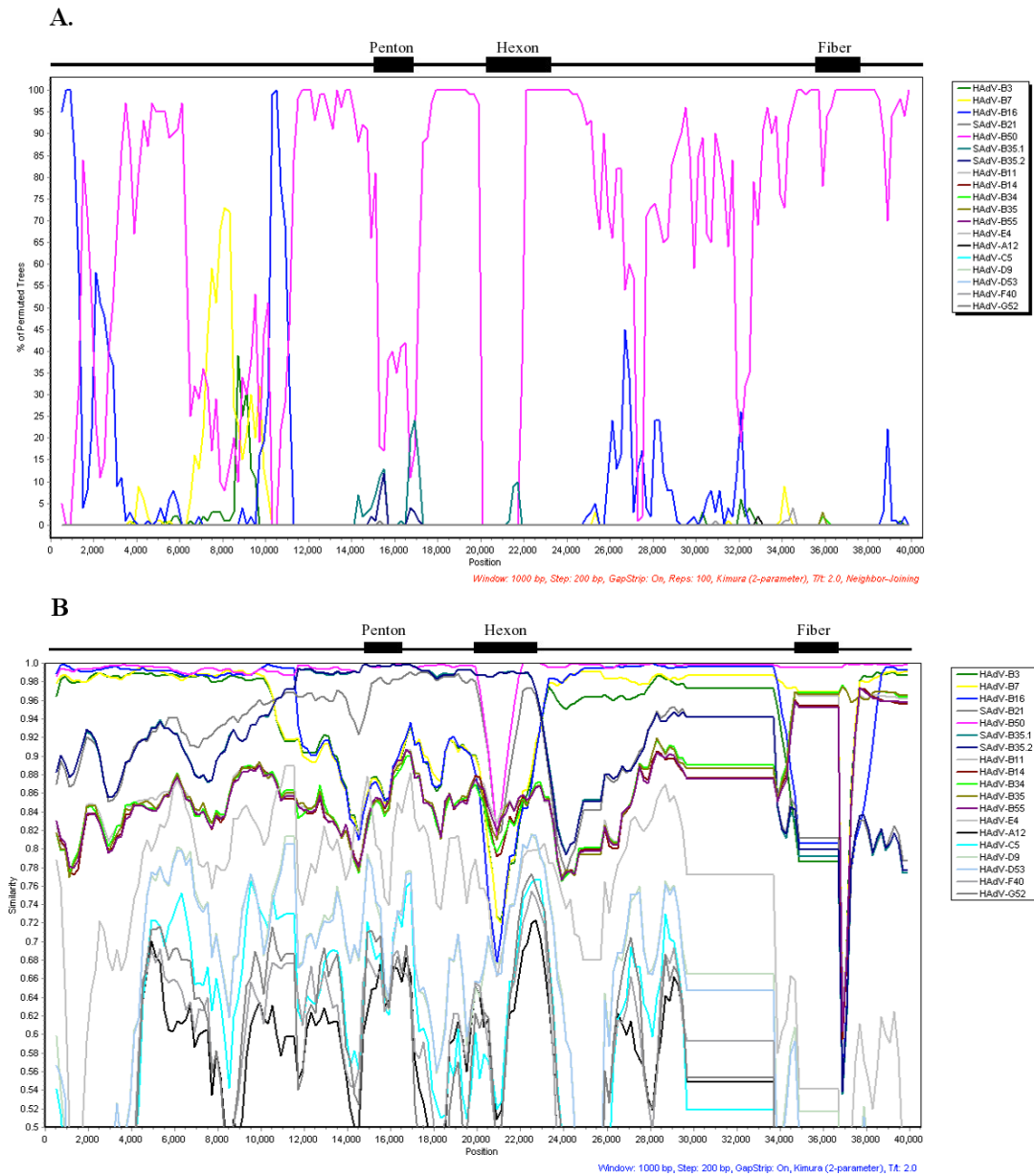


Figure 11 - Whole genome recombination analysis of HAdV-21.

A) Whole genome Bootscan analysis of HAdV-B21 with representative HAdV genomes, showing sequence similarity to HAdV-B50 at the hexon sequence. (window size 1000bp, step size 200bp, repeat 100) **B)** Whole genome SimPlot analysis of HAdV-B21 with representative HAdV genomes, shows high similarity of SAAdV-B35.1 and B35.2 at the distal region of hexon (window size 1000bp, step size 200bp, repeat 100).

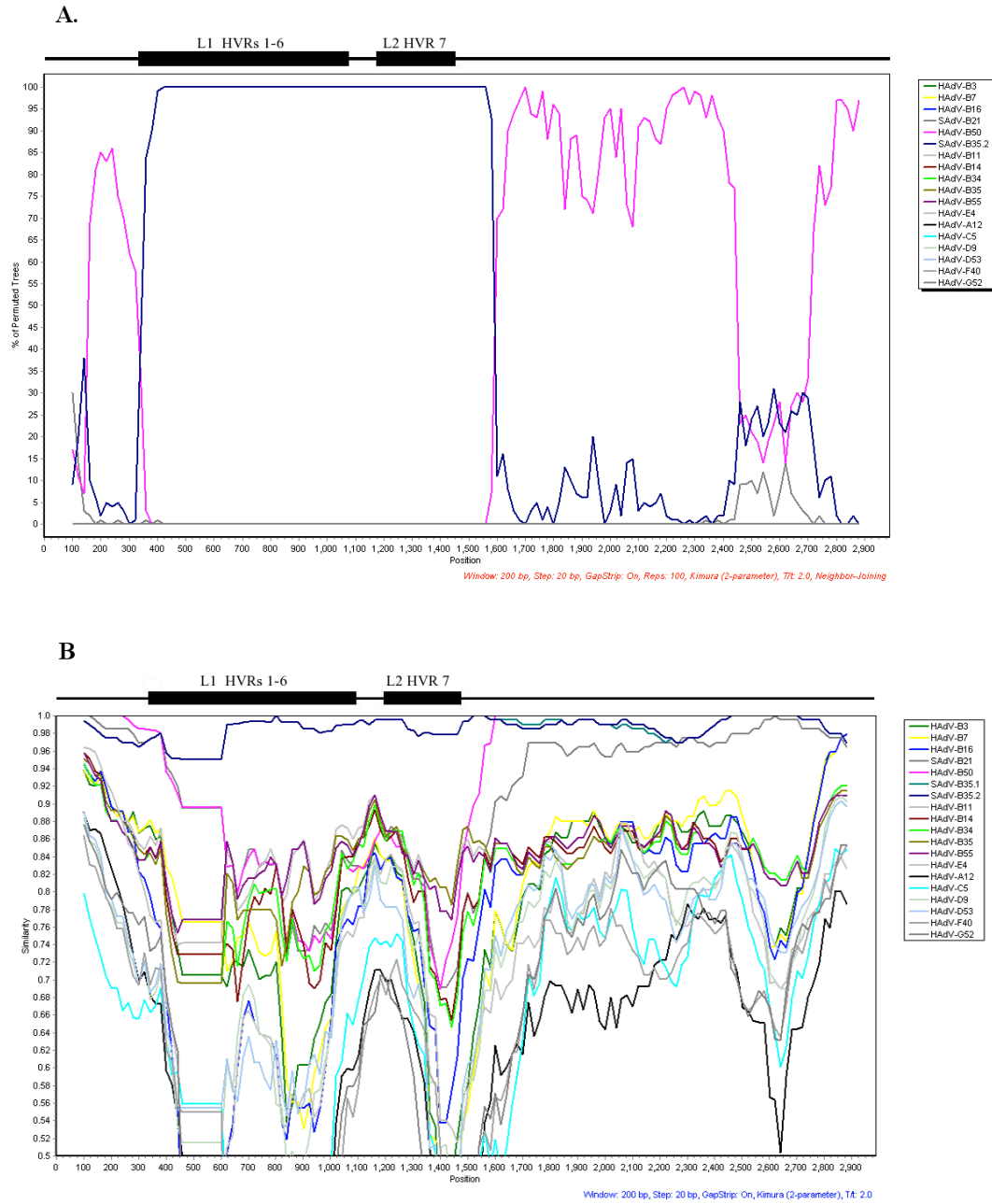


Figure 12 - Whole genome recombination analysis of HAdV-21.

A) Hexon Bootscan analysis of HAdV-B21 with representative HAdVs. This is a composite, using HAdV-B35.2 as a representative of the SAdV-B35. Additional iterations with each SAdV-B35.1 shows the identical pattern, and inclusion of all SAdV-B35 members “competed” out the high similarity. (window size 200bp, step size 20bp) **B)** Hexon SimPlot analysis of HAdV-B21 with representative HAdVs, shows the SAdV-B35 sequence contributions. (window size 200bp, step size 20bp)

Phylogenetic analysis

MEGA was used for phylogenetic analysis with bootstrap confirmed neighbor joining trees of the HAdV genomes, penton base, hexon and its parts, and the fiber knob region. This provides an additional view of the recombination events and allowing a closer examination of the HAdV evolution. Whole genome phylogenetic analysis shows B1 and B2 genomes forming a subclade together as part of the species B but also branching separately with SAdV-Bs in between (**Figure 13**).

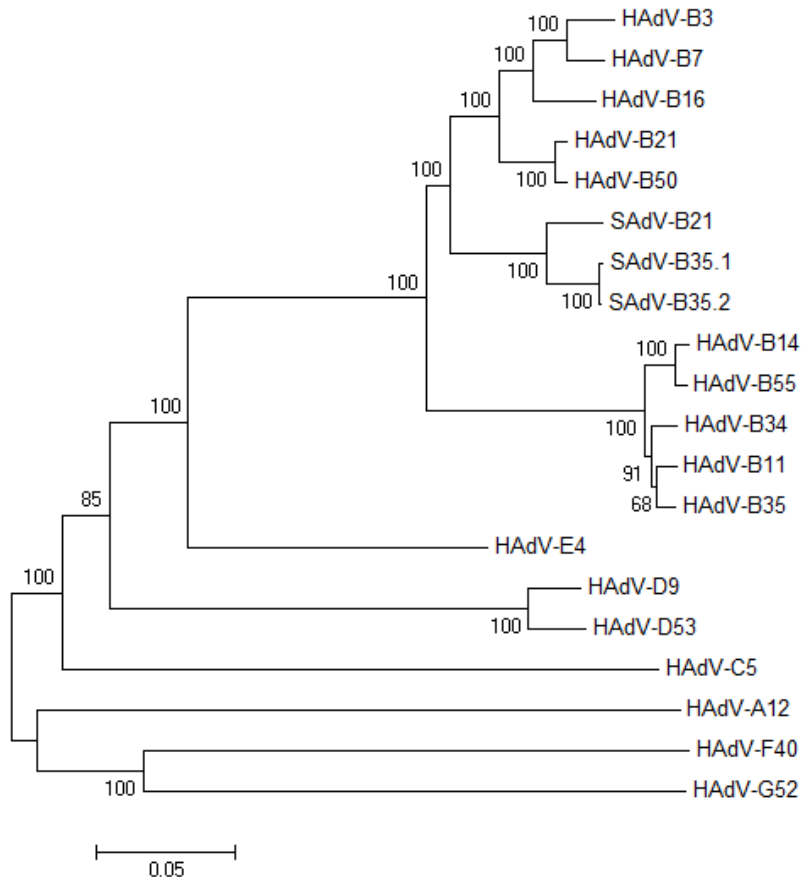


Figure 13 - Whole genome phylogenetic analysis of HAdV-B21.

The phylogenetic tree was constructed from aligned sequences using MEGA, via the neighbor-joining methods and a bootstrap test of phylogeny. Bootstrap values shown at the branching points indicate the percentages of 1000 replications produce the clade. A Bootstrap value of 70 and above is considered to be robust.

Phylogenetic analysis of the hexon gene (**Figure 14**) shows HAdV-B21 in the same clade but branching away from, as rest of the B1s. It has higher similarity to SAdV-B35.1, B35.2 subclades and SAdV-B21, HAdV-B50 subclades. For detailed analysis, the hexon sequence is then divided into the proximal and distal subsequences (**Figure 15A and 15B**) defined in zPicture. This shows that the proximal portion subclades with SAdV-B35.1 and B35.2 while the distal portion subclades with HAdV-B50. These results are consistent with and support the findings from the zPicture and SimPlot Bootscan analyses.

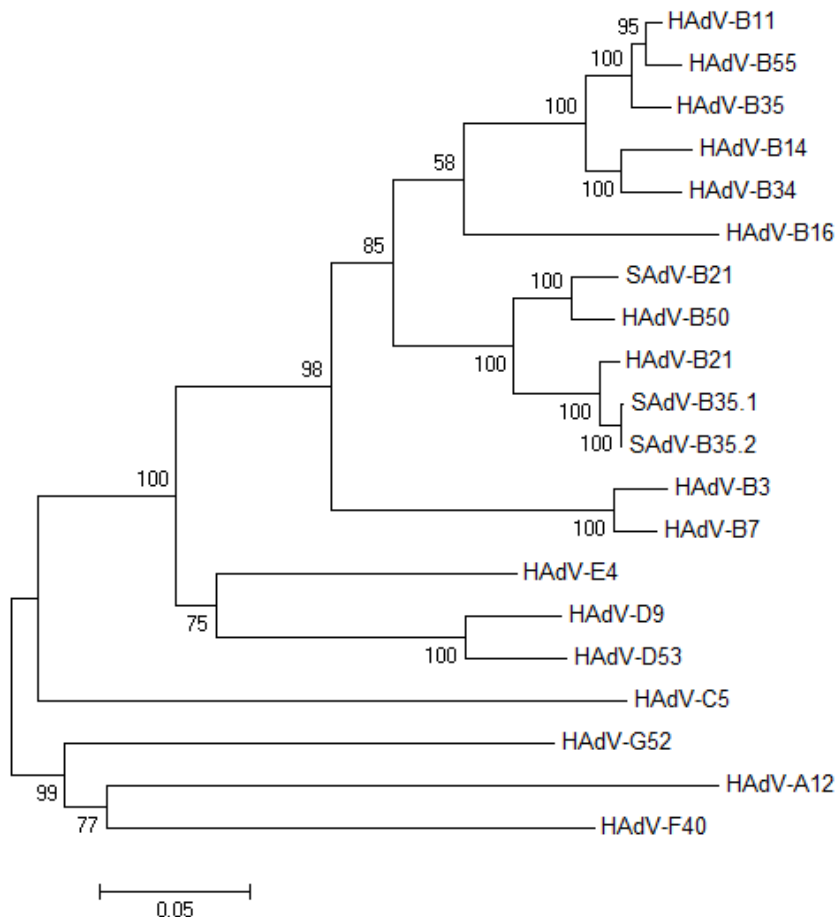


Figure 14 - Hexon phylogenetic analysis of HAdV-B21. Bootstrap neighbor-joining trees hexon gene sequence relationships, with the species B members and representatives of the other HAdV species for reference.

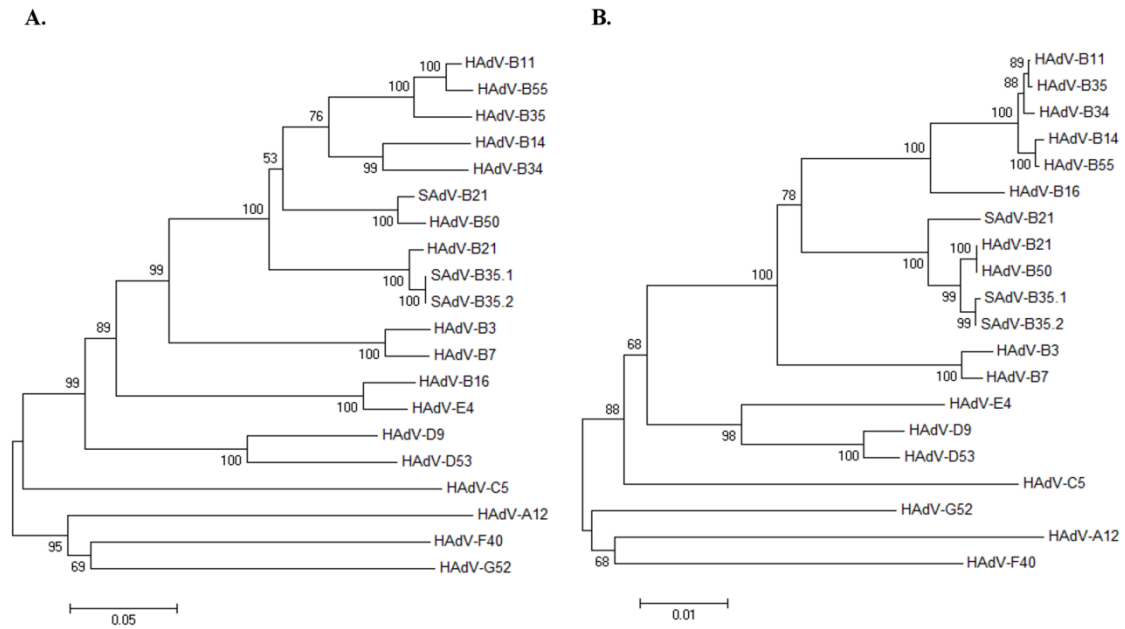


Figure 15 - Hexon halves phylogenetic analysis of HAdV-B21. **A)** Proximal portion of Hexon sequence phylogenetic relationships. **B)** Distal portion of Hexon sequence phylogenetic relationships. Phylogenetic trees were generated from aligned sequences using MEGA, via the neighbor-joining method and a bootstrap test of phylogeny.

The phylogenetic analysis of the fiber knob region shows HAdV-B21 clade closest to HAdV-B50 then to HAdV-B34 and B35. It branches away from rest of the Simian Bs and away from rest of the B1s and B2s. (**Figure 16**).

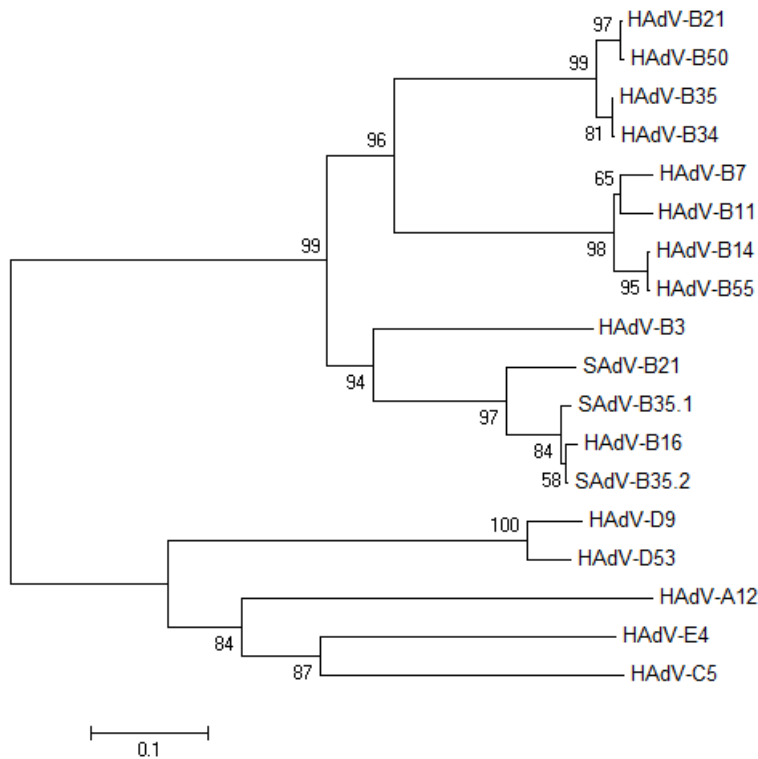


Figure 16 - Fiber knob phylogenetic analysis of HAdV-B21. Phylogenetic tree was constructed from aligned sequences using MEGA, via the neighbor-joining method and a bootstrap test of phylogeny.

Discussion

With the applications of bioinformatics methods and genomic technology available today, we understand the relationship of infectious diseases and microbial pathogens in greater detail (Relman et al., 2011). Potentially emergent pathogens may be predicted using genome sequence data, as noted for a HAdV (Robinson et al., 2013). The recently availability of a large quantity of genomic data also provides valuable information to understanding the molecular evolution mechanism for emergent and novel pathogens (Robinson et al., 2009; Walsh et al., 2009).

In this study, HAdV-B21 presents a unique recombination event between human and simian adenoviruses. Although the recombination direction cannot be determined, DNA sequences are transferred laterally between human and simian viruses. These viruses from zoonotic transfers may become optimized and adapted through evolution to the new host (Dehghan et al., 2013). HAdVs are important biomedical tools as vectors for gene and epitope delivery (Darr et al., 2009; Stone et al., 2006). Simian adenoviruses are increasingly being considered as alternative to its human strains due to seroprevalence concerns. Their recent attentions to SAdVs as potential gene vectors have brought more genomes sequences into the data set (Roy et al., 2004), allowing for more comprehensive analyses and deeper insights, such as reported here for HAdV-B21.

Conclusion

Bioinformatics and genomic comparisons of HAdV-B21 genome to other HAdV genomes revealed two recombination events in the penton base and hexon genes with both chimpanzee (SAdV-B35.1) and bonobo (SAdV-B35.2) AdVs. Comprehending the role that recombination plays in adenovirus evolution is important in gene therapy vector development and in the development of vaccines. Similar to the previous chapter, this also represents another exception to the previous observations and hypotheses that HAdV genome did not exchange sequences across species.

CHATPER 4 – HUMAN ADENOVIRUS TYPE 58

Introduction

Human adenovirus type 58 (HAdV-D58) is the first novel gastrointestinal adenoviral pathogen identified since the 1990s (De Jong et al., 1999). It was isolated from the stool sample of an AIDS patient who presented with severe chronic diarrhea. The virus contains a novel hexon gene coding sequence as well as a novel recombinant fiber gene. Serological analysis also demonstrated that HAdV-D58 has a different neutralization profile than all other HAdV characterized to date since the 1990s. The characterization of this gastrointetinal pathogen has public health significance, especially in Argentina and South America where it was first isolated, as well as other “developing” regions.

Genetic Analysis of a Novel Human Adenovirus with a Serologically Unique Hexon and a Recombinant Fiber Gene

Elizabeth B. Liu¹, Leonardo Ferreyra², Stephen L. Fischer³, Jorge V. Pavan², Silvia V. Nates², Nolan Ryan Hudson⁴, Damaris Tirado⁴, David W. Dyer⁵, James Chodosh⁶, Donald Seto¹, Morris S. Jones^{7*}

1 Department of Bioinformatics and Computational Biology and Department of Systems Biology, George Mason University, Manassas, Virginia, United States of America, **2** Virology Institute, School of Medical Sciences, National University of Cordoba, Cordoba, Argentina, **3** Naval Hospital Camp Pendleton, Camp Pendleton, California, United States of America, **4** Clinical Investigation Facility, David Grant USAF Medical Center, Travis AFB, Fairfield, California, United States of America, **5** Department of Microbiology and Immunology, University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma, United States of America, **6** Howe Laboratory, Massachusetts Eye and Ear Infirmary, Department of Ophthalmology, Harvard Medical School, Boston, Massachusetts, United States of America, **7** Viral and Rickettsial Disease Laboratory, California Department of Public Health, Richmond, California, United States of America

Abstract

In February of 1996 a human adenovirus (formerly known as Ad-Cor-96-487) was isolated from the stool of an AIDS patient who presented with severe chronic diarrhea. To characterize this apparently novel pathogen of potential public health significance, the complete genome of this adenovirus was sequenced to elucidate its origin. Bioinformatic and phylogenetic analyses of this genome demonstrate that this virus, heretofore referred to as HAdV-D58, contains a novel hexon gene as well as a recombinant fiber gene. In addition, serological analysis demonstrated that HAdV-D58 has a different neutralization profile than all previously characterized HAdVs. Bootscan analysis of the HAdV-D58 fiber gene strongly suggests one recombination event.

Citation: Liu EB, Ferreyra L, Fischer SL, Pavan JV, Nates SV, et al. (2011) Genetic Analysis of a Novel Human Adenovirus with a Serologically Unique Hexon and a Recombinant Fiber Gene. PLoS ONE 6(9): e24491. doi:10.1371/journal.pone.0024491

Editor: Damien P. Martin, Institute of Infectious Disease and Molecular Medicine, South Africa

Received: June 22, 2011; **Accepted:** August 11, 2011; **Published:** September 7, 2011

Copyright: © 2011 Liu et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This research was supported by R01EY013124 (DS, MSJ and JC) and P30EY014104 (JC, MSJ) was also funded by the United States Air Force Surgeon General-approved Clinical Investigation No. FDG20040024E. JC was also funded by an unrestricted grant to the Department of Ophthalmology, Harvard Medical School, from Research to Prevent Blindness, Inc. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: morris.jones@cdph.ca.gov

Introduction

Human adenoviruses (HAdVs) were first isolated in 1953 from pediatric adenoid tissue and from a military basic trainee as respiratory pathogens [1, 2]. Since then, 56 types have been isolated and characterized [3, 4, 5, 6, 7]. Currently, there are 56 HAdVs in the genus Mastadenovirus in the family Adenoviridae, that are organized into seven species (A–G) [3, 4, 7, 8]. Individual HAdV types were originally differentiated based on immunochemical or serological methods, but more recently, genomics and bioinformatic analyses have supplanted serology [8]. Members of each species are highly similar at the nucleotide level, and do not appear to recombine readily with members of another species. Species grouping also reflect a tendency for specific human diseases: for example many HAdVs within species HAdV-D cause epidemic keratoconjunctivitis [9], whereas HAdVs in species HAdV-B are known to cause respiratory infections [10].

Currently there are three human adenoviruses (HAdV-F40, HAdV-F41 and HAdV-G52) that are associated with gastroenteritis [7, 8]. Gastroenteritis is associated with an estimated 5,000 deaths per year in United States [11]. It is likely that the etiological agents of gastroenteritis include yet-to-be identified pathogenic agents.

In this report we examined an adenovirus isolated from the stool of an AIDS patient who presented with severe chronic diarrhea.

Based upon whole genomic and bioinformatic analysis, this virus appears to belong to species HAdV-D, with the proposed name HAdV-D58.

Results

Microbiological Investigation

In February of 1996 an adenovirus was isolated from the stool of a 31-year-old AIDS patient who presented with severe chronic diarrhea and was subsequently hospitalized. *Cryptosporidium parvum* and *Giardia lamblia* were also found in the fecal matter of the patient; therefore, the clinical symptoms cannot be exclusively linked with the adenovirus infection.

Amplification and sequencing of the novel adenovirus

Partial sequencing of HAdV-D58, previously published as the Ad-Cor-96-487 strain [12], via imputed serum neutralization, demonstrated that portions of the hexon and fiber genes resembled HAdV-D33 and HAdV-D29, respectively [12]. This suggested that this novel HAdV isolated from an AIDS patient originated at least in part by recombination. To elucidate the genetic characteristics of HAdV-D58, the entire genome has been sequenced and analyzed.

a low of 90.72% (HAdV-D8; phylogenetic distance of 0.0711) to 93.97% (HAdV-D49; phylogenetic distance of 0.0341).

Genomic recombination analysis

Comparison of HAdV-D58 with the full-length genomes of viruses in species HAdV-D using SimPlot analysis revealed significant sequence divergence in the hexon, E3, and fiber coding sequences (Fig. 2).

Genetic analysis of the novel adenovirus hexon coding sequences

Analysis of the HAdV-D58 genome via pairwise comparison suggested that the hexon coding sequence was unlike any other known human adenovirus hexon sequence (Fig. 2). To determine if the hexon gene was novel, we performed SimPlot analysis using all hexon loop 1 (L1) and loop 2 (L2) coding sequences in species HAdV-D. L1 and L2 contain the epsilon (ϵ) determinant, which contain the epitopes for serum neutralization [13]. SimPlot analysis confirmed that the hexon gene of HAdV-D58 is unique compared with all other hexon genes in species HAdV-D (Fig. 3). In terms of nucleotide identity, the L1 and L2 of HAdV-D33 were most similar to HAdV-D58 with 84.4 and 89.8% nucleotide identity, respectively (Table S1). No substantial evidence of recombination in the hexon coding sequence was revealed.

Analysis of the E3 genes

In the E3 region 19K, RID α , RID β , and 14.7K are the only genes that have been investigated. The function of the E3/19K protein is to prevent human MHC class I molecules from being transported to the cell surface [14]. Specifically, amino acids W52, M87, and W96 were shown to be important for HLA-I modulation [14]. A second function of E3/19K is to inhibit NK cells from recognizing HAdV-infected cells by sequestering MHC-I chain-related proteins A and B (MICA/B) [15]. The 14.7K protein product inhibits the internalization of TNF receptor 1 [16]. The RID α and RID β proteins down-modulate the apoptosis receptor Fas/Apo-1 [17].

Bootscan analysis strongly suggests that there was a recombination event in the middle of the open reading frames of 19K and CR1- γ (Fig. 4). These recombination events did not disrupt any of the E3 open reading frames in the HAdV-D58 genome. Analysis of the 19K open reading frame in HAdV-D58 demonstrated that amino acids W52, M87, and W96 were present (data not shown). The percent identities of the HAdV-D58 19K, RID α , RID β , and 14.7K open reading frames were 96.6, 98.9, 98.4, and 97 percent identical to the homologous open reading frames of E3 coding sequences for HAdV-D49-19K, HAdV-D36-RID α , HAdV-D15-RID β , and HAdV-D15-14.7K, respectively (Table S2).

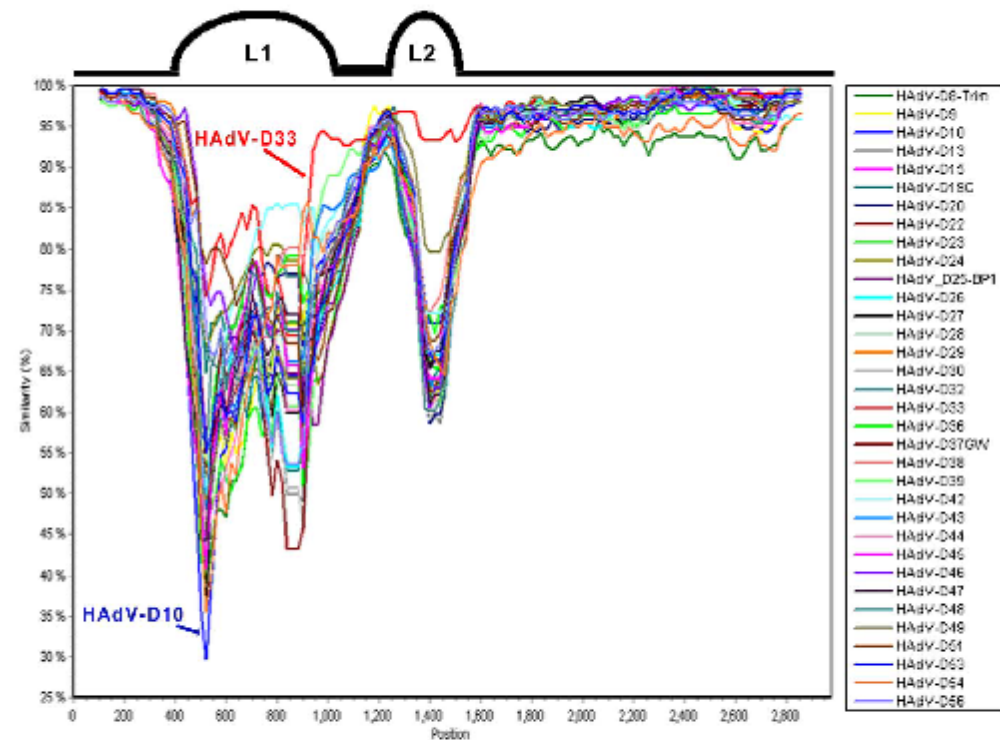


Figure 3. SimPlot analysis of the HAdV-D58 hexon coding sequence. L1 and L2 correspond to loops 1 and 2 of the hexon gene, which contain the epsilon (ϵ) fragment, an important determinant of neutralization.
doi:10.1371/journal.pone.0024491.g003

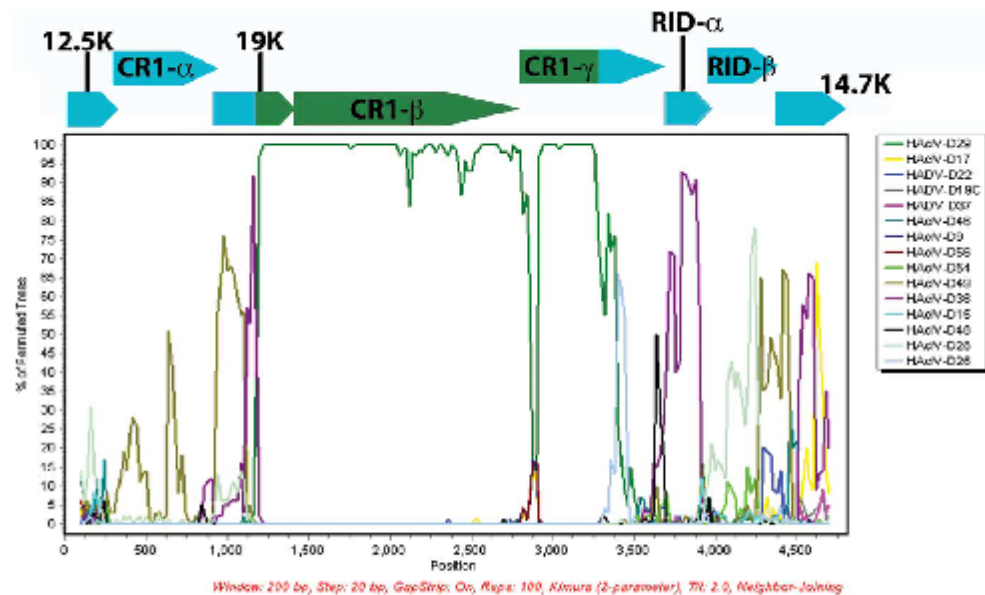


Figure 4. Computational analysis of the E3 region. SimPlot analysis of the E3 region of HAdV-D58 compared to fully sequenced E3 regions from species HAdV-D. The arrows over the Bootscan demarcate the approximate positions of the E3 coding sequences.
doi:10.1371/journal.pone.0024491.g004

Fiber recombination analysis

To determine whether or not there was recombination in the fiber gene of HAdV-D58, we performed Bootscan and SimPlot analysis using the fiber sequences in GenBank. Our results suggested the fiber gene of HAdV-D58 contains two recombination sites (Fig. 5). The first was in the middle of the shaft coding sequence and the second was in the shaft/knob boundary. The possible recombination at the shaft/knob boundary is tenuous since it is not possible to differentiate between HAdV-D25 and HAdV-D29 at this junction as evidenced by SimPlot analysis (Fig. 5B).

Phylogenetic analysis

Detailed phylogenetic analysis of completely sequenced HAdV genomes and selected coding sequences, performed with nucleotide data, confirmed that HAdV-D58 was a novel adenovirus (Figs. 6–8). The tree topology of HAdV-D58 was different depending on the protein analyzed. The whole genome sequence of HAdV-D58 was closest to HAdV-D29 (Fig. 6A). Using sequences available in GenBank, along with unpublished sequences, the penton base of HAdV-D58 grouped with HAdV-D8 "Trim", which is a prototype genome (Fig. 6B). Hexon loops 1 (L1) and 2 (L2) both clustered to HAdV-D33 (Fig. 7A and 7B), which was similar to results reported by Ferreyra et al [12]. The fiber knob was tightly linked to HAdV-D29 (Fig. 8).

Viral neutralization

Since bioinformatic analysis showed that HAdV-D58 is genetically similar to previously typed HAdVs, correlating this data to its serum neutralization profile is important. Only antiserum to HAdV-D29, at a dilution of 1:32, was able to

neutralize HAdV-D58 (Table 1). In contrast, antiserum to HAdV-D29 neutralized HAdV-D29 at 1:512 (Table 1). These results demonstrated that HAdV-D58 has a unique neutralization profile.

Serum neutralization vs. Phylogenetic analysis

A previous study proved that when the nucleotide identity of L2 in the hexon differs by $\geq 2.5\%$, a new HAdV type is suspected [13]. To provide a correlation between serum neutralization data, molecular typing (i.e., imputed serum neutralization), and phylogenomics data for the determination of a new HAdV type, the hexon L2 sequence of the proposed novel HAdV-D58 was compared against the L2 sequences of HAdV-D33, -D49, and -D38 (the closest phylogenetic relatives of the HAdV-D58 L2). The difference in percent nucleotide identity between L2 of HAdV-D33, -D49 and -D38, and that of HAdV-D58 was 10.18, 20.73, and 25.09 percent, respectively. Thus, using the L2 sequencing criteria established by Madisch et al also demonstrates that HAdV-D58 is a new type.

Discussion

In the past, human adenoviruses were characterized primarily based on their serological profile and hemagglutination properties [18]. Today the classification methods used for novel adenoviruses has been expanded to include whole genome sequencing and bioinformatic analysis [19]. We used whole genome sequencing, bioinformatic analysis, and serology to irrefutably demonstrate that HAdV-D58 is a novel human adenovirus type.

The serological and genomic characteristics of HAdV-D58 are unique. Specifically, the hexon gene of HAdV-D58 was genetically dissimilar to all known HAdV hexon genes (Fig. 3). Furthermore,

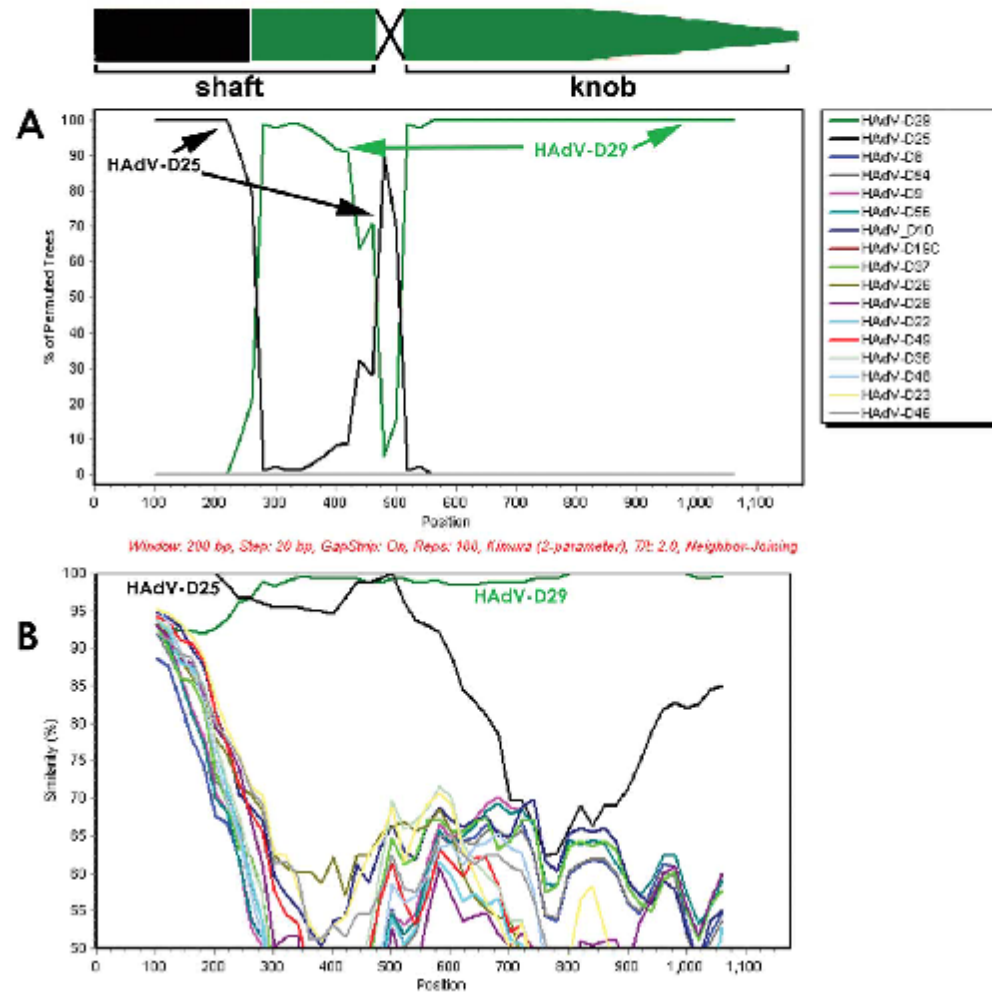


Figure 5. Computational analysis of the fiber regions. (A) Bootscan and (B) SimPlot analysis of the fiber region of HAdV-D58 compared to fully sequenced E3 and fiber regions from species HAdV-D. doi:10.1371/journal.pone.0024491.g005

this was corroborated by neutralization data that demonstrated both 16- and 64-fold differences with antiserum from HAdV-D29 and HAdV-D33, respectively (Table 1).

We found that the fiber gene of HAdV-D58 contains at least one recombination event and possibly a second (Fig. 5). The second possible recombination site is located at the shaft/knob boundary. It is not currently possible to determine if recombination happened at the shaft/knob boundary (Fig. 5B). A prior study described a recombination hotspot in the fiber gene of species HAdV-D at the shaft/knob boundary [20]. However, our Bootscan analysis on the same fiber coding sequences listed in Darr et al [20], did not reveal evidence of recombination (Fig. 9). This result was also corroborated

independently (personal communication Jason Seto). The analysis describing recombination in the fiber proteins of HAdV-D47 and HAdV-D30 utilized consensus sequences for two of four alignments [20]. The problem with this analysis is that consensus sequences do not exist in nature and could induce artificial data when introduced into recombination analysis. Furthermore, the only way to re-create the supposed recombination events (proposed by Darr et al [20] that created HAdV-D20 was by combining sequences HAdV-D20-FM210561 and HAdV-D23-FM210540, which are 100% identical (Table 2), with the 3' sequences of HAdV-D20-AJ811444 and HAdV-D23-AJ811446 (see Materials and Methods), respectively (Table 2). We were also able to recreate

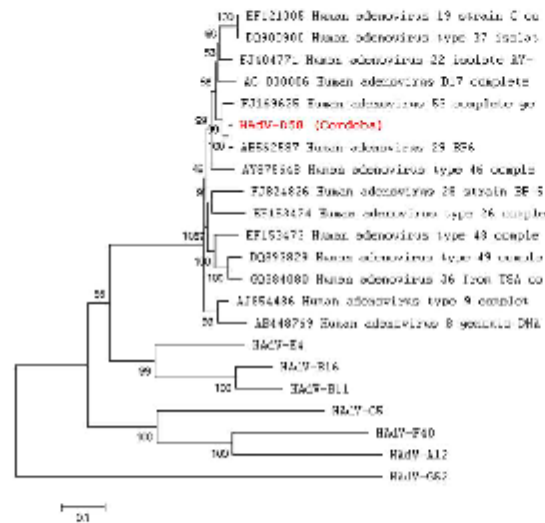
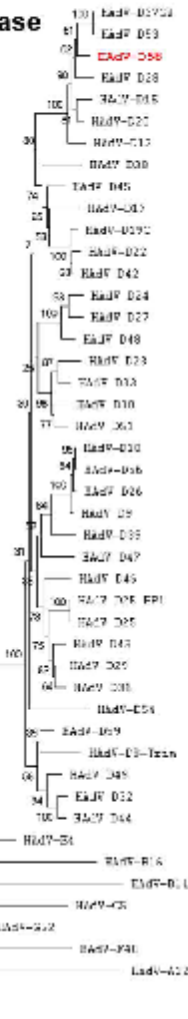
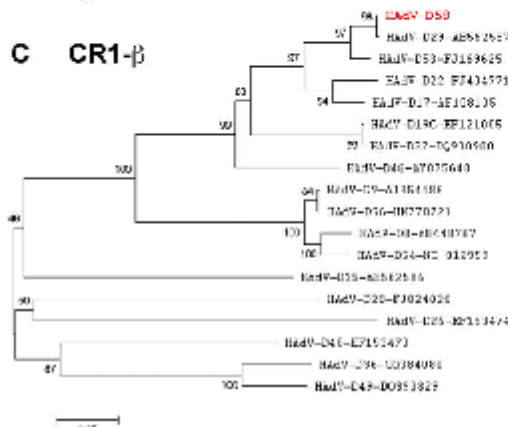
A Whole genome**B Penton base****C CR1-β**

Figure 6. Phylogenetic analysis of whole genome, penton base, and E3 CR1-β in HAdV-D58. Phylogenetic analysis is based on the nucleic acid sequence of (A) whole genomes, (B) penton base, and (C) CR1-β. Phylogenetic trees were constructed from aligned sequences using MEGA via the neighbor-joining methods and a bootstrap test of phylogeny. Bootstrap values shown at the branching points indicate the percentages of 1000 replications produced the data.
doi:10.1371/journal.pone.0024491.g006

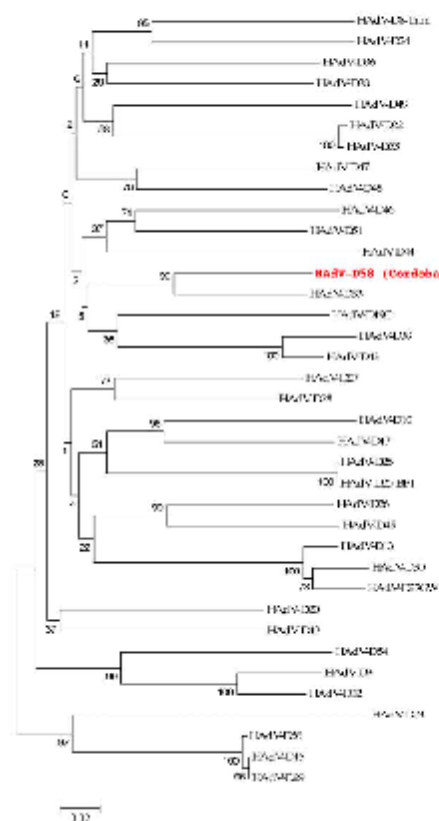
the proposed recombination event that created HAdV-D25 when we combined the sequences of HAdV-D25-FM210542 and HAdV-D25-FM210543, which are also 100% identical (Table 2), with the 3' sequences of HAdV-D25-AJ811448 and HAdV-D25-AJ811449 (see Materials and Methods), respectively (Table 2). When this data is considered together, we find no concrete evidence that the shaft/knob junction is a hot-spot for recombination.

For HAdVs, the number of E3 ORF's ranges between 6 and 9 [7,21]. HAdVs in species HAdV-D and HAdV-G contain the

49K/CR1-β ORF [7,21]. Interestingly, Bootscan analysis suggests that the E3 region of HAdV-D58 was created by recombination with HAdV-D29 (Fig. 4). However, analysis of all sequenced E3 regions in species HAdV-D demonstrates that recombination hot spots do not exist in this part of the genome for species HAdV-D (data not shown). Thus, it is difficult to speculate what advantage there is for a seemingly random recombination in the E3 region.

Phylogenetic analysis of the HAdV-D58 genome showed a close relationship to HAdV-D29, however individual proteins of HAdV-

A Hexon Loop1



B Hexon Loop2

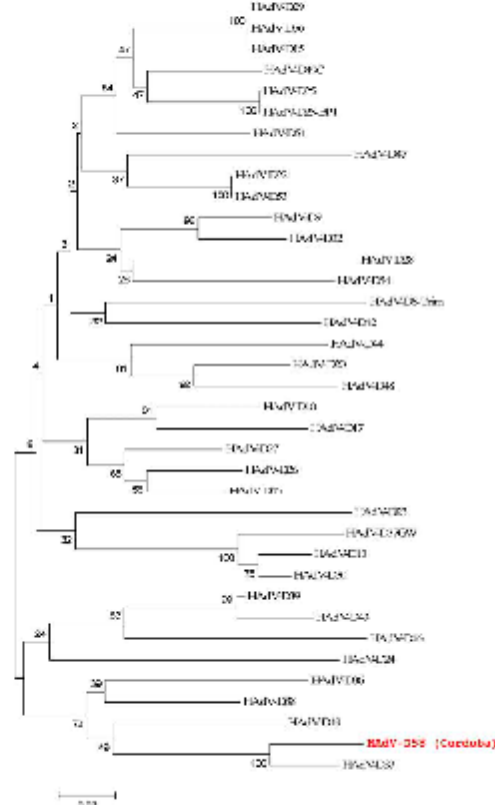


Figure 7. Phylogenetic analysis of HAdV-D58 hexon loops 1 and 2. Analysis of HAdV-D58 hexon L1 and L2 is based on the nucleic acid sequence of (A) hexon and (B) hexon L2. Phylogenetic trees were constructed from aligned sequences using MEGA, via the neighbor-joining methods and a bootstrap test of phylogeny. Bootstrap values shown at the branching points indicate the percentages of 1000 replications produced the data. doi:10.1371/journal.pone.0024491.g007

D58 clade with different types from species HAdV-D (Fig. 6A). For example, the HAdV-D58 penton, hexon, and fiber coding sequences showed close relationships to HAdV-D53, HAdV-D33, and HAdV-D29, respectively (Figs. 6B, 7, and 8). These results are consistent with results from other studies that also used bioinformatic analysis to identify novel adenovirus status to recently sequenced adenovirus genomes [3,4,5,7]. Moreover, phylogenetic analysis results are also consistent with the findings from our SimPlot analysis (Figs. 2, 4, and 5).

Conclusions

In this study, we sequenced the genome of an apparently novel adenovirus. The novel hexon coding sequence, coupled with bioinformatic analysis, demonstrated that this genome is different from all previously characterized HAdVs, and is a novel human adenovirus.

Materials and Methods

Ethics Statement

The work reported herein was performed under United States Air Force Surgeon General-approved Clinical Investigation No. FDG20040024E, by the Institutional Review Board at the David Grant USAF Medical Center. Informed Consent was not required, because we did not use clinical samples.

Viruses, cells and neutralization test

The isolation of HAdV-D58 (previously known as Ad-Cor-96-487) was previously described [12]. In brief, the stool sample was inoculated into Hep-2 cells and subcultured in Earle's MEM supplemented with 10% of fetal bovine serum (FBS), penicillin (200 U/ml), L-glutamine (2 mM), Fungizone (1 µg/ml), and streptomycin (200 µg/ml). HAdV-D58 was investigated serologi-

Fiber knob

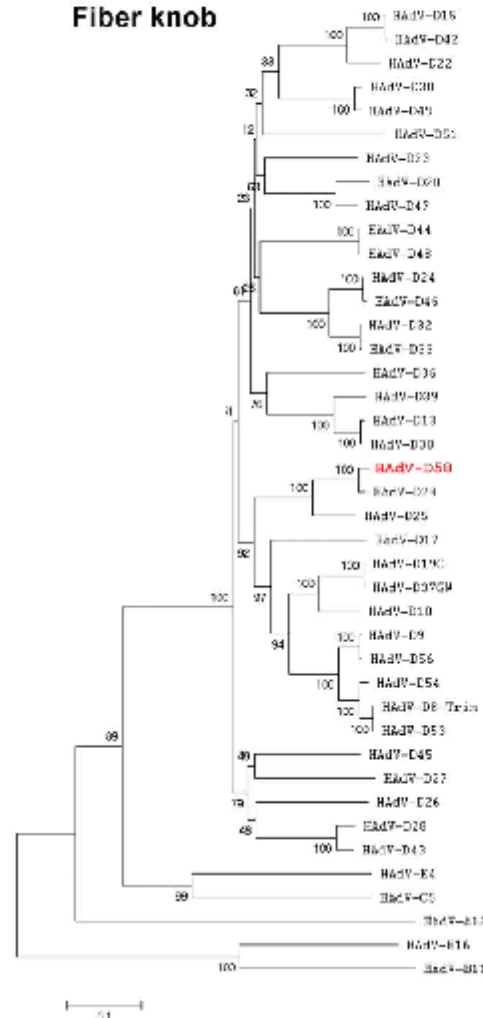


Figure 8. Phylogenetic analysis of the fiber coding sequence in HAdV-D58. Analysis of HAdV-D58 is based on the nucleic acid sequence of the fiber knob. Phylogenetic trees were constructed from aligned sequences using MEGA, via the neighbor-joining methods and a bootstrap test of phylogeny. Bootstrap values shown at the branching points indicate the percentages of 1000 replications produced the data.
doi:10.1371/journal.pone.0024491.g008

cally by viral neutralization assay (VN) using horse polydonal antisera directed against prototype strains of HAdV-D8, -D9, -D10, -D13, -D15, -D17, -D29, -D33, -D43, -D44, -D45, -D46 and -D47. VN tests were conducted on Hep-2 cells grown in 96-well microplates. The Hep-2 cells used in this study were used previously [22] and are a common cell line used for adenovirus research. Type-

Table 1. Serum neutralization of HAdV-D58 with hyper immune serum.

Antiserum	HAdV-D58	HAdV-D29	HAdV-D33
αHAdV-D8	<8		
αHAdV-D9	<8		
αHAdV-D10	<8		
αHAdV-D13	<8		
αHAdV-D15	<8		
αHAdV-D17	<8		
αHAdV-D29	32	512	
αHAdV-D33	<8		512
αHAdV-D43	<8		
αHAdV-D44	8		
αHAdV-D45	<8		
αHAdV-D46	8		
ααHAdV-D47	<8		

doi:10.1371/journal.pone.0024491.t001

specific antisera were inactivated at 56°C for 30 min and serially diluted twofold, 50 µl per well with four replicate wells per dilution. A working dilution of virus (HAdV-D58) containing 100 TCID₅₀ in 50 µl was added to each well, and the plates were incubated at 37°C in 5% CO₂ for 1 h. During the incubation period, Hep-2 cells were trypsinized and resuspended at 5 × 10⁶ cells per ml. After the incubation, 100 µl of cell suspension were added to each well. The contents of each well were mixed, and the plates were incubated at 37°C in 5% CO₂ for 6 days. After 6 days, the medium was removed and cells were stained with crystal violet solution (1.46 g crystal violet, 50 ml ethanol, 300 ml formaldehyde, 650 ml distilled water). The neutralization titer was calculated as the maximum dilution of antiserum that completely inhibited viral growth as evidenced by the lack of cytopathic effects.

Nomenclature

This virus was named HAdV-D58 because the number 57 was already taken in GenBank (HQ003817). For rules of adenovirus nomenclature, see <http://hadwv.gmu.edu/>.

Nucleotide sequence accession numbers

The HAdV-D58 genome and annotation have been deposited in GenBank prior to manuscript submission; accession number HQ883276. The following HAdV genomes (GenBank accession numbers) were used for comparative analysis: HAdV-D8 (AB448767), HAdV-D9 (AJ854486), HAdV-D19C (EF121005), HAdV-D22 (FJ404771), HAdV-D26 (EF153474), HAdV-D28 (FJ824826), HAdV-D36 (GQ384080), HAdV-D37 (DQ900900), HAdV-D46 (AY875648), HAdV-D48 (EF153473), HAdV-D49 (DQ393829), HAdV-D53 (FJ169625), HAdV-D54 (AB333801), and HAdV-D56 (HM770721).

Amplification of the HAdV-D58 genome

To amplify regions of HAdV-D58 flanking the sequences previously described by Ferreyra et al. [12], we designed primers based on conserved adenovirus sequences in species HAdV-D. All amplicons were then sequenced using primer walking. The genome was assembled using SeqMan, which is an assembly program inside of the Lasergene 8 software suite.

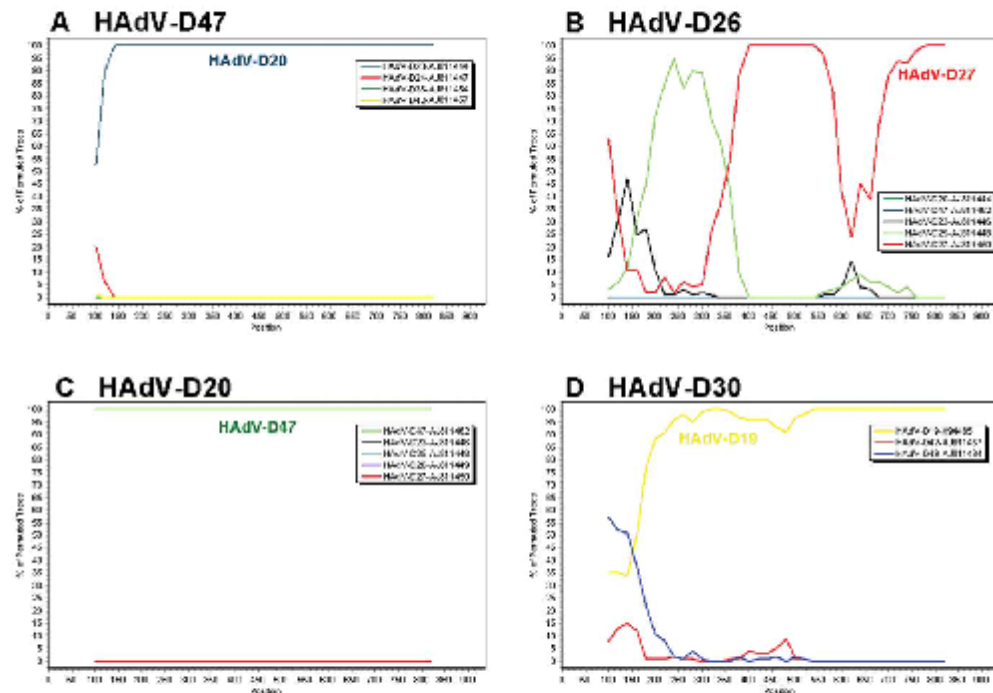


Figure 9. Bootscan analysis of selected fiber genes in species HAdV-D. (A) HAdV-D47, (B) -D26, (C) -D20, and (D) -D30. This figure is a corrected repeat of Figure 2 in Darr et al [20]. doi:10.1371/journal.pone.0024491.g009

Nucleic Acid Isolation

HAdV-D58 particles were separated from Hep-2 cells by ultracentrifugation. Genomic DNA was acquired from viral particles using *AcuPrep* Genomic DNA Extraction Kit (Bioneer Corporation). Finally, the viral DNA was resuspended in deionized water and stored at -20°C until use.

Bioinformatics

The available genomes from species HAdV-D were aligned using the clustalW [23] alignment method which is available through a

web interface at <http://www.ebi.ac.uk/Tools/clustalw2/index.html>. The default parameters for gap open penalty and gap extension penalty were used.

Hexon coding sequences used for analysis were: HAdV-D8 (AB448767), HAdV-D9 (AJ854486), HAdV-D10 (AB369368), HAdV-D13 (DQ149616.1), HAdV-D15 (AB330096.1), HAdV-D19C (AB448774), HAdV-D20 (AB330101.1), HAdV-D22 (FJ619037), HAdV-D24 (AB330105.1), HAdV-D25 (AB330106.1), HAdV-D26 (EF153474), HAdV-D27 (AB330108.1), HAdV-D28 (FJ824826), HAdV-D29 (AB562587), HAdV-D30 (AB330111.1), HAdV-D32

Table 2. Comparison of the nucleotide sequences used by Darr et al to show recombination events in the fiber/knob junction.

Sequences compared	Nucleotide identity
HAdV-D20-FM210561 vs. HAdV-D23-FM210540	100%
HAdV-D20-AJ811444 vs. HAdV-D23-AJ811446	73%
HAdV-D20-AJ811444 vs. the last 442 base pairs of HAdV-D20 used in Fig. 2C from Darr et al.	100%
HAdV-D23-AJ811446 vs. the last 442 base pairs of HAdV-D23 used in Fig. 2C from Darr et al.	100%
HAdV-D25-FM210542 vs. HAdV-D26-FM210543	100%
HAdV-D25-AJ811448 vs. HAdV-D26-AJ811440	70%
HAdV-D25-FM210542 vs. the last 442 base pairs of HAdV-D25 used in Fig. 2B from Darr et al.	100%
HAdV-D26-FM210543 vs. the last 442 base pairs of HAdV-D26 used in Fig. 2B from Darr et al.	100%

doi:10.1371/journal.pone.0024491.t002

(AB330113.1), HAdV-D33 (AB330114.1), HAdV-D36 (GQ394080), HAdV-D37 (DQ900900), HAdV-D38 (AB330119.1), HAdV-D39 (AB330120.1), HAdV-D42 (AB330123.1), HAdV-D43 (AB330124.1), HAdV-D44 (AB330125.1), HAdV-D45 (AB330126.1), HAdV-D46 (AY875648), HAdV-D47 (AB330128.1), HAdV-D48 (EF153473), HAdV-D49 (DQ393829), HAdV-D51 (AB330132.1), HAdV-D53 (FJ169625.1), HAdV-D54 (AB333801) and HAdV-D56 (FM770721). SimPlot software was used to complete a bootscan analysis of the aligned hexon genes of the available HAdV-D genomes [24]. The default settings for window size, a step size, replicates used, gap stripping, distance model, and tree model were, respectively, 200, 20, 100, "on", "Kimura", and "Neighbor Joining". The HAdV-D58 hexon was chosen as the reference sequence for the analysis [24].

Fiber coding sequences used for analysis were HAdV-D8 (AB448767), HAdV-D9 (AJ854486), HAdV-D10 (AB369368), HAdV-D19C (AB448774), HAdV-D19a (CS301726), HAdV-D22 (FJ619037), HAdV-D26 (EF153474), HAdV-D28 (FJ824826), HAdV-D29 (AB562587), HAdV-D36 (GQ394080), HAdV-D37 (DQ900900), HAdV-D46 (AY875648), HAdV-D48 (EF153473), HAdV-D49 (DQ393829), HAdV-D54 (AB333801), and HAdV-D56 (FM770721). Fiber genes from species HAdV-D genomes were aligned using ClustalW [23]. The default gap opening and gap extension penalties were used (15.0 and 6.66).

To analyze the results by Darr et al [20], we used fiber genes HAdV-D19 (X94485), HAdV-D20 (AJ811444), HAdV-D23 (AJ811446), HAdV-D24 (AJ811447), HAdV-D25 (AJ811448), HAdV-D26 (AJ811449), HAdV-D27 (AJ811450), HAdV-D30 (AF473933), HAdV-D36 (AJ811454), HAdV-D42 (AJ811457), HAdV-D47 (AJ811462), and HAdV-D49 (AJ811464). To recreate the results acquired by Darr et al in figure 2C, we amalgamated HAdV-D20-FM210561 to nucleotides 375 through 816 of HAdV-D20-AB811444.1, HAdV-D23-FM210540 to nucleotides 364 through 821 of HAdV-D23-AJ811446, and HAdV-D26-FM210543 to nucleotides 373 through 799 of HAdV-D26-AJ811449. Next, we generated an alignment using the amalgamated sequences of HAdV-D20, -D23, and -D26 with AJ811448, AJ811450, and AJ811462, then performed BootScans with SimPlot. Furthermore, figures 2A and 2D of Darr et al used several "consensus sequences" which contain wobble bases [20].

References

- Hileman MR, Werner JH, Gauld RL (1953) Influenza antibodies in the population of the USA: an epidemiological investigation. *Bull World Health Organ* 8: 613-631.
- Rowe WP, Hasbner RJ, Gilmore LJ, Permet RH, Ward TG (1953) Isolation of a cytopathogenic agent from human adenoids undergoing spontaneous degeneration in tissue culture. *Proc Soc Exp Biol Med* 84: 570-573.
- Walsh MP, Seto J, Jones MS, Chodosh J, Xu W, et al. (2010) Computational analysis identifies human adenovirus type 55 as a re-emergent acute respiratory disease pathogen. *J Clin Microbiol* 48: 991-993.
- Robinson CM, Singh G, Henquell C, Walsh MP, Peigne-Lafaille H, et al. (2010) Computational analysis and identification of an emergent human adenovirus pathogen implicated in a respiratory fatality. *Virology* 409: 141-147.
- Walsh MP, Chintakundawar A, Robinson CM, Madiach I, Herrach B, et al. (2009) Evidence of molecular evolution driven by recombination events influencing tropism in a novel human adenovirus that causes epidemic keratoconjunctivitis. *PLoS One* 4: e5635.
- Ishiko H, Aoki K (2009) Spread of epidemic keratoconjunctivitis due to a novel serotype of human adenovirus in Japan. *J Clin Microbiol* 47: 2678-2679.
- Jones MS, 2nd, Herrach B, Casac RD, Gorman MM, Dela Cruz WP, et al. (2007) New adenovirus species found in a patient presenting with gastroenteritis. *J Virol* 81: 5978-5984.
- Echavria M (2009) Adenoviruses. In: Zuckerman AJ, Banatvala JE, Pattison JR, eds. *Principles and Practice of Clinical Virology*. Sixth Edition ed. Chichester: John Wiley & Sons, pp 463-488.
- Robinson CM, Shariati F, Zaitchik J, Gilguy AF, Dyer DW, et al. (2009) Human adenovirus type 19: genomic and bioinformatics analysis of a keratoconjunctivitis isolate. *Virus Res* 139: 122-126.

Phylogenetic analysis of HAdV-D58

Nucleotide alignment of whole genome, penton, L1 and L2 of hexon, hexon, fiber and fiber knob, ORF2 and ORF3 regions were performed using the multiple alignment software based on a "Fast Fourier Transform" algorithm (MAFFT, <http://www.ebi.ac.uk/Tools/mafft/index.html>). Phylogenetic distances and trees were generated from these aligned sequences by the Molecular Evolutionary Genetic Analysis software (MEGA v4.1; <http://www.megasoftware.net/>). Distances were obtained with pairwise distance calculation using maximum composite likelihood model. Subsequent phylogenetic trees were obtained using bootstrap tests of phylogeny of 1000 replicates with neighbor-joining method featured in the program.

Supporting Information

Table S1 Percent identities of the nucleotide coding sequences of loop1 (L1) and loop2 (L2) HAdV-D58 coding regions to homologous sequences from other viruses in species HAdV-D.
(PDF)

Table S2 Percent identities of the nucleotide coding sequences of selected E3 HAdV-D58 coding sequences and their homologs.
(PDF)

Acknowledgments

We thank Carl Gibbins for help in sequencing this virus. The views expressed in this material are those of the authors, and do not reflect the official policy or position of the U.S. Government, the Department of Defense, or the Department of the Air Force.

Author Contributions

Conceived and designed the experiments: DS MSJ LF. Performed the experiments: NRH EBL SVN JVP DT. Analyzed the data: SIF MSJ DS EBL JC DWD. Contributed reagents/materials/analysis tools: MSJ DS JC DWD. Wrote the paper: MSJ DS JC DWD LF.

- Metzger D, Gibbins C, Hadem NR, Jones MS (2010) Evaluation of multiplex type-specific real-time PCR assays using the LightCycler and joint biological agent identification and diagnostic system platform for detection and quantitation of adult human respiratory adenoviruses. *J Clin Microbiol* 48: 1397-1403.
- Fremont PD (2003) Mortality due to gastroenteritis of unknown etiology in the United States. *J Infect Dis* 187: 441-452.
- Ferreira I, Giordano M, Martinez I, Ito MR, Baril P, et al. (2010) A novel human adenovirus hexon protein of species D found in an AIDS patient. *Arch Virol* 155: 27-35.
- Madiach I, Harste G, Pommer H, Albert H (2005) Phylogenetic Analysis of the Main Neutralization and Hemagglutination Determinants of All Human Adenovirus Prototypes as a Basis for Molecular Classification and Taxonomy. *Journal of Virology* 79: 15265-15276.
- Sester M, Koebnick K, Owen D, Ao M, Bromberg V, et al. (2010) Conserved Amino Acid within the Adenovirus 2 E3/19K Protein Differentially Affect Downregulation of MHC Class I and MICB/B Proteins. *Journal of Immunology* 184: 255-267.
- McSharry BP, Burgert H-G, Owen DP, Stanton RJ, Prod'homme V, et al. (2008) Adenovirus E3/19K promotes evasion of NK cell recognition by intracellular sequestration of the NK2D ligand major histocompatibility complex class I chain-related protein A and B. *Journal of Virology* 82: 4385-4394.
- Schneider-Brachet W, Tchikv V, Merkle O, Jakob M, Halks C, et al. (2006) Inhibition of TNF receptor 1 internalization by adenovirus 14.7K as a novel immune escape mechanism. *Journal of Clinical Investigation* 116: 2901-2913.
- Elving A, Burgert HG (1998) The adenovirus E3/10.4K-14.5K proteins down-modulate the apoptosis receptor Fas/Apo-1 by inducing its internalization. *Proc Natl Acad Sci U S A* 95: 10072-10077.

18. Hienhofer JC, Stone VO, Broderick JR (1991) Antigenic relationships among the 47 human adenoviruses determined in reference home antisera. *Arch Virol* 121: 179–197.
19. Torres S, Chodosh J, Seto D, Jones MS (2010) The Revolution in Viral Genomics as Exemplified by the Bioinformatic Analysis of Human Adenoviruses. *Virology* 2: 1367–1381.
20. Durr S, Madlch I, Hofmayer S, Rohm F, Heim A (2009) Phylogeny and primary structure analysis of fiber shafts of all human adenovirus types for rational design of adenoviral gene-therapy vectors. *J Gen Virol* 90: 3840–3854.
21. Windheim M, Högendorf A, Bugert HG (2004) Immune evasion by adenovirus E3 proteins: exploitation of intracellular trafficking pathways. *Current Topics in Microbiology and Immunology* 273: 29–85.
22. Li H, Zhou R, Chen J, Tian X, Zhang Q, et al. (2009) A recombinant replication-defective human adenovirus type 3: a vaccine candidate. *Vaccine* 27: 116–122.
23. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, et al. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23: 2947–2948.
24. Lele KS, Bollinger RC, Paranjape RS, Gadgil D, Kulkarni SS, et al. (1999) Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J Virol* 73: 152–160.

CHAPTER 5 – HUMAN ADENOVIRUS TYPE 59

Introduction

Human adenovirus type 59 (HAdV-D59) is a novel human respiratory pathogen that was isolated from an AIDS patient's bronchoalveolar lavage (BAL) biopsy sample. The patient presented with a fever, cough, tachycardia and expiratory wheezes. Using bioinformatics and genomic analysis, the genome sequence was examined for insight into its molecular origins. The results suggest that HAdV-D59 is an emergent pathogen with a molecular evolution pathway that includes multiple recombination events in the penton base, hexon and fiber genes. These were apparently transferred from parental genomes HAdV-D19C, HAdV-D25 and HAdV-D56 respectively. Furthermore, serological analysis shows a neutralization profile for the hexon epitope that is similar to but not identical to HAdV-D25, suggesting a divergent but common ancestor.

Computational and Serologic Analysis of Novel and Known Viruses in Species Human Adenovirus D in Which Serology and Genomics Do Not Correlate

Elizabeth B. Liu^{1*}, Debra A. Wadford^{2*}, Jason Seto¹, Maria Vu², Nolan Ryan Hudson³, Lisa Thrasher³, Sarah Torres³, David W. Dyer⁴, James Chodosh⁵, Donald Seto¹, Morris S. Jones^{1*}

1 School of Systems Biology, George Mason University, Manassas, Virginia, United States of America, **2** Viral and Rickettsial Disease Laboratory, California Department of Public Health, Richmond, California, United States of America, **3** Clinical Investigation Facility, David Grant USAF Medical Center, Travis AFB, Fairfield, California, United States of America, **4** Department of Microbiology and Immunology, University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma, United States of America, **5** Howe Laboratory, Massachusetts Eye and Ear Infirmary, Department of Ophthalmology, Harvard Medical School, Boston, Massachusetts, United States of America

Abstract

In November of 2007 a human adenovirus (HAdV) was isolated from a bronchoalveolar lavage (BAL) sample recovered from a biopsy of an AIDS patient who presented with fever, cough, tachycardia, and expiratory wheezes. To better understand the isolated virus, the genome was sequenced and analyzed using bioinformatic and phylogenomic analysis. The results suggest that this novel virus, which is provisionally named HAdV-D59, may have been created from multiple recombination events. Specifically, the penton, hexon, and fiber genes have high nucleotide identity to HAdV-D19C, HAdV-D25, and HAdV-D56, respectively. Serological results demonstrated that HAdV-D59 has a neutralization profile that is similar yet not identical to that of HAdV-D25. Furthermore, we observed a two-fold difference between the ability of HAdV-D15 and HAdV-D25 to be neutralized by reciprocal antiserum indicating that the two hexon proteins may be more similar in epitopic conformation than previously assumed. In contrast, hexon loops 1 and 2 of HAdV-D15 and HAdV-D25 share 79.13 and 92.56 percent nucleotide identity, respectively. These data suggest that serology and genomics do not always correlate.

Citation: Liu EB, Wadford DA, Seto J, Vu M, Hudson NR, et al. (2012) Computational and Serologic Analysis of Novel and Known Viruses in Species Human Adenovirus D in Which Serology and Genomics Do Not Correlate. PLoS ONE 7(3): e33212. doi:10.1371/journal.pone.0033212

Editor: Michael A. Barry, Mayo Clinic, United States of America

Received: July 23, 2011; **Accepted:** February 12, 2012; **Published:** March 13, 2012

Copyright: © 2012 Liu et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This research was financially supported by grant R01EY013124 (DWD, DS, MSJ and JC) and P30EY014104 (JC, LS, NRH, and MSJ) were supported by United States Air Force Surgeon General grant FDG20040024E. JC is also funded by an unrestricted grant to the Department of Ophthalmology, Harvard Medical School, from Research to Prevent Blindness, Inc. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: dmorris@yahoo.com

† These authors contributed equally to this work.

Introduction

The first human adenoviruses (HAdVs) were isolated in 1953 from a military basic trainee and the adenoid tissue of a pediatric patient as respiratory pathogens [1]. Presently there are greater than 60 types that have been isolated and characterized either with serological methods or, more recently, genomic methods [2,3,4,5,6]. HAdVs are classified into the Mastadenovirus genus and are further parsed into seven species (A–G) [3,4,5,6,7,8]. Originally, HAdV types were identified, characterized and classified based on serum neutralization and hemagglutination inhibition assays among other biological attributes [9]; however recently, bioinformatics and genomic analysis of the whole genome have replaced serology-based methods for typing novel HAdVs [3,4,5,6,8,10]. At the nucleotide level, the members of each adenovirus species are highly similar to each other, and do not commonly recombine with members of other species. The species groupings, in part, may reflect the cell tropism of the viruses, as well as the resulting symptoms and diseases caused by the individual HAdV types. For example, species HAdV-B1 viruses are known to cause respiratory infections of the lower lung [11] whereas viruses in species HAdV-D can cause ocular disease, including epidemic keratoconjunctivitis [12], and gastrointestinal disease [3,13].

In this report, an adenovirus isolated from a bronchoalveolar lavage (BAL) sample that was biopsied from an AIDS patient who presented with fever, cough, tachycardia and expiratory wheezes is examined using genomics and bioinformatics. Based upon the whole genome analysis and supported by limited serological data, this adenovirus belongs to species HAdV-D, and is a 'never seen before' novel virus, to be given the name of HAdV-D59.

Materials and Methods

Ethics Statement

The work reported herein was performed under United States Air Force Surgeon General-approved Clinical Investigation No. FDG20040024E, by the Institutional Review Board at the David Grant USAF Medical Center. Informed Consent was not required, because we did not use clinical samples.

Viruses, cells, and serum neutralization assay

Adenovirus neutralization assays were run as previously described [14]. Briefly, serotyping of adenovirus isolates was performed using a standard dose of virus against specific rabbit

antisera raised against reference stock adenoviruses types 1-49 from the collection maintained by the Viral and Rickettsial Disease Laboratory of the California Department of Public Health, Richmond, CA. Reference viruses were originally obtained from the reporting investigators: the Research Reference Reagents Branch, National Institutes of Health; the Respiratory Viral Disease Unit, Centers for Disease Control and Prevention; or the American Type Culture Collection. Stock virus cultures were passaged in A549 cells (American Type Culture Collection, Rockville, MD), the cells were disrupted by vortexing, and cell-free supernatant fluid was then frozen at -70°C .

Equal volumes of diluted virus and immune serum were mixed and incubated for one hour in 5% CO₂ at 37°C. Thereafter A549 cells were added, mixed, and incubated at 37°C in 5% CO₂ for 7 days. Each assay contained a back titration of the virus used. Living cells were distinguished from dead cells by measuring the amount of Finter's Neutral Red [15] present as indicated by absorbance at 550 nm using a microplate spectrophotometer (Bio-Tek Instruments, Winooski, VT). Virus neutralization titers were determined by equating cell death to virus growth (no virus neutralization). Neutralization was plotted as a percentage of cell control absorbance, to determine endpoint virus and serum titers. Three independent experiments were run yielding similar results.

Nucleotide sequence accession numbers

The HAdV-D59 genome sequence and its annotation are deposited in GenBank and retrievable as accession number JF799911. In addition, the following HAdV genomes (GenBank accession numbers) were used for comparative computational analyses: HAdV-D8 (AB448767), HAdV-D9 (AJ554486), HAdV-D15 (AB562586), HAdV-D17 (AF108105), HAdV-D19C (EF121005), HAdV-D22 (FJ404771), HAdV-D25 (unpublished), HAdV-D26 (EF153474), HAdV-D28 (FJ248262), HAdV-D36 (GQ384008), HAdV-D37 (DQ909000), HAdV-D46 (AY875648), HAdV-D48 (EF153473), HAdV-D49 (DQ393829), HAdV-D53 (FJ169625), HAdV-D54 (AB333801), HAdV-D56 (HM770721), and HAdV-D58 (HQ883276).

Fiber coding sequences used for analysis were as follows: HAdV-D8 (AB448767), HAdV-D9 (AJ854486), HAdV-D10 (AB369368), HAdV-D19C (AB448774), HAdV-D19a

IC3301 726), HdV-D22 (FJ619037), HAdV-D26 (EF153474), HAdV-D27 (FJ824826), HAdV-D29 (AB562587), HAdV-D36 (GQ384080), HAdV-D37 (DQ090800), HAdV-D46 (A Y875648), HAdV-D48 (EF153473), HAdV-D49 (DQ393829), HAdV-D54 (AB333801), HAdV-D56 (FM770721), and HAdV-D58 (HQ883276). Fiber genes from species HAdV-D genomes were aligned using ClustalW [16]. For this analysis, the default gap opening and gap extension penalties were 15.0 and 6.66.

Amplification and DNA sequencing of the HAdV-D59 genome

To amplify regions of HAdV-D59 using the polymerase chain reaction (PCR) protocol, conserved adenovirus sequences in species HAdV-D were used to design primers. All amplicons were then sequenced on an ABI 3130xl using a primer walking strategy. The HAdV-D59 genome was sequenced to 8-fold coverage following PCR amplification, with both strands represented.

Bioinformatics

The HAIV-D59 genome was compared against a select number of viral genomes from the HAIV-D group based on its GC content, which is indicative of HAIV species. The selection of which genomes was based on initial overall high nucleotide identity to HAIV-D59. The data presented are final iterations of analyses that initially included all of the sequenced genomes in species HAIV-D.

Recombination analysis

Whole genome sequences of HAdV-D59 and members of *se* species HAdV-D were first aligned with kalign (<http://www.zib.ac.uk/Tools/msd/kalign/>) for a broad perspective of the genome. SimPlot [17] was then used to construct a Bootscan analysis of the aligned sequences. The window size and step size were set to 1000 and 200 respectively.

Following this, to provide a detailed close inspection of recombination events, the penton base gene, hexon gene, E3 coding region and fiber gene from the HAdV-D genomes were aligned to their counterparts using ClustalW [16]. This was also followed by recombination analysis using SimPlot with the window size and step size set to 250 and 50, respectively.

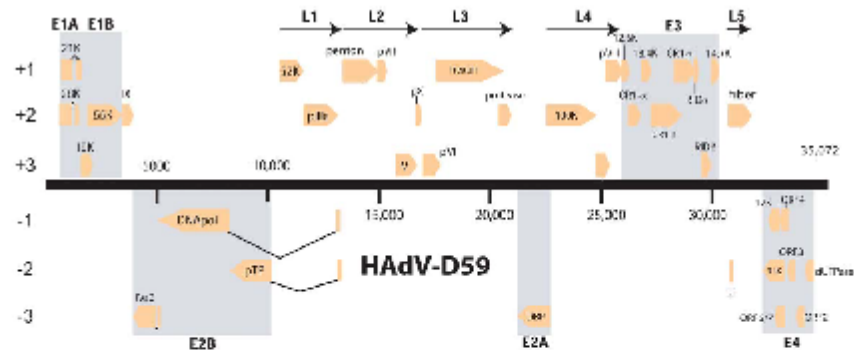


Figure 1. Genome organization of HAdV-D59. The HAdV-D59 genome is represented by a black horizontal line marked at 5-kbp intervals. Protein encoding regions are shown as arrows indicating transcriptional orientation above and below the genome. Spliced genes are indicated by V-shaped lines.

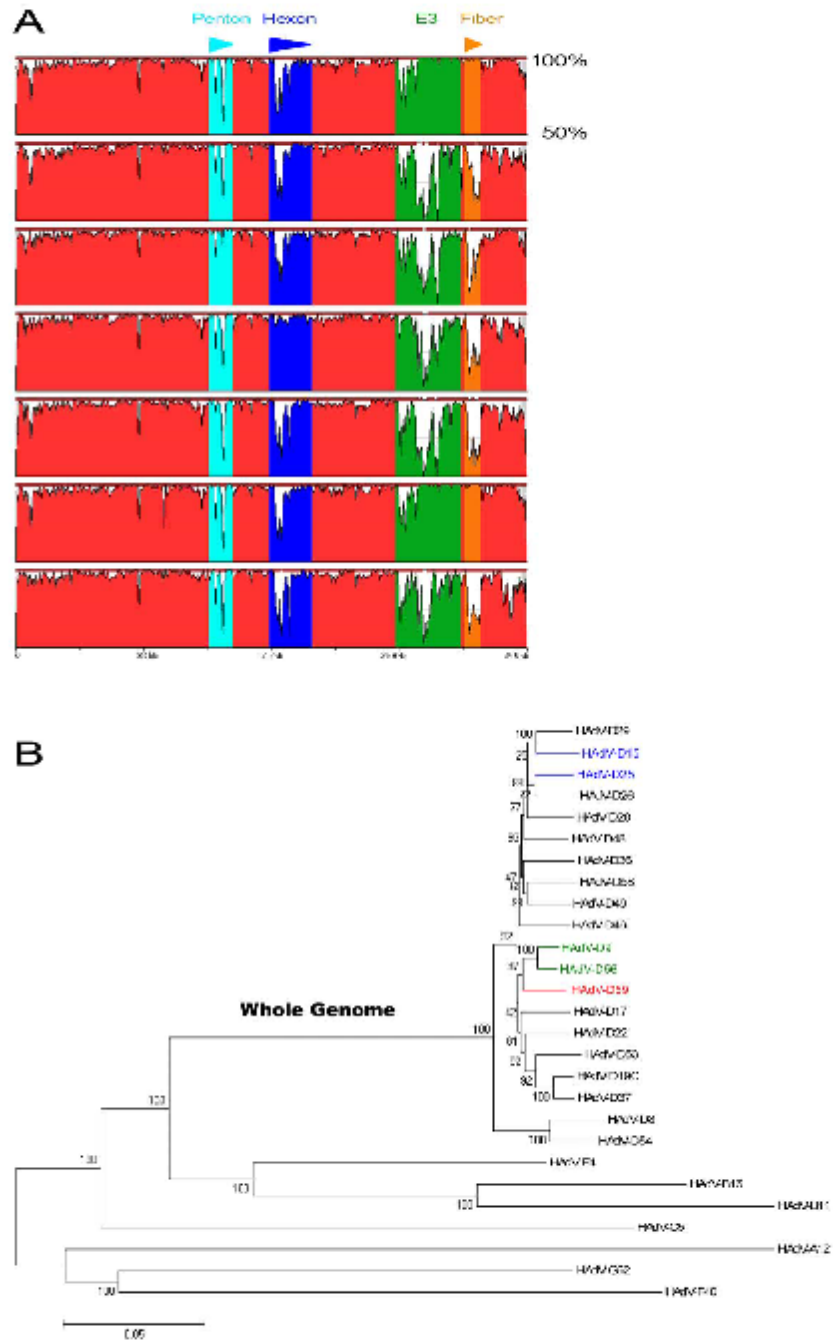


Figure 2. Comparative genomic analysis. (A) Pairwise nucleotide comparison of selected HAdV-D genomes to HAdV-D59 using zPicture. The arrows above the x-axis demarcate the positions of penton base, hexon, E3 region, and fiber coding sequences in the genome of HAdV-D59. The y-axis notes the percent identity. HAdV-D9, HAdV-D22, HAdV-D19C, HAdV-D25, HAdV-D28, HAdV-D36, and HAdV-D56 were used for comparison to HAdV-D59 because they share high nucleotide identity to the aforementioned virus in different sections of their genomes. (B) Whole genome phylogenetic analysis. The phylogenetic tree was constructed from aligned sequences using MEGA, via the neighbor-joining methods and a bootstrap test of phylogeny. Bootstrap values shown at the branching points indicate the percentages of 1000 replications produced the clade. A Bootstrap value of 70 and above is considered to be robust. doi:10.1371/journal.pone.0033212.g002

Percent Identity

Whole genome, penton base gene, hexon gene, E3 coding region and fiber gene nucleotide sequences of HAdV-D59, along with members of species HAdV-D, were aligned using kalign; these were then compared to each other based on percent identity values calculated with Chimera [18].

Phylogenomic analysis of HAdV-D59

Sequence alignments for phylogenomic analysis were generated using the kalign method noted earlier. Phylogenetic trees were constructed from these aligned sequences using Molecular Genetic Analysis Software (MEGA 4.1; <http://www.megasoftware.net>), via neighbor-joining methods and bootstrap test of phylogeny with replicates set to 1000.

Results

Clinical Investigation

In November 2007, an AIDS patient was admitted to San Francisco General Hospital, presenting with fever. The patient also complained of a cough productive of yellow sputum and blood. Clinical examination revealed a body temperature of 101°F, tachycardia and expiratory wheezes. During the hospital stay, a CT scan displayed results suggestive of a cavitary lung lesion. This prompted a bronchoalveolar lavage (BAL) for a diagnostic specimen (via bronchoscopy). A virus was cultured from the BAL sample and sent to the California Department of Public Health (CDPH) for further analysis and identification. No other pathogens were isolated from this patient.

The virus was propagated at the Viral and Rickettsial Disease Laboratory at the California Department of Public Health, and was identified as an unknown adenovirus by serum neutralization assay. Initial sequence analysis of amplicons derived from the hexon and the fiber genes revealed similarity to gene sequences from HAdV-D25 and HAdV-D56, respectively. The possibility

that this virus might represent a novel, recombinant pathogen provoked whole genome analysis in order to characterize this isolate more thoroughly.

Amplification, sequencing, and genetic characteristics of the novel adenovirus

To elucidate the genetic characteristics of this pathogen (HAdV-D59), we sequenced and analyzed the entire genome. The genome length of HAdV-D59 is 35,072 base pairs (Figure 1), with a base composition of 22.4% A, 20.4% T, 28.5% G, 28.7% C. The GC content of 57.2% is consistent with the values found for members of species HAdV-D (mean of 57.0%). The organization of the 36 open reading frames (ORFs) that were annotated had a genome organization similar to other mastadenoviruses (Fig. 1). The inverted terminal repeat (ITR) sequences for HAdV-D59 were determined to be 151 bp in length. Within species HAdV-D, HAdV-D59 has a genome percent identity ranging from a low of 92.06% (HAdV-D8) to 96.48% (HAdV-D9).

Genome Analysis

Since initial DNA sequencing suggested evidence of recombination, we analyzed the HAdV-D59 genome using zPicture, a dynamic blastz alignment visualization program designed for comparative analysis (<http://zpicture.dcode.org/>). Comparisons of the HAdV-D59 genome with the whole genome sequences of HAdV-D9, -D22, -D19C, -D25, -D28, -D36, -D56 and -D58 were performed. Consistent with other previously reported viruses in species HAdV-D [4,19], HAdV-D59 shows heterogeneity in the penton base, hexon, E3, and fiber coding sequences (Fig. 2A). Pairwise alignment suggested that these regions had highest nucleotide identities with sequences from HAdV-D9, HAdV-D19C, HAdV-D25, HAdV-D28, and HAdV-D56 (Table 1).

Comprehensive phylogenomic analyses of whole genome HAdVs were performed. Using sequences available in GenBank as well as the unpublished sequence of HAdV-D25, the whole genome phylogenetic tree analysis resulted in a subclade that includes HAdV-D59, HAdV-D9, and HAdV-D56 with a high confidence bootstrap value of 97 (Fig. 2B).

Penton Base Gene Analysis

Recently it was shown that two coding sequences for the external hypervariable loops in the penton base gene contain hotspots for recombination in species HAdV-D [20]. Analysis of the primary amino acid sequences in species HAdV-D showed that the most similar loop1 sequences (nucleotides 200–600) to that of HAdV-D59 were HAdV-D28 and HAdV-D36 (80.95%). In addition, the most similar RGD loop (nucleotides 650–1150) to HAdV-D59 was HAdV-D22 with 100% amino acid identity. Bootscan analysis [17] with penton base sequences from species HAdV-D confirmed the aforementioned relationships (Fig. 3). Phylogenetic analysis of the HAdV-D59 penton base hypervariable loop 1 also confirmed that it is a close relative of HAdV-D28 and HAdV-D36 with a robust bootstrap value of 91 (Fig. 3B). Phylogenetic analysis also demonstrated that the HAdV-D59

Table 1. Percent identities of the nucleotide coding sequences of selected HAdV-D59 coding regions to homologous sequences from other viruses in species HAdV-D.

	Penton	Hexon L1	Hexon L2	E3	Fiber
HAdV-D9	92.63%	79.14%	85.69%	96.35%	99.54%
HAdV-D15	93.03%	78.10%	91.48%	80.57%	77.45%
HAdV-D19C	93.03%	75.41%	88.89%	80.52%	76.59%
HAdV-D22	97.55%	78.23%	81.85%	83.02%	70.75%
HAdV-D25	92.64%	96.48%	95.91%	82.82%	71.06%
HAdV-D28	92.31%	77.44%	82.22%	77.61%	69.97%
HAdV-D36	93.60%	75.59%	75.19%	78.30%	69.89%
HAdV-D56	92.44%	78.23%	91.11%	95.98%	99.82%
HAdV-D58	92.63%	74.97%	77.29%	82.88%	70.61%

doi:10.1371/journal.pone.0033212.t001

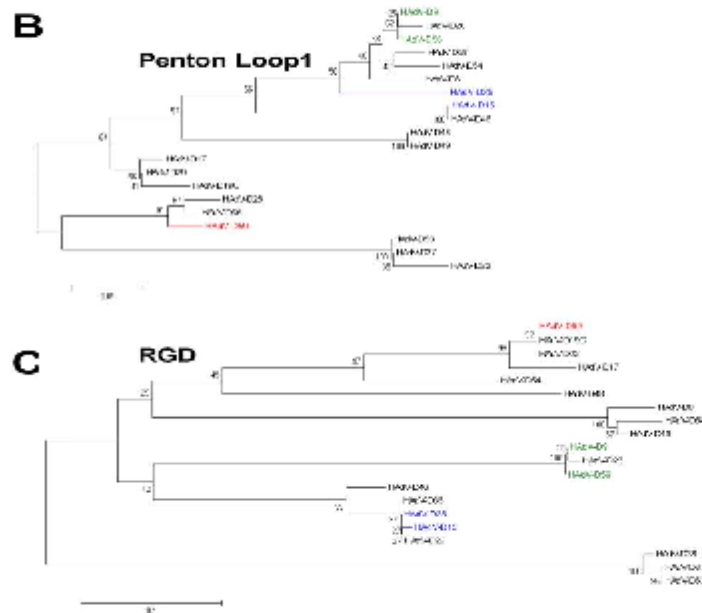
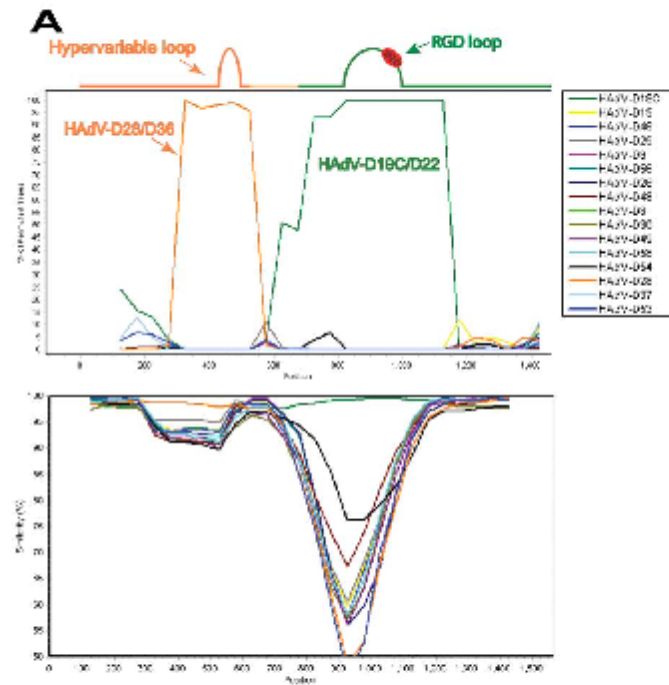


Figure 3. Computational analysis of the penton base gene. (A) Bootscan analysis of the HAdV-D59 penton base gene with fully sequenced penton genes in species HAdV-D using a window size of 250 bp and step size of 50 bp. (B) Phylogenetic analysis of the hypervariable loop 1 penton base gene sequences in species HAdV-D. (C) Phylogenetic analysis of the RGD motif and surrounding variable region of sequences in species HAdV-D. The phylogenetic trees were generated from aligned sequences using MEGA, via the neighbor-joining method and a bootstrap test of phylogeny. doi:10.1371/journal.pone.0033212.g003

RGD loop segregates with the subclade that includes HAdV-D19C and HAdV-D22 with a bootstrap value of 92 (Fig. 3C).

Hexon Gene Analysis

A previous study showed that recombination in HAdVs can occur within the hexon gene of viruses in species HAdV-D [8]. This has also been demonstrated in other species of HAdVs [4,5]. To determine whether or not the HAdV-D59 hexon coding region was either novel or the result of a recombination event, SimPlot analysis was performed. SimPlot results were not consistent with a recent recombination event in the hexon of HAdV-D59 (Fig. 4A). In addition, the nucleotide percent identity to other hexon coding sequences in species HAdV-D was determined (Table 1). HAdV-D59 loops 1 and 2 (L1 and L2) were 96.48% and 95.19% identical to L1 and L2 of HAdV-D25, respectively (Table 1). Phylogenetic analysis demonstrated that L1 of HAdV-D59 and HAdV-D25 are as distantly related as the following HAdV pairs which were shown to be distinct via serology [21]: HAdV-D39 and HAdV-D43; HAdV-D37 and HAdV-D13; HAdV-D15 and HAdV-D30; HAdV-D29 and HAdV-D30; HAdV-D9 and HAdV-D32, as well as HAdV-D45 and HAdV-D26 (Fig. 4B). Furthermore, the same phenomenon was observed for L2 with the exception of the HAdV-D45/D26 pair (Fig. 4C).

The difference in nucleotide identity between HAdV-D59 and HAdV-D25 (the nearest phylogenetic relative to HAdV-D59) in the L1 and L2 domains are greater than 2.5% (3.52% and 4.81%, respectively). Madisch et al stated that percent nucleotide identity differences greater than 2.4% and 2.5% in L1 and L2, respectively, strongly suggests identification of a novel HAdV [22]. Therefore the percent of nucleotide identity differences in L1 and L2 of HAdV-D59 further suggests that the aforementioned virus is novel.

E3 Genome Region Analysis

SimPlot and Bootscan results suggest that a large portion of the E3 transcription region (genes encoding for the CR1 β , 18.4 k, CR1 γ , RID α , RID β , and 14.7 k proteins) in HAdV-D59 may have originated from a recombination event between either HAdV-D56 or HAdV-D9 and another yet to be described HAdV (Fig. 5). Interestingly, the SimPlot results suggest that a recombination event took place within the open reading frame of the CR1 β gene (Fig. 5A). We also examined other E3 genes in species HAdV-D and did not detect common recombination loci (data not shown).

Phylogenetic analyses of two genes (CR1 β and CR1 γ genes) in the E3 region demonstrate that the coding regions for CR1 β and CR1 γ in HAdV-D59 are closely related to those of HAdV-D9 and HAdV-D56 (Fig. 5B, 5C).

Fiber Gene Analysis

SimPlot analyses of the HAdV-D59 fiber was performed on fiber sequences extracted from GenBank. The results suggested that the fiber gene of HAdV-D59 is nearly identical with sequences from both HAdV-D56 and HAdV-D9 (Fig. 5A). Furthermore, the fiber of HAdV-D59 had 99.54% and 99.84% nucleotide identities with HAdV-D9 and HAdV-D56 (Table 1), respectively. Phylogenetic analysis of the fiber genes in species HAdV-D confirms that the fiber of HAdV-D59 was closest in

sequence to the corresponding sequences in HAdV-D56 and HAdV-D9 (Fig. 5D).


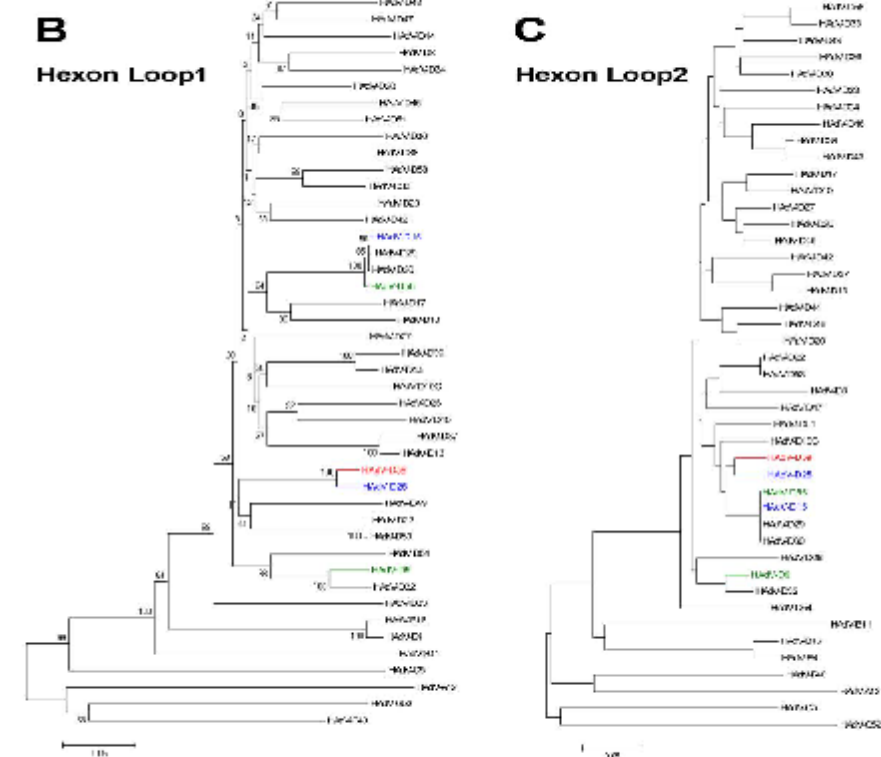
Serum Neutralization

HAdV-D59 was neutralized by both HAdV-D25 and HAdV-D15 antiserum, yet not by HAdV-D9 antiserum (Table 2). Interestingly, HAdV-D25 antiserum showed at least a two-fold higher neutralization titer to HAdV-D59 (greater than 1:4096) than it did against its cognate antigen HAdV-D25 (1:2048) (Table 2). These results demonstrate that the antigenic profile of HAdV-D59 differs from that of HAdV-D25 (Table 2). HAdV-D15 antiserum demonstrated only a two-fold increase in its ability to neutralize its cognate HAdV-D15 (1:1024) relative to the heterologous HAdV-D25 antigen (1:512). Similar reciprocal results were obtained with HAdV-D25 antiserum against HAdV-D15 (1:1024) and HAdV-D25 (1:2048). When tested with rabbit serum our results indicate that HAdV-D15 and HAdV-D25 are not as serologically distinct as previously reported [21].

Discussion

Our results demonstrated that HAdV-D25 antiserum was more effective at neutralizing HAdV-D59 than HAdV-D25 (Table 2). Since L1 and L2 protrude from the surface of HAdVs [23], it is not surprising that there is a difference between the ability of HAdV-D25 antiserum to neutralize the different viruses. One possibility for the differences in neutralization may be that the few differences in the primary amino acid structure present the HAdV-D59 hexon three-dimensional structure in such a way that the neutralizing epitopes are enhanced, thus making the virus easier to neutralize. Interestingly, we also observed a two-fold difference between the ability of HAdV-D15 and HAdV-D25 to be neutralized by reciprocal antiserum. This contradicts one study that showed antiserum to HAdV-D15 and HAdV-D25 did not cross-react in reciprocal neutralization experiments [21]. However, our data are consistent with the original characterization of HAdV-D15 and HAdV-D25 (previously called BP-1) [24], thus we conclude that HAdV-D15 and HAdV-D25 are not separate serotypes according to the traditional methods used for differentiating serotypes. These results also demonstrate that using neutralization as a criterion to type novel adenoviruses is complicated by non-standard serology methods and reagents that may yield interlaboratory variability of neutralization results. In contrast, using genomics as a method for typing HAdVs is consistent regardless of which laboratory generates the results.

Even though HAdV-D15 and HAdV-D25 showed only a two-fold difference via serum neutralization, they were recognized as different serotypes by Rosen et al, because they had different fiber proteins [24]. HAdV-D15 and HAdV-D25 share 79.13 and 92.56 percent nucleotide identity in L1 and L2, respectively. Thus, bioinformatic analysis demonstrates that they are actually different types using the criteria established by Madisch et al, which states that the nucleotide identity of L2 must differ by greater than 2.5 percent to type a novel virus [22]. Furthermore, pairwise nucleotide comparison of the hexon coding sequences for HAdV-D15 and HAdV-D25 show how genetically divergent they are (Fig. 6). Neutralization assays measure the overall effect of various antibodies that bind to multiple epitopes and may yield

 PLoS ONE | www.plosone.org

representatives from the other species. Phylogenetic trees were generated from aligned sequences using MEGA, via the neighbor-joining method and a bootstrap test of phylogeny.

doi:10.1371/journal.pone.0033212.g004

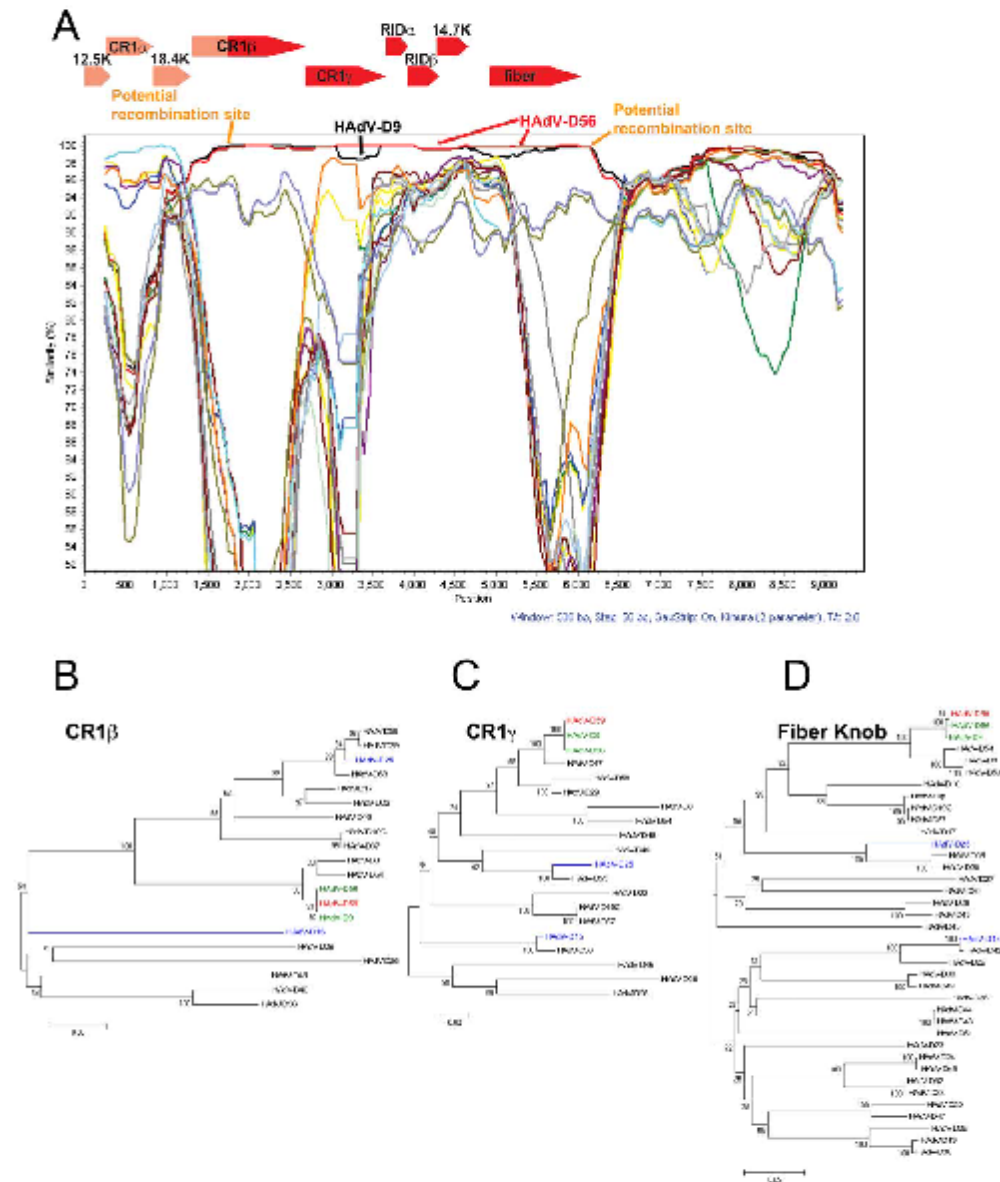


Figure 5. Computational analysis of the E3 and Fiber regions. (A) SimPlot analysis of the E3 and fiber region of HAdV-D59 compared to fully sequenced E3 and fiber regions from species HAdV-D. The arrows over the bootstrap demarcate the approximate positions of the E3 coding sequences. Phylogenetic analysis of (B) HAdV-D59 CR1β (C) HAdV-D59 CR1γ, and (D) fiber knob. The phylogenetic trees were generated from aligned sequences using MEGA, via the neighbor-joining method and a bootstrap test of phylogeny.

doi:10.1371/journal.pone.0033212.g005

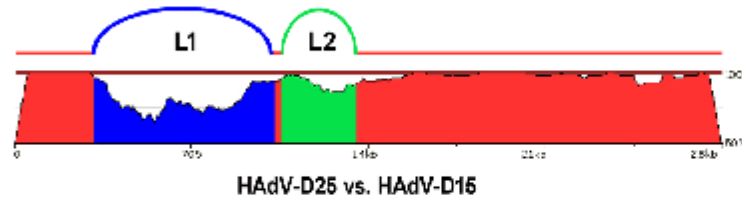


Figure 6. Pairwise nucleotide comparison of the hexon genes of HAdV-D15 and HAdV-D25. The blastz program zPicture was used to compare the hexon genes of HAdV-D15 and HAdV-D25. The blue region represents L1 and the green region represents L2.
doi:10.1371/journal.pone.0033212.g006

Table 2. Neutralization of HAdV-D59 with hyper immune serum.

	Antiserum		
	aHAdV-D9	aHAdV-D15	aHAdV-D25
HAdV-D9	1:128	<8*	<8*
HAdV-D15	<8*	1:1024	1:1024
HAdV-D25	<8*	1:512	1:2048
HAdV-D59	<8*	1:128	>14096

*No neutralization.
doi:10.1371/journal.pone.0033212.t002

variable interlaboratory results due to non-standard methods and reagents. Genomic analyses measure genetic differences in the genome that have the potential to affect the pathogenicity of the virus and can be independently verified by most laboratories. The contrasting serology and genomics results for HAdV-D15 and HAdV-D25 demonstrate that these two methods do not always yield concordant results.

SimPlot analysis of the HAdV-D59 L1 and L2 regions demonstrates high nucleotide identity between the hexons of HAdV-D25 and HAdV-D59 and suggests that they may be derived from a yet undiscovered common ancestor. If L1 and L2 of HAdV-D25 and HAdV-D59 are from a common ancestor, the recombination event may be ancient, as evidenced by 3.52 and 4.81 percent nucleotide differences in the L1 and L2 sequences, respectively. If the recombination events were recent, SimPlot analysis would illustrate near 100% nucleotide identity, which was shown for HAdV-D53 and HAdV-D56 [4,5,19]. With a distant past recombination event the hexon genes from HAdV-D25 and HAdV-D59 would have mutated over many replication cycles, after the initial recombination, to result in the variation we detected.

Multiple studies have shown that HAdVs in species HAdV-D recombine with one another in the penton base and hexon genes [4,8,20]. In this paper, we demonstrate that recombination may have occurred in the E3 region of HAdV-D59; however, after

examining all of the sequenced E3 genes in species HAdV-D, we found that there was not a predictable pattern of recombination (data not shown). Viruses in species HAdV-D show variability in their cell tropism ranging from growth in ocular tissues to gastrointestinal and/or respiratory tissues [4,25,26]. Given that the fiber knob is an important determinant of cell tropism, it may be concluded that recombination is an important molecular evolution pathway for the diversity observed within species HAdV-D.

The section of the HAdV-D59, -D56, and -D9 genomes that encodes for CR1 β , CR1 γ , RID α , RID β , 14.7K, and fiber show high nucleotide identity (Fig. 5A). From the nucleotide data, it is impossible to tell whether or not the 3' end of HAdV-D59 came from HAdV-D56 or HAdV-D9. Although HAdV-D9 was discovered in 1957 [25], there has been no disease associated with this virus. In contrast, prior serological evidence suggests that HAdV-D56, an ocular and respiratory pathogen (with hexon and fiber coding sequences similar to HAdV-D15 and HAdV-D9, respectively) [12,26] has been implicated in human disease as early as 1960 [27] and at other points in time as well [27,28,29,30,31,32]. HAdV-D59 may have also existed prior to our current description yet had gone undetected during the same time periods. Thus, it is impossible to say with absolute certainty which of the aforementioned viruses existed first and/or whether they evolved from a common ancestor. Future genomic analysis of known and unknown adenoviruses is needed to elucidate further the evolutionary history of HAdVs.

Acknowledgments

The views expressed in this material are those of the authors, and do not reflect the official policy or position of the U.S. Government, the Department of Defense, or the Department of the Air Force. We thank Drs. Michael Walsh and Shoaleh Delghan for advice, critical discussion and preliminary analyses.

Author Contributions

Conceived and designed the experiments: MSJ DS DAW. Performed the experiments: EBL JS MV NRH LT ST. Analyzed the data: EBL DAW JS DWD JC DS MSJ. Contributed reagents/materials/analysis tools: DWD JC DS MSJ. Wrote the paper: EBL DAW DWD JC DS MSJ.

References

1. Rowe WP, Hachner RJ, Gilmore LK, Pierotti RH, Ward TG (1963) Isolation of a cytopathogenic agent from human adenoids undergoing spontaneous degeneration in tissue culture. *Proceedings of the Society for Experimental Biology and Medicine* 84: 570–573.
2. Ishiko H, Aoki K (2009) Spread of epidemic keratoconjunctivitis due to a novel serotype of human adenovirus in Japan. *J Clin Microbiol* 47: 2678–2679.
3. Jones MS, 2nd, Harsach R, Gense RD, Goram MM, Dela Cruz WP, et al. (2007) New adenovirus species found in a patient presenting with gastroenteritis. *J Virol* 81: 5978–5984.
4. Robinson CM, Singh G, Henquell C, Walsh MP, Prigent-Lafeuille H, et al. (2011) Computational analysis and identification of an emergent human adenovirus pathogen implicated in a respiratory facility. *Virology* 409: 141–147.

5. Walsh MP, Seto J, Jones MS, Chodosh J, Xu W, et al. (2010) Computational analysis identifies human adenovirus type 55 as a re-emergent acute respiratory disease pathogen. *J Clin Microbiol* 48: 991–999.
6. Walsh MP, Seto J, Liu EB, Dehghan S, Hudson NR, et al. (2011) Computational analysis of two species C human adenoviruses provides evidence of a novel virus. *J Clin Microbiol* 49: 3483–3490.
7. Liu EB, Ferreira L, Fischer SL, Ryan JV, Nates SV, et al. (2011) Genetic analysis of a novel human adenovirus with a serologically unique hexon and a recombinant fiber gene. *PLoS One* 6: e24491.
8. Walsh MP, Chinokundawar A, Robinson CM, Madich I, Harach B, et al. (2009) Evidence of molecular evolution driven by recombination events influencing tropism in a novel human adenovirus that causes epidemic keratoconjunctivitis. *PLoS One* 4: e3635.
9. Rosen L. (1960) A hemagglutination-inhibition technique for typing adenoviruses. *Am J Hyg* 71: 120–128.
10. Seto D, Chodosh J, Brier JR, Jones MS (2011) Using the whole-genome sequence to characterize and name human adenoviruses. *J Virol* 85: 5701–5702.
11. Metzgar D, Gibbins C, Hudson NR, Jones MS (2010) Evaluation of multiplex type-specific real-time PCR assays using the LightCycler and joint biological agent identification and diagnostic system platform for detection and quantitation of adult human respiratory adenoviruses. *J Clin Microbiol* 48: 1397–1403.
12. Robinson CM, Shattil F, Zaishik J, Gilstrap AF, Dyer DW, et al. (2009) Human adenovirus type 19: genomic and bioinformatics analysis of a keratoconjunctivitis isolate. *Virus Res* 139: 122–126.
13. Echavarria M In Principles and Practice of Clinical Virology: John Wiley & Sons.
14. Crawford-Miksa LK, Schurr DP (1994) Quantitative colorimetric microneutralization assay for characterization of adenoviruses. *Journal of clinical microbiology* 32: 2331–2334.
15. Montefiori DC, Robinson WE, Jr., Schuffman SS, Mitchell WM (1988) Evaluation of antiviral drugs and neutralizing antibodies to human immunodeficiency virus by a rapid and sensitive microneutralization assay. *Journal of clinical microbiology* 26: 231–235.
16. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, et al. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23: 2947–2948.
17. Lele KS, Bolinger RC, Paragipe RS, Gadikeri D, Kulkarni SS, et al. (1999) Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J Virol* 73: 152–160.
18. Pettenes EF, Goddard TD, Huang OC, Couch GS, Greenblatt DM, et al. (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* 25: 1605–1612.
19. Robinson CM, Seto D, Jones MS, Dyer DW, Chodosh J (2011) Molecular evolution of human species D adenoviruses. *Infect Genet Evol* 11: 1208–1217.
20. Robinson CM, Raju J, Walsh MP, Seto D, Dyer DW, et al. (2009) Computational analysis of human adenovirus type 22 provides evidence for recombination among species D human adenoviruses in the penton base gene. *J Virol* 83: 8980–8985.
21. Hierholzer JC, Sone YO, Braden JR (1991) Antigenic relationships among the 47 human adenoviruses determined in reference horse antisera. *Archives of virology* 121: 179–197.
22. Madich I, Hante G, Pommer H, Heim A (2005) Phylogenetic analysis of the main neutralization and hemagglutination determinants of all human adenovirus prototypes as a basis for molecular classification and taxonomy. *J Virol* 79: 15265–15276.
23. Rue JJ, Burnett RM (2004) Adenovirus structure. *Hum Gene Ther* 15: 1167–1176.
24. Rosen L, Ramo S, Bell JA (1961) Four newly recognized adenoviruses. *Proc Soc Exp Biol Med* 107: 434–437.
25. Kibick S, Melander L, Enders JF (1957) Clinical associations of enteric viruses with particular reference to agents exhibiting properties of the ECHO group. *Ann N Y Acad Sci* 67: 311–325.
26. Henquell C, Bouff B, Mirand A, Bucher C, Tesore O, et al. (2009) Fatal adenovirus infection in a neonate and transmission to health-care workers. *J Clin Virol* 45: 345–348.
27. Campbell HG, Kaul JA, Langmack M, Wilkes FD (1960) Illnesses in children infected with an adenovirus antigenically related to types 9 and 15. *Pediatrics* 25: 823–828.
28. Adrian T, Bastian B, Benoit W, Hierholzer JC, Wigand R (1983) Characterization of adenovirus 15/H9 intermediate strains. *Intervirology* 23: 15–22.
29. Adrian T, Bastian B, Wagner V (1989) Restriction site mapping of adenovirus types 9 and 15 and genome types of intermediate adenovirus 15/H9. *Intervirology* 30: 169–176.
30. Hierholzer JC, Rodriguez FH, Jr. (1981) Antigenically intermediate human adenovirus strains associated with conjunctivitis. *J Clin Microbiol* 13: 395–397.
31. Wigand R, Fliedner D (1968) Serologically intermediate adenovirus strains: a regular feature of group II adenoviruses. *Arch Gesamte Virusforsch* 24: 245–256.
32. Wigand R (1987) Pitfalls in the identification of adenoviruses. *Journal of virological methods* 16: 161–169.

CHAPTER 6 – BIOINFORMATICS TOOLS DEVELOPMENT TO ENHANCE VIRAL GENOME ANALYSIS

Introduction

Currently there are limited numbers of efficient and accurate computational software tools for the bioinformatics analysis of small genomes such as viruses, including human adenoviruses specifically. Data presentation is an important tool necessary to convey complex genomic information. Many of the current genome data presentation methods include manually-drawn figures that are created based on an estimation of genome location. To provide better visualization of the data, two tools were developed during the course of adenoviral genome investigations described in the preceding chapters. The two tools are as follows: 1) DrawBar (<http://binf.gmu.edu/eliu1/drawbar/>) and 2) GeneMap (<http://binf.gmu.edu/eliu1/genemap/>). DrawBar is a web application that generates a visual representation to portray accurately gene locations. The result is a valuable addition to the recombination analysis and presentation of SimPlot and Bootscan data. GeneMap was developed to display accurately and efficiently the locations of any proteins sequences encoded in a genome. This enables “landmarks” to be placed along the nucleotide sequence as references, for example to mark recombination event.

Finally, as a way to convey “state-of-the-art” HAdV information to the research community and the public, a resource website was created to house human adenovirus

data and HAdV typing and identification information provided by the Human Adenovirus Working Group and international human adenovirus research community (hadvwg.gmu.edu). This resource was suggested by a virology taxonomy expert/representative from the National Library of Medicine (National Institutes of Health). It is an effort to provide standard for typing and identification, and to coordinate the process of naming novel human adenoviruses prior to GeneBank deposit. Furthermore, its goal is to reduce confusion and conflicts and competing deposits of GeneBank data.

Discussion

DrawBar

In scientific research papers and publications, figures are powerful and important tools to present data in a visual manner that allows reader to understand the easier large quantity of complex data. In this sense, the efficient and accurate labeling in a figure is equally important for the in successful presentation of data. This is especially important in genome analysis. Currently, most of the sequence analysis tools such as SimPlot, Bootscan and zPicture, provide figures in which the protein location labeling is done manually. This can be time consuming and often inaccurate, as the placements of protein locations were hand-drawn on the figures based on approximation. As an alternative, DrawBar was developed. It is a web application written in PHP, a general-purpose

scripting language that is suited especially to web development (PHP, 2015). DrawBar provides accurate placements of protein location in a pixel to base pair ratio.

DrawBar requires a few specific parameters in order to generate an accurate labeling for genome sequence data figures. There are the pixel length of the graph that requires labeling, base pair size of the displayed genome or protein and the specific base pair position of each target gene's labeling region. With these, the DrawBar program then calculates the percentage of each gene position to the total number of base pairs that the figure displayed, then translates this calculated percentage to pixel positions in relations to the total pixel length of the figure. In the example provided in **Figure 17**, protein labeling is required for a whole genome SimPlot graph that has a width of 600 pixels. The target genome has a size of 35000 base pairs, and the specific locations of E1A, E1B, polymerase, penton base, hexon and fiber proteins are required for reference. The required format for DrawBar is that the gene name has to be a single string without any spacing. A single blank space separates the gene name and the gene location. The gene location's start and end positions must be a single string numeric that is separated by a hyphen, and each gene name and location must start on a new line in the input box.

DrawBar

Bar Length in Pixels:

Genome/protein size:

Example of Gene and Location entry format:
 E1A 574-1496
 E1B 1541-3383
 Polymerase 5409-8621
 Penton 13878-15563
 Hexon 18434-21303
 Fiber 31406-34191
 Protein names need to be one entity, ie no spaces:
 Protein_1 or Second_protein_1
 NOT Gene 1 and NOT Second Gene 1

E1A 574-1496
 E1B 1541-3383
 Polymerase 5409-8621
 Penton 13878-15563
 Hexon 18434-21303
 Fiber 31406-34191

To save the above image:
 For PC User: Right-click, Save Image / Copy Image.
 For Mac User: Click on image and drag to desktop or desired folder.
 This is a beta version, if any issue please contact: eliu1@gmu.edu

Figure 17 - DrawBar input screen.

DrawBar (<http://binf.gmu.edu/eliu1/drawbar/>) input page. Parameters include pixel length of the graph that requires labeling, base pair size of the displayed genome or protein and the specific base pair position of each target gene's labeling region.

The Results page provides a legend bar that can be placed on top of the figure, with accurately calculated gene locations. Calculated gene locations, in pixel units, are displayed on the result page for debugging and accuracy check purposes (**Figure 18**). The user can right click on the resulting legend bar and save the image, which then can be pasted on to the SimPlot figure. The resulting legend bar serves as an accurate placement reference, and the user can then add any additional styling to the legend bar if needed.



Figure 18 - DrawBar results screen.
Resulting legend bar image contains accurately calculated gene locations and serves as an placement reference.

There are five php files required for DrawBar in order to run on a server, index.php; index.display.php; header.php; footer.php; image.php and map.php. First, index.php file contains code that initializes the application and handles the order of operation. The Index.display.php file contains code that handles the display the input screen content, and initializes the input variables for later processing. Header.php and footer.php contains code that displays content in the header and footer of the webpage; in this case, application title and footer note information. Image.php and map.php handle the actual calculation of the base pair to the pixel ratio along with the actual drawing of the image that is displayed in the result page. All of the source codes for DrawBar are included in **Appendix 1**.

GeneMap

GeneMap is a web application developed using PHP (www.php.org). Its structure and concept is very similar to that of the DrawBar tool, providing a map that presents the locations of all of the important proteins encoded in a genome. This is usually drawn by hand or through graphic software, again by estimation. The location of each protein is based on approximation its sequence location in relations to other proteins. This process can be tedious and extremely inaccurate since the location of the proteins can be overlapping, and also may include reverse or complement strands and spliced genes. GeneMap was developed to automate this process and to present an accurate portrayal of each proteins location. It is used for data presentation in publication, and also can serve as an accurate protein placement reference for user to draw on if additional styling of the

image is needed. This tool has been used in publications to enhance the visual display of adenovirus sequence data and to support novel findings (Seto et al., 2010)

GeneMap has three input fields that take specific required parameters in order to generate an accurate protein placement map in a genome. The first required field is a map name that will be displayed on top of the resulting map. A second required field is the base pair length of the target genome. The last input box is for specific base pair position of each target gene's mapping region with special syntax that can be used to indicate if reverse complement strand or splice gene is present. This GeneMap program then calculates the percentage of each gene position to the total number of base pair the figure displayed, translating this calculated percentage to pixel position in relations to the total pixel length of the figure, then accurately plot the protein locations on the final map image. This is very similar to the DrawBar design, but GeneMaps requires additional features that handle the placements of each protein on the map when there is a potential of overlapping. GeneMap calculates if there is another protein within the vicinity of the current protein; if so, it draws the new protein either above or below the existing protein to avoid overlapping. Also the placement and presentation of reverse or complement strands and splice genes are different from the regular forward strand proteins. In the example provided in **Figure 19**, a genome mapping figure is required for HAdV-B16. The target genome has a size of 35522 base pairs, with specific locations for all of the annotated proteins noted. The required format for GeneMap is that the gene name has to be a single string without any spacing. A single blank space separates the gene name and the gene location. The gene location start and end positions must be a single string

numeric that is separated by a hyphen. Each gene name and location must start on a new line in the input box. In addition, “, c” is added in the end of the protein location to indicate if a strand is reverse or complement. And in the case of a spliced gene, the two coding or exon locations are separated by a comma.

GeneMap

Map Name:

Genome Length:

Example of Gene and Location entry format:
Gene1 1-10
Gene2 20-50, 55-80 (for spliced genes)
Gene3 100-160, c (c indicates opposite strand coding)
Gene4 100-130, 145-160, c (complement and opposite strand coding)

Gene names need to be one entity, ie no spaces:
Gene_1 or Second_Gene_1
NOT Gene 1 and NOT Second Gene 1

To save the above image:
For PC User: Right-click, Save Image / Copy Image.
For Mac User: Click on image and drag to desktop or desired folder.
This is a beta version, if any issue please contact: eliu1@gmu.edu

Figure 19 - GeneMap input screen.

GeneMap (<http://binf.gmu.edu/eliu1/genemap/>) input page. Parameters include title, pixel length of the map, base pair size of the displayed genome and the specific base pair position of each target gene's labeling region.

The Results page provides final image of map, with accurately calculated and displayed gene locations. The calculated gene locations, in pixel units, are displayed on the Results page for debugging and accuracy check purposes (**Figure 20**).

GeneMap
E1A_28kDa 574 1153 1247 1452 forward E1A_25.7kDa 574 1060 1247 1452 forward E1A_6.3kDa 574 645 1247 1348 forward E1B_smallT 1598 2134 forward E1B_largeT 1903 3381 forward IX 3476 3892 forward IVa2 3945 5278 5557 5569 reverse E2B_10.4kDa 7130 7417 forward E2B_binding 7826 8422 forward E2B_12.6kDa 8226 8570 reverse E2B_terminal 8419 10386 13843 13851 reverse L1_55kDa 10851 12020 forward L1_IIIa 12045 13811 forward L2_penton 13902 15569 forward L2_VII 15582 16160 forward L2_V 16203 17255 forward L2_X 17284 17511 forward L3_VI 17586 18284 forward L3_hexon 18450 21272 forward L3_23kDa 21309 21938 forward E2A_DNAbinding 22027 23580 reverse L4_100kDa 23611 26097 forward I.4

Figure 20 - GeneMap result screen 1.

The calculated gene locations, in pixel units, are displayed on the Results page for debugging and accuracy check purposes.

The final image indicate all forward strands on the top half of the map and the reverse strands on the bottom half. Spliced genes have a green connector indicator (**Figure 21**). The user can then “right click” on the resulting image to save for later use. The resulting gene map can also serve as an accurate placement reference, and the user can then draw on any additional styling to the legend bar if needed.

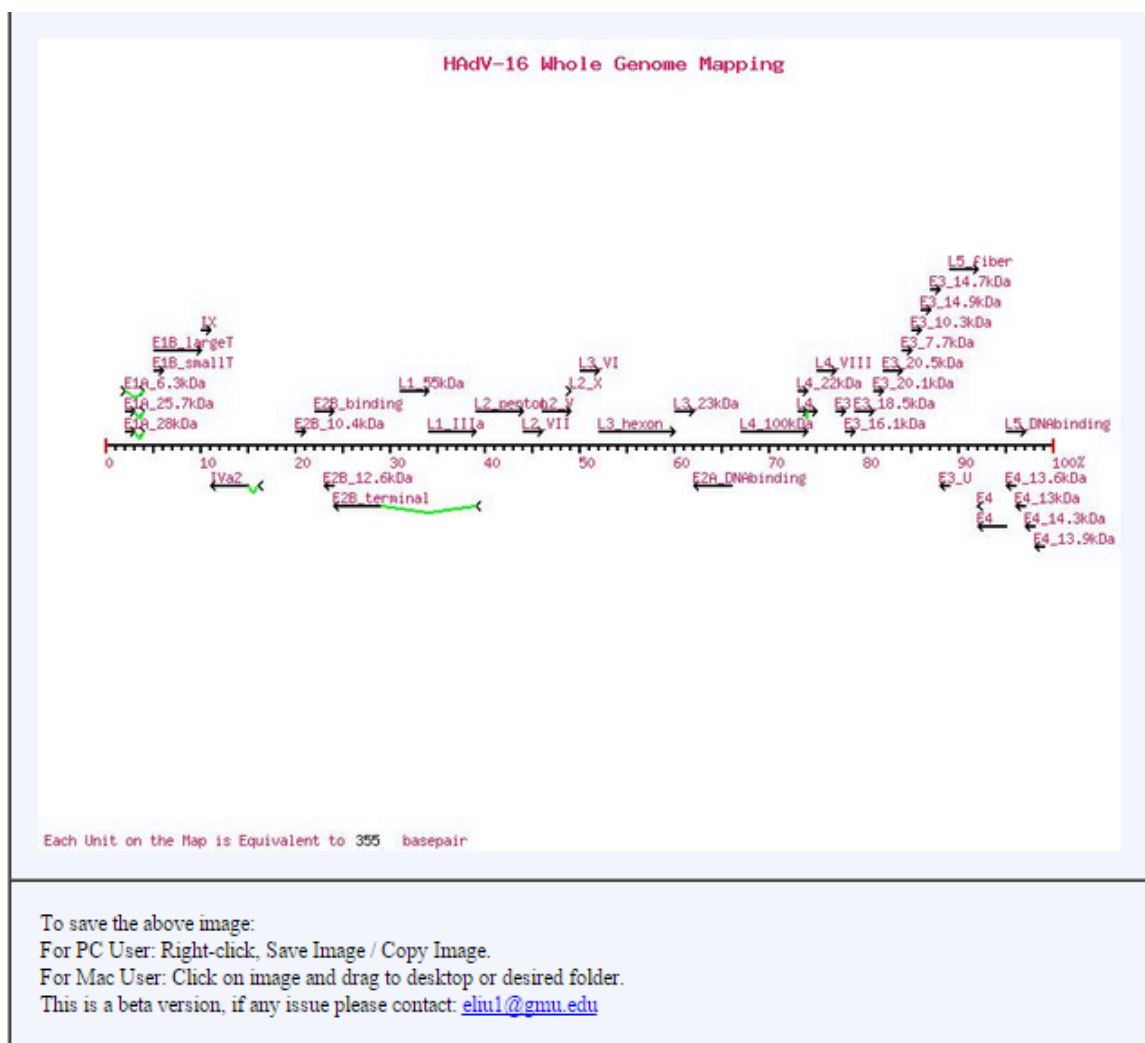


Figure 21 - GeneMap result screen 2.
Result image indicate all forward strands on the top half of the map and the reverse strands on the bottom half.

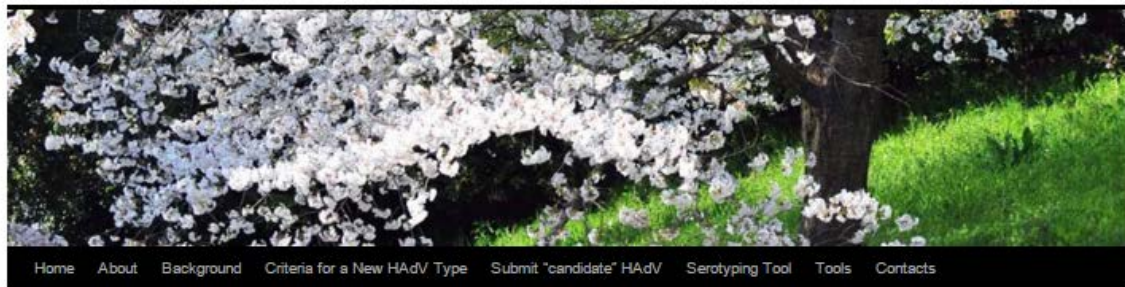
There are five php files required for GeneMap to run on a server: index.php; index.display.php; header.php; footer.php; image.php and map.php. These files provide the following functions: the index.php file is used to initialize the application and handles the order of operation; the index.display.php file processes the input screen content display, and initializes the input variables for later processing; the Header.php and footer.php files include code that displays the content in the header and footer section of the webpage; and the remaining two files, similar to DrawBar, image.php and map.php files handle the actual calculation of the base pair to pixel ratio and creates the drawing of the image that is displayed in the result page. In GeneMap, these two files also have additional functions that handle the calculation and differentiate the placements for the spliced genes, forward and reverse or complement genes. Those functions also provide placement calculation to avoid overlapping genes on the final image. All of the source codes for GeneMap are included in **Appendix 2**.

Human Adenovirus Working Group Reference Website

The Human Adenovirus Working Group has been formed as collaboration between several adenoviral researchers and a viral genome representative of the National Center for Biotechnology Information (NIH) in an effort to coordinate and standardize the process of assigning names to novel HAdVs based on genome data. The website was built with WordPress (WordPress, 2015), and is a basic web content hosting site. It provides background information regarding adenovirus typing, criteria for a new HAdV type and name, instructions regarding how to submit a candidate HAdV and a serotyping

tool. The serotyping tool was constructed using a DataTables JavaScript library based plugin that is freely available in the WordPress plugin library. The resulting tool displays all potential types corresponding to the query serotype entered by a user (**Figure 22**).

HAdV Working Group



Serotyping Tool

(or molecular typing for the un-serotyped)

Search:

Genome Type	Serotype	Accession No.	Species
HAdV-4	HAdV-4; 16	AY594253	E
HAdV-11	HAdV-11; 35	AY863756	B2
HAdV-16	HAdV-16; 4	AY601636	B1
HAdV-21	HAdV-21	AY601633	B1
HAdV-53	HAdV-53	FJ869625	D
HAdV-62	HAdV-62*	JN162671	D

Showing 1 to 6 of 6 entries (filtered from 70 total entries)

Figure 22 - HAdV Working group website serotyping tool. Serotyping tool in HAdV Working group website (hadvvg.gmu.edu) displays the result table from the search feature.

CHAPTER 7 – FUTURE DIRECTIONS

Future Directions

There are many follow-up questions remaining to be answered following this study of the molecular evolution and characterization of four human adenoviruses pathogens. Since the directionality of the cross species recombination cannot be determined, additional studies may provide insight in to these recombination events. This includes additional genome analysis with a larger data set in order to obtain a more complete and clearer picture of human adenovirus evolution, and directionality of horizontal gene transfer.

Standardized typing with genomics and bioinformatics protocols can also facilitate accurate identification of human adenoviruses. This will reduce long-standing confusion and conflicts in the naming and typing of novel and archived human adenoviruses. In addition to standardizing typing of human adenoviruses, standardized annotation of the genome is also necessary.

Currently the process of sequence annotation and isolation is repetitive and labor intensive. Also, genome annotation of human adenovirus can vary among research groups. This can provide inconsistent results in applying various bioinformatics tools such as sequence comparisons and percent identity calculations. To standardize the

annotation protocol for human adenovirus, either a manual protocol that can provide consistent results should be developed or an automated algorithm and tool should be developed. Both would help facilitate the process to provide a systematic standard that allows future analyses and understanding of any human adenovirus genome.

Additional research can be directed to software development for the graphical representation of sequence analysis. Current software such as SimPlot is extremely helpful and sufficient in graphic data representation, but are not without its caveats. As demonstrated in the HAdV-B16 and HAdV-B21 analyses and discussions, when there is more than one genome with high similarity to each other, the Bootscan results became unreadable as the highly similar sequences interfere with each other. Researchers have to eliminate the competing sequences manually in order to provide a clear presentation of the targeted data. It would be much more efficient for the Bootscan software to handle such competitive similarities and to display accurate results without the need to eliminate manually a number of sequences.

Software such as GeneMap can also be improved by having higher resolution and better quality of the resulting map. As PHP has limitations in its graphical outputs, other languages such as Java or Objective C can be considered in order to improve on the final results and to provide better data presentation.

All of the methods and tools mentioned in this dissertation have already generated vast amount of valuable and quality data. As the study of human adenoviruses, and other viruses, progress, these methods will definitely be improved and advanced.

APPENDICES

Appendix 1 DrawBar code

```
<?
//index.php by Elizabeth Liu for Draw Bar graphical output

require_once("header.php");
require_once("index.display.php");
//extract ($_GET);
$task = $_GET['task'];

switch ($task){
    case process:
        DISPLAY_Index::ProcessData();
        DISPLAY_Index::OutputData();
        break;

    default:

        DISPLAY_Index::DefaultView();

        break;
}
require_once("footer.php");
?>
```

```

<?

// index.display.php by Elizabeth Liu

class DISPLAY_Index{

    // original index page with 2 forms
    // form one is genemap title
    // form two is input box for the gene names and length
    // two buttons, one clear one submit
    function DefaultView(){

        ?>
        <form method="post" action="index.php?task=<?=process?>">
        Bar Length in Pixels: <br>
        <input type="text" name="barlength" value=600><br><br>
        Genome/protein size:<br>
        <input type="text" name="genomesize" value=35000><br><br>
        Example of Gene and Location entry format: <br>
        E1A 574-1496<br>
        E1B 1541-3383<br>
        Polymerase 5409-8621<br>
        Penton 13878-15563<br>
        Hexon 18434-21303<br>
        Fiber 31406-34191<br>
        Protein names need to be one entity, ie no spaces:<br>
        Protein_1 or Second_protein_1 <br>
        NOT Gene 1 and NOT Second Gene 1<br>

        <textarea name="proteinlocations" rows="20" cols="80">
        E1A 574-1496
        E1B 1541-3383
        Polymerase 5409-8621
        Penton 13878-15563
        Hexon 18434-21303
        Fiber 31406-34191
        </textarea><br>
        <!--<input type="reset" value="Clear the form"/>--!>
        <input type="submit" value="Submit"/>
        </form>

        <?

    }

    function ProcessData(){

        $barlength=$_REQUEST['barlength'];
        $genomesize=$_REQUEST['genomesize'];
        $data=$_REQUEST['proteinlocations'];
        $lines=split("\n", $data);
        $i=0;

        foreach($lines as $nextLine){
            // split input
            $test=preg_split("/[\\n\\t\\s,-]+/", $nextLine);

```

```

//remove blank space at end of array
// store gene name in a sepearte array
if(is_numeric($test[1])!=false){
    break;
}
$genename[$i]=$test[0];
//remove genename and leave $test just as gene locations
array_splice($test, 0,1);

echo $genename[$i];
echo "<br>";

foreach($test as $item){
    echo $item;
    echo "\t";
}
// sort the array
//sort($test);
$genePosition[$i]=$test;
    echo "<br>";
$i+=1;
}
echo "<br>";
echo $barlength;
echo "<br>";
echo $genomesize;
echo "<br>";
$_SESSION['barlength']=$barlength;
$_SESSION['genomesize']=$genomesize;
$_SESSION['names']=$genename;
$_SESSION['genePosition']=$genePosition;

$lengthratio = 1;
if($genomesize>100&$barlength!=0){
    $lengthratio=round($genomesize/$barlength);
}
echo "length ratio is: ";
echo $lengthratio;
echo "<br>";

}

function OutputData(){
// echo "output result here";
echo "<br>";
echo "<img src=image.php>";

}

}

?>

```

```

<?
// image.php by Elizabeth Liu
session_start();
header("Content-type: image/png");
//=====
// DRAW BACKGROUND IMAGE
//=====
//Calculate image size base on bar length. height of
//the image is 1/50 for now
$imgh=0.05*$_SESSION['barlength'];
$imgl=$_SESSION['barlength'];
$im=@imagecreate($imgl,$imgh)
    or die("cannot initialize new GD image stream");
$backgournd_color=imagecolorallocate($im, 255,255,255);
$black=imagecolorallocate($im, 0, 0, 0);
//calculate

//=====
// DRAW BAR
//=====
// draw a straight line with 2 red stoppers
$linethickness=0.004*$_SESSION['barlength'];
imageSetThickness($im, $linethickness);
imageline($im, 0, round($imgh/1.5), $imgl, round($imgh/1.5),
$black);

//=====
// Draw Actual Genes
//=====
$pxlratio=round($_SESSION['genomesize']/$_SESSION['barlength']);
$blockh=4*$linethickness;
$blockyvalb=round($imgh/1.5)-$blockh/2;
$blockyvale=round($imgh/1.5)+$blockh/2;

$txth=$blockyvalb-15;
for ($m=0; $m<sizeof($_SESSION['names']); $m++){

    $beg=round($_SESSION['genePosition'][$m][0]/$pxlratio);
    $end=round($_SESSION['genePosition'][$m][1]/$pxlratio);

    imagefilledrectangle($im, $beg, $blockyvalb, $end, $blockyvale,
$black);

    imagestring($im, 2, $beg, $txth, $_SESSION['names'][$m],
$black);
}

imagejpeg($im);
imagedestroy($im);
?>

```

Header.php

```
<?
session_start();
//header.php by Elizabeth Liu

?>
<html>
<head>
<meta http-equiv="Pragma" content="no-cache">
<meta http-equiv="Cache-Control" content="no-cache">
<meta http-equiv="Expires" content="Sat, 01 Dec 2001 00:00:00 GMT">
<title>DrawBar-Elizabeth Liu</title>
</head>
<body>
<center>
<table border=2px width=800px bgcolor="#f2f5fc" cellpadding=20
cellspacing=0>
<tr><td><h1>DrawBar</h1></td></tr>
<tr><td>
```

Footer.php

```
<?
// footer.php by Elizabeth Liu
?>
</td></tr>
<tr><td colspan=2>
<table width=100%>
<tr><td>To save the above image:<br>
For PC User: Right-click, Save Image / Copy Image. <br>
For Mac User: Click on image and drag to desktop or desired folder.<br>
This is a beta version, if any issue please contact: <a href="mailto:
eliul@gmu.edu?subject=Troubleshooting
drawbar">eliul@gmu.edu</a></td></tr>
</table>
</td></tr></table>
</body>
</html>
```

Appendix 2 GeneMap code

```
<?
//index.php by Elizabeth Liu for gene map graphical output

require_once("header.php");
require_once("index.display.php");
//extract ($_GET);
$task = $_GET['task'];

switch ($task){
    case process:
        DISPLAY_Index::ProcessData();
        DISPLAY_Index::OutputData();
        break;

    default:

        DISPLAY_Index::DefaultView();

        break;
}
require_once("footer.php");
?>
```

```

<?

// index.display.php by Elizabeth Liu

class DISPLAY_Index{

    // original index page with 2 forms
    // form one is genemap title
    // form two is input box for the gene names and length
    // two buttons, one clear one submit
    function DefaultView(){

        ?>
        <form method="post" action="index.php?task=<?=process?>">
        Map Name: <br>
        <input type="text" name="mapname"><br><br>
        Genome Length:<br>
        <input type="text" name="maxlength"><br><br>
        Example of Gene and Location entry format: <br>
        Gene1 1-10<br>
        Gene2 20-50, 55-80 (for spliced genes)<br>
        Gene3 100-160, c (c indicates opposite strand coding) <br>
        Gene4 100-130, 145-160, c (complement and opposite strand
coding)<br><br>
        Gene names need to be one entity, ie no spaces:<br>
        Gene_1 or Second_Gene_1 <br>
        NOT Gene 1 and NOT Second Gene 1<br>

        <textarea name="geneinfo" rows="20" cols="80"></textarea><br>
        <input type="reset" value="Reset"/>
        <input type="submit" value="Submit"/>
        </form>
        <?

    }

    function ProcessData(){

        $mapname=$_REQUEST['mapname'];
        $maxlength=$_REQUEST['maxlength'];
        $data=$_REQUEST['geneinfo'];
        $lines=split("\n", $data);
        $i=0;
        $max=$maxlength;
        $min=0;
        $reverse=array_fill(0, 500, 0);
        $fwd=0;
        $rev=0;

        foreach($lines as $nextLine){
            // split input
            $test=preg_split("/[\\n\\t\\s,-]+/", $nextLine);
            //remove blank space at end of array
            // store gene name in a sepearte array
            if(is_numeric($test[1])==false){

```

```

        break;
    }
    $genename[$i]=$test[0];
    //remove genename to calc max and min
    array_splice($test, 0,1);
    for($a=0; $a < sizeof($test); $a++){
        if(is_numeric($test[$a])== false){
            if($test[$a]=='c'){
                $reverse[$i]=1;
            }
            array_splice($test, $a);
        }
    }
    echo $genename[$i];
    echo "<br>";

    foreach($test as $item){
        echo $item;
        echo "\t";
    }
    // sort the array
    sort($test);
    //reverse
    if($reverse[$i]==1){
        $comp[$rev]=$test;
        $rname[$rev]=$genename[$i];
        echo "\t";
        echo "reverse";
        $rev+=1;
    }
    //forward array
    else{
        $forward[$fwd]=$test;
        $fname[$fwd]=$genename[$i];
        echo "\t";
        echo "foward";
        $fwd+=1;
    }
    $genePosition[$i]=$test;
    echo "<br>";

    // if($max<max($test)){
    //     $max=max($test);
    // }
    // if($min>min($test)){
    //     $min=min($test);
    // }
    $i+=1;
}
echo "max is : ";
echo $max;
echo " min is : ";
echo $min;

```



```

    echo "<br>";
$_SESSION['minMaxFwd'][0]=$min;
$_SESSION['minMaxFwd'][1]=$max;
$_SESSION['mapName']=$mapname;
$_SESSION['names']=$genename;
$_SESSION['genePosition']=$genePosition;
$_SESSION['reverse']=$reverse;
$_SESSION['comp']=$comp;
$_SESSION['forward']=$forward;
$_SESSION['fname']=$fname;
$_SESSION['rname']=$rname;
$ratio = 1;
if($max>100){
    $ratio=round(($max-$min)/100);
}
echo "ratio is: ";
echo $ratio;
echo "<br>";
$yval=290;
for($i=0; $i<sizeof($genename); $i++){
    if(($i!=0) && ($genename[$i] == $genename[$i-1])){
        $yval-=5;
    }
    for ($k=0; $k<sizeof($genePosition[$i])-1; $k+=2){
        $end=intval(($genePosition[$i][$k+1]-$min)/$ratio)*7+50;
        $beg=intval(($genePosition[$i][$k]-$min)/$ratio)*7+50;
        echo "pixel beg and ends are";
        echo $beg;
        echo " ";
        echo $end;
        echo "<br>";
    }
}
}
function OutputData(){
//  echo "output result here";
    echo "<br>";
    echo "<img src=image.php>";
}

}

?>

```

```

<?
// image.php by Elizabeth Liu
session_start();
header("Content-type: image/png");
$im=@imagecreate(1600,1200)
    or die("cannot initialize new GD image stream");
$background_color=imagecolorallocate($im, 255,255,255);
$text_color=imagecolorallocate($im, 233, 14, 91);
$black=imagecolorallocate($im, 0, 0, 0);
$red=imagecolorallocate($im, 255, 0, 0);
$green=imagecolorallocate($im, 0, 255, 0);
imageSetThickness($im, 2);
imagestring($im, 5, 300, 10, $_SESSION['mapName'], $text_color);
//=====
// DRAW RULER
//=====
// draw a straight line with 2 red stoppers
imageline($im, 50, 300, 750, 300, $black);
imageline($im, 50, 295, 50, 305, $red);
imageline($im, 750, 295, 750, 305, $red);
$rulecount=1;
$numcount=0;
$stickcount=1;
imagestring($im, 2, 50, 306, "0", $text_color);
imagestring($im, 2, 750, 306, "100%", $text_color);
for ($r=57; $r<750; $r+=7){
// at five, output taller tick mark
    if($rulecount==5){
        imageline($im, $r, 300, $r, 305, $black);
        $rulecount=0;
        $numcount+=1;
    }
    else{
        imageline($im, $r, 300, $r, 303, $black);
    }
    if($numcount==2){
        imagestring($im, 2, $r, 306, $stickcount, $text_color);
        $numcount=0;
    }
    $rulecount+=1;
    $stickcount+=1;
}
//=====
// Calculate Porportion
//=====
//by default ratio is 1, but not true in most cases
$ratio = 1;
if($_SESSION['minMaxFwd'][1]>100){
    $ratio=round(($_SESSION['minMaxFwd'][1]-
$_SESSION['minMaxFwd'][0])/100);
}
imagestring($im, 2, 5, 585, "Each Unit on the Map is Equivalent to
", $text_color);
imagestring($im, 2, 235, 585, $ratio, $black);

```

```

imagestring($im, 2, 270, 585, "basepair", $text_color);
//=====
// Draw Actual Genes, Fwd and Comp strands
//=====
$yval=290;
$ryval=330;
//=====
// for comp strand
//=====
for($i=0; $i<sizeof($_SESSION['rname']); $i++){
    // if the names is the same as previous one, then probably very
    // close together, move up yval
    if(($i!=0) && ($_SESSION['rname'][$i]==$_SESSION['rname'][$i-1])){
        $ryval+=15;
    }
    else{ // if different name reset to 290
        $ryval=330;
        if($i!=0){
            // calc new beg pixel
            // if spliced
            $newbeg=round(($_SESSION['comp'][$i][0]-
$_SESSION['minMaxFwd'][0])/$ratio)*7+50;

            for($r=0; $r<sizeof($rrecord); $r++){
                // now check record see if too close or not
                if($rrecord[$r][2]==$ryval){
                    $namelen=strlen($_SESSION['rname'][$r]);
                    $namelen*=10;

                    if(($newbeg<$rrecord[$r][0]+$namelen)||($newbeg<$rrecord[$r][1]+7)){
                        $ryval+=15;
                    }
                }
            }
        }
    }
}

for ($k=0; $k<sizeof($_SESSION['comp'][$i])-1; $k+=2){
    // if almost out of frame, reset
    if($ryval>=585){
        $ryval=330;
    }
    $obeg=round(($_SESSION['comp'][$i][0]-
$_SESSION['minMaxFwd'][0])/$ratio)*7+50;
    $send=round(($_SESSION['comp'][$i][$k+1]-
$_SESSION['minMaxFwd'][0])/$ratio)*7+50;
    $beg=round(($_SESSION['comp'][$i][$k]-
$_SESSION['minMaxFwd'][0])/$ratio)*7+50;
    imageline($im, $beg, $ryval, $send, $ryval, $black);
    // arrow
    imageline($im, $beg, $ryval, $beg+3, $ryval-3, $black);
    imageline($im, $beg, $ryval, $beg+3, $ryval+3, $black);
    imagestring($im, 2, $obeg, $ryval-13, $_SESSION['rname'][$i],
$text_color);
}

```

```

    }
    $rrecord[$i][0]=$obeg;
    $rrecord[$i][1]=$end;
    $rrecord[$i][2]=$ryval;
    if(is_numeric($_SESSION['comp'][$i][3])){
        // draw V for spliced gene, calc midpoint
        $sbeg=round(($_SESSION['comp'][$i][1]-
$_SESSION['minMaxFwd'][0])/$ratio)*7+50;
        $send=round(($_SESSION['comp'][$i][2]-
$_SESSION['minMaxFwd'][0])/$ratio)*7+50;
        $mid=round(($send+$sbeg)/2);
        imageline($im, $sbeg, $ryval, $mid, $ryval+5, $green);
        imageline($im, $send, $ryval, $mid, $ryval+5, $green);
    }
}
//=====
// for Forward strand
//=====
for($l=0; $l<sizeof($_SESSION['fname']); $l++){
    // if names as same as prevoius one, move up, likely to be close
    if(($l!=0) && ($_SESSION['fname'][$l]==$_SESSION['fname'][$l-
1])){
        $yval-=15;
    }
    else{ // if different name reset to 290
        $yval=290;
        if($l!=0){
            // calc new beg pixel
            $newbeg=round(($_SESSION['forward'][$l][0]-
$_SESSION['minMaxFwd'][0])/$ratio)*7+50;
            for($f=0; $f<sizeof($frecord); $f++){
                // now check record see if too close or not
                if($frecord[$f][2]==$yval){
                    $namelen=strlen($_SESSION['fname'][$f]);
                    $namelen*=8;

                    if(($newbeg<$frecord[$f][0]+$namelen)||($newbeg<$frecord[$f][1]+7)){
                        $yval-=15;
                    }
                }
            }
        }
    }
}
for ($m=0; $m<sizeof($_SESSION['forward'][$l]); $m+=2){
    if($yval<=15){
        $yval=290;
    }
    $obeg=round(($_SESSION['forward'][$l][0]-
$_SESSION['minMaxFwd'][0])/$ratio)*7+50;
    $send=round(($_SESSION['forward'][$l][$m+1]-
$_SESSION['minMaxFwd'][0])/$ratio)*7+50;
    $beg=round(($_SESSION['forward'][$l][$m]-
$_SESSION['minMaxFwd'][0])/$ratio)*7+50;
    imageline($im, $beg, $yval, $send, $yval, $black);
}

```

```

        // arrow
        imageline($im, $end, $yval, $end-3, $yval-3, $black);
        imageline($im, $end, $yval, $end-3, $yval+3, $black);
        imagestring($im, 2, $obeg, $yval-13, $_SESSION['fname'][$l],
$text_color);
    }
    $frecord[$l][0]=$obeg;
    $frecord[$l][1]=$end;
    $frecord[$l][2]=$yval;
    if(sizeof($_SESSION['forward'][$l])==4){
        //draw V
        $sbeg=round(($_SESSION['forward'][$l][1]-
$_SESSION['minMaxFwd'][0])/$ratio)*7+50;
        $send=round(($_SESSION['forward'][$l][2]-
$_SESSION['minMaxFwd'][0])/$ratio)*7+50;
        $mid=round(($send+$sbeg)/2);
        imageline($im, $sbeg, $yval, $mid, $yval+5, $green);
        imageline($im, $send, $yval, $mid, $yval+5, $green);
    }
}
imagejpeg($im);
imagedestroy($im);
?>

```

```

<?
// map.php by Elizabeth Liu
echo "image here shown";
echo "<br>";
$mapname=$_REQUEST['mapname'];
$data=$_REQUEST['geneinfo'];
$lines=split("\n", $data);
$i=0;
$max=0;
$min=100000;
foreach($lines as $nextLine){
    // echo $nextLine;
    //problem, adding an additional space at end of the array still
questionable
    //problem temporarily fixed
    $test=preg_split("/[\n\s,-]+/", $nextLine);
    //remove blank space at end of array
    $n=sizeof($test);
    // if last item in array is numeric, then keep, if not then
splice
    if(is_numeric($test[$n-1])!=false){
        array_splice($test, -1);
    }
    //store gene name at seperate array
    $genename[i]=$test[0];
    //remove genename to calc max and min
    array_splice($test, 0,1);
    // if next item is not numeric, then break
    if(is_numeric($test[0])!=false){
        break;
    }
    if($max<max($test)){
        $max=max($test);
    }
    if($min>min($test)){
        $min=min($test);
    }
    $array[i]=$test;
    /*
    echo "number of genes: ";
    echo sizeof($array[i])/2;
    echo "<br>";*/
    $i+=1;
}
echo "max is : ";
echo $max;
echo "min is : ";
echo $min;
echo "<br>";

```

?>

Header.php

```
<?
session_start();
//header.php by Elizabeth Liu

?>
<html>
<head>
<meta http-equiv="Pragma" content="no-cache">
<meta http-equiv="Cache-Control" content="no-cache">
<meta http-equiv="Expires" content="Sat, 01 Dec 2001 00:00:00 GMT">
<title>GeneMap-Elizabeth Liu</title>
</head>
<body>
<center>
<table border=2px width=800px bgcolor="#f2f5fc" cellpadding=20
cellspacing=0>
<tr><td><h1>GeneMapBeta</h1></td></tr>
<tr><td>
```

Footer.php

```
<?
// footer.php by Elizabeth Liu
?>
</td></tr>
<tr><td colspan=2>
<table width=100%>
<tr><td>To save the above image:<br>
For PC User: Right-click, Save Image / Copy Image. <br>
For Mac User: Click on image and drag to desktop or desired folder.<br>
This is a beta version, if any issue please contact: <a href="mailto:
eliul@gmu.edu?subject=Troubleshooting
genemap">eliul@gmu.edu</a></td></tr>
</table>
</td></tr></table>
</body>
</html>
```

REFERENCES

- Bell, S. D., Jr., C. D. Mc, E. S. Murray, R. S. Chang, and J. C. Snyder. 1959. Adenoviruses isolated from Saudi Arabia. I. Epidemiologic features. *Am J Trop Med Hyg* 8:492-500.
- Binn, L. N., J. L. Sanchez, and J. C. Gaydos. 2007. Emergence of adenovirus type 14 in US military recruits-a new challenge. *J Infect Dis* 196:1436-7.
- Bournsnell, M., and V. Mautner. 1981. In vitro construction of a recombinant adenovirus Ad2:Ad5. *Gene* 13:311-7.
- Bournsnell, M. E., and V. Mautner. 1981. Recombination in adenovirus: crossover sites in intertypic recombinants are located in regions of homology. *Virology* 112:198-209.
- Crawford-Miksza, L. K., R. N. Nang, and D. P. Schnurr. 1999. Strain variation in adenovirus serotypes 4 and 7a causing acute respiratory disease. *J Clin Microbiol* 37:1107-12.
- Crawford-Miksza, L. K., and D. P. Schnurr. 1996. Adenovirus serotype evolution is driven by illegitimate recombination in the hypervariable regions of the hexon protein. *Virology* 224:357-67.
- D'Ambrosio, E., N. Del Grosso, A. Chicca, and M. Midulla. 1982. Neutralizing antibodies against 33 human adenoviruses in normal children in Rome. *J Hyg (Lond)* 89:155-61.
- Dehghan et al., 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* 447, 265(2013).
- Dehghan et al., computational analysis of four human adenovirus type 4 genomes reveals molecular evolution through two interspecies recombination events. *Virology* 443, 197(2013).
- De Jong, J. C., A. G. Wermenbol, M. W. Verweij-Uijterwaal, K. W. Slaterus, P. Wertheim-Van Dillen, G. J. Van Doornum, S. H. Khoo, and J. C. Hierholzer. 1999. 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 37:3940-5.

Echavarria, M. 2009. Adenovirus. In Principles and Practice of Clinical Virology, Sixth Edition. (A. J. Zuckerman, J. E. Banatvala, B. D. Schoub, P. D. Griffiths and P. Mortimer, eds.). John Wiley and Sons (San Diego,CA):463-88.

EMBL-EBI 2015 http://www.ebi.ac.uk/Tools/psa/emboss_needle/nucleotide.html

Engelmann, I., I. Madisch, H. Pommer, and A. Heim. 2006. An outbreak of epidemic keratoconjunctivitis caused by a new intermediate adenovirus 22/H8 identified by molecular typing. Clin Infect Dis 43:e64-6.

Feng, M., R. S. Chang, T. R. Smith, and J. C. Synder. 1959. Adenoviruses isolated from Saudi Arabia. II. Pathogenicity of certain strains for man. Am J Trop Med Hyg 8:501-4.

Fenner, Frank J.; Gibbs, E. Paul J.; Murphy, Frederick A.; Rott, Rudolph; Studdert, Michael J.; White, David O. (1993). *Veterinary Virology (2nd ed.)*. Academic Press, Inc.

Hierholzer, J. C., T. Adrian, L. J. Anderson, R. Wigand, and J. W. Gold. 1988. Analysis of antigenically intermediate strains of subgenus B and D adenoviruses from AIDS patients. Arch Virol 103:99-115.

Hierholzer, J. C., Y. O. Stone, and J. R. Broderson. 1991. Antigenic relationships among the 47 human adenoviruses determined in reference horse antisera. Arch Virol 121:179-97.

Hilleman, M. R., and J. H. Werner. 1954. Recovery of new agent from patients with acute respiratory illness. Proc Soc Exp Biol Med 85:183-8.

Imelli, N., Z. Ruzsics, D. Puntener, M. Gastaldelli, and U. F. Greber. 2009. Genetic reconstitution of the human adenovirus type 2 temperature-sensitive 1 mutant defective in endosomal escape. Virol. J. 6:174.

Ishiko, H., Y. Shimada, T. Konno, A. Hayashi, T. Ohguchi, Y. Tagawa, K. Aoki, S. Ohno, and S. Yamazaki. 2008. Novel human adenovirus causing nosocomial epidemic keratoconjunctivitis. J Clin Microbiol 46:2002-8.

Jones, M. S., 2nd, B. Harrach, R. D. Ganac, M. M. Gozum, W. P. Dela Cruz, B. Riedel, C. Pan, E. L. Delwart, and D. P. Schnurr. 2007. New adenovirus species found in a patient presenting with gastroenteritis. J Virol 81:5978-84.

Kajon, A. E., and G. Wadell. 1996. Sequence analysis of the E3 region and fiber gene of human adenovirus genome type 7h. Virology 215:190-6.

Tim Lahm, Helena N. Spartz, Dean A. Hawley, Diane S. Leland, Karen M. Wolf, Homer L. Twigg III, Michael D. Ober, Fatal adenovirus serotype 21 infection associated with hemophagocytic lymphohistiocytosis and multiorgan failure, *Respiratory Medicine CME*, Volume 3, Issue 4, 2010, Pages 223-225, ISSN 1755-0017

Li, Q. G., J. Hambræus, and G. Wadell. 1991. Genetic relationship between thirteen genome types of adenovirus 11, 34, and 35 with different tropisms. *Intervirology* 32:338-50

Lion T. et al., Adenovirus Infections in Immunocompetent and Immunocompromised Patients. *Clin. Microbio. Rev.* July 2014 vol. 27 no. 3 441-462

Liu EB, Ferreyra L, Fischer SL, Pavan JV, Nates SV, et al. (2011) Genetic analysis of a novel human adenovirus with a serologically unique hexon and a recombinant fiber gene. *PLoS One* 6: e24491.

Liu EB, Wadford DA, Seto J, Vu M, Hudson NR, et al. (2012) Computational and Serologic Analysis of Novel and Known Viruses in Species Human Adenovirus D in Which Serology and Genomics Do Not Correlate. *PLoS ONE* 7(3): e33212. doi:10.1371/journal.pone.0033212

Lole KS, Bollinger RC, Paranjape RS, Gadkari D, Kulkarni SS, Novak NG, Ingersoll R, Sheppard HW, Ray SC. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J Virol.* 1999 Jan;73(1):152-60.

Louie, J. K., A. E. Kajon, M. Holodniy, L. Guardia-LaBar, B. Lee, A. M. Petru, J. K. Hacker, and D. P. Schnurr. 2008. Severe pneumonia due to adenovirus serotype 14: a new respiratory threat? *Clin Infect Dis* 46:421-5.

Lukashev, A. N., O. E. Ivanova, T. P. Eremeeva, and R. D. Iggo. 2008. Evidence of frequent recombination among human adenoviruses. *J Gen Virol* 89:380-8.

Mahadevan, P., J. Seto, C. Tibbetts, and D. Seto. 2010. Natural variants of human adenovirus type 3 provide evidence for relative genome stability across time and geographic space *Virology ePub*.:Nov. 23, 2009. 397:113-8.

Mautner, V., and N. Mackay. 1984. Recombination in adenovirus: analysis of crossover sites in intertypic overlap recombinants. *Virology* 139:43-52.

Mei, Y. F., K. Lindman, and G. Wadell. 1998. Two closely related adenovirus genome types with kidney or respiratory tract tropism differ in their binding to epithelial cells of various origins. *Virology* 240:254-66.

Metzgar, D., M. Osuna, S. Yingst, M. Rakha, K. Earhart, D. Elyan, H. Esmat, M. D. Saad, A. Kajon, J. Wu, G. C. Gray, M. A. Ryan, and K. L. Russell. 2005. PCR analysis of egyptian respiratory adenovirus isolates, including identification of species, serotypes, and coinfections. *J Clin Microbiol* 43:5743-52.

Morgan, P. N., E. B. Moses, E. P. Fody, and A. L. Barron. 1984. Association of adenovirus type 16 with Reye's-syndrome-like illness and pneumonia. *South Med J* 77:827-30.

Numazaki, Y., S. Shigeta, T. Kumasaka, T. Miyazawa, M. Yamanaka, N. Yano, S. Takai, and N. Ishida. 1968. Acute hemorrhagic cystitis in children. Isolation of adenovirus type II. *N Engl J Med* 278:700-4.

Ogawa K. 1989. Embryonal neuroepithelial tumors induced by human adenovirus type 12 in rodents. 1. Tumor induction in the peripheral nervous system. *Acta Neuropathol* 7(3):244-53.

Pereira, H. G., R. J. Huebner, H. S. Ginsberg, and J. Van Der Veen. 1963. A Short Description of the Adenovirus Group. *Virology* 20:613-20.

PHP <https://secure.php.net/> 2015

Purkayastha, A., S. E. Ditty, J. Su, J. McGraw, T. L. Hadfield, C. Tibbetts, and D. Seto. 2005. Genomic and bioinformatics analysis of HAdV-4, a human adenovirus causing acute respiratory disease: implications for gene therapy and vaccine vector development. *J Virol* 79:2559-72.

Purkayastha, A., J. Su, S. Carlisle, C. Tibbetts, and D. Seto. 2005. Genomic and bioinformatics analysis of HAdV-7, a human adenovirus of species B1 that causes acute respiratory disease: implications for vector development in human gene therapy. *Virology* 332:114-29.

Purkayastha, A., J. Su, J. McGraw, S. E. Ditty, T. L. Hadfield, J. Seto, K. L. Russell, C. Tibbetts, and D. Seto. 2005. Genomic and bioinformatics analyses of HAdV-4vac and HAdV-7vac, two human adenovirus (HAdV) strains that constituted original prophylaxis against HAdV-related acute respiratory disease, a reemerging epidemic disease. *J Clin Microbiol* 43:3083-94.

Relman, D.A., 2001. Microbial genomics and infectious diseases. *New Engl. J. Med.* 365, 347-357.

Rowe, W. P., R. J. Huebner, L. K. Gilmore, R. H. Parrott, and T. G. Ward. 1953. Isolation of a cytopathogenic agent from human adenoids undergoing spontaneous degeneration in tissue culture. *Proc Soc Exp Biol Med* 84:570-3.

Robinson, C. M., J. Rajaiya, M. P. Walsh, D. Seto, D. W. Dyer, M. S. Jones, and J. Chodosh. 2009. Computational analysis of human adenovirus type 22 provides evidence for recombination between human adenoviruses species D in the penton base gene. *J Virol*. 83:8980-5.

Rutherford, K et al., "Artemis: sequence visualization and annotation," *Bioinformatics (Oxford,England)* 16, no. 10 (October 2000): 944-5.

Seto, J., Walsh, M. P., Mahadevan, P., Zhang, Q., and Seto, D. Applying genomic and bioinformatics resources to human adenovirus genomes for use in vaccine development and for applications in vector development for gene delivery. *Viruses* 2, 1 (2010), 1-26.

Seto, J., M. P. Walsh, P. Mahadevan, A. Purkayastha, J. M. Clark, C. Tibbetts, and D. Seto. 2009. Genomic and bioinformatics analyses of HAdV-14p, reference strain of a re-emerging respiratory pathogen and analysis of B1/B2. *Virus Res* 143:94-105.

Seto, Donald et al., "Evidence of molecular evolution driven by recombination events influencing tropism in a novel human adenovirus that causes epidemic keratoconjunctivitis," *PloS One* 4, no. 6 (2009): e5635.

Skog, J., K. Edlund, A. T. Bergenheim, and G. Wadell. 2007. Adenoviruses 16 and CV23 efficiently transduce human low-passage brain tumor and cancer stem cells. *Mol Ther* 15:2140-5.

Tamura, K., J. Dudley, M. Nei, and S. Kumar. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* 24:1596-9.

Van Der Veen, J., and G. Kok. 1957. Isolation and typing of adenoviruses recovered from military recruits with acute respiratory disease in The Netherlands. *Am J Hyg* 65:119-29.

Tcherepanov, V., Ehlers, A., and Upton, C. 2006. Genome Annotation Transfer Utility (GATU): rapid annotation of viral genomes using a closely related reference genome. *BMC Genomics* 7: 150.

Toogood, C. I., J. Crompton, and R. T. Hay. 1992. Antipeptide antisera define neutralizing epitopes on the adenovirus hexon. *J Gen Virol* 73:1429-35.

Wadell, G. 1984. Molecular epidemiology of human adenoviruses. *Curr Top Microbiol Immunol* 110:191-220.

Wadell, G., M. L. Hammarskjold, G. Winberg, T. M. Varsanyi, and G. Sundell. 1980. Genetic variability of adenoviruses. *Ann N Y Acad Sci* 354:16-42.

Walsh, M. P., A. Chintakuntlawar, C. M. Robinson, I. Madisch, B. Harrach, N. R. Hudson, D. Schnurr, A. Heim, J. Chodosh, D. Seto, and M. S. Jones. 2009. Evidence of molecular evolution driven by recombination events influencing tropism in a novel human adenovirus that causes epidemic keratoconjunctivitis. *PLoS One* 4:e5635.

Walsh, M. P., J. Seto, M. S. Jones, J. Chodosh, W. Xu, and D. Seto. 2010. Computational analysis identifies human adenovirus type 55 as a re-emergent acute respiratory disease pathogen. *J Clin Microbiol ePub*:Dec. 30, 2009.

Waye, M.M.Y.; Sing, C.W. Anti-Viral Drugs for Human Adenoviruses. *Pharmaceuticals* **2010**, 3, 3343-3354.

Williams, J., T. Grodzicker, P. Sharp, and J. Sambrook. 1975. Adenovirus recombination: physical mapping of crossover events. *Cell* 4:113-9.

Williams, J., H. Young, and P. Austin. 1975. Complementation of human adenovirus type 5 ts mutants by human adenovirus type 12. *J Virol* 15:675-8.

WordPress <https://wordpress.com> 2015

Yang, Z., Z. Zhu, L. Tang, L. Wang, X. Tan, P. Yu, Y. Zhang, X. Tian, J. Wang, Y. Zhang, D. Li, and W. Xu 2009. Genomic analyses of recombinant adenovirus type 11a in China. *J Clin Microbiol*.

Zhu, Z., Y. Zhang, S. Xu, P. Yu, X. Tian, L. Wang, Z. Liu, L. Tang, N. Mao, Y. Ji, C. Li, Z. Yang, S. Wang, J. Wang, D. Li, and W. Xu. 2009. Outbreak of acute respiratory disease in China caused by B2 species of adenovirus type 11. *J Clin Microbiol* 47:697-703.

BIOGRAPHY

Elizabeth Liu grew up in Beijing and Virginia. She attended the Virginia Polytechnic Institute & State University, where she received her Bachelor of Science in 2005. She went on to receive her Master of Science from George Mason University in 2006. She then received her Doctorate in Science from George Mason University in 2015