MOLECULAR EXAMINATION OF THE EVOLUTION AND SEROTYPES OF HUMAN ADENOVIRUSES

by

Michael P. Walsh

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:

Donald Seto — Dr. Donald Seto, Dissertation Director

Morris Jones — Dr. Morris Jones II, Committee Member

Patrick Gillevet — Dr. Patrick Gillevet, Committee Member

James D. Willett — Dr. James Willett, Committee Member

Saleet Jafri — Dr. Saleet Jafri, Department Chairperson

Richard Diecchio — Dr. Richard Diecchio, Associate Dean for Academic and Student Affairs, College of Science

Vikas Chandhoke — Dr. Vikas Chandhoke, Dean, College of Science

Date: 9/22/2010 — Fall Semester 2009
George Mason University
Fairfax, VA

Molecular Examination of the Evolution and Serotypes of Human Adenoviruses

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

By

Michael P. Walsh
PhD, MS, BS
George Mason University, 2010

Director: Donald Seto, Professor
Bioinformatics and Computational Biology

Fall Semester 2010
George Mason University
Fairfax, VA

DEDICATION

This is dedicated to Irene K. Walsh, and Mary K. Kearney whose sacrifices made this document possible.

ACKNOWLEDGEMENTS

I would like to thank Dr. Donald Seto and the members of his lab (Jason Seto, Liz Liu, and Shoaleh Dehghan) for their help and support.

TABLE OF CONTENTS

LIST OF FIGURES

ABSTRACT

MOLECULAR EXAMINATION OF THE EVOLUTION AND SEROTYPES OF HUMAN ADENOVIRUSES

Michael P. Walsh, PhD

George Mason University, 2010

Dissertation Director: Dr. Donald Seto

Human adenoviruses (HAdVs) are important human pathogens. Recent advances in genome sequencing and sequence analysis have made it possible to study the evolution of these viruses in new and interesting ways. These technologies have been used to study recombinant and non-recombinant HAdVs. Data from studies of these genomes have also revealed possibilities for improving the current methods for classifying HAdVs.

CHAPTER 1: Introduction

1.1. HAdV biology

1.1.1. AdV

Adenoviruses (AdV) are DNA viruses that comprise the family *Adenoviridae*. The first members of this family of viruses were simultaneously isolated as a non-specific human respiratory infectious agent from the adenoid tissue of a child in 1953 [1] and also simultaneously by the military [2]. Since their identification, many additional members of the *Adenoviridae* family have been identified. These viruses have been shown to affect all vertebrates, from birds and fishes to higher primates including chimpanzees and humans. Members of this family are separated into five genera, with the taxonomy still undergoing revision, particularly those based on new genome and molecular-based data [3]. Human adenoviruses (HAdV) can cause pathologies or may have no apparent consequence to the host. The symptoms of adenovirus infections vary, and are organ and tissue specific, causing illnesses ranging from the common cold to severe gastrointestinal distress and including death [4]. In spite of the diverse nature of the *Adenoviridae* family, the members of this family share elements of a common morphology and similar genomic organization.

AdVs are non-enveloped or “naked” and have an icosahedral shape. Their genetic information consists of a double-stranded linear DNA genome covalently linked to a protein (all enclosed in a protein nucleocapsid). The diameter of the virus particles is usually 60 to 90 nanometers. Adenovirus genomes have variable sizes ranging approximately 30 to 38 kilobases in length. All adenovirus genomes produce a core set of proteins, transcribed as “early” or “late” genes, that play essential roles in the virus' replication, expression, structure and function [5].

One of the polypeptides produced by AdVs is called “II”, based on a transcriptional scheme (Figure 1). This protein is often referred to as the “hexon monomer” because trimers of this polypeptide have a hexonal shape. These trimers contribute the outer structure of the virus. Three other proteins, which are called “VI”, “VIII”, and “IX”, are thought to associate with the hexon protein to provide structural stability to the virus particles. AdVs also produce a polypeptide called “III”. This protein forms a pentamer that is referred to as a penton base. The penton base associates with 5 molecules of polypeptide IIIa and the resulting complex is essential to the virus' penetration into host cells. AdVs also produce protein IV that forms trimers called fiber proteins. These fiber proteins are involved in host interaction and receptor binding

resulting in cell tropism. Polypeptides V and VII of the adenovirus are called core proteins and are thought to associate with the virus DNA to serve a histone-like packing function. Finally, the TP or terminal protein is cross linked to the genome DNA and serves a function in adenovirus genome replication [5].

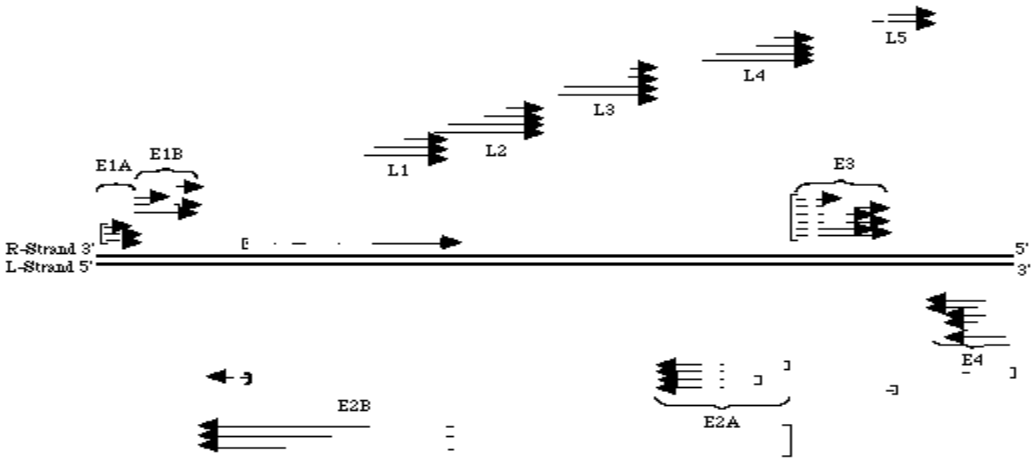


Figure 1. Arrangement of the genes of an adenovirus genome. The multiple elements (lines) associated with each gene represent alternative splicing products [5].

The hexon, penton and fiber proteins are of great interest as they serve to define the individual virus. As they are external, they serve as recognition site for interactions with the host and are subject to evolutionary forces. They are useful for 'typing' such as with antibodies and serve as the first method for separating serotypes and species.

The overall structure of the adenovirus particle contains 240 hexon complexes and 12 pentons. In addition, each of the pentons has a fiber protein attached to it. A schematic representation of a cross-section of the particle is contained in the Figure 2.

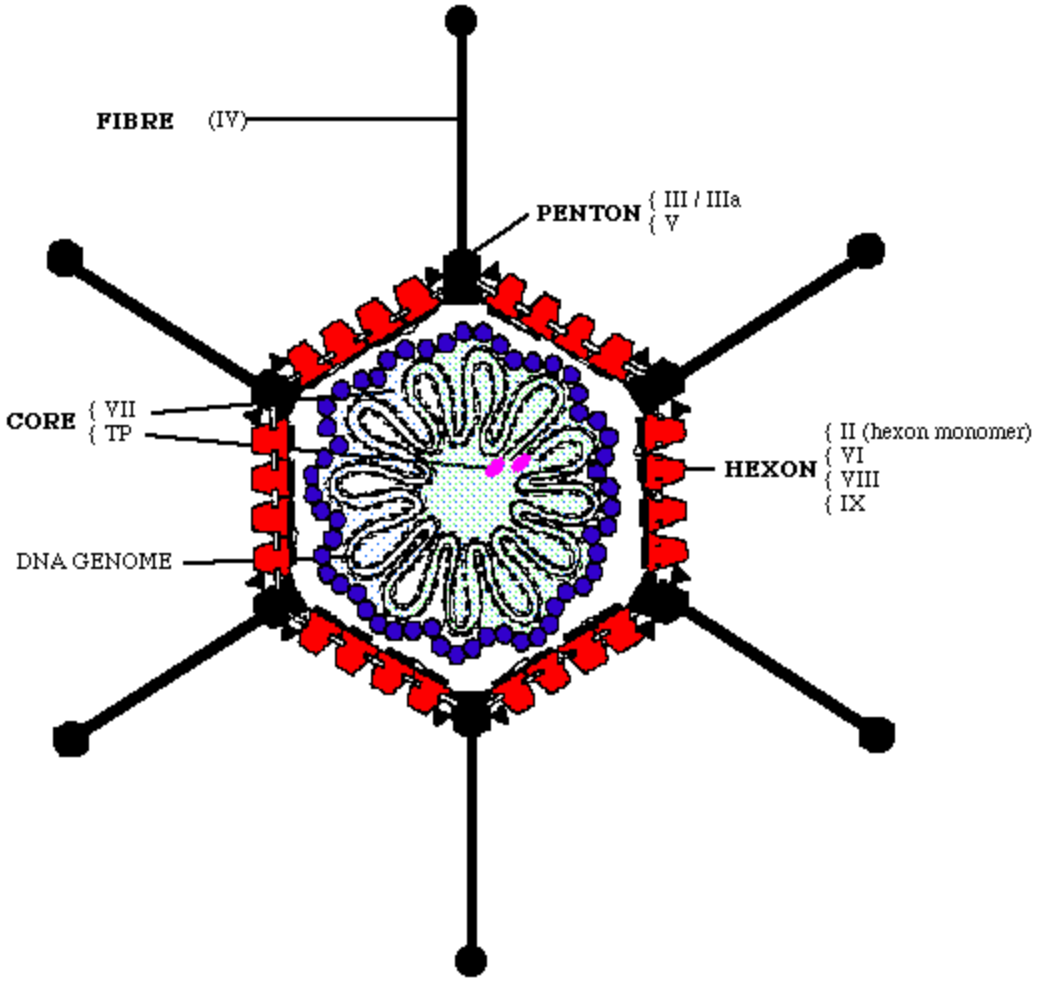


Figure 2. A cross-section of the adenovirus particle. The hexon proteins provide the particle with structural stability. The penton and fiber proteins are involved in virus penetration and host recognition. The core proteins provide a histone-like packing capability for the DNA [5].

1.1.2. Transcriptional Circuitry

The life cycle and genomic arrangement of AdVs are conserved from species to species. The adenovirus life cycle is separated into three categories based on the transcriptional pathways that these viruses use. The first portion of the cycle is often referred to as the “immediate early” phase of virus replication. During this phase the virus produces trans-acting regulatory factors that are responsible for activating the

downstream genes in its genome [5]. The second phase of the life cycle is the “early” phase. The genes involved in this phase produce products that help the virus to evade the host immune response, prevent host cell apoptosis, and activate other genes [6]. The genes of the immediate early and early phases also alter the host cell so that its proteins can be used to replicate the adenovirus genome. The last phase of the adenovirus cycle is called the “late” phase. This phase is responsible for the production of the virus' structural proteins and their assembly into finished virus particles. All of these phases express genes that have alternatively spliced products. This allows the virus to produce proteins that can easily interact with the host cells machinery [6].

Each adenovirus genome contains complementary inverted terminal repeats (ITRs) at each end. The ITRs allow single strands of the virus genome to form origins of replication. These single strands are dissociated from the DNA duplex during the virus’s asymmetric strand synthesis [7]. AdV ITRs also provide transcriptional elements. These elements have highly conserved sequences including core replication and host transcription binding sites [8].

Adjacent to the 5’ ITR, the first portion of the adenovirus genome encodes the E1A and E1B genes. These genes are part of the immediate early phase of the virus life cycle and code for products that play a role in the transcriptional activation of genes that are located later in the genome. Further along in the genome are the E2A and E2B genes. These genes play a role in the early phase of virus replication. Finally, located at the end of the genome, there is a cluster of L genes and E3 genes. Products from these genes are responsible for the late phase of virus replication. L genes encode for structural proteins and the E3 products have been shown to play a role in preventing host cell apoptosis [5]. Figure 1 shows a schematic representation of an adenovirus genomic arrangement.

All adenovirus replication occurs inside the host cell nucleus. Prior to replication, the virus particle must penetrate the cell. As shown in Figure 3, in the first step in this process, a fiber (knob) molecule of the adenovirus particle binds to a specific host cell receptor, hence cell tropism. Next, the virus particle is taken into the host cell vacuole through phagocytosis. The pentons of the adenovirus particle have properties that are toxic to host cells and this toxicity causes the vacuole to break open, releasing the virus particle into the host cell cytoplasm. Finally, the virus particle migrates to the nuclear membrane and injects its DNA into the nucleus through a nuclear pore [5].
Note, the mechanism(s) of virus binding is(are) still vastly unknown. The penton may also serve as a recognition protein for cell binding.

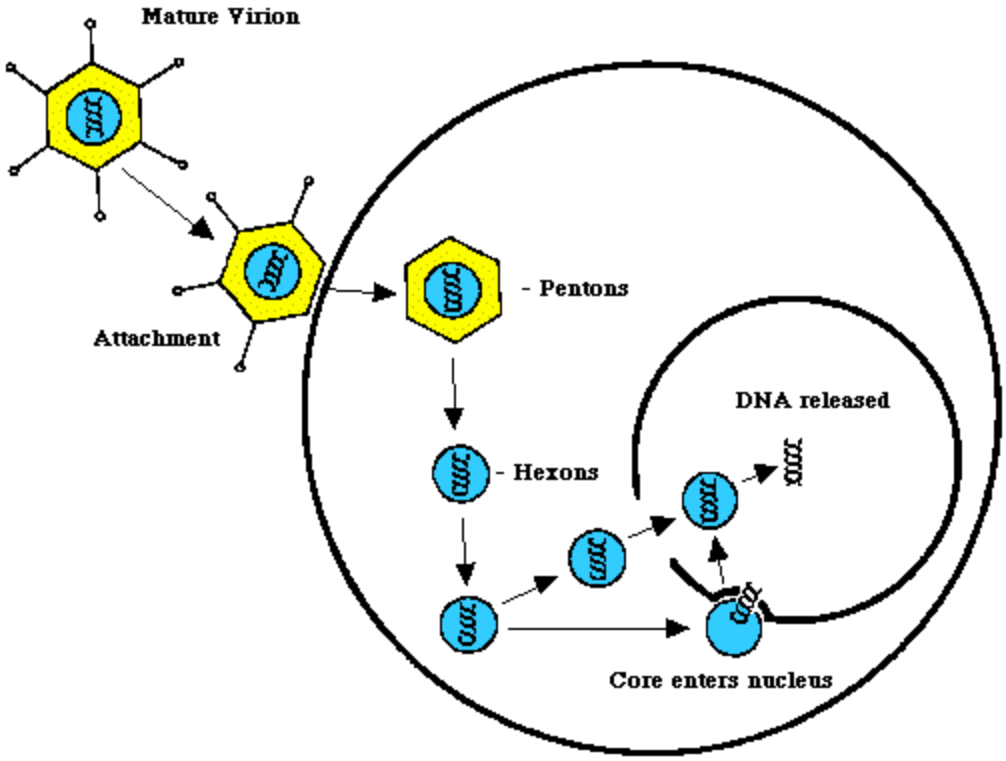


Figure 3. Penetration of an adenovirus particle into a host cell [5].

1.1.3. HAdV

HAdVs are members of the genus Mastadenovirus [3]. There are currently 55 different types of HAdVs that are separated into 7 species based on biology, proteome, immunochemistry and DNA homology. Three additional novel types are reported based on genomics [9,10,11]. Both the type number and the species designations are still being debated, based on changing definitions and genome and molecular data. For example, Number 52 and species G are recent additions [12]. Species are noted by the letters A through G. These species are of importance because they infect humans (and other primates, depending on definitions) and many of them have been studied extensively, from a biological and clinical point of view.

Species HAdV-A contains three types (HAdV-A12, -A18, and -A31) [13]. The most studied of these viruses is HAdV-A12. This serotype has been completely sequenced and annotated. HAdV-A18 and -A31 have also been sequenced and the analysis of HAdV-A18 will be discussed in Chapter 3 of this document [14]. The most notable characteristic of this species is that it has been shown to cause the development of sarcomas in experimentally infected hamsters [15].

Viruses in species HAdV-B are divided into two sub-species labeled B1 and B2. The types of the B1 subspecies are -B3, -B7, -B16, -B21, and -B50. The genomes of these types are sequenced and, with the exception of HAdV-B50, have been shown to cause upper respiratory infections. The members of the B2 species are types -B11, -B14, -B34, -B35, and -B55. These genomes have also been sequenced and, with the exception of HAdV-B14 and B55, have been shown to infect the human urinary tract and kidneys [16]. HAdV-B14 and HAdV-B55 have been implicated in respiratory infections, a clinical finding which has been explored further with genomics and bioinformatics [17,11].

Viruses in species HAdV-C include types -C1, -C2, -C5, and -C6. HAdV-C1, -C2 and -C5 have been sequenced. The members of this species have been implicated in a large number of respiratory and asymptomatic infections. Infections caused by members of the HAdV-C species may have latent effects that are not well-understood. The specific cell types that host this species of adenovirus have not been identified [18]. In addition, HAdV-C5 is a potential vaccine and vector candidate which will be discussed in greater detail in Chapter 3 [19].

Species HAdV-D contains approximately 32 types. Thirteen of these serotypes (8, 9, 17, 19, 22, 26, 28, 36, 37, 46, 48, 49 and 53) are sequenced. The types of this species have been linked to a variety of symptoms. For example, HAdV-D36 has been suspected as a cause of weight gain and obesity in other vertebrates [20]. Also, several members of the HAdV-D species are known to cause severe eye infections, such as epidemic keratoconjunctivitis (EKC) [21,10]. One recent isolate, HAdV-D53, is highly contagious, causes EKC and is the result of molecular recombinations [10]. HAdV type 49 has been explored as a possible genetic vector [22].

The only virus in species HADV-E that infects humans is HAdV-E4. In addition, there are also four chimpanzee-hosted AdVs that are contained in species HAdV-E4. HAdV-E4 is one of the primary agents of an acute respiratory disease (ARD) that affects a number of military recruits. Some reports suggest that HAdV-E4 is responsible for as much as 99.9% of the adenovirus-caused ARD in the US military [23]. This particular type is interesting as the recent versions appear to have a single recombination in the Inverted Terminal Repeat (ITR) that appears to result in a more robust infection (personal communication with Donald Seto).

As a result of the public health implications of HAdV-E4, it is one of the few HAdVs for which there is a vaccine [4].

Species HAdV-F contains two viruses, HAdV-F40 and HAdV-41, which are linked to gastrointestinal ailments [24]. They are also described as fastidious because they are difficult to culture *in vitro* [25]. Both of these HAdV have been fully sequenced.

HAdV species G is the most recently reported and accepted HAdV species. The sole human representative of this species, HAdV-G52, has been implicated in outbreaks of gastroenteritis. This species is also the first to be characterized and identified using bioinformatics techniques [12].

1.2. HAdV Evolution

Molecular evolution is the study of evolution at the nucleic and amino acid level. It is integral to the study of HAdV diversity. Changes at the genome level of HAdV have led to an expanding number of new types [9,12,10,11]. These changes range from small insertion/deletions (indels) and substitutions to homologous recombination events. The study of the HAdV diversity is advancing quickly due to technological advancements in whole genome sequencing and bioinformatics methods.

HAdV genomes are relatively small and encode less than fifty proteins. As a result of this, small indels and substitutions in HAdV genomes can have a large effect on these viruses. Recent studies have been conducted on the consequences of mutations in the fiber gene of HAdV-B11. These studies found that a substitution in a single codon, which encodes for an arginine residue at position 279 of the protein, is capable of changing the phenotype of the virus to that of HAdV-B7 [26]. Studies such as these help to illustrate the importance that indel mutations play in HAdV diversity.

Another important driving force in the molecular evolution in HAdV is homologous recombination. Recombination events are emerging as a primary source of HAdV diversity. Several recent studies have described recombination events that have led to changes in cell tropism and phenotype of HAdV [27,10,11]. These events represent an important source of new HAdV types and their study can help researchers to a greater understanding of how HAdV evolve.

In the past, the detailed study of HAdV evolution has been hindered by technology. The limited number of available HAdV genome sequences limited high resolution studies. Two examples of low resolution HAdV analyses are serum neutralization (SN) and hemagglutination inhibition (HI) assays. Serum neutralization assays have been used to explore epitopes contained within the HAdV hexon protein. This type of assay has been used to classify fifty one distinct “serotypes” of HAdV [21,28]. Hemagglutination inhibition assays have been used to examine the characteristics of the HAdV fiber protein and further classify the viruses [21,28]. Until recently, these assays were the only methods available to differentiate HAdVs.

SN and HI assays possess inherent limitations that can interfere with the study of HAdV. One of these limitations is that the assays are logistically difficult to perform. The tests take time and reagents, such as the viruses themselves, can be difficult (or impossible) to culture or manufacture. The difficulty in growing the reagents can slow down research. Also, the results of the assays (like many antibody assays) can be somewhat subjective. This can lead to incorrect or inconclusive results. Another limitation of these assays is the proportion of the genome that they examine. SN targets epitopes that are contained within two major loops of the hexon protein [29]. The portion of the genome that encodes for these loops is less than three percent of the whole genome. HI tests target the knob region of the fiber protein [30]. The fiber knob is encoded for by less than two percent of the genome. Even when used together, these assays provide only a narrow view of a HAdV genome which is only moderately useful for describing the origins and evolution of the virus.

A more recently applied technique for the study of HAdV genomes employs restriction endonuclease (RE) analysis [31]. In this method, a genome is digested with a panel of defined REs. The resulting genome fragments are subjected to electrophoresis which creates a "restriction map" from the unique sized fragments. Restriction maps from different HAdVs are compared to explore the degree of similarity and differences between genomes.

RE analysis has advantages when compared to antibody-based methods such as SN and HI. One advantage is that RE analysis explores the whole primary nucleotide sequence that is ignored by antibody assays. Another advantage of RE maps is that they can be used to examine the entire length of a genome rather than the smaller portions that are covered by SN and HI assays.

A disadvantage of RE analysis is the result of low resolution of RE maps. RE analysis only examines RE sites, which constitute a small proportion of a genome. This means that two genomes with identical RE maps are not necessarily identical. Also, small differences in RE sites can lead to large differences in RE maps. Relationships between genomes must be inferred from RE maps rather than identified from a base by base examination of the genome sequences. Higher resolution techniques must be used to characterize accurately HAdV sequences and their evolution.

Advancements in whole genome sequencing are providing an abundance of data for HAdV evolution research. Combined with bioinformatics tools, whole genome sequences provide a high-resolution picture of how HAdVs change or evolve in time. It is no longer necessary to infer relationships based on antibody or RE data. With sequence alignments, recombination scans and other tools (discussed in Chapter 2 of this document), the process and pathways by which HAdV species evolve can be characterized in more accurate and meaningful ways.

Whole genome studies of HAdV are currently providing data that are important in several different ways. First, these data are important because they provide information about human pathogens. HAdV-B14 is a recently emerging pathogen that has caused fatal acute respiratory disease (ARD) outbreaks among civilian and military personnel in the US [32,33,17]. The whole genome analysis of HAdV-B14 (described in Chapter 3 of this document) may provide information that can help researchers to determine the origins of this virus. Second, these data are important because they provide information about potential human vaccines and vectors. HAdV-C5 is commonly used in vaccines and genetic vectors because of the virus's stability over time [19,34,35]. Data from bioinformatics studies of HAdV-C5 (Chapter 3) will help scientists to ascertain the source of the virus's resistance to mutation/recombination and will aid in the search for other HAdV vaccine/vector candidates. Third, these data are important because they provide information about less commonly researched HAdVs. HAdV-A18 belongs to HAdV species A which has been understudied in the past [14]. The data from examinations of viruses such as HAdV-A18 can be useful as a reference in the exploration of vaccines/vectors and in recombination studies.

Whole genome studies have also led to a greater understanding of the role that recombination plays in the evolution of HAdVs. Inconsistent SN and HI results have

caused researchers to suspect that certain HAdVs are recombinant [36,37,38]. Recently, an emerging recombinant HAdV ocular pathogen was isolated in Germany [39]. The resulting bioinformatics characterization of this virus, HAdV-D53 (discussed in Chapter 4 of this document), provided the first computational confirmation of recombination events in HAdVs at this high-resolution genome level [10].

Recombination studies have been informative as to the nature of re-emerging HAdV pathogens. HAdV-B55 (discussed in Chapter 4) has been linked to sporadic outbreaks of ARD in China and elsewhere [40,41]. Early sequenced based examinations of the hexon of HAdV-B55 indicated that the virus was a variant of HAdV-B11 [41]. However, HAdV-B11 is a renal pathogen that is not associated with respiratory illnesses [16]. Recombination studies have revealed that the majority (97%) of the genome of HAdV-B55 comes directly from HAdV-B14, a respiratory pathogen. Furthermore, the virus contains hexon a recombination with HAdV-B11 [11].

Data from the examinations of HAdV-D53 and B55 demonstrate the potential of recombination studies to resolve paradoxes and questions from early HAdV research. Hexon sequence data indicated that both HAdV-D53 and B55 were variants of existing HAdV types (HAdV-D22 and B11, respectively). However, these data were confusing because neither virus shared symptoms or tropism with its supposed parents. Recombination analysis data revealed that these viruses both contain hexon recombinations that led to mischaracterizations of the viruses. HAdV-D53 and B55 do not share the properties of their parent viruses because only a small portion of their genomes are related to their parents. This small portion, the hexon gene, is the source of the apparent paradox in identification.

CHAPTER 2: Sequence Analysis

2.1. Introduction

In the past, HAdV genome research has been hindered by a lack of sequence data and bioinformatics tools. Without this data, the study of HAdVs has been limited to low resolution data such as that gained from restriction enzyme digests and serology-based interpretations of the genome [21,31]. Recently, advances in whole genome sequencing technologies have changed this paradigm. As a result of the dramatic increase in the availability of genome sequence data, the analysis of sequences has become a limiting factor in HAdV genome research. To solve this problem, a systematic approach to the analysis of annotated HAdV genomes, including software tools, has been developed.

2.2. Genome Analysis Process

The sequence analysis process developed begins with a newly sequenced HAdV and includes annotation. In the first step in the analysis of a new sequence, a whole genome multiple sequence alignment (MSA) is created that contains the new sequence along with every other currently sequenced HAdV. As part of this step, percent identity values are calculated from the MSA. Since HAdVs of the same species are similar to one another, the percent identity calculations are informative as to the species of the newly sequenced HAdV. The percent identity values also give researchers the ability to use data from previous studies to elucidate properties of the new sequence. For example, if HAdV-D37 shares a high identity with the new sequence, and HAdV-D37 causes eye infections, it may be worthwhile to explore the possibility that the new sequence causes eye infections as well. Establishing these types of relationships can increase the quality of future studies and reduce the time needed to complete them by giving researchers a guide to studying a new sequence.

Phlyogenomics can be defined as the phylogenetic examination of the molecular data from a genome [42]. The phylogenomics of the newly sequenced HAdV are explored in the next step of the analysis process. In this step, MSA's of the three serological determinants (penton, hexon, and fiber) of the new virus are created. Next, two sections of the hexon MSA are extracted, corresponding to previously published studies and primers that isolate hexon loop 1, the variable and defining region [30], and a portion of the hexon conserved region [43]. The fiber MSA is then analyzed in a similar manner using primers [43] designed to extract the variable knob region of the fiber gene. Finally, bootstrapped (one thousand replicates) neighbor joined phylogenetic trees and

distance matrices are created from the five resulting MSA's (whole genome, penton, hexon loop1, hexon conserved 3 and fiber knob).

Phlyogenetic trees of the new sequence provide several different types of information. The whole genome phylogenetic tree and distance matrix values reconfirm data obtained from genome percent identity measurements. The hexon loop1 and fiber knob trees mimic results from serum neutralization (SN) and hemagglutination (HI) assays which are the classical HAdV typing methods. Using phylogenetic trees in lieu of these assays greatly reduces the time and effort required to gain information pertaining to the hexon and fiber of the new genome.

When examined as a whole, the five phylogenetic trees can be used to determine if a new genome is a variant of an existing genome type or a novel, and perhaps even recombinant, genome type. Figure 4 shows an excerpt from a phylogenomic examination of a sequenced HAdV isolate called SGN-1222. SGN-1222 was originally designated as a variant of HAdV-B11 but forms a clade with HAdV-B55 in all trees. This reveals that SGN-1222 is a variant of HAdV-B55 rather than a novel HAdV or a variant of HAdV-B11. The difference in the clades formed between the two hexon trees reveals that both HAdV-B55 and SGN-1222 contain an intra-hexon recombination between HAdV-B11 and HAdB-B14. Finally, the branch lengths in the clade that contains HAdV-B55, SGN-1222 and HAdV-B11 in the hexon loop1 tree indicates that these three HAdVs will provide identical results in serum neutralization assays.

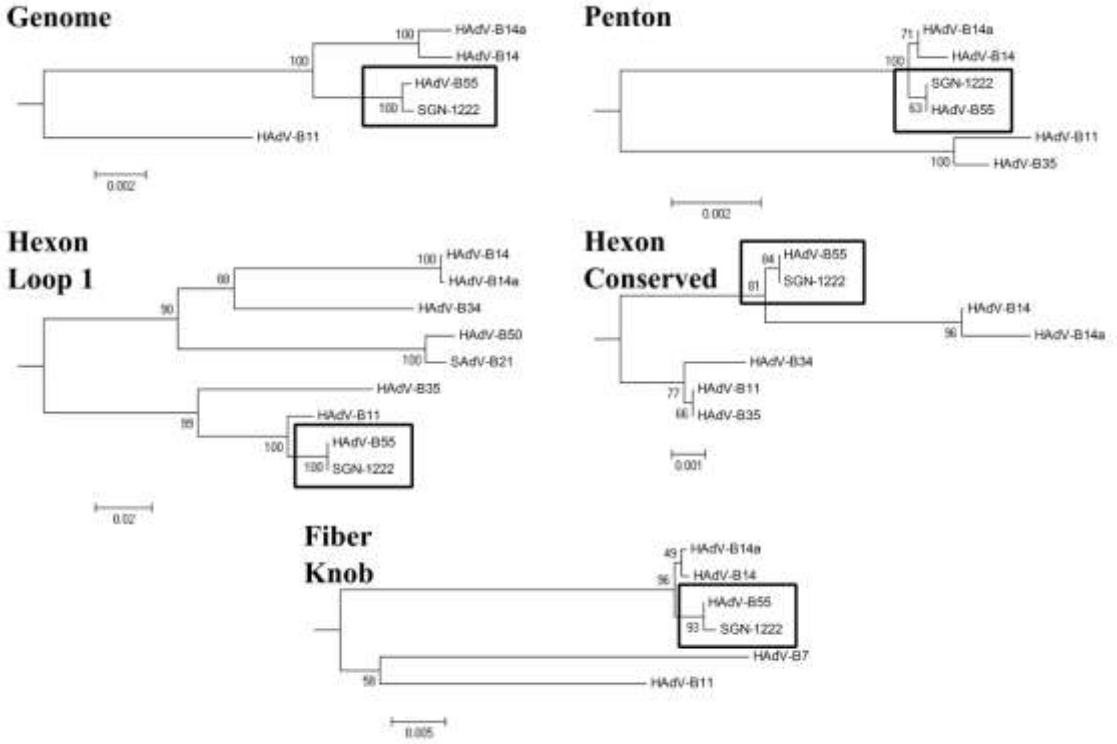


Figure 4. A phylogenomic examination of (SGN-1222) . Selected phylogenetic trees, containing HAdV isolate SGN-1222, are pictured. Clades formed in the trees reveal that the isolate is a variant of HAdV-B55 and that it contains an intra-hexon recombination.

In the third step in the process, the new genome is compared to its closest relative, as established by percent identity and phylogenomic data, using pairwise whole genome alignment visualization. Whole genome alignment visualization allows researchers to examine regions of a new genome to determine relationships and indentify possible recombination events. Figure 5 shows an example of an alignment visualization that examines HAdV-B55 using zPicture software [44]. In the alignment, a close relationship between HAdV-B55 and HAdV-B14 is visible (y-axis 90 – 100% identity). However, HAdV-B55 shows higher identity to HAdV-B11 in the proximal 3rd of the hexon region. This pattern indicates a potential recombination event.

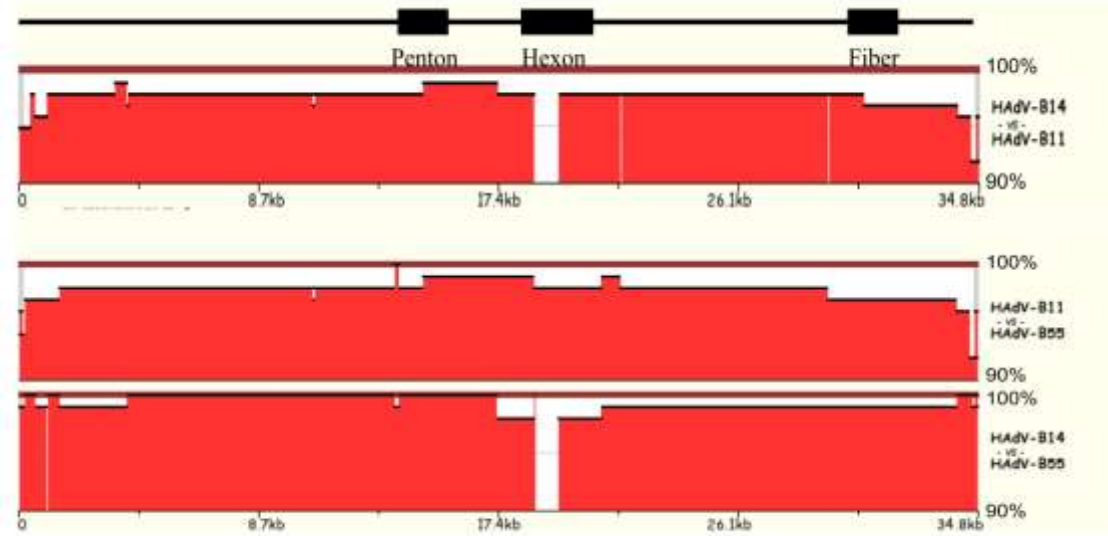


Figure 5. Whole genome alignment visualization. Pairwise alignment visualization of HAdV-B55 versus its closest relatives is shown. HAdV-B14 and HAdV-B55 show high identity to each other throughout most of their genomes. HAdV-B55 shows higher identity to HAdV-B11 in the proximal third of the hexon region indication a possible recombination event in this region.

In the next step in the sequence analysis process, all of the genes and predicted proteins in the newly sequenced HAdV are aligned with homologs from other HAdVs and percent identity values are calculated. The percent identities from the genes are useful in confirming recombination events or other relationships that have been observed

earlier in the sequence analysis process (data not shown). Amino acid percent identities can be used to determine whether or not the relationships are translated as differences at the proteome level.

In the final stage of the analysis, the new genome is formally examined for recombination events using computational tools, such as Simplot [45]. In the first step of this recombination analysis a whole genome alignment containing the new sequences and several related sequences is examined for such events. When recombination events are found in a region of the alignment, that region is excised so that it can be examined using a scan with higher resolution. Finally, results from the recombination scans are combined with whole genome percent identities, phylogenomic data, genome visualization results and amino acid and genes. All of the resulting information is used conjunction with whatever is known about the biological context of the virus to determine the potential for a recombination event.

2.3. Computational Tools

The genome analyses described make use of a variety of computational tools. Many of these are freely available in the public domain. However, some of the tools used were not available and had to be designed and developed.

2.3.1. Genome Alignment and Percent Identity Calculation

Two tools were used to align genomes and calculate percent identities. An internet accessible and relatively new alignment program, called Multiple Alignment via Fast Fourier Transforms (MAFFT) [46], was used to align whole genomes. Percent identity values were calculated, based on these alignments, using the percent identity tool available as part of the UCSF Chimera program [47]. Both MAFFT and Chimera were useful in reducing the time required to complete the genome alignment stage of the sequence analysis process.

MAFFT is a sequence alignment program capable of aligning sequences quickly. The speed of the program is the result of the algorithm's incorporation of Fast Fourier Transforms (FFT), a mathematical formula, to align sequences.

The use of MAFFT to align sequences was convenient, and possibly necessary, due to the size of the alignments used in this project. There a currently over fifty-five sequenced HAdV genomes and the length of these genome ranges from 30 to 38 kilobases. Creating an alignment of this size can take several hours using Clustal, the current standard tool. The same alignment can be constructed in less than ten minutes using MAFFT. Furthermore, MAFFT has been shown to be as, or more, accurate than Clustal [48]. The combination of high speed and accuracy of MAFFT makes it the best option to create HAdV whole genome alignments.

A program called UCSF Chimera was used to calculate percent identities between HAdV genomes [47]. Chimera contains a tool that allows for the computation of percent identities from a MSA. This is in contrast to commonly used software, such as EMBOSS

[49], which requires pairwise alignment of sequences for calculations. The ability to compute genome percent identities from one MSA, rather than the several dozen pairwise alignments that would be required by other programs, reduced the time and effort needed to complete genome identity calculations.

2.3.2. Phylogenomic Analysis

The primary tool used for phylogenomic analysis of HAdV genomes was Molecular Evolutionary Genetics Analysis (MEGA) [50]. MEGA is an open source alignment viewer/editor that provides options for creating phylogenetic distance matrices and trees. All trees were constructed using the bootstrap, neighbor joining option available within the program. Default parameters were used in all analysis.

The distances, upon which the distance matrices and phylogenetic trees are based, were calculated from MAFFT MSA's using the Maximum Composite Likelihood (MCL) method [50,51]. This method maximizes the sum of the log likelihoods for all sequence pairs according to a nucleotide substitution pattern. This method was chosen because it is the default setting in MEGA, a widely used and well tested phylogenetic program.

2.3.3. Genome Alignment Visualization

Whole genome pairwise alignment visualization was completed using a web accessible alignment software suite called zPicture [44]. This program is similar to other types of software, such as mLAGAN [52], in that it uses a BlastZ algorithm [53] to align sequences. The resulting alignment can be visualized in two different ways. The first visualization uses a sliding window (e.g., 100 bp with a 25 bp moving step) to create a smoothed graph of the similarities between sequences. This method of visualization is useful for examining sequences of high similarity because it allows for minute differences between the sequences to be displayed. The second visualization method uses blocks to show regions of similarity between sequences. The size of the blocks is determined by gaps in the alignment, and the percent identity is determined and graphed for each block. This visualization is useful for comparing regions of less similar sequences and can be used to spot potential regions of recombination.

zPicture and other programs (mLAGAN, PipMaker) use identical algorithms for alignments. The choice to use zPicture, rather than other programs, was based on several factors. One of these factors is that zPicture is available on the web (http://zpicture.dcode.org/) which increases the ease with which it can be used. Also, zPicture contains certain options that are not available with comparable programs. The most notable of these options is the ability to visualize alignments using blocks. This capability allowed for the visualization to be used as a further check for potential recombination events.

2.3.4. Gene/Protein Alignment and Percent Identity Calculation

Two tools were used to calculate percent identity values for individual genes/proteins. The first of these tools was a local database (db) of annotated HAdV genomes from GenBank. This db was used to facilitate the retrieval of gene/protein sequences. The second tool was a Java-based protein alignment percent identity calculation program. This program was used to compare homologs from different HAdV genomes. Both tools were developed during this project and will be made available to the general public via web server.

Currently the majority of sequenced and annotated HAdV genomes are stored in GenBank. GenBank, however, is not limited to HAdV and encompasses a large body of all sequence data. Retrieving HAdV sequence data from GenBank is difficult because users must circumvent many thousands of sequences. To alleviate this problem, a local HAdV specific db was created for this project.

The schema for the local HAdV-DB, shown in Figure 6, has four tables that store information from HAdV genomes. The Adenovirus table houses reference information relating to each genome. This table allows user to retrieve general information about the HAdV including the species, serotype, strain, etc. The remaining tables store the sequence information for the virus. The Genome table contains the genome sequences of the viruses so that they can be easily accessed for genome alignments and recombination studies. The NonCodingFtrs table stores sequence information for all of the non-coding features of HAdV genomes. Non-coding sequences are beyond the scope of this investigation but were included in the db in the event that they become useful for future studies. Finally, CodingFtrs table contains nucleotide sequences and amino acid translations of all of the coding sequences in HAdV genomes. This table allows for retrieval of sequences by gene/protein name which is useful when aligning homologs from different HAdV genomes.

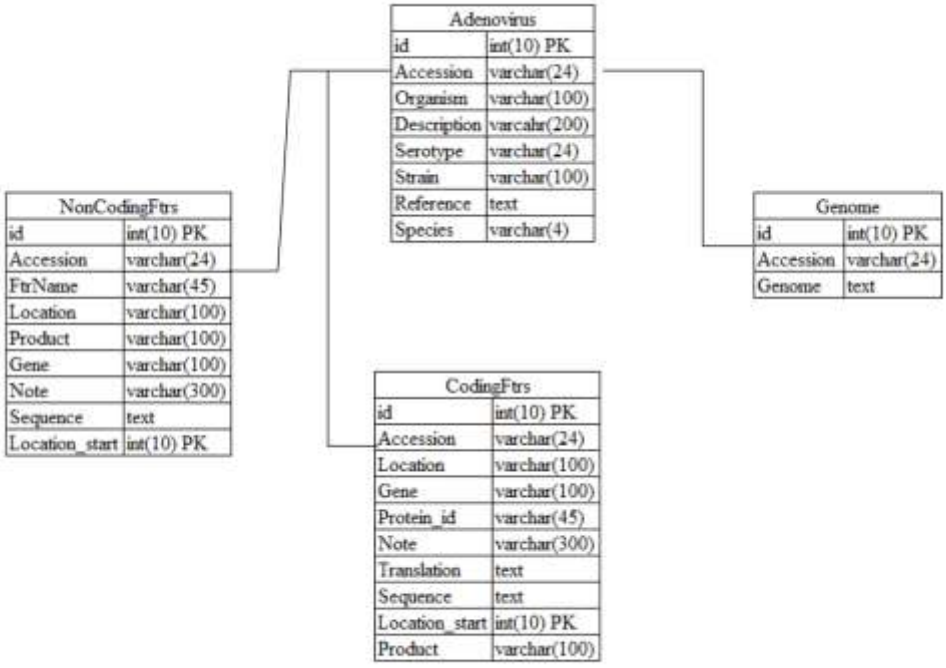


Figure 6. HAdV DB Schema. An entity relationship diagram of the schema of a local HAdV DB is displayed. Lines indicate foreign key relationships.

A series of Java programs was used to populate the HAdV DB. A Java framework called BioJava [54] was used to write a script that retrieves genomes from GenBank; parses out the necessary information; and feeds that data into the appropriate tables of the DB. Another Java framework, Hibernate [55], was used to communicate programmatically with the DB. Hibernate has many functions that simplify the process of ensuring that the information in the DB remains consistent and synchronized.
The majority of the programs available for alignment and calculation of percent identities were insufficient for this project because they focus on determining relationships between one set of homologs rather than all sets of homologs. For example, the percent identity between all HAdV hexons can be easily calculated by isolating hexons from the genomes of interest and running EMBOSS to determine percent identities. However, determining the percent identities between all of the proteins in HAdV-D37 and their homologs in HAdV-D19 would require more than thirty runs of EMBOSS. A Java-based program was written to decrease the amount of time and effort required to compare homologs of different HAdV.

The percent identity program has three main purposes. First, given two HAdV genome annotation files (in GenBank format), the program can align homologs from the

two HAdV genomes and calculate percent identities. Second, the program can compare homologs that are input as lists via fasta formatted files. Finally, the program is capable of parsing coding sequences out of GenBank formatted files.
The results of the program are calculated from alignments constructed using a BioJava implementation of a Needleman-Wunsch algorithm. The software was developed using the Java Spring framework [56] so that it could be made available on the WWW.

2.3.5. Recombination Analysis

There are many different types of recombination analysis software [45,57]. These software use a variety of different algorithms to identify potential recombination events. Among recombination detection programs, Simplot [45] stands out because of its ease of use, logical/understandable algorithm and presentable results. For these reasons, it was the primary program used in the recombination studies discussed in this document.

The Simplot software is capable of producing two types of analyses that, in complementation, can help researchers to identify putative recombination events. The first type of scan is called a similarity plot. This scan examines a nucleotide MSA, using a user-defined sliding window before calculating and graphing the percent identity among the sequences in the alignment. The second type of scan is called a bootscan. A bootscan uses a phylogenetic algorithm to determine which of the regions of a MSA are most likely to contain recombination events.

It is important to understand the bootscan algorithm because any unfamiliarity with respect to the algorithm can lead to misinterpretation of the results from bootscans. In the first step of the bootscan process, an alignment is input into the program and a query sequence is chosen by the user. The query sequence is the sequence that will be searched for recombination events. This alignment is then separated into overlapping windows based on user input parameters. Next, a user defined number of bootstrapped neighbor-joining trees are constructed for each of the resulting overlapping windows. The percentage of the trees in which each sequence in the alignment forms a clade with the query sequence is calculated for each window and graphed as colored lines.

Potential recombination events are easily visible in bootscan graphs as “changes in peaks”. Figure 7, is an example of a bootscan that illustrates this change in peaks. In most of the first half of the bootscan (positions ~300 to 1400), the query sequence clades with HAdV-E4 (the green line). However, the second half of the scan shows a strong relationship between the query sequence and HAdV-B11, indicated by a high blue peak. This change in peaks, from green to blue, suggests that a recombination event may have taken place in the query sequence.

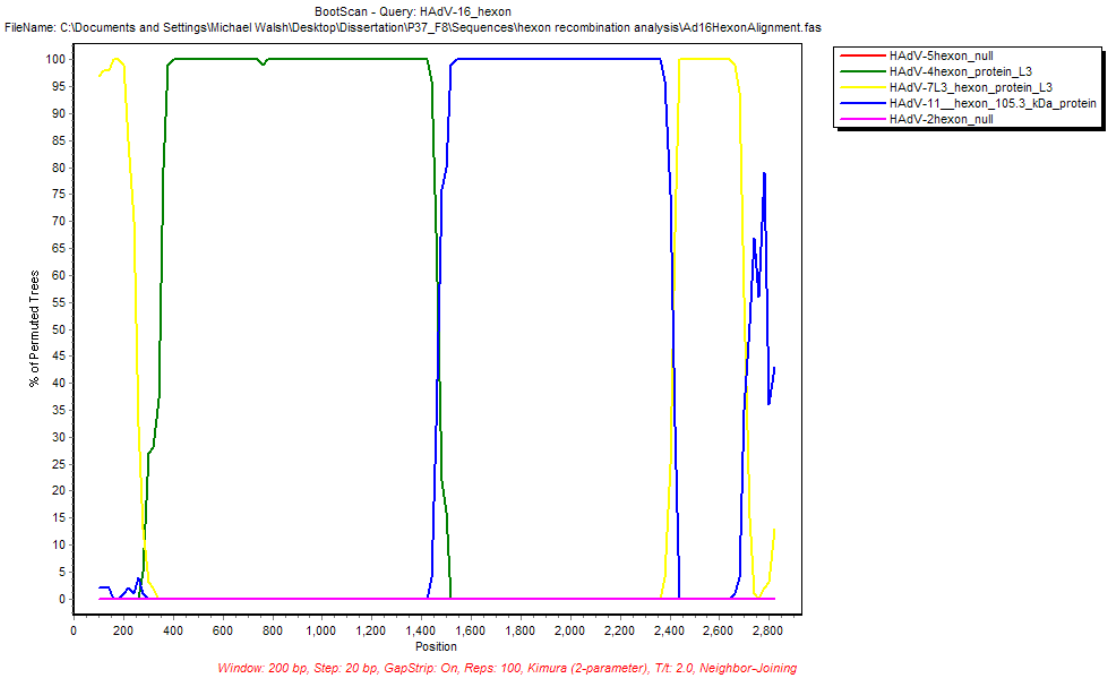


Figure 7. Bootscan example of "change in peaks" phenomenon. A bootscan graph shows potential recombination between HAdV-B16 and HAdV-E4 at the hexon gene.

Identification of potential recombination events in a bootscan is relatively easy given the simple graph output. However, overreliance on bootscan output can lead to misinterpretation of results. Bootscan output measures the (phylogenetic) relationships between sequences on a relative scale. In order to identify correctly a recombination event, constant (percent identity, similarity plot) measurements must also be taken into account. Figure 8, shows an example that illustrates this point. This figure shows a bootscan (top panel) and similarity plot (bottom panel) of HAdV-A18. If one were to examine the bootscan alone, it would be easy to conclude that HAdV-A18 has a recombination event with HAdV-A31 in the first half of the graph. However, the similarity plot graph reveals that HAdV-A18 and HAdV-A31 are only ~85% identical in the (potentially) recombinant region. This type of result could mean that not enough data was included in the initial scan, or that the recombination is very old and that the sequences have since diverged. Either way, the initial bootscan results, taken alone, are misleading and could lead to confusion.

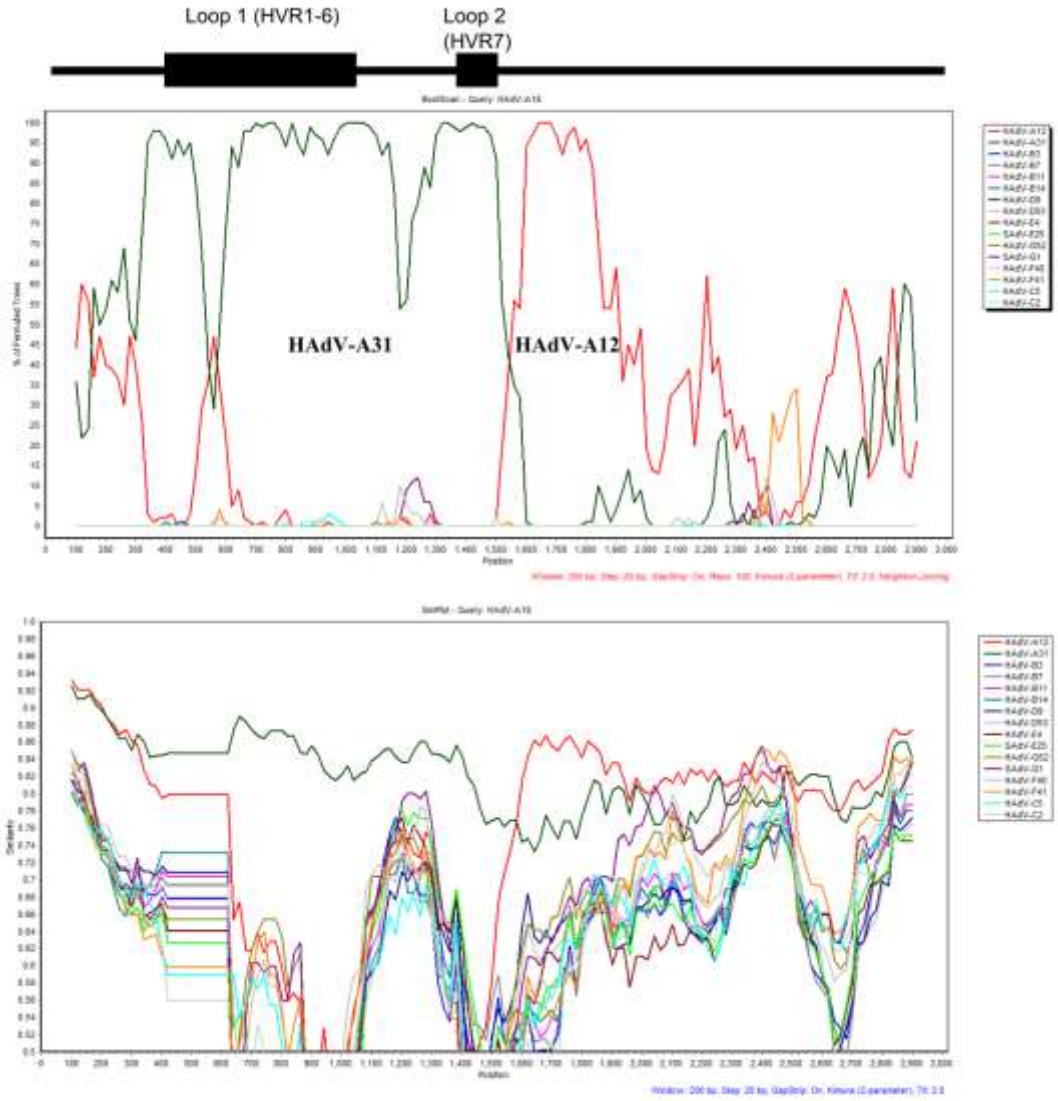


Figure 8. Recombination results for HAdV-A18. The top panel shows the bootscan result for the hexon of HAdV-A18. The panel shows a similarity plot result for the same query. The major loops of the hexon are noted at the top of the figure for reference purposes.

Another potential source of confusion involved in recombination scan results involves the amount of sequence data that is included in the scan. Problems in this area can be caused by including either too much or too little sequence data. Too much sequence data can be a problem if the sequence similarities interfere with the scan, e.g., compete with each other. For example, if two variants of HAdV-E4 are included in the

scan, any potential recombinations in HAdV-E4 will be masked. The extra HAdV-E4 will act as a de facto copy of the query and will always be more similar and more potentially "recombinant" than the other sequences in the alignment. Too little sequence data can become a problem because it forces researchers to draw conclusions based on incomplete data. Figure 9 shows an example of the use of too little sequence data in a bootscan. The figure shows two bootscans of HAdV-D54, a recently sequenced member of the HAdV-D species where the top panel shows an early version of the bootscan that contains the limited number of HAdV-D genomes that were available at the time. This panel shows a potential recombination event with HAdV-D9 in the middle of the graph (in the hexon region). The bottom panel shows a bootscan done on the same query, when more HAdV-D sequences were available. The hexon recombination does not appear in this scan, indicating that this recombination was an artifact caused by a lack of sequence data.

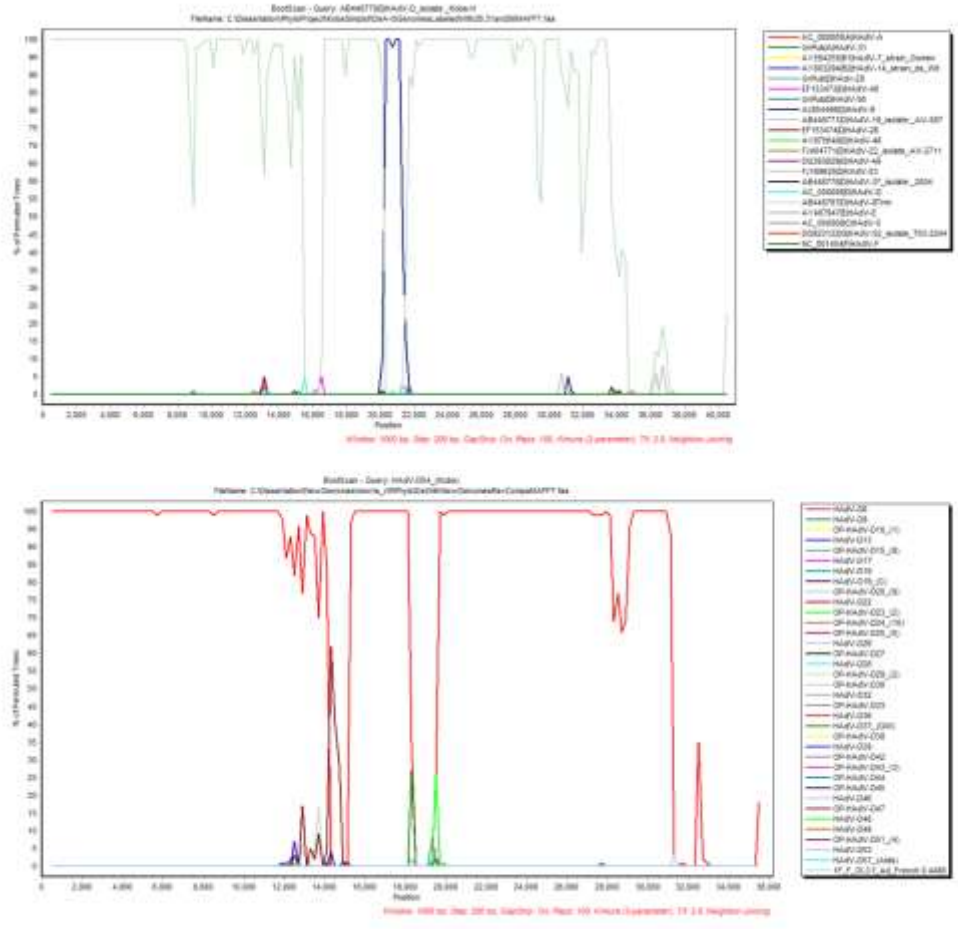


Figure 9. Bootscan of HAdV-D54. The top panel shows a bootscan of the HAdV-D54 genome that contains limited sequence data. The bottom panel shows the same bootscan after more sequence data was included.

Finally, Figure 10 shows an ideal result (in terms of the number of sequences queried and the clarity of the recombination evidence). The figure shows a bootscan (top panel) and similarity plot (bottom panel) of the hexon of HAdV-D53. All available sequences were included for analysis and both graphs show strong evidence of a recombination event between HAdV-D53 and HAdV-D22. Furthermore, the recombinant region has biological significance. This region is biologically significant because it encodes for two major loops of the hexon protein. These loops contain the epitopes for serum neutralization assays [29,30] and, presumably, for antibody binding. A change in this region could give HAdV-D53 a selective advantage over other HAdVs in evading the host immune response.

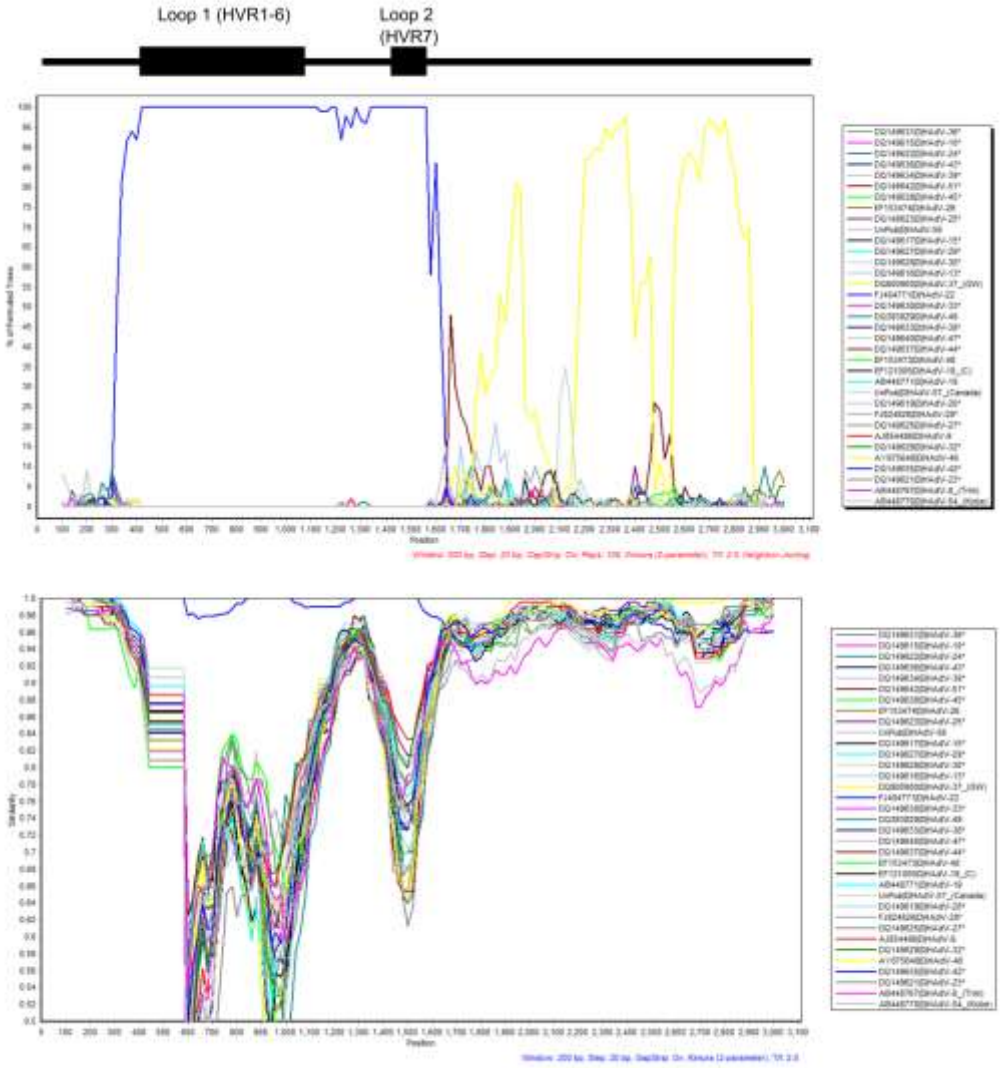


Figure 10. HAdV-D53 hexon recombination. The top panel shows bootscan result for the hexon of HAdV-D53. The panel shows similarity plot results for the same query. The major loops of the hexon are noted at the top of the figure for reference purposes.

When determining whether or not a recombination event has taken place, it is important not to rely too heavily on one type of data. Similarity plot/bootscan results must be independently confirmed using phylogenomic, percent identity, and clinical data.

2.4. Conclusion

The sequence analysis process described in this chapter allows researchers to examine and characterize newly annotated genome quickly and thoroughly. However, the true advantage to this method lies in the fact that every step of the process independently confirms results from the other steps. These data and analyses complement each other. Percent identity data confirms zPicture and Simplot data, Phylogenomic data confirms bootscan and zPicture data, etc. This allows researchers to be more confident about conclusions that are drawn based on the sequence analysis process.

CHAPTER 3: Non-Recombinant Genomes

The study of non-recombinant HAdV genomes is important for several different reasons. First, many of these genomes are human pathogens that are capable of causing severe or even fatal diseases. Exploration of HAdV genomes, such as these, can provide information about the source of the pathogenecity which may be embedded in the genome. Studies of these genomes may also lead to possible treatments or vaccines. Second, HAdV genomes have been used in vaccine and vector development projects. Analyses of genomes used in vaccines and vectors will provide information about the stability of these genomes and their rate of molecular evolution. That data can be used to improve the development process. Finally, many HAdV types are understudied because they are relatively benign or they are difficult to culture. Studies of these genomes can provide data that are useful as a reference in comparative genomic studies of other HAdV. Parts 1-4 of this chapter contain publications that explore each of these aspects of HAdV research.

A detailed genomic and bioinformatics analysis of HAdV-B14 comprises Part 1 of this chapter [17]. HAdV-B14 is a member of subspecies B2, which has been implicated in outbreaks of acute respiratory disease (ARD) throughout the world [32,58,59]. The B2 subspecies is of interest because of the pathogenecity of many of its members. However, unlike HAdV-B14, most B2 HAdVs cause renal ailments. The publication describes the most detailed examination of the B2 subspecies to date and will be useful in determining why HAdV-B14 causes respiratory rather than renal symptoms.

Parts 2 and 3 of this chapter contain publications that examine the vaccine and vector development potential of HAdV. The publication presented in Part 2 describes an analysis of HAdV-C5 [19]. HAdV-C5 is a respiratory pathogen that is often asymptomatic and has been studied for over fifty years. The fifty-year body of data provided by HAdV-C5 analyses provided a unique opportunity to examine the stability of the genome over time. The findings presented in the paper demonstrate the unusual molecular stability of HAdV-C5 and its continued use as a vector or vaccine.

The publication presented in Part 3 of this chapter provides a review of the current data and tools used for HAdV vaccines and vector research [60]. This paper describes detailed analyses of HAdV-B3, B7, HAdV-E4, and HAdV-C5 including recently isolated field strains for comparison. All of the HAdV have been explored as possible genetic vectors and HAdV-B7 and E4 have been used in vaccine development. The paper explores the stability of these strains along with many other factors that may affect their suitability as vaccine/vector candidates.

Part 4 of this chapter contains a paper that describes analysis of HAdV-A18 [14]. HAdV-A18 is a member of species A. HAdV species A is less studied than other species as a result of the limited effects of its members. However, HAdV-A12, another member of species A, has been implicated in tumor formation in rodent models. The analysis of HAdV-A18 reveals that it contains sequence similarity with HAdV-A12 in the regions thought to be linked to the transformation of rodent cells [61,62,14].

The analysis of HAdV-A18 also serves as a reference in present and future comparative genomics studies. Comparative genomics studies, such as recombination and *in silico* RE analyses, depend on having available reference genomes. For example, it is not possible to study fully the recombination history of HAdV species A without data from all of the members of that species. HAdV-A18 was the last member of species A to be sequenced and data from its analysis made a recombination study of that species possible. The results of the recombination study are presented in the publication.

Part 1

Genomic and bioinformatics analyses of HAdV-14p, reference strain of a re-emerging respiratory pathogen and analysis of B1/B2

Virus Research 143 (2009) 94–105

Contents lists available at ScienceDirect

Virus Research

journal homepage: www.elsevier.com/locate/virusres

Genomic and bioinformatics analyses of HAdV-14p, reference strain of a re-emerging respiratory pathogen and analysis of B1/B2

Jason Seto[a], Michael P. Walsh[a], Padmanabhan Mahadevan[a], Anjan Purkayastha[a,b,1], James M. Clark[b], Clark Tibbetts[b,2], Donald Seto[a,b,*]

[a] Department of Bioinformatics and Computational Biology, George Mason University, 10900 University Blvd., MSN 5B3, Manassas, VA 20110, USA
[b] Office of the Surgeon General, US Air Force, Directorate of Modernization (SGR) and the Epidemic Outbreak Surveillance (EOS) Consortium, 5203 Leesburg Pike, Suite 1401, Falls Church, VA 22041, USA

ARTICLE INFO

Article history:
Accepted 23 March 2009
Available online 2 April 2009

Keywords:
Human adenovirus
B1, B2 subspecies
Viral bioinformatics
Re-emerging pathogen
Respiratory disease
Human therapy vector

ABSTRACT

Unlike other human adenovirus (HAdV) species, B is divided into subspecies B1 and B2. Originally this was partly based on restriction enzyme (RE) analysis. B1 members, except HAdV-50, are commonly associated with respiratory diseases while B2 members are rarely associated with reported respiratory diseases. Recently two members of B2 have been identified in outbreaks of acute respiratory disease (ARD). One, HAdV-14, has re-emerged after an apparent 52-year absence. Genomic analysis and bioinformatics data are reported for HAdV-14 prototype for use as a reference and to understand and counter its re-emergence. The data complement and extend the original criteria for subspecies designation, unique amongst the adenoviruses, and highlight differences between B1 and B2, representing the first comprehensive analysis of this division. These data also provide finer granularity into the pathoepidemiology of the HAdVs. Whole genome analysis uncovers heterogeneous identity structures of the hexon and fiber genes amongst the HAdV-14 and the B1/B2 subspecies, which may be important in prescient vaccine development. Analysis of cell surface proteins provides insight into HAdV-14 tropism, accounting for its role as a respiratory pathogen. This HAdV-14 prototype genome is also a reference for applications of B2 adenoviruses as vectors for vaccine development and gene therapy.

1. Introduction

Human adenoviruses (HAdVs) make up the largest group of adenoviruses that have been identified and defined within the family Adenoviridae (Benkö et al., 2005), comprising potentially 53 serotypes (Green et al., 1979; Jones et al., 2007) (unpublished data) which are divided into seven species, based on biology, serology and DNA homology (Green et al., 1979; Wadell, 1984; Wadell et al., 1980). Of the seven species defined, only the members of the B species have been further subdivided into two subspecies, B1 and B2, based in part upon differences and similarities of their restriction enzyme patterns (Wadell, 1984; Wadell et al., 1980). There have been questions of the validity of this subdivision as both may cause similar pathologies; these may be answered given complete genome data and an in depth *in silico* analyses of the proteome of these members.

The B1 and E members (HAdV-3, -7, -16, -21 and -4) are given special attention due to their roles as pathogens causing respiratory diseases. These HAdVs, particularly HAdV-3, -4 and -7, are the most commonly reported agents in respiratory diseases worldwide, particularly in the U.S. military (Cheng et al., 2008; Gray et al., 2007; Larranaga et al., 2000; Lin et al., 2006; Metzgar et al., 2007). The importance of their roles as pathogens in the U.S. is reflected in the recent expedited re-engineering of vaccines against HAdV-4 and -7 (Lyons et al., 2008) and the preliminary development of a vaccine against HAdV-3 in China (Zhang et al., 2009).

While the B1 subspecies has been traditionally associated with respiratory diseases, the B2 subspecies (HAdV-11, -34 and -35) has been mainly associated with infections of the urinary tract and kidney. However, there are reports of the B2 subspecies associated with respiratory disease as well (Metzgar et al., 2005); a report from Turkey found the presence of HAdV-11 in two cases of acute respiratory infections in a military training venue (Chmielewicz et al., 2005). HAdV-14, also a B2 member, has been shown to cause

* Corresponding author at: 10900 University Blvd., MSN 5B3, Department of Bioinformatics and Computational Biology, George Mason University, Manassas, VA 20110, USA. Tel.: +1 703 993 8403; fax: +1 703 993 8401.
E-mail addresses: jseto@gmu.edu (J. Seto), mwalsh5@gmu.edu (M.P. Walsh), pmahadev@gmu.edu (P. Mahadevan), purkayas@wi.mit.edu (A. Purkayastha), jayemmsee@gmail.com (J.M. Clark), clark.tibbetts@tessarae.com (C. Tibbetts), dseto@gmu.edu, dseto99@gmail.com (D. Seto).
[1] Present address: Whitehead Institute, Nine Cambridge Center, Cambridge, MA 02142, USA.
[2] Present address: TessArae, LLC., 46090 Lake Center Plaza, Suite 304, Potomac Falls, VA 20165, USA.

0168-1702/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.virusres.2009.03.011

several epidemic outbreaks of ARD in the U.S. recently, in both military (Metzgar et al., 2007) and civilian settings (Louie et al., 2008; Metzgar et al., 2007). That HAdV-14 is a respiratory pathogen is consistent with the original isolation and identification of HAdV-14 in 1955, as the "de Wit" strain, from throat washings of military recruit with acute respiratory illness in the Netherlands (Van Der Veen and Kok, 1957). It was also associated, in the same year, with an outbreak in a civilian setting in England (Kendall et al., 1957). Its recent appearances in epidemic outbreaks are reminders to not forget about apparent non-circulating pathogens as they may re-emerge unexpectedly, in this case, after 52 years.

The genomes of the B2 subspecies are of interest for three reasons: (1) they cause human diseases; (2) they are candidates for vaccine development; and (3) they may serve as vectors for the gene therapy and vaccine development. Hence, more data and understanding of them will be reflected in applications across these areas. The prototype HAdV-14 genome and its comparative genomics, genome annotation data and *in silico* proteome are presented here as a bioinformatics-based analysis to serve as a reference and a resource both for further respiratory disease pathogen investigations and understanding the biology of the species B adenoviruses and for applications in vector development.

2. Materials and methods

2.1. Strains

The prototype HAdV-14 (de Wit strain), isolated in 1955 (Van Der Veen and Kok, 1957), was obtained from the American Type Culture Collection (ATCC; Manassas, VA).

2.2. Cells, virus stocks and DNA preparation

Protocols for virus growth and DNA production were outsourced to Virapur, LLC. (San Diego, CA) and were essentially as reported previously (Lauer et al., 2004; Purkayastha et al., 2005b,c).

2.3. DNA sequencing

2.3.1. "Shotgun" sequencing of the HAdV-14 genome

Genome sequence determination was outsourced to Commonwealth Biotechnologies, Inc. (Richmond, VA) using methodology as described in an earlier series of related adenovirus genomics and bioinformatics studies (Lauer et al., 2004; Purkayastha et al., 2005a,b,c). A minimum threefold coverage with both strands represented provided high-quality data, with a fivefold redundancy average. Quality control included reconciliation of sequence data with annotation data.

2.4. PCR strategy and methodology

PCR-amplifications and sequencing of these products were as per protocols reported earlier (Purkayastha et al., 2005a,b,c). Problematic regions, generally highlighted during annotation, were resolved and re-sequenced using these protocols. These problems include regions with polynucleotide runs, e.g., polyN and GC compressions, etc. Moving sequencing primers closer and/or resequencing the complement strand also helped to resolve these problems.

2.5. Sequence bioinformatics

The genome was analyzed as described for HAdV genomes earlier (Lauer et al., 2004; Purkayastha et al., 2005a,b,c). For visualizing and recording the annotation progress, the web-accessible Sanger Center tool Artemis was used (Berriman and Rutherford, 2003). Comparative genomics analysis against archived adenovirus genomes representing sequenced members of all species was carried out. These sequences and associated data were retrieved from GenBank. Table 1 displays the serotypes, species, accession numbers, genome sizes and GC content of these adenoviruses. For ease of visualization, a single member of the non-B species was used to represent the species cluster for presentation in the figures.

Multiple-sequence alignments of the proteins were generated by CLUSTALX (Thompson et al., 1997). Phylogeny trees were constructed by the neighbor-joining method described in the literature (Saitou and Nei, 1987). Additional web-based bioinformatics tools include PipMaker, RDP, SimPlot, zPicture and EMBOSS.

Further details of Section 2 are presented in the Supplemental Materials.

3. Results and discussion

3.1. Genome DNA sequence analysis

The genome of HAdV-14p has been sequenced and analyzed using Internet-accessible bioinformatics tools. Selected human and simian adenovirus genomes are presented in the bioinformatics and comparative genomics data reported in these analyses, following initial computational analyses which included the larger set of genomes. This is to keep the figures from being overwhelming. Generally one "type" genome was used for representing a species, with the exception of the B species which are of greater relevance in the discussion of the B1/B2 parsings. This representative genome either was the first of its species to sequenced, or was the best annotated (and arbitrarily chosen (species D)).

Shown in Table 1 is the list of the genomes used, along with physical genome data; these include values from all of the B1 and B2 members (including unpublished data). For the genome size and GC content, a larger set of genome data was used in order to lend weight to the numbers. No significant difference was noted between the larger set of genomes and the subset shown in Table 1. The genome size of HAdV-14p (GenBank accession no. AY803294) is 34,764 nucleotides, which is slightly lower than the mean of 34,789 for the B2 members (nine samples). However, the B2 data are skewed with multiple HAdV-11 and -35 data points. The B2 sizes are slightly shorter than the mean for the B1 subspecies, at 35,363 (14 samples), and clusters to the lower range of sizes for HAdVs in general (33,248–36,204 nucleotides). This slight difference is one indication of the correct division of B1 and B2.

Table 1
HAdV genomes and physical genome data. Human and simian adenovirus genomes and their Genbank accession numbers are accessible from GenBank.

Species	Serotype	Accession no.	Size	GC%	Notes
A	HAdV-12	AC_000005	34,125	46.5	
B1	HAdV-3	AY599834	35,345	51.05	GB
	HAdV-7	AY594255	35,306	51.0	
	HAdV-16	AY601636	35,522	51.28	ch. 91
	HAdV-21	AY601633	35,382	51.25	AV-1645
	HAdV-50	AY737798	35,385	51.23	Wan
	SAdV-21	AC_000010	35,524	51.16	Simian, chimp
B2	HAdV-11	AC_000015	34,794	48.9	
	HAdV-14	AY803294	34,764	48.81	
	HAdV-34	AY737797	34,775	48.94	Compton
	HAdV-35	AY128640	34,794	48.88	Holden
C	HAdV-1	AF534906	36,001	55.3	
D	HAdV-9	AJ854486	35,083	57.1	
E	HAdV-4	AY599837	35,994	57.6	
F	HAdV-40	NC_001454	34,214	51.2	
G	HAdV-52	DQ923122	34,250	55.1	

GC content is a metric accepted for adenovirus classification (International Committee on Taxonomy of Viruses (ICTV); Benkö et al., 2005). The HAdV-14 genome has a GC% of 48.8%, fitting into the B2 subspecies mean of 48.9%. Subspecies B1 has a mean of 51.1%. GC% of other species of human adenoviruses are as follows: A (46.5%), C (55.2%), D (57.0%), E (58.1%), F (51.1%) and G (55.1%).

Thus, both the genome length and the GC content data support the division of species B into two subspecies originally based on RE data (Wadell, 1984; Wadell et al., 1980) and re-affirm the inclusion of HAdV-14 in the B2 group. There may be slight variations for additional isolates, but likely they will fit within the parameters observed to date. For example, it is reported that the recently circulating outbreak HAdV-14 strain, noted as HAdV-14a, is a different genome type from this prototype on the basis of restriction enzyme (RE) digest patterns (Louie et al., 2008). Its publication and comparison with this prototype will be of interest and importance.

3.2. Whole genome-based comparisons and phylogeny analysis

Genomics and bioinformatics, as a combined approach, are allowing detailed insights into pathogens and pathogenicity. The tools available allow fine dissection of related organisms, perhaps accounting for differences in infectivity, robustness, tropism and host responses. Recently this approach has been used to examine the role of HAdV-37 in eye disease (Robinson et al., 2008) and also previously to examine HAdV-4 and -7 roles in acute respiratory disease (Purkayastha et al., 2005a,b), and to examine the vaccine strains of both HAdV-4 and -7 (Purkayastha et al., 2005c), as well as to account for the origin of the E species (Purkayastha et al., 2005b).

Using whole genome data, and detailed in Table 1, additional computational analyses were carried out: whole genome comparisons, multiple-sequence alignment (MSA) and phylogenetic analysis. Software tools designed for the detection of sequence recombination, homology and similarity were used to characterize the HAdV-14 genome further. These include EMBOSS, zPicture, Bootscanning, SimPlot, PipMaker and MAVID. PipMaker and zPicture are used to compare the nucleotide similarities between pairs of genomes. MAVID is an alignment program and is used to perform whole genome phylogeny analysis. Bootscanning and SimPlot are tools used for recombination analysis (http://sray.med.som.jhmi.edu/SCRoftware). EMBOSS is used to calculate percent identities (http://emboss.sourceforge.net/).

3.2.1. Whole genome alignments

Pairs of genome sequences were aligned using zPicture, which is an alignment tool based on the BLASTZ algorithm utilized by PipMaker (http://zpicture.dcode.org/). This tool outputs regions of conservation. Fig. 1 shows relevant subset of these alignments. For this analysis, the *y*-axis comprises the baseline, set at 50% identity (which is adjustable), and the top line, set at 100%; the entire genome length is displayed along the *x*-axis. The darker shading represents (user added) landmarks in the genome, as noted at the top horizontal line. The *y*-axis is plotted to scale so that identity values from the range of 50–100% may be extracted from the plot.

Alignments of HAdV-14 to other B2 members, HAdV-11, -34 and -35, produce similar profiles: high level of identity across the genome, with regions containing partial heterogeneous identity/dissimilarity in the E1A, penton, hexon and fiber regions. This is shown in Fig. 1. The penton region is interesting, with a bipartite division of similarity between HAdV-14 and -11 and -35, unlike a homogeneous lower level of similarity with -34. Bipartite refers to the apparent two blocks of levels of similarity across this gene. The hexons of the B2 show a similar pattern, but of a seemingly

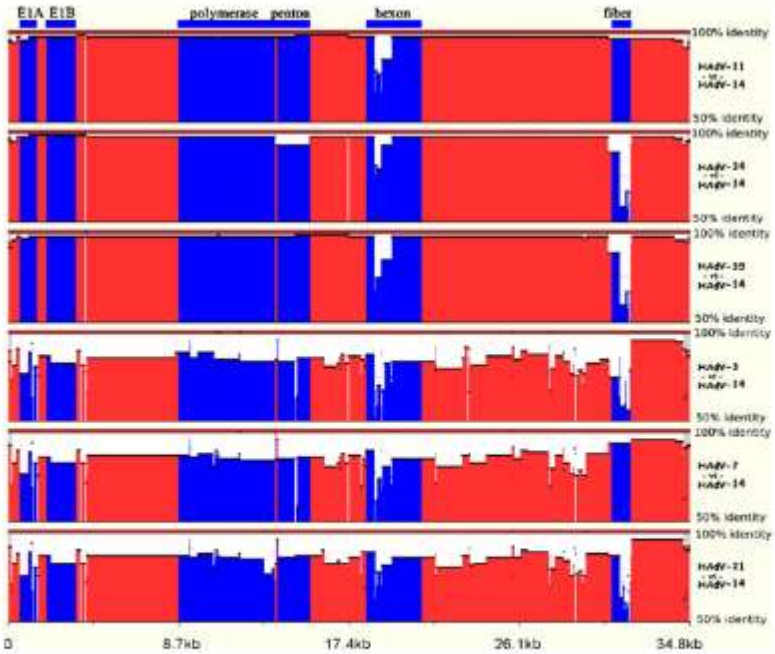


Fig. 1. Genome comparisons. zPicture was used to compare whole genomes for regions of homology/similarity. Along the y-axis, the baseline is set at 50% identity and the top line is set at 100%; the entire genome length is displayed on the x-axis. The darker shading represents landmarks in the genome, as noted at the top.

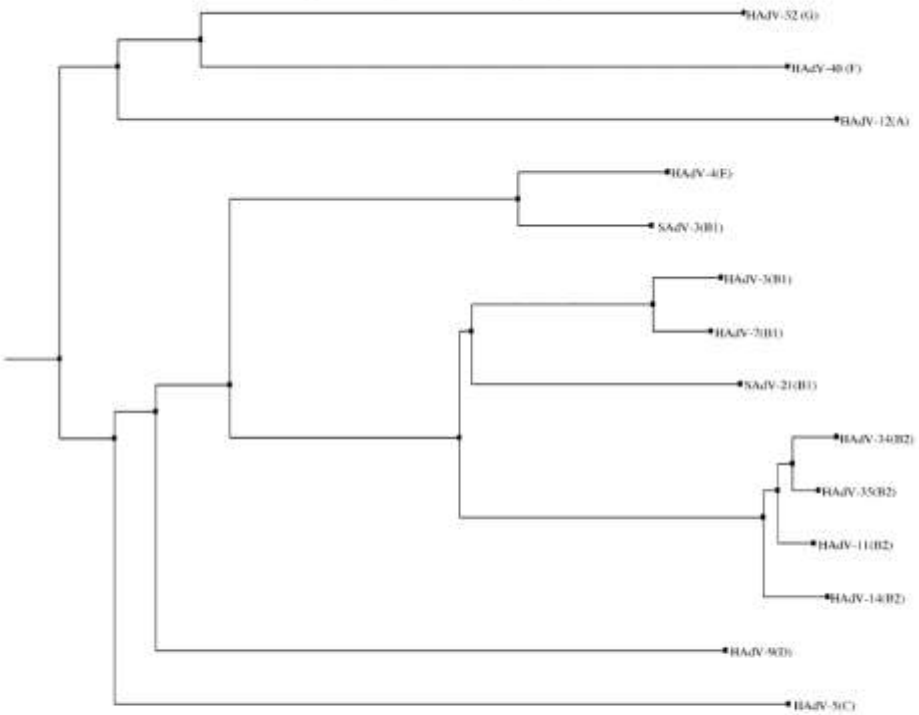


Fig. 2. Whole genome sequence phylogeny. MAVID was used to align a subset of human and simian (chimpanzee) adenoviruses, using the default parameters. (http://baboon.math.berkeley.edu/mavid).

tripartite organization of similarity; that is, two blocks of identity separated by one block that appears to contain areas heterogenous in similarity. HAdV-11 is distinguished at the fiber region from ~34 to ~35 as having a homogeneous level of similarity to ~14 unlike the heterogenous pattern for the other two.

Against the B1s, HAdV-3, -7 and -21, the analysis shows similar profiles, albeit with a lower level of identity across the genomes. The E1A dissimilarity is even more pronounced, with the proximal portion less similar than for the B2s and a shorter distal similarity region, and less similar on the whole. Penton similarity includes a small region in HAdV-3 and -7, but not -21, that is 50% or less in identity. The hexons follow a similar pattern as seen for the B2 comparisons; that is, of a seemingly tripartite organization of similarity, that is two blocks of identity and one block that contains regions that are heterogenous in similarity. A noteworthy observation is that the fiber from HAdV-7 has a constant high level of identity with HAdV-14, like that from HAdV-11; HAdV-7 does not show the heterogenous profile as do HAdV-34 and -35. HAdV-3 fiber is similar to the latter two, as is HAdV-21. These genome alignments and similarities, on the whole, also provide support for the B1/B2 subdivision, at the whole genome level.

3.2.2. Phylogeny alignments of whole genomes

Fig. 2 displays the results of the whole genome phylogeny analysis of a representative group of HAdV as a tree; species designations are included for reference (including two representative chimpanzee SAdVs that are included formally into the HAdV species classifications). MAVID is used; this analysis generates a global multiple alignment and constructs a neighbor-joining phylogeny tree using the default parameters (http://baboon.math.berkely.edu/mavid) (Bray and Pachter, 2004).

These whole genome-based data reveal separate clades of B1 and B2, with branching for both off a common point from the other species. Thus, these data also reconfirm the original inclusion of HAdV-14 into the B2 subspecies and the partitioning of the B1 and B2 subspecies based on RE analysis, and originally designated as "DNA Homology Cluster 1 and Cluster 2" (Wadell, 1984; Wadell et al., 1980). They also confirm the species differentiation at the whole genome level.

3.2.3. ITRs

The inverted terminal repeats (ITRs) are well-characterized genome landmarks found across all adenoviruses infecting vertebrates (Dan et al., 2001). Critical viral DNA replication and transcription functions are imbedded in the ITRs (Rawlins et al., 1984). Fig. 3 shows a multiple-sequence alignment (MSA) of ITRs from a selection of HAdVs. All B2 members have identical ITRs, starting with "CATCATCAAT". This differs from the B1, D, E and A, but is similar to the ones from species C, F and G. The canonical DNA core origin of replication motif (Temperley and Hay, 1992) and the transcription factor binding sites for NFI (Mul et al., 1990) and NFIII (Hatfield and Hearing, 1993) are all present, as displayed in Fig. 3. As noted in the literature, NFI and NFIII binding increases the efficiency of HAdV replication (Hatfield and Hearing, 1993; Mul et al., 1990; Nagata et al., 1982). All three motifs conform to their consensus with the exception of a substitution of a "C" for the conserved "T" at the beginning of the NFIII motif for the two B1 members. HAdV-14 Sp1 and ATF transcription factor binding sites are also present

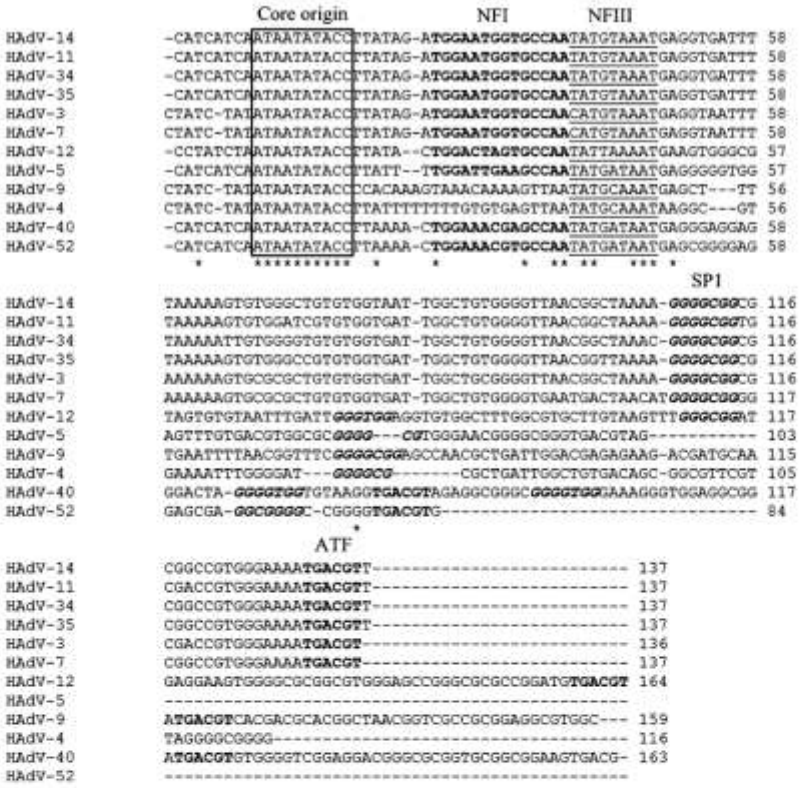


Fig. 3. ITR analysis. ITR sequences are extracted from the GenBank annotation files or from the sequences directly. Alignment was using CLUSTALW. Features are noted.

and identical in sequence and position with respect to their counterparts from the B1 and B2 subspecies, unlike the equivalent sites across the other species. These sequences diverge slightly and their positions vary relative to the same motifs from the other species. These host-supplied proteins also enhance viral replication as well (Hatfield and Hearing, 1993; Mul et al., 1990).

3.3. Bioinformatics of the in silico proteome

The proteome was determined and analyzed using query genomes as noted above, with a goal of understanding further the relationship of the HAdV-14 genome to the other genomes comprising the B species. In particular the E1a, penton, hexon and fiber proteins were analyzed in depth as these proteins, which include variable surface proteins, may serve as references for comparisons that may be useful for understanding the differences between serotypes (or genome types) and species. Because of their importance in defining the B1/B2 subspecies split, the number and similarities of the fiber proteins and the VA-RNA coding regions are also examined. The entire comprehensive set of coding and non-coding sequences was determined and analyzed (data presented in Supplemental Materials).

3.3.1. Protein and coding sequence annotation

Protein and RNA coding sequences annotated in the HAdV-14 genome are displayed in Table 2 in the context of a comparison cataloging these coding sequences for members of the published B2 and B1 genomes. All genomes have been re-examined, i.e., coding regions that are not present in the GenBank record have been "located" in the genome sequence to verify their absence. The HAdV-14 proteome contains all the proteins found in the other members of the B2 subspecies, with synteny retained, and differs from the B1 members slightly as per the B1/B2 differences.

3.3.2. Overview of core genes and the B1/B2 subspecies division

The identification of a set of core genes between organisms of a group may define better common origins and relationships. This concept is emerging as more genome sequences are being acquired. Recently it was noted that bacterial species, for example, can be described by a "pan-genome", which comprises a "core" set of genes that contains all the genes in common amongst related strains and a "dispensable" set of genes, ones that are present in subsets of this group. The dispensable genome allows specialization into niches (Medini et al., 2005). A recent on-going re-evaluation of all of the bacteriophages, starting with the Podoviridae, is using core genes and proteomes as a method of revisiting traditional bacteriophage classifications, allowing clarity across genome sequence recombination for updated classifications of the bacteriophages (Lavigne et al., 2008).

The adenoviruses, as classified into species, have different cell tropisms and different pathologies; in part, this may be due to core and dispensible genes in the genomes. Table 2 displays the coding sequences from the members of both the B1 and B2 subspecies,

Table 2
Annotation of coding sequences from the B2 and B1 genomes.

Gene	Product	HAdV-14	HAdV-11	HAdV-35	HAdV-3	HAdV-7
E1A	29.1 kDa protein	+	+	+	+	+
E1A	25.7 kDa protein	+	+	+	+	+
E1A	6.5 kDa protein	+	+	+	+	+
E1B	20 kDa protein	+	+	+	+	+
E1B	10 kDa protein	+	+	+	+	+
E1B	54.9 kDa protein	+	+	+	+	+
pIX	pIX protein	+	+	+	+	+
IVa2	IVa2 protein	+	+	+	+	+
E2B	DNA polymerase	+	+	+	+	+
Hypothetical protein	A-106 11.5kDa protein	+	+	+	+	+
Hypothetical protein	19 kDa protein	–	–	–	+	+
Hypothetical protein	10.4 kDa protein	–	–	–	+	+
Hypothetical protein	15.3 kDa agnoprotein	+	+	+	+	+
Hypothetical protein	12.6 kDa protein	–	–	–	+	+
E2B	pTP protein	+	+	+	+	+
Hypothetical protein	14.5 kDa protein	–	–	–	+	+
Hypothetical protein	11.3 kDa protein	–	–	–	+	+
Hypothetical protein	9.7 kDa protein	–	–	–	+	+
VA-RNA I	VA-RNA I	+	+	+	+	+
VA-RNA II	VA-RNA II	–	–	–	+	+
L1	43 kDa protein	+	+	+	+	+
L1	pIIIa protein	+	+	+	+	+
Hypothetical protein	6.1 kDa protein	–	–	–	–	–
L2	Penton protein	+	+	+	+	+
L2	pVII protein	+	+	+	+	+
L2	pV protein	+	+	+	+	+
L2	pX protein	+	+	+	+	+
L3	pVI protein	+	+	+	+	+
L3	Hexon protein	+	+	+	+	+
L3	23 kDa protein	+	+	+	+	+
E2A	DNA Binding protein	+	+	+	+	+
L4	100 kDa protein	+	+	+	+	+
L4	22 kDa protein	+	+	+	+	+
L4	33 kDa protein	+	+	+	+	+
L4	pVIII protein	+	+	+	+	+
E3	11.7 kDa protein	+	+	+	+	+
E3	14.6 kDa protein	+	+	+	+	+
E3	18.4 kDa protein	+	+	+	+	+
E3	20.1 kDa protein	+	+	+	+	+
E3	20.8 kDa protein	+	+	+	+	+
E3	7.7 kDa protein	–	–	–	–	+
E3	10.3 kDa protein	+	+	+	+	+
E3	14.9 kDa protein	+	+	+	+	+
E3	15 kDa protein	+	+	+	+	+
U	U protein	+	+	+	+	+
L5	Fiber protein	+	+	+	+	+
E4	Orf6/7 15.9 kDa protein	+	+	+	+	+
E4	Orf6 34.7 kDa protein	+	+	+	+	+
E4	Orf4 14.2 kDa protein	+	+	+	+	+
L5	DNA binding agnoprotein	+	+	+	+	+
E4	Orf3 13.6 kDa protein	+	+	+	+	+
E4	Orf2 14.3 kDa protein	+	+	+	+	+
E4	Orf1 14.2 kDa protein	+	+	+	+	+

HAdV-14, -11, -35, -3 and -7. Data were compiled for all members, including unpublished members, and the others do not differ from the B1/B2 pattern presented. All members of species B have the same set of core genes common to the Bs. There are differences between B1 and B2. Aside from six "hypothetical" proteins found in the B1s and missing in the B2s there are two non-hypothetical coding sequences absent in the B2s that are present in the B1s: one is the E3 7.7 kDa protein and the other is the second VA-RNA gene.

Dispensible genes within the HAdVs may include the E3 transcript-derived proteins, which are critical to the survival of the particular genome type and/or species, as they mediate the evasion from the host immune system (Kajon et al., 2005). They seem to vary across the genomes. Identical to the B2 genomes, HAdV-14 contains eight of the nine genes common to the B species. On the other hand, all the B1s have the full complement. These coding regions indicate the common ancestry of the B species, yet also support the separation into subspecies as well.

3.3.3. *Detailed analyses of E1A, penton, hexon and fiber genes*

As noted in the whole genome comparisons and alignments, E1A, penton, hexon and fiber genes seem to have heterogeneous identity structures relative to HAdV-14; that is, there are local regions of homology and regions of dissimilarity. This may reflect the divergent evolutionary events that are driven by recombination. To address this critical issue, these have been subjected to further analysis, especially the hexon and the fiber. All of the bioinformatics tools, including MSA, phylogeny analysis (bootstrapping and neighbor-joining) and genome alignments (bootscanning, SimPlot and zPicture) reveal the similar patterns for local regions of identities. Each of the blocks of identity and dissimilarity have been identified and extracted and re-analyzed with zPicture, MSA and phylogeny analyses. The resulting data allows higher resolution analysis. These reinforce the idea that there are recombination events in very specific regions of the genome, especially in the fiber and hexon regions (Crawford-Miksza and Schnurr, 1996; Madisch

Table 3
Percent identities of select HAdV-14 proteins spanning the genome. To measure the level of homology between the HAdV-14 and the B2s, relative to the B1s and the other HAdV species, percent identities of the proteins were determined using EMBOSS.

Protein	E1A 29.1 kDa protein	E1B 20 kDa protein	DNA polymerase	pTP	L1 43-kDa protein	L2 penton	L3 hexon	pVIII	L5 fiber	E4 34 kDa protein
HAdV-11 B2	96	99	99	99	99	98	92	100	93	99
HAdV-34 B2	97	99	99	99	100	95	94	100	63	98
HAdV-35 B2	95	98	99	99	100	98	92	100	63	98
HAdV-3 B1	79	88	90	94	95	85	86	94	57	98
HAdV-7 B1	79	88	90	94	95	85	86	94	92	97
HAdV-12 A	44	42	70	77	72	72	77	77	20	49
HAdV-5 C	37	48	74	81	71	69	85	80	20	57
HAdV-9 D	42	54	74	78	78	77	82	80	30	65
HAdV-4 E	56	59	85	90	84	83	83	89	27	70
HAdV-40 F	38	46	70	73	76	72	78	79	24	45
HAdV-52 G	36	44	73	77	74	72	79	78	27	48

et al., 2005). Recombination may be an important component of HAdV evolution and pathogenicity. An earlier report on HAdV-4, for example, suggested recombination events of the fiber gene as a mechanism in its evolution (Gruber et al., 1993); subsequent genome determination and bioinformatics re-affirmed this recombination at higher resolution and revealed a zoonosis with chimpanzee as a possible origin of the species (Purkayastha et al., 2005a). This has been re-confirmed recently again by sequencing of the fiber genes (Madisch et al., 2005).

Data from the zPicture analysis are presented in Fig. 1 at the global whole genome level. These are all relative to HAdV-14 and the full analysis included the range of all HAdV species, with only the selection of species B presented (B2 HAdV-11, -34, and -35; B1 HAdV-3, -7, and -21), as the other species do not add to the discussion. Three B1 members are presented to show a range of differences, i.e., HAdV-21 was included as an additional reference for the B1 subspecies and differs slightly from the HAdV-3 and -7 alignments to HAdV-14. Percent identities were calculated by global alignments using the EMBOSS package ("needle" program).

3.3.3.1. E1A protein. This transcription factor is an essential protein and should therefore be highly conserved across the range of HAdVs. From the amino acid analysis (Table 3) and phylogenetic analysis (data not shown), the E1A protein follows the current accepted classification scheme of the HAdV genomes (Fig. 2). The whole genome alignments displayed in Fig. 1 reveal differences, expanded in the Supplemental Figure, that may reflect origins or regions of absolute conservation. This analysis can be localized to the E1A sub-regions with each expanded for better resolution using zPicture, MSA and phylogeny analyses. The in depth zPicture alignment is presented as Supplemental Fig. 1. The E1A may be broken into three local blocks for further analysis, especially with the B1 counterparts: (1–443), (444–589) and (590–786 (end)). Although not as dramatic as will be discussed for the hexon and fiber pro-

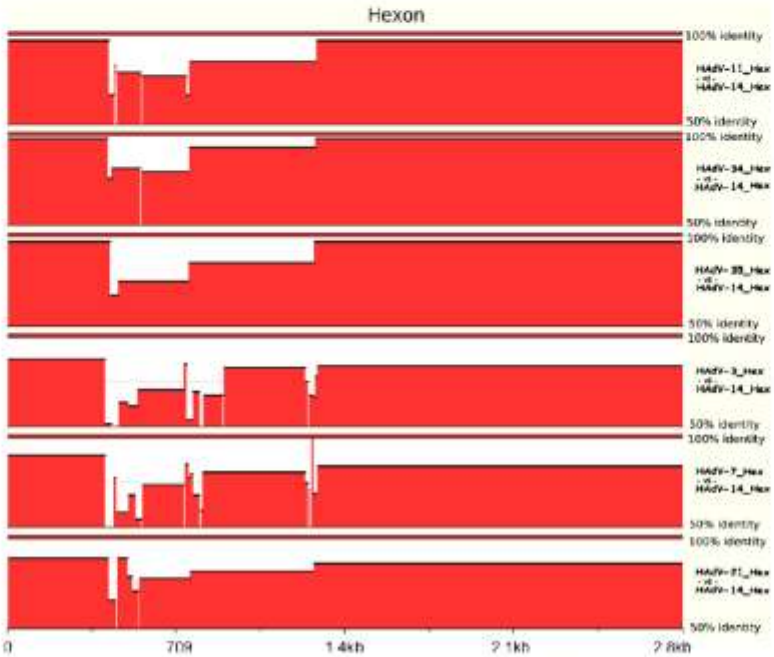


Fig. 4. Hexon protein analysis using zPicture. Hexon sequences were aligned in pairs relative to HAdV-14. Breakpoints of divergence can be estimated and analyzed individually.

Table 4
Fiber proteins from B1 and B2. Gene lengths and copy number varies across the HAdV species.

Serotype	Species	# of fibers	Length of fiber 1 protein	Fiber 1% identity	Fiber 1 gene length	Length of fiber 2 protein	Fiber 2% identity	Fiber 2 gene length
HAdV-14	B2	1	325	100	978	–	–	–
HAdV-11	B2	1	325	93	978	–	–	–
HAdV-34	B2	1	323	63	972	–	–	–
HAdV-35	B2	1	323	63	972	–	–	–
HAdV-3	B1	1	319	57	960	–	–	–
HAdV-7	B1	1	325	92	978	–	–	–
HAdV-12	A	1	587	20	1764	–	–	–
HAdV-5	C	1	581	20	1746	–	–	–
HAdV-9	D	1	362	30	1089	–	–	–
HAdV-4	E	1	425	27	1278	–	–	–
HAdV-40	F	2	387	24	1164	547	18	1644
HAdV-52	G	2	363	27	1092	560	19	1683

teins, B2 E1A proteins show bipartite identity for HAdV-11 and -34 relative to HAdV-14: there is a region from 1 to 443 that is 96–98%, followed by a region of 97–99% identity from 444 to end. In contrast, HAdV-35 shows a high and unbroken level of identity across the entire gene.

For the B1s (HAdV-3, -7 and -21, respectively), it is a tripartite pattern with three blocks of similarity from 1 to 443, 444 to 589 and 590 to end: 71%, 86% and 82% identities, respectively. There is also a small gap, for all three B1s, between the second and third blocks that has less than 50% identity with HAdV-14. These patterns support a B1/B2 division.

3.3.3.2. Penton. A less striking but still informative picture is seen for the penton coding region (Fig. 1). The B2s, HAdV-11 and -35, have a identity of ~98% constant across the gene; -34 is lower at 93%. Pentons from the B1s have a much lower, and nearly constant, identity (~90%) across the gene, broken by a gap of divergent sequences, less than 50%. HAdV-21 is similar to the -3 and -7 sequences overall, without this gap.

3.3.3.3. Hexon. As shown in Fig. 4, the hexon region displays a "roughly" tripartite pattern of homology. These may reflect recombination events, with the more heterogeneous regions being more distant events in the past. The "tripartite" structure is defined as two regions of apparent constant levels of identity with one region of heterogeneous similarities: 1–441; 442–1320; and 1321–2836 (end). For more a more detailed analysis, breaking this region into four blocks for further analysis (that is, the heterogeneous region

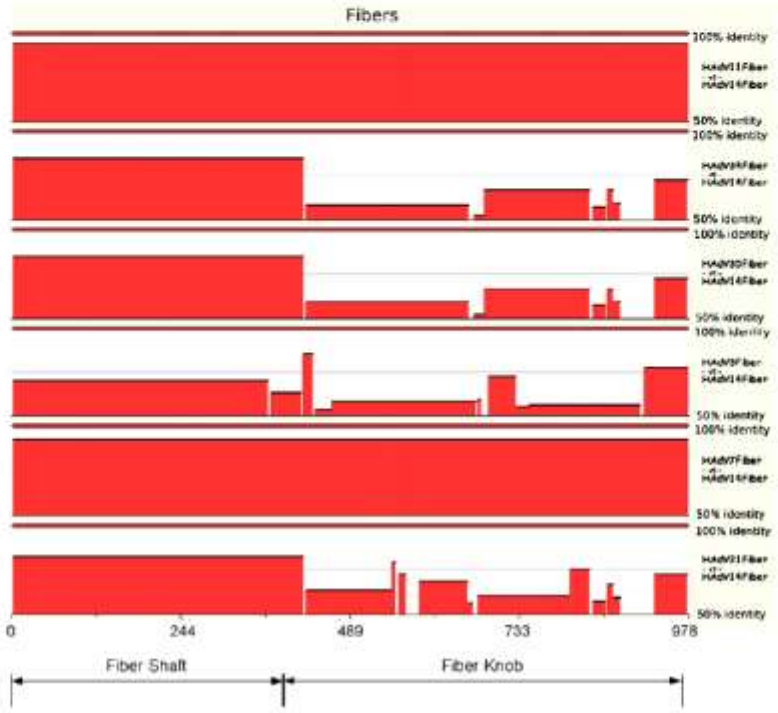


Fig. 5. Fiber protein analysis using zPicture. Fiber sequences were aligned in pairs relative to HAdV-14. Fiber can be divided into two domains: shaft and knob. The knob which may determine cell tropism varies in sequence.

into two parts) using MSA and phylogeny tools, the B2s have in common pattern, with similar levels of homology in similar locations: 1–441 at 95%; 554–803 at 64–71%; 863–1320 at 81–85%; and 1327–2836 (end) 92–97%.

For the members of the B1s shown, the percent identities are similar but lower, but again have similar physical locations: 83–87%; 51–75% (high number for HAdV-21); 70–78%; and 82–86%, respectively. The phylogeny analysis shows the B1 HAdV-21 "block 2" moving into the sub-branch with the B2s, again perhaps reflecting similar origins or recombination events (Supplemental Fig. 2a and b).

3.3.3.4. Fiber. Table 4 displays the comparisons of the fiber protein and gene from the B2s and a selection of genomes representing the species of HAdVs. One fiber is found in the majority of the species, with the exceptions being two for F and G. On the whole, HAdV-14 and -11 have fibers that are 93% identical, differing substantially from -34 and -35, at 63%. HAdV-3 is 57% identical, while HAdV-7 is 92%. The lengths of the B species fibers are comparable to each other and different from the other species, at 319–325 nucleotides. The fibers from the other species have identities in the 20–30% range when compared with HAdV-14, with lengths ranging from 362 to 587 nucleotides. The second fiber protein of F and G are at 18–19% identity compared with HAdV-14, with lengths of 387 and 560, respectively.

The fiber protein may be segregated into two functional domains: shaft and knob. This partition occurs putatively at position 384 of the gene. Fig. 5 shows the zPicture analysis of these fiber regions, with two domains defined. The shaft portion is less heterogeneous than the knob portion, in terms of homology with HAdV-14.

Analyses with zPicture, Bootscanning and SimPlot genome alignments present a striking view of the fiber gene region. Fig. 5 shows that, in relations to HAdV-14, HAdV-11 has a fiber region that is between 94% and 98% identity over the entire region; HAdV-7 has a similar profile, at a lower level of identity at 92–95%. In contrast, HAdV-34 and -35 have a bipartite identity, with the proximal portion (shaft) at 85% (1–423) and a distal portion (knob) that is heterogeneous: three main blocks at between 50% and 65% and several other smaller less similar regions. HAdV-3 and -21 are similar, with the proximal portion at 68% and 83%, respectively, and the distal three main blocks at between 50% and 63%, again with other smaller less similar regions. The phylogeny analysis reflects these data, with HAdV-7 sub-branching amongst HAdV-14 and -11 in both the whole fiber alignment as well as the alignments of each of the four main blocks (data not shown). The heterogeneity may reflect recombination events.

3.3.3.5. VA-RNA coding sequences. As noted earlier, there was a second non-"hypothetical protein" coding difference between the B2 and the B1s subspecies. This coding difference is in the virus-associated RNA or VA-RNA coding sequences, noted as 'genes' here. As shown in Table 5, HAdV-14 and other B2s contain one VA-RNA gene whereas B1s contain two. Genome sequences were re-examined to account for the missing sequences: none was found. From the genomes deposited in GenBank, species A, F and G all contain one VA-RNA gene whereas species B1, C, D and E contain two VA-RNA genes. VA-RNA from the other three B2s show percent identities of ~98% each relative to the one from HAdV-14; their sizes are similar at ~162 nucleotides. The equivalent VA-RNA I from the B1 members HAdV-3 and -7 show a percent identity of ~80% and are longer at ~168, with the VA-RNA II at 177 and 175 nucleotides, and 60% and 49% identity, relative to HAdV-14 VA-RNA I.

These are in contrast with the less similar ones from species A (141; 51%), C (160; 55%), D (163; 77%), E (159; 78%), F (171; 53%) and G 164; 56%). The VA-RNA II relative to VA-RNA I from HAdV-14 are C (158; 63%), D (154; 58%) and E (101; 36%).

VA-RNAs play an important role in the pathogenicity of HAdVs as they inhibit the host anti-viral functions, e.g., interferons (Mathews and Shenk, 1991). The differences in identity, number and lengths of the VA-RNA coding sequences between the B2, B1, E and C adenoviruses may have important bearing in the current discussions of tissue and cell tropism, epidemiology, pathoepidemiology and HAdV vaccine development directed at specific respiratory pathogens, as well as in the development of adenoviruses as vaccine and gene therapy vectors.

3.3.3.6. Short summary of the coding sequences analysis. The detailed analysis of the span of proteins across the HAdV-14 genome demonstrates a shared origin for the four members of the B2 subspecies and a close relationship with the B1 subspecies, setting these apart from other species. These include the coat proteins that are expected to differ due to potential interactions with the host but balanced by the need for conservation due to their potential interactions with receptors for entry into cells. The percent identity data are intriguing, especially in the context of the whole genome alignment, particularly for the hexon and fiber proteins, with their multi-partite identity/dissimilarity structures.

3.4. Bioinformatics support of the B1, B2 subspecies designation

Whole genome data and comparisons, along with comparisons of the individual genes and proteins, were used to support and to strengthen the original subdivision of the B species into subspecies. That is, data and analyses presented in this report reflect and confirm, in part, the use of restriction enzyme digest patterns in defining the subdivision, which may now be more defining given the virtual restriction enzyme analyses possible.

In addition, the availability of compete genome sequences allowed the *in silico* examination and comparison of the proteome, again supporting the original designation of two subspecies. Table 2 provides a global accounting of all of the putative coding regions and, as discussed, highlight differences between the B1 and B2 members as subspecies. Table 3 provides further support to this in the form of protein homologies. These computational results complement and enhance clinical and molecular biological observations, lending an additional and independent view.

3.5. Human disease and current outbreaks

As noted, several HAdVs cause respiratory diseases, with certain specific and commonly encountered species and serotypes implicated and reported. Occasionally, less commonly associated and even unexpected strains emerge or re-emerge in the global population to cause respiratory disease. There does not seem to be a division in the diseases caused by members of the B1 versus the B2. In the U.S., during 2006–2007, HAdV-14 was the unexpected pathogenic agent identified in outbreaks of severe and sometimes fatal acute respiratory disease (ARD) in both civilian and military populations (Louie et al., 2008; Metzgar et al., 2007). As reported in the Morbidity and Mortality Weekly Report (November 2007) and by two groups, the isolates were distinct from the original 1955 prototype strain (Louie et al., 2008; Metzgar et al., 2007). As noted earlier, the prototype HAdV-14 was isolated from a military population as an agent causing ARD (Van Der Veen and Kok, 1957) and was also identified in an outbreak in a civilian population in the same year but elsewhere (Kendall et al., 1957). The genomics and bioinformatics data pertaining to this prototype strain are presented here.

Table 5
VA-RNA coding sequences are present as one or two copies in the adenovirus genomes.

Serotype	Species	# of VA-RNAs	Length of VA-RNA I	VA-RNA I % identity	Length of VA-RNA II	VA-RNA II % identity
HAdV-14	B2	1	162	100	–	–
HAdV-11	B2	1	158	96	–	–
HAdV-34	B2	1	162	99	–	–
HAdV-35	B2	1	161	98	–	–
HAdV-3	B1	2	175	81	177	60
HAdV-7	B1	2	161	79	175	49
HAdV-12	A	1	141	51	–	–
HAdV-5	C	2	160	55	158	63
HAdV-9	D	2	163	77	154	58
HAdV-4	E	2	159	78	101	36
HAdV-40	F	1	171	53	–	–
HAdV-52	G	1	164	56	–	–

3.6. Re-emerging respiratory disease agent and vaccine development

In general, the adenoviruses that are associated with and reported for respiratory diseases have been mainly confined to the B1s (HAdV-3 and -7; with HAdV-16 and -21 to a lesser extent) and HAdV-4 (E) (Blasiole et al., 2004; Kajon et al., 2007; Lin et al., 2006; Metzgar et al., 2005; Vora et al., 2006). Sporadic reports of 'non-specific' B2 (Metzgar et al., 2005; Vora et al., 2006), and specific HAdV-11 (Chmielewicz et al., 2005), associated with patients suggest a role of some B2 serotypes in respiratory diseases. HAdV-14 is a recent re-emerging pathogen, reported in epidemic outbreaks in the U.S., in both military and civilian settings (Louie et al., 2008; Metzgar et al., 2007). With this re-emergence as a pathogenic respiratory agent, thoughts of a specific vaccine are appropriate, especially in the context of the pending deployment of vaccines against HAdV-4 and -7 (Lyons et al., 2008). Despite the observation that the use of the vaccines against HAdV-4 and -7 apparently gave immunity against other serotypes, such as HAdV-3 (Binn et al., 2007), insight into the differences (and similarities) of the genome and proteome of HAdV-14, particularly the coat proteins may lend weight to vaccine development, especially in the context of the development and validation time.

Given the similarities in genome data and especially in protein identities between these HAdV-11 and -14, particularly in the penton (98%), hexon (92%) and fiber (93%) proteins, and given the re-emergence of HAdV-11 as a respiratory pathogen, it may be possible to engineer a vaccine that is useful against both serotypes should they persist as global re-emergent pathogens.

3.7. Recombinant protein-bearing vectors

Recombinant HAdVs are used as recombinant protein-bearing vectors for gene therapy and vaccine delivery (Berg et al., 2005; Graham and Prevec, 1992). The application of HAdV-2 and -5-based vectors may be limited due to widespread early childhood infections and pre-existing immunity. Other species of HAdVs may be better candidates. Vectors derived from species B serotypes are being examined (Stone et al., 2006). For example, a subspecies B1 HAdV-3-based vector has been constructed and evaluated as a first generation gene transfer vector, in part, because of its high virulence and use of different receptor to enter cells (Sirena et al., 2005). Alternatively, a B2 serotype has been considered for use as vectors as well; HAdV-11 has been sequenced by two groups independently (Mei et al., 2003; Stone et al., 2003) and examined for gene transfer vector applications (Stone et al., 2005). Another B2, HAdV-35, has been sequenced as well and also studied in this context, again by two groups independently (Gao et al., 2003; Shott et al., 2008; Vogels et al., 2003). A recombinant HAdV-35 vector for gene therapy has been used in brain cancer trials, presumably "bypassing pre-existing anti-vector immunity in cancer patients" (Brouwer et al., 2007). Each has unique merits but also detractions, for example, a report in the literature described a protective immune response in two mouse strains against a recombinant HAdV-35-derived tuberculosis vaccine (Radosevic et al., 2007).

A summary of the differences between human adenoviruses in 1984 mentions that "with the exception of [HAdV-14], members [of the subspecies B2] are characterized by a predilection for shedding via the urine. This may be protracted, and can occur with or without symptoms" (Wadell, 1984). Noted also in this report is that subspecies B2 members were rarely isolated then, in general. Perhaps this still holds true today and HAdV-14 is still uncommon amongst the uncommon B2 within the population of HAdVs in humans. If this is the case, then due to its rarity in the general population, HAdV-14 may be an excellent candidate as a gene transfer and/or vaccine vector.

4. Summary

The prototype HAdV-14p genome, genome annotation and comparative genomics data are presented as a genomics and bioinformatics study to serve as a reference and a resource for both respiratory disease investigations and possible vaccine development to deal with a re-emerging pathogen, as well as for applications in vector development. Complete genome and proteome data provide finer granularity into the pathoepidemiology of the HAdVs, for example, the analysis of the putative cell entry proteins, penton and fiber genes, gives insight into the tropism of HAdV-14, accounting for its role as a respiratory pathogen. These data provide a comprehensive analysis and the bioinformatics complementation and support for dividing the B species into two subspecies, B1 and B2.

Acknowledgments

During the collection of the genome data and preliminary data analyses, AP (2003–2005), JMC (2004), CT (2001–2005) and DS (2002–2004) were affiliated with the HQ USAF Surgeon General Office, Directorate of Modernization (SGR) and the Epidemic Outbreak Surveillance (EOS) Consortium, Falls Church, VA. Portions of this work were funded, during the time periods noted, by a grant (DAMD17-03-2-0089) from the U.S. Army Medical Research and Material Command (USAMRMC) and additional support was provided by the EOS Project, funded through HQ USAF Surgeon General Office, Directorate of Modernization (SGR) and the Defense Threat Reduction Agency. Therefore, 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.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.virusres.2009.03.011.

References

Benkő, M.H.B., Both, G.W., Russell, W.C., Adair, B.M., Ádám, É., de Jong, J.C., Hess, M., Johnson, M., Kajon, A., Kidd, A.H.D.L.H.Q.-G.L., Mautner, V., Pring-Akerblom, P., Wadell, G., 2005. In: Fauquet, C.M., Mayo, M.A., Maniloff, J., Desselberger, U., Ball, L.A. (Eds.), Family Adenoviridae in the Eighth Report of the International Committee on Taxonomy of Viruses, vol. 8. Academic Press, San Diego, CA, pp. 213–228.

Berg, M., Difatta, J., Hoiczyk, E., Schlegel, R., Ketner, G., 2005. Viable adenovirus vaccine prototypes: high-level production of a papillomavirus capsid antigen from the major late transcriptional unit. Proc. Natl. Acad. Sci. U.S.A. 102 (12), 4590–4595.

Berriman, M., Rutherford, K., 2003. Viewing and annotating sequence data with Artemis. Brief Bioinform. 4 (2), 124–132.

Binn, L.N., Sanchez, J.L., Gaydos, J.C., 2007. Emergence of adenovirus type 14 in US military recruits—a new challenge. J. Infect. Dis. 196 (10), 1436–1437.

Blasiole, D.A., Metzgar, D., Daum, L.T., Ryan, M.A., Wu, J., Wills, C., Le, C.T., Freed, N.E., Hansen, C.J., Gray, G.C., Russell, K.L., 2004. Molecular analysis of adenovirus isolates from vaccinated and unvaccinated young adults. J. Clin. Microbiol. 42 (4), 1686–1693.

Bray, N., Pachter, L., 2004. MAVID: constrained ancestral alignment of multiple sequences. Genome Res. 14 (4), 693–699.

Brouwer, E., Havenga, M.J., Ophorst, O., de Leeuw, B., Gijsbers, L., Gillissen, G., Hoeben, R.C., ter Horst, M., Nanda, D., Dirven, C., Avezaat, C.J., Goudsmit, J., Sillevis Smitt, P., 2007. Human adenovirus type 35 vector for gene therapy of brain cancer: improved transduction and bypass of pre-existing anti-vector immunity in cancer patients. Cancer Gene Ther. 14 (2), 211–219.

Cheng, C.C., Huang, L.M., Kao, C.L., Lee, P.I., Chen, J.M., Lu, C.Y., Lee, C.Y., Chang, S.Y., Chang, L.Y., 2008. Molecular and clinical characteristics of adenoviral infections in Taiwanese children in 2004–2005. Eur. J. Pediatr. 167 (6), 633–640.

Chmielewicz, B., Benzler, J., Pauli, G., Krause, G., Bergmann, F., Schweiger, B., 2005. Respiratory disease caused by a species B2 adenovirus in a military camp in Turkey. J. Med. Virol. 77 (2), 232–237.

Crawford-Miksza, L.K., Schnurr, D.P., 1996. Adenovirus serotype evolution is driven by illegitimate recombination in the hypervariable regions of the hexon protein. Virology 224 (2), 357–367.

Dan, A., Elo, P., Harrach, B., Zadori, Z., Benko, M., 2001. Four new inverted terminal repeat sequences from bovine adenoviruses reveal striking differences in the length and content of the ITRs. Virus Genes 22 (2), 175–179.

Gao, W., Robbins, P.D., Gambotto, A., 2003. Human adenovirus type 35: nucleotide sequence and vector development. Gene Ther. 10 (23), 1941–1949.

Graham, F.L., Prevec, L., 1992. Adenovirus-based expression vectors and recombinant vaccines. Biotechnology 20, 363–390.

Gray, G.C., McCarthy, T., Lebeck, M.G., Schnurr, D.P., Russell, K.L., Kajon, A.E., Landry, M.L., Leland, D.S., Storch, G.A., Ginocchio, C.C., Robinson, C.C., Demmler, G.J., Saubolle, M.A., Kehl, S.C., Selvarangan, R., Miller, M.B., Chappell, J.D., Zerr, D.M., Kiska, D.L., Halstead, D.C., Capuano, A.W., Setterquist, S.F., Chorazy, M.L., Dawson, J.D., Erdman, D.D., 2007. Genotype prevalence and risk factors for severe clinical adenovirus infection, United States 2004–2006. Clin. Infect. Dis. 45 (9), 1120–1131.

Green, M., Mackey, J.K., Wold, W.S., Rigden, P., 1979. Thirty-one human adenovirus serotypes (Ad1–Ad31) form five groups (A–E) based upon DNA genome homologies. Virology 93 (2), 481–492.

Gruber, W.C., Russell, D.J., Tibbetts, C., 1993. Fiber gene and genomic origin of human adenovirus type 4. Virology 196 (2), 603–631.

Hatfield, L., Hearing, P., 1993. The NFIII/OCT-1 binding site stimulates adenovirus DNA replication in vivo and is functionally redundant with adjacent sequences. J. Virol. 67 (7), 3931–3939.

Jones 2nd, M.S., Harrach, B., Ganac, R.D., Gozum, M.M., Dela Cruz, W.P., Riedel, B., Pan, C., Delwart, E.L., Schnurr, D.P., 2007. New adenovirus species found in a patient presenting with gastroenteritis. J. Virol. 81 (11), 5978–5984.

Kajon, A.E., Moseley, J.M., Metzgar, D., Huong, H.S., Wadleigh, A., Ryan, M.A., Russell, K.L., 2007. Molecular epidemiology of adenovirus type 4 infections in US military recruits in the postvaccination era (1997–2003). J. Infect. Dis. 196 (1), 67–75.

Kajon, A.E., Xu, W., Erdman, D.D., 2005. Sequence polymorphism in the E3 7.7K ORF of subspecies B1 human adenoviruses. Virus Res. 107 (1), 11–19.

Kendall, E.J., Riddle, R.W., Tuck, H.A., Rodan, K.S., Andrews, B.E., McDonald, J.C., 1957. Pharyngo-conjunctival fever; school outbreaks in England during the summer of 1955 associated with adenovirus types 3,7 and 14. Br. Med. J. 2 (5037), 131–136.

Larranaga, C., Kajon, A., Villagra, E., Avendano, L.F., 2000. Adenovirus surveillance on children hospitalized for acute lower respiratory infections in Chile (1988–1996). J. Med. Virol. 60 (3), 342–346.

Lauer, K.P., Llorente, I., Blair, E., Seto, J., Krasnov, V., Purkayastha, A., Ditty, S.E., Hadfield, T.L., Buck, C., Tibbetts, C., Seto, D., 2004. Natural variation among human adenoviruses: genome sequence and annotation of human adenovirus serotype 1. J. Gen. Virol. 85 (Pt 9), 2615–2625.

Lavigne, R., Seto, D., Mahadevan, P., Ackermann, H.W., Kropinski, A.M., 2008. Unifying classical and molecular taxonomic classification: analysis of the Podoviridae using BLASTP-based tools. Res. Microbiol. 159 (5), 406–414.

Lin, B., Wang, Z., Vora, G.J., Thornton, J.A., Schnur, J.M., Thach, D.C., Blaney, K.M., Ligler, A.G., Malanoski, A.P., Santiago, J., Walter, E.A., Agan, B.K., Metzgar, D., Seto, D., Daum, L.T., Kruzelock, R., Rowley, R.K., Hanson, E.H., Tibbetts, C., Stenger, D.A., 2006. Broad-spectrum respiratory tract pathogen identification using resequencing DNA microarrays. Genome Res. 16 (4), 527–535.

Louie, J.K., Kajon, A.E., Holodniy, M., Guardia-LaBar, L., Lee, B., Petru, A.M., Hacker, J.K., Schnurr, D.P., 2008. Severe pneumonia due to adenovirus serotype 14: a new respiratory threat? Clin. Infect. Dis. 46 (3), 421–425.

Lyons, A., Longfield, J., Kuschner, R., Straight, T., Binn, L., Seriwatana, J., Reitstetter, R., Froh, I.B., Craft, D., McNabb, K., Russell, K., Metzgar, D., Liss, A., Sun, X., Towle, A., Sun, W., 2008. A double-blind, placebo-controlled study of the safety and immunogenicity of live, oral type 4 and type 7 adenovirus vaccines in adults. Vaccine 26 (23), 2890–2898.

Madisch, I., Harste, 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 (24), 15265–15276.

Mathews, M.B., Shenk, T., 1991. Adenovirus virus-associated RNA and translation control. J. Virol. 65 (11), 5657–5662.

Medini, D., Donati, C., Tettelin, H., Masignani, V., Rappuoli, R., 2005. The microbial pan-genome. Curr. Opin. Genet. Dev. 15 (6), 589–594.

Mei, Y.F., Skog, J., Lindman, K., Wadell, G., 2003. Comparative analysis of the genome organization of human adenovirus 11, a member of the human adenovirus species B, and the commonly used human adenovirus 5 vector, a member of species C. J. Gen. Virol. 84 (Pt 8), 2061–2071.

Metzgar, D., Osuna, M., Kajon, A.E., Hawksworth, A.W., Irvine, M., Russell, K.L., 2007. Abrupt emergence of diverse species B adenoviruses at US military recruit training centers. J. Infect. Dis. 196 (10), 1465–1473.

Metzgar, D., Osuna, M., Yingst, S., Rakha, M., Earhart, K., Elyan, D., Esmat, H., Saad, M.D., Kajon, A., Wu, J., Gray, G.C., Ryan, M.A., Russell, K.L., 2005. PCR analysis of Egyptian respiratory adenovirus isolates, including identification of species, serotypes, and coinfections. J. Clin. Microbiol. 43 (11), 5743–5752.

Mul, Y.M., Verrijzer, C.P., van der Vliet, P.C., 1990. Transcription factors NFI and NFIII/oct-1 function independently, employing different mechanisms to enhance adenovirus DNA replication. J. Virol. 64 (11), 5510–5518.

Nagata, K., Guggenheimer, R.A., Enomoto, T., Lichy, J.H., Hurwitz, J., 1982. Adenovirus DNA replication in vitro: identification of a host factor that stimulates synthesis of the preterminal protein-dCMP complex. Proc. Natl. Acad. Sci. U.S.A. 79 (21), 6438–6442.

Purkayastha, A., Ditty, S.E., Su, J., McGraw, J., Hadfield, T.L., Tibbetts, C., Seto, D., 2005a. 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 (4), 2559–2572.

Purkayastha, A., Su, J., Carlisle, S., Tibbetts, C., Seto, D., 2005b. 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 (1), 114–129.

Purkayastha, A., Su, J., McGraw, J., Ditty, S.E., Hadfield, T.L., Seto, J., Russell, K.L., Tibbetts, C., Seto, D., 2005c. 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 (7), 3083–3094.

Radosevic, K., Wieland, C.W., Rodriguez, A., Weverling, G.J., Mintardjo, R., Gillissen, G., Vogels, R., Skeiky, Y.A., Hone, D.M., Sadoff, J.C., van der Poll, T., Havenga, M., Goudsmit, J., 2007. Protective immune responses to a recombinant adenovirus type 35 tuberculosis vaccine in two mouse strains: CD4 and CD8 T-cell epitope mapping and role of gamma interferon. Infect. Immun. 75 (8), 4105–4115.

Rawlins, D.R., Rosenfeld, P.J., Wides, R.J., Challberg, M.D., Kelly Jr., T.J., 1984. Structure and function of the adenovirus origin of replication. Cell 37 (1), 309–319.

Robinson, C.M., Shariati, F., Gillaspy, A.F., Dyer, D.W., Chodosh, J., 2008. Genomic and bioinformatics analysis of human adenovirus type 37: new insights into corneal tropism. BMC Genomics 9, 213.

Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4 (4), 406–425.

Shott, J.P., McGrath, S.M., Pau, M.G., Custers, J.H., Ophorst, O., Demoitie, M.A., Dubois, M.C., Komisar, J., Cobb, M., Kester, K.E., Dubois, P., Cohen, J., Goudsmit, J., Heppner, D.G., Stewart, V.A., 2008. Adenovirus 5 and 35 vectors expressing Plasmodium falciparum circumsporozoite surface protein elicit potent antigen-specific cellular IFN-gamma and antibody responses in mice. Vaccine 26 (23), 2818–2823.

Sirena, D., Ruzsics, Z., Schaffner, W., Greber, U.F., Hemmi, S., 2005. The nucleotide sequence and a first generation gene transfer vector of species B human adenovirus serotype 3. Virology 343 (2), 283–298.

Stone, D., di Paolo, N.C., Lieber, A., 2006. Development of group B adenoviruses as gene transfer vectors. Biotechnol. Genet. Eng. Rev. 22, 101–123.

Stone, D., Furthmann, A., Sandig, V., Lieber, A., 2003. The complete nucleotide sequence, genome organization, and origin of human adenovirus type 11. Virology 309 (1), 152–165.

Stone, D., Ni, S., Li, Z.Y., Gaggar, A., DiPaolo, N., Feng, Q., Sandig, V., Lieber, A., 2005. Development and assessment of human adenovirus type 11 as a gene transfer vector. J. Virol. 79 (8), 5090–5104.

Temperley, S.M., Hay, R.T., 1992. Recognition of the adenovirus type 2 origin of DNA replication by the virally encoded DNA polymerase and preterminal proteins. EMBO J. 11 (2), 761–768.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25 (24), 4876–4882.

Van Der Veen, J., Kok, G., 1957. Isolation and typing of adenoviruses recovered from military recruits with acute respiratory disease in The Netherlands. Am. J. Hyg. 65 (2), 119–129.

Vogels, R., Zuijdgeest, D., van Rijnsoever, R., Hartkoorn, E., Damen, I., de Bethune, M.P., Kostense, S., Penders, G., Helmus, N., Koudstaal, W., Cecchini, M., Wetterwald, A., Sprangers, M., Lemckert, A., Ophorst, O., Koel, B., van Meerendonk, M., Quax, P., Panitti, L., Grimbergen, J., Bout, A., Goudsmit, J., Havenga, M., 2003. Replication-deficient human adenovirus type 35 vectors for gene transfer and vaccination: efficient human cell infection and bypass of preexisting adenovirus immunity. J. Virol. 77 (15), 8263–8271.

Vora, G.J., Lin, B., Gratwick, K., Meador, C., Hansen, C., Tibbetts, C., Stenger, D.A., Irvine, M., Seto, D., Purkayastha, A., Freed, N.E., Gibson, M.G., Russell, K., Metzgar, D., 2006. Co-infections of adenovirus species in previously vaccinated patients. Emerg. Infect. Dis. 12 (6), 921–930.

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

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

Zhang, Q., Su, X., Seto, D., Zheng, B.J., Tian, X., Sheng, H., Li, H., Wang, Y., Zhou, R., 2009. Construction and characterization of a replication-competent human adenovirus type 3-based vector as a live-vaccine candidate and a viral delivery vector. Vaccine 27 (8), 1145–1153.

Part 2

Computational analysis of adenovirus serotype 5 (HAdV-C5) from an HAdV coinfection shows genome stability after 45 years of circulation

ARTICLE IN PRESS

YVIRO-05744; No. of pages: 7; 4C:

Virology xxx (2010) xxx–xxx

Contents lists available at ScienceDirect

Virology

journal homepage: www.elsevier.com/locate/yviro

Computational analysis of adenovirus serotype 5 (HAdV-C5) from an HAdV coinfection shows genome stability after 45 years of circulation

Jason Seto [a,1], Michael P. Walsh [a,1], David Metzgar [b], Donald Seto [a,*]

[a] Department of Bioinformatics and Computational Biology, George Mason University, 10900 University Blvd., MSN 5B3, Manassas, VA 20110, USA
[b] Department of Respiratory Diseases Research, Naval Health Research Center, 140 Sylvester Road, San Diego, CA 92106, USA

ARTICLE INFO

Article history:
Received 12 January 2010
Returned to author for revision 1 March 2010
Accepted 13 May 2010
Available online xxxx

Keywords:
Adenovirus evolution
Coinfection
Genome stability
Viral bioinformatics
Viral genomics
Sequencing error

ABSTRACT

Adenovirus coinfections present opportunities for genome recombination. Computational analysis of an HAdV-C5 field strain genome, recovered from a patient with acute respiratory disease and coinfected with HAdV-B21, shows that there was no exchange of genomic material into HAdV-C5. Comparison of this genome to the sparsely amplified prototype demonstrates a high level of sequence conservation and stability of this genome across 45 years. Further, comparison to a version of the prototype that had been passaged in laboratory settings shows stability as well. HAdV genome stability and evolution are considerations for applications as vaccines and as vectors for gene delivery. In the annotation analysis, a single sequencing error in the HAdV-C5_ARM (Adenovirus Reference Material) genome is noted and may lead to erroneous annotation and biological interpretations.

Introduction

Human adenoviruses (HAdVs) are important pathogens (Echavarria, 2009) and are model organisms for understanding biological systems (Russell, 2009). As public health concerns, HAdVs are isolated from highly contagious respiratory and ocular infections with emergent strains characterized as "intermediate" and "intertypic" using limited molecular typing and/or serology (Engelmann et al., 2006; Ishiko et al., 2008; Zhu et al., 2009). One has been documented as more virulent than its prototype and other variants, e.g., HAdV-B7h (Kajon et al., 1996; Kajon and Wadell, 1996; Larranaga et al., 2000), and another characterized with an apparent change in cell tropism, "HAdV-B11a" (Hierholzer et al., 1974; Li et al., 1991; Zhu et al., 2009). However, none has been described in comprehensive and exact genomic details until recently: HAdV-D53 (Walsh et al., 2009) and B55 (Walsh et al., 2010; Yang et al., 2009), which are emergent "types" with characterization based on non-serological techniques; and HAdV-B16, D22, and E4, which are archived "serotypes", originally defined with serology and recently recharacterized using genomics (Robinson et al., 2009; Seto, unpublished results).

HAdV-C5 is a pathogen that causes respiratory symptoms of variable severity including acute, mild, and none, i.e., asymptomatic (Echavarria, 2009; Edwards et al., 1985; Fox et al., 1969; Garnett et al., 2009). Species C members may be present in individuals as latent and persistent infections, in mucosal lymphocytes that could be reactivated if these cells are stimulated (Garnett et al., 2009). Field strain variants are found in outbreaks globally (Gray et al., 2007; Kajon et al., 1993, 1996, 1999; Larranaga et al., 2000; Metzgar et al., 2005).

HAdV genomes may be unstable, subjected to changes observed as antigenic shifts and resulting from recombination, and subjected to nucleotide variations, defined as genetic drift (Crawford-Miksza and Schnurr, 1996b) and resulting from base substitutions, insertions, and deletions (indels). Reports in the literature of HAdV recombination support this (Boursnell and Mautner, 1981a,b; Lukashev et al., 2008; Mautner and Mackay, 1984; Williams et al., 1975). These observations led, in part, to a hypothesis that recombination drives serotype evolution (Crawford-Miksza and Schnurr, 1996b). A recent study analyzing sixteen species C field strains suggests that recombination is frequent (Lukashev et al., 2008). Genome instability, particularly recombination, is a major potential problem for the long-term effectiveness of HAdV vaccines (Crawford-Miksza et al., 1999). It is also a concern in using HAdV genomes as gene delivery vectors. Two issues are critical: (1) frequency of these recombination and other genome changes and (2) conditions allowing recombination.

This report presents genomic and computational evidences of a stable genome from an HAdV-C5 field strain, complementing work done with similarly stable genomes such as HAdV-B3 (Mahadevan et al., 2010) and HAdV-B7, but in contrast to earlier observations of HAdV-E4 variants (Crawford-Miksza et al., 1999). These studies suggest HAdV genomes may be very stable, contrary to other reports, e.g., of frequent recombination events (Lukashev et al., 2008). Three

* Corresponding author. Fax: +1 703 993 8401.
E-mail addresses: jseto@gmu.edu (J. Seto), mwalsh5@gmu.edu (M.P. Walsh), david.metzgar@med.navy.mil (D. Metzgar), dseto@gmu.edu (D. Seto).
[1] These authors contributed equally to this paper.

0042-6822/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.virol.2010.05.010

Please cite this article as: Seto, J., et al., Computational analysis of adenovirus serotype 5 (HAdV-C5) from an HAdV coinfection shows genome stability after 45 years of circulation, Virology (2010), doi:10.1016/j.virol.2010.05.010

views of the stability of an HAdV-C5 genome are presented here: (1) circulation in the population over 45 years; (2) laboratory amplification over 39 years; and (3) coinfection with a potential recombination partner. All of these observations on genome stability (and instability) contribute to the overall and complete understanding of HAdV evolution, phylogeny, and pathoepidemiology, at the primary nucleotide sequence level, with implications to vaccine and gene delivery vector development.

Results and discussion

HAdV-C5 was among the first HAdVs to be isolated in 1953 (Rowe et al., 1953) and one of the first to be sequenced in 1991, albeit as a "composite" sequence of laboratory-circulating strain(s), with contributions from several researchers using different methodologies; the final 7558 nucleotides completed the contig (Chroboczek et al., 1992). In reflection of its importance as a gene delivery vector, as well as for other applications, a designation as the "Adenovirus Reference Material" (ARM) has been established (Sugarman et al., 2003), with the original stock amplified and made available by the American Type Culture Collection (ATCC; Manassas, VA). This "nearly pristine" genome, representing the original isolate as a "minimally-amplified" specimen, has been sequenced (Sugarman et al., 2003). A unique opportunity existed, to compare this prototype with two versions of its genome: an archival versus a laboratory-circulated, 39 years later (1953 to ca. 1992). Both are compared to a circulating field strain isolated 45 years later (1998).

Nucleotide sequence analysis

The genomes are nearly identical: Field strain "HAdV-C5_FS" has ~100% nucleotide identity to the reference genome "HAdV-5_ARM" and 99.9% with the circulating laboratory composite "HAdV-C5p". This suggests laboratory amplifications allow the accumulation of mutations, absent the selection pressures found in a host infection. The genome identities to other species C members are higher (ca. 95%), as expected, than to representatives of other HAdV species (A, B, D, E, F, and G; 46%–57%). The lengths are nearly identical: HAdV-C5_ARM at 35,934 nucleotides; HAdV-C5p at 35,938; and HAdV-C5_FS at 35,931, with GC contents of 55.2%.

Using HAdV-C5_ARM as the reference, the two prototype versions differ from each other by thirty nucleotides, and the HAdV-C5_FS (1998) differs by four nucleotides (Fig. 1). From a genomics "point of view", three of these four differences are insertion/deletions ("indels") occurring in poly-T and poly-A tracts, which may be problematic, i.e., sequencing artifacts, in Sanger chemistry protocols and which, as errors, would mean an even much higher degree of conservation between the two genomes. On the other hand, if not sequencing artifacts, they may be understood biochemically as adenovirus DNA polymerase that is prone to replication infidelity in regions of homopolymers due to polymerase slippage (Crawford-Miksza and Schnurr, 1996b; Lindenbaum et al., 1986); polymerase slippage is hypothesized as a mechanism for HAdV evolution in general (Crawford-Miksza and Schnurr, 1996b) and in the E3 region specifically (Kajon et al., 2005). It should be noted that similar naturally occurring polymorphisms are noted for poly-A and poly-T tracts, which supports these observed SNPs as "non-sequencing artifacts" and as part of "naturally occurring" hypervariable regions (Houng et al., 2009). These signatures are used for "tracking" the circulation, transmission and evolution of pathogens within and from outbreak to outbreak (Houng et al., 2009).

On the whole, however, these data suggest that this HAdV-C5_FS genome is highly and remarkably conserved after 45 years of passage through the population and after 39 years of laboratory culturing (Fig. 2). Relative to the HAdV-5C_ARM genome, with the exception of the hexon, the nucleotide changes are few and are located outside of the coding regions of the capsid coat proteins. The majority of the hexon differences are in the proximal region (at nucleotide positions 642, 672, 816, 817, 2322, and 2789 of the hexon), perhaps defining a critical "hot spot" that contains the seven hypervariable loops (at nucleotide positions 409–543, 557–579, 631–654, 739–780, 799–846, 910–945, and 1261–1347 of the hexon), which are targets for serum neutralization. This is also the region found to be involved in documented recombination (Walsh et al., 2009; Walsh et al., 2010; Yang et al., 2009; Seto, unpublished results).

This view of genome stability may not be errant; comparisons of the genomes of both the prototype and vaccine strains of HAdV-B7 and HAdV-E4 to currently circulating field strains documented similar limited genome changes; these strains are separated by a similar time span (Purkayastha et al., 2005c; Seto, unpublished results). If this holds true for most HAdV strains, i.e., limited mutations and infrequent recombination, it would explain the long-term success of the HAdV-B7 and E4 vaccines that were highly effective and support the reimplementation of these same vaccines (Crawford-Miksza et al., 1999; Gaydos and Gaydos, 1995; Gray et al., 2000; Lyons et al., 2008; Top et al., 1971).

Proteome differences and sequence error

Upon *in silico* analyses, the three HAdV-C5 proteomes are nearly identical. Results included data from annotating the full and equivalent complement of coding sequences among the three HAdV-C5 genomes as well as from the rest of the species C members, revealing the same set of coding sequences. Percent ID comparisons,

```
HAdV-C5_FS      CCCCAGGCTGGGGAGAATGTTTTAAAA-AAAAAAAAGCATGATGCAAAAT 14099
HAdV-C5_ARM     CCCCAGGCTGGGGAGAATGTTTTAAAAAAAAAAAAAAGCATGATGCAAAAT 14100
                **************************** ********************

HAdV-C5_FS      CGCCAGAGGAGCTGTTGAGCCGCCGCGCGCCCGCTTTCCAAGATGGCTAC 18848
HAdV-C5_ARM     CGCCAGAGGAGCTGCTGAGCCGCCGCGCGCCCGCTTTCCAAGATGGCTAC 18849
                ************** ************************************

HAdV-C5_FS      TTATATTACTGACCCTTGTTGCGC-TTTTTTGTGCGTGCTCCACATTGGC 29841
HAdV-C5_ARM     TTATATTACTGACCCTTGTTGCGCTTTTTTTGTGCGTGCTCCACATTGGC 29843
                ************************ ************************

HAdV-C5_FS      TTTT-TTATTCCAAAAGATTATCCAAAACCTCAAAATGAAGATCTATTAA 34390
HAdV-C5_ARM     TTTTTTTATTCCAAAAGATTATCCAAAACCTCAAAATGAAGATCTATTAA 34393
                **** **************************************************
```

Fig. 1. Alignment of four regions containing nucleotide differences between HAdV-C5_FS and HAdV-C5_ARM. Each pair of rows is a different region of the aligned genomes, with the locations noted and showing the genome mismatches. The top two panels show deletions that do not fall within coding sequences. The third, a base substitution, falls in the E3 coding region and the bottom shows a deletion that occurs after the stop codon of E4 ORF. All three deletions are found in HAdV-C5_FS.

Please cite this article as: Seto, J., et al., Computational analysis of adenovirus serotype 5 (HAdV-C5) from an HAdV coinfection shows genome stability after 45 years of circulation, Virology (2010), doi:10.1016/j.virol.2010.05.010

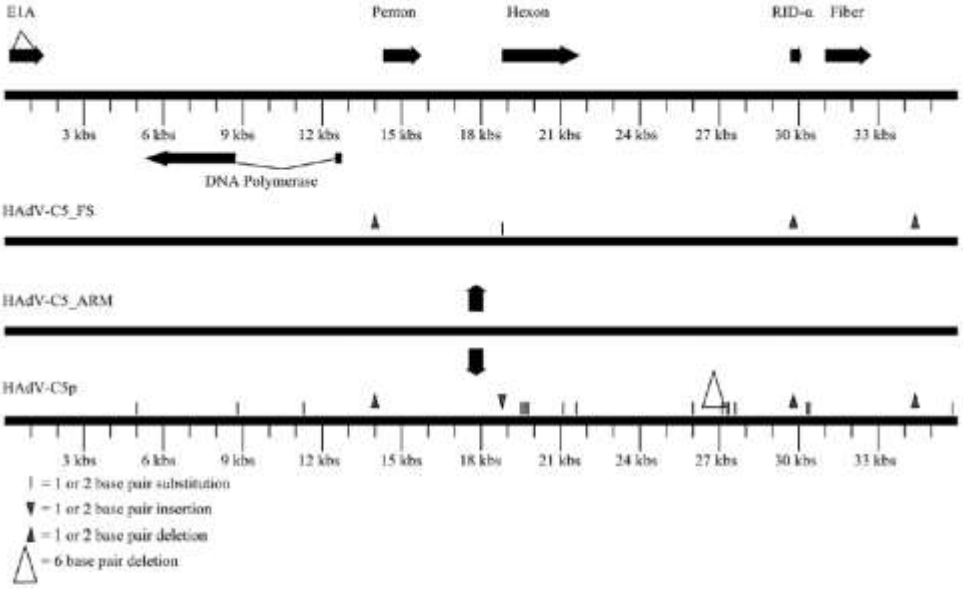


Fig. 2. Genome difference schematic map. For reference, selected genes are noted, including the three surface coat proteins: penton, hexon, and fiber. Differences in the genomes are noted by triangles (deletions), wedges (insertions), and vertical lines (base substitutions). The center genome is and represents the HAdV-C5_ARM sequence, as the reference for comparisons, with HAdV-C5p below and HAdV-C5_FS above.

with HAdV-C5_FS as the reference, of the coding sequences revealed identities of ca. 99–100% with the two HAdV-C5 proteomes (data not shown) with lower levels of identity with other species C members (ca., 70–100%). Two exceptions are the hexon, with apparently inconsequential nucleotide changes, and the E3 RID-α protein, with an apparent SNP leading to a significant frameshift.

HAdV-C5_FS and HAdV-C5_ARM hexons, which are identical, share two differences with HAdV-C5p, the laboratory strain. These are T273A and R930Q, with the T273A mutation of potential interest. Amino acid 273 is within the fifth hypervariable region (HVR5) that spans amino acids 267–282 and is a potential epitope for antisera. The other changes do not lead to amino acid alterations. The hexon from the laboratory-circulating strain contains a larger cluster of mutations relative to the other parts of the genome; this phenomenon is not manifested in the field strain, perhaps reflecting the role selection plays.

The field strain E3-encoded RID-α protein (aka "E3 10.4 kDa protein"; Wold et al., 1999) showed 100% identity with the laboratory-circulating genome, and 78% and 79% with HAdV-C1 and C2. However, HAdV-C5_FS is only 29% identical to its counterpart in the ARM strain. Sequence analysis revealed a nucleotide insertion (T) that causes a frameshift mutation, resulting in a truncation. This may be significant, particularly as embedded in the original "pristine" prototype. However, given the importance of the E3 proteins for virus evasion from the host immune response (Horwitz, 2004; Wold et al., 1999), this is likely a sequencing error rather than a natural variation. Using Artemis to view the three alignments, correcting for the single nucleotide change corrects the frame (data not shown). However, if this is indeed an SNP in HAdV-C5_ARM, then the recently circulating HAdV-C5_FS has a compensating mutation to put the coding region back into frame, identical to the laboratory-circulating strain, which in itself would be interesting, as an example of adaptive and convergent evolution.

This is a striking illustration of how sequencing errors may affect the interpretation of genomes and biology and reinforces the importance of correct and thoughtful annotation, especially as a criterion for sequencing quality control and assurance, rather than the deposit of a genome sequence. With genome sequencing technology available in a high-throughput and cost-effective format, data analysis experience is critical in providing correct and relevant contributions from the entire genome data, even as seemingly inconsequential as a single base.

Phylogenomics

Phylogenetic analysis by bootstrap-confirmed neighbor-joining trees of the HAdV genomes was performed using Molecular Evolutionary Genetics Analysis (MEGA) 4.0.2 (Tamura et al., 2007). An alignment showing the three HAdV-C5 genomes as they form a subclade with other members of species C is displayed as the relevant portion of a larger comprehensive whole genome phylogeny analysis tree (Fig. 3). The branch lengths of the species C subclade are very short and are also reflected in phylogenetic analyses and trees of the hexon, penton, and fiber, allowing a detailed examination of HAdV evolution (data not shown). For comparison, each HAdV species forms its own clade, reconfirming original characterizations and partitions based on biology and serology studies, but with different branching lengths; members of subspecies B1 and B2 form separate subclades as well, indicating the resolution of the branching distances. These results are consistent with and support the findings of the whole genome comparisons, indicating highly conserved genomes, as well as the Bootscan analyses, indicating no recombination events with other HAdV genomes.

Genome recombination analysis

Determination of potential recombination events was by examining protein identities, with the HAdV-C5_FS as reference, and using recombination detection software tools (Simplot and Bootscan; Lole

Please cite this article as: Seto, J., et al., Computational analysis of adenovirus serotype 5 (HAdV-C5) from an HAdV coinfection shows genome stability after 45 years of circulation, Virology (2010), doi:10.1016/j.virol.2010.05.010

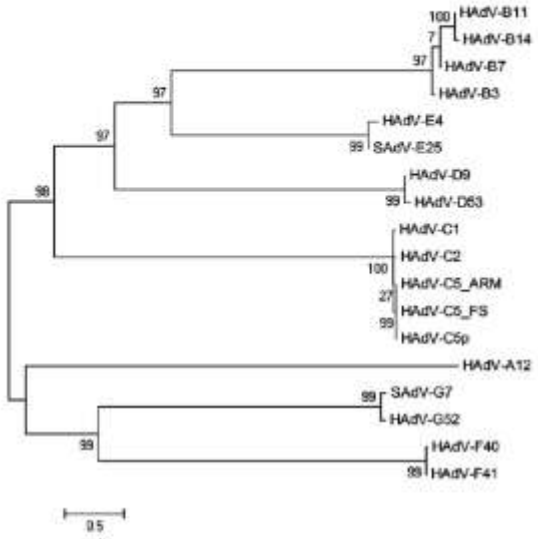


Fig. 3. Whole genome phylogenetic analysis. A neighbor-joining tree of a selection of HAdV genomes shows the HAdV-C5 strains clading with the other members of the C species. Numbers shown indicate the percentages of 1000 bootstrap replications producing the clade. The scale bar is in units of nucleotide substitutions per site.

et al., 1999). Both revealed no evidence of recombination with other serotypes (data not shown). This is another indication that the genome of HAdV-C5_FS genome is stable, even beyond the 45 years of circulation in the population as discussed in this manuscript, i.e., the prototype was not a product of recombination, at least not in association with any sequenced HAdV genomes. In contrast, species C does contain members that show recombination events: the HAdV-C6 genome contains a hexon recombination (Seto, unpublished results) and recent characterizations of sixteen field strains of species C display evidence of recombination, leading to the hypothesis that recombination in HAdVs may be frequent (Lukashev et al., 2008). Recent reanalysis of archived genomes of established serotype identity have provided revelations of genome instability, including recombination in the penton (Robinson et al., 2009) and the hexon (Seto, unpublished results).

The lack of genome recombination in this HAdV-C5 field strain is interesting since it has been presumably circulating in the general population and, in this case, was isolated as a coinfection with HAdV-B21. It is noted that the context of this specific coinfection may be misleading as the length of coinfection time is not known, but the potential for genome exchange was present. Although the coinfecting HAdV is a species B serotype, erroneous concerns and previous conclusions that interspecific recombination may not occur (Lukashev et al., 2008) are now proving premature and unsubstantiated as two serotypes, HAdV-B16 and HAdV-E4, are demonstrated to contain interspecies recombination using genomics and computational means (Seto, unpublished results).

HAdV coinfection

As the first HAdV coinfection identified, characterized, and recently reported formally, NHRC 7151 has been vetted in great detail, by molecular typing with species-specific PCR and also with a microarray assay to confirm that the two coinfecting genomes are HAdV-C5 and B21 (Lin et al., 2006; Vora et al., 2006). Additionally, NHRC 7151 was subjected to multiple cycles of cell culturing in A-549 (lung cell line; ATCC CCL-185) and plaque purification to demonstrate that the strains coinfected the same cell at high MOI. Additional HAdV coinfection strains have been reported in surveys of respiratory pathogens and outbreaks (Echavarria et al., 2006; Metzgar et al., 2005).

Whole genome sequencing produced initially two assemblages of divergent contigs, one that was consistent with HAdV-C5 and one that was identified as HAdV-B21. Table 1 presents the BLAST results of four segments against GenBank entries, resulting in the identification of HAdV-B21 as the coinfectant. Each was further analyzed and showed the identity as well as nucleotide differences; for example, an alignment of "segment-4" from NHRC 7151 with HAdV-B21 DNA polymerase shows eight nucleotide substitutions (Fig. 4). This suggests that the HAdV-B21 field strain contained within this NHRC 7151 coinfection is not identical to its prototype, unlike HAdV-C5_FS. Additionally, it is noted that originally, NHRC 7151 showed a microneutralization result as an HAdV-E4. One explanation is that neutralization is predictably poor at identifying the component serotypes in coinfections. Another explanation is perhaps this particular HAdV-B21 field strain contains a partial hexon recombination that serum neutralizes as HAdV-E4, similar to a recombination found in HAdV-B55, resulting in a serum neutralization change (Walsh et al., 2010; Yang et al., 2009; Zhu et al., 2009). Three other partial hexon recombinant HAdV genomes have been reported:

Table 1
BLAST analysis and identification of HAdV-B21-like DNA segments from NHRC 7151. DNA sequences were generated during the shotgun sequencing phase and BLAST-analyzed against GenBank sequences. Additional sequence alignments confirmed these results.

	Segment 1		Segment 2		Segment 3		Segment 4	
Gene ID	E1B 55K		IVa2		IVa2		DNA polymerase	
HAdV	Score	E-value	Score	E-value	Score	E-value	Score	E-value
HAdV-B21	2806	0	531	E−152	240	3.00E−65	834	0
HAdV-B16	2775	0	500	E−143	240	3.00E−65	816	0
HAdV-B50	2767	0	531	E−152	240	3.00E−65	825	0
HAdV-B3	2656	0	515	E−148	232	7.00E−63	820	0
HAdV-B7	2648	0	515	E−148	240	3.00E−65	825	0
HAdV-B11	1083	0	278	3.00E−76	145	1.00E−36	536	3.00E−149
HAdV-B34	1035	0	248	3.00E−67	137	3.00E−34	527	2.00E−146
HAdV-B35	1029	0	248	3.00E−67	137	3.00E−34	542	7.00E−151
HAdV-B14	1017	0	236	1.00E−63	131	2.00E−32	527	2.00E−146
HAdV-E4	125	1.00E−29	127	7.00E−31	161	2.00E−41	455	8.00E−125
HAdV-C1	40	6.00E−04	70	1.00E−13	56	9.00E−10	284	9.00E−78
HAdV-C2	40	6.00E−04	70	1.00E−13	56	9.00E−10	271	6.00E−74
HAdV-C5	40	6.00E−04	66	2.00E−12	56	9.00E−10	271	6.00E−74
HAdV-F40	34	0.038	44	8.00E−06	60	5.00E−11	237	3.00E−64
HAdV-A12	36	0.01	58	5.00E−10	70	6.00E−14	226	3.00E−61

Please cite this article as: Seto, J., et al., Computational analysis of adenovirus serotype 5 (HAdV-C5) from an HAdV coinfection shows genome stability after 45 years of circulation, Virology (2010), doi:10.1016/j.virol.2010.05.010

```
HAdV-B21        GCATTGGACAGAAGCTTGGCGATGGAGCGCATGGTTTGGTTCTTTTCCTTGTCCGCGCGC
segment-4       GCATTGGACAGAAGCTTGGCGATGGAGCGCATGGTTTGGTTCTTTTCCTTGTCCGCGCGC
                ************************************************************

HAdV-B21        TCCTTGGCGGCGATGTTAAGCTGGACGTACTCGCGCGCCACACATTTCCATTCAGGGAAG
segment-4       TCCTTGGCGGCGATGTTAAGCTGGACGTACTCGCGCGCCACACATTTCCATTCAGGGAAG
                ************************************************************

HAdV-B21        ATGGTTGTCAGTTCATCCGGAACTATTCTGACTCGCCATCCCCTATTGTGCAGGGTTATC
segment-4       ATGGTTGTCAGTTCATCCGGAACTATTCTGACTCGCCATCCCCTATTGTGCAGTGTTATC
                ************************************************** ******

HAdV-B21        AGATCCACACTGGTGGCCACCTCGCCTCGGAGGGGCTCATTGGTCCAGCAGAGTCGACCT
segment-4       AGATCCACACTGGTGGCCACCTCGCCTCGGAGGGGCTCATTGGTCCAGCAGAGTCGACCT
                ************************************************************

HAdV-B21        CCTTTTCTTGAACAGAAAGGTGGGAGGGGGTCTAGCATGAACTCATCAGGGGGGTCCGCA
segment-4       CCTTTCTTGAAACAGAAAGGGGGGAGGGGGTCTAGCATGAACTCATCAGGGGGGTCCGCA
                *****  *  ********** ****************************************

HAdV-B21        TCTATGGTAAATATTCCCGGTAGCAAATCTTTGTCAAAATAGCTGATGGTGGCGGGATCA
segment-4       TCTATGGTAAATATTCCCGGGAGTAAATCTTTGTCAAAATAGCTGATGGTGGCGGGATCA
                ******************** ** ************************************

HAdV-B21        TCCAAGGTCATCTGCCATTCTCGAACTGCCAGCGCGCGCTCATAGGGGTTAAGAGGGGTG
segment-4       TCCAAGGTCATCTGCCATTCTCGAACTGCCAGCGCGCGCTCATAGGGGTTAAGAGGGGTG
                ************************************************************

HAdV-B21        CCCCAGGGCATGGGGTGGGTGAGCGCGGAGGCATACATGCCACAGATATCGTAGACATAG
segment-4       CCCCAGGGCATGGGGTGGGTGAGCGCGGAGGCATACATGCCACAGATATCGTAGACATAG
                ************************************************************
```

Fig. 4. NHRC 7151 coinfection "segment 4" DNA sequence alignment. Sequence alignment and analysis identifies this 480 nucleotide sequence from NHRC 7151 as from the HAdV-B21 DNA polymerase gene.

HAdV-D53, serum neutralizing as HAdV-D22 (Engelmann et al., 2006; Walsh et al., 2009), and one each found in HAdV-E4 and HAdV-B16 (Seto, unpublished results), providing serological cross-reaction with each other historically (Hierholzer, Stone, and Broderson, 1991; Wigand et al., 1985).

For reference, and as a measure of different genome stabilities, a contiguous number of 200 nucleotides in a sequenced region of HAdV-B21 showed eight base substitutions and no indels (Fig. 4). Within the same number of nucleotides, but as four noncontiguous segments in HAdV-C5_FS, there were one base substitution and three indels (Figs. 1–3). This was also the total number of genome changes in the HAdV-C5_FS genome.

Bioinformatics of genome stability

HAdV genome recombination has been reported in the literature (Boursnell and Mautner, 1981a,b; Crawford-Miksza and Schnurr, 1996b; Lukashev et al., 2008; Mautner and Mackay, 1984; Williams et al., 1975). High-resolution analysis of primary nucleotide sequences has only recently provided comprehensive and exact documentation of genome changes, including recombination (Robinson et al., 2009; Walsh et al., 2009, 2010; Yang et al., 2009; Seto, unpublished results), leading to questions of where they may occur and how frequently these and other genome changes occur.

The isolation of HAdV strains, including variants, novel serotypes and "intermediate" or "intertypic" strains from immunocompromised patients provided a potential answer as to "where they may occur" (Crawford-Miksza and Schnurr, 1996a; De Jong et al., 1999; Hierholzer et al., 1988a,b; Khoo et al., 1995; Koopmann et al., 2000; Schnurr et al., 1995). However, the hypothesis of genome shuffling in this "bioreactor" was not supported formally until HAdV coinfections were observed and characterized (Echavarria et al., 2006; Lin et al., 2006; Metzgar et al., 2005; Vora et al., 2006), providing a means to how shuffling between genomes may occur, that is, identifying a "clinically relevant environment suitable for the generation of new recombinational variants" (Vora et al., 2006). The question now is how frequent are these genome changes.

Overall genome instability and resulting strain variation are important for both vaccine and vector development and for understanding HAdV biology, pathology, and outbreaks. Using limited molecular typing, outbreaks have been studied to identify which strains and what changes are associated with disease and the severity of the infection. For example, strain variations in species C members have been examined using restriction enzyme (RE) digestion patterns (Kajon et al., 1993, 1999). The same approach is used for a survey of 724 HAdV-E4 strains causing acute respiratory disease (ARD) in the US, showing seven genome types from 1997 to 2003 (Kajon et al., 2007). Results include "very different, and often stable, genome type distributions at different geographic sites, despite the homogeneity of the recruit source population" (Kajon et al., 2007). Outbreaks caused by ARD-associated pathogens, HAdV-B3 and B7, have been characterized using RE patterns (Li and Wadell, 1986, 1988; Lin et al., 2004), resulting in the characterization of variants that are more virulent or have more serious impact on public health, for example HAdV-B7h was observed to cause more severe symptoms than other variants, including causing 17 of 18 fatal cases and outbreak (Kajon et al., 1996; Kajon and Wadell, 1996; Larranaga et al., 2000).

An example of the relevance of genome stability studies is a survey of HAdV-E4 and B7 strains by the sequencing of the hexon gene and using neutralization tests to assess their suitability as vaccine candidates. This study showed "no significant strain variation in the neutralization epitopes of the AV7a genome over a 42-year period" but found that "the current strain of AV 4, which has been in circulation since 1995, is significantly different from the AV 4 prototype and the vaccine strains" (Crawford-Miksza et al., 1999). Despite this, current vaccine reimplementation uses the original HAdV-E4 vaccine strain (Lyons et al., 2008). Whole genome bioinformatics and genome stability analysis of the HAdV-E4 prototype (Purkayastha et al., 2005a), vaccine (Purkayastha et al., 2005c), and two current field strains (Seto, unpublished results) supports this reimplementation effort and presents an alternative view to the earlier study (Crawford-Miksza et al., 1999).

Additional genomic and computational studies provide extending and complementing comprehensive understanding of genome

Please cite this article as: Seto, J., et al., Computational analysis of adenovirus serotype 5 (HAdV-C5) from an HAdV coinfection shows genome stability after 45 years of circulation, Virology (2010), doi:10.1016/j.virol.2010.05.010

stability, e.g., five HAdV-B3 strains spanning 50 years (Mahadevan et al., 2010); HAdV-B7 and E4 prototype and vaccine strains spanning a minimum of 10 years (Purkayastha et al., 2005c); HAdV-B7 prototype and field strain spanning 50 years (Seto, unpublished results); and HAdV-E4 field strains spanning 50 years (Seto, unpublished results). These and other studies show that assumptions based, in part, on incomplete, i.e., limited molecular typing, data are now being refined in the context of whole genome data, for example the recent genomic and computational characterization of HAdV interspecies recombination (Seto, unpublished results).

Conclusion

Presented here is the resolution and documentation of a coinfecting genome sequence of the HAdV-C5 field strain that was remarkably conserved despite 45 years of circulation in the human population when compared to the "pristine", archived sequence of the prototype strain of this serotype. Of note, this field strain genome revealed no recombination events, particularly with its coinfectant HAdV-B21, and showed few nucleotide changes *in toto*. Mirrored against this is another version of the prototype genome, sequenced as a composite of strains that had been circulating and amplified in the research community for 39 years. A similar analysis of other coinfection partners will be important to provide an understanding of the role of HAdV coinfections in genome shuffling, and how this provides for the evolution of HAdV, particularly in the context of novel types and in producing emerging and reemerging HAdV pathogens. These observations and understanding of HAdV genome stability (and instability) impact the development and long-term efficacy of vaccines and the selection and use of HAdV genomes as gene delivery vectors.

As a footnote, the importance of accurate genome sequencing and annotation is highlighted by a single putative sequencing error in the Genbank "HAdV-C5_ARM" reference sequence (E3-encoded RID-α protein frameshift). This reinforces the critical importance of bioinformatics, particularly careful and thoughtful annotation and sequence analysis, rather than having sequence data deposits as a goal.

Materials and methods

A coinfection sample "NHRC 7151.AV5.V.98.FJ" was obtained from the US Naval Health Research Center (Kevin Russell, San Diego, CA); HAdV-C5 was sequenced from this sample (GenBank AY601635). Two versions, but of the same original source, of the HAdV-C5 prototype genomes are available: one is a laboratory-circulating strain (HAdV-C5p; AC_000008) and one is an ATCC archived "minimally amplified" strain (HAdV-C5_ARM; AY339865; Chroboczek et al., 1992; Sugarman et al., 2003).

Virus growth in A-549 cells and DNA production were outsourced to Virapur (San Diego, CA). Genome sequencing was outsourced to Commonwealth Biotechnologies (Richmond, VA). Protocols for both were the same as ones proven for a series of HAdV genomes analyzed (Lauer et al., 2004; Purkayastha et al., 2005a,b,c; Seto et al., 2009). The initial shotgun strategy was replaced by a directed primer walk strategy using HAdV-C5p GenBank data, once the coinfection was detected; HAdV-B21 sequences were confirmed with additional PCR-based amplification and sequencing, and by computational analysis.

Bioinformatic analyses were performed using protocols noted earlier (Seto et al., 2009; Walsh et al., 2009): (1) MAFFT was used to compute alignments of DNA sequences (Katoh and Toh, 2008); (2) sequence percent identities were calculated using the UCSF Chimera package (Pettersen et al., 2004); (3) Artemis was used to view and analyze genome pairs (Rutherford et al., 2000); (4) percent identity values for annotated proteins were calculated using a BioJava implementation of a Needleman–Wunsch algorithm (Holland et al., 2008); (5) phylogenetic neighbor-joining trees were constructed using MEGA4 (Tamura et al., 2007); and (6) recombination analysis was performed using SimPlot (Lole et al., 1999).

Acknowledgments

We thank Clark Tibbetts for initial discussions and for providing the impetus and opportunity to pursue this line of research. Anjan Purkayastha, James M. Clark, and Kevin Russell provided appreciated preliminary analyses and discussions. During the collection of the genome data and preliminary data 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 US 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 US Department of Defense.

NHRC 7151 was collected with the support of the Global Emerging Infections Surveillance and Response System (GEIS), a Division of the Armed Forces Health Surveillance Center, Silver Spring, MD (www.afhsc.mil).

References

Boursnell, M., Mautner, V., 1981a. In vitro construction of a recombinant adenovirus Ad2:Ad5. Gene 13 (3), 311–317.
Boursnell, M.E., Mautner, V., 1981b. Recombination in adenovirus: crossover sites in intertypic recombinants are located in regions of homology. Virology 112 (1), 198–209.
Chroboczek, J., Bieber, F., Jacrot, B., 1992. The sequence of the genome of adenovirus type 5 and its comparison with the genome of adenovirus type 2. Virology 186 (1), 280–285.
Crawford-Miksza, L., Schnurr, D.P., 1996a. Seroepidemiology of new AIDS-associated adenoviruses among the San Francisco Men's Health Study. J. Med. Virol. 50 (3), 230–236.
Crawford-Miksza, L.K., Schnurr, D.P., 1996b. Adenovirus serotype evolution is driven by illegitimate recombination in the hypervariable regions of the hexon protein. Virology 224 (2), 357–367.
Crawford-Miksza, L.K., Nang, R.N., Schnurr, D.P., 1999. Strain variation in adenovirus serotypes 4 and 7a causing acute respiratory disease. J. Clin. Microbiol. 37 (4), 1107–1112.
De Jong, J.C., Wermenbol, A.G., Verweij-Uijterwaal, M.W., Slaterus, K.W., Wertheim-Van Dillen, P., Van Doornum, G.J., Khoo, S.H., Hierholzer, J.C., 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 (12), 3940–3945.
Echavarria, M., 2009. Adenovirus, In: Zuckerman, A.J., Banatvala, J.E., Schoub, B.D., Griffiths, P.D., Mortimer, P. (Eds.), Principles and Practice of Clinical Virology, Sixth edition. John Wiley and Sons, San Diego, CA, pp. 463–488.
Echavarria, M., Maldonado, D., Elbert, G., Videla, C., Rappaport, R., Carballal, G., 2006. Use of PCR to demonstrate presence of adenovirus species B, C, or F as well as coinfection with two adenovirus species in children with flu-like symptoms. J. Clin. Microbiol. 44 (2), 625–627.
Edwards, K.M., Thompson, J., Paolini, J., Wright, P.F., 1985. Adenovirus infections in young children. Pediatrics 76 (3), 420–424.
Engelmann, I., Madisch, I., Pommer, H., Heim, A., 2006. An outbreak of epidemic keratoconjunctivitis caused by a new intermediate adenovirus 22/H8 identified by molecular typing. Clin. Infect. Dis. 43 (7), e64–e66.
Fox, J.P., Brandt, C.D., Wassermann, F.E., Hall, C.E., Spigland, I., Kogon, A., Elveback, L.R., 1969. The virus watch program: a continuing surveillance of viral infections in metropolitan New York families. VI. Observations of adenovirus infections: virus excretion patterns, antibody response, efficiency of surveillance, patterns of infections, and relation to illness. Am. J. Epidemiol. 89 (1), 25–50.
Garnett, C.T., Talekar, G., Mahr, J.A., Huang, W., Zhang, Y., Ornelles, D.A., Gooding, L.R., 2009. Latent species C adenoviruses in human tonsil tissues. J. Virol. 83 (6), 2417–2428.
Gaydos, C.A., Gaydos, J.C., 1995. Adenovirus vaccines in the U.S. military. Mil. Med. 160 (6), 300–304.
Gray, G.C., Goswami, P.R., Malasig, M.D., Hawksworth, A.W., Trump, D.H., Ryan, M.A., Schnurr, D.P., 2000. Adult adenovirus infections: loss of orphaned vaccines

Please cite this article as: Seto, J., et al., Computational analysis of adenovirus serotype 5 (HAdV-C5) from an HAdV coinfection shows genome stability after 45 years of circulation, Virology (2010), doi:10.1016/j.virol.2010.05.010

precipitates military respiratory disease epidemics. For the Adenovirus Surveillance Group. Clin. Infect. Dis. 31 (3), 663–670.
Gray, G.C., McCarthy, T., Lebeck, M.G., Schnurr, D.P., Russell, K.L., Kajon, A.E., Landry, M.L., Leland, D.S., Storch, G.A., Ginocchio, C.C., Robinson, C.C., Demmler, G.J., Saubolle, M.A., Kehl, S.C., Selvarangan, R., Miller, M.B., Chappell, J.D., Zerr, D.M., Kiska, D.L., Halstead, D.C., Capuano, A.W., Setterquist, S.F., Chorazy, M.L., Dawson, J.D., Erdman, D.D., 2007. Genotype prevalence and risk factors for severe clinical adenovirus infection, United States 2004-2006. Clin. Infect. Dis. 45 (9), 1120–1131.
Hierholzer, J.C., Pumarola, A., Rodriguez-Torres, A., Beltran, M., 1974. Occurrence of respiratory illness due to an atypical strain of adenovirus type 11 during a large outbreak in Spanish military recruits. Am. J. Epidemiol. 99 (6), 434–442.
Hierholzer, J.C., Adrian, T., Anderson, L.J., Wigand, R., Gold, J.W., 1988a. Analysis of antigenically intermediate strains of subgenus B and D adenoviruses from AIDS patients. Arch. Virol. 103 (1–2), 99–115.
Hierholzer, J.C., Wigand, R., Anderson, L.J., Adrian, T., Gold, J.W., 1988b. Adenoviruses from patients with AIDS: a plethora of serotypes and a description of five new serotypes of subgenus D (types 43–47). J. Infect. Dis. 158 (4), 804–813.
Hierholzer, J.C., Stone, Y.O., Broderson, J.R., 1991. Antigenic relationships among the 47 human adenoviruses determined in reference horse antisera. Arch. Virol. 121 (1–4), 179–197.
Holland, R.C., Down, T.A., Pocock, M., Prlic, A., Huen, D., James, K., Foisy, S., Drager, A., Yates, A., Heuer, M., Schreiber, M.J., 2008. BioJava: an open-source framework for bioinformatics. Bioinformatics 24 (18), 2096–2097.
Horwitz, M.S., 2004. Function of adenovirus E3 proteins and their interactions with immunoregulatory cell proteins. J. Gene Med. 6 (Suppl 1), S172–S183.
Houng, H.S., Lott, L., Gong, H., Kuschner, R.A., Lynch, J.A., Metzgar, D., 2009. Adenovirus microsatellite reveals dynamics of transmission during a recent epidemic of human adenovirus serotype 14 infection. J. Clin. Microbiol. 47 (7), 2243–2248.
Ishiko, H., Shimada, Y., Konno, T., Hayashi, A., Ohguchi, T., Tagawa, Y., Aoki, K., Ohno, S., Yamazaki, S., 2008. Novel human adenovirus causing nosocomial epidemic keratoconjunctivitis. J. Clin. Microbiol. 46 (6), 2002–2008.
Kajon, A.E., Wadell, G., 1996. Sequence analysis of the E3 region and fiber gene of human adenovirus genome type 7h. Virology 215 (2), 190–196.
Kajon, A.E., Suarez, M.V., Avendano, L.F., Hortal, M., Wadell, G., 1993. Genome type analysis of South American adenoviruses of subgenus C collected over a 7-year period. Arch. Virol. 132 (1–2), 29–35.
Kajon, A.E., Mistchenko, A.S., Videla, C., Hortal, M., Wadell, G., Avendano, L.F., 1996. Molecular epidemiology of adenovirus acute lower respiratory infections of children in the south cone of South America (1991–1994). J. Med. Virol. 48 (2), 151–156.
Kajon, A.E., Portes, S.A., de Mello, W.A., Nascimento, J.P., Siqueira, M.M., 1999. Genome type analysis of Brazilian adenovirus strains of serotypes 1, 2, 3, 5, and 7 collected between 1976 and 1995. J. Med. Virol. 58 (4), 408–412.
Kajon, A.E., Xu, W., Erdman, D.D., 2005. Sequence polymorphism in the E3 7.7K ORF of subspecies B1 human adenoviruses. Virus Res. 107 (1), 11–19.
Kajon, A.E., Moseley, J.M., Metzgar, D., Huong, H.S., Wadleigh, A., Ryan, M.A., Russell, K.L., 2007. Molecular epidemiology of adenovirus type 4 infections in US military recruits in the postvaccination era (1997-2003). J. Infect. Dis. 196 (1), 67–75.
Katoh, K., Toh, H., 2008. Recent developments in the MAFFT multiple sequence alignment program. Brief. Bioinform. 9 (4), 286–298.
Khoo, S.H., Bailey, A.S., de Jong, J.C., Mandal, B.K., 1995. Adenovirus infections in human immunodeficiency virus-positive patients: clinical features and molecular epidemiology. J. Infect. Dis. 172 (3), 629–637.
Koopmann, J., Dombrowski, F., Rockstroh, J.K., Pfeifer, U., Sauerbruch, T., Spengler, U., 2000. Fatal pneumonia in an AIDS patient coinfected with adenovirus and Pneumocystis carinii. Infection 28 (5), 323–325.
Larranaga, C., Kajon, A., Villagra, E., Avendano, L.F., 2000. Adenovirus surveillance on children hospitalized for acute lower respiratory infections in Chile (1988–1996). J. Med. Virol. 60 (3), 342–346.
Lauer, K.P., Llorente, I., Blair, E., Seto, J., Krasnov, V., Purkayastha, A., Ditty, S.E., Hadfield, T.L., Buck, C., Tibbetts, C., Seto, D., 2004. Natural variation among human adenoviruses: genome sequence and annotation of human adenovirus serotype 1. J. Gen. Virol. 85 (Pt 9), 2615–2625.
Li, Q.G., Wadell, G., 1986. Analysis of 15 different genome types of adenovirus type 7 isolated on five continents. J. Virol. 60 (1), 331–335.
Li, Q.G., Wadell, G., 1988. Comparison of 17 genome types of adenovirus type 3 identified among strains recovered from six continents. J. Clin. Microbiol. 26 (5), 1009–1015.
Li, Q.G., Hambraeus, J., Wadell, G., 1991. Genetic relationship between thirteen genome types of adenovirus 11, 34, and 35 with different tropisms. Intervirology 32 (6), 338–350.
Lin, K.H., Lin, Y.C., Chen, H.L., Ke, G.M., Chiang, C.J., Hwang, K.P., Chu, P.Y., Lin, J.H., Liu, D.P., Chen, H.Y., 2004. A two decade survey of respiratory adenovirus in Taiwan: the reemergence of adenovirus types 7 and 4. J. Med. Virol. 73 (2), 274–279.
Lin, B., Wang, Z., Vora, G.J., Thornton, J.A., Schnur, J.M., Thach, D.C., Blaney, K.M., Ligler, A.G., Malanoski, A.P., Santiago, J., Walter, E.A., Agan, B.K., Metzgar, D., Seto, D., Daum, L.T., Kruzelock, R., Rowley, R.K., Hanson, E.H., Tibbetts, C., Stenger, D.A., 2006. Broad-spectrum respiratory tract pathogen identification using resequencing DNA microarrays. Genome Res. 16 (4), 527–535.
Lindenbaum, J.O., Field, J., Hurwitz, J., 1986. The adenovirus DNA binding protein and adenovirus DNA polymerase interact to catalyze elongation of primed DNA templates. J. Biol. Chem. 261 (22), 10218–10227.
Lole, K.S., Bollinger, R.C., Paranjape, R.S., Gadkari, D., Kulkarni, S.S., Novak, N.G., Ingersoll, R., Sheppard, H.W., Ray, S.C., 1999. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. J. Virol. 73 (1), 152–160.
Lukashev, A.N., Ivanova, O.E., Eremeeva, T.P., Iggo, R.D., 2008. Evidence of frequent recombination among human adenoviruses. J. Gen. Virol. 89 (Pt 2), 380–388.
Lyons, A., Longfield, J., Kuschner, R., Straight, T., Binn, L., Seriwatana, J., Reitstetter, R., Froh, I.B., Craft, D., McNabb, K., Russell, K., Metzgar, D., Liss, A., Sun, X., Towle, A., Sun, W., 2008. A double-blind, placebo-controlled study of the safety and immunogenicity of live, oral type 4 and type 7 adenovirus vaccines in adults. Vaccine 26 (23), 2890–2898.
Mahadevan, P., Seto, J., Tibbetts, C., Seto, D., 2010. Natural variants of human adenovirus type 3 provide evidence for relative genome stability across time and geographic space. Virology 397, 113–118.
Mautner, V., Mackay, N., 1984. Recombination in adenovirus: analysis of crossover sites in intertypic overlap recombinants. Virology 139 (1), 43–52.
Metzgar, D., Osuna, M., Yingst, S., Rakha, M., Earhart, K., Elyan, D., Esmat, H., Saad, M.D., Kajon, A., Wu, J., Gray, G.C., Ryan, M.A., Russell, K.L., 2005. PCR analysis of Egyptian respiratory adenovirus isolates, including identification of species, serotypes, and coinfections. J. Clin. Microbiol. 43 (11), 5743–5752.
Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E., 2004. UCSF Chimera--a visualization system for exploratory research and analysis. J. Comput. Chem. 25 (13), 1605–1612.
Purkayastha, A., Ditty, S.E., Su, J., McGraw, J., Hadfield, T.L., Tibbetts, C., Seto, D., 2005a. 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 (4), 2559–2572.
Purkayastha, A., Su, J., Carlisle, S., Tibbetts, C., Seto, D., 2005b. 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 (1), 114–129.
Purkayastha, A., Su, J., McGraw, J., Ditty, S.E., Hadfield, T.L., Seto, J., Russell, K.L., Tibbetts, C., Seto, D., 2005c. 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 (7), 3083–3094.
Robinson, C.M., Rajaiya, J., Walsh, M.P., Seto, D., Dyer, D.W., Jones, M.S., Chodosh, J., 2009. Computational analysis of human adenovirus type 22 provides evidence for recombination between human adenoviruses species D in the penton base gene. J. Virol.
Rowe, W.P., Huebner, R.J., Gilmore, L.K., Parrott, R.H., Ward, T.G., 1953. Isolation of a cytopathogenic agent from human adenoids undergoing spontaneous degeneration in tissue culture. Proc. Soc. Exp. Biol. Med. 84 (3), 570–573.
Russell, W.C., 2009. Adenoviruses: update on structure and function. J. Gen. Virol. 90 (Pt 1), 1–20.
Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.A., Barrell, B., 2000. Artemis: sequence visualization and annotation. Bioinformatics 16 (10), 944–945.
Schnurr, D., Bollen, A., Crawford-Miksza, L., Dondero, M.E., Yagi, S., 1995. Adenovirus mixture isolated from the brain of an AIDS patient with encephalitis. J. Med. Virol. 47 (2), 168–171.
Seto, J., Walsh, M.P., Mahadevan, P., Purkayastha, A., Clark, J.M., Tibbetts, C., Seto, D., 2009. Genomic and bioinformatics analyses of HAdV-14p, reference strain of a re-emerging respiratory pathogen and analysis of B1/B2. Virus Res. 143 (1), 94–105.
Sugarman, B.J., Hutchins, B.M., McAllister, D.L., Lu, F., Thomas, B.K., 2003. The complete nucleotide acid sequence of the adenovirus type 5 reference material (ARM) genome. Bioprocessing J. September/October, 27–32.
Tamura, K., Dudley, J., Nei, M., Kumar, S., 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol. Biol. Evol. 24 (8), 1596–1599.
Top Jr., F.H., Grossman, R.A., Bartelloni, P.J., Segal, H.E., Dudding, B.A., Russell, P.K., Buescher, E.L., 1971. Immunization with live types 7 and 4 adenovirus vaccines. I. Safety, infectivity, antigenicity, and potency of adenovirus type 7 vaccine in humans. J. Infect. Dis. 124 (2), 148–154.
Vora, G.J., Lin, B., Gratwick, K., Meador, C., Hansen, C., Tibbetts, C., Stenger, D.A., Irvine, M., Seto, D., Purkayastha, A., Freed, N.E., Gibson, M.G., Russell, K., Metzgar, D., 2006. Co-infections of adenovirus species in previously vaccinated patients. Emerg. Infect. Dis. 12 (6), 921–930.
Walsh, M.P., Chintakuntlawar, A., Robinson, C.M., Madisch, I., Harrach, B., Hudson, N.R., Schnurr, D., Heim, A., Chodosh, J., Seto, D., Jones, M.S., 2009. Evidence of molecular evolution driven by recombination events influencing tropism in a novel human adenovirus that causes epidemic keratoconjunctivitis. PLoS ONE 4 (6), e5635.
Walsh, M.P., Seto, J., Jones, M.S., Chodosh, J., Xu, W., Seto, D., 2010. Computational analysis identifies human adenovirus type 55 as a re-emergent acute respiratory disease pathogen. J. Clin. Microbiol. 48, 991–993.
Wigand, R., Sehn, N., Hierholzer, J.C., de Jong, J.C., Adrian, T., 1985. Immunological and biochemical characterization of human adenoviruses from subgenus B. I. Antigenic relationships. Arch. Virol. 84 (1–2), 63–78.
Williams, J., Grodzicker, T., Sharp, P., Sambrook, J., 1975. Adenovirus recombination: physical mapping of crossover events. Cell 4 (2), 113–119.
Wold, W.S., Doronin, K., Toth, K., Kuppuswamy, M., Lichtenstein, D.L., Tollefson, A.E., 1999. Immune responses to adenoviruses: viral evasion mechanisms and their implications for the clinic. Curr. Opin. Immunol. 11 (4), 380–386.
Yang, Z., Zhu, Z., Tang, L., Wang, L., Tan, X., Yu, P., Zhang, Y., Tian, X., Wang, J., Zhang, Y., Li, D., Xu, W., 2009. Genomic analyses of recombinant adenovirus type 11a in China. J. Clin. Microbiol.
Zhu, Z., Zhang, Y., Xu, S., Yu, P., Tian, X., Wang, L., Liu, Z., Tang, L., Mao, N., Ji, Y., Li, C., Yang, Z., Wang, S., Wang, J., Li, D., Xu, W., 2009. Outbreak of acute respiratory disease in China caused by B2 species of adenovirus type 11. J. Clin. Microbiol. 47 (3), 697–703.

Please cite this article as: Seto, J., et al., Computational analysis of adenovirus serotype 5 (HAdV-C5) from an HAdV coinfection shows genome stability after 45 years of circulation, Virology (2010), doi:10.1016/j.virol.2010.05.010

Part 3

Applying Genomic and Bioinformatic Resources to Human Adenovirus Genomes for Use in Vaccine Development and for Applications in Vector Development for Gene Delivery

Viruses **2010**, *2*, 1-26; doi:10.3390/v2010001

OPEN ACCESS

viruses

ISSN 1999-4915

www.mdpi.com/journal/viruses

Review

Applying Genomic and Bioinformatic Resources to Human Adenovirus Genomes for Use in Vaccine Development and for Applications in Vector Development for Gene Delivery

Jason Seto [1], Michael P. Walsh [1], Padmanabhan Mahadevan [2], Qiwei Zhang [3] and Donald Seto [1,*]

[1] Department of Bioinformatics and Computational Biology, George Mason University, 10900 University Blvd., MSN 5B3, Manassas, VA 20110, USA; E-Mails: jseto@gmu.edu (J.S.); mwalsh5@gmu.edu (M.P.W.)

[2] Department of Biological Sciences, Vanderbilt University, Nashville, TN 37235, USA; E-Mail: padmahadevan@gmail.com

[3] Department of Microbiology, The University of Hong Kong, Pokfulam, Hong Kong SAR, China; E-Mail: zhang.qiwei@yahoo.com

* Author to whom correspondence should be addressed; E-Mail: dseto@gmu.edu; Tel.: +1 703 993-8403; Fax: +1 703 993-8401.

Received: 28 September 2009; in revised form: 5 December 2009 / Accepted: 17 December 2009 / Published: 6 January 2010

Abstract: Technological advances and increasingly cost-effect methodologies in DNA sequencing and computational analysis are providing genome and proteome data for human adenovirus research. Applying these tools, data and derived knowledge to the development of vaccines against these pathogens will provide effective prophylactics. The same data and approaches can be applied to vector development for gene delivery in gene therapy and vaccine delivery protocols. Examination of several field strain genomes and their analyses provide examples of data that are available using these approaches. An example of the development of HAdV-B3 both as a vaccine and also as a vector is presented.

Keywords: genomics; bioinformatics; molecular evolution; pathoepidemiology; field strain; adenovirus vaccine; human gene therapy vector; vaccine delivery vector

---

1. Introduction

Recombinant DNA technology has provided a molecular medicine scenario and opportunity where a disease or illness may be treated or prevented by the introduction of a specific gene or a fragment of a gene into the patient using a recombinant vector. This may be either as gene therapy, where a missing or non-functional gene is substituted with a functional version, or as vaccination, where a specific defined antigen is introduced [1,2]. Adenoviruses (AdVs) have been and are currently used as the basis for vectors delivering these genes [1-4], despite concerns of pre-existing immune responses [5,6]. In the way of understanding and perhaps avoiding these problems, a confluence of genomics and bioinformatics approaches is useful by providing the primary DNA sequence data and analyses of the original genomes from which the vectors are derived, allowing for a better understanding of the "starting material"- the genome- and its "expression"- the virus. These data also provide detailed relationships of the genes found within the genomes, as well as how these gene products may relate, computationally, to the antigens that may have been presented in the past, through HAdV infections, *i.e.*, the problem of pre-existing AdV immunity [5-7].

A contrasting, but related and also important, topic is the development of human adenovirus (HAdV) genomes and viruses as a vaccine against HAdV infections [8-10]. Epidemics are caused by the respiratory and ocular HAdVs, especially in large, concentrated and vulnerable, *i.e.*, unvaccinated, populations [11-17]. With their resulting high morbidity and occasional mortality, this preventable public health problem is amenable to a vaccine, similar to ones already produced, validated and deployed previously in the U.S. military basic trainee population [8-10,18]. The same concerns and 'mechanical problems' affecting the development, safety and wide-spread applications of these HAdV vaccines, which are based on the use of the genomes themselves as both vector and antigen, allow lessons learned, using genomics and bioinformatics, to be transferred to their applications as gene delivery vectors in gene therapy and in vaccine therapy. Hence, the HAdV genomes themselves will be the focus and context of this survey of the genomics and bioinformatics resources, which provide an information-laden basis for the rational design of HAdV vaccines and vectors for gene delivery.

There are two contrasting observations of the pathoepidemiology of these viruses that have important implications for both vaccine development and in its use as a gene delivery vector. First, HAdV is a pathogen that causes a range of human illnesses, including respiratory, ocular, gastrointestinal and metabolic diseases [19,13]; however, they may also infect aymptomatically [13]. Species C outbreaks have been reported as latent infections [20], which again may lead concerns of seroprevalence. This presents problems of pre-existing immune responses and the pathogenicity itself. Additionally, as the specific symptoms and illnesses are reflections of the particular organ or tissue infected, gene delivery roles are limited to the cell tropism of the particular HAdV.

Second, genome stability is an important consideration. Recently, there are two contrasting observations of genome stability that have been presented, at the genome sequence resolution level, for five recent field isolates (manuscripts under review or in press) and three emergent pathogens [14,21,22]. Coupled with the literature that multiple simultaneous HAdV coinfections have been reported as inseparable mixtures [23] and documented using molecular typing methods [24-27], concerns of possible genome transfer between the coninfectants, perhaps resulting in new strains and serotypes [23], are raised. These new strains may arise as recombinants, particularly under conditions

that may be ideal, such as an immunocompromised individual [23,19,28]. Recombination has been suggested as a pathway for new serotypes [29]; and in general, recombination is reported for HAdV under different conditions [30-33]. This phenomenon has been taken advantage of in constructing vectors [34,35]. As an infectious disease pathogen, putative recombinants are reported as "intertypic" and "intermediate" [36-40], with epitopes mapped by serum neutralization and/or molecular typing methods. These techniques are limited and allow only a partial analysis. Recent whole genome and bioinformatic analyses document several recombinants as highly contagious emergent pathogens [38,14,41,21,22,17]. These observations suggest that, perhaps at a very low rate, recombination should be a consideration for the safety and efficacy of HAdV vaccines and use as vectors.

This is, and should be, balanced by a survey of several field strains which, along with their prototype genomes and other similar isolates, suggest that HAdV genomes may be stable in general, and, in one case of HAdV-C5, remarkably stable with four base changes across fifty years circulation both in the population and in a laboratory context (manuscript submitted).

1.1. Synopsis of HAdV biology

Adenoviruses are double-stranded DNA viruses that infect all vertebrates spanning fish to snakes to birds to humans [42,43]. Three recent reviews survey HAdVs as human pathogens, viruses and a "basic biology model subject" [19,13,44], so the reader is referred to these excellent sources for further detailed information. In brief, although their ca. 35,000-nucleotide genomes are relatively similar, there are many sequence differences (for example, species A is 58% identical to species C) along with differences in the proteomes encoded that are reflected by differences in their individual biology. These account for their tissue tropism, virulence, pathogenicity, host response/immune systems evasion and other biological characteristics. HAdVs are partitioned into seven "species" A-G, with species B separated into subspecies B1 and B2. Species G is recently recognized with the identification and description of a novel type 52 that differs from previously characterized and defined HAdVs, using genomics and computational methods [45]. Fifty-one serotypes based on serum neutralization and molecular typing techniques [23,13], were recognized until HAdV-G52 [19,13,44,45]. With, and based on, the genomic and bioinformatic analyses of several emergent HAdV pathogens from recent ocular and respiratory outbreaks, there are now 55 types reported in the literature [14,21,22]. As an aside, at the recent 9th International Adenovirus Meeting held in Dobogókő, Hungary (April 2009) where some of these genomic and bioinformatic data were presented, along with descriptions of two of the novel types, a formal "open floor" discussion led to a consensus of using "type" as part of the HAdV nomenclature scheme, allowing genome data to differentiate HAdV and keeping the original serotype names for 1-51. This has been touched upon in the literature earlier with opposing viewpoints [46]. As will be reported, and as suggested by a subcommittee at the meeting, genome data and computational analyses support and reconfirm the existing "serotype" nomenclature and classification in the context of the proposed "type" classification and nomenclature (manuscript in preparation).

1.2. Early genomics of HAdV

Despite the early recognition of the importance of HAdV as an infectious disease pathogen [19,13], its continued role as a globally circulating pathogen [13-15,38,26,17] and its previous role as a model

"organism" in DNA replication biochemistry, cell and molecular biology [44], HAdV genomes and their analyses have lagged behind other genomes in the genomics era until recently. As DNA sequencing technologies have improved and as the focus of DNA sequencing targets shifted from smaller "feasible" genomes, e.g., Phi-X174 and mitochondrion, to the larger, e.g., *H. influenza* and *E. coli*, and then to the much larger, "more relevant" (to human health and well-being) and "sexier and exotic" genomes, e.g., human, rice, silkworm and panda, adenoviruses were left behind. These thoughts do "not hold water" as adenoviruses have a tremendous impact on human biology both as a pathogen and as a biotechnology tool! As of 2002, there were only five HAdV genomes archived in GenBank, consisting of genomes that were "cobbled together" as composites of earlier published data that were coupled with "final pieces" sequenced to put together a complete genome. The original sequences, deposited by different researchers, were obtained using different sequencing methodologies and using, likely, in-house laboratory-circulating version of the prototypes. These include genomes for HAdV-C2 [47], HAdV-C5 [48], HAdV-A12 [49], HAdV-D17 (Genzyme Corp.; GenBank 1998) and HAdV-F40 [50]. With the exception of HAdV-D17, having documented sequencing errors [43], these genomes are still useful as reference sequences and relevant to understanding HAdV biology and evolution, especially with the continued annotation of their genomes [43] and the recent high resolution descriptions of recombination events in HAdV genomes [41,51,21,22]. The importance of the HAdV-C5 genome is demonstrated in the resequencing of its genome using a single methodology, and by the preparation, designation and distribution of an industry standard, "Adenovirus Reference Material" (ARM), available from ATCC (Manassas, VA) [52].

1.3. Current genomics of HAdV

In contrast to the "tsunami" of genome sequences generated for bacteriophage, viral, mitochondrial, bacterial and eukaryotic genomes paralleling the rapid and continued development of faster, robust, more efficient and more cost-effective DNA sequencing technologies and its application to many organisms and groups of organisms, the HAdV genomes database had been sparsely populated. Given the increasing and improving number of genomic and bioinformatic tools and methods, it seemed sensible to examine the well-studied AdVs using these approaches. In particular, one application is in biotechnological applications including vaccine development and diagnostics platforms. Apparently this was of interest to several research groups as well, as a "seiche" (rather than a tsunami) of HAdV sequences appeared. HAdV-B35 was sequenced independently by two different groups, both interested in vector development and the use of HAdV-B35 as an alternative to HAdV-C5 in gene therapy and gene transfer applications [53,7], and, in part, to bypass pre-existing HAdV-C immunity. The HAdV-B11 genome was sequenced independently twice at this time as well [54,55]. In addition, the first HAdV to be described clinically and historically was also sequenced in this time period, as a prelude to sequencing and analyzing the rest of the prototypes and several field strains responsible for respiratory diseases, especially acute respiratory disease (ARD) [56]. A practical consideration for these genomes was for the identification of sequence diagnostic probes for the development of a microarray-based surveillance and diagnostics assay [25]. Subsequently, the release into GenBank of 16 additional genome data sets from the authors, HAdV-B3 (two genomes), HAdV-B7 (three genomes), HAdV-B16, HAdV-B21, HAdV-B50, HAdV-B14, HAdV-B34, HAdV-C5, HAdV-C6 (embargoed) and HAdV-E4 (four genomes), added to the growing number of available HAdV genomes [25,56-60]. Also

available were five chimpanzee AdV genomes [61,62], which are of interest as alternative gene delivery vectors. Currently, there are 31 prototype and 37 field isolate genomes deposited in GenBank, with the number expected to grow as clinical investigations of HAdV-associated illnesses are leading to the identification of putatively interesting HAdVs. One such interesting observation involved a recent fatal case of ARD, with an identified HAdV isolate that also was a highly contagious ocular pathogen [63].

2. Tools and methodologies of bioinformatics for adenovirus genomes

2.1. Genomics: acquisition of data

The advent of genomics and bioinformatics has complemented and extended the available data and tools for characterizing and developing vaccines, as well as for vector development in gene transfer and delivery based on HAdVs. Having in hand the exact nucleotide sequence allows more precise and defined manipulation of the genome. Knowing exactly where and how much of the original genome to delete to allow the insertion of expressible heterologous sequences and having rapid access to the set of in silico proteome allows comparisons across similar genomes to ascertain critical and "unimportant" genes, and allows non-essential genes to be deleted. Limited resequencing ensures no genome and gene insert changes are inadvertently in the final product. Finally, as cell tropism is embedded in the primary sequence and plays a role in the delivery of genes to targeted cell types [64-66], it may be possible to alter tropism.

Sanger-based dideoxy-sequencing chemistry has been the standard methodology to date. Its several variations, conveniently and uniformly converted into kit formats, allow for high-throughput automation; for example, DYEnamic ET Terminator Cycle Sequencing kits (Amersham Biosciences; Piscataway, NJ, USA) generated ladders that were resolved on an ABI Prism 377 Sequencer (Applied Biosystems; Foster City, CA, USA), and were the basis for the genomes sequenced by the authors. The development and application of "Next Generation" instrumentation and protocols are creating opportunities to obtain much greater numbers of HAdV genomes, allowing for detailed examinations of HAdV, including pathoepidemiology and molecular evolution. Given the number and flux of new technologies, and the lack of use with these systems in AdV genomic studies, we will not discuss their use here. It is noted that each technology has its own particular strengths and weaknesses, and these will need to be understood with regards to AdV genomics.

In general, regardless of the technology, there are two common considerations for HAdV genome sequencing. One is accuracy and quality control. For the Sanger-based method, a minimum three-fold coverage across the entire genome is required, with problematic regions and potentially relevant SNPs to be covered by additional re-sequencing. A "2+1" strategy, that of obtaining sequences comprising both strands, allows high confidence in the genome data. Another method to ensure high quality sequence data is to analyze the proteome with computational means. For all sequencing methods, sequence assembly, particularly from "short read" ladders of some protocols, and the quality control of these genome sequences may be augmented by the annotation process. In other words, there are essential genes, such as hexon, penton, fiber, and genome features, such as the inverted terminal repeats (ITRs), that must be present and within a certain conserved range of sequences. Having an

annotation allows unreliable data to be re-sequenced for resolution, allowing a ten-fold increase in sequence accuracy.

Second, there are regions that may be difficult to sequence, given a particular technology. One example for the Sanger methodology is the ends of the AdV genome. Both ends of this linear HAdV molecule contain inverted terminal repeats (ITRs). These should complement each other, and serve as a quality control check. One end may be sequenced by DNA polymerase "running off the template"; the other end is problematic, as DNA polymerase requires a template to initiate. An additional complication is that HAdV genomes contain a covalently linked protein that is attached to the end. One solution to circumvent these obstacles is to use a "rapid amplification of cDNA ends" kit (5'/3' RACE; Roche Diagnostics Corp.), with modifications [22].

2.2. Bioinformatics

2.2.1. Genome analysis

Computational analyses of the genome involve examining the nucleotide sequence of the genome and the protein sequences of the proteome. There are many software tools available, both as commercial packages and as public resources on the Internet. One site housing the URLs for many public resource software tools is http://molbiol-tools.ca/. Specific tools used in our studies are described here and detailed in Table 1.

For DNA sequencing ladders assembly, DNA Sequencher (Gene Codes Corp.; Ann Arbor, MI, USA) was used for the completion of the seventeen genomes noted earlier as submitted by the authors. This software can also be used to align sequences and, importantly, manually move them, allowing visual characterization of recombinant or deleted sequences.

As noted above, once a consensus contig genome is produced, a quick examination of certain "landmarks" is useful, for example, the ITRs, pTP and Pol genes are difficult to sequence. A first pass annotation, to be described later, is recommended as a sequence quality control step before additional effort and excitement are expended, or as admonished by a biochemist as one of the "Ten commandments of enzymology": "Don't waste clean thinking on dirty enzyme" [67].

The attributes of the genome include GC content, with the percent GC diagnostic of species: A (47%), B1 (51%), B2 (49%), C (55%), D (57%), E (57%), F (51%) and G (55%). Genome lengths are not indicative of species and range from 34,125 bases (HAdV-A12) to 36,015 (HAdV-E4). The percent identity of a new genome may be determined relative to sequenced genomes; for example, species A is 58% identical to species C. The genome nucleotide sequence can be examined for repeats using PipMaker. This software uses a BlastZ algorithm to compute the local alignments of pairs of genomes and produces dot plots that give an indication of the similarity of the two genomes, as well as highlights any genome rearrangements [68]; the highest similarity scoring fragments will align on a diagonal. On the same site, zPicture gives another version of this genome identity analysis.

Multiple whole genome alignments, of genomes the size of HAdV, can be made using MAVID (http://baboon.math.berkeley.edu/mavid) [69]. MAVID, in turn, produces alignment outputs that may be ported into phylogeny tree analysis algorithm. For our studies, neighbor-joining trees [70] are constructed using MEGA4 [71]. Genome recombination events can be found in the genome alignments. Recombination is a contributing factor in the evolution of HAdV [14,51,21,22], as noted

for driving serotype evolution based on serum neutralization studies [29]. Sequence recombination can be detected whole and partial genome alignments, using Bootscan and SimPlot [72], with other comparable software also available [73].

Although seemingly anachronistic, restriction enzyme (RE) pattern analysis is still very useful, particularly to understand the context of the genomes relative to previously reported RE patterns of isolates reported in the literature. It should also be noted that as a "whole genome scan" tool, *i.e.*, a "genotyping" tool, RE patterns are effective for a rapid visual overview and comparison of the nucleotide genomes. The pDRAW32 software is one example for this in silico RE analysis. Also, the availability of genome data allows unlimited RE patterns and resolution of "faint" and multiple bands, and allows much better resolution than gel-based and photograph-based gel data in the literature.

2.2.2. Proteome analysis

A full-length annotation of coding and non-coding sequences completes the presentation of the genome sequence and extraction of information from the nucleotide string. In the past, HAdV and simian adenovirus (SAdV) sequences deposited in GenBank were incompletely annotated, with only a minimal annotation associated, particularly if submitted for patent purposes. We have developed a beta version of an automated genome annotator for our studies. This gives a "first pass" annotation that is suitable for assessing sequencing data quality. Refinement of the annotation manually as well as the examination of genome differences can be done using a genome viewer such as Artemis [74].

The proteome may be examined computationally, using percent identity comparisons of the nucleotide sequence and the amino acid percent identities of the proteins. These are manually calculated using the EMBOSS package [75]; more recently a beta version of an automated tool allows the same calculations. This provides an independent view to any recombination events.

Individual proteins and genome landmarks, e.g., ITRs, may be analyzed phylogenetically, using CLUSTAL for multiple sequence alignments (MSA) [76] and porting into phylogeny tree analysis software. One example of the relevance of this approach is a report showing the zoonotic origin of HAdV-E4 from chimpanzees [57]. An implication of the proteome analysis is a suggestion that the use of chimpanzee AdVs as alternatives in gene delivery in order to bypass pre-existing immune response may not be advisable, or should be done with caution.

2.2.3. Informatics support

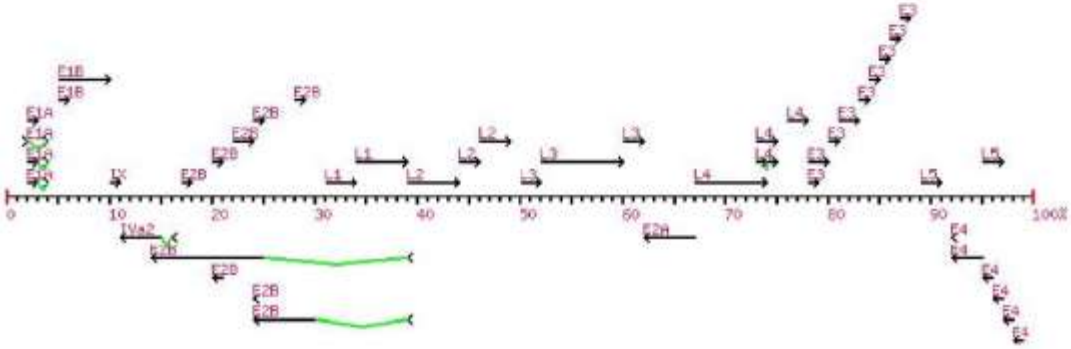
The Internet provides opportunities for worldwide interactions and collaborations, and for community-based resources. An "AdenovirusWiki" has been developed as an open resource for adenovirus research. In addition to the software tools available on the Internet, several local tools are also available, as beta versions: automated genome annotation; proteome percent identity analysis; and gene mapping tool (Figure 1); (www.irgolf.com/genemapv2). These provide for a pipeline to take a genome nucleotide sequence through analyses to produce genome annotations, proteome identification and analyses and a presentation of the coding sequences on a genome schematic.

A local tool developed in the authors' research group is "Virus Genome Annotation Tool" (VGAT). This is a beta version that is publicly available, with the caveat that it is a test version, at http://binf.gmu.edu/zenith/tool/lghmms.php. VGAT uses "Hidden Markov Models" (HMMs) to

annotate virus genomes. Currently, this software tool has been trained to annotate members of the HAdV-D species. It can be expanded to include members of other HAdV species, and the ability to add user defined training sets will be added.

A tool for automatically comparing the protein percent similarities in proteomes relative to their homologs in other proteomes is available in a beta form. The protein alignments and percent identities are calculated using a BioJava implementation [77] of a Needleman and Wunsch algorithm. When completed, the tool will be publicly available via the Internet; currently this tool is available upon request.

Figure 1. Automated gene-mapping tool. A field strain of HAdV-B7 (accession number AY601635; strain designation #NHRC 7151) has been sequenced and annotated. Its coding sequences are displayed using a gene-mapping tool using a derived annotation table.



2.2.4. Bioinformatics Tools Summary

Again, all of the computational tools noted for our studies are summarized in Table 1. As mentioned in the text, two are beta versions and need to be optimized. There are additional and equivalent software tools available over the internet, including multiple independent tools for similar analyses. A caveat is that some may be limited to certain computer platforms and need to be compiled. In some cases, the original contributors may no longer support some tools (orphans); however, some tools may be very useful and still supported, albeit at a different URL due to the contributor changing physical addresses. These may be found by 'googling' the tool name and/or the author on the Internet to locate the tool.

Table 1. Summary of Bioinformatics Tools.

Tool	Purpose	Availability
Sequencher	sequence assembly	Commercial
PipMaker	genome sequence analysis	http://pipmaker.bx.psu.edu/cgi-bin/pipmaker?basic
zPicture	genome sequence analysis	http://zpicture.dcode.org
MAVID	whole genome alignment	http://baboon.math.berkeley.edu/mavid

Table 1. *Cont.*

MEGA4	alignment viewer, phylogeny	http://www.megasoftware.net/
Simplot	recombination analysis	http://sray.med.som.jhmi.edu/SCRoftware/simplot/
pDRAW32	*in silico* restriction enzyme digest	http://www.acaclone.com/
Artemis	sequence viewer, annotation tool	http://www.sanger.ac.uk/Software/Artemis/
EMBOSS	sequence analysis	http://emboss.sourceforge.net/
Auto % Id *beta*	sequence % id	available upon request
Clustal	sequence alignment	http://www.clustal.org/
Adenovirus Wiki	repository of adenovirus data	http://www.binf.gmu.edu/wiki/index.php/Main_Page
Mapping Tool *beta*	create gene maps	www.irgolf.com/genemapv2
VGAT *beta*	automated virus genome annotation	http://binf.gmu.edu/zenith/tool/lghmms.php

3. Considerations of HAdVs for vaccine development and for vectors development for gene transfer and delivery

Limited molecular typing, e.g., PCR amplification coupled with DNA sequencing of certain targets, is a quick informative method to be applied in rationally designing adenoviral gene delivery vectors and in screening HAdVs and constructs as vector candidates [78-80]. Molecularly typing the outer coat proteins, hexon, penton and fiber, is important as they have critical roles in tissue tropism as well as in the host immune response to the virus. A caveat is that genome recombination may occur at other locations that may have subsequent and important consequences in the biotechnological application of the genome.

Although HAdV species C was initially the focus as vectors for biomedical and biotechnological applications, current interest range beyond this group. There have been many vectors based on other human, and even non-human, AdV serotypes that have developed as vectors for gene delivery and vaccine development. One review by Stone *et al*. [81] is an example. Primary literature citations include [2,4,7,53,62,82-89]. It is anticipated that genomics and bioinformatics resources will aid in these on-going work and development.

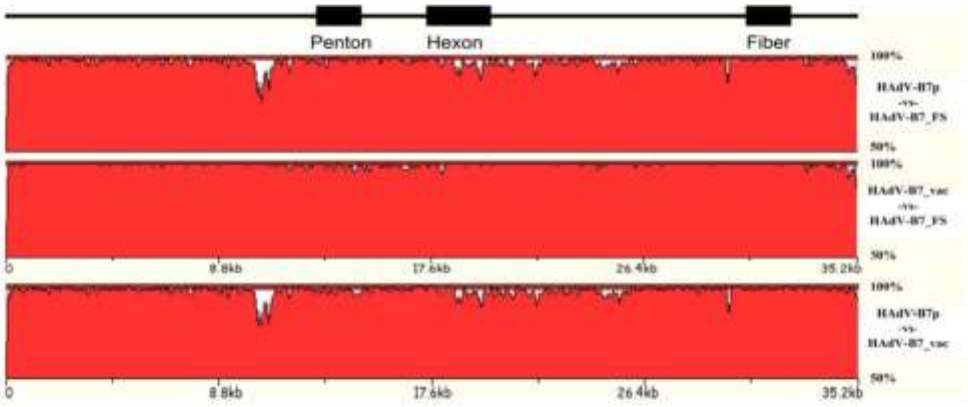
3.1. Natural variation of HAdV genomes

As noted earlier, recombination events are hallmarks of HAdV *in vitro*, and have now been documented in whole genome studies. The rate of recombination is not yet known. It may be that some species or types may be amenable to recombination based on sequence, e.g., hotspots, and biology, e.g., cell tropism and coinfection. Mutations as nucleotide changes, such as insertions and deletions (indels) and substitutions, are more common. However, given the fidelity of the DNA polymerase, the relevance of these genome changes remains to be elucidated. The question relevant to both HAdV vaccine development and gene delivery vector design is: How stable are these genomes?

Genomes from some isolates appear to be very stable, at least from the viewpoint of their antigenic epitopes. For example, the HAdV-B7 and E4 vaccines were highly effective in the U.S. military basic trainee population [8-10,18] for over twenty-five years [9]. This suggests, at least, a conservation of the epitopes for these two serotypes. A genome comparison of the prototype *versus* the "vaccine" (presumably the "then-circulating" and dominant) strains of both HAdV-B7 and HAdV-E4 showed

few genome changes, mainly indels and base substitutions [57-59]. These strains were of the 1950s and 1960s era. A pair of more recent HAdV-E4 field strains (accession number AY599837; strain designation #NHRC 3) and (AY599835; NHRC 42606), from two different outbreaks, and a recent HAdV-B7 field strain (AY601634; NHRC 1315) were sequenced, as 1990s isolates, to allow a comparison of their genomes. These showed similar limited mutations as well. Figure 2 displays whole genome comparisons of the three HAdV-B7 strains, across approximately forty-five years. Based on these whole genomes analyses and based on molecular typing, e.g., PCR amplification coupled with sequencing, of critical epitopes, it is likely the vaccines in production will be effective, as these genomes appear relatively stable.

Figure 2. HAdV-B7 genome comparisons. Three HAdV-B7 genomes, spanning ca. 1950s to 1990s, are analyzed using zPicture. Locations of the outer coat proteins, which are critical putative epitopes and cell tropism determinants, are arrayed for reference.



Proteome analysis allowed a detailed examination of all three HAdV-B7 genomes. Table 2 shows that despite the nucleotide differences, the protein percent identities are high, with the exception of the agnoprotein and the E3 7.7 kDa protein. The E3 difference may be important as those proteins may have roles in host immune response [90].

Similar genome and proteome data were obtained from analyses with HAdV-E4p, HAdV-E4_vac, HAdV-E4_FS1 and HAdV-E4_FS2 (data not shown). Two other prototype (ca. 1960s) and field strain (ca. 1990s) genomes have been paired and sequenced (or extracted from GenBank) as well: HAdV-B3 and HAdV-B3_FS (AY599836; NHRC 1276); and HAdV-C5 and HAdV-C5_FS (AY601635; NHRC 7151). Additionally, the HAdV-B3 genomes were compared with two field strains sequenced and described in China as well as a laboratory-circulating strain [91]. Genomes available for the HAdV-C5 analysis are even more interesting, as two prototype genomes were available from GenBank: 1) the original report, a composite presumably of several laboratory-circulating strains; and 2) an amplification of the original prototype from ATCC (Manassas, VA, USA) and now available as an "Adenovirus Reference Material" (ARM). Shown in Figure 3, these versions of the prototype were compared to the genome of a field strain, HAdV-C5_FS, which was isolated as one of a pair of

coinfecting HAdVs (manuscript submitted). The other coinfectant was HAdV-B21 and no signs of recombination in the HAdV-C5 genome were observed. Only four genome changes (one substitution and three indels) separated the 1998 field strain from the 1953 prototype (ARM). HAdV-C5_FS differs slightly more from the circulating laboratory strain (99.9%), suggesting laboratory passages allow some unselected mutations to accumulate.

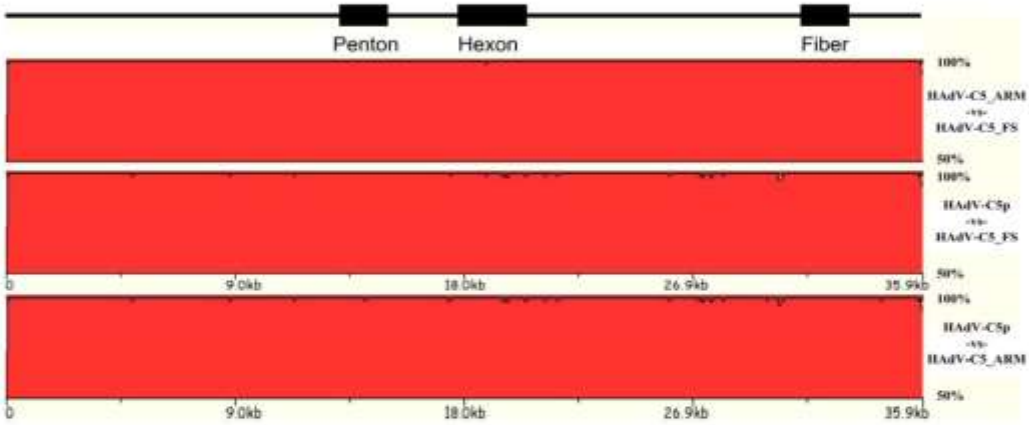
Table 2. Proteome Analysis of HAdV-B7. Percent identities of proteins amongst the reported three genomes of HAdV-B7 are calculated, relative to the HAdV-B7_FS proteome.

HAdV-B7_FS	HAdV-B7_vac	HAdV-B7p
E1A 28 kDa protein	98.9	99.6
E1A 32 kDa protein	98.3	99.6
E1A 6 kDa protein	98.3	100.0
E1B 20 kDa protein	100.0	98.3
E1B 55 kDa protein	100.0	99.0
IX protein	100.0	100.0
IVa2 protein	99.3	98.9
DNA polymerase	99.7	98.2
Hypothetical	99.1	98.1
agnoprotein	99.5	72.7
Hypothetical	100.0	68.4
Terminal protein	100.0	98.8
Hypothetical	99.2	95.5
Hypothetical	98.9	93.4
52 kDa protein	99.2	97.4
IIIa protein	99.1	99.7
penton base protein	98.7	99.3
VII protein	98.4	99.5
V protein	98.3	98.3
X protein	98.3	98.3
VI protein	96.4	96.8
hexon	99.8	97.0
protease	100.0	97.6
DBP	99.8	97.7
100 kDa protein	99.6	98.2
33 kDa protein	98.3	83.9
22 kDa protein	100.0	97.5
VIII protein	86.8	86.3
E3 12.1 kDa	100.0	100.0
E3 CR1-α	100.0	99.3
E3 glycoprotein	100.0	100.0
E3 CR1-β	95.5	93.3
E3 CR1-γ	99.5	98.4
E3 7.7 kDa protein	100.0	59.1
E3 RID-α	100.0	100.0
E3 RID-β	92.4	100.0
E3 14.7 kDa protein	100.0	99.3
U protein	100.0	98.1
fiber	99.7	98.8
E4 ORF 6/7 protein	100.0	100.0
E4 32 kDa protein	99.3	98.7
E4 ORF 4 protein	98.4	95.9

Table 2. *Cont.*

agnoprotein	99.4	97.0
E4 ORF 3 protein	99.1	99.1
E4 ORF 2 protein	96.9	98.4
E4 ORF 1 protein	98.4	96.0

Figure 3. HAdV-C5 genome comparison with zPicture. Genomes of a laboratory-circulating strain and a coinfectant field strain of HAdV-C5 are compared with the reference prototype (ARM). The field strain genome has ~100% sequence identity with the ARM strain, with four nucleotide differences across 45 years.



The apparent stability of these examples suggests that, in some cases, HAdV genomes are not as vulnerable to large-scale genome changes, such as recombination events. Accumulation of indels and base substitutions do occur, as would be expected, although in one case, HAdV-C5, it can be surprisingly few in number. All of these observations, including the highly effective nature of the HAdV-B7 and E4 vaccines earlier, imply that vaccines developed and vectors developed using HAdV genomes may be stable and useful for a period of time.

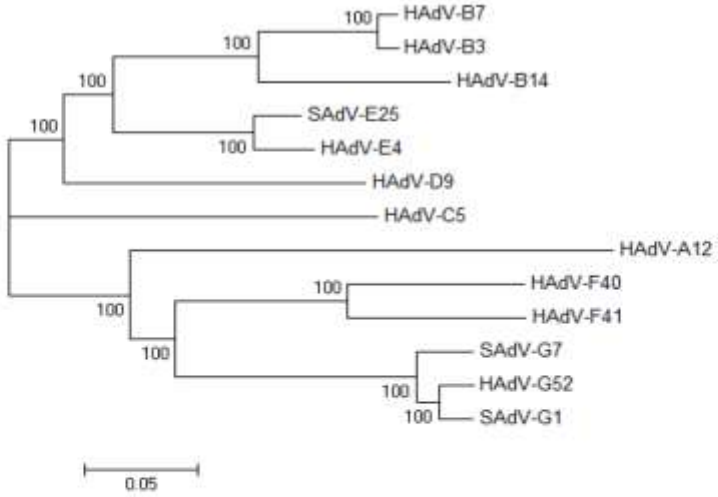
3.2. Natural variation of HAdV genomes: new types, new species, and vector candidate

Novel HAdV may be candidates in the quest for effective, appropriate (e.g., tissue and organ specific) and safe (e.g., asymptomatic and non-immunogenic) vectors. Genomics and bioinformatics have provided the identification and characterization of HADV-52, isolated from the stool of a patient with gastroenteritis [45]. As it was distinguished on the basis of genomics and bioinformatics rather than the traditional immunochemical techniques, it is referred to as "type" rather than the inappropriate "serotype". In addition, the case has been made for it as both a new type and a new species as well; the opposing view has been discussed in the literature [46]. There are additional computational data, derived from bioinformatic methods noted earlier, that were not reported in the original report which

provide strong support for this as well. HAdV-G52 shows a very high whole genome percent nucleotide identity with SAdV-G1, a simian (monkey) AdV, at 95.5%. This has been proposed as a member of species G. In contrast, the percent identity with SAdV-G7 is also high at 82.9%, another proposed species G member. The percent identities between the next phylogenetically closest, and also gastrointestinal, viruses HAdV-F40 and HAdV-F41 are much lower at approximately 69%. For reference, the percent identity between HAdV-F40 and HAdV-F41 is 85.8%, both members of species F. GC analysis shows a clustering of HAdV-G52 (55.1%), SAdV-G1 (55.2%) and SAdV-G7 (56.3%) as opposed to HAdV-F40 (51.2%) and HAdV-F41 (51.0%); again, GC contents seem to correlate with species grouping when surveyed across all of the sequenced genomes (data not shown).

Moreover, whole genome phylogenetic analysis (Figure 4) shows that HAdV-G52 subclades with SAdV-G1 and SAdV-G7. HAdV-F40 and HAdV-F41 forms a separate subclade. Proteome analysis, in the form of percent similarities, also shows closer relationships between the proposed "G" species members than with the other HAdV (data shown in Table 3). HAdV-G52 is missing the RL3 protein, which is unique to HAdV-F proteomes; this is also not in genomes of the other available genomes including the two simian ones grouped as species G. The RL3 protein is a 6.7 kDa protein that is encoded by E3 [92]. The function of RL3 is thought to be in directing glycoproteins to the endoplasmic reticulum [92].

Figure 4. Whole genome phylogenetic tree. HAdV-G52 subclades with SAdV-G1 and SAdV-G7. HAdV-F40 and HAdV-F41 subclade separately. Numbers presented are percent bootstrap replicates.



These results, taken with the original data, strongly suggest that HAdV-G52 is more related to SAdV-G1 and SAdV-G7 than to HAdV-F40 and HAdV-F41, and sufficiently different to merit designation as a new type and species. It would be inappropriate to refer to this strain as a "serotype" HAdV-G52 as no comprehensive serotyping data were provided to distinguish it from the accepted

ones; however, it is clearly different from the 52 established serotypes. These data also suggest the two monkey AdVs are likely members of species G and may suggest a zoonotic origin for HAdV-G52. Given the lower percent identity scores with the other HAdVs (data not shown), HAdV-G52 may be a strong candidate as a vector for gene delivery, possibly avoiding pre-existing immunity issues.

Table 3. Protein percent identities of species F and G, relative to HAdV-G52. To assess the relationships between proteomes of species G and F, protein percent identities were calculated. One, denoted as "*" was not found in HAdV-G52.

Adenovirus	Percent Identity, relative to HAdV-G52														
	E1A	E1B 55 kDa protein	IVa2	DNA Pol	pTP	L1 55 kDa protein	L2 penton	L3 hexon	E2A DBP	CR1-alpha1 (RL1)	CR1-beta1 (RL2)	RL3*	L5 fiber1	L5 fiber2	E4 34 kDa
SAdV-G1	92.3	99.2	99.1	98.6	99.5	100	99.2	92.3	94.6	97.7	97.0	-	82.6	98.6	100
SAdV-G7	38.6	90.3	98.9	97.6	98.8	99.7	93.3	90.0	94.8	-	-	-	59.3	72.1	98.3
HAdV-F40	47.1	68.2	87.7	78.5	82.4	82.6	86.1	84.7	62.3	44.3	35.7	+	37.8	52.7	62.3
HAdV-F41	42.8	68.5	89.5	79.9	84.4	83.8	87.1	87.8	66.1	44.3	34.8	+	38.5	53.4	61.2

3.3. Natural variation of AdV genomes: non-human primate AdV genomics and vector candidates

The persistence, infectivity and wide distribution of HAdV in general lead to concerns of pre-existing immunity, as characterized by seroprevalence [7]. This, in turn, leads to concerns with the use of HAdV as vectors. In an attempt to develop vectors that may be free of potential problems, alternative non-human AdVs may be appropriate substitutes as vector candidates, especially those from the great apes which presumably can infect human cells due to similarities.

A growing number of such AdVs are beginning to be isolated, characterized and examined for use as vectors. A recent contribution of 33 novel non-human primate genomes (30 ape and three macaque) has been reported and deposited in GenBank [93]. These are in addition to several monkey AdV genomes, sporadically deposited as simian AdV (SAdV) since 2004, and the five original chimpanzee AdV genomes, also noted as SAdVs, deposited into GenBank in 2004 [61,62]. The first chimpanzee AdVs were originally deposited at the American Type Culture Collection (ATCC) in the 1960s-70s, so these 33 additional genomes represent a recent, renewed and directed interest in novel non-human primate AdVs. Biotechnological applications, including vector applications, appear to drive the enthusiasm for the collection and characterization of these genomes. At least three groups are contributing to this seiche of monkey and great ape AdV genomes (noted collectively in the past as "simian"), as per several reports at the recent 9th International Adenovirus Meeting (Dobogókö, Hungary; April 2009). Given the wide diversity of genomes and this larger collection, it was suggested then by one of authors that it may be appropriate to standardize the nomenclature, to one that is also discriminatory and informative. Rather than classifying them all as "simian" AdVs (SAdVs; for example, [93]), subclassification into chimpanzee (ChAdV; for example, [82]), bonobo (BoAdV),

gorilla (GoAdV) and perhaps monkey (MoAdV), *etc*. may be more appropriate, especially for eventual "Big Picture" analyses of all genomes and for discussions of HAdV origins, molecular evolution, natural histories, taxonomy and virus reservoirs.

As a result of recent interests, the inventory of non-human primate AdV genomes is growing at a rapid pace, and the genomes are providing alternative biotechnology tools as well as providing resources for a more detailed glimpse into the biology, genomics and bioinformatics of HAdVs. The availability of these and other primate genomes allow more thorough computational analysis and finer resolutions of earlier observations. For example, discussed earlier were the zoonotic origins of HAdV-E4 and species E from the chimpanzee [94,95,57]. The recent 33 novel non-human primate genomes are parsed by genome analysis into HAdV species B, C and E [93], complementing the B and E species partitioning of the original SAdV-21 through SAdV-25, and confirming a close phylogenetic relationship with the HAdVs. As the recently described AdVs were collected as samples from substantial and persistent shedding in the stools of asymptomatic and apparently healthy primates, a comment (and caution) as to a zoonosis potential was noted. Noted also was the possibility and observation of intraspecies recombination in one of these primate AdV genomes [93], echoing recent reports of genome recombination in HAdV [14,51,21]. These two possibilities, zoonosis and recombination, may have relevance in understanding pre-existing human seroprevalence as well, as the earlier computational analysis of HAdV-E4 showed protein homologies to the chimpanzee AdVs (SAdV-21 through SAdV-25). Further bioinformatic analyses should be applied to these new genomes.

Chimpanzee AdVs have been developed into vectors for potential human applications recently and in the near past [82-84]. It is anticipated that both genomics and bioinformatics will play large roles in the further and continuing development and applications of these AdVs of human uses, especially within the context of the rational design of vectors for gene delivery.

4. Applications

Genomics and bioinformatics have provided a more detailed understanding and another dimension of these viruses through their genomes and proteomes. These are very useful in the context of vaccine development and also in the continuing biotechnological development of HAdV genomes as a vector for gene delivery. As an example, HAdV-B3 is discussed as a subject for both applications. This particular ARD infectious disease agent remains a global pathogen and is a public health problem, particularly in high-density populations [96]. Data suggest that a vaccine developed against a particular HAdV may be cost-effective and will be an effective prophylactic for an extended period of time, e.g., stable genome, similar to the original HAdV-B7 and HAdV-E4 vaccines. On the other hand, seroprevalence due to its circulation is a concern and limits its use as a vector [7].

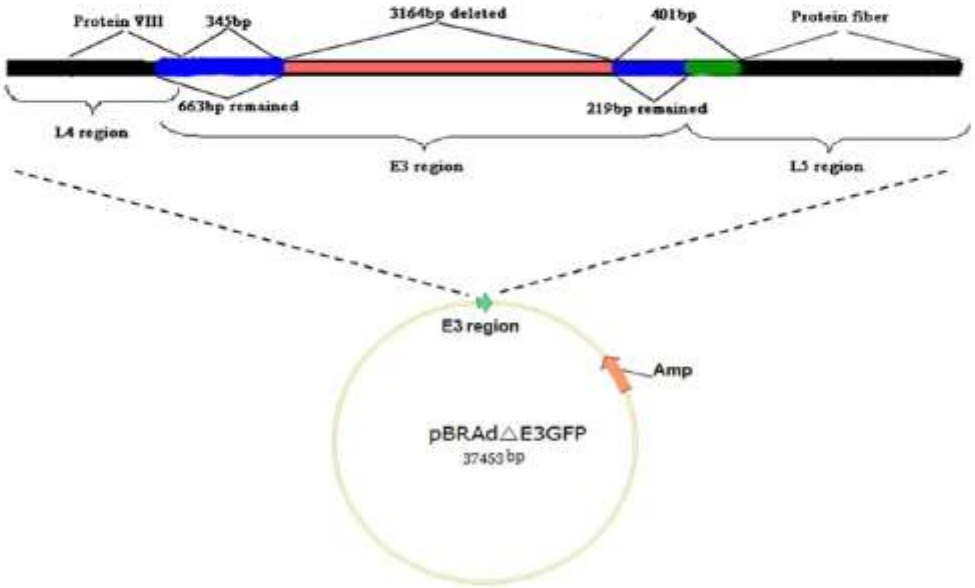
4.1. Applications: development of HAdV-B3 vaccine

HAdV-B3 remains an important human pathogen for ARD [13,96]. The serotype is considered highly virulent and has been associated with high morbidity, due to pharyngoconjunctival fever and residual lung damage, as well as mortality in children [97,98]. Currently there is no effective vaccine against HAdV-B3 infection. Therefore, it makes sense to have a safe, effective, readily available and inexpensive vaccine available against this virus in the countries with very dense and vulnerable

populations, especially in populations where the prevalence of HAdV-B3-specific neutralizing antibodies is very low. To this end, Zhang *et al.* have developed a replication-competent recombinant HAdV-B3 rAdΔE3GFP vector expressing eGFP as a vaccine candidate [85].

As shown in Figure 5, a recombinant virus was constructed by deleting a 3,164 nucleotide segment in the non-essential E3 region (nucleotides 27,737–30,900), yielding the rAdΔE3GFP genome [85]. A CMV-eGFP-SV40 expression cassette (1,616 nucleotides) was inserted into the E3 region by recombination. The left 663 bp and right 219 bp flanking E3 regions remained in place. In theory, a maximum size of 4,800 nucleotides (foreign gene) can be inserted into this E3-deletion vector. Mice immunized with the recombinant eGFP AdV by either intramuscular injection, intragastric or intranasal inoculation routes raised a significant antibody response to eGFP and to the wild-type HAdV-B3 GZ1 strain at the same time [96,85]. Alternatively, Li *et al.* constructed another replication-defective HAdV-B3 by the deletion of the entire E1 region for use also as a vaccine candidate [86]. Wild-type HAdV-B3 can be neutralized by the sera from the mice intramuscularly immunized with this recombinant virus [86].

Figure 5. Deletion map of the E3 region in the recombinant pBRAdΔE3GFP vector. The red region indicates the deletion in the HAdV-B3. Blue regions indicate remaining left and right regions of the E3 region. Black regions indicate VIII protein and the fiber, L4 and L5 regions respectively. The green indicates the gap between the E3 region and the L5 fiber region.



In the U.S., deployment of live enteric-coated oral vaccines against HAdV-B7 and HAdV-E4 was successful in removing both pathogens as agents for ARD for approximately twenty-five years [99]. Both vaccine strains presumably and selectively infected the lower intestinal tract, as administered in

these enteric-coated capsules, and stimulated the production of appropriate neutralizing and circulating antibodies. No adverse signs or symptoms of illness were associated with these two vaccines [8,100]. Similarly, the replication-competent HAdV-B3 may act as an effective and safe vaccine candidate as well when administered in enteric-coated oral capsules. No helper cells are needed and the necessary virus titers should be easy to obtain. Additionally and importantly, this vaccine genome could be used either as a bivalent or trivalent vaccine for the delivery of more viral antigens. Continuing work with the heterologous expression of the HAdV-B7 hexon in this vector is underway [85].

4.2. Applications: development of HAdV-B3 as a vector for gene delivery

Many gene therapy vectors, to date, used in human clinical applications are currently based on species C members: HAdV-C5 and HAdV-C2. However, the apparent pre-existing immunity against them, from previous infections [101,7], and the lack of the coxsackie and adenovirus receptor (CAR) or integrin expression in target cells may be of concern, for the safety of the patient and the efficacy of these species C-based AdV vectors [64,66], respectively.

As a consequence and an attempt at rational design of vectors, species B have been explored as alternatives, both as to increase the range of cells infected and to bypass pre-existing immunity [53-55,87,81,7]. For example, HAdV-B3 is reported to gain entry into cells through alternate receptors: CDX, CD46, CD80 or CD86 [65,102-106]. These are expressed in a multitude of cell types, including important gene therapy target cells that express either no or low levels of CAR. Therefore, HAdV-B3 may be an alternative to HAdV-C5-based gene-transfer vectors. A recombinant E1-deleted HAdV-B3 vector has been engineered on a bacterial artificial chromosome [88]. It is efficiently transduced into CD46-positive rodent and human cells. Another replication-defective HAdV-B3 vector was also constructed independently by molecular cloning [86]. These viruses were shown to replicate in an E1-complementing cell line. Other recombinant species B-based replication-defective vectors have also been developed: HAdV-B7 [89], HAdV-B11 [107,87] and HAdV-B35 [53,7]. However, seroprevalence due to presumably previous HAdV-B3 infections may be of concern. More studies are still needed for the effective and appropriate safe applications of these vectors for *in vivo* gene transfer.

5. Conclusions

Taking advantage of the recent and continuing improvements in high-throughput DNA sequencing technology and methodology, coupled with a myriad of bioinformatic tools developed for other organisms and areas of research, the HAdV researchers now have a wealth of genome and proteome resources to apply to understanding the comprehensive and integrated biology of the virus, including deeper and finer points concerning viral origins, putative reservoirs, molecular evolution, natural histories and taxonomy. These data have been applied to the development of vectors for gene delivery, either for gene therapy applications or for the delivery of antigens in vaccine development. Several of the early genomes sequenced were done so for this purpose. Research examining the natural variation of HAdVs, as well as the molecular evolution of their genomes, particularly in the context of emerging pathogens, has shown that the genomes are seemingly stable, that is accumulating indels and base substitutions commonly and recombination less commonly. These observations have relevance in understanding the biology and the pathoepidemiology of adenoviruses as a whole, and, importantly,

also have relevance in the development of vaccines against these pathogens as well as the biotechnological applications in vector development.

Acknowledgments

P.M. was supported by graduate student stipend and travel funds by George Mason University. We thank Gopy Naren Kammila for initiating, and Meher Pree Boorgula for continuing, the development of an AdenovirusWiki site and Zenith Maddipatla developing the beta HMM-based autoannotator software, all as partial fulfillment of their M.S. degree requirements. Elizabeth Liu provided the beta gene mapping software tool as partial fulfillment of her Ph.D. degree. James M. Clark, Shoaleh Dehghan and Shahnaz Sepehrband contributed to the preliminary partial analysis of several field strains, in partial fulfillment of their M.S. degree requirements. Anjan Purkayastha contributed to the preliminary partial analysis of some of these genomes. We also thank Clark Tibbetts for initial adenovirus discussions and for providing the opportunity to pursue this line of research. During the collection of the noted genome data set (particularly, the ones in the text with accession numbers) and preliminary data analyses, DS (2002-2004), JMC (2004), AP (2003-2005) 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.

References

1. Berg, M.; Difatta, J.; Hoiczyk, E.; Schlegel, R.; Ketner, G. Viable Adenovirus Vaccine Prototypes: High-Level Production of a Papillomavirus Capsid Antigen from the Major Late Transcriptional Unit. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 4590-4595.
2. Radosevic, K.; Wieland, C.W.; Rodriguez, A.; Weverling, G.J.; Mintardjo, R.; Gillissen, G.; Vogels, R.; Skeiky, Y.A.; Hone, D. M.; Sadoff, J.C.; van der Poll, T.; Havenga, M.; Goudsmit, J. Protective Immune Responses to a Recombinant Adenovirus Type 35 Tuberculosis Vaccine in Two Mouse Strains: CD4 and CD8 T-Cell Epitope Mapping and Role of Gamma Interferon. *Infect. Immun.* **2007**, *75*, 4105-4115.
3. Graham, F. L.; Prevec, L. Adenovirus-Based Expression Vectors and Recombinant Vaccines. *Biotechnology* **1992**, *20*, 363-390.
4. Shott, J.P.; McGrath, S.M.; Pau, M.G.; Custers, J.H.; Ophorst, O.; Demoitie, M.A.; Dubois, M.C.; Komisar, J.; Cobb, M.; Kester, K.E.; Dubois, P.; Cohen, J.; Goudsmit, J.; Heppner, D.G.; Stewart, V.A. Adenovirus 5 and 35 Vectors Expressing Plasmodium Falciparum Circumsporozoite Surface Protein Elicit Potent Antigen-Specific Cellular IFN-Gamma and Antibody Responses in Mice. *Vaccine* **2008**, *26*, 2818-2823.

5. Lowenstein, P.R.; Mandel, R.J.; Xiong, W.D.; Kroeger, K.; Castro, M.G. Immune Responses to Adenovirus and Adeno-Associated Vectors Used for Gene Therapy of Brain Diseases: The Role of Immunological Synapses in Understanding the Cell Biology of Neuroimmune Interactions. *Curr. Gene. Ther.* **2007**, *7*, 347-360.
6. Thacker, E.E.; Timares, L.; Matthews, Q.L. Strategies to Overcome Host Immunity to Adenovirus Vectors in Vaccine Development. *Expert Rev. Vaccines* **2009**, *8*, 761-777.
7. Vogels, R.; Zuijdgeest, D.; van Rijnsoever, R.; Hartkoorn, E.; Damen, I.; de Bethune, M.P.; Kostense, S.; Penders, G.; Helmus, N.; Koudstaal, W.; Cecchini, M.; Wetterwald, A.; Sprangers, M.; Lemckert, A.; Ophorst, O.; Koel, B.; van Meerendonk, M.; Quax, P.; Panitti, L.; Grimbergen, J.; Bout, A.; Goudsmit, J.; Havenga, M. Replication-Deficient Human Adenovirus Type 35 Vectors for Gene Transfer and Vaccination: Efficient Human Cell Infection and Bypass of Preexisting Adenovirus Immunity. *J. Virol.* **2003**, *77*, 8263-8271.
8. Couch, R.B.; Chanock, R.M.; Cate, T.R.; Lang, D.J.; Knight, V.; Huebner, R.J. Immunization with Types 4 and 7 Adenovirus by Selective Infection of the Intestinal Tract. *Am. Rev. Respir. Dis.* **1963**, *88*, SUPPL 394-403.
9. Gray, G.C.; Goswami, P.R.; Malasig, M.D.; Hawksworth, A.W.; Trump, D.H.; Ryan, M.A.; Schnurr, D.P. Adult Adenovirus Infections: Loss of Orphaned Vaccines Precipitates Military Respiratory Disease Epidemics. For the Adenovirus Surveillance Group. *Clin. Infect. Dis.* **2000**, *31*, 663-670.
10. Hilleman, M.R.; Stallones, R.A.; Gauld, R.L.; Warfield, M.S.; Anderson, S.A. Vaccination Against Acute Respiratory Illness of Adenovirus (RI-APC-ARD) Etiology. *Am. J. Public Health Nations Health* **1957**, *47*, 841-847.
11. Binn, L.N.; Sanchez, J.L.; Gaydos, J.C. Emergence of Adenovirus Type 14 in US Military Recruits--a New Challenge. *J. Infect. Dis.* **2007**, *196*, 1436-1437.
12. Blasiole, D.A.; Metzgar, D.; Daum, L.T.; Ryan, M.A.; Wu, J.; Wills, C.; Le, C.T.; Freed, N.E.; Hansen, C.J.; Gray, G.C.; Russell, K.L. Molecular Analysis of Adenovirus Isolates from Vaccinated and Unvaccinated Young Adults. *J. Clin. Microbiol.* **2004**, *42*, 1686-1693.
13. Echavarria, M. Adenovirus. In *Principles and Practice of Clinical Virology*, 6th ed.; Zuckerman, A.J., Banatvala, J.E., Pattison, J.R., Eds.; John Wiley & Sons, Ltd, Country? 2009; 463-488.
14. Ishiko, H.; Aoki, K. Spread of Epidemic Keratoconjunctivitis Due to a Novel Serotype of Human Adenovirus in Japan. *J. Clin. Microbiol.* **2009**, *47*, 2678-2679.
15. Louie, J.K.; Kajon, A.E.; Holodniy, M.; Guardia-LaBar, L.; Lee, B.; Petru, A.M.; Hacker, J.K.; Schnurr, D.P. Severe Pneumonia Due to Adenovirus Serotype 14: A New Respiratory Threat? *Clin. Infect. Dis.* **2008**, *46*, 421-425.
16. Metzgar, D.; Osuna, M.; Kajon, A.E.; Hawksworth, A.W.; Irvine, M.; Russell, K.L. Abrupt Emergence of Diverse Species B Adenoviruses at US Military Recruit Training Centers. *J. Infect. Dis.* **2007**, *196*, 1465-1473.
17. Zhu, Z.; Zhang, Y.; Xu, S.; Yu, P.; Tian, X.; Wang, L.; Liu, Z.; Tang, L.; Mao, N.; Ji, Y.; Li, C.; Yang, Z.; Wang, S.; Wang, J.; Li, D.; Xu, W. Outbreak of Acute Respiratory Disease in China Caused by B2 Species of Adenovirus Type 11. *J. Clin. Microbiol.* **2009**, *47*, 697-703.
18. Lyons, A.; Longfield, J.; Kuschner, R.; Straight, T.; Binn, L.; Seriwatana, J.; Reitstetter, R.; Froh, I.B.; Craft, D.; McNabb, K.; Russell, K.; Metzgar, D.; Liss, A.; Sun, X.; Towle, A.; Sun, W. A

Double-Blind, Placebo-Controlled Study of the Safety and Immunogenicity of Live, Oral Type 4 and Type 7 Adenovirus Vaccines in Adults. *Vaccine* **2008**, *26*, 2890-2898.

19. Echavarria, M. Adenoviruses in Immunocompromised Hosts. *Clin. Microbiol. Rev.* **2008**, *21*, 704-715.
20. Garnett, C.T.; Talekar, G.; Mahr, J.A.; Huang, W.; Zhang, Y.; Ornelles, D.A.; Gooding, L.R. Latent Species C Adenoviruses in Human Tonsil Tissues. *J. Virol.* **2009**, *83*, 2417-2428.
21. Walsh, M.P.; Chintakuntlawar, A.; Robinson, C.M.; Madisch, I.; Harrach, B.; Hudson, N.R.; Schnurr, D.; Heim, A.; Chodosh, J.; Seto, D.; Jones, M.S. Evidence of Molecular Evolution Driven by Recombination Events Influencing Tropism in a Novel Human Adenovirus That Causes Epidemic Keratoconjunctivitis. *PLoS ONE* **2009**, *4*, e5635.
22. Yang, Z.; Zhu, Z.; Tang, L.; Wang, L.; Tan, X.; Yu, P.; Zhang, Y.; Tian, X.; Wang, J.; Zhang, Y.; Li, D.; Xu, W. Genomic Analyses of Recombinant Adenovirus Type 11a in China. *J. Clin. Microbiol.* **2009**, *47*, 3082-3090.
23. De Jong, J.C.; Wermenbol, A.G.; Verweij-Uijterwaal, M.W.; Slaterus, K.W.; Wertheim-Van Dillen, P.; Van Doornum, G.J.; Khoo, S.H.; Hierholzer, J.C. Adenoviruses from Human Immunodeficiency Virus-Infected Individuals, Including Two Strains That Represent New Candidate Serotypes Ad50 and Ad51 of Species B1 and D, Respectively. *J. Clin. Microbiol.* **1999**, *37*, 3940-3945.
24. Echavarria, M.; Maldonado, D.; Elbert, G.; Videla, C.; Rappaport, R.; Carballal, G. Use of PCR to Demonstrate Presence of Adenovirus Species B, C, or F as Well as Coinfection with Two Adenovirus Species in Children with Flu-Like Symptoms. *J. Clin. Microbiol.* **2006**, *44*, 625-627.
25. Lin, B.; Wang, Z.; Vora, G.J.; Thornton, J.A.; Schnur, J.M.; Thach, D.C.; Blaney, K.M.; Ligler, A.G.; Malanoski, A.P.; Santiago, J.; Walter, E.A.; Agan, B.K.; Metzgar, D.; Seto, D.; Daum, L. T.; Kruzelock, R.; Rowley, R.K.; Hanson, E.H.; Tibbetts, C.; Stenger, D.A. Broad-Spectrum Respiratory Tract Pathogen Identification Using Resequencing DNA Microarrays. *Genome Res.* **2006**, *16*, 527-535.
26. Metzgar, D.; Osuna, M.; Yingst, S.; Rakha, M.; Earhart, K.; Elyan, D.; Esmat, H.; Saad, M.D.; Kajon, A.; Wu, J.; Gray, G.C.; Ryan, M.A.; Russell, K.L. PCR Analysis of Egyptian Respiratory Adenovirus Isolates, Including Identification of Species, Serotypes, and Coinfections. *J. Clin. Microbiol.* **2005**, *43*, 5743-5752.
27. Vora, G.J.; Lin, B.; Gratwick, K.; Meador, C.; Hansen, C.; Tibbetts, C.; Stenger, D.A.; Irvine, M.; Seto, D.; Purkayastha, A.; Freed, N.E.; Gibson, M.G.; Russell, K.; Metzgar, D. Co-Infections of Adenovirus Species in Previously Vaccinated Patients. *Emerg. Infect. Dis.* **2006**, *12*, 921-930.
28. Hierholzer, J.C.; Wigand, R.; Anderson, L.J.; Adrian, T.; Gold, J.W. Adenoviruses from Patients with AIDS: A Plethora of Serotypes and a Description of Five New Serotypes of Subgenus D (types 43-47). *J. Infect. Dis.* **1988**, *158*, 804-813.
29. Crawford-Miksza, L.K.; Schnurr, D.P. Adenovirus Serotype Evolution Is Driven by Illegitimate Recombination in the Hypervariable Regions of the Hexon Protein. *Virology* **1996**, *224*, 357-367.
30. Boursnell, M.; Mautner, V. *In Vitro* Construction of a Recombinant Adenovirus Ad2:Ad5. *Gene* **1981**, *13*, 311-317.
31. Boursnell, M.E.; Mautner, V. Recombination in Adenovirus: Crossover Sites in Intertypic Recombinants Are Located in Regions of Homology. *Virology* **1981**, *112*, 198-209.

32. Mautner, V.; Mackay, N. Recombination in Adenovirus: Analysis of Crossover Sites in Intertypic Overlap Recombinants. *Virology* **1984**, *139*, 43-52.
33. Williams, J.; Grodzicker, T.; Sharp, P.; Sambrook, J. Adenovirus Recombination: Physical Mapping of Crossover Events. *Cell* **1975**, *4*, 113-119.
34. Bernt, K.; Liang, M.; Ye, X.; Ni, S.; Li, Z.Y.; Ye, S.L.; Hu, F.; Lieber, A. A New Type of Adenovirus Vector That Utilizes Homologous Recombination to Achieve Tumor-Specific Replication. *J. Virol.* **2002**, *76*, 10994-11002.
35. Ling, L.E.; Manos, M.M.; Gluzman, Y. Sequence of the Junction in Adenovirus 2-SV40 Hybrids: Examples of Illegitimate Recombination. *Nucleic. Acids. Res.* **1982**, *10*, 8099-8112.
36. Adrian, T.; Bastian, B.; Benoist, W.; Hierholzer, J.C.; Wigand, R. Characterization of Adenovirus 15/H9 Intermediate Strains. *Intervirology* **1985**, *23*, 15-22.
37. Chmielewicz, B.; Benzler, J.; Pauli, G.; Krause, G.; Bergmann, F.; Schweiger, B. Respiratory Disease Caused by a Species B2 Adenovirus in a Military Camp in Turkey. *J. Med. Virol.* **2005**, *77*, 232-237.
38. Engelmann, I.; Madisch, I.; Pommer, H.; Heim, A. An Outbreak of Epidemic Keratoconjunctivitis Caused by a New Intermediate Adenovirus 22/H8 Identified by Molecular Typing. *Clin. Infect. Dis.* **2006**, *43*, e64-66.
39. Hierholzer, J.C.; Adrian, T.; Anderson, L.J.; Wigand, R.; Gold, J.W. Analysis of Antigenically Intermediate Strains of Subgenus B and D Adenoviruses from AIDS Patients. *Arch. Virol.* **1988**, *103*, 99-115.
40. Hierholzer, J.C.; Pumarola, A. Antigenic Characterization of Intermediate Adenovirus 14-11 Strains Associated with Upper Respiratory Illness in a Military Camp. *Infect. Immun.* **1976**, *13*, 354-359.
41. Ishiko, H.; Shimada, Y.; Konno, T.; Hayashi, A.; Ohguchi, T.; Tagawa, Y.; Aoki, K.; Ohno, S.; Yamazaki, S. Novel Human Adenovirus Causing Nosocomial Epidemic Keratoconjunctivitis. *J. Clin. Microbiol.* **2008**, *46*, 2002-2008.
42. Benkö, M.H.B.; Both, G.W.; Russell, W.C.; Adair, B.M.; Ádám, É.; de Jong, J.C.; Hess, M.; Johnson, M.; Kajon, A.; Kidd, A.H.; Lehmkuhl, H.D.; Li, Q.; Mautner, V.; Pring-Akerblom, P.; Wadell, G. Family Adenoviridae. In *The Eighth Report of the International Committee on Taxonomy of Viruses*. Fauquet C.M., Mayo M.A., Maniloff J., Desselberger U., Ball, L.A., Eds.; Academic Press: San Diego, CA, USA, 2005; pp. 213-228.
43. Davison, A. J.; Benko, M.; Harrach, B. Genetic Content and Evolution of Adenoviruses. *J. Gen. Virol.* **2003**, *84*, 2895-2908.
44. Russell, W.C. Adenoviruses: Update on Structure and Function. *J. Gen. Virol.* **2009**, *90*, 1-20.
45. Jones, M.S.; Harrach, B.; Ganac, R.D.; Gozum, M.M.; Dela Cruz, W.P.; Riedel, B.; Pan, C.; Delwart, E.L.; Schnurr, D.P. New Adenovirus Species Found in a Patient Presenting with Gastroenteritis. *J. Virol.* **2007**, *81*, 5978-5984.
46. de Jong, J.C.; Osterhaus, A.D.; Jones, M.S.; Harrach, B. Human Adenovirus Type 52: A Type 41 in Disguise? *J. Virol.* **2008**, *82*, 3809; author reply 3809-3810.
47. Roberts, R.J.; Akusjarvi, G.; Alestrom, P.; Gelinas, R.E.; Gingeras, T.R.; Sciaky, D.; Pettersson, U. A Consensus Sequence for the Adenovirus-2 Genome. In *Adenovirus DNA*. Doerfler W., Ed.; Martinus Nijhoff: Boston, MA, USA, 1986; pp. 1-51.

48. Chroboczek, J.; Bieber, F.; Jacrot, B. The Sequence of the Genome of Adenovirus Type 5 and Its Comparison with the Genome of Adenovirus Type 2. *Virology* **1992**, *186*, 280-285.
49. Sprengel, J.; Schmitz, B.; Heuss-Neitzel, D.; Doerfler, W. The Complete Nucleotide Sequence of the DNA of Human Adenovirus Type 12. *Curr. Top. Microbiol. Immunol.* **1995**, *199 (Pt 2)*, 189-274.
50. Davison, A.J.; Telford, E.A.; Watson, M.S.; McBride, K.; Mautner, V. The DNA Sequence of Adenovirus Type 40. *J. Mol. Biol.* **1993**, *234*, 1308-1316.
51. Robinson, C.M.; Rajaiya, J.; Walsh, M.P.; Seto, D.; Dyer, D.W.; Jones, M.S.; Chodosh, J. Computational Analysis of Human Adenovirus Type 22 Provides Evidence for Recombination Between Human Adenoviruses Species D in the Penton Base Gene. *J. Virol.* **2009**, *83*, 8980-8985.
52. Sugarman, B.J.; Hutchins, B.; McAllister, D.L.; Lu, F.; Thomas, B.K. The Complete Nucleotide Acid Sequence of the Adenovirus Type 5 Reference Material (ARM) Genome. *Bioprocessing J.* **2003**, *September/October*, 27-32.
53. Gao, W.; Robbins, P.D.; Gambotto, A. Human Adenovirus Type 35: Nucleotide Sequence and Vector Development. *Gene Ther.* **2003**, *10*, 1941-1949.
54. Mei, Y.F.; Skog, J.; Lindman, K.; Wadell, G. Comparative Analysis of the Genome Organization of Human Adenovirus 11, a Member of the Human Adenovirus Species B, and the Commonly Used Human Adenovirus 5 Vector, a Member of Species C. *J. Gen. Virol.* **2003**, *84*, 2061-2071.
55. Stone, D.; Furthmann, A.; Sandig, V.; Lieber, A. The Complete Nucleotide Sequence, Genome Organization, and Origin of Human Adenovirus Type 11. *Virology* **2003**, *309*, 152-165.
56. Lauer, K.P.; Llorente, I.; Blair, E.; Seto, J.; Krasnov, V.; Purkayastha, A.; Ditty, S.E.; Hadfield, T.L.; Buck, C.; Tibbetts, C.; Seto, D. Natural Variation Among Human Adenoviruses: Genome Sequence and Annotation of Human Adenovirus Serotype 1. *J. Gen. Virol.* **2004**, *85*, 2615-2625.
57. Purkayastha, A.; Ditty, S.E.; Su, J.; McGraw, J.; Hadfield, T.L.; Tibbetts, C.; Seto, D. Genomic and Bioinformatics Analysis of HAdV-4, a Human Adenovirus Causing Acute Respiratory Disease: Implications for Gene Therapy and Vaccine Vector Development. *J. Virol.* **2005**, *79*, 2559-2572.
58. Purkayastha, A.; Su, J.; Carlisle, S.; Tibbetts, C.; Seto, D. 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* **2005**, *332*, 114-129.
59. Purkayastha, A.; Su, J.; McGraw, J.; Ditty, S.E.; Hadfield, T.L.; Seto, J.; Russell, K.L.; Tibbetts, C.; Seto, D. 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.* **2005**, *43*, 3083-3094.
60. Seto, J.; Walsh, M.P.; Mahadevan, P.; Purkayastha, A.; Clark, J.M.; Tibbetts, C.; Seto, D. Genomic and Bioinformatics Analyses of HAdV-14p, Reference Strain of a Re-Emerging Respiratory Pathogen and Analysis of B1/B2. *Virus Res.* **2009**, *143*, 94-105.
61. Roy, S.; Gao, G.; Clawson, D.S.; Vandenberghe, L.H.; Farina, S.F.; Wilson, J.M. Complete Nucleotide Sequences and Genome Organization of Four Chimpanzee Adenoviruses. *Virology* **2004**, *324*, 361-372.

62. Roy, S.; Gao, G.; Lu, Y.; Zhou, X.; Lock, M.; Calcedo, R.; Wilson, J.M. Characterization of a Family of Chimpanzee Adenoviruses and Development of Molecular Clones for Gene Transfer Vectors. *Hum. Gene Ther.* **2004**, *15*, 519-530.
63. Henquell, C.; Boeuf, B.; Mirand, A.; Bacher, C.; Traore, O.; Dechelotte, P.; Labbe, A.; Bailly, J.L.; Peigue-Lafeuille, H. Fatal Adenovirus Infection in a Neonate and Transmission to Health-Care Workers. *J. Clin. Virol.* **2009**, *45*, 345-348.
64. Fechner, H.; Haack, A.; Wang, H.; Wang, X.; Eizema, K.; Pauschinger, M.; Schoemaker, R.; Veghel, R.; Houtsmuller, A.; Schultheiss, H.P.; Lamers, J.; Poller, W. Expression of Coxsackie Adenovirus Receptor and Alphav-Integrin Does Not Correlate with Adenovector Targeting *in Vivo* Indicating Anatomical Vector Barriers. *Gene Ther.* **1999**, *6*, 1520-1535.
65. Gaggar, A.; Shayakhmetov, D.M.; Lieber, A. CD46 Is a Cellular Receptor for Group B Adenoviruses. *Nat. Med.* **2003**, *9*, 1408-1412.
66. Tomko, R.P.; Xu, R.; Philipson, L. HCAR and MCAR: The Human and Mouse Cellular Receptors for Subgroup C Adenoviruses and Group B Coxsackieviruses. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 3352-3356.
67. Kornberg, A. Ten Commandments of Enzymology, Amended. *Trends Biochem. Sci.* **2003**, *28*, 515-517.
68. Schwartz, S.; Zhang, Z.; Frazer, K. A.; Smit, A.; Riemer, C.; Bouck, J.; Gibbs, R.; Hardison, R.; Miller, W. PipMaker--a Web Server for Aligning Two Genomic DNA Sequences. *Genome Res.* **2000**, *10*, 577-586.
69. Bray, N.; Pachter, L. MAVID: Constrained Ancestral Alignment of Multiple Sequences. *Genome Res.* **2004**, *14*, 693-699.
70. Saitou, N.; Nei, M. The Neighbor-Joining Method: A New Method for Reconstructing Phylogenetic Trees. *Mol. Biol. Evol.* **1987**, *4*, 406-425.
71. Tamura, K.; Dudley, J.; Nei, M.; Kumar, S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) Software Version 4.0. *Mol. Biol. Evol.* **2007**, *24*, 1596-1599.
72. Lole, K.S.; Bollinger, R.C.; Paranjape, R.S.; Gadkari, D.; Kulkarni, S.S.; Novak, N.G.; Ingersoll, R.; Sheppard, H.W.; Ray, S.C. Full-Length Human Immunodeficiency Virus Type 1 Genomes from Subtype C-Infected Seroconverters in India, with Evidence of Intersubtype Recombination. *J. Virol.* **1999**, *73*, 152-160.
73. Etherington, G.J.; Dicks, J.; Roberts, I.N. Recombination Analysis Tool (RAT): A Program for the High-Throughput Detection of Recombination. *Bioinformatics* **2005**, *21*, 278-281.
74. Berriman, M.; Rutherford, K. Viewing and Annotating Sequence Data with Artemis. *Brief. Bioinform.* **2003**, *4*, 124-132.
75. Rice, P.; Longden, I.; Bleasby, A. EMBOSS: The European Molecular Biology Open Software Suite. *Trends Genet.* **2000**, *16*, 276-277.
76. Thompson, J.D.; Gibson, T.J.; Plewniak, F.; Jeanmougin, F.; Higgins, D.G. The CLUSTAL_X Windows Interface: Flexible Strategies for Multiple Sequence Alignment Aided by Quality Analysis Tools. *Nucleic. Acids. Res.* **1997**, *25*, 4876-4882.
77. Holland, R.C.; Down, T.A.; Pocock, M.; Prlic, A.; Huen, D.; James, K.; Foisy, S.; Drager, A.; Yates, A.; Heuer, M.; Schreiber, M.J. BioJava: An Open-Source Framework for Bioinformatics. *Bioinformatics* **2008**, *24*, 2096-2097.

78. Darr, S.; Madisch, I.; Hofmayer, S.; Rehren, F.; Heim, A. Phylogeny and Primary Structure Analysis of Fiber Shafts of All Human Adenovirus Types for Rational Design of Adenoviral Gene Therapy Vectors. *J. Gen. Virol.* **2009**, *90*, 2849-2854.
79. Madisch, I.; Harste, G.; Pommer, H.; Heim, A. Phylogenetic Analysis of the Main Neutralization and Hemagglutination Determinants of All Human Adenovirus Prototypes as a Basis for Molecular Classification and Taxonomy. *J. Virol.* **2005**, *79*, 15265-15276.
80. Madisch, I.; Hofmayer, S.; Moritz, C.; Grintzalis, A.; Hainmueller, J.; Pring-Akerblom, P.; Heim, A. Phylogenetic Analysis and Structural Predictions of Human Adenovirus Penton Proteins as a Basis for Tissue-Specific Adenovirus Vector Design. *J. Virol.* **2007**, *81*, 8270-8281.
81. Stone, D.; di Paolo, N.C.; Lieber, A. Development of Group B Adenoviruses as Gene Transfer Vectors. *Biotechnol. Genet. Eng. Rev.* **2006**, *22*, 101-123.
82. Peruzzi, D.; Dharmapuri, S.; Cirillo, A.; Bruni, B.E.; Nicosia, A.; Cortese, R.; Colloca, S.; Ciliberto, G.; La Monica, N.; Aurisicchio, L. A Novel Chimpanzee Serotype-Based Adenoviral Vector as Delivery Tool for Cancer Vaccines. *Vaccine* **2009**, *27*, 1293-1300.
83. Farina, S.F.; Gao, G.; Xiang, Z.Q.; Rux, J.J.; Burnett, R.M.; Alvira, M.R.; Marsh, J.; Ertl, H.C.J.; Wilson, J.M. Replication-Defective Vector Based on a Chimpanzee Adenovirus. *J. Virol.* **2001**, *75*, 11603-11613.
84. Roy, S.; Zhi, Y.; Kobinger, G.P.; Figueredo, J.; Calcedo, R.; Miller, J.R.; Feldmann, H.; Wilson, J.M. Generation of an Adenoviral Vaccine Vector Based on Simian Adenovirus 21. *J. Gen. Virol.* **2006**, *87*, 2477-2485.
85. Zhang, Q.; Su, X.; Seto, D.; Zheng, B.J.; Tian, X.; Sheng, H.; Li, H.; Wang, Y.; Zhou, R. Construction and Characterization of a Replication-Competent Human Adenovirus Type 3-Based Vector as a Live-Vaccine Candidate and a Viral Delivery Vector. *Vaccine* **2009**, *27*, 1145-1153.
86. Li, H.; Zhou, R.; Chen, J.; Tian, X.; Zhang, Q.; Zeng, Q.; Gong, S. A Recombinant Replication-Defective Human Adenovirus Type 3: A Vaccine Candidate. *Vaccine* **2009**, *27*, 116-122.
87. Stone, D.; Ni, S.; Li, Z.Y.; Gaggar, A.; DiPaolo, N.; Feng, Q.; Sandig, V.; Lieber, A. Development and Assessment of Human Adenovirus Type 11 as a Gene Transfer Vector. *J. Virol.* **2005**, *79*, 5090-5104.
88. Sirena, D.; Ruzsics, Z.; Schaffner, W.; Greber, U.F.; Hemmi, S. The Nucleotide Sequence and a First Generation Gene Transfer Vector of Species B Human Adenovirus Serotype 3. *Virology* **2005**, *343*, 283-298.
89. Nan, X.; Peng, B.; Hahn, T.W.; Richardson, E.; Lizonova, A.; Kovesdi, I.; Robert-Guroff, M. Development of an Ad7 Cosmid System and Generation of an Ad7deltaE1deltaE3HIV(MN) Env/Rev Recombinant Virus. *Gene Ther.* **2003**, *10*, 326-336.
90. Kajon, A.E.; Xu, W.; Erdman, D.D. Sequence Polymorphism in the E3 7.7K ORF of Subspecies B1 Human Adenoviruses. *Virus Res.* **2005**, *107*, 11-19.
91. Mahadevan, P.; Seto, J.; Tibbetts, C.; Seto, D. Natural Variants of Human Adenovirus Type 3 Provide Evidence for Relative Genome Stability Across Time and Geographic Space. *Virology* **2009**, E-pub 23 November 2009.
92. Yeh, H.Y.; Pieniazek, N.; Pieniazek, D.; Luftig, R.B. Genetic Organization, Size, and Complete Sequence of Early Region 3 Genes of Human Adenovirus Type 41. *J. Virol.* **1996**, *70*, 2658-2663.

93. Roy, S.; Vandenberghe, L.H.; Kryazhimskiy, S.; Grant, R.; Calcedo, R.; Yuan, X.; Keough, M.; Sandhu, A.; Wang, Q.; Medina-Jaszek, C.A.; Plotkin, J.B.; Wilson, J.M. Isolation and Characterization of Adenoviruses Persistently Shed from the Gastrointestinal Tract of Non-Human Primates. *PLoS Pathog.* **2009**, *5*, e1000503.
94. Gruber, W.C.; Russell, D.J.; Tibbetts, C. Fiber Gene and Genomic Origin of Human Adenovirus Type 4. *Virology* **1993**, *196*, 603-611.
95. Li, Q.G.; Wadell, G. The Degree of Genetic Variability Among Adenovirus Type 4 Strains Isolated from Man and Chimpanzee. *Arch. Virol.* **1988**, *101*, 65-77.
96. Zhang, Q.; Su, X.; Gong, S.; Zeng, Q.; Zhu, B.; Wu, Z.; Peng, T.; Zhang, C.; Zhou, R. Comparative Genomic Analysis of Two Strains of Human Adenovirus Type 3 Isolated from Children with Acute Respiratory Infection in Southern China. *J. Gen. Virol.* **2006**, *87*, 1531-1541.
97. Fujimoto, T.; Hamamoto, I.; Taniguchi, K.; Chikahira, M.; Okabe, N. Molecular Epidemiology of Adenovirus Type 3 Detected from 1994 to 2006 in Hyogo Prefecture, Japan. *Jpn. J. Infect. Dis.* **2008**, *61*, 143-145.
98. Ryan, M.A.; Gray, G.C.; Smith, B.; McKeehan, J.A.; Hawksworth, A.W.; Malasig, M.D. Large Epidemic of Respiratory Illness Due to Adenovirus Types 7 and 3 in Healthy Young Adults. *Clin. Infect. Dis.* **2002**, *34*, 577-582.
99. Dudding, B.A.; Wagner, S.C.; Zeller, J.A.; Gmelich, J.T.; French, G.R.; Top, F.H. Fatal Pneumonia Associated with Adenovirus Type 7 in Three Military Trainees. *N. Eng. J. Med.* **1972**, *286*, 1289-1292.
100. Edmondson, W.P.; Purcell, R.H.; Gundelfinger, B.F.; Love, J.W.; Ludwig, W.; Chanock, R.M. Immunization by Selective Infection with Type 4 Adenovirus Grown in Human Diploid Tissue Culture. II. Specific Protective Effect Against Epidemic Disease. *JAMA.* **1966**, *195*, 453-459.
101. Su, X.B.; Ma, X.; Hong, Z.Y.; Wu, H.B.; Wang, Y.S.; Chen, L. [Epidemiological Study of Human Type 5 Adenovirus in Guangzhou Using Chemiluminescence for Neutralizing Antibody Assay]. *Nan Fang Yi Ke Da Xue Xue Bao (J. Southern Med. Univ.)* **2007**, *27*, 1323-1326.
102. Fleischli, C.; Sirena, D.; Lesage, G.; Havenga, M.J.; Cattaneo, R.; Greber, U.F.; Hemmi, S. Species B Adenovirus Serotypes 3, 7, 11 and 35 Share Similar Binding Sites on the Membrane Cofactor Protein CD46 Receptor. *J. Gen. Virol.* **2007**, *88*, 2925-2934.
103. Short, J.J.; Pereboev, A.V.; Kawakami, Y.; Vasu, C.; Holterman, M.J.; Curiel, D.T. Adenovirus Serotype 3 Utilizes CD80 (B7.1) and CD86 (B7.2) as Cellular Attachment Receptors. *Virology* **2004**, *322*, 349-359.
104. Short, J.J.; Vasu, C.; Holterman, M.J.; Curiel, D.T.; Pereboev, A. Members of Adenovirus Species B Utilize CD80 and CD86 as Cellular Attachment Receptors. *Virus Res.* **2006**, *122*, 144-153.
105. Sirena, D.; Lilienfeld, B.; Eisenhut, M.; Kalin, S.; Boucke, K.; Beerli, R.R.; Vogt, L.; Ruedl, C.; Bachmann, M.F.; Greber, U.F.; Hemmi, S. The Human Membrane Cofactor CD46 Is a Receptor for Species B Adenovirus Serotype 3. *J. Virol.* **2004**, *78*, 4454-4462.
106. Tuve, S.; Wang, H.; Ware, C.; Liu, Y.; Gaggar, A.; Bernt, K.; Shayakhmetov, D.; Li, Z.; Strauss, R.; Stone, D.; Lieber, A. A New Group B Adenovirus Receptor Is Expressed at High Levels on Human Stem and Tumor Cells. *J. Virol.* **2006**, *80*, 12109-12120.

107. Holterman, L.; Vogels, R.; van der Vlugt, R.; Sieuwerts, M.; Grimbergen, J.; Kaspers, J.; Geelen, E.; van der Helm, E.; Lemckert, A.; Gillissen, G.; Verhaagh, S.; Custers, J.; Zuijdgeest, D.; Berkhout, B.; Bakker, M.; Quax, P.; Goudsmit, J.; Havenga, M. Novel Replication-Incompetent Vector Derived from Adenovirus Type 11 (Ad11) for Vaccination and Gene Therapy: Low Seroprevalence and Non-Cross-Reactivity with Ad5. *J. Virol.* **2004**, *78*, 13207-13215.

Part 4

Computational analysis of human adenovirus serotype 18

ARTICLE IN PRESS

YVIRO-05747; No. of pages: 9; 4C: 5

Virology xxx (2010) xxx–xxx

Contents lists available at ScienceDirect

Virology

journal homepage: www.elsevier.com/locate/yviro

Computational analysis of human adenovirus serotype 18

Michael P. Walsh [a], Jason Seto [a], Damaris Tirado [b], James Chodosh [c], David Schnurr [d], Donald Seto [a,*], Morris S. Jones [b]

[a] Department of Bioinformatics and Computational Biology, George Mason University, 10900 University Blvd., MSN 5B3, Manassas, VA 20110, USA
[b] Clinical Investigation Facility, David Grant USAF Medical Center, Travis, CA 94535, USA
[c] Department of Ophthalmology, Howe Laboratory, Massachusetts Eye and Ear Infirmary, Harvard Medical School, Boston, MA 02114, USA
[d] Viral and Rickettsial Disease Laboratory, California Department of Public Health, Richmond, CA, USA

ARTICLE INFO

Article history:
Received 26 January 2010
Returned to author for revision 1 March 2010
Accepted 13 May 2010
Available online xxxx

Keywords:
Human adenovirus
Molecular evolution
Species A
Oncogenesis

ABSTRACT

The genome of the sole remaining unsequenced member of species A, human adenovirus type 18 (HAdV-A18), has been sequenced and analyzed. Members of species A are implicated as gastrointestinal pathogens and were shown to be tumorigenic in rodents. These whole genome and in silico proteome data are important as references for reexamining and integrating earlier work and observations based on lower resolution techniques, such as restriction enzyme digestion patterns, particularly for hypotheses based on pre-genomics data. Additionally, the genome of HAdV-A18 will also serve as reference for current studies examining the molecular evolution and origins of human and simian adenoviruses, particularly genome recombination studies. Applications of this virus as a potential vector for gene delivery protocols may be practical as data accumulate on this and other adenovirus genomes.

Introduction

Human adenoviruses (HAdVs) are partitioned into seven species, based on characteristics including biochemical and genome properties and homologies, tissue tropism, and immunochemical examination of serum neutralizing and hemagglutinating epitopes that are found on the virion capsid proteins. Species A contains three members, HAdV-A12, -A18 and -A31, that generated earlier interest due to their oncogenic capabilities, with HAdV-A12 and -A18 shown to cause tumors in newborn hamsters (Huebner et al., 1962; Trentin et al., 1962) and mice (Yabe et al., 1964). The precise basis for this oncogenesis remains unknown; however, the HAdV-A12 E1A proteins have been shown to repress class I major histocompatibility complex expression in transformed rodent cells, and these E1A proteins modulate class I gene expression by similar mechanisms in both transformed rodent and human cells (Vasavada et al., 1986). Having the complete genomes of HAdV-A18 and -A31 (Hofmayer et al., 2009) allows additional detailed analyses by researchers.

HAdV-A18 was originally isolated from an anal swab of a child with Niemann–Pick syndrome and characterized in the 1950s (Rowe et al., 1956, 1958). Its role in human health is not well characterized; however, another species A member, HAdV-A31, isolated originally from the stool of a healthy child (Pereira et al., 1965), has been shown to be causative agents in gastrointestinal disease with the identification of subsequent field isolates (Adrian and Wigand, 1989; Adrian et al., 1987; Brown et al., 1984; Hammond et al., 1985; Kidd et al., 1982). Several other HAdV species are associated with gastroenteritis and are isolated from stool samples as well. Among these, the members of species F (de Jong et al., 1983; Uhnoo et al., 1984) are of particular importance because they are the genomes most closely similar to those from species A, as reported in the phylogenetic and percent identity data of this report. The association of HAdV-A31 with gastroenteritis, its genome identity with HAdV-A18, and the close relationship between the HAdV-F and HAdV-A species all suggest a potential pathogenic role as a gastroenteritis agent for HAdV-A18.

Reports of HAdV coinfections (Vora et al., 2006), coupled with the isolation of HAdV mixtures and HAdV types (De Jong et al., 1999) from immunocompromised individuals suggest the possibility of the host acting as a bioreactor, and this provides a molecular pathway for HAdV-type evolution and pathogen emergence by recombination. Novel types of HAdV, isolated as emergent pathogens causing highly contagious outbreaks, have been reported recently. These have been characterized using a genomics and bioinformatics approach to take advantage of the high-resolution methodologies (Ishiko and Aoki, 2009; Walsh et al., 2009; Yang et al., 2009). The pathogens include a respiratory, HAdV-B55 (Yang et al., 2009), three ocular pathogens, HAdV-D53 and -D54, and a reanalysis of HAdV-D22 (Ishiko and Aoki, 2009; Robinson et al., 2009; Walsh et al., 2009). Reexamination of several prototype strains of HAdV respiratory pathogens archived at the American Type Culture Collection (ATCC; Manassas, VA) by genome sequencing and bioinformatic analysis shows a role of recombination in their genesis (HAdV-E4,

* Corresponding author. Fax: +1 703 993 8401.
E-mail addresses: mwalsh5@gmu.edu (M.P. Walsh), jseto@gmu.edu (J. Seto), Damaris.Tirado@travis.af.mil (D. Tirado), James_Chodosh@meei.harvard.edu (J. Chodosh), dseto@gmu.edu (D. Seto), Morris.Jones@travis.af.mil (M.S. Jones).

0042-6822/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.virol.2010.05.013

Please cite this article as: Walsh, M.P., et al., Computational analysis of human adenovirus serotype 18, Virology (2010), doi:10.1016/j.virol.2010.05.013

manuscript in preparation; HAdV-C6, manuscript in preparation; HAdV-B16, manuscript in preparation). However, currently there are limited numbers of whole HAdV genome sequences available in GenBank, handicapping a full comprehensive genomic analysis, particularly of recombination events. In several recombination analysis profiles of HAdV-A18, a suggestion of recombination is present, but the data are inconclusive perhaps lacking the appropriate genome sequence(s) for comparison. An example of this problem is demonstrated in the recently characterized emergent keratoconjuntivitis pathogen, HAdV-D53, in which a span of recombinant sequences showed similarity to an unsequenced and/or unknown HAdV genome. This reflects also the recent applications of these tools to the HAdV genomes, as well as the limited number of genomes archived.

To support and enhance continuing computational and comparative studies on the molecular evolution and pathoepidemiology of HAdV, the genome and analysis of HAdV-A18 is presented. Reexamining the oncogenesis data reported for species A genomes in the context of these newer genome data will allow further understanding of the biology of these rarely reported HAdVs.

Results and discussion

Genome

There are three members of the HAdV-A species, originally differentiated by serum neutralization and clustered with respect to biological and genome-derived similarities. All are now sequenced, annotated, and analyzed. SmaI in silico restriction enzyme digest patterns are consistent with previously reported data, with additional small-sized bands visible (data not shown; Wadell et al., 1980). The genome of HAdV-A18 has a length of 34,177 nucleotides compared to 34,125 for HAdV-A12 and 33,763 for HAdV-A31. The GC content, of 46.5%, is consistent with species A, even though the three genomes in this species provide a small sample size for such an analysis. Percent GC is a metric for species identification. This and the percent of identity of HAdV-A18 relative to representative HAdVs are noted in Table 1.

An overview of other HAdV species, particularly those with more members, validates grouping by this metric (data not shown). Species A genomes have been noted as being the most diverse within species groupings than the other species. Still, HAdV-A18 shows the highest genome nucleotide identity to the other HAdV-A members (83% to HAdV-A12 and 80.5% to HAdV-A31) than to other HAdV prototypes. For example, lower levels of genome identities are observed in comparison with genomes from other species with HAdV-F40 and -F41 having the next highest values of identity to HAdV-A18 (61.1%).

Table 1
GC content and percent identity of HAdV-A18, relative to representative HAdVs.

HAdV	Accession no.	% Identity	% GC
HAdV-A18	GU191019	100.0	46.5
HAdV-A12	AC_000005	83.0	46.5
HAdV-A31	In press	80.5	46.4
HAdV-B3	AY599834	60.3	51.1
HAdV-B7	AY594255	60.5	51.0
HAdV-B11	AY163756	60.6	48.9
HAdV-B14	AY803294	60.6	48.1
HAdV-C2	AC_000007	58.7	55.2
HAdV-C5	AC_000008	58.7	55.2
HAdV-D9	AJ854486	57.0	57.1
HAdV-D53	FJ169625	56.6	56.2
HAdV-E4	AY594253	58.3	57.7
SAdV-E25	AC_000011	58.5	59.8
HAdV-F40	NC_001454	61.1	51.2
HAdV-F41	DQ315364	61.1	51.0
SAdV-G1	NC_006879	58.8	55.2
HAdV-G52	DQ923122	58.4	55.1

Species A members have been implicated in tumor formation in rodents, initially observed for HAdV-A12 (Huebner et al., 1962). It has also been observed that transformation of cells in vitro by HAdV-A12 requires both E1A and E1B genes based on restriction enzyme digestion products (Graham et al., 1974). These low-resolution studies though "state-of-the-art at the time, i.e., using restriction enzyme fragments, can be reinterpreted given the whole genome data. HAdV-A12 was one of the first HAdV genomes sequenced, archived as X73487 (Sprengel et al., 1994). Its genome has been reannotated as AC_000006 and has been useful as a reference genome for ongoing HAdV research. The two remaining species A members were recently sequenced and shown to contain sequences in common with those from HAdV-A12 but are still distinctly different. Significantly, the sequences that constitute the "transforming segment", originally defined as a restriction enzyme fragment comprising the left 7.2% of the genome, have a high level of similarity with each other, ca. 80%. This similarity is not present in other HAdVs, for example in neither HAdV-B7 nor in HAdV-C2 and -C5 (Fujinaga et al., 1977; Sawada et al., 1979), which were discussed in the literature as not hybridizing to the HAdV-A12 counterpart; these show less than 61% identity (61%, 58.66%, and 59.8%, respectively).

Analysis of selected genes

Selected and presumably important and differentiating nucleotide coding sequences from HAdV-A18 were chosen for further analysis and comparison to homologs from other HAdV species from the preliminary computational analysis of all coding sequences. These selected coding sequences include E1A, DNA polymerase (Pol), penton, hexon, fiber, and E4 34 kDa. The penton, hexon, and fiber genes of HAdV are of particular importance because they are outer coat proteins, which play a significant role in the tropism of the virus, as well as in the definition and identification of these prototypes based on serum neutralization and hemagglutination epitopes. E1A, Pol, and E4 34 kDa were added to this analysis to provide a more complete picture of HAdV-A18, across the genome from its proximal to distal ends. The Pol sequence is important as a conserved essential gene, while the E1A protein has been implicated as a determinant of the tumorigenic properties of these HAdVs. Whole genome sequences are included in the analysis as a further reference.

Alignments and percent identities

Multiple sequence alignments provide a more detailed understanding of the genes relative to other HAdVs. The nucleotide coding sequences of HAdV-A18 selected genes were analyzed with respect to percent identities and are presented in comparison with homologs found in other HAdVs (Fig. 1).

The percent identity values for the selected coding sequences follow a pattern similar to that of the identity values noted (Table 1) for the HAdV whole genome. HAdV-A18 shows the highest identity relationship to the other HAdV-A serotypes, with lower identity values observed for other species. HAdV-A18 shows the highest identity to HAdV-A12 in most of the coding sequences presented. However, the hexon and E4 34 kDa genes of HAdV-A18 are more similar to homologs in HAdV-A31. This type of pattern suggests the possibility of a recombination event in one or both of these genes. The genome of HAdV-A18 was examined for recombination events (data described later in this report) and none were found.

Phylogenetic relationships of selected genes

Phylogenetic trees were used to explore the relationship of the coding sequences of HAdV-A18 to homologs in other HAdV species (Fig. 2). Members of the same species form distinct clades in all of the trees. In every tree except the hexon, members of the HAdV-A species

Please cite this article as: Walsh, M.P., et al., Computational analysis of human adenovirus serotype 18, Virology (2010), doi:10.1016/j.virol.2010.05.013

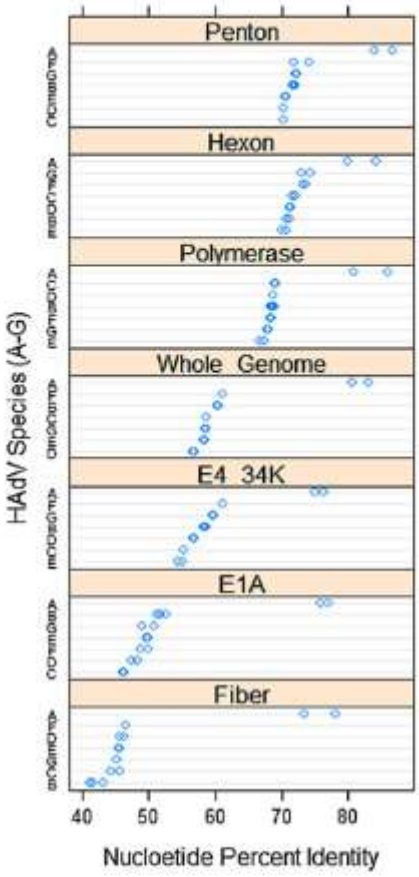


Fig. 1. Dot plot analysis of percent identity of selected genes. The percent identities of selected genes of HAdV-A18, to homologs in other species, are presented in the form of a dot plot. Two members of each species (A, B1, B2, C, D, E, F, and G) were chosen for the analysis to provide range and completeness. The resulting values were grouped by gene and sorted in descending order according to the median percent identity values of the genes. Values in each gene panel are grouped by species and sorted according to descending median percent identity. The chart was constructed using the R statistical computing environment (http://www.R-project.org/).

form a subclade in the same pattern with HAdV-A12 and -A31 pairing together, and HAdV-A18 pairing alone within the species A clade. However, in the hexon tree HAdV-A31 subclades with HAdV-A18 rather than HAdV-A12. This change in pattern among the phylogenetic trees is suggestive of a possible recombination event between the hexon sequences of HAdV-A18 and -A31.

Hexon recombination analysis

Patterns observed in percent identity values and phylogenetic trees for the coding sequences of HAdV-A18 suggest a possible recombination event with the hexon of HAdV-A31. To explore this observation further, analyses with Bootscan and Simplot (Lole et al., 1999) software were conducted with the hexon of HAdV-A18 (Fig. 3). Bootscan data show a possible recombination event. However, the Bootscan results are not as clear-cut and conclusive as those used to support previously reported recombination events (Robinson et al., 2009; Walsh et al., 2009). For example, in the previous cases demonstrating recombination, peaks in the Bootscans showed long stretches of values at 100%, indicating a strong close phylogenetic relationship between the recombinant sequences. The HAdV-A18 Bootscan shows no such pattern, which suggests a weaker phylogenetic relationship or perhaps a more ancient one with genetic drift contributing to and complicating this result. However, the Bootscan analysis does show two distinct sets of peaks that indicate that the first half of the hexon of HAdV-A18 is closely related to HAdV-A31 while the second half is more related to HAdV-A12. This pattern is common among the HAdV hexons that have been shown to contain a recombination event. In contrast, Simplot and percent identity data reveal that there is only 84.5% identity between the hexons of HAdV-A18 and -A31. Recombinations in other HAdV genomes from the literature have shown percent identity values of greater than 95% (Walsh et al., 2009). Therefore, the sum of the data from these analyses, e.g., Simplots, Bootscans, phylogenetic trees, and percent identity values, suggests that if there is a recombination event between the hexons of HAdV-A18 and -A31, it is either due to an ancient event or it is to another "yet-to-be-sequenced" HAdV genome.

Other sequences

VA RNAs

The HAdV genomes contain coding sequences for one or two "virus associated RNA" (VA RNA) sequences. These are transcribed by the host RNA polymerase III (Gonçalves and de Vries, 2006) and are noted to play a role in the inhibition of the host anti-viral functions (Mathews and Shenk, 1991). The number of VA RNAs varies according to species, with the genomes of HAdV species B2, C, D, and E encoding for two VA RNAs (VA RNA I and VA RNA II) and the members of species A, B1, F, and G encoding for a single VA RNA (VA RNA I).

Table 2 lists the lengths and percent identities of VA RNA I of HAdV-A18 relative to its homolog from other HAdVs. VA RNA I of HAdV-A18 has an identical length to its homolog in species A members and shows the highest identity scores to the VA RNA I of HAdV-A12. Among the non-HAdV-A species, HAdV-A18 shows the highest identity to the VA RNA of HAdV-G52 (65.3%).

Inverted terminal repeats

Both ends of the HAdV genomes contain "Inverted Terminal Repeat" (ITR) sequences (Dán et al., 2001; Hatfield and Hearing, 1993). The ITRs are conserved and are crucial for the replication of the virus as they contain the sequence motifs for DNA replication (Dán et al., 2001; Hatfield and Hearing, 1993; Temperley and Hay, 1992). Three important components of the ITR are the canonical core origin (Temperley and Hay, 1992), NF I binding site, and NF III binding site (Hatfield and Hearing, 1993). The core origin binds a pre-terminal protein (pTP)-Pol heterodimer, and the NF I and NF III binding sites interact with the host transcription factors, which are required for adenovirus replication.

Fig. 4 shows a schematic representation of an alignment of the ITRs of several HAdVs. The ITR of HAdV-A18 contains all of the requisite components: a core origin and both NF I/NF III binding sites. However, the core origin of HAdV-A18 contains mutations in its second and third bases, "TA" to "AT", not found elsewhere to date. This mutation causes the core origin of HAdV-A18 to be unique among sequenced HAdV ITRs, but its significance and effects are undetermined. It has been shown that the core origins of non-human adenoviruses are not as well conserved, in contrast to the HAdVs, and observations from the literature suggest that an "AT-rich" domain is all that is required for the core origin to be functional (Dán et al., 2001). This indicates that while this mutation is unique among the HAdVs, it may not have noticeable effect on the replication of the virus. Alternatively, it may

Please cite this article as: Walsh, M.P., et al., Computational analysis of human adenovirus serotype 18, Virology (2010), doi:10.1016/j.virol.2010.05.013

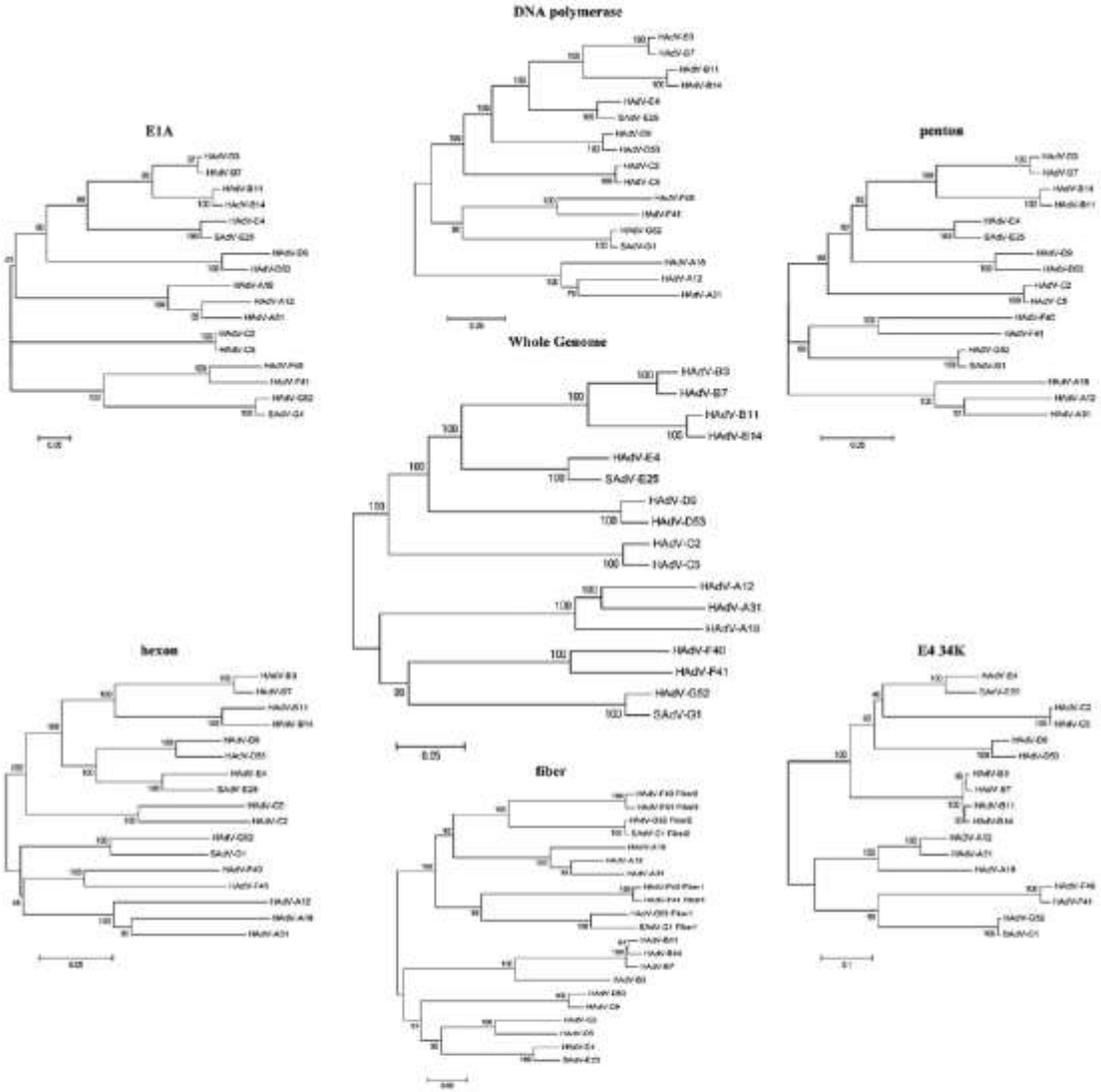


Fig. 2. Phylogenetic relationships. A neighbor-joining algorithm was used to determine the phylogenetic relationships of the E1A, DNA polymerase, penton, hexon, fiber, and E4 34 kDa nucleotide coding sequences. For reference, a whole genome analysis is presented. Numbers next to the branches of the trees indicate the percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates).

also indicate a closer relationship with and a more recent origin from the non-human adenoviruses.

Proteome analysis

Examining the in silico proteome of HAdV-A18 can provide information on the effects of the differences observed in nucleotide alignments discussed earlier. To this end, the proteins of HAdV-A18 were aligned with their homologs in HAdV-A12 and -A31. Percent identities, which were calculated based on these alignments, are displayed in Fig. 5.

Protein percent identities

The percent identity values show that the proteome of HAdV-A18 is most similar to that of HAdV-A12, Within the two proteomes, the highest identity values are in the E4 ORF 3 protein and the least similar values are in the CR1-α and CR1-β proteins.

Virtual 2D gel electrophoresis

To examine the differences in the proteomes of HAdV-A18, -A31, and -A12, all sequence-based predicted proteins from the genomes

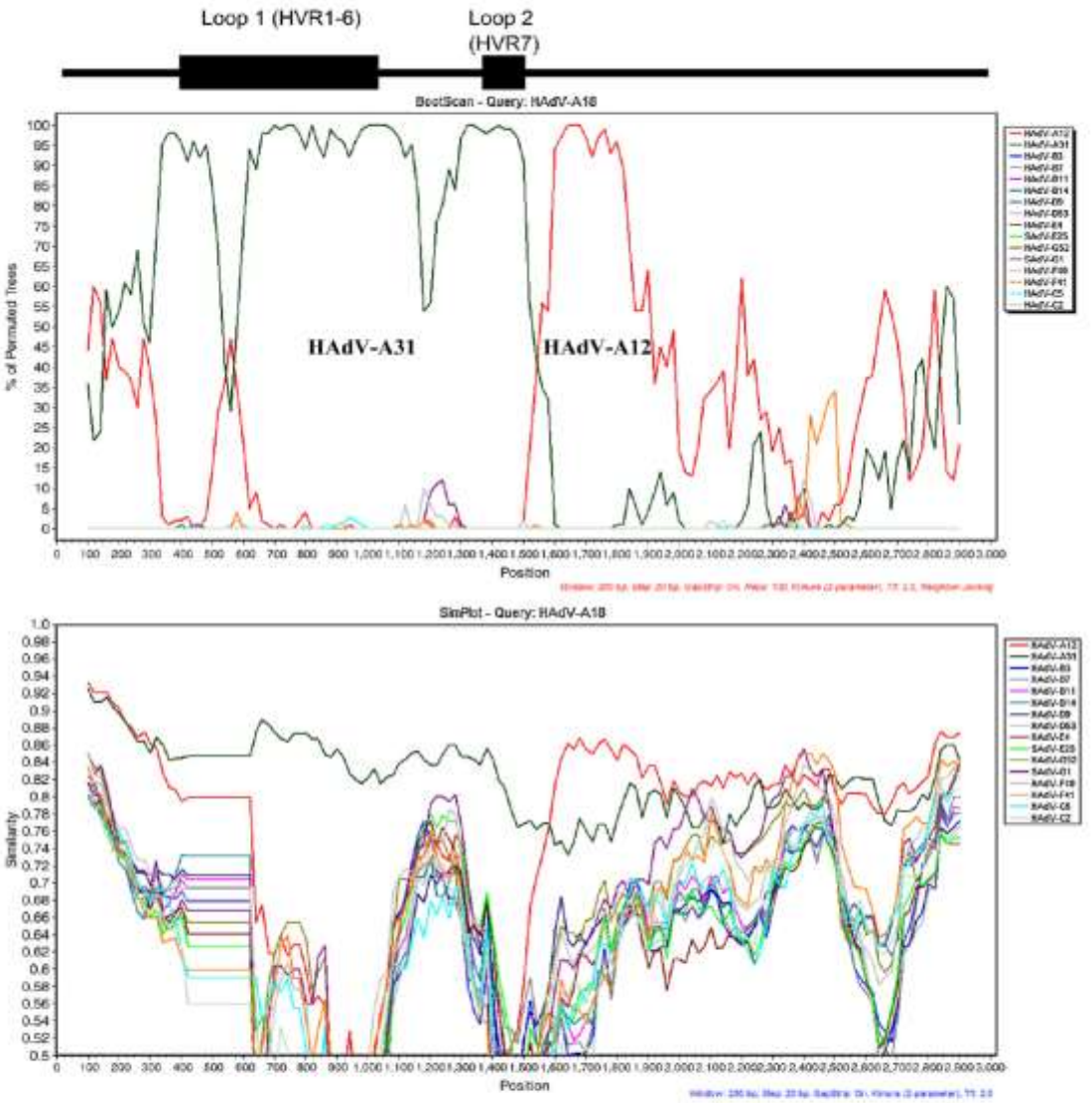


Fig. 3. Hexon recombination analysis. (Top) Bootscan and (bottom) Simplot recombination results are shown. The positions of the putative serotyping epitopes and the two hypervariable loops, of the hexon, are noted above the graphs as reference. The parameters are listed in red or blue lettering under each.

were examined via in silico 2D gel electrophoresis, which parses the proteins based on their computed isoelectric point (pI) and molecular weight. None of the homologs from the three genomes shows significant differences in molecular weight. However, some homologs display charge differences. Selected proteins are noted and presented in Fig. 6: the coat proteins that are putatively involved in tissue tropism and serve as serum neutralization and hemagglutination epitopes for laboratory serotyping; E1A protein, which is reported in the literature as the transforming or tumorigenic factor; DNA polymerase, a conserved essential protein; and several E3 transcript-derived proteins, which are reported to be involved in host immune system evasion. Of particular interest, the fiber and CR1-β proteins show the greatest discrepancy in charge density and molecular size, which is not unexpected given the low identity values of these proteins to one another. The fiber protein provides recognition of host cell surface proteins and provides entry into the cell, as a cell tropism parameter. CR1-β is important as it is noted to have a role in escaping the host immunosurveillance (Burgert and Blusch, 2000). Interestingly, the RID-alpha proteins of HAdV-A18 and -A31 show very similar charge/mass ratios in spite of the fact that they share only

Please cite this article as: Walsh, M.P., et al., Computational analysis of human adenovirus serotype 18, Virology (2010), doi:10.1016/j.virol.2010.05.013

Table 2
Comparison of VA RNA coding regions.

	Length of VA RNA I	VA RNA I	Length of VA RNA II	VA RNA II
HAdV-A12	141	97.2	–	–
HAdV-A31	141	90.1	–	–
HAdV-B3	175	52.5	177	50.4
HAdV-B7	175	52.5	171	52.5
HAdV-B11	158	53.2	–	–
HAdV-B14	162	51.8	–	–
HAdV-C2	160	48.9	158	54.6
HAdV-C5	160	48.9	158	54.6
HAdV-D9	163	58.9	154	47.5
HAdV-D53	163	58.2	179	47.5
HAdV-E4	159	58.2	101	33.7
SAdV-E25	159	56.0	160	46.8
HAdV-F40	171	63.1	–	–
HAdV-F41	174	63.8	–	–
HAdV-G52	164	65.3	–	–
SAdV-G1	164	63.8	–	–

Percent identities and coding lengths of VA RNA I from HAdV-A18, relative to counterparts in other species.

85.7% identity. The 2D gel supports an earlier report of species A polypeptide analysis that the patterns are "clearly distinct from each other" (Wadell et al., 1980). The hexon protein analysis, both 2D gel and phylogeny, also support the past observation that these three viruses cross-react via serum neutralization, while the phylogeny analysis agrees with the observation of similar hemagglutination–inhibition assay results for these species A members; the 2D gel shows physical differences between the three fibers (Wadell et al., 1980).

Conclusion

This report of the HAdV-A18 genomic sequence data and analysis completes the set of genomes classified within the HAdV-A species. One of the major contributions of this HAdV-A18 genome sequence is that it serves as a reference for further genomic and wet-bench studies, both as a reference for recombination analysis in understanding the phylogeny and molecular evolution and for studies of the origins of HAdVs. It is also important for understanding the putative roles of several HAdV genomes and their genes in reported and potential oncogenesis in rodents; the exact nucleotide sequences will allow additional wet-bench investigations.

Recombination has been implicated as a mechanism in the emergence of HAdVs as pathogens (Walsh et al., 2009). HAdV-A18 was examined for recombination events, but no overwhelming evidence of an event was found. However, as these types of in silico analyses are novel, further and additional genome studies are required to define the exact metrics for recombination. The few recent HAdV recombination events reported in the literature to date are very clearly supported by these data (Robinson et al., 2009; Walsh et al., 2009). However, in the case of HAdV-A18, while the Bootscan analysis evidence revealed the strong possibility of a recombination event in the hexon with HAdV-A31, complementary Simplot analysis and percent identity data are not supportive. Therefore, either the

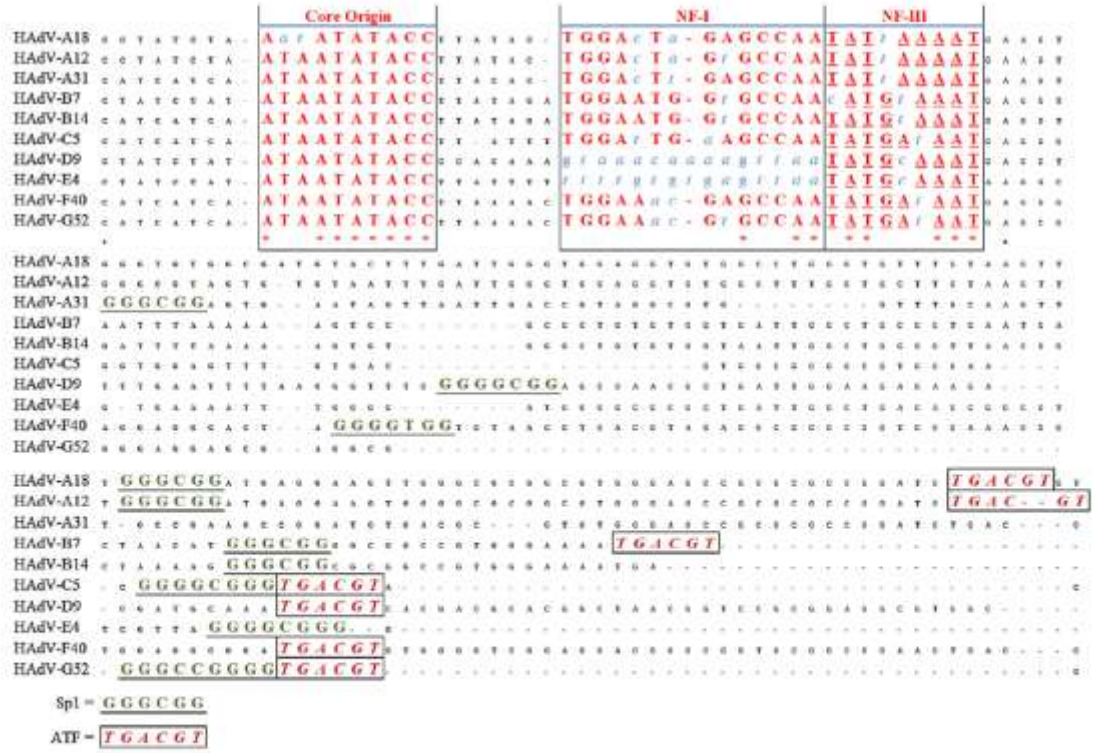


Fig. 4. HAdV ITR analysis. An alignment of the ITRs of selected HAdV is presented, with critical sequence motifs marked: core origin, NF I binding site, and NF III binding site. Non-conserved nucleotides within the boxes appear as blue italics, with the Sp1 motif (GGGCGG) underlined and ATF motif (TGACGT) boxed.

Please cite this article as: Walsh, M.P., et al., Computational analysis of human adenovirus serotype 18, Virology (2010), doi:10.1016/j.virol.2010.05.013

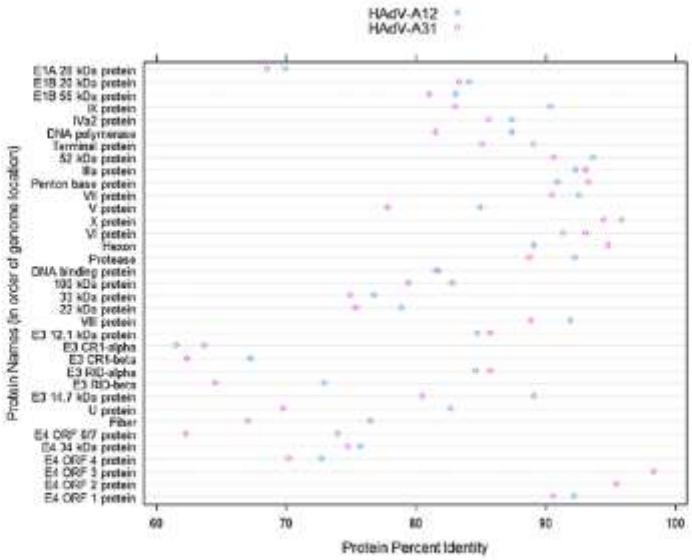


Fig. 5. Dot plot of protein percent identities. The percent identities of proteins of HAdV-A18 to homologs in the other members of the HAdV-A species (HAdV-A12 is in light blue and-A31 is in dark red) are plotted in a dot plot format. The proteins are sorted according to the location of their nucleotide coding sequence. This chart was constructed using the R statistical computing environment (http://www.R-project.org/).

hexon of HAdV-A18 has no recombination event or it does have a recombination event that is from an ancient event that has been complicated by genetic drift. As more HAdV recombination events are revealed and studied, the predicament of conflicting results between phylogeny-based methods (Bootscan) and similarity-based methods (Simplot and percent identity data) will likely to become a more prominent problem.

Materials and methods

HAdV-A18 was obtained from the collection at the Viral and Rickettsial Disease Laboratory (VRDL; California Department of Public Health), archived as TC-81190. The passage history of this stock is given as follows: five times through HeLa cells and eight times through KB cells at the National Institutes of Health, where the virus

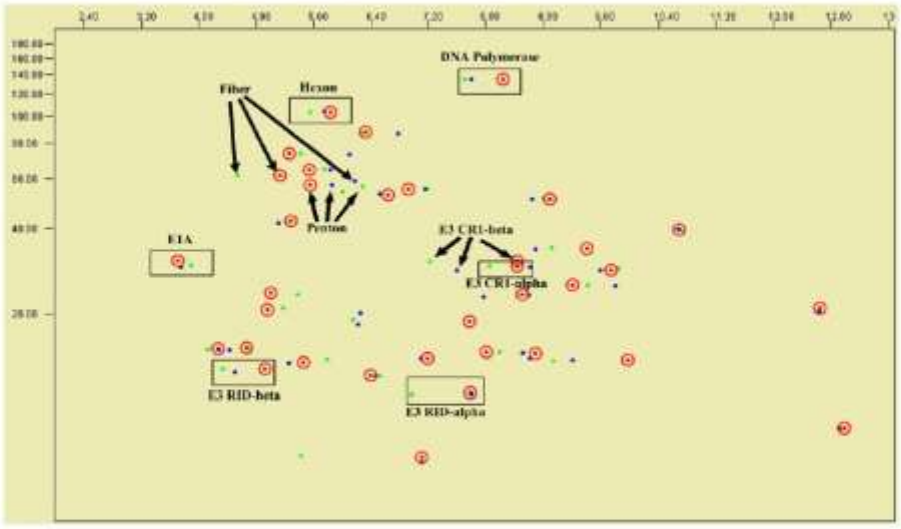


Fig. 6. Virtual 2D gel electrophoresis analysis. The in silico migration patterns of select proteins of HAdV-A18 (red and circled), HAdV-A12 (green or lighter shade of gray), and HAdV-A31 (blue or darker shade of gray) are noted, with arrows or boxes clustering the homologs.

Please cite this article as: Walsh, M.P., et al., Computational analysis of human adenovirus serotype 18, Virology (2010), doi:10.1016/j.virol.2010.05.013

was first cultured; and followed by passage through HeLa cells twice and human fetal diploid kidney cells four times at VRDL.

Preparation of HAdV-A18 for genome sequencing

For DNA production and collection, the HAdV-A18 stock was passed once onto monolayers of A549 cells in 25-cm^2 flasks to verify cytopathic effect, and then subsequently amplified in monolayers of A549 cells in 75-cm^2 flasks for intracellular viral DNA extraction, using a protocol described by Kajon and modified from Shinagawa (Kajon and Erdman, 2007). Cells collected from this passage were concentrated two-fold, and viral DNA was extracted and further processed using a MagnaNA Pure LC DNA Isolation Kit I (Roche; Indianapolis, IN), in accordance to the manufacturers' instructions.

Amplification and sequencing of the HAdV-A18 genome

To amplify regions of HAdV-A18, primers based on the conserved adenovirus sequences of HAdV-A12 were designed to flank the known sequences. All amplicons were sequenced using a primer-walking strategy and the Sanger sequencing chemistry with ladders resolved on an ABI3730x.

Sequence analysis and genome annotation

DNA sequence data were parsed and assembled using SeqMan software (Lasergene 8, Madison, WI). In toto, the genome assembly contained 1718 high-quality reads with an average length of 300 bps. Nucleotide coverage for both strands of the genome was 8-fold.

The annotated sequence of HAdV-A18 has been deposited in GenBank and is accessioned as GU191019. For the computational analyses, the following HAdV genomes were used: HAdV-A18 (reported here), HAdV-A12 (AC_000005), HAdV-A31 (Heim, personal communication; AM749299), HAdV-B3 (AY599834), HAdV-B7 (AY594255), HAdV-B11 (AY163756), HAdV-B14 (AY803294), HAdV-C2 (AC_000007), HAdV-C5 (AC_000008), HAdV-D9 (AJ854486), HAdV-D53 (FJ169625), HAdV-E4 (AY594253), SAdV-E25 (AC_000011), HAdV-F40 (NC_001454), HAdV-F41 (DQ315364), SAdV-G1 (NC_006879), and HAdV-G52 (DQ923122).

Bioinformatic analyses were performed using protocols and net-accessible software tools similar to those described in earlier publications (Walsh et al., 2009; Seto et al., 2009): (1) MAFFT (http://www.ebi.ac.uk/Tools/mafft/; Katoh et al., 2002) was used to compute alignments of nucleotide sequences; (2) nucleotide percent identities were calculated using the UCSF Chimera package (Pettersen et al., 2004); (3) Artemis (Rutherford et al., 2000) was used to view, annotate, and compare genomes; (4) percent identity values for annotated proteins were calculated using a BioJava (Holland et al., 2008) implementation of a Needleman–Wunsch algorithm; (5) whole genomes were aligned using MEGA (Tamura et al., 2007); (6) phylogenetic bootstrapped neighbor-joining trees (1000 replicates) were constructed using MEGA4 (Tamura et al., 2007); (7) recombination analysis was performed using Simplot (Lole et al., 1999); and (8) in silico proteome analyses were performed using an in-house beta software tool to calculate percent identities and applying a virtual 2D gel electrophoresis analysis to compare protein charge and size (JVirGel software; http://www.jvirgel.de; Hiller et al., 2006).

Acknowledgments

The work reported herein was performed with the support of the United States Air Force Surgeon General, Clinical Investigation No. FDG20040024E. This was also supported in part by US Public Health Service NIH grants EY013124 and P30EY013104, Massachusetts Lions Eye Research Fund, and an unrestricted grant to the Department of Ophthalmology, Harvard Medical School from Research to Prevent Blindness. DS and MPW were also supported by a USAF contract, FA442709P0262. The views expressed in this material are those of the authors and do not reflect the official policy or position of the US Government, the Department of Defense, or the Department of the Air Force.

References

Adrian, T., Wigand, R., 1989. Genome type analysis of adenovirus 31, a potential causative agent of infants' enteritis. Arch. Virol. 105 (1–2), 81–87.
Adrian, T., Wigand, R., Richter, J., 1987. Gastroenteritis in infants, associated with a genome type of adenovirus 31 and with combined rotavirus and adenovirus 31 infection. Eur. J. Pediatr. 146 (1), 38–40.
Brown, M., Petric, M., Middleton, P.J., 1984. Diagnosis of fastidious enteric adenoviruses 40 and 41 in stool specimens. J. Clin. Microbiol. 20 (3), 334–338.
Burgert, H., Blusch, J.H., 2000. Immunomodulatory functions encoded by the E3 transcription unit of adenoviruses. Virus Genes 21 (1), 13–25.
Dán, Á., Élő, P., Harrach, B., Zádori, Z., Benkő, M., 2001. Four new inverted terminal repeat sequences from bovine adenoviruses reveal striking differences in the length and content of the ITRs. Virus Genes 22 (2), 175–179.
De Jong, J.C., Wermenbol, A.G., Verweij-Uijterwaal, M.W., Slaterus, K.W., Wertheim-Van Dillen, P., Van Doornum, G.J., Khoo, S.H., Hierholzer, J.C., 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 (12), 3940–3945.
Fujinaga, K., Ojima, S., Yano, S., Shinagawa, M., 1977. The transforming DNA sequence common to highly oncogenic human adenoviruses, type 12, 18 and 31. Proc. Jpn Acad. B Phys. Biol. Sci. 53 (4), 152–155.
Gonçalves, M.A.F.V., de Vries, A.A.F., 2006. Adenovirus: from foe to friend. Rev. Med. Virol. 16 (3), 167–186.
Graham, F.L., van der Eb, A.J., Heijneker, H.L., 1974. Size and location of the transforming region in human adenovirus type 5 DNA. Nature 251 (5477), 687–691.
Hammond, G.W., Mauthe, G., Joshua, J., Hannan, C.K., 1985. Examination of uncommon clinical isolates of human adenoviruses by restriction endonuclease analysis. J. Clin. Microbiol. 21 (4), 611–616.
Hatfield, L., Hearing, P., 1993. The NFIII/OCT-1 binding site stimulates adenovirus DNA replication in vivo and is functionally redundant with adjacent sequences. J. Virol. 67 (7), 3931–3939.
Hiller, K., Grote, A., Maneck, M., Munch, R., Jahn, D., 2006. JVirGel 2.0: computational prediction of proteomes separated via two-dimensional gel electrophoresis under consideration of membrane and secreted proteins. Bioinformatics 22 (19), 2441–2443.
Hofmayer, S., Madisch, I., Darr, S., Rehren, F., Heim, A., 2009. Unique sequence features of the adenovirus 31 complete genomic sequence are conserved in clinical isolates. BMC Genomics 10 (1), 557.
Holland, R.C.G., Down, T.A., Pocock, M., Prlić, A., Huen, D., James, K., Foisy, S., Dräger, A., Yates, A., Heuer, M., Schreiber, M.J., 2008. BioJava: an open-source framework for bioinformatics. Bioinformatics 24 (18), 2096–2097.
Huebner, R.J., Rowe, W.P., Lane, W.T., 1962. Oncogenic effects in hamsters of human adenovirus types 12 and 18. Proc. Natl Acad. Sci. U. S. A. 48 (12), 2051–2058.
Ishiko, H., Aoki, K., 2009. Spread of epidemic keratoconjunctivitis due to a novel serotype of human adenovirus in Japan. J. Clin. Microbiol. 47 (8), 2678–2679.
de Jong, J.C., Wigand, R., Kidd, A.H., Wadell, G., Kapsenberg, J.G., Muzerie, C.J., Wermenbol, A.G., Firtzlaff, R.G., 1983. Candidate adenoviruses 40 and 41: fastidious adenoviruses from human infant stool. J. Med. Virol. 11 (3), 215–231.
Kajon, A.E., Erdman, D.D., 2007. Assessment of genetic variability among subspecies b1 human adenoviruses for molecular epidemiology studies. Meth. Mol. Med. 131, 335–355.
Katoh, K., Misawa, K., Kuma, K., Miyata, T., 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucl. Acids Res. 30 (14), 3059–3066.
Kidd, A.H., Cosgrove, B.P., Brown, R.A., Madeley, C.R., 1982. Faecal adenoviruses from Glasgow babies. Studies on culture and identity. J. Hyg. (Lond.). 88 (3), 463–474.
Lole, K.S., Bollinger, R.C., Paranjape, R.S., Gadkari, D., Kulkarni, S.S., Novak, N.G., Ingersoll, R., Sheppard, H.W., Ray, S.C., 1999. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. J. Virol. 73 (1), 152–160.
Mathews, M.B., Shenk, T., 1991. Adenovirus virus-associated RNA and translation control. J. Virol. 65 (11), 5657–5662.
Pereira, M.S., Pereira, H.G., Clarke, S.K., 1965. Human adenovirus type 31. A new serotype with oncogenic properties. Lancet 1 (7375), 21–23.
Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E., 2004. UCSF Chimera—a visualization system for exploratory research and analysis. J. Comput. Chem. 25 (13), 1605–1612.
Robinson, C.M., Rajaiya, J., Walsh, M.P., Seto, D., Dyer, D.W., Jones, M.S., Chodosh, J., 2009. Computational analysis of human adenovirus type 22 provides evidence for recombination between human adenoviruses species D in the penton base gene. J. Virol. Accessed August 11, 2009.
Rowe, W.P., Hartley, J.W., Huebner, R.J., 1956. Additional serotypes of the APC virus group. Proc. Soc. Exp. Biol. Med. 91 (2), 260–262.
Rowe, W.P., Hartley, J.W., Huebner, R.J., 1958. Serotype composition of the adenovirus group. Proc. Soc. Exp. Biol. Med. 97 (2), 465–470.
Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.A., Barrell, B., 2000. Artemis: sequence visualization and annotation. Bioinformatics 16 (10), 944–945.

Please cite this article as: Walsh, M.P., et al., Computational analysis of human adenovirus serotype 18, Virology (2010), doi:10.1016/j.virol.2010.05.013

Sawada, Y., Ojima, S., Shimojo, H., Shiroki, K., Fujinaga, K., 1979. Transforming DNA sequences in rat cells transformed by DNA fragments of highly oncogenic human adenovirus type 12. J. Virol. 32 (2), 379–385.

Seto, J., Walsh, M.P., Mahadevan, P., Purkayastha, A., Clark, J.M., Tibbetts, C., Seto, D., 2009. Genomic and bioinformatics analyses of HAdV-14p, reference strain of a re-emerging respiratory pathogen and analysis of B1/B2. Virus Res. 143 (1), 94–105.

Sprengel, J., Schmitz, B., Heuss-Neitzel, D., Zock, C., Doerfler, W., 1994. Nucleotide sequence of human adenovirus type 12 DNA: comparative functional analysis. J. Virol. 68 (1), 379–389.

Tamura, K., Dudley, J., Nei, M., Kumar, S., 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol. Biol. Evol. 24 (8), 1596–1599.

Temperley, S.M., Hay, R.T., 1992. Recognition of the adenovirus type 2 origin of DNA replication by the virally encoded DNA polymerase and preterminal proteins. EMBO J. 11 (2), 761–768.

Trentin, J.J., Yabe, Y., Taylor, G., 1962. The quest for human cancer viruses: a new approach to an old problem reveals cancer induction in hamsters by human adenovirus. Science 137 (3533), 835–841.

Uhnoo, I., Wadell, G., Svensson, L., Johansson, M.E., 1984. Importance of enteric adenoviruses 40 and 41 in acute gastroenteritis in infants and young children. J. Clin. Microbiol. 20 (3), 365–372.

Vasavada, R., Eager, K.B., Barbanti-Brodano, G., Caputo, A., Ricciardi, R.P., 1986. Adenovirus type 12 early region 1A proteins repress class I HLA expression in transformed human cells. Proc. Natl Acad. Sci. U. S. A. 83 (14), 5257–5261.

Vora, G.J., Lin, B., Gratwick, K., Meador, C., Hansen, C., Tibbetts, C., Stenger, D.A., Irvine, M., Seto, D., Purkayastha, A., Freed, N.E., Gibson, M.G., Russell, K., Metzgar, D., 2006. Co-infections of adenovirus species in previously vaccinated patients. Emerg. Infect. Dis. 12 (6), 921–930.

Wadell, G., Hammarskjold, M.L., Winberg, G., Varsanyi, T.M., Sundell, G., 1980. Genetic variability of adenoviruses. Ann. NY Acad. Sci. 354, 16–42.

Walsh, M.P., Chintakuntlawar, A., Robinson, C.M., Madisch, I., Harrach, B., Hudson, N.R., Schnurr, D., Heim, A., Chodosh, J., Seto, D., Jones, M.S., 2009. Evidence of molecular evolution driven by recombination events influencing tropism in a novel human adenovirus that causes epidemic keratoconjunctivitis. PLoS ONE 4 (6), e5635.

Yabe, Y., Samper, L., Bryan, E., Taylor, G., Trentin, J.J., 1964. Oncogenic effect of human adenovirus type 12, in mice. Science 143 (3601), 46–47.

Yang, Z., Zhu, Z., Tang, L., Wang, L., Tan, X., Yu, P., Zhang, Y., Tian, X., Wang, J., Zhang, Y., Li, D., Xu, W., 2009. Genomic analyses of recombinant adenovirus type 11a in China. J. Clin. Microbiol. 47 (10), 3082–3090.

Please cite this article as: Walsh, M.P., et al., Computational analysis of human adenovirus serotype 18, Virology (2010), doi:10.1016/j.virol.2010.05.013

CHAPTER 4: Recombinant Genomes

Recent genomics studies indicate that recombination may be a driving force in the evolution of HAdV types [63,38,64,10,11]. Recombination events among HAdV have been implicated in the genesis of novel types [5]. These recombinations have also been linked to changes in virus tropism among re-emergent pathogens [6]. Parts 1-3 of this chapter contain publications that describe three of the first recombination studies of HAdVs that use genomics in attempt to understand molecular evolution at a fine resolution.

A genomics, bioinformatics and recombination analysis of HAdV-D53, a severe, highly contagious, eye pathogen, is described in the publication that comprises Part 1 of this chapter [5]. The paper describes the first computational and detailed examination of recombination events among HAdV types. Data from this examination reveal that the HAdV-D53 genome is the result of at multiple recombination events. Furthermore, each of the genes that encode for the HAdV-D53's surface coat proteins (penton, hexon, and fiber) originates from a different HAdV (HAdV-D37, HAdV-D22, and HAdV-D8, respectively). The evidence presented in the paper suggests that recombinations among the HAdV-D species have led to the emergence of a novel pathogen, in this instance, HAdV-D53.

Part 2 of this chapter contains a paper that describes a bioinformatics analysis of HAdV-D22 [27]. The publication's major finding is the first description of a penton recombination, between HAdV-D22, D37 and D19. The paper also explores the possibility of events in other HAdV pentons and suggests that a possible "hot spot" for recombinations exists within the gene. These data are interesting because it implies that recombination events among HAdV types (especially in the D species) occur more frequently than originally thought.

The final part of this chapter contains a publication describing the computational examination of HAdV-B55, a re-emergent and historically sporadic respiratory pathogen [6]. HAdV-B55 was originally mis-identified as a variant of HAdV-B11 because of molecular typing results discussed within this publication. Data from this publication reveal that HAdV-B55 is most closely related to HAdV-B14. In retrospect and with whole genome data, the molecular typing results were confounded by a recombination event in the portion of the hexon (both variable loops) that was amplified and sequenced as part of the typing process. The variable loops of the hexon constitute less than three percent of the genome and are the only part of the virus that is closely related to HAdV-B11.

HAdV-B55 is an example that a small hexon recombination may have had profound effects on the virus. This recombination event allows HAdV-B55 to disguise itself as a renal pathogen and evade the host defenses.

Part 1

Evidence of Molecular Evolution Driven by Recombination Events Influencing Tropism in a Novel Human Adenovirus that Causes Epidemic Keratoconjunctivitis

OPEN ACCESS Freely available online

PLoS one

Evidence of Molecular Evolution Driven by Recombination Events Influencing Tropism in a Novel Human Adenovirus that Causes Epidemic Keratoconjunctivitis

Michael P. Walsh[1], Ashish Chintakuntlawar[2], Christopher M. Robinson[2], Ijad Madisch[3], Balázs Harrach[4], Nolan R. Hudson[5], David Schnurr[6], Albert Heim[3], James Chodosh[2], Donald Seto[1], Morris S. Jones[5]*

1 Department of Bioinformatics and Computational Biology, George Mason University, Manassas, Virginia, United States of America, 2 Howe Laboratory, Massachusetts Eye and Ear Infirmary, Harvard Medical School, Boston, Massachusetts, United States of America, 3 Institut für Virologie, Medizinische Hochschule, Hannover, Germany, 4 Veterinary Medical Research Institute, Hungarian Academy of Sciences, Budapest, Hungary, 5 Clinical Investigation Facility, David Grant USAF Medical Center, Travis AFB, California, United States of America, 6 Viral and Rickettsial Disease Laboratory, California Department of Public Health, Richmond, California, United States of America

Abstract

In 2005, a human adenovirus strain (formerly known as HAdV-D22/H8 but renamed here HAdV-D53) was isolated from an outbreak of epidemic keratoconjunctititis (EKC), a disease that is usually caused by HAdV-D8, -D19, or -D37, not HAdV-D22. To date, a complete change of tropism compared to the prototype has never been observed, although apparent recombinant strains of other viruses from species *Human adenovirus D* (HAdV-D) have been described. The complete genome of HAdV-D53 was sequenced to elucidate recombination events that lead to the emergence of a viable and highly virulent virus with a modified tropism. Bioinformatic and phylogenetic analyses of this genome demonstrate that this adenovirus is a recombinant of HAdV-D8 (including the fiber gene encoding the primary cellular receptor binding site), HAdV-D22, (the ε determinant of the hexon gene), HAdV-D37 (including the penton base gene encoding the secondary cellular receptor binding site), and at least one unknown or unsequenced HAdV-D strain. Bootscanning analysis of the complete genomic sequence of this novel adenovirus, which we have re-named HAdV-D53, indicated at least five recombination events between the aforementioned adenoviruses. Intrahexon recombination sites perfectly framed the ε neutralization determinant that was almost identical to the HAdV-D22 prototype. Additional bootscan analysis of all HAdV-D hexon genes revealed recombinations in identical locations in several other adenoviruses. In addition, HAdV-D53 but not HAdV-D22 induced corneal inflammation in a mouse model. Serological analysis confirmed previous results and demonstrated that HAdV-D53 has a neutralization profile representative of the ε determinant of its hexon (HAdV-D22) and the fiber (HAdV-D8) proteins. Our recombinant hexon sequence is almost identical to the hexon sequences of the HAdV-D strain causing EKC outbreaks in Japan, suggesting that HAdV-D53 is pandemic as an emerging EKC agent. This documents the first genomic, bioinformatic, and biological descriptions of the molecular evolution events engendering an emerging pathogenic adenovirus.

Citation: Walsh MP, Chintakuntlawar A, Robinson CM, Madisch I, Harrach 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(6): e5635. doi:10.1371/journal.pone.0005635

Editor: Wanda Markotter, University of Pretoria, South Africa

Received January 7, 2009; **Accepted** April 8, 2009; **Published** June 3, 2009

Funding: The work reported herein was performed under United States Air Force Surgeon General-approved Clinical Investigation No. FDG20040024E, and partially supported by Hungarian Research Fund grant K72484, U.S. Public Health Service NIH grants EY013124 and T32 A1007633, and a Research to Prevent Blindness Physician-Scientist Merit Award (to JC). 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: drmorrisj@yahoo.com

Introduction

Epidemic keratoconjunctivitis (EKC), characterized by inflammation of the conjunctiva and cornea, produces a sudden onset of acute follicular conjunctivitis and stromal keratitis and is a worldwide problem causing significant and sometime lasting morbidity [1]. Human adenoviruses (HAdVs) HAdV-D8, -D19, and -D37 are the most common pathogens causing EKC [1].

Adenoviruses were first isolated from civilians and military trainees who had respiratory disease in the early 1950s [2,3]. They were the first respiratory viruses to be isolated and characterized. Epidemiological studies confirmed that adenoviruses are the cause of acute febrile respiratory disease among military recruits [4,5] and have been persistent in the global population. Since then, 52 human adenovirus (HAdV) genotypes have been characterized and classified according to their immunochemical properties, nucleic acid similarities, hexon and fiber protein characteristics, biological properties, and phylogenetic analysis, and placed in the genus *Mastadenovirus* [6,7]. These 52 adenovirus genotypes that infect humans are classified into seven species (*Human adenovirus A* to *G*) [6,8] and are known to cause a range of diseases specific to the tropisms of the viruses: keratoconjunctivitis (HAdV-D8, HAdV-D19, and HAdV-D37) [9,10], gastroenteritis (HAdV-A31, HAdV-F40, HAdV-F41, and HAdV-G52) [6], acute

respiratory disease (HAdV-B3, HAdV-E4, HAdV-B7, HAdV-B14, and HAdV-B21) [5], and perhaps obesity (HAdV-D36) [11].

Adenoviruses have linear double-stranded DNA genomes that generally range from 26 to 45 kb and are encapsidated in an icosahedral protein shell that ranges from 70 to 100 nm [8]. The primary components of the protein shell are the hexon, penton base, and fiber proteins. Through genome sequence analysis, it has been demonstrated that the genomes of all human adenoviruses have similar genetic organization [12,13,14].

In the past, human adenovirus serotype and species classification were defined by reactivity of outer coat proteins to discriminating antibodies (e.g., immunochemistry/virus neutralization) as well as by other biological properties (e.g. oncogenic potential, hemagglutination properties). Today, given the availability of DNA sequencing and analysis technology, phylogenetics (based on comparative nucleic acid and amino acid sequence analysis of informative viral proteins or/and their genes, as well as analysis of genomic organization) is a highly quantitative, cost-effective, expedient method and the preferred and reliable method for classifying adenoviruses. It is a preferred and reliable method for demonstrating how viruses are related through molecular evolution as it provides and relies on the primary sequence data [15,16,17,18].

In this study we sought to characterize a unique intermediate recombinant HAdV isolate, at the molecular level. This novel strain was isolated from a patient who, along with eleven other patients, presented with highly contagious EKC outbreak in Germany was described [19]. Since HAdV-D22 was never associated with EKC, we performed whole genome sequencing, complemented with bioinformatics, including phylogenetic and in silico proteome analysis, as well as *in vivo* studies in a mouse model to characterize this unique recombinant virus. To reflect this novel and different genome and because of the multiple recombination events and several unique sequence segments in the genome of this virus, we renamed this virus HAdV-D53.

Results

Amplification and sequencing of the new adenovirus

Initial and partial sequencing of HAdV-D53 (previously HAdV-D22/H8) demonstrated that portions of the penton and fiber genes were similar to HAdV-D37 and HAdV-D8, respectively [19]; thus suggesting that this disease causing virus was the result of recombination. To understand clearly the genetic characteristics and the nature of HAdV-D53, the entire genome has been sequenced and analyzed.

Physical features of new adenovirus genome

The genome length of HAdV-D53 is 34,909 base pairs, with a base composition of 23% A, 20.8% T, 28.2% G, 28% C and the GC content was 56.2%. The GC content is consistent with members of species *Human adenovirus D* (HAdV-D) (57.0% mean). The organization of the 36 open reading frames (ORF's) that were found had a genome organization similar to other mastadenoviruses (Fig. 1). The inverted terminal repeat (ITR) sequences for HAdV-D53 were determined to be 212 bp in length.

The nucleotide and amino acid identities for selected genes in the genome of HAdV-D53 to its nearest relatives are shown in Tables 1 and 2, respectively. Interestingly, the pVII and protein V genes were dissimilar to homologous genes in any adenovirus species with 83 and 87% nucleotide identity, respectively to HAdV-D37 (Table 1), the nearest relative in that region of the HAdV-D53 genome. For pVII, the low nucleotide identity is partially due to a 99 bp deletion, resulting in a 33 amino acid deletion of the predicted pVII protein. When compared to HAdV-D37, the protein V gene contains 2 deletions. The first deletion is 18 bp and the second is 93 bp.

Genomic recombination analysis

To determine if recombination occurred within the HAdV-D53 genome, several software tools were applied. A bootscanning

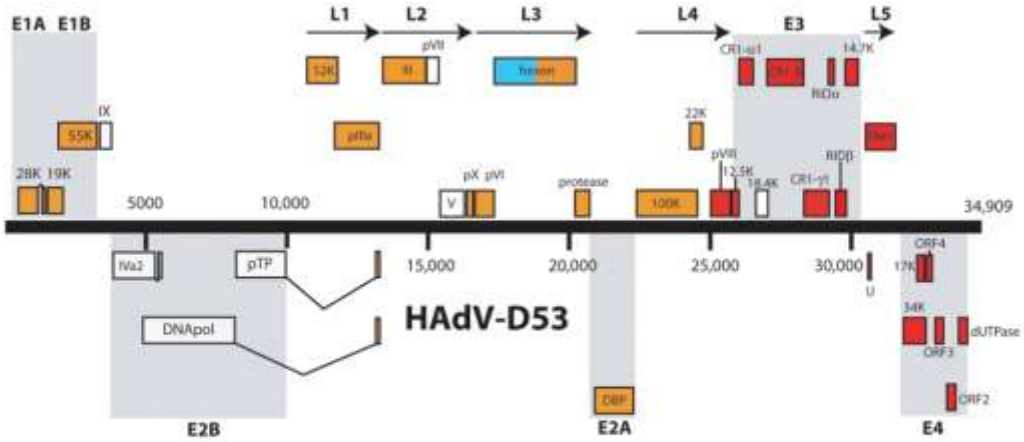


Figure 1. Genome organization of HAdV-D53. Genome is represented by a central black horizontal line marked at 5-kbp intervals. Protein-encoding regions are shown as boxes. Boxes above the black line represent open reading frames (ORFs) that are encoded on the forward (or upper) strand. Boxes underneath the black line represent ORFs that are encoded on the reverse (or lower) strand. The colors of the boxes correspond to which adenovirus the protein is most likely descended from: red – HAdV-D8, aqua – HAdV-22, orange – HAdV-D37, white – dissimilar to all known adenoviruses.
doi:10.1371/journal.pone.0005635.g001

Table 1. Percent identities of the nucleotide coding sequences of selected HAdV-D53 proteins and their homologs[a].

	HAdV-D8p[b]	HAdV-D19	HAdV-D22	HAdV-D37	HAdV-D48	HAdV-D49
E1B 19K	94.9	99.3	99	**99.5**	97.8	97.8
E1B 55K	96	99.2	98.3	**99.3**	97.4	96.9
IX	95.1	96.1	98.5	96.1	97.1	97.1
IVa2	96.7	98	97.4	98	98.3	91.3
DNA polymerase	95.2	98.1	97.6	98	98	98
pTP	93.3	97.6	96.8	97.9	96.4	96.8
52K	95.7	98	98.1	**100**	97.9	98.3
penton base	89.2	91.2	92.4	**100**	90.5	90.1
pVII	79.2	80.4	80.4	83.2	80	80.1
V	85.2	87	87.5	87.2	87.7	87.3
pX	95.1	100	100	**100**	98.7	99.6
hexon	89.4	90.2	*98.4*	*98.5*	90.6	90.4
protease	95.4	96	95.4	**99.8**	96.5	96.5
22K	89.9	100	98.8	**100**	99.3	99
pVIII	95.6	98.4	98.5	98.4	98.3	97.7
12.2K	93.2	96	97.5	96	96.9	96.9
CR1-α	52.5	80.5	97	80.5	75.4	77.6
18.4K	95.2	91.1	98.1	91.1	96.2	94.1
CR1-β	**100**	74.5	85.1	74.5	64.5	58
CR1-γ	86.4	75.3	Nd[c]	75.3	75.1	80.5
RID-α	93.5	94.2	94.2	94.2	98.2	98.2
RID-β	90.6	87.4	93.2	87.4	94.2	99.2
14.7K	96.7	95.2	97.2	95.2	96.7	97.7
fiber	**100**	75.1	67.6	75	69	67.6
dTPase	**100**	48.4	88.9	48.4	85.1	85

Standard nomenclature has been applied so that orthologs have the same name (Davison et al., 2003). Numbers in bold reflect the proposed origin. Italics note the gene with supposed double origin.

[a]Percent identities and similarities were determined by global alignment using the EMBOSS needle program with a gap penalty of 10.0 and a gap extension penalty of 5.0.

[b]Not present in the genome.

[c]Prototype HAdV-D8 strain is Trim isolate – ATCC VR-1604.

doi:10.1371/journal.pone.0005635.t001

program [20] was used to determine the relationship of HAdV-D53 to all of the fully sequenced HAdV-D genotypes. According to the alignments, several regions indicated recombination events; nucleotides (as genome coordinates) 1–1000, 1500–3250 (E1A, E1B, 55K), 8500–15,750 (52K, pIIIa, penton base), 17,000–17,750, (pX, pVI) and 19,500–25,000 (second half of hexon, protease, DBP, 100K, 22K) showed a strong relationship to HAdV-D37; nucleotides 17,750–19,500 (first half of hexon) showed a strong relationship to HAdV-D22; nucleotides 27,375–29,750 (CR1-β) and 30,500–34,909 (14.7K, fiber, E4 ORF6) showed a strong relationship to HAdV-D8 (Fig. 2). Although the bootscan analysis showed that nucleotides 29750–30,500 have a strong relationship to HAdV-D49, we believe that this region comes from an unsequenced HAdV-D8 strain, because that region (RIDβ) in a partially sequenced HAdV-D8 strain has 100% amino acid identity to the Hiroshima HAdV-D8 isolate (Table 2). These relationships were confirmed by comparison with nucleotide identity in Table 1, as well as BLASTP similarity analysis of the proteins (Table 2). In contrast, nucleotides 1000–1500, 3250–8500, 15,750–17,000, and 25,000–27,375 showed slightly lower similarity to several known adenoviruses, suggesting that this region of species HAdV-D adenoviruses are both well conserved and so far unique for the studied strain. Thus, bootscan analysis of the HAdV-D53 genome shows evidence of multiple recombination events.

Hexon recombination analysis

The results of our whole genome bootscan indicated that a recombination event occurred inside the hexon gene. The hexon contains loops 1 (L1) and 2 (L2), which are the most important determinants of neutralization via antibodies as well as immune escape. Since L1 and L2 are the most relevant for serotyping, we performed bootscan analysis to pinpoint where the recombination events occurred in the hexon gene of HAdV-D53. The results of the bootscan analysis shown in Figure 3A and 3B reveal that a recombination event occurred between nucleotide 380 and 1400–1620 which are the amino terminus of L1 and the conserved C terminus of the highly variable L2, respectively (Table 3). Thus, the complete neutralization epitope ε, which is nearly identical to the sequenced HAdV-D22, is framed by non HAdV-D22 sequences in the recombinant strain HAdV-D53.

Based on the nucleotide identity of HAdV-D53 to other adenoviruses, we believe that the previous name HAdV-D22/H8 is not appropriate due to the fact that the fully sequenced genome and the bioinformatic analyses demonstrate that HAdV-D53 is the

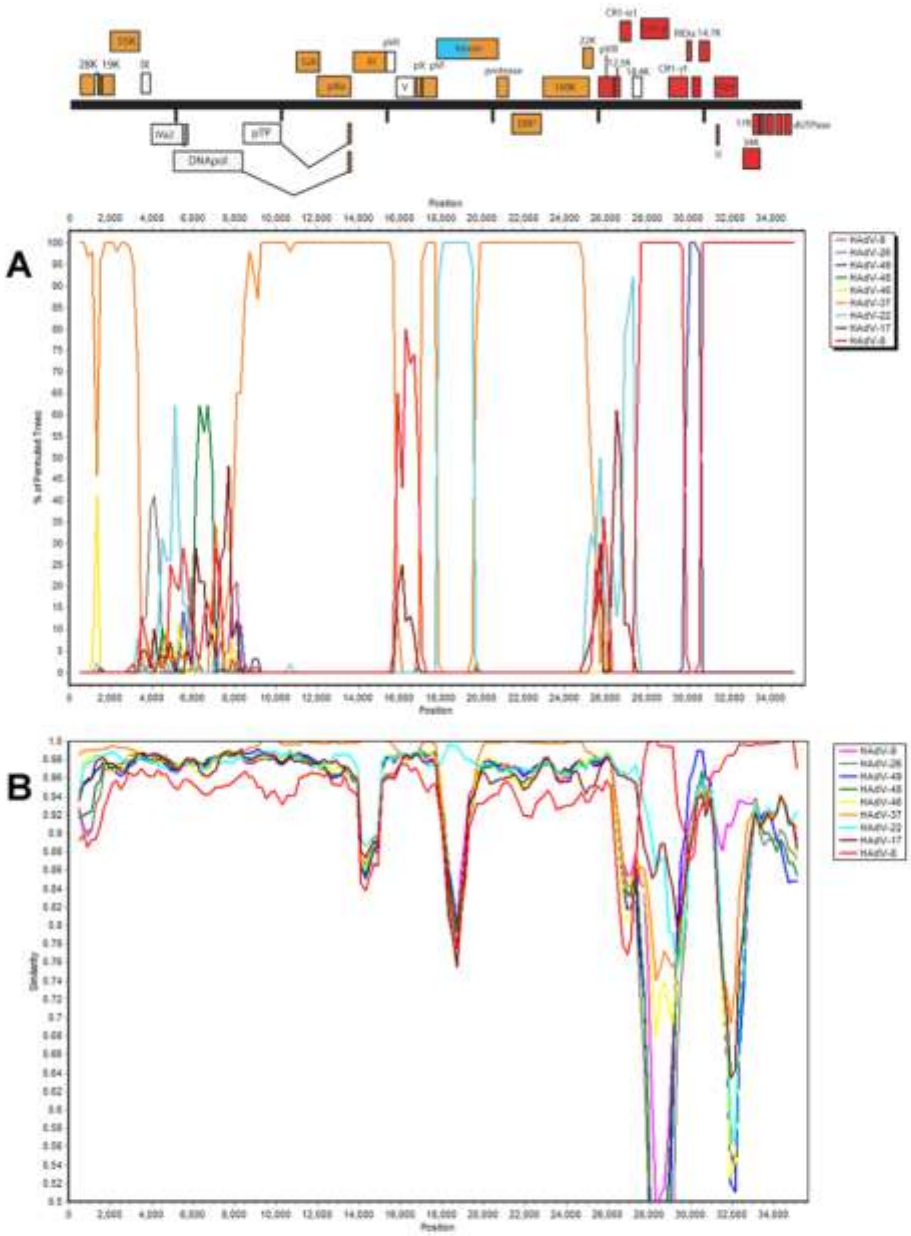


Figure 2. Whole genome (A) bootscan and (B) simplot of HAdV-D53 compared to fully sequenced HAdV-D genomes.
doi:10.1371/journal.pone.0005635.g002

product of multiple recombinations of known and perhaps undiscovered and/or yet unsequenced adenoviruses. Taken together, we propose the name "HAdV-D53" for this novel recombinant adenovirus, reflecting its genome divergence from other human adenoviruses. We also believe this "genome type" designation is more appropriate in light of the current and future DNA sequencing

Table 2. Percent identities of selected amino acid sequences of HAdV-D53 proteins and their homologs.

	HAdV-D8	HAdV-D8p	HAdV-D19	HAdV-D22	HAdV-D37	HAdV-D48	HAdV-D49
E1B19K		92	99	99	**100**	97	97
E1B55K		96	99	98	**99**	97	96
IX		95	97	99	97	97	98
IVa2		97	99	99	**99**	99	98
DNA polymerase		96	98	98	98	99	98
pTP		94	97	97	98	97	97
52K		96	99	99	**100**	98	98
penton base		89	91	92	**100**	89	89
pVII		79	80	80	81	79	80
V		84	87	87	87	87	87
pX		100	100	100	**100**	100	100
hexon		92	90	99	90	92	90
protease		97	100	99	**100**	100	100
DBP		96	99	98	**100**	97	97
22K		81	100	99	**100**	99	98
pVIII	**100**	97	98	99	98	98	98
12.5K	**100**	95	95	99	95	98	98
CR1-α	**99**	54	73	94	73	67	71
18.4K	97	95	91	97	91	94	92
CR1-β	97	**100**	74	81	74	48	44
CR1-γ	**100**	**81**	62	65	62	65	73
RID-α	**100**	94	29	96	96	100	100
RID-β	**100**	89	92	93	92	89	97
14.7K	**100**	96	94	97	94	97	98
fiber	**100**	**100**	74	62	74	63	59
dUTPase		**100**	92	87	92	84	82

Numbers in bold reflect the proposed origin. Italics note the gene with supposed double origin.

doi:10.1371/journal.pone.0005635.t002

and analysis technology, superseding the importance of the previous classifications based on serology (e.g., serotypes).

Hexon recombination is common in species *Human adenovirus D*

To determine whether or not this phenomenon was common in other adenoviruses, the available hexon genes of all HAdV-D genotypes were cross-examined. Recombination events at similar nucleotide locations of the hexon gene in HAdV-D13, HAdV-D32, and HAdV-D39 (Fig. 3C–E) were found. The HAdV-D13 recombination is especially interesting regarding the present study, as HAdV-D13 acquired L1 and L2 from HAdV-D37 and the same region in HAdV-D53 was presumably exchanged for L1 and L2 of HAdV-D22. To demonstrate the validity of our recombination predictions, we included the bootscan analysis of the HAdV-D49 hexon gene, which does not show evidence of any recombination events (Fig. 3F). Taken together, these data suggest that adenoviruses in HAdV-D species are susceptible to recombination events at the amino terminus of L1 and the carboxy terminus of L2 of the hexon gene; implicating a mechanism which allows adenoviruses to switch neutralization epitopes.

In vivo HAdV-D53 induced keratitis

Since HAdV-D53 was isolated from a patient with EKC and appeared to be corneotropic [19], we tested its ability to induce corneal innate immune responses in a previously described mouse model of adenovirus keratitis, in which EKC viruses induce a keratitis similar to human EKC, but without viral replication [21]. HAdV-D53 infection induced a clinically evident keratitis (corneal opacity) as early as 1 day post-infection (dpi) that peaked by 3–4 dpi (Fig. 4A). In contrast, mock and HAdV-D22 injection did not induce corneal opacity at any time post-infection. Neither virus replicated in the mouse cornea (data not shown). Hematoxylin and eosin staining of corneal cross sections at 4 dpi with HAdV-D53 showed thinning of the epithelial cell layer, stromal edema, and infiltration by leukocytes (Fig. 4B). In contrast, HAdV-D22 infection induced only modest cellular infiltration. We next assessed corneal myeloperoxidase (MPO) levels after infection as a measure of the presence of infiltrating neutrophils and monocytes [22,23]. HAdV-D53 infection induced significantly higher levels of MPO when compared to HAdV-D22 and mock infected corneas (Fig. 4C). By flow cytometry, corneal infection with HAdV-D53 caused a significantly greater number of infiltrating neutrophils (Gr1+F4/80−) [24,25], similar to previous studies with HAdV-D37 [21], than with HAdV-D22 infection. Inflammatory monocytes (Gr1+F4/80+) [26,27] and resident macrophages (Gr1-F4/80+) [28] did not increase significantly after infection with either virus (Figs. 4F and G). Because neutrophils appeared by histology and flow cytometry to be the predominant infiltrating cell in HAdV-D53 keratitis, we also tested

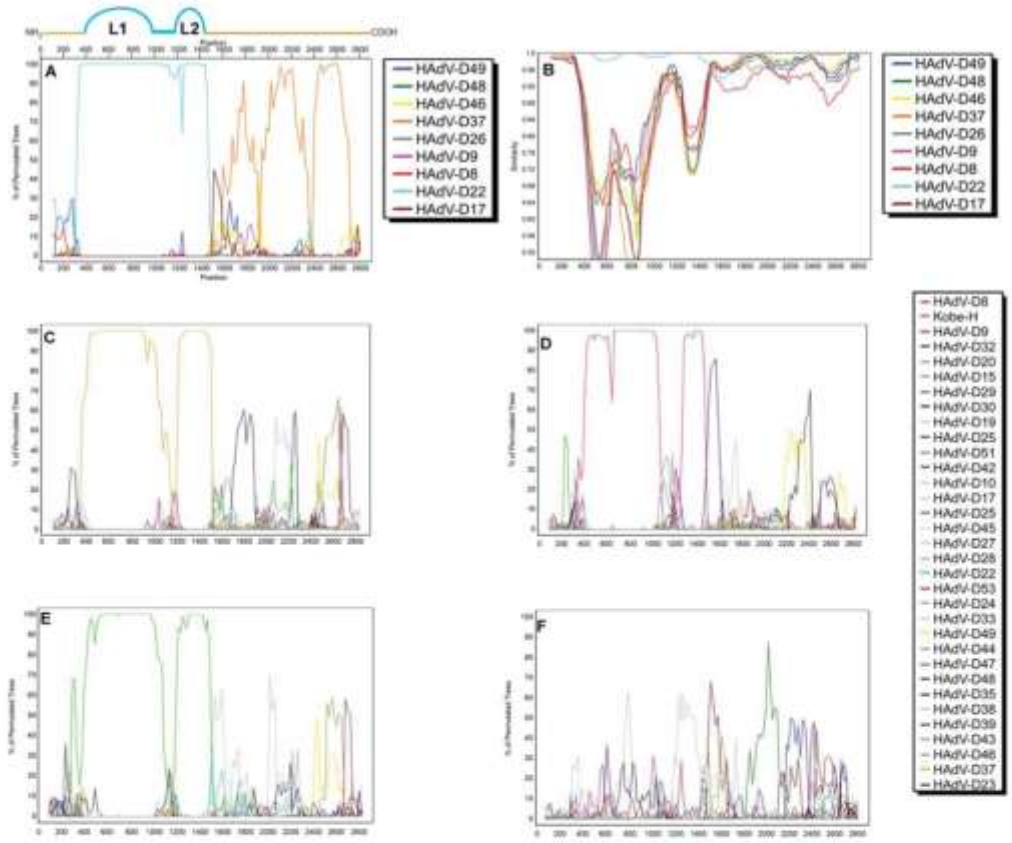


Figure 3. Bootscan of HAdV-D species hexon genes demonstrating recombination events. Comparison of HAdV-D53 by (A) bootscan and (B) simplot with the HAdV-D types which have a fully sequenced genome. (C) HAdV-D13, (D) HAdV-D32, (E) HAdV-D39, and (F) HAdV-D49 were compared to all hexon genes in species HAdV-D by bootscan analysis.
doi:10.1371/journal.pone.0005635.g003

the expression of neutrophil chemokines CXCL1 and CXCL2 [29,30]. Both CXCL1 (Fig. 4D) and CXCL2 (Fig. 4E) were expressed at significantly higher levels after infection with HAdV-D53 than with HAdV-D22.

Phylogenetic analysis

Detailed phylogenetic analysis of selected proteins, performed with nucleotide data and deduced amino acid sequences confirmed that HAdV-D53 was an unusual recombinant adenovirus. The tree topology of HAdV-D53 was different depending on which protein was tested. The penton base had the closest relationship to HAdV-D37, whereas the fiber gene was closest to HAdV-D8 (Fig. 5). Interestingly, DNA polymerase, protein V, and pVII genes did not cluster tightly with any other virus, which reflects their unique sequences. As expected the L1 and L2, which are responsible for the neutralization ε determinant, clustered with HAdV-D22. In L1, HAdV-D22 was 1.8% distant to HAdV-D53 and L2 was identical to HAdV-D53. In the β-determinant, HAdV-D53, supported by a strong bootstrap value (83%), clustered to HAdV-D37. Using sequence data that was available in GenBank, the hexon sequences of HAdV-D53 clustered tightly with the Japanese isolates 1/Yamaguchi/2004, C075/Matsuyama/2003, and FS161/Fukui/2004 suggesting that HAdV-D53 and the Japanese isolates represent different isolates of the same HAdV genotype.

Viral neutralization

Since our sequence analysis shows that HAdV-D53 is genetically similar to HAdV-D8, -D22, and -D37, we wanted to determine its serum neutralization profile. Antisera to HAdV-D8 and HAdV-D22 neutralized HAdV-D53 at dilutions of 1:128 and 1:256, respectively. In contrast, antisera to HAdV-D37 was unable to neutralize HAdV-D53 at a dilution of equal to or less than 1:8. These results confirm previous results and demonstrated that HAdV-D53 has a neutralization profile representative of its hexon and fiber proteins whereas the penton base did not contribute to neutralization.

Table 3. An excerpt from the plot values of a Simplot Bootscan of the HAdV-22D and HAdV-D37 hexons.

Center Pos	HAdV-D22 hexon	HAdV-D37 hexon
1320	**100**	0
1340	**100**	0
1360	**100**	0
1380	**100**	0
1400	**100**	0
1420	**100**	0
1440	**100**	0
1460	80	0
1480	18	1
1500	1	12
1520	0	7
1540	0	7
1560	0	3
1580	0	2
1600	0	46
1620	0	51
1640	0	45
1660	0	51
1680	0	60
1700	0	57
1720	0	50
1740	0	80

doi:10.1371/journal.pone.0005635.t003

Discussion

The initial report of HAdV-D53 described that this novel, possibly emergent disease-causing, strain comprised a HAdV-D22 strain that had recombined with HAdV-D8 and HAdV-D37 [19]. Full genome sequencing of this isolate, HAdV-D53, and bioinformatic analysis have demonstrated that this genome is so different both from HAdV-D22 and all the other officially accepted serotypes that it must be seen as a novel human adenovirus which we have re-named HAdV-D53 based on its primary sequence data and analyses.

This genome is based on a highly probable homologous recombination between HAdV-D37 and HAdV-D8; however (probably after the initial recombination event), other parts of the genome have been replaced by genome parts from several known or unknown HAdVs. Altogether, we assume the occurrence of at least five major recombination events: (1) recombination of HAdV-D37 and HAdV-D8 (occurring between the end of the HAdV-D37 22K gene and the beginning of the HAdV-8 pVIII gene); (2) exchange with an unknown or unsequenced adenovirus from species HAdV-D from the beginning of the protein IX gene to the end of the pTP gene; (3) replacement of the pVII and protein V genes with the same genome fragment from an unknown or unsequenced adenovirus from species HAdV-D; (4) exchange of L1 and L2 between HAdV-D22 and the recombinant virus; (5) and replacement of the 18.4K gene from an unknown source.

Here we presented evidence that the neutralization epitope ε of HAdV-D53, highly homologous to HAdV-D22, was generated by two recombination events which brought about the complete exchange of L1 and L2. This phenomenon is apparent in three other adenoviruses from HAdV-D (HAdV-D13, -D32, and -D39), as well as HAdV-B16, which is a member of species HAdV-B [31]. This detailed analysis at the complete genome level demonstrates that recombination may be a common event within adenoviruses, especially in species HAdV-D, as a general mechanism driving molecular evolution and immunogenicity. The neutralization epitope is framed by highly conserved sequences, which are also used for generic detection of most HAdVs by PCR [32,33,34]. These conserved sequences allow homologous recombination when a cell is infected with two different adenovirus types. Our results demonstrate that within HAdV-D, the neutralization epitopes ε are exchangeable in nature leading to immune escape of a highly virulent and prevalent HAdV type. This resembles the antigenic shift mechanism of influenza A viruses which is caused by reassortment, a more efficient way of gene transfer.

To date, this is the first fully sequenced recombinant adenovirus to be associated with EKC. Bootscan analysis showed that several regions of HAdV-D53 (IVa2, DNA polymerase, pTP, pVII, V, and 18.4K) were dissimilar to any known adenovirus. These sequences are either from an undiscovered adenovirus or a known yet unsequenced HAdV-D isolate. Additional whole genome sequencing studies of adenoviruses will shed light on this important question.

In light of its association with EKC, it seems significant that experimental corneal infection with HAdV-D53 induced inflammation, while infection with HAdV-D22, a virus not associated with EKC but highly related to HAdV-D53, did not. Those areas of the genome unique to EKC-causing viruses represent likely sources of corneal tropism. Full genome sequencing, bioinformatics analysis, and genome wide comparisons between EKC and non-EKC inducing HAdV-D strains are beginning to yield clues to corneal tropism and pathogenesis [8,9]. Further experiments recombining different adenovirus genes will determine which genes are crucial for EKC.

Early genotyping of HAdV-D53 by sequencing of the hexon (the major neutralization determinant) and other determinants (fiber and penton) gave results of a recombinant strain HAdV-D22/H8 [19]. Thus, HAdV-D53 fulfilled the hexon L1 and L2 criteria for typing as HAdV-D22 [18], with a fiber knob (hemagglutination determinant) sequence identical to HAdV-D8. In contrast to the classical concept of a recombinant strain, HAdV-D53 was cross reactive with a HAdV-D8 specific antiserum (Table 4). This confirms that some of the neutralization antibodies in the HAdV-D8 antiserum bind to the HAdV-D8-like fiber of HAdV-D53 and block infectivity by interfering with virus/primary cellular receptor interaction.

Phylogenetic analysis of the complete genomic sequence of HAdV-D53 showed similar genetic distances to the other available HAdV-D types (6.1% to 9.3% nucleic acid sequence divergence) as observed between other prototypes of species HAdV-D (6.0% to 9.5%) (Fig. 5). This supports the idea that HAdV-D53 is the prototype of a new genotype. Therefore, phylogeny deduced from complete genomic sequence data supports that HAdV-D53 is a new prototype. However, HAdV-D53 is a recombinant virus and its genome is not of monophyletic origin. For most parts of its genome the ancestors of its sequence (HAdV-D8, -D22, -D37) could be identified by bootscan analysis and confirmed by building phylogenetic trees of the corresponding sequence stretches. For example, L1 and L2 of the neutralization determinant ε are highly variable and evolved rapidly by immune escape mechanisms. L1 and L2 of HAdV-D53 were (except for a single point mutation) identical to HAdV-D22 suggesting a recent recombination event in the phylogeny of HAdV-D53. However, bootscan analysis suggested that several regions of HAdV-D53 (IX, IVa2, DNA polymerase, pTP, pVII, protein V, and 18.4K) were dissimilar to all

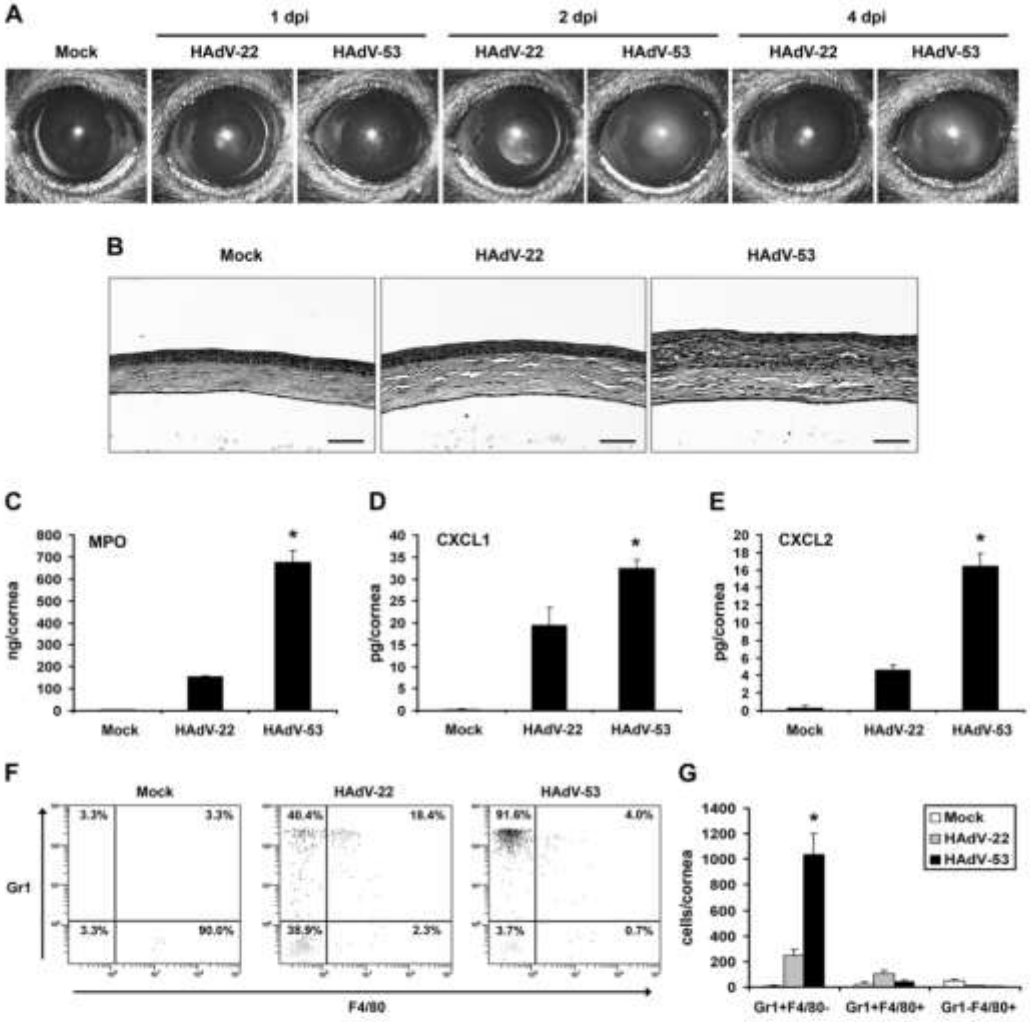


Figure 4. HAdV-D53 induces keratitis. (A) Clinical appearance of HAdV-D53 keratitis. Virus-free buffer (mock), 10^4 TCID of HAdV-D22, or HAdV-D53 was injected in the corneal stroma of C57BL/6 mice (n = 8 corneas/group). Corneas were examined under a surgical microscope up to 4 days post-infection. One representative picture from each group is shown at the indicated time points. (B) Histopathology of HAdV-D53 keratitis. Representative histopathological sections at 4 days post-infection of mouse corneas injected with buffer, HAdV-D22, or HAdV-D53 are shown (hematoxylin and eosin stain; scale bar = 50 μ). (C) Myeloperoxidase (MPO) expression in HAdV-D53 keratitis. Mock, HAdV-D22, and HAdV-D53 infected corneas were analyzed by ELISA at 24 hours post-infection for the expression of myeloperoxidase enzyme. (D, E) Chemokine expression in HAdV-D53 keratitis. Expression of CXCL1 (D) and CXCL2 (E) in mock, HAdV-D22, and HAdV-D53 infected corneas were analyzed by ELISA at 16 hpi. Data is mean ± SEM from three individual experiments (n = 9 corneas/group). (F, G) Phenotypic analysis of inflammatory cells in HAdV-D53 keratitis. Mock, HAdV-D22, and HAdV-D53 infected corneas at 24 hours post-infection were homogenized and single cell preparations were stained with anti-CD45, anti-Gr1, and anti-F4/80 antibodies. Cells were gated on CD45-positive staining. (F) Representative dot plots or (G) quantification of three separate experiments is shown for each group (mean cells/cornea ± SEM, n = 9 corneas/group). In all experiments statistical significance is denoted by *, P<.05 as determined by ANOVA with Scheffe's multiple comparison test.
doi:10.1371/journal.pone.0005635.g004

known adenoviruses. Construction of phylogenetic trees supported that these parts of the genome are either from an undiscovered adenovirus or a known yet unsequenced HAdV-D isolate. However, these genome regions are well conserved in HAdV-D and thus led to low, non significant bootstrap values (see Fig. 5 polymerase, protein V and pVII). Additional whole genome sequencing studies of adenovirus prototypes may elucidate whether some of these parts of the HAdV-D53 genome are also derived from recombination

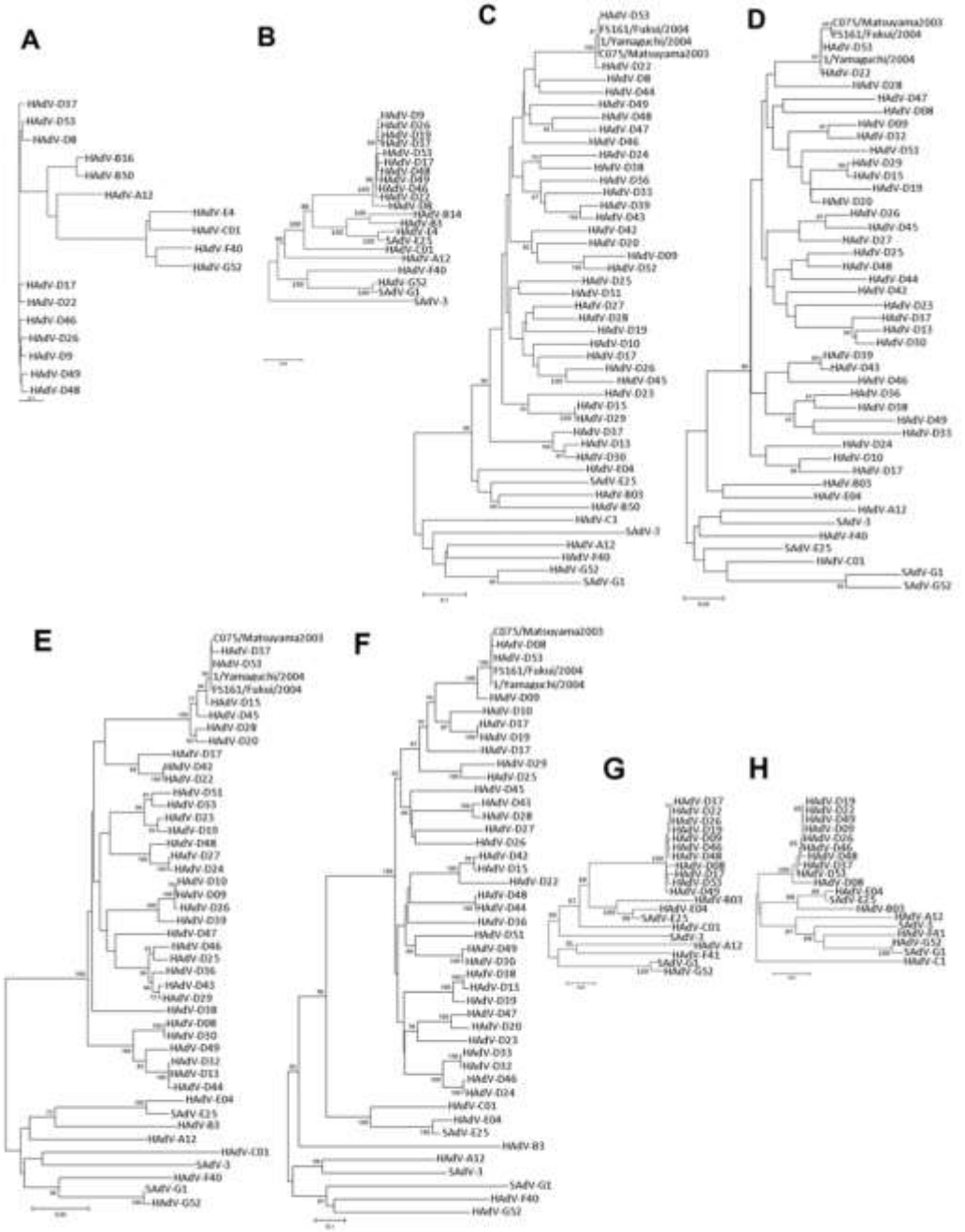


Figure 5. Phylogenetic analysis of HAdV-D53. Analysis of HAdV-D53 is based on the nucleic acid sequence of (**A**) complete genomes, as well as the predicted amino acid sequences of (**B**) polymerase, (**C**) L1 and (**D**) L2 of the hexon protein penton, (**E**) β-determinant, (**F**) γ-determinant, (G) pV and (**H**) pVII. Numbers denote human adenovirus serotypes. HAdV-D53 (in bold) shows the new isolate. The numbers close to the nodes represents bootstrap pseudoreplicates.
doi:10.1371/journal.pone.0005635.g005

Table 4. Neutralization of HAdV-D53 with hyper immune serum.

Antiserum	HAdV-D53	HAdV-D8	HAdV-D22	HAdV-D37
αHAdV-D8	1/128	1/1024	<1/8	<1/8
αHAdV-D22	1/256	<1/8	1/128	<1/8
αHAdV-D37	<1/8	<1/8	<1/8	1/4096

doi:10.1371/journal.pone.0005635.t004

events. Interestingly, protein V, a minor capsid protein, was significantly smaller than the homologous proteins of all other members of HAdV-D (e.g. 297 aa vs. 334 aa in HAdV-D46). Moreover, pVII also contained several deletions, nevertheless phylogenetic trees clearly supported clustering of HAdV-D53 protein V and pVII with species HAdV-D in spite of these deletions.

The 5′-ITR sequence contains highly conserved critical motifs that are required for adenovirus replication [37]. These motifs include the canonical 'core origin,' defined as the minimal DNA requirement for the initiation of replication, binding the terminal protein-DNA polymerase complex [38], and several host transcription factor binding sequences which are required for efficient adenovirus replication [39,40]. For example, it has been shown that Oct-1 binds to the NF III motif to stimulate transcription by 6–8 fold [41]. Within most HAdV species, both NF I and NF III binding sites are conserved except for species HAdV-E as seen in HAdV-E4 and HAdV-E4 vaccine strain [42,43] and simian AdV-21 (SAdV-B21), SAdV-E22 through E25 (unpublished observations) which lack the NF I binding site. Significantly, HAdV-D53 is also missing the NF I motif, like the other members of the sequenced HAdV-D types (Fig. 6). Previous annotations of the sequenced HAdV-D members do not remark upon this absence of the NF I. Perhaps this absence of an NF I site is an indication of a different evolutionary line of origin for species HAdV-D, as opposed to the other HAdVs with both NF I and NF III motifs. The latter half of the ITR contains motifs for binding Sp1 (GGG GGT GG) and ATF (TGA CGT). These motifs are also reported to contribute to the efficiency of viral DNA replication [44]. While the Sp1 motif seems to be less conserved (GGG CGg/t gg), they are similar for HAdV-D types. The ATF motif is conserved and present in HAdV-D (Fig. 6).

As new strains of adenoviruses appear and are isolated, usually with an accompanying pathology, initial attempts at understanding the clinical relevance involves characterizing the isolate with respect to structural features. These include the traditional serological methods and reagents. However, in some cases the isolates are difficult to culture and/or the reagents are not readily available. In the past, the isolate is either characterized as much as possible or archived in a laboratory as an unculturable, yet interesting isolate. Today, when an interesting adenovirus isolate arises, full-genome sequencing, phylogenetic analysis, and other state-of-the-art methodology and technology provide alternatives to these limitations. As a recent example, when HAdV-G52 was discovered, it was found that the virus grew too slowly in tissue culture to be 'properly' serotyped. This and the lack of readily available serotyping reagents limited a 'traditional' characterization. However, phylogenetic analysis, only made possible through whole genome sequencing, demonstrated that it was a novel adenovirus isolate that was quite divergent from all other species of human adenoviruses [6]. Similarly, if serology and limited sequence analysis, e.g., limited hexon, penton and fiber data, were the only tools that we had available for the original characterization of this proposed HAdV-D53, the reported original conclusion in regards to HAdV-D53, that it is a variant of HAdV-D22 albeit with minor genetic modifications in the penton and fiber genes [19], would have been and remained incorrect. In order to conclusively characterize a suspected novel adenovirus, whole genome sequencing and bioinformatics analysis of the resultant and complete reference primary nucleotide sequence should be performed.

The fact that the genes associated with serum neutralization are from known viruses raises a central question, "What are the criteria for defining and naming a new "type" of adenovirus?" Although serology has been crucial in the pre-genomic era, it can not be used as the gold-standard for the typing of novel adenoviruses that will be sequenced and characterized in the future. If serology was the only tool that we had in our typing toolbox, we would not have determined that HAdV-D53 was due to several recombinations of known and perhaps unknown adenoviruses. In the past, the "serotype" designation was used to distinguish different and separate adenoviruses. However, due to the fact that there are about 200 known adenovirus types, this approach is impractical. Moreover, the neutralization of recombinants such as HAdV-B16, and -D53 would yield inconclusive data. Full-genome sequencing and bioinformatic analyses should be the primary methods used when proclaiming novel adenovirus genotypes as it is quicker and a less cumbersome alternative for adenovirus typing, especially given the cost-effective technology to obtain genome sequences rapidly and the growing array of bioinformatics tools, along with the growing adenovirus database.

We propose using "genotype" rather than "serotype" as a means for identifying, characterizing and differentiating adenoviruses, based on genome sequence analyses. This fits into the currently accepted classification of adenovirus "genome types," in which substrains of adenoviruses are designated by lower case alphabetic designations in addition to their primary designation, e.g., HAdV-7a, b, c…, if their restriction enzyme digestion patterns differ from the reference prototype genome, "HAdV-7p."

Recently, partial genome sequences from HAdV-D strains causing EKC outbreaks in Japan were published [35,36]. These were almost identical to HAdV-D53 (including the intrahexon recombination sites) suggesting that HAdV-D53 has already spread around the globe as an emerging EKC agent, reflecting the epidemiology of a globally connected population and a newly emergent pathogen.

Materials and Methods

Ethics Statement

The animals involved in this study were procured, maintained, and used in accordance with the Laboratory Animal Welfare Act of 1966, as amended, and NIH 80-23, Guide for the Care and Use of Laboratory Animals, National Research Council.

Nucleotide sequence accession numbers

The HAdV-D53 genome and annotation have been deposited in GenBank prior to manuscript submission; accession number FJ169625. The following HAdV genomes (GenBank accession numbers) were used: HAdV-A12 (AC_000005), HAdV-B7 (AY594255), HAdV-D8 (AB110079), HAdV-B11 (AY163756), HAdV-C5 (AC_000008), HAdV-E4 (AY599837), HAdV-D49 (DQ393829), HAdV-D53 (FJ169625), HAdV-D9 (AJ854486), HAdV-B16 (AY601636), HAdV-D17 (AC_000006), HAdV-D19 (ER121005), HAdV-D22 (FJ404771) HAdV-D26 (EF153474), HAdV-D37 (DQ900900), HAdV-D46 (AY875648), HAdV-D48 (EF153473), HAdV-D22 (unpublished genome sequence), HAdV-D8 (published partial sequences (AB110079) and unpublished whole genome sequence).

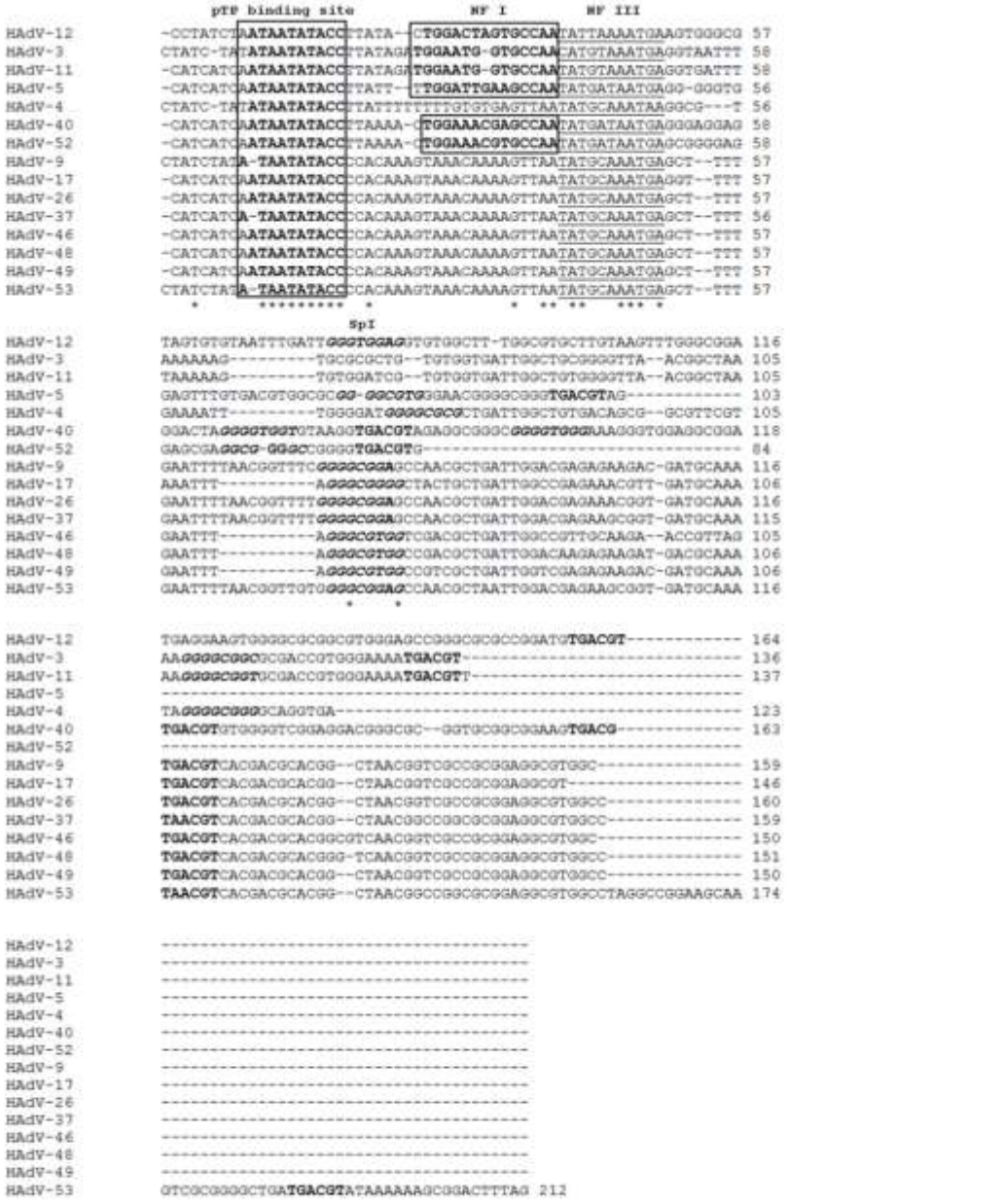


Figure 6. Analysis of the HAdV-D53 inverted terminal repeat (ITR). NF I, NF III, SpI, and pTP binding motifs are marked. The ATF binding site is TGACGT.
doi:10.1371/journal.pone.0005635.g006

Amplification of the HAdV-D53 genome

To amplify regions of HAdV-D53 flanking the sequences described by Engelmann et al. [19], we designed primers based on conserved adenovirus sequences of types in HAdV-D. All amplicons were then sequenced using primer walking.

Viruses, cells and neutralization test

Viral neutralization assays were run as previously described [45]. Rabbit antisera to prototype strains were standardized in cross-neutralization tests against adenovirus prototype viruses 1–49. Prototype viruses were from archives maintained at the State of California, Department of Public Health, Viral and Rickettsial Disease Laboratory.

Nucleic Acid Isolation

Viral DNA was extracted from tissue culture and processed stool samples using the MagNA Pure LC DNA Isolation Kit I (Roche, Indianapolis, IN) according to the manufacturers' recommendations for the MagNA Pure LC automated nucleic acid extraction system.

Bioinformatics

Percent identities for HAdV-53 genes/proteins. The global alignment were performed using the EMBOSS [46] needle program. The proteins and genes of HAdV-53 were compared to homologs in other HAdV-D genomes. In cases were a genome lacked sufficient annotation, genes and proteins were found manually using the Artemis [47] annotation program. The percent identities for the proteins (Table 2) of the HAdV-D sequences were obtained via BLASTP [48]. The percent identities for the nucleotide sequences (Table 2) that code for these proteins were determined using a BioJava [49] implementation of a Needleman-Wunsch algorithm.

Recombination analysis of hexon genes (Figure 3). Hexons genes from the HAdV-D genomes were aligned using ClustalW [50] alignment option available in the MEGA 4 program [51]. The default gap opening and gap extension penalties were used (15.0 and 6.66). SimPlot [20] software was used to complete a bootscan analysis of the aligned hexon genes of the available HAdV-D genomes. 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-53 hexon was chosen as the reference sequence for the analysis.

Recombination analysis of HAdV-D whole genomes (Figure 2). The available HAdV-D genomes were aligned using the MAFFT [52] alignment method which is available through a web interface at http://www.ebi.ac.uk/Tools/mafft/. The default parameters for gap open penalty, gap extension penalty, and perform fft were used (1.53, 0.12, "localpair").

Simplot [20] software was used to complete a bootscan analysis of the aligned HAdV-D genomes. The default parameters for window size and step size were altered (1000, 200). All other default parameters were left unchanged.

Recombination Analysis

Two groups of hexon coding nucleotide sequences were analyzed for recombination events. The first group consisted of the hexon genes of the human adenovirus D species (HAdV-D8, -D9, -D17, -D22, -D26, -D37, -D46, -D48, -D49, -D53). This group is referred to as the HAdV-D53 hexon group. The second group consisted of hexon genes from HAdV-B16, -C5, -E4, -B7, -B11, and -C2. The following accession numbers were used for the hexon recombination analyses. HAdV-A12 (AC_000005), HAdV-B7 (AY594255), HAdV-B11 (AY163756), HAdV-C5 (AC_000008), HAdV-E4 (AY599837), HAdV-D49 (DQ393829), HAdV-D22, (AB330103), HAdV-D53 (FJ169625), HAdV-D9 (AJ854486), HAdV-16/B1 (AY601636), HAdV-D17 (AC_000006), HAdV-D26 (EF153474), HAdV-D37 (DQ900900), HAdV-D46 (AY875648), HAdV-D48 (EF153473). The two groups of sequences were aligned using the ClustalW [50] alignment option available in the MEGA 4 program [51]. The default gap opening and gap extension penalties were used. Those penalties were 15.0 and 6.66 respectively.

Two different programs were used to analyze the two alignments for recombination events. The first program is SimPlot [20]. The bootscan option of SimPlot was used to analyze the alignments. The default settings were used. These included a window size = 200, a step size = 20, replicates used = 100, gap stripping = "on", distance model = "Kimura", tree model = "-Neighbor Joining". The HAdV-D53 hexon was chosen as the reference sequence HAdV-D53 hexon group. HAdV-D16's hexon was chosen as the reference in the HAdV-16 hexon group.

The second program is the Recombination Detection Program (RDP) [53]. This program uses several different algorithms (including bootscanning) to determine the presence of recombination events. 1 of the "general recombination detection options" was changed so that the program would recognize that the sequences in the alignment were linear and not circular. No other default options were changed.

Phylogenetic analysis of HAdV-D53

DNA polymerase, penton base (β-determinant), pVII, protein V, L1 and L2 of the hexon, and fiber knob (γ-determinant) nucleotide sequences were compared by sequential pairwise alignment with the Clustal Algorithm implemented in the BioEdit software package (version 6.0.5) and adjusted manually to conform to the optimized alignment of deduced amino acid sequences. Phylogenetic relationships were inferred from the aligned nucleic acid as well as from the amino acid sequences by the neighbour-joining method implemented in the programs DNAdist and Neighbor integrated in the MEGA software package (version 3.1) using the Kimura two-parameter substitution model and a transition/transversion ratio of 10. Support for specific tree topologies was estimated by bootstrap analysis with 1000 pseudoreplicate data sets.

In vivo model of adenovirus keratitis

Eight to 12 week old C57BL/6J mice (stock # 000664) were purchased from Jackson Laboratory (Bar Harbor, ME). Animal housing and care were in accordance with Animal Care and Use Committee guidelines. Mice were anesthetized for virus infection by intramuscular injection of ketamine (85 mg/kg) and xylazine (14 mg/kg) and later euthanized by CO_2 inhalation. For infection, 1 microliter of virus-free dialysis buffer, cesium chloride gradient purified HAdV-D22, or purified HAdV-D53 (10^8 tissue culture infectious dose) was injected in the central corneal stroma as previously described [21]. After euthanasia, corneas were removed and fixed in 10% neutral buffered formalin, embedded in paraffin, and sections cut at 5 μ thick prior to staining. For ELISA, corneas were harvested at indicated time points and homogenized using phosphate buffered saline (PBS) with protease inhibitors, and the reactions performed as per manufacturer's instructions (R&D Systems, Minneapolis, MN). ELISA plates were analyzed on a microplate reader (Molecular Devices, Sunnyvale, CA) with limits of detection of <2 pg/mL for CXCL1 and <1.5 pg/mL for CXCL2. Flow cytometry was performed as described by Carr and coworkers [54]. Corneas were dissected at indicated time points, and digested with 1 mg/ml collagenase type I (Sigma, St. Louis,

MO). Non-specific binding was blocked by anti-mouse Fc (BD Pharmingen, San Diego, CA) and 5% normal rat serum (Jackson Immuno Research, West Grove, PA). Cells were labeled with FITC-conjugated anti-mouse F4/80 (clone CEA3-1), phycoerythrin-Cy5-conjugated anti-CD45 (clone 30-F11), and PE-conjugated anti-mouse Gr-1 (clone RB6-8C5) (all from BD Biosciences, San Jose, CA), and incubated in the dark on ice for 30 min, washed 3× with PBS/1% BSA, resuspended in PBS containing 1% paraformaldehyde, and incubated overnight. CountBright absolute counting beads (Invitrogen, Eugene, OR) were added (21,600 beads/sample), cell suspensions gated on $CD45^{high}$ labeled cells, and the numbers of each cell type determined at this gate setting. A second gate was established to count the number of beads that passed through during each run (300 sec). The absolute number of cells per cornea was determined by calculating the number of input beads/21,600× number of cells in the $CD45^{high}$-gated sample.

Acknowledgments

We thank Sarah Torres and Lisa Thrasher for help with sequencing HAdV-D8 and HAdV-D22.

The animals involved in this study were procured, maintained, and used in accordance with the Laboratory Animal Welfare Act of 1966, as amended, and NIH 80-23, Guide for the Care and Use of Laboratory Animals, National Research Council.

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: IM DS AH JC DS MSJ. Performed the experiments: MPW AVC CMR IM NRH. Analyzed the data: MPW AVC CMR IM BH DS AH JC DS MSJ. Contributed reagents/materials/analysis tools: DS AH JC DS MSJ. Wrote the paper: MPW AVC CMR IM BH DS AH JC DS MSJ.

References

1. Gordon YJ, Araki K, Kinchington PR (1996) Adenovirus keratoconjunctivitis. In: Pepose JS, Holland GN, Wilhelmus KR, eds. Ocular Infection and Immunity. St. Louis: Mosby. pp 877–94.
2. Rowe WP, Huebner RJ, Gilmore LK, Parrott 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.
3. Hilleman MR, Werner JH (1954) Recovery of new agent from patients with acute respiratory illness. Proc Soc Exp Biol Med 851: 183–188.
4. Ginsberg HS, Gold E, Jordan WS Jr, Katz S, Badger GF, Dingle JH (1955) Relation of the new respiratory agents to acute respiratory diseases. Am J Pub Health 45: 915–922.
5. Dingle JH, Langmuir AD (1968) Epidemiology of acute, respiratory disease in military recruits. Am Rev Respir Dis 976: Suppl 1–65.
6. Jones MS, Harrach B, Ganac RD, Gozum MMA, dela Cruz WP, et al. (2007) New adenovirus species found in patient presenting with gastroenteritis. J Virol 81: 5978–5984.
7. Louie JK, Kajon AE, Holodniy M, Guardia-LaBar L, Lee B, et al. (2008) Severe pneumonia due to adenovirus serotype 14: a new respiratory threat? Clin Infect Dis 46: 421–425.
8. Benkő M, Harrach B, Both GW, Russell WC, Adair BM, et al. (2005) Family *Adenoviridae*. In: Fauquet CM, Mayo MA, Maniloff J, Desselberger U, Ball LA, eds. Virus Taxonomy. VIIIth Report of the International Committee on Taxonomy of Viruses. New York: Elsevier. pp 213–228.
9. Robinson CM, Shariati F, Gillaspy AF, Dyer DW, Chodosh J (2008) Genomic and bioinformatics analysis of human adenovirus type 37: new insights into corneal tropism. BMC Genomics 9: 213.
10. Robinson CM, Shariati F, Zaitchik J, Gillaspy AF, Dyer DW, Chodosh J (2009) Human adenovirus type 19: Genomic and bioinformatics analysis of a keratoconjunctivitis isolate. Virus Res 139: 122–129.
11. Dhurandhar NV, Israel BA, Kolesar JM, Mayhew GF, Cook ME, Atkinson RL (2000) Int J Obes Relat Metab Disord 24: 989–996.
12. Davison AJ, Benkő M, Harrach B (2003) Genetic content and evolution of adenoviruses. J Gen Virol 84: 2895–2908.
13. Lauer KP, Llorente I, Blair E, Seto J, Kramov V, et al. (2004) Natural variation among human adenoviruses: genome sequence and annotation of human adenovirus serotype 1. J Gen Virol 85: 2615–2625.
14. Kovács GM, Harrach B, Zakhartchouk AN, Davison AJ (2005) The complete genome sequence of simian adenovirus 1 – an Old World monkey adenovirus with two fiber genes. J Gen Virol 86: 1681–1686.
15. Harrach B, Benkő M (1998) Phylogenetic analysis of adenovirus sequences. Proof of the necessity of establishing a third genus in the Adenoviridae family. In: Wold WSM, ed. Adenovirus Methods and Protocols. vol 21. Totowa: Humana Press Inc. pp 309–339.
16. Benkő M, Harrach B (2003) Molecular evolution of adenoviruses. In: Doerfler W, Böhm P, eds. Adenoviruses: Model and Vectors in Virus-Host Interactions. Current Topics in Microbiology and Immunology, vol. 272. Berlin: Springer. pp 3–35.
17. Kovács GM, Davison AJ, Zakhartchouk AN, Harrach B (2004) Analysis of the first complete genome sequence of an Old World monkey adenovirus reveals a lineage distinct from the six human adenovirus species. J Gen Virol 85: 2799–2807.
18. Madisch I, Harste 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.
19. Engelmann I, Madisch I, Pommer H, Heim A (2006) An outbreak of epidemic keratoconjunctivitis caused by a new intermediate adenovirus 22/H8 identified by molecular typing. Clin Infect Dis 43: 64–66.
20. Lole KS, Bollinger RC, Paranjape RS, Gadkari 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.
21. Chintalanellawar AV, Astley R, Chodosh J (2007) Adenovirus type 37 keratitis in the C57BL/6J mouse. Invest Ophthalmol Vis Sci 48: 781–788.
22. Shultz J, Kaminker K (1963) Myeloperoxidase of the leukocyte of normal human blood. Arch Biochem Biophys 96: 465–467.
23. Bos A, Wever R, Roos D (1978) Characterization and quantification of the peroxidase in human monocytes. Biochim Biophys Acta 523: 37–48.
24. Hendal K, Ruscetti FW, Ihle JN, Jacobsen SE, Dubois CM, et al. (1991) Characterization and regulation of RB6-8C5 antigen expression on murine bone marrow cells. J Immunol 147: 22–28.
25. Del Rio L, Bennouna S, Salinas J, Denkers EY (2001) CXCR2 deficiency confers impaired neutrophil recruitment and increased susceptibility during *Toxoplasma gondii* infection. J Immunol 167: 6503–6509.
26. Geissmann F, Jung S, Littman DR (2003) Blood monocytes consist of two principal subsets with distinct migratory properties. Immunity 19: 71–82.
27. Gordon S, Taylor PR (2005) Monocyte and macrophage heterogeneity. Nat Rev Immunol 5: 953–964.
28. Hirsch S, Austyn JM, Gordon S (1981) Expression of the macrophage-specific antigen F4/80 during differentiation of mouse bone marrow cells in culture. The Journal of Experimental Medicine 154: 713–725.
29. Bozic CR, Kolakowski LF Jr, Gerard NP, Garcia-Rodriguez C, von Uexkull-Guldenband C, et al. (1995) Expression and biologic characterization of the murine chemokine KC. J Immunol 154: 6048–6057.
30. Wolpe SD, Sherry B, Juers D, Davatelis G, Yurt RW, Cerami A (1989) Identification and characterization of macrophage inflammatory protein 2. Proc Natl Acad Sci U S A 86: 612–616.
31. Ebner K, Pinsker W, Lion T (2005) Comparative sequence analysis of the hexon gene in the entire spectrum of human adenovirus serotypes: phylogenetic, taxonomic, and clinical implications. J Virol 79: 12635–12642. Erratum in: J Virol 80: 2586 (2006).
32. Allard A, Girones R, Juto P, Wadell G (1990) Polymerase chain reaction for detection of adenoviruses in stool samples. J Clin Microbiol 28: 2659–2667. Erratum in: J Clin Microbiol 29: 2683 (1991).
33. Kiss I, Matiz K, Allard A, Wadell G, Benkő M (1996) Detection of homologous DNA sequences in animal adenoviruses by polymerase chain reaction. Acta Vet Hung 44: 243–251.
34. Heim A, Ebnet C, Harste G, Pring-Akerblom P (2003) Rapid and quantitative detection of human adenovirus DNA by real-time PCR. J Med Virol 70: 228–239. Erratum in: J Med Virol 71: 320 (2003).
35. Adhikary AK, Inada T, Banik U, Mukoyama A, Ikeda Y, et al. (2004) Serological and genetic characterization of a unique strain of adenovirus involved in an outbreak of epidemic keratoconjunctivitis. J Clin Pathol 57: 411–416.
36. Aoki K, Ishiko H, Konno T, Shimada Y, Hayashi A, et al. (2008) Epidemic keratoconjunctivitis due to the novel hexon-chimeric-intermediate 22,37/H8 human adenovirus. J Clin Microbiol 46: 3259–3269.
37. Dán A, Ehó P, Harrach B, Zádori Z, Benkő M (2001) Four new inverted terminal repeat sequences from bovine adenoviruses reveal striking differences in the length and content of the ITRs. Virus Genes 22: 173–179.
38. Temperley SM, Hay RT (1992) Recognition of the adenovirus type 2 origin of DNA replication by the virally encoded DNA polymerase and preterminal protein. EMBO J 11: 761–768.
39. Mul YM, Verrijzer CP, van der Vliet PC (1990) Transcription factors NFI and NFIII/oct-1 function independently, employing different mechanisms to enhance adenovirus DNA replication. J Virol 64: 5510–5518.

40. Hatfield L, Hearing P (1991) Redundant elements in the adenovirus type 5 inverted terminal repeat promote bidirectional transcription in vitro and are important for virus growth in vivo. Virology 184: 265–76.
41. Evans JD, Hearing P (2002) Adenovirus replication. In: Curiel DT, Douglas JT, eds. Adenoviral vectors for gene therapy. London: Academic Press. pp 59–70.
42. Purkayastha A, Ditty SE, Su J, McGraw J, Tedl L, et al. (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: 2559–2572.
43. Purkayastha A, Su J, McGraw J, Ditty SE, Hadfield TL, et al. (2005) Genomic and bioinformatics analyses of HAdV-4-vac 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: 3083–3094.
44. Hatfield L, Hearing P (1993) The NFIII/OCT-1 binding site stimulates adenovirus DNA replication in vivo and is functionally redundant with adjacent sequences. J Virol 67: 3931–3939.
45. Crawford-Miksza LK, Schnurr D (1994) Quantitative colorimetric microneutralization assay for characterization of adenoviruses. J Clin Microbiol 32: 2331–2334.
46. Rice P, Longden I, Bleasby A (2000) EMBOSS: The European Molecular Biology Open Software Suite. Trends in Genetics 16: 276–277.
47. Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P (2000) Artemis: sequence visualization and annotation. Bioinformatics 16: 944–945.
48. Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, et al. (2008) NCBI BLAST: a better web interface. Nucleic Acids Research 36(Web Server issue): W5–9.
49. Holland RCG, Down T, Pocock M, A. PrliÄ A, Huen D, James K (2008) BioJava: an open-source framework for bioinformatics. Bioinformatics 24: 2096–2097. doi: 10.1093/bioinformatics/btn397.
50. 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.
51. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Molecular Biology and Evolution 24: 1596–1599.
52. Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucl Acids Res 30: 3059–3066.
53. Martin D, Rybicki E (2000) RDP: detection of recombination amongst aligned sequences. Bioinformatics 16: 562–563.
54. Carr DJ, Wuest T, Ash J (2008) An increase in herpes simplex virus type 1 in the anterior segment of the eye is linked to a deficiency in NK cell infiltration in mice deficient in CXCR3. J Interferon Cytokine Res 28: 245–251.

Part 2

Computational Analysis of Human Adenovirus Type 22 Provides Evidence for Recombination among Species D Human Adenoviruses in the Penton Base Gene

JOURNAL OF VIROLOGY, Sept. 2009, p. 8980–8985 Vol. 83, No. 17
0022-538X/09/$08.00+0 doi:10.1128/JVI.00786-09

Computational Analysis of Human Adenovirus Type 22 Provides Evidence for Recombination among Species D Human Adenoviruses in the Penton Base Gene▽

Christopher M. Robinson,[1] Jaya Rajaiya,[1] Michael P. Walsh,[2] Donald Seto,[2] David W. Dyer,[3] Morris S. Jones,[4] and James Chodosh[1]*

Department of Ophthalmology, Howe Laboratory, Massachusetts Eye and Ear Infirmary, Harvard Medical School, Boston, Massachusetts[1]; Department of Bioinformatics and Computational Biology, George Mason University, Manassas, Virginia[2]; Department of Microbiology and Immunology, University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma[3]; and Clinical Investigation Facility, David Grant USAF Medical Center, Travis, California[4]

Received 17 April 2009/Accepted 18 June 2009

Recombination in human adenoviruses (HAdV) may confer virulence upon an otherwise nonvirulent strain. The genome sequence of species D HAdV type 22 (HAdV-D22) revealed evidence for recombination with HAdV-D19 and HAdV-D37 within the capsid penton base gene. Bootscan analysis demonstrated that recombination sites within the penton base gene frame the coding sequences for the two external hypervariable loops in the protein. A similar pattern of recombination was evident within other HAdV-D types but not other HAdV species. Further study of recombination among HAdVs is needed to better predict possible recombination events among wild-type viruses and adenoviral gene therapy vectors.

Adenoviruses were first isolated in 1953 (18, 31) and currently cause an array of human diseases. These include respiratory, gastrointestinal, and ocular surface infections, opportunistic infections in immune-deficient individuals, and possibly, obesity (9, 13, 14, 20, 40). Adenoviruses have also recently been used as gene therapy vectors (19, 35). Thus, while adenoviruses continue to cause significant morbidity and mortality in the human population, their existence also provides a potential benefit for the treatment of patients with an even broader range of ailments.

Since the characterization of the first human adenovirus (HAdV), 53 types have been identified and subsequently classified into seven species (A to G) on the basis of serology, restriction endonuclease digestion patterns, and to a lesser degree, genotyping. Recently, high-throughput sequencing technology has made whole-genome sequencing both rapid and affordable (27). However, the genomic sequences of 23 out of 53 HAdV types remain to be determined, with most of those within species D.

Species D HAdV type 22 (HAdV-D22) was originally isolated from a child in 1956 (3). While it is not clear what role HAdV-D22 plays in human disease, one report revealed a possible tropism for the eye (32). Recently, recombination with HAdV-D22 has been identified as the source of a novel HAdV, HAdV-D53, causing an outbreak of keratoconjunctivitis in Germany (37). HAdV-D22 recombination was also identified in a possible variant of HAdV-D53 that was isolated from a patient in Japan (2). Therefore, HAdV-D22 has shown the propensity to recombine with other viruses, with clinically important consequences. The emergence of new pathogenic HAdV genotypes, along with continued interest in HAdVs as vectors for human gene therapy, make adenovirus recombination a critically important issue.

HAdV-D22 was acquired from the American Type Culture Collection (Manassas, VA). The complete genome of the prototype strain AV-2711 (ATCC VR-1100) was sequenced on an Applied Biosystems (Foster City, CA) 3730 XL DNA sequencer in the Laboratory for Genomics and Bioinformatics at the University of Oklahoma Health Sciences Center using a previously described protocol (29). The sequence was validated by sequencing on an ABI SOLiD DNA sequencer. Sequences from both methodologies were 100% identical and provided 7,727-fold coverage for the genome.

The mVISTA Limited Area Global Alignment of Nucleotides (LAGAN) tool (http://genome.lbl.gov/vista/index.shtml) was used to align and compare the whole HAdV genomes (6) of HAdV-D22 and each of the other nine completely sequenced HAdV-Ds. Analysis revealed sequence diversity in the penton base, hexon, E3, and fiber open reading frames (Fig. 1). Surprisingly, comparison of HAdV-D22 to HAdV-D19 strain C (30) and HAdV-D37 (29) revealed considerable sequence conservation in the penton base gene (Fig. 1 and 2A and B).

Based on the sequence conservation seen in the penton base, Simplot 3.5.1 (http://sray.med.som.jhmi.edu/SCRoftware/simplot/) and Recombination Detection Program (RDP) version 3.34 (http://darwin.uvigo.es/rdp/rdp.html) were used to identify possible recombination sites (21, 24). Bootscan analysis identified two possible recombination loci in the HAdV-D22 penton base, the first in HAdV-D37, encompassing nucleotides 400 to 600, and the second in HAdV-D19 at nucleotides 750 to 1350 (Fig. 3A). In silico amino acid analysis showed that these two probable recombination areas code for the two variable loops in the penton base protein (Fig. 3A), located on the exterior of the viral capsid (16, 44).

* Corresponding author. Mailing address: Massachusetts Eye and Ear Infirmary, 243 Charles Street, Boston, MA 02114. Phone: (617) 573-6398. Fax: (617) 573-4324. E-mail: james_chodosh@meei.harvard.edu.

▽ Published ahead of print on 24 June 2009.

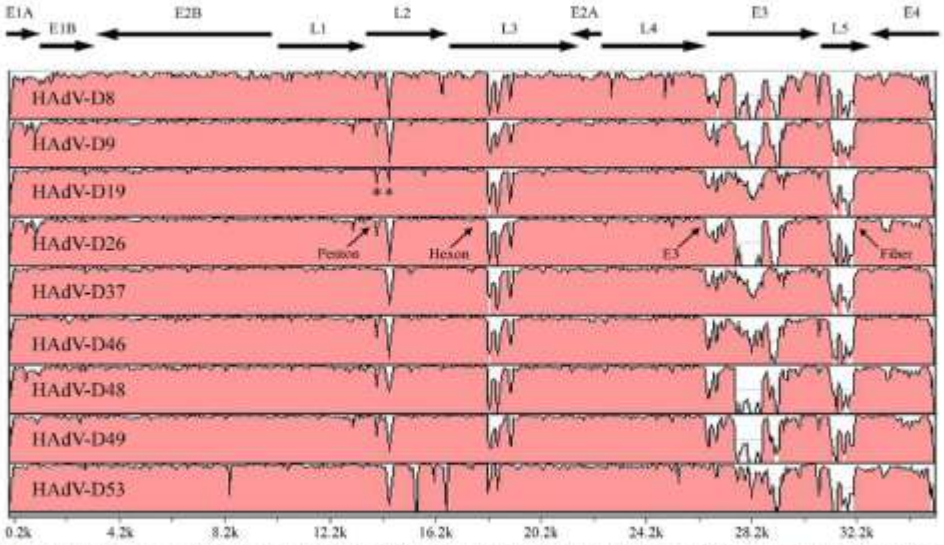


FIG. 1. Global pairwise sequence alignment of HAdV-D22 with nine other HAdV-D types. The percent sequence conservation is reflected in the height of each data point along the *y* axis. A conserved sequence in the penton base open reading frame is designated by an asterisk. GenBank accession numbers are as follows: HAdV-D8, AB448768; HAdV-D9, AJ854486; HAdV-D19, EF121005; HAdV-D26, EF153474; HAdV-D37, DQ900900; HAdV-D46, AY875648; HAdV-D48, EF153473; HAdV-D49, DQ393829; and HAdV-D53, FJ169625.

We extended our investigation to determine if these recombination sites were common to penton base genes of other HAdV genomes. We found evidence for recombination between HAdV-D9 and HAdV-D26 in both the nucleotide 400 to 600 and 750 to 1350 regions (Fig. 3B), between HAdV-D49 and HAdV-D48 in the nucleotide 400 to 600 region, and between HAdV-D49 and HAdV-D8 in the nucleotide 750 to 1350 region (Fig. 3C). A similar pattern was observed in one or both of

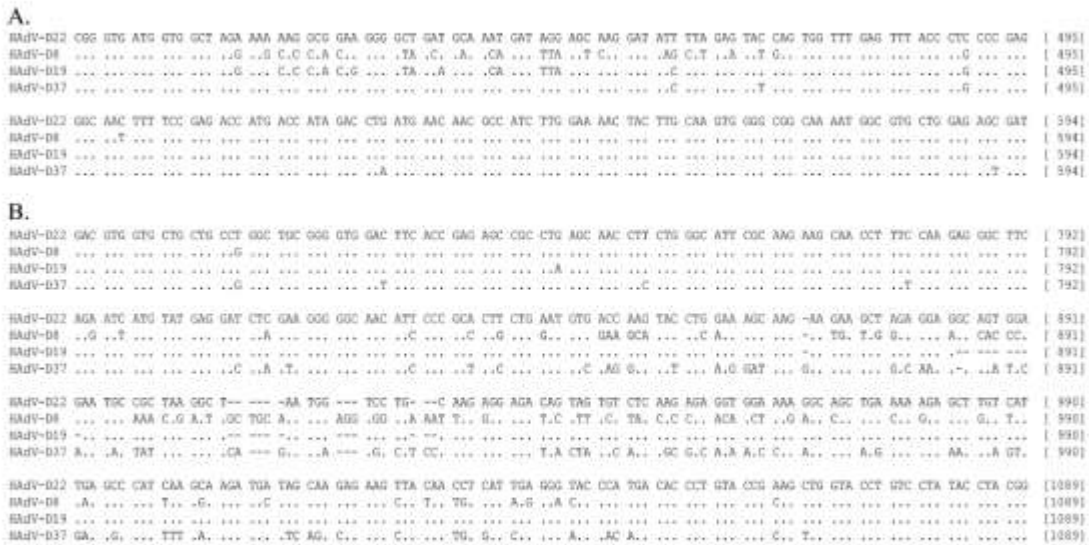
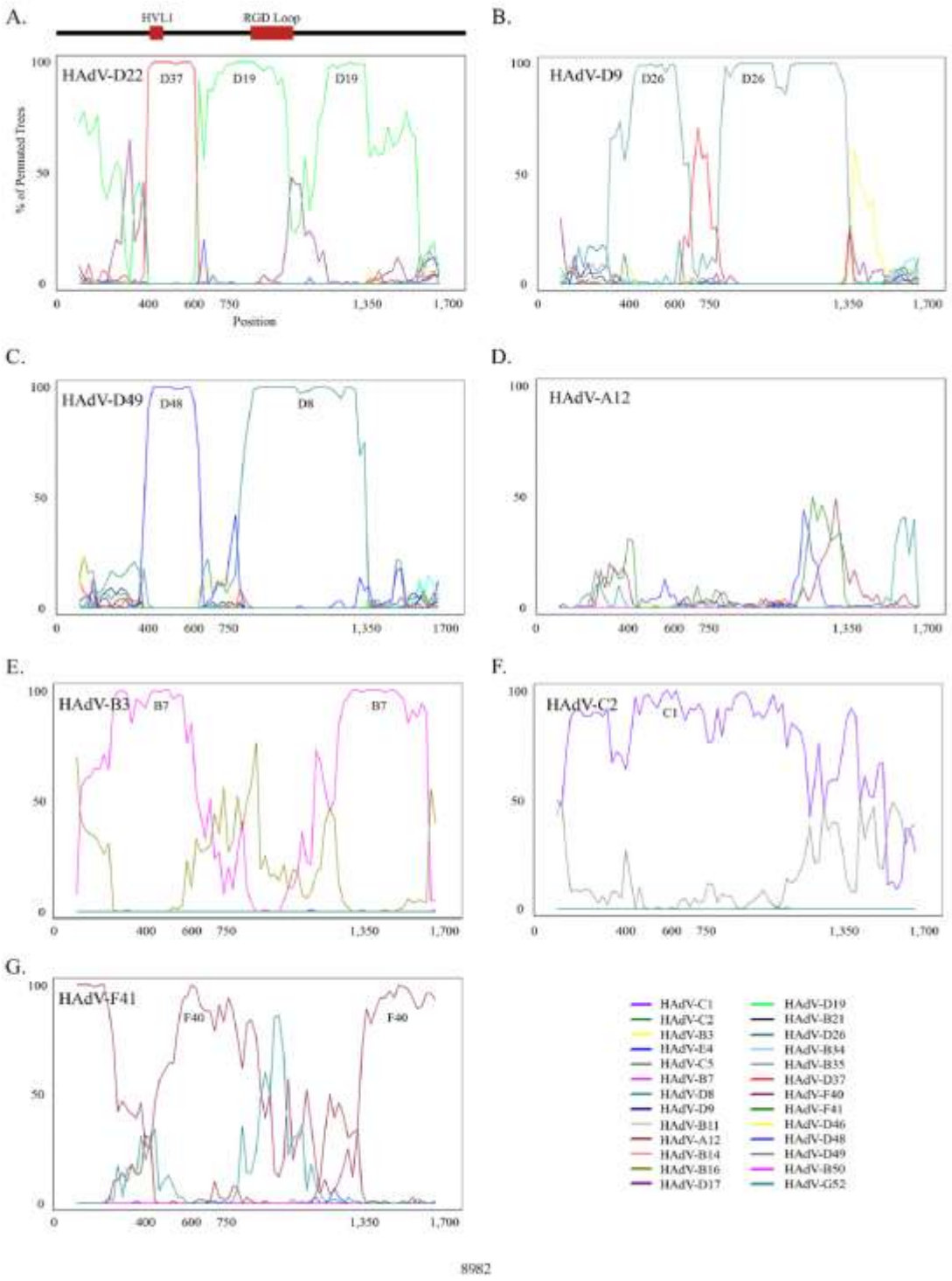


FIG. 2. Multisequence alignment of HAdV-D penton base genes. Nucleotide alignment of the bp 397 to 594 (A) and bp 694 to 1089 (B) regions of the penton base gene.



8982

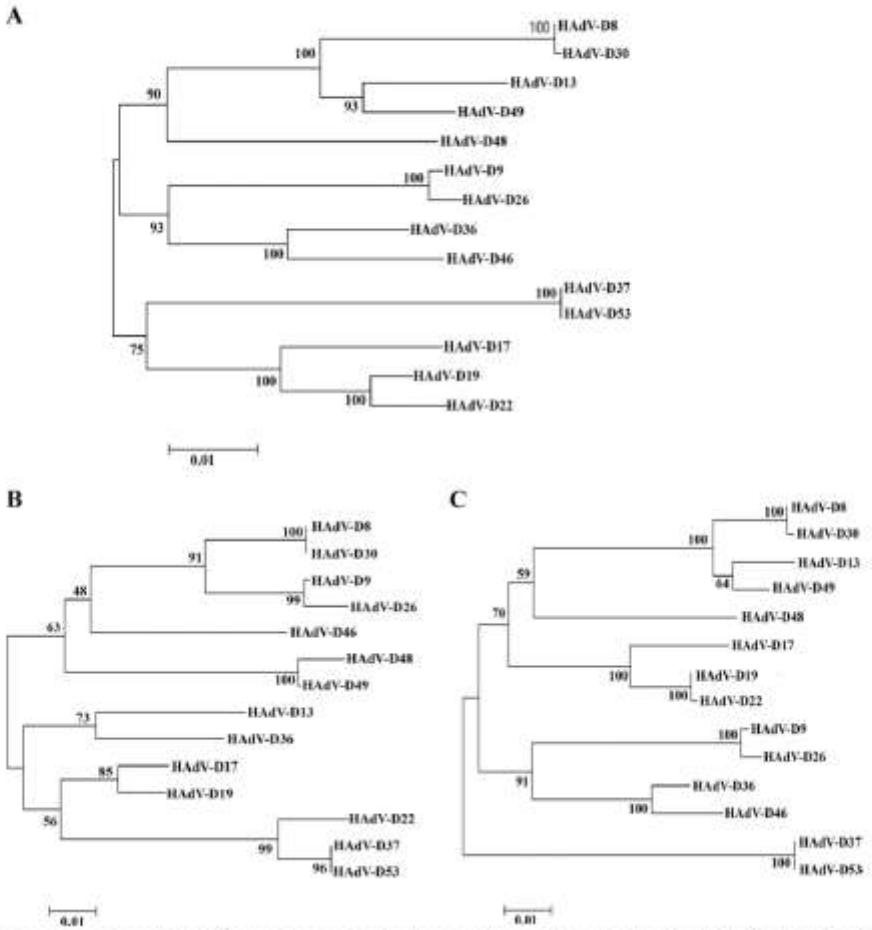


FIG. 4. Phylogenetic analysis of HAdV-D penton base genes. Bootstrap neighbor-joining tree designed using MEGA 4.0.2. The Gonnet protein weight matrix in ClustalX alignment was used, along with complete deletion options. Percent bootstrap confidence levels (1,000 replicates) are shown on the relevant branches. (A) Analysis of sequenced HAdV-D penton base genes. (B) Analysis of the nucleotide 400 to 600 region of sequenced HAdV-D penton base genes. (C) Analysis of the nucleotide 750 to 1350 region of sequenced HAdV-D penton base genes.

these nucleotide regions for HAdV-D8, -17, -19, -26, -37, and -48 (data not shown). Remarkably, this pattern of recombination in the penton base gene was unique to HAdV species D (Fig. 3D, E, F, and G).

Bootstrap-confirmed neighbor-joining phylogenetic trees of HAdV-D penton base genes were constructed with Molecular Evolutionary Genetics Analysis (MEGA) 4.0.2 (http://www.megasoftware.net/index.html) to examine viral evolution in HAdV-D (34). Analysis of the entire penton gene revealed a close relationship of HAdV-D22 and HAdV-D19 strain C (Fig. 4A) (30). Additional phylogenetic trees were constructed to encompass the two proposed recombination sites within this

FIG. 3. Bootscan analysis of HAdV-D penton base genes. Comparison of HAdV-D22 (A) with completely sequenced HAdV types. Similar comparisons of the same region were performed with HAdV-D9 (B), HAdV-D49 (C), HAdV-A12 (D), HAdV-B3 (E), HAdV-C2 (F), or HAdV-F41 (G) as the reference type. HAdV-D53 was left out of the analysis due to the 100% identity of its penton base to that of HAdV-D37. The axes of all panels are as labeled in panel A. GenBank accession numbers are as follows: HAdV-A12, NC_001460; HAdV-B3, DQ086466; HAdV-C2, AC_000007; and HAdV-F41, DQ315364.

gene. Phylogenetic analysis of the nucleotide 400 to 600 region revealed a close identity among HAdV-D22, HAdV-D37, and HAdV-D53 (Fig. 4B). Analysis of the nucleotide 750 to 1350 region of the penton base gene revealed a close identity of HAdV-D22 and HAdV-D19 (Fig. 4C).

In summary, comparison of the HAdV-D22 genome to other sequenced HAdV-D genomes identified substantial sequence divergence in the penton, hexon, E3, and fiber open reading frames. Differences between HAdV-D genomes in these areas have previously been described by us in reports on HAdV-D19 strain C and HAdV-D37 genomes (29, 30). However, the sequence conservation among HAdV-D22, HAdV-D19, and HAdV-D37 in the penton base gene was unexpected and suggests recombination among these viruses. Bootscan analysis identified recombination events at two regions within the penton base gene, encompassing nucleotides 400 to 600 and 750 to 1350. Further analysis of other HAdV-D penton genes suggests that these areas represent hot spots for recombination.

The propensity for recombination among different HAdV-Ds in the penton base gene can be understood in the context of penton base function. The adenovirus penton base acts as the ligand for a secondary attachment event that is critical to host cell internalization (38) and thus is critical to infection. The penton base protein contains two hypervariable loops located on the surface of the viral capsid (16, 44). One loop contains the host cell integrin binding Arg-Gly-Asp (RGD) motif mediating attachment to host cell integrins (8, 38). The RGD motif is conserved in almost every HAdV penton base protein, with the exception of HAdV-F40 and HAdV-F41, which do not use host cell integrins for internalization (1, 11). The second exterior loop, known as the variable loop (HVL1), has no known function. Both of these regions are highly variable among different HAdV types. We identified recombination around both of these loops for multiple viruses within HAdV-D (Fig. 3A).

In recent epidemiological studies, patients were identified with coincident clinical infections with two or more adenovirus types (17, 36). Simultaneous infection by more than a single HAdV type is possible because of conserved host receptor affinity, common tissue tropisms, and an absence of immunity across serotypes. Our data, along with evidence from previous studies (22, 23, 37, 43), identified recombination events among viruses with similar tissue tropisms, providing evidence that the restriction of tissue tropism might determine in part the observed recombination within adenoviruses of the same species. Recently, a novel HAdV, HAdV-D53, was isolated from an outbreak of keratoconjunctivitis. Subsequent analysis revealed recombination among HAdV-D22, HAdV-D37, HAdV-D8, and a previously unknown adenoviral sequence, suggesting the potential for the emergence of new pathogens, with important ramifications for human disease.

Previous work provides evidence for recombination in HAdVs (10, 12, 22, 25, 26, 37, 39, 41–43), but the mechanisms of recombination have yet to be identified. Recombination may result from selective pressure from the host immune system relative to surface capsid proteins, a nucleotide motif that directs cellular recombination machinery to the local sites on the viral DNA, or a combination of both. Two eukaryotic recombinases, RAD51 and Dmc1, both homologues of the bacterial recombinase RecA, act in host cell DNA recombination (5, 33). RAD51 mediates recombination during mitosis, while Dmc1 acts during meiosis. RAD51 is of potential interest because it colocalizes with promyelocytic leukemia nuclear bodies in the nucleus (4). Adenoviral proteins E1A and E4 Orf3 have been shown to interact with promyelocytic leukemia nuclear bodies, which play an important role in adenoviral replication (7, 15). An interaction between RAD51 and adenoviral DNA has not been studied. A proposed motif consisting of CCNCCNTNNCCNC was recently identified as being associated with loci of recombination and genome instability in humans (28). Although not present in the viruses we studied, sequencing of other viruses within HAdV-D may yet reveal a consensus site for recombination. The elucidation of recombination mechanisms for HAdVs should allow a better understanding of adenoviral evolution.

In conclusion, our analysis of the penton base gene of HAdV-D identified a potential paradigm for adenovirus recombination and the emergence of pathogenic strains. An in depth understanding of adenovirus recombination and evolution is critical to ensure the safety of adenoviral gene therapy.

Nucleotide sequence accession number. The HAdV-D22 genome sequence obtained in this study has been deposited in GenBank under accession number FJ404771.

REFERENCES

1. **Albinsson, B., and A. H. Kidd.** 1999. Adenovirus type 41 lacks an RGD alpha(v)-integrin binding motif on the penton base and undergoes delayed uptake in A549 cells. Virus Res. **64**:125–136.
2. **Aoki, K., H. Ishiko, T. Kuno, Y. Shimada, A. Hayashi, H. Kaneko, T. Ohgoshi, Y. Tagawa, S. Ohno, and S. Yamazaki.** 2008. Epidemic keratoconjunctivitis due to the novel hexon-chimeric-intermediate 22,37/H8 human adenovirus. J. Clin. Microbiol. **46**:3259–3269.
3. **Bell, S. D., Jr., T. R. Bota, and D. E. McComb.** 1960. Adenoviruses isolated from Saudi Arabia. III. Six new serotypes. Am. J. Trop. Med. Hyg. **9**:523–526.
4. **Bischof, O., S. H. Kim, J. Irving, S. Beresten, N. A. Ellis, and J. Campisi.** 2001. Regulation and localization of the Bloom syndrome protein in response to DNA damage. J. Cell Biol. **153**:367–380.
5. **Bishop, D. K., D. Park, L. Xu, and N. Kleckner.** 1992. DMC1: a meiosis-specific yeast homolog of E. coli recA required for recombination, synaptonemal complex formation, and cell cycle progression. Cell **69**:439–456.
6. **Brudno, M., C. B. Do, G. M. Cooper, M. F. Kim, E. Davydov, E. D. Green, A. Sidow, and S. Batzoglou.** 2003. LAGAN and Multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA. Genome Res. **13**:721–731.
7. **Carvalho, T., J. S. Seeler, K. Ohman, P. Jordan, U. Pettersson, G. Akusjarvi, M. Carmo-Fonseca, and A. Dejean.** 1995. Targeting of adenovirus E1A and E4-ORF3 proteins to nuclear matrix-associated PML bodies. J. Cell Biol. **131**:45–56.
8. **Cheresh, D. A., and R. C. Spiro.** 1987. Biosynthetic and functional properties of an Arg-Gly-Asp-directed receptor involved in human melanoma cell attachment to vitronectin, fibrinogen, and von Willebrand factor. J. Biol. Chem. **262**:17703–17711.
9. **Chu, W., and D. Pavan-Langston.** 1979. Ocular surface manifestations of the major viruses. Int. Ophthalmol. Clin. **19**:135–167.
10. **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**:357–367.
11. **Davison, A. J., E. A. Telford, M. S. Watson, K. McBride, and V. Mautner.** 1993. The DNA sequence of adenovirus type 40. J. Mol. Biol. **234**:1308–1316.
12. **de Jong, P. J., G. Valderrama, I. Spigland, and M. S. Horwitz.** 1983. Adenovirus isolates from urine of patients with acquired immunodeficiency syndrome. Lancet **i**:1293–1296.
13. **Dhanandhar, N. V., P. Kulkarni, S. M. Ajinkya, and A. Sherikar.** 1992. Effect of adenovirus infection on adiposity in chicken. Vet. Microbiol. **31**:101–107.
14. **Dingle, J. H., and A. D. Langmuir.** 1968. Epidemiology of acute, respiratory disease in military recruits. Am. Rev. Respir. Dis. **97**(Suppl):1–65.
15. **Doucas, V., A. M. Ishov, A. Romo, H. Juguilon, M. D. Weitzman, R. M. Evans, and G. G. Maul.** 1996. Adenovirus replication is coupled with the dynamic properties of the PML nuclear structure. Genes Dev. **10**:196–207.
16. **Fuschiotti, P., G. Schoehn, P. Fender, C. M. Fabry, E. A. Hewat, J. Chroboczek, R. W. Ruigrok, and J. F. Conway.** 2006. Structure of the dodecahedral penton particle from human adenovirus type 3. J. Mol. Biol. **356**:510–520.
17. **Gray, G. C., T. McCarthy, M. G. Lebeck, D. P. Schnurr, K. L. Russell,**

A. E. Kajon, M. L. Landry, D. S. Leland, G. A. Storch, C. C. Ginocchio, C. C. Robinson, G. J. Demmler, M. A. Saubolle, S. C. Kehl, R. Selvarangan, M. B. Miller, J. D. Chappell, D. M. Zerr, D. L. Kiska, D. C. Halstead, A. W. Capuano, S. F. Setterquist, M. L. Chorazy, J. D. Dawson, and D. D. Erdman. 2007. Genotype prevalence and risk factors for severe clinical adenovirus infection, United States 2004–2006. Clin. Infect. Dis. **45**:1120–1131.

18. **Hilleman, M. R., and J. H. Werner.** 1954. Recovery of new agent from patients with acute respiratory illness. Proc. Soc. Exp. Biol. Med. **85**:183–188.
19. **Jager, L., and A. Ehrhardt.** 2007. Emerging adenoviral vectors for stable correction of genetic disorders. Curr. Gene Ther. **7**:272–283.
20. **Kojaoghlanian, T., P. Flomenberg, and M. S. Horwitz.** 2003. The impact of adenovirus infection on the immunocompromised host. Rev. Med. Virol. **13**:155–171.
21. **Lole, K. S., R. C. Bollinger, R. S. Paranjape, D. Gadkari, S. S. Kulkarni, N. G. Novak, R. Ingersoll, H. W. Sheppard, and S. C. Ray.** 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.
22. **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**:380–388.
23. **Madisch, I., S. Hofmayer, C. Moritz, A. Grintzalis, J. Hainmueller, P. Pring-Akerblom, and A. Heim.** 2007. Phylogenetic analysis and structural predictions of human adenovirus penton proteins as a basis for tissue-specific adenovirus vector design. J. Virol. **81**:8270–8281.
24. **Martin, D. P., C. Williamson, and D. Posada.** 2005. RDP2: recombination detection and analysis from sequence alignments. Bioinformatics **21**:260–262.
25. **Mautner, V., and M. E. Bourssnell.** 1983. Recombination in adenovirus: DNA sequence analysis of crossover sites in intertypic recombinants. Virology **131**:1–10.
26. **Mautner, V., and N. Mackay.** 1984. Recombination in adenovirus: analysis of crossover sites in intertypic overlap recombinants. Virology **139**:43–52.
27. **Moruzova, O., and M. A. Marra.** 2008. Applications of next-generation sequencing technologies in functional genomics. Genomics **92**:255–264.
28. **Myers, S., C. Freeman, A. Auton, P. Donnelly, and G. McVean.** 2008. A common sequence motif associated with recombination hot spots and genome instability in humans. Nat. Genet. **40**:1124–1129.
29. **Robinson, C. M., F. Shariati, A. F. Gillaspy, D. W. Dyer, and J. Chodosh.** 2008. Genomic and bioinformatics analysis of human adenovirus type 37: new insights into corneal tropism. BMC Genomics **9**:213.
30. **Robinson, C. M., F. Shariati, J. Zaitshik, A. F. Gillaspy, D. W. Dyer, and J. Chodosh.** 2009. Human adenovirus type 19: genomic and bioinformatics analysis of a keratoconjunctivitis isolate. Virus Res. **139**:122–126.
31. **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**: 570–573.
32. **Sever, J. L., and R. G. Traub.** 1962. Conjunctivitis with follicles associated with adenovirus type 22. N. Engl. J. Med. **266**:1375–1376.
33. **Shinohara, A., H. Ogawa, and T. Ogawa.** 1992. Rad51 protein involved in repair and recombination in S. cerevisiae is a RecA-like protein. Cell **69**: 457–470.
34. **Tamura, K., J. Dudley, M. Nei, and S. Kumar.** 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol. Biol. Evol. **24**:1596–1599.
35. **Thirion, C., H. Lochmuller, Z. Ruzsics, M. Boelhauve, C. Konig, C. Thedieck, S. Kuttk, C. Geiger, S. Kochanek, C. Volpers, and H. G. Burgert.** 2006. Adenovirus vectors based on human adenovirus type 19a have high potential for human muscle-directed gene therapy. Hum. Gene Ther. **17**: 193–205.
36. **Vora, G. J., B. Lin, K. Gratwick, C. Meador, C. Hansen, C. Tibbetts, D. A. Stenger, M. Irvine, D. Seto, A. Purkayastha, N. E. Freed, M. G. Gibson, K. Russell, and D. Metzgar.** 2006. Co-infections of adenovirus species in previously vaccinated patients. Emerg. Infect. Dis. **12**:921–930.
37. **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.
38. **Wickham, T. J., P. Mathias, D. A. Cheresh, and G. R. Nemerow.** 1993. Integrins alpha v beta 3 and alpha v beta 5 promote adenovirus internalization but not virus attachment. Cell **73**:309–319.
39. **Williams, J., T. Grodzicker, P. Sharp, and J. Sambrook.** 1975. Adenovirus recombination: physical mapping of crossover events. Cell **4**:113–119.
40. **Wood, D. J.** 1988. Adenovirus gastroenteritis. Br. Med. J. **296**:229–230.
41. **Young, C. S., G. Cachianes, P. Munz, and S. Silverstein.** 1984. Replication and recombination in adenovirus-infected cells are temporally and functionally related. J. Virol. **51**:571–577.
42. **Young, C. S., and S. J. Silverstein.** 1980. The kinetics of adenovirus recombination in homotypic and heterotypic genetic crosses. Virology **101**:503–515.
43. **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.
44. **Zubieta, C., G. Schoehn, J. Chroboczek, and S. Cusack.** 2005. The structure of the human adenovirus 2 penton. Mol. Cell **17**:121–135.

Part 3

Computational Analysis Identifies Human Adenovirus Type 55 as a Re-Emergent Acute Respiratory Disease Pathogen

JOURNAL OF CLINICAL MICROBIOLOGY, Mar. 2010, p. 991–993 Vol. 48, No. 3
0095-1137/10/$12.00 doi:10.1128/JCM.01694-09

Computational Analysis Identifies Human Adenovirus Type 55 as a Re-Emergent Acute Respiratory Disease Pathogen[▽]

Michael P. Walsh,[1] Jason Seto,[1] Morris S. Jones,[2] James Chodosh,[3] Wenbo Xu,[4] and Donald Seto[1]*

Department of Bioinformatics and Computational Biology, George Mason University, 10900 University Blvd., MSN 5B3, Manassas, Virginia 20110[1]; Clinical Investigation Facility, David Grant USAF Medical Center, Travis, California 94535[2]; Department of Ophthalmology, Howe Laboratory, Massachusetts Eye and Ear Infirmary, Harvard Medical School, Boston, Massachusetts 02114[3]; and State Key Laboratory for Molecular Virology and Genetic Engineering, National Institute for Viral Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Nanwei Road 27, Xuanwu District, Beijing 100050, People's Republic of China[4]

Received 31 August 2009/Returned for modification 12 December 2009/Accepted 17 December 2009

Novel human adenoviruses (HAdVs) arise from genome recombination. Analysis of HAdV type 55 from an outbreak in China shows a hexon recombination between HAdV-B11 and HAdV-B14, resulting in a genome that is 97.4% HAdV-B14. Sporadic appearances as a re-emergent pathogen and misidentification as "HAdV-B11a" are due to this partial hexon.

Human adenoviruses (HAdVs) were first identified as respiratory pathogens (7, 18) but are now recognized as causing a range of diseases, including those that are ocular, gastrointestinal, and metabolic (4). There are 51 "serotypes" defined by using biological characteristics, including immunochemical methods, e.g., serum neutralization and hemagglutination (4). A novel HAdV was characterized using nonimmunochemical methods, e.g., genomics and bioinformatics, and was named HAdV-G52 (9). Recently, genomics and bioinformatics data have defined two emergent, pathogenic, and recombinant HAdVs (D53 and D54) (8, 21). All three are characterized by genomics and therefore should be termed appropriately as "type" rather than "serotype," as discussed at the 9th International Adenovirus Meeting (Dobogókő, Hungary, April 2009) (our unpublished data). Described here is a re-emergent respiratory pathogen, HAdV-B55, which contains a partial hexon recombination conferring a change in serotype and, hence, an escape from immune reactivity against HAdV-B14. It was isolated recently from an acute respiratory disease (ARD) outbreak in China and incorrectly termed "HAdV-B11-like" (22, 23).

Genomes of "HAdV-B11 strain QS-DLL," noted here as HAdV-B55 (accession no. FJ643676), HAdV-11p (accession no. AF532578), and HAdV-B14p (accession no. AY803294), were obtained from GenBank. Computational analysis included sequence comparisons with zPicture (15), restriction enzyme pattern analysis (http://www.acaclone.com/), phylogeny analysis using MAVID (1), and recombination analysis with SimPlot and Bootscan (12). Genome percent identities were determined using zPicture, MAFFT (10), and Chimera (16).

The genome "chassis" of HAdV-B55 is HAdV-B14 (19), an ARD pathogen, with a partial HAdV-B11 hexon providing the misleading serum neutralization result and incorrect, incomplete original identification, as well as an incorrect clinical diagnosis. HAdV-B55 hexon contains 907 nucleotides from HAdV-B11 out of 2,841 nucleotides (31.9%), embedded in the 34,755 base pair genome (2.6%) (Fig. 1A and B). Sequence comparisons reflect the following: HAdV-B55 versus HAdV-B14 at 98.86%; HAdV-B55 versus HAdV-B11 at 97.64%; and HAdV-B11 versus HAdV-B14 at 97.21% (Fig. 1C). HAdV-B55 appears to have evolved from a recombination between HAdV-B14 and -B11 and should be noted as a new type, in the context of both the recombinant HAdVs that are accepted as novel types (8, 21) and its change in serotype (22). In phylogeny analyses, HAdV-B55 groups into a subclade with HAdV-B14 and HAdV-B14a, whereas hexon (loops 1 and 2) groups into a subclade with HAdV-B11. As "HAdV-B11a" did not cross-react with HAdV-B14 antisera, it is not a simple variant of HAdV-B14. Detailed restriction enzyme (RE) digestion pattern analysis, seemingly anachronistic but highly effective visually, confirms this, showing HAdV-B55 as closer to the HAdV-B14 genome than to the HAdV-B11 genome (data not shown), and also matching the original RE analysis for HAdV-B11a (11).

The divergent hexon comprises two regions of identity indicating a recombination event within the gene (Fig. 1B). Analysis of an earlier described "HAdV-B11a" hexon (GenBank accession no. AY972815) (2) shows it is identical to the one embedded in the "QS" genome (23), suggesting that "HAdV-B11a" was incorrectly characterized earlier as a variant of HAdV-B11 (2, 6, 11, 22) by the use of "then-available" assays. Serum neutralization has the hexon as its target, as does limited molecular typing (3, 13). Since genome rearrangements are missed by serological and molecular assays against a limited repertoire, it is problematic to rely solely on these methods to characterize a virus thoroughly. Thus, due to the limitations of these assays, the original "HAdV-B11a" isolates have been misnamed as well (2, 6, 11).

As "HAdV-B11a," this virus represented a paradox. HAdV-

* Corresponding author. Mailing address: 10900 University Blvd., MSN 5B3, Department of Bioinformatics and Computational Biology, George Mason University, Manassas, VA 20110. Phone: (703) 993-8403. Fax: (703) 993-8401. E-mail: dseto@gmu.edu.

▽ Published ahead of print on 30 December 2009.

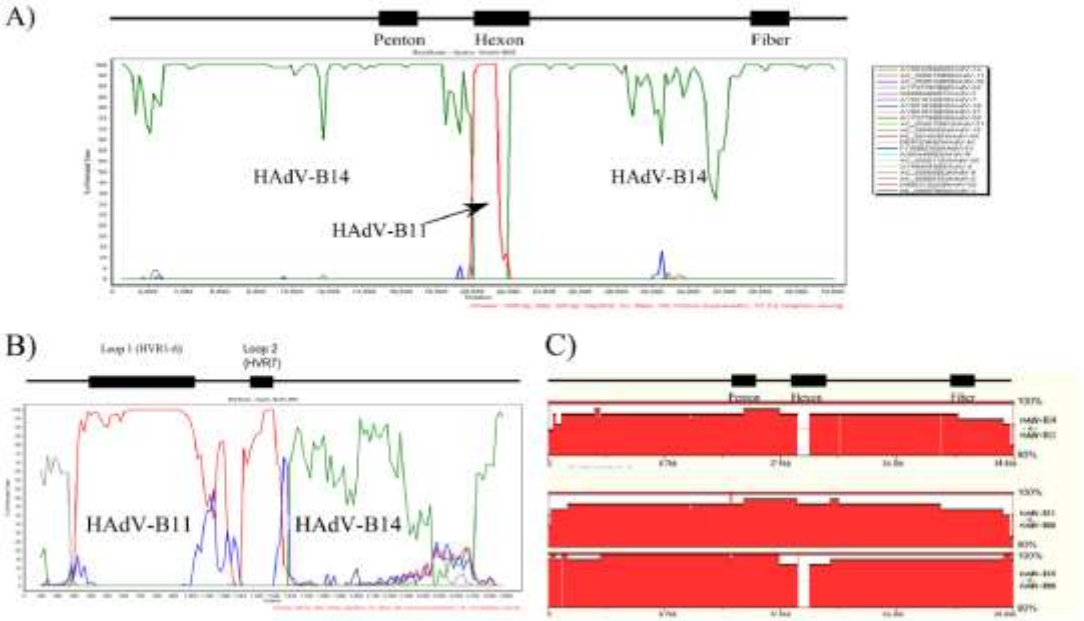


FIG. 1. (A) Bootscan recombination analysis of the HAdV-B55 genome. The dark green reflects HAdV-B14, and the red indicates HAdV-B11 sequences. Parameters are as follows: 1,000-bp window; 200-bp step; 100 repetitions; neighbor-joining algorithm. The HAdV types analyzed and the colors are the same for both panels A and B, in descending order: HAdV-B14 (top; dark green), B11, B35, B34, B3, B7, B16, B21, B50, simian adenovirus (SAdV)-B21, A12, F40, F41, D53, D9, SAdV-E25, E4, C5, C2, G52, and SAdV-G1 (bottom; light green). (B) Fine-resolution analysis of the partial hexon gene recombination. Bootscan analysis of the hexon gene shows that the proximal region, containing HVR1-6 ("Loop 1"; 407 to 1,034 nucleotide gene location) and HVR7 ("Loop 2"; 1,363 to 1,484 nucleotide gene location), derives from HAdV-B11. The distal portion derives from HAdV-B14, as does the rest of the genome. Parameters are as follows: 200-bp window; 20-bp step; 100 repetitions; neighbor-joining algorithm. (C) Comparative genomics. zPicture compares the nucleotide sequences of pairs of genomes by using a moving overlapping window and scoring percent identities. The *y* axis is set between 90 and 100%, indicating a very high level of identity.

B11 is a member of subspecies B2 and was not originally associated with respiratory disease (14); only HAdV-B14 in this group is associated with respiratory disease (20). A clue as to the incorrect association of HAdV-B11 as a respiratory pathogen lies in the reports that "HAdV-B11a" was unusual in having cellular tropism to respiratory epithelial cells rather than renal cells, as observed for original and subsequent isolates of HAdV-B11 (11, 23). In addition, the original "HAdV-B11a" did not agglutinate monkey erythrocytes, unlike the true HAdV-B11 field strains and the prototype Slobitski strain (5). If HAdV-B55 is truly a variant of HAdV-B11 and an ARD pathogen, it presents a drastic change in cell and tissue tropism—from renal and urinary tract to lung, with high morbidity and some mortality (11, 23). This paradox is resolved by the analysis of its genome, that it is HAdV-B14-like rather than HAdV-B11-like. By inference, the cell recognition epitope is either the distal 68.1% of the HAdV-B14 hexon or the fiber, rather than the proximal 31.9% encompassing "loops 1 and 2" of the HAdV-B11 hexon.

Although originally identified in sporadic and infrequent historical ARD outbreaks by serum neutralization as "HAdV-B11-like" and recently recharacterized by immunochemistry (serum neutralization and enzyme-linked immunosorbent assay [ELISA]) and limited molecular typing (PCR and hexon sequencing) also as "HAdV-B11-like" (2, 6, 11, 22), genomics and bioinformatics demonstrate that it is a novel type. All of these older techniques, assaying only the hexon gene, simply reconfirmed a HAdV-B11-like hexon epitope. Therefore, it is incorrectly named and noted in GenBank (accession no. FJ643676) as "HAdV-B11a" or "HAdV-B11 strain QS-DLL" (23). Based on the reported and additional computational analyses and within the context of recent reports of HAdV molecular evolution based on recombination (8, 17, 21), this pathogen is a novel adenovirus type that should be named HAdV-B55.

A tsunami of genome sequence information from both newly isolated and presciently archived HAdV strains and their accompanying bioinformatics are leading to an in-depth understanding of the biology of HAdVs. Correct identification and nomenclature of viruses are critical for defining and understanding them as well as their pathogenic variants, particularly in GenBank. The correction of the name of this re-emergent type is urgent, as research groups characterizing other ARD outbreaks by similar or the same HAdV pathogens are on the verge of reporting their findings, adding to the confusion. Genome recombination plays an important role in the

molecular evolution of HAdVs, with consequences of newly emerging strains and new understandings of re-emerging pathogens that have tropism change or that become more virulent. An intriguing thought for HAdV-B55 is that perhaps this region of the hexon is a recombination hot spot, driving the evolution, selection, and appearance of a recurrent re-emergent pathogen by altering its serotype and giving it an advantage in a population with seroprevalence against HAdV-B14. As more HAdV genomes are elucidated, they will lead to better understandings of other pathogens that may utilize similar evolution pathways in appearing as emergent and re-emergent pathogens.

We thank Qiwei "Kevin" Zhang, The University of Hong Kong, for assistance in obtaining the prepublication genome of HAdV-B55.

This work was supported in part by U.S. Public Health Service NIH grants EY013124 (D.S., M.P.W., M.S.J., and J.C.) and P30EY013104 (J.C.). J.C. is also supported by funding from the Massachusetts Lions Eye Research Fund, Inc., and an unrestricted grant to the Department of Ophthalmology, Harvard Medical School, from Research to Prevent Blindness, Inc. M.S.J. is supported by the United States Air Force Surgeon General, Clinical Investigation no. FDG20040024E.

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.

REFERENCES

1. **Bray, N., and L. Pachter.** 2004. MAVID: constrained ancestral alignment of multiple sequences. Genome Res. **14**:693–699.
2. **Chmielewicz, B., J. Benzler, G. Pauli, G. Krause, F. Ilergmann, and B. Schweiger.** 2005. Respiratory disease caused by a species B2 adenovirus in a military camp in Turkey. J. Med. Virol. **77**:232–237.
3. **Crawford-Miksza, L., and D. P. Schnurr.** 1996. Analysis of 15 adenovirus hexon proteins reveals the location and structure of seven hypervariable regions containing serotype-specific residues. J. Virol. **70**:1836–1844.
4. **Echavarria, M.** 2009. Adenovirus, p. 463–488. *In* A. J. Zuckerman, J. E. Banatvala, B. D. Schoub, P. D. Griffiths, and P. Mortimer (ed.), Principles and practice of clinical virology, 6th ed. John Wiley and Sons, San Diego, CA.
5. **Hierholzer, J. C., and A. Pumarola.** 1976. Antigenic characterization of intermediate adenovirus 14-11 strains associated with upper respiratory illness in a military camp. Infect. Immun. **13**:354–359.
6. **Hierholzer, J. C., A. Pumarola, A. Rodriguez-Torres, and M. Beltran.** 1974. Occurrence of respiratory illness due to an atypical strain of adenovirus type 11 during a large outbreak in Spanish military recruits. Am. J. Epidemiol. **99**:434–442.
7. **Hilleman, M. R., and J. H. Werner.** 1954. Recovery of new agent from patients with acute respiratory illness. Proc. Soc. Exp. Biol. Med. **85**:183–188.
8. **Ishiko, H., and K. Aoki.** 2009. Spread of epidemic keratoconjunctivitis due to a novel serotype of human adenovirus in Japan. J. Clin. Microbiol. **47**:2678–2679.
9. **Jones, M. S., II, 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**:5978–5984.
10. **Katoh, K., and H. Toh.** 2008. Recent developments in the MAFFT multiple sequence alignment program. Brief. Bioinformatics **9**:286–298.
11. **Li, Q. G., J. Hambraeus, and G. Wadell.** 1991. Genetic relationship between thirteen genome types of adenovirus 11, 34, and 35 with different tropisms. Intervirology **32**:338–350.
12. **Lole, K. S., R. C. Bollinger, R. S. Paranjape, D. Gadkari, S. S. Kulkarni, N. G. Novak, R. Ingersoll, H. W. Sheppard, and S. C. Ray.** 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.
13. **Madisch, I., G. Harste, H. Pommer, and A. Heim.** 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.
14. **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 11. N. Engl. J. Med. **278**:700–704.
15. **Ovcharenko, I., G. G. Loots, R. C. Hardison, W. Miller, and L. Stubbs.** 2004. zPicture: dynamic alignment and visualization tool for analyzing conservation profiles. Genome Res. **14**:472–477.
16. **Pettersen, E. F., T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin.** 2004. UCSF Chimera—a visualization system for exploratory research and analysis. J. Comput. Chem. **25**:1605–1612.
17. **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 among species D human adenoviruses in the penton base gene. J. Virol. **83**:8980–8985.
18. **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**: 570–573.
19. **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.
20. **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**:119–129.
21. **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.
22. **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. **47**:3082–3090.
23. **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**:269?–703.

CHAPTER 5: A Necessary Paradigm Shift in HAdV Nomenclature

The current accepted taxonomy scheme of HAdV classifies the HAdVs into species based on biological and clinical data [21]. Within each species, HAdVs are further delineated into different "serotypes". The serotype of a HAdV is determined using specific properties of the hexon and fiber genes of the organism and was originally based on antibody – antigen reactions or serology (hence, "sero" type)[21].

The properties used to ascertain the serotype of a HAdV can be measured using antibody or molecular typing. Antibody typing uses serum neutralization (SN) and hemagglutination inhibition (HI) assays to test whether or not the hexon and fiber proteins of a query HAdV are similar to homologs in a previously established serotype [1]. In molecular typing, portions of a query HAdV genome that encode for the epitopes targeted in SN and HI assays (hexon and fiber genes) are amplified via PCR. The PCR products are then sequenced and compared to established serotypes using sequence homology software (Blast) [66,30,67,4]. In cases where data from the hexon and fiber characterizations disagree, hexon data is given precedence and the virus is often referred to as an intermediate strain [36].

Antibody and molecular typing techniques were the best and only available HAdV classification mechanisms for many years. However, these methods are limited in the proportion of the genome and virus they explore. SN and hexon molecular typing examine a portion of the hexon that constitutes approximately a third of the gene and less than three percent of the genome each. HI and fiber molecular typing explore less than a third of the fiber gene which accounts for approximately two percent of the genome. The ninety-five percent of the genome that is ignored by the assays certainly holds data that would be useful in the classification of HAdV.

The limitations of the current typing techniques have led to contradictions in the study of some HAdV. For example, early molecular typing results indicated that HAdV-D53 (discussed in Chapter 4) was a variant of HAdV-D22 [39]. However, the fact that HAdV-D53 had been implicated in outbreaks of severe epidemic keratoconjunctivitus (EKC) distinguished it from HAdV-D22 which is not normally pathogenic [39,27]. This apparent paradox was not resolved until whole genome analysis of HAdV-D53 revealed that the majority of its genome was similar to HAdV-D37, a known causal agent of EKC [65,10]. The data from the HAdV-D53 analysis led to its designation as a novel HAdV type. HAdV-D53 is referred to as a novel "type" as opposed to "serotype" to distinguish it from HAdVs that have been classified using serological methods.

Advancements in genome sequencing and bioinformatics methods allow for the development of a more "comprehensive typing algorithm" (CTA). This algorithm makes

use of all available genome analysis data from phylogenomics (as defined in Chapter 2) and recombination analysis to clinical data. The method also incorporates the principles of the current typing conventions (which make use of the virus capsid proteins: penton, hexon, fiber) so that the contradictions with past HAdV literature can be avoided. The sum of the information from all parts of the CTA is combined to determine the degree of novelty of a query HAdV sequence and its place in the HAdV taxonomy.

The phylogenomic examination of a HAdV begins with a whole genome phylogenetic tree. The whole genome tree is based on a sequence alignment of all available HAdV genomes. This tree can be used to determine the species of a query HAdV. Phylogenomic species determination is an improvement over earlier methods because it uses the molecular relationship between HAdVs, rather than clinical data that can be incomplete or misinterpreted, to discern the species of a HAdV. Furthermore, and perhaps surprisingly, data from this method reflects the findings of past literature. Figure 11 shows a whole genome phylogenetic tree that demonstrates this fact. In the tree, all taxa form clades based on the species determinations made by the original authors of the reference publication for each HAdV.

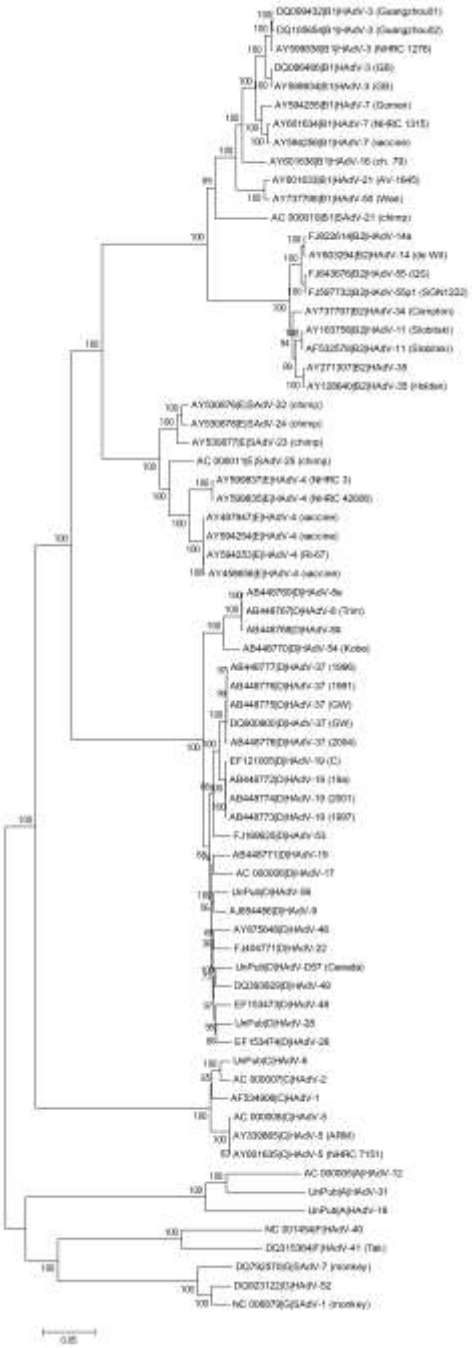


Figure 11. A whole genome phylogenetic tree of seventy HAdV genomes. The name of each taxa has a format that list the GenBank accession number for the genome, followed by a "|", followed by the species of the genome and followed by the serotype of the HAdV. The neighbor joining, bootstrapped tree (1000 replicates) was constructed using Mega 4. Phylogenetic distances for the tree were calculated using the Maximum

Composite Likelihood method (MCL). Taxa from the same species form distinct clades within the tree.

It would be convenient, and therefore tempting, to use some metric of the whole genome phylogeny (percent identity or phylogenetic distance values) to determine the type of a query HAdV. In this type of method, the shortest distance between any two different established HAdV types could be measured and used as a sort of "low bar". Then any query HAdV with a distance from its nearest relative that is larger than the "low bar" would be called novel.

The "low bar" method was explored and inconsistencies in the current genome versus type data made it impossible to employ it as a comprehensive typing strategy. The most common types of inconsistencies found were cases in which the distance between two established HAdVs of the same type was larger than the distance between HAdVs of different types. For example, the Maximum Composite Likelihood (MCL) [50,51] calculated distance between the reference and field strains of HAdV-B7 is 0.0176, which is larger than the distance of 0.0139 between HAdV-B35 and HAdV-B11. Percent identity values of these HAdVs show the same relationships indicating that the inconsistencies observed are not artifacts of the method of distance calculation. The source of the contradictions in the distance values is not yet known but may be related to different evolutionary rates among disparate species of HAdVs. Nonetheless, a more refined method must be used to determine HAdV types.

To determine the type of query HAdVs, the phylogenomics of individual genes must be examined. The hexon, fiber and penton genes are explored in this step of the CTA. These genes encode the surface proteins of the virus. Their role in determining the cell tropism of the virus makes them a logical choice for HAdV typing. Furthermore, the genes are well studied and information from previous examinations can be incorporated into the CTA so that its results remain consistent with past research.

The HAdV hexon protein contains two major loops (referred to as L1 and L2) that hold the epitope targeted by SN assays. Primers that are capable of isolating the portion of the hexon gene that encodes for these loops have been developed as part of a molecular typing strategy [30]. Only one of the hexon loops is necessary to mimic SN results (personal communication). Given the choice of using either loop, it is logical to select L1 because it is larger and any phylogeny based on L1 will be less likely to be affected by small differences in alignment methods. For these reasons, the phylogenomics of the L1 region of the hexon gene have been included in the CTA.

Recombination studies have revealed that the all of the known HAdV hexon recombination events bracket the L1 and L2 regions [10,11]. Downstream of L1 and L2 is a conserved region (referred to as C4) that constitutes approximately half of the gene and shows very little recombination potential. Incorporating phylogenomics data from this region into the CTA allows for the identification of potential recombination events. Primers that isolate a portion of the C4 region have been developed previously [43] and were integrated in to the analysis.

The HAdV fiber protein contains a variable region, called the knob. This contains the epitope that is targeted in HI assays. This region also contains the determinants for cell entry and cell tropism. Primers bracketing the knob region of the fiber gene have been published [30] and so the phylogenomics of this region is included as part of the CTA in order to mimic HI results.

The penton protein and gene are often overlooked by antibody and molecular typing methods. As a result of this, primers for important portions of the penton gene have not been developed. The phylogenetics of the entire penton gene is incorporated into the CTA representing this portion of the genome.

Figure 12 shows a phylogenomic examination of HAdV-B55 and HAdV-B55p1. Both of these isolates are members of the HAdV-B2 subspecies, as indicated by the whole genome tree. The fact that HAdV-B55 and B55p1 form clades with one another in all trees suggests that they are variants of the same HAdV type.

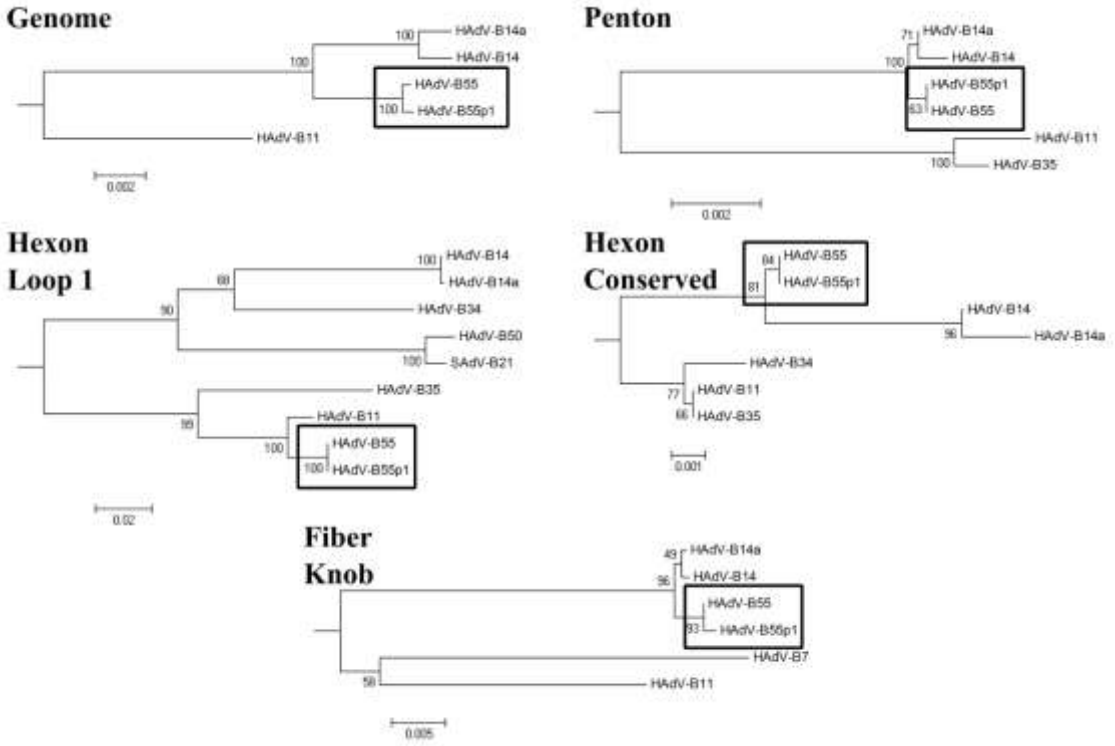


Figure 12. A phylogenetic examination of HAdV-B55 and HAdV-B55p1. The genome, penton, hexon conserved, and fiber knob trees reveal a close relationship between HAdV-B55, B55p1, HAdV-B14, and HAdV-B14a. The hexon loop 1 tree shows a close relationship with HAdV-B11, indicating a recombination event. HAdV-B55 and HAdV-B55p1 form clades together in all trees, demonstrating that they are variants of the same type.

The HAdV-B55 isolates were originally mis-identified as variants of HAdV-B11 (HAdV-B11a) based on molecular typing data. However, phylogenetic analysis of the whole genome, penton, hexon conserved region and fiber knob region clearly indicate that HAdV-B55 and HAdV-B55p1 are most closely related to HAdV-B14. Furthermore, the discrepancy between the hexon loop 1 and conserved region trees shows a hexon recombination (discussed in Chapter 4) that led to the misleading molecular typing results.

Only two logical conclusions can be drawn from the phylogenomic examination of the HAdV-B55 isolates. The first conclusion is that HAdV-B55 and B55p1 are variants of HAdV-B14. The second conclusion is that the HAdV-B55 isolates are variants of a novel HAdV type due to changes in pathology.

To determine whether or not the isolates of HAdV-B55 represent a novel HAdV type, the CTA incorporates biological or clinical data. The phylogenetic trees of both hexon regions indicate that the HAdV-B55 variants contain a hexon recombination. This data is confirmed by a formal recombination analysis (Figure 5 in chapter 2, also discussed in Chapter 4) [11]. However, a recombination event alone is not enough to designate a novel HAdV type. Otherwise, biologically insignificant recombination events could be used as a basis to name novel types which could lead to confusion in the literature. In order for a new type to be defined based on a recombination event, that event must have some biological or clinical relevance. It is this relevance that makes the claim of a novel type scientifically defensible.

The recombination event within the genome of HAdV-B55 causes it (and its variants) to have a hexon that resembles HAdV-B11, a renal pathogen. However, the rest of the HAdV-B55 genome is similar to HAdV-B14 which, like HAdV-B55, causes respiratory symptoms. The renal hexon of HAdV-B55 could allow the virus to evade a host immune system that is primed to fight respiratory pathogens that are similar to HAdV-B14. The biological implications of this evasion are intriguing and lead to the conclusion that HAdV-B55 and its variants represent a novel type of HAdV.

The analysis of HAdV-B55 demonstrates how a new CTA can be used to combine phylogenomic, recombination, biological and clinical data to designate a novel HAdV type. Furthermore, the analysis of HAdV-B55p1 shows that variants of existing types can also be identified using the CTA. This new algorithm makes use of existing typing methods and augments them with whole genome analysis to create an exact and thorough typing strategy.

CHAPTER 6: Future Directions

The previous chapters of this document discuss a new algorithm to study comprehensively HAdV genomes and their evolutionary history. This method has been useful in the examination of HAdVs with pathogenic or vaccine/vector development potential and viruses from species that have been understudied in the past. The algorithm has also been used to examine the growing number of newly isolated recombinant HAdV genomes. Furthermore, recombination studies have revealed the need for a new comprehensive typing method for HAdV which has been developed and incorporated into the genome analysis process.

Each aspect of this new analysis algorithm can be improved with future research. First, the study of non-recombinant genomes can be enhanced by standardizing the annotation of HAdV genomes. Second, the study of recombinant genomes can be improved by more accurately defining the amount of data required to identify a recombination event and the best way to deal with the growing amount of data being produced by recombination studies. Finally, the comprehensive typing algorithm (CTA) discussed in this document can be improved thorough the development of a standardized nomenclature for HAdV variants.

The non-recombinant aspects of HAdV genome analysis can be greatly improved with the development of standardized annotation for genes and proteins. An annotation of HAdV sequences varies widely from genome to genome. Products from the E1A gene, for example, can be named according to their size in kilodaltons (eg. E1A 21K in HAdV-D53), length in residues (eg. E1A 253R in HAdV-D46), or rate of centrifugal sedimentation rate (eg. E1A 13S in HAdV-B3). The disparate naming conventions hinder homology (and other) studies because they make it more difficult to determine whether or not genes or products are, in fact, homologs of one another.

The diverse HAdV research community will have to play a large role in any effort to standardize the annotations process. One possible method of involving this community in an annotation review is to create a publicly available and editable database of HAdV sequence data. This Wikipedia-like resource could create an environment where the annotation of specific HAdV sequences is standardized by the researchers who are most familiar with these sequences. The first step toward developing this type of HAdV catalog is the creation of a sequence database similar to the one that is described in Chapter 2 of this document.

Answering two important questions will improve HAdV recombination studies in the future. The first question is “How convincing must data be to support adequately a

claim of recombination?" The second question is "What is the best way to undertake recombination studies that contain a large number of reference sequences?"

The recombination studies discussed in this document describe HAdVs that contain clear, almost undeniable recombination events. In all cases, the recombinant sequences have a very high degree of similarity (>95%) and bootscan analysis graphs show prolonged plateaus at the maximum possible level (100%). In short, these recombination claims are very easy to defend scientifically. However, it may be the case that these recombination events represent the so called "low fruit" that has been picked from the tree. Future recombination studies may have to rely on data that supports a claim of recombination in a less convincing manner. For example, the analysis of HAdV-A18 (discussed in Chapter 3) presents bootscan evidence that indicates the presence of hexon recombination. This bootscan data is contradicted by the similarity plot graphs, showing only eighty-five percent (85%) similarity in the recombinant region. This contradiction raises a question as to what degree of homology is necessary for a defensible claim of recombination. Questions, such as these, are likely to become more common as the number of HAdV recombination studies increases.

The question of how convincing data must be to support a claim of recombination will ultimately have to be answered by the HAdV research community through the process of peer review. However, it may be logical to incorporate clinical and biological data into recombination studies to aid in answering this question. The recombination events in HAdV-D53 and HAdV-B55 changed the tropism of these viruses, which has biological and clinical implications. These implications make a claim of recombination more scientifically convincing.

The number of HAdV genomes that are fully sequenced is growing rapidly. The large number of sequences available can make recombination studies more difficult. The problem with using a high number of sequences in a recombination study is that it becomes more difficult to differentiate those sequences in graphs, such as bootscans or similarity plots. However, eliminating sequences from the analysis could decrease the quality of the analysis. Answering the question of how to deal with numerous sequences will be an important part of future recombination studies.

One strategy for dealing with a large number of sequences in a recombination study is to "pare down" the data in a step-wise fashion. In this method, a first-pass recombination analysis is run that contains every possible sequence. Next, the sequences that contribute little or nothing to the analysis are removed from the sequence group and the recombination scan is run again. This process is repeated until the recombination plots are sufficiently clarified. The "pare-down" strategy was employed, to great effect, in the recombination studies that are discussed in this document. As the body of HAdV sequences grows, this strategy will, most likely, have to be refined in some manner.

The comprehensive typing algorithm (CTA) that is discussed in Chapter 5 of this document can be improved by standardizing the nomenclature that will be used for variants of existing HAdV types. The CTA has been shown to be capable of determining the novelty of a query HAdV. The current convention on naming novel HAdVs is to name them in sequential order. Since there are currently fifty-five HAdV types, the next

novel HAdV to be characterized will be named HAdV-56. This convention is suitable for most purposes and is accepted among the HAdV community. However, the typing convention for variants of HAdVs is not standardized. This variability of variant names could lead to confusion with regard to the identity of HAdVs in the future.

It may be possible to standardize variant names in the future, by using less variable methods to measure the differences between variants. A popular typing convention for variant names that has been used in the past is based on restriction enzyme (RE) map differences [Heim ref]. Variants that differ significantly from another strain are given letters. This is the case for HAdV-B14**a** which is a variant of HAdV-B14. Variants that differ only slightly from another strain are given numbers. This is the case with HAdV-B55p**1** which is a variant of HAdV-B55p (the "p" stands for prototype and is commonly omitted from the name).

A major problem with using the RE typing convention is that the degree of difference which is required to give a variant a letter name, rather than a number, is ill-defined. The primary reason that this degree of difference is not defined is that RE maps measure sequence identity in a variable way. In a RE analysis, the restriction maps of two sequences may be more or less similar, depending upon which enzymes are used to construct the maps.

A less variable method of measuring the difference between sequences could be used to overcome the limitations of RE analysis. Percent identity and phylogenetic distance measurements between sequences remain relatively constant even when different alignment methods are used or extra sequences are added to the analysis. These measurements could be used to define a metric that is required to name a variant with a letter. This technique would be similar to the "low bar" method discussed in Chapter 5. It is possible that using "low bar" measurements to define variants will prove to be impossible, just as they cannot be used to define novel HAdV types. However, the use of these measurements is a logical approach that should be explored.

The genome analysis method described in this document represents a comprehensive way to characterize HAdVs quickly and accurately. This system has already produced a wealth of valuable data. As HAdV research progresses, the genome analysis algorithm will undoubtedly, evolve and improve.

Note: All references to HAdV-B14a, in this manuscript, describe HAdV-B14p1. The virus was originally incorrectly named.

APPENDIX

Java code for the Protein Percent Identities program that is discussed in this Dissertation:

dispacter-servelet.xml

```
<?xml version="1.0" encoding="UTF-8"?>
<beans xmlns="http://www.springframework.org/schema/beans"
      xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
      xmlns:p="http://www.springframework.org/schema/p"
      xmlns:aop="http://www.springframework.org/schema/aop"
      xmlns:tx="http://www.springframework.org/schema/tx"
      xsi:schemaLocation="http://www.springframework.org/schema/beans
http://www.springframework.org/schema/beans/spring-beans-2.5.xsd
      http://www.springframework.org/schema/aop
http://www.springframework.org/schema/aop/spring-aop-2.5.xsd
http://www.springframework.org/schema/tx
http://www.springframework.org/schema/tx/spring-tx-2.5.xsd">

   <bean
class="org.springframework.web.servlet.mvc.support.ControllerClassNameHandlerMapp
ing"/>

   <!--
   Most controllers will use the ControllerClassNameHandlerMapping above, but
   for the index controller we are using ParameterizableViewController, so we must
   define an explicit mapping for it.
   -->
   <bean id="urlMapping"
class="org.springframework.web.servlet.handler.SimpleUrlHandlerMapping">
      <property name="mappings">
         <props>
            <prop key="FastaUploadForm.htm">fastaAlign</prop>
            <prop key="FastaConvertForm.htm">fastaConvert</prop>
            <prop key="FileUploadForm.htm">seqAlign</prop>
            <prop key="index.htm">indexController</prop>
         </props>
      </property>
   </bean>

   <bean id="multipartResolver"
class="org.springframework.web.multipart.commons.CommonsMultipartResolver"/>
```

```
    <bean id="viewResolver"
        class="org.springframework.web.servlet.view.InternalResourceViewResolver"
        p:prefix="/WEB-INF/jsp/"
        p:suffix=".jsp" />

    <!--
    The index controller.
    -->
    <bean name="indexController"
        class="org.springframework.web.servlet.mvc.ParameterizableViewController"
        p:viewName="index" />

    <bean id="seqAlign" class="controller.SeqAlignController" p:seqAlignService-
ref="seqAlignService">
        <property name="commandName" value="seqAlignForm"/>
        <property name="pages">
            <list>
                <value>FileUploadForm</value>
                <value>SeqSelectForm</value>
                <value>SeqConfirmationForm</value>
            </list>
        </property>
    </bean>

    <bean id="fastaAlign" class="controller.FastaAlignController" p:seqAlignService-
ref="seqAlignService">
        <property name="commandName" value="fastaAlignForm"/>
        <property name="pages">
            <list>
                <value>FastaUploadForm</value>
                <value>FastaSeqSelectForm</value>
                <value>FastaSeqConfirmationForm</value>
            </list>
        </property>
    </bean>

    <bean id="fastaConvert" class="controller.FastaConvertFormController"
    p:seqAlignService-ref="seqAlignService">
        <property name="commandName" value="fastaForm"/>
        <property name="commandClass" value="controller.FastaConvertBean"/>
        <property name="formView" value="FastaConvertForm"/>
        <property name="successView" value="FastaConvertConfirmation"/>
```

```
    </bean>


</beans>
```

applicationContext.xml

```
<?xml version="1.0" encoding="UTF-8"?>
<beans xmlns="http://www.springframework.org/schema/beans"
       xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
       xmlns:p="http://www.springframework.org/schema/p"
       xmlns:aop="http://www.springframework.org/schema/aop"
       xmlns:tx="http://www.springframework.org/schema/tx"
       xsi:schemaLocation="http://www.springframework.org/schema/beans
http://www.springframework.org/schema/beans/spring-beans-2.5.xsd
       http://www.springframework.org/schema/aop
http://www.springframework.org/schema/aop/spring-aop-2.5.xsd
       http://www.springframework.org/schema/tx
http://www.springframework.org/schema/tx/spring-tx-2.5.xsd">

    <!--bean id="propertyConfigurer"
          class="org.springframework.beans.factory.config.PropertyPlaceholderConfigurer"
          p:location="/WEB-INF/jdbc.properties" />

    <bean id="dataSource"
          class="org.springframework.jdbc.datasource.DriverManagerDataSource"
          p:driverClassName="${jdbc.driverClassName}"
          p:url="${jdbc.url}"
          p:username="${jdbc.username}"
          p:password="${jdbc.password}" /-->
    <bean name="seqAlignService" class="service.SeqAlignService">
      <constructor-arg value="/WEB-INF/Res/BLOSUM62.txt"/>
    </bean>
    <!-- ADD PERSISTENCE SUPPORT HERE (jpa, hibernate, etc) -->

</beans>
```

AlignConf.jsp

```
<?xml version="1.0" encoding="UTF-8"?>
<beans xmlns="http://www.springframework.org/schema/beans"
       xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
       xmlns:p="http://www.springframework.org/schema/p"
       xmlns:aop="http://www.springframework.org/schema/aop"
```

```
      xmlns:tx="http://www.springframework.org/schema/tx"
      xsi:schemaLocation="http://www.springframework.org/schema/beans
http://www.springframework.org/schema/beans/spring-beans-2.5.xsd
      http://www.springframework.org/schema/aop
http://www.springframework.org/schema/aop/spring-aop-2.5.xsd
      http://www.springframework.org/schema/tx
http://www.springframework.org/schema/tx/spring-tx-2.5.xsd">

   <!--bean id="propertyConfigurer"
         class="org.springframework.beans.factory.config.PropertyPlaceholderConfigurer"
         p:location="/WEB-INF/jdbc.properties" />

   <bean id="dataSource"
         class="org.springframework.jdbc.datasource.DriverManagerDataSource"
         p:driverClassName="${jdbc.driverClassName}"
         p:url="${jdbc.url}"
         p:username="${jdbc.username}"
         p:password="${jdbc.password}" /-->
   <bean name="seqAlignService" class="service.SeqAlignService">
      <constructor-arg value="/WEB-INF/Res/BLOSUM62.txt"/>
   </bean>
   <!-- ADD PERSISTENCE SUPPORT HERE (jpa, hibernate, etc) -->

</beans>
```

FastaConvertConfirmation.jsp

```
<%--
   Document   : FastaConvertConfirmation
   Created on : May 7, 2009, 12:53:03 PM
   Author     : Michael
--%>
<%@ taglib prefix="c" uri="http://java.sun.com/jsp/jstl/core" %>
<%@ taglib prefix="form" uri="http://www.springframework.org/tags/form" %>
<%@ taglib prefix="spring" uri="http://www.springframework.org/tags" %>
<%@page contentType="text/html" pageEncoding="UTF-8"%>
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN"
  "http://www.w3.org/TR/html4/loose.dtd">

<html>
   <head>
      <meta http-equiv="Content-Type" content="text/html; charset=UTF-8">
      <title>Fasta Conversion Confirmation</title>
   </head>
```

```
<body>
   <h1>Fasta Conversion Confirmation</h1>
   <p>
      Sequence Type: ${fastaForm.seqType}
   </p>
   <em>The results are presented in tablular format so that they can easily be<br/>
   "screen scraped" into a text file</em>
   <c:if test="${fastaForm.seqType == 'proteins'}">
      <table border="1">
         <thead>
            <tr>
               <th align="left">Protein Seqs</th>
            </tr>
         </thead>
         <tbody>
            <c:forEach items="${fastaForm.seqList}" var="seq">
            <tr>
               <td>>${seq.geneProd}<br/></td>
            </tr>
            <tr>
               <td>${seq.cdsTransltn}</td>
            </tr>
            </c:forEach>
         </tbody>
      </table>
      </c:if>

      <c:if test="${fastaForm.seqType == 'nucleotides'}">
      <table border="1">
         <thead>
            <tr>
               <th align="left">Nucleotide Seqs</th>
            </tr>
         </thead>
         <tbody>
            <c:forEach items="${fastaForm.seqList}" var="seq">
            <tr>
               <td>>${seq.geneProd}<br/></td>
            </tr>
            <tr>
               <td>${seq.cdsSeq}</td>
            </tr>
            </c:forEach>
```

```
            </tbody>
          </table>
          </c:if>
   </body>
</html>
```

FastaConvertForm.jsp

```
<%--
   Document   : FastaConvertForm
   Created on : May 7, 2009, 12:52:27 PM
   Author     : Michael
--%>
<%@ taglib prefix="c" uri="http://java.sun.com/jsp/jstl/core" %>
<%@ taglib prefix="form" uri="http://www.springframework.org/tags/form" %>
<%@ taglib prefix="spring" uri="http://www.springframework.org/tags" %>
<%@page contentType="text/html" pageEncoding="UTF-8"%>
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN"
  "http://www.w3.org/TR/html4/loose.dtd">

<html>
   <head>
      <meta http-equiv="Content-Type" content="text/html; charset=UTF-8">
      <title>Fasta Convert</title>
   </head>
   <body>
      <h1>Input GenBank Files and Output Features in Fasta Format.</h1>
      <form:form commandName="fastaForm" action="FastaConvertForm.htm"
      enctype="multipart/form-data">
         <input type="file" name="file"/><br/>
         <form:checkbox path="seqType" label="get protein seqs" value="proteins"/>
         <form:checkbox path="seqType" label="get nuleotide seqs"
value="nucleotides"/>
         <br/>
         <input type="submit" value="submit"/>
      </form:form>
   </body>
</html>
```

FastaSeqConfirmation.jsp

```
<%--
   Document   : FastaSeqConfirmationForm
   Created on : May 10, 2009, 3:44:11 PM
   Author     : Michael
```

```
--%>

<%@ taglib prefix="c" uri="http://java.sun.com/jsp/jstl/core" %>
<%@ taglib prefix="form" uri="http://www.springframework.org/tags/form" %>
<%@ taglib prefix="spring" uri="http://www.springframework.org/tags" %>
<%@page contentType="text/html" pageEncoding="UTF-8"%>
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN"
  "http://www.w3.org/TR/html4/loose.dtd">

<html>
   <head>
      <meta http-equiv="Content-Type" content="text/html; charset=UTF-8">
      <title>JSP Page</title>
   </head>
   <body>
      <h1>Hello World!Page3</h1>
       <form:form commandName="fastaAlignForm" action="FastaUploadForm.htm">

          <table border="1">
             <thead>
                <tr>
                   <th></th>
                   <th></th>
                </tr>
             </thead>
             <tbody>
                <tr>
                   <td valign="top"><c:forEach items="${fastaAlignForm.alteredSeqList}"
var="seq">
                         ${seq.fastaLabel}
                         <strong>
                            Will be compared to->
                         </strong><br/> <hr/>
                      </c:forEach></td>
                   <td valign="top"><c:forEach
items="${fastaAlignForm.alteredSeqList2}" var="seq1">
                         ${seq1.fastaLabel}  <br/> <hr/>

                      </c:forEach></td>
                </tr>
             </tbody>
          </table>
         <input type="submit" value="submit" name="_finish"/>
```

```
      <!--</form>-->

      </form:form>
   </body>
</html>
```

FastaSeqSelectForm.jsp

```
<%--
   Document   : FastaSeqSelectForm
   Created on : May 10, 2009, 3:41:33 PM
   Author     : Michael
--%>

<%@ taglib prefix="form" uri="http://www.springframework.org/tags/form" %>
<%@page contentType="text/html" pageEncoding="UTF-8"%>
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN"
  "http://www.w3.org/TR/html4/loose.dtd">

<html>
   <head>
      <meta http-equiv="Content-Type" content="text/html; charset=UTF-8">
      <title>JSP Page</title>
   </head>
   <body>
      <h1>Hello World!FastaSeqselect</h1>
      <form:form commandName="fastaAlignForm" action="FastaUploadForm.htm">
         <font color="red"><form:errors path="*"
delimiter="<br/>"></form:errors></font>
         <table border="1">
            <thead>
               <tr>
                  <th>seq1</th>
                  <th>seq2</th>
               </tr>
            </thead>
            <tbody>
               <tr>
                  <td valign="top">
                     <form:checkboxes path="alteredIds"
items="${fastaAlignForm.seqList}"
                        itemLabel="fastaLabel" itemValue="id" delimiter="<br/><hr/>"/></td>
                  <td valign="top"><form:checkboxes path="alteredIds2"
items="${fastaAlignForm.seqList2}"
```

```
                    itemLabel="fastaLabel" itemValue="id" delimiter="<br/><hr/>"/></td>
                </tr>
            </tbody>
        </table>

        <input type="submit" value="submit" name="_target2"/>
    </form:form>
  </body>
</html>
```

FastaUploadForm.jsp

```
<%--
   Document   : FastaUploadForm
   Created on : May 10, 2009, 3:39:31 PM
   Author     : Michael
--%>

<%@ taglib prefix="form" uri="http://www.springframework.org/tags/form" %>
<%@page contentType="text/html" pageEncoding="UTF-8"%>
<%@ taglib prefix="c" uri="http://java.sun.com/jsp/jstl/core" %>
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN"
  "http://www.w3.org/TR/html4/loose.dtd">

<html>
  <head>
    <meta http-equiv="Content-Type" content="text/html; charset=UTF-8">
    <title>JSP Page</title>
  </head>
  <body>
    <h1>Hello World! This is the fastaUpload</h1>
    <form:form commandName="fastaAlignForm" action="FastaUploadForm.htm"
    enctype="multipart/form-data">
    <!--<form action="" method="POST">-->
    <table border="1">
        <thead>
          <tr>
            <th>1st GenBank File</th>
            <th>2nd GenBank File</th>
          </tr>
        </thead>
        <tbody>
          <tr>
            <td><input type="file" name="file"/></td>
```

```
                <td><input type="file" name="file2"/></td>
              </tr>
            </tbody>
          </table>
        <input type="submit" value="submit" name="_target1"/><br/>
        <font color="red"><form:errors path="*"
delimiter="<br/>"></form:errors></font>
      <!--</form>-->
      </form:form>
   </body>
</html>
```

FileUploadForm.jsp

```
<%--
   Document   : FileUploadForm
   Created on : Apr 29, 2009, 3:51:55 PM
   Author     : Michael
--%>
<%@ taglib prefix="form" uri="http://www.springframework.org/tags/form" %>
<%@page contentType="text/html" pageEncoding="UTF-8"%>
<%@ taglib prefix="c" uri="http://java.sun.com/jsp/jstl/core" %>
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN"
  "http://www.w3.org/TR/html4/loose.dtd">

<html>
   <head>
      <meta http-equiv="Content-Type" content="text/html; charset=UTF-8">
      <title>JSP Page</title>
   </head>
   <body>
      <h1>Hello World! This is the fileupload</h1>
      <form:form commandName="seqAlignForm" action="FileUploadForm.htm"
      enctype="multipart/form-data">
     <!--<form action="" method="POST">-->
     <table border="1">
            <thead>
               <tr>
                  <th>1st GenBank File</th>
                  <th>2nd GenBank File</th>
               </tr>
            </thead>
            <tbody>
               <tr>
```

```
                    <td><input type="file" name="file"/></td>
                    <td><input type="file" name="file2"/></td>
                </tr>
            </tbody>
        </table>
        <input type="submit" value="submit" name="_target1"/><br/>
        <font color="red"><form:errors path="*"
delimiter="<br/>"></form:errors></font>
        <!-- <c:forEach items="${errors}" var="e">
            <strong font="red">${e}</strong><br/>
        </c:forEach>-->
    <!--</form>-->
    </form:form>
  </body>
</html>
```

SeqConfirmationForm.jsp

```
<%--
  Document   : Page3
  Created on : Apr 30, 2009, 12:15:57 PM
  Author     : Michael
--%>
<%@ taglib prefix="c" uri="http://java.sun.com/jsp/jstl/core" %>
<%@ taglib prefix="form" uri="http://www.springframework.org/tags/form" %>
<%@ taglib prefix="spring" uri="http://www.springframework.org/tags" %>
<%@page contentType="text/html" pageEncoding="UTF-8"%>
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN"
  "http://www.w3.org/TR/html4/loose.dtd">

<html>
  <head>
    <meta http-equiv="Content-Type" content="text/html; charset=UTF-8">
    <title>JSP Page</title>
  </head>
  <body>
    <h1>Hello World!Page3</h1>
     <form:form commandName="seqAlignForm" action="FileUploadForm.htm">
        <table border="1">
            <thead>
                <tr>
                    <th></th>
                    <th></th>
                </tr>
```

```
            </thead>
            <tbody>
               <tr>
                  <td valign="top"><c:forEach items="${seqAlignForm.alteredSeqList}"
var="seq">
                        ${seq.geneProdTrans}
                        <strong>
                           Will be compared to->
                        </strong><br/> <hr/>
                     </c:forEach></td>
                  <td valign="top"><c:forEach items="${seqAlignForm.alteredSeqList2}"
var="seq1">
                        ${seq1.geneProdTrans}  <br/> <hr/>

                     </c:forEach></td>
               </tr>
            </tbody>
         </table>
        <input type="submit" value="submit" name="_finish"/>
     <!--</form>-->

     </form:form>
   </body>
</html>
```

SeqSelectForm.jsp

```
<%--
   Document   : SeqSelectForm
   Created on : Apr 29, 2009, 3:57:01 PM
   Author     : Michael
--%>
<%@ taglib prefix="form" uri="http://www.springframework.org/tags/form" %>
<%@page contentType="text/html" pageEncoding="UTF-8"%>
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN"
  "http://www.w3.org/TR/html4/loose.dtd">

<html>
   <head>
      <meta http-equiv="Content-Type" content="text/html; charset=UTF-8">
      <title>JSP Page</title>
   </head>
   <body>
      <h1>Hello World!Seqselect</h1>
```

```
      <form:form commandName="seqAlignForm" action="FileUploadForm.htm">
         <font color="red"><form:errors path="*"
delimiter="<br/>"></form:errors></font>
         <table border="1">
            <thead>
               <tr>
                  <th>seq1</th>
                  <th>seq2</th>
               </tr>
            </thead>
            <tbody>
               <tr>
                  <td valign="top">
                     <form:checkboxes path="alteredIds" items="${seqAlignForm.seqList}"
                     itemLabel="geneProdTrans" itemValue="id"
delimiter="<br/><hr/>"/></td>
                  <td valign="top"><form:checkboxes path="alteredIds2"
items="${seqAlignForm.seqList2}"
                     itemLabel="geneProdTrans" itemValue="id"
delimiter="<br/><hr/>"/></td>
               </tr>
            </tbody>
         </table>

        <input type="submit" value="submit" name="_target2"/>
      </form:form>
   </body>
</html>
```

redirect.jsp

```
<%--
Views should be stored under the WEB-INF folder so that
they are not accessible except through controller process.

This JSP is here to provide a redirect to the dispatcher
servlet but should be the only JSP outside of WEB-INF.
--%>
<%@page contentType="text/html" pageEncoding="UTF-8"%>
<% response.sendRedirect("index.htm"); %>
```

FastaAlignController

```
/*
 * To change this template, choose Tools | Templates
```

```
 * and open the template in the editor.
 */

package controller;

import java.util.ArrayList;
import java.util.HashMap;
import java.util.Map;
import javax.servlet.http.HttpServletRequest;
import javax.servlet.http.HttpServletResponse;
import org.springframework.beans.propertyeditors.CustomCollectionEditor;
import org.springframework.validation.BindException;
import org.springframework.validation.Errors;
import org.springframework.web.bind.ServletRequestDataBinder;
import org.springframework.web.servlet.ModelAndView;
import org.springframework.web.servlet.mvc.AbstractWizardFormController;
import service.Featr;
import service.SeqAlignService;

/**
 *
 * @author Michael
 */
public class FastaAlignController extends AbstractWizardFormController {
    private SeqAlignService seqAlignService;

    public void setSeqAlignService(SeqAlignService seqAlignService) {
        this.seqAlignService = seqAlignService;
    }

    public FastaAlignController() {
        //Initialize controller properties here or
        //in the Web Application Context

        //setCommandClass(controller.SeqAlignBean.class);
        //setCommandName("MyCommandName");
        //setSuccessView("successView");
        //setFormView("formView");
        //setCommandName("seqAlignForm");
                    //setPages(new String[] {"FileUploadForm", "SeqSelectForm", "Page3"});
        System.out.println("Constructor:::::::");
    }
```

```
    @Override
    protected Object formBackingObject(HttpServletRequest request) throws Exception {
        SeqAlignBean fastaAlignForm = new SeqAlignBean();
        //seqAlignForm.setDum("dum");
        System.out.println("FBO:::::::: "+ fastaAlignForm.toString());
        return fastaAlignForm;
    }

    @Override
    protected Map referenceData(HttpServletRequest request, Object command, Errors
errors, int page) throws Exception {
        SeqAlignBean  bean = (SeqAlignBean) command;
        if (getCurrentPage(request)==1){
            Map<Object, Object> dataMap = new HashMap<Object, Object>();
            bean.setSeqList(seqAlignService.makeFeatureListFromFastaFile(bean.getFile()));

bean.setSeqList2(seqAlignService.makeFeatureListFromFastaFile(bean.getFile2()));
            //seqAlignService.getProtAlignmentsAndIds("MMMPPVVVM",
"MMMVVVM");
            //bean.setDum(Integer.toString(bean.getSeqList().size()));
            //dataMap.put("seqList", bean.getSeqList());
            System.out.println("RefDataIF0::::::::::::"+getCurrentPage(request));
            return dataMap;
        }
        if (getCurrentPage(request)==2){
            Map<Object, Object> dataMap = new HashMap<Object, Object>();
            bean.setAlteredSeqList();
            bean.setAlteredSeqList2();
            //bean.setDum(Integer.toString(bean.getAlteredSeqList().size()));
            //dataMap.put("seqList", bean.getSeqList());
            System.out.println("RefDataIF2::::::::::::"+getCurrentPage(request));
            return dataMap;
        }
        System.out.println("RefData::::::::::::"+getCurrentPage(request));


        return super.referenceData(request, command, errors, page);
    }



    @Override
```

```
   protected void onBind(HttpServletRequest request, Object command, BindException
errors) throws Exception {
      SeqAlignBean fastaAlignForm = (SeqAlignBean) command;
      if (getCurrentPage(request)==0){

//fastaAlignForm.setSeqList(seqAlignService.makeFeatureListFromFastaFile(fastaAlign
Form.getFile()));

//fastaAlignForm.setSeqList2(seqAlignService.makeFeatureListFromFastaFile(fastaAlig
nForm.getFile2()));
         //seqAlignService.getProtAlignmentsAndIds("MMMPPVVVM",
"MMMVVVM");
         //fastaAlignForm.setDum(Integer.toString(fastaAlignForm.getSeqList().size()));
      }
      if (getCurrentPage(request)==1){
         System.out.println("errors:::"+errors.getAllErrors().toString());

//seqAlignForm.setSeqList(seqAlignService.makeCodingFeatureListFromFile(seqAlignF
orm.getFile()));
//          fastaAlignForm.setAlteredSeqList();
//          fastaAlignForm.setAlteredSeqList2();
//
fastaAlignForm.setDum(Integer.toString(fastaAlignForm.getAlteredSeqList().size()));
      }
      //seqAlignForm.setDum(Integer.toString(seqAlignForm.getPst().length));
      System.out.println("OnBind::::::"+ getCurrentPage(request));
   }

   @Override
   protected void validatePage(Object command, Errors errors, int page) {
      SeqAlignBean seqAlignForm = (SeqAlignBean) command;
      System.out.println("Validate::::::::");
      if (page==0){
         System.out.println("Validate::::::::0" + errors.getErrorCount());
//          ValidationUtils.rejectIfEmpty(errors, "file", "2 genbank files must be included");
         if(seqAlignForm.getFile().getOriginalFilename().isEmpty()){
            System.out.println("Validate::::::::File");
            errors.rejectValue("file", "2 genbank files must be included", "The 1st fasta file
must be included");
         }if(seqAlignForm.getFile2().getOriginalFilename().isEmpty()){
            System.out.println("Validate::::::::File");
            errors.rejectValue("file2", "2 genbank files must be included", "The 2nd fasta
file must be included");
```

```
            }
            System.out.println("Validate::::::::0afte" + errors.getErrorCount());
        }
        if (page==1){
            if(seqAlignForm.getSeqList().size() -
                    seqAlignForm.getAlteredIds().length!=
                    seqAlignForm.getSeqList2().size() -
                    seqAlignForm.getAlteredIds2().length){
                System.out.println("Validate::::::::page1");
                errors.reject("", "Both columns of the table must contain the same number of
seqs!");

            }
        }
        //super.validatePage(command, errors, page);
    }


    @Override
    protected ModelAndView processFinish(HttpServletRequest request,
HttpServletResponse response, Object command, BindException errors) throws
Exception {
        //throw new UnsupportedOperationException("Not supported yet.");
        SeqAlignBean seqAlignForm = (SeqAlignBean) command;
        seqAlignForm.setAligns(
                seqAlignService.getProtAlignsFromFastaFeatrLists(
                seqAlignForm.getAlteredSeqList(),
                seqAlignForm.getAlteredSeqList2()));
        System.out.println("Finish::::::");
        return new ModelAndView("AlignConf", "seqAlignForm", seqAlignForm);
    }



}
```

FastaCovertBean.java

```
/*
 * To change this template, choose Tools | Templates
 * and open the template in the editor.
 */

package controller;
```

```
import java.util.ArrayList;
import org.springframework.web.multipart.MultipartFile;
import service.Featr;

/**
 *
 * @author Michael
 */
public class FastaConvertBean {
    private MultipartFile file;
    private ArrayList<Featr> seqList;
    private String seqType;

    public String getSeqType() {
        return seqType;
    }

    public void setSeqType(String seqType) {
        this.seqType = seqType;
    }



    public MultipartFile getFile() {
        return file;
    }

    public void setFile(MultipartFile file) {
        this.file = file;
    }

    public ArrayList<Featr> getSeqList() {
        return seqList;
    }

    public void setSeqList(ArrayList<Featr> seqList) {
        this.seqList = seqList;
    }


}
```

FastaConvertFormController.java

```
/*
 * To change this template, choose Tools | Templates
 * and open the template in the editor.
 */

package controller;

import javax.servlet.http.HttpServletRequest;
import javax.servlet.http.HttpServletResponse;
import org.springframework.validation.BindException;
import org.springframework.web.multipart.MultipartFile;
import org.springframework.web.servlet.ModelAndView;
import org.springframework.web.servlet.mvc.SimpleFormController;
import service.SeqAlignService;

/**
 *
 * @author Michael
 */
public class FastaConvertFormController extends SimpleFormController {
    private SeqAlignService seqAlignService;

    public void setSeqAlignService(SeqAlignService seqAlignService) {
        this.seqAlignService = seqAlignService;
    }

    public FastaConvertFormController() {
        //Initialize controller properties here or
        //in the Web Application Context

        //setCommandClass(MyCommand.class);
        //setCommandName("MyCommandName");
        //setSuccessView("successView");
        //setFormView("formView");
    }

//    @Override
//    protected void doSubmitAction(Object command) throws Exception {
//        throw new UnsupportedOperationException("Not yet implemented");
//    }

    //Use onSubmit instead of doSubmitAction
```

```
   //when you need access to the Request, Response, or BindException objects

   @Override
   protected ModelAndView onSubmit(
         HttpServletRequest request,
         HttpServletResponse response,
         Object command,
         BindException errors) throws Exception {
      FastaConvertBean fastaForm = (FastaConvertBean) command;
      MultipartFile file = fastaForm.getFile();
      fastaForm.setSeqList(seqAlignService.makeCodingFeatureListFromFile(file));
      ModelAndView mv = new ModelAndView(getSuccessView(),"fastaForm",
fastaForm);
      //Do something...
      return mv;
   }


}
```

SeqAlignBean.java

```
/*
 * To change this template, choose Tools | Templates
 * and open the template in the editor.
 */

package controller;

import java.util.ArrayList;
import org.springframework.web.multipart.MultipartFile;
import service.Align;
import service.Featr;

/**
 *
 * @author Michael
 */
public class SeqAlignBean {
   private MultipartFile file;
   private MultipartFile file2;
   private ArrayList<Featr> seqList;
   private ArrayList<Featr> seqList2;
   private String[] alteredIds;
```

```
private String[] alteredIds2;
private ArrayList<Featr> alteredSeqList;
private ArrayList<Featr> alteredSeqList2;
private ArrayList<Align> aligns;
private String dum;

public ArrayList<Align> getAligns() {
   return aligns;
}

public void setAligns(ArrayList<Align> aligns) {
   this.aligns = aligns;
}


public String[] getAlteredIds2() {
   return alteredIds2;
}

public void setAlteredIds2(String[] alteredIds2) {
   this.alteredIds2 = alteredIds2;
}

public ArrayList<Featr> getAlteredSeqList2() {
   return alteredSeqList2;
}

public void setAlteredSeqList2() {
   ArrayList<Featr> alteredSeqLists= new ArrayList<Featr>();
   for(Featr f: this.seqList2){
      String go = "go";
      for(String id:this.alteredIds2){
         if(f.getId().equals(id)){
            go = "nogo";
         }
      }
      if(go.equals("go")){
         alteredSeqLists.add(f);
      }
   }
   this.alteredSeqList2 = alteredSeqLists;
}
```

```
public ArrayList<Featr> getSeqList2() {
   return seqList2;
}

public void setSeqList2(ArrayList<Featr> seqList2) {
   this.seqList2 = seqList2;
}


public String[] getAlteredIds() {
   return alteredIds;
}

public void setAlteredIds(String[] alteredIds) {
   this.alteredIds = alteredIds;
}

public String getDum() {
   return dum;
}

public void setDum(String dum) {
   this.dum = dum;
}

public ArrayList<Featr> getAlteredSeqList() {
   return alteredSeqList;
}

public void setAlteredSeqList() {
   ArrayList<Featr> alteredSeqLists= new ArrayList<Featr>();
   for(Featr f: this.seqList){
      String go = "go";
      for(String id:this.alteredIds){
         if(f.getId().equals(id)){
            go = "nogo";
         }
      }
      if(go.equals("go")){
         alteredSeqLists.add(f);
      }
   }
   this.alteredSeqList = alteredSeqLists;
```

```
    }

    public MultipartFile getFile() {
        return file;
    }

    public void setFile(MultipartFile file) {
        this.file = file;
    }

    public MultipartFile getFile2() {
        return file2;
    }

    public void setFile2(MultipartFile file2) {
        this.file2 = file2;
    }

    public ArrayList<Featr> getSeqList() {
        return seqList;
    }

    public void setSeqList(ArrayList<Featr> seqList) {
        this.seqList = seqList;
    }


}
```

SeqAlignController.java

```
/*
 * To change this template, choose Tools | Templates
 * and open the template in the editor.
 */

package controller;

import java.util.ArrayList;
import java.util.HashMap;
import java.util.Map;
import javax.servlet.http.HttpServletRequest;
import javax.servlet.http.HttpServletResponse;
import org.springframework.beans.propertyeditors.CustomCollectionEditor;
```

```
import org.springframework.validation.BindException;
import org.springframework.validation.Errors;
import org.springframework.validation.ValidationUtils;
import org.springframework.web.bind.ServletRequestDataBinder;
import org.springframework.web.servlet.ModelAndView;
import org.springframework.web.servlet.mvc.AbstractWizardFormController;
import service.Featr;
import service.SeqAlignService;

/**
 *
 * @author Michael
 */
public class SeqAlignController extends AbstractWizardFormController {
    private SeqAlignService seqAlignService;

    public void setSeqAlignService(SeqAlignService seqAlignService) {
        this.seqAlignService = seqAlignService;
    }

    public SeqAlignController() {
        //Initialize controller properties here or
        //in the Web Application Context

        //setCommandClass(controller.SeqAlignBean.class);
        //setCommandName("MyCommandName");
        //setSuccessView("successView");
        //setFormView("formView");
        //setCommandName("seqAlignForm");
                    //setPages(new String[] {"FileUploadForm", "SeqSelectForm", "Page3"});
        System.out.println("Constructor:::::::");
    }

    @Override
    protected Object formBackingObject(HttpServletRequest request) throws Exception {
        SeqAlignBean seqAlignForm = new SeqAlignBean();
        //seqAlignForm.setDum("dum");
        System.out.println("FBO:::::::: "+ seqAlignForm.toString());
        return seqAlignForm;
    }

    @Override
```

```
   protected Map referenceData(HttpServletRequest request, Object command, Errors
errors, int page) throws Exception {
      SeqAlignBean  bean = (SeqAlignBean) command;
      System.out.println("Ref::::");
      if (getCurrentPage(request)==1){
         System.out.println("RefDataIF0::::::::::::"+getCurrentPage(request));
         Map<Object, Object> dataMap = new HashMap<Object, Object>();

bean.setSeqList(seqAlignService.makeCodingFeatureListFromFile(bean.getFile()));

bean.setSeqList2(seqAlignService.makeCodingFeatureListFromFile(bean.getFile2()));
         //seqAlignService.getProtAlignmentsAndIds("MMMPPVVVM",
"MMMVVVM");
         //bean.setDum(Integer.toString(bean.getSeqList().size()));
         //dataMap.put("seqList", bean.getSeqList());
         return dataMap;
      }
      if (getCurrentPage(request)==2){
         Map<Object, Object> dataMap = new HashMap<Object, Object>();
         bean.setAlteredSeqList();
         bean.setAlteredSeqList2();
         //bean.setDum(Integer.toString(bean.getAlteredSeqList().size()));
         //dataMap.put("seqList", bean.getSeqList());
         System.out.println("RefDataIF2::::::::::::"+getCurrentPage(request));
         return dataMap;
      }
      System.out.println("RefData::::::::::::"+getCurrentPage(request));


      return super.referenceData(request, command, errors, page);
   }

   @Override
   protected void validatePage(Object command, Errors errors, int page) {
      SeqAlignBean seqAlignForm = (SeqAlignBean) command;
      System.out.println("Validate::::::::");
      if (page==0){
         System.out.println("Validate::::::::0" + errors.getErrorCount());
//          ValidationUtils.rejectIfEmpty(errors, "file", "2 genbank files must be included");
         if(seqAlignForm.getFile().getOriginalFilename().isEmpty()){
            System.out.println("Validate::::::::File");
            errors.rejectValue("file", "2 genbank files must be included", "The 1st genbank
files must be included");
```

```
        }if(seqAlignForm.getFile2().getOriginalFilename().isEmpty()){
           System.out.println("Validate::::::::File");
           errors.rejectValue("file2", "2 genbank files must be included", "The 2nd
genbank files must be included");
        }
        System.out.println("Validate::::::::0afte" + errors.getErrorCount());
     }
     if (page==1){
        if(seqAlignForm.getSeqList().size() -
              seqAlignForm.getAlteredIds().length!=
              seqAlignForm.getSeqList2().size() -
              seqAlignForm.getAlteredIds2().length){
           System.out.println(seqAlignForm.getAlteredIds().length+
                 ", "+ seqAlignForm.getAlteredIds2().length);
           System.out.println("Validate::::::::page1");
           errors.reject("", "Both columns of the table must contain the same number of
seqs!");

        }
     }
     //super.validatePage(command, errors, page);
  }



  @Override
  protected void onBind(HttpServletRequest request, Object command, BindException
errors) throws Exception {
     SeqAlignBean seqAlignForm = (SeqAlignBean) command;
     if (getCurrentPage(request)==0){

//seqAlignForm.setSeqList(seqAlignService.makeCodingFeatureListFromFile(seqAlignF
orm.getFile()));

//seqAlignForm.setSeqList2(seqAlignService.makeCodingFeatureListFromFile(seqAlign
Form.getFile2()));
        ////seqAlignService.getProtAlignmentsAndIds("MMMPPVVVM",
"MMMVVVM");
        //seqAlignForm.setDum(Integer.toString(seqAlignForm.getSeqList().size()));
     }
     if (getCurrentPage(request)==1){
        System.out.println("errors:::"+errors.getAllErrors().toString());
```

```
//seqAlignForm.setSeqList(seqAlignService.makeCodingFeatureListFromFile(seqAlignF
orm.getFile()));
//          seqAlignForm.setAlteredSeqList();
//          seqAlignForm.setAlteredSeqList2();
//
seqAlignForm.setDum(Integer.toString(seqAlignForm.getAlteredSeqList().size()));
        }
        //seqAlignForm.setDum(Integer.toString(seqAlignForm.getPst().length));
        System.out.println("OnBind::::::"+ getCurrentPage(request));
    }



    @Override
    protected ModelAndView processFinish(HttpServletRequest request,
HttpServletResponse response, Object command, BindException errors) throws
Exception {
        //throw new UnsupportedOperationException("Not supported yet.");
        SeqAlignBean seqAlignForm = (SeqAlignBean) command;
        seqAlignForm.setAligns(
                seqAlignService.getProtAlignsFromFeatrLists(
                seqAlignForm.getAlteredSeqList(),
                seqAlignForm.getAlteredSeqList2()));
        System.out.println("Finish::::::");
        return new ModelAndView("AlignConf", "seqAlignForm", seqAlignForm);
    }



}
```

Align.java

```
/*
 * To change this template, choose Tools | Templates
 * and open the template in the editor.
 */

package service;

/**
 *
 * @author Michael
```

```
 */
public class Align {
    private String queryLabel;
    private String targetLabel;
    private String queryAligned;
    private String targetAligned;
    private double percentId;

    public double getPercentId() {
        return percentId;
    }

    public void setPercentId(double percentId) {
        this.percentId = percentId;
    }

    public String getQueryAligned() {
        return queryAligned;
    }

    public void setQueryAligned(String queryAligned) {
        this.queryAligned = queryAligned;
    }

    public String getQueryLabel() {
        return queryLabel;
    }

    public void setQueryLabel(String queryLabel) {
        this.queryLabel = queryLabel;
    }

    public String getTargetAligned() {
        return targetAligned;
    }

    public void setTargetAligned(String targetAligned) {
        this.targetAligned = targetAligned;
    }

    public String getTargetLabel() {
        return targetLabel;
    }
```

```
    public void setTargetLabel(String targetLabel) {
        this.targetLabel = targetLabel;
    }



}
```

Featr.java

```
/*
 * To change this template, choose Tools | Templates
 * and open the template in the editor.
 */

package service;

/**
 *
 * @author Michael
 */
public class Featr {
    private String srcAccNum;
    private String id;
    private String cdsGene = null;
    private String cdsProduct;
    private String cdsLoc = null;
    private int ftrStart;
    private String cdsNote = null;
    private String cdsProtID = null;
    private String cdsSeq = null;
    private String cdsTransltn = null;
    private String ftrName = null;
    private String ftrNote = null;
    private String ftrGene = null;
    private String ftrLoc = null;
    private String ftrProduct = null;
    private String ftrSeq = null;
    private String geneProdTrans = null;
    private String geneProd = null;
    private String fastaLabel = null;

    public Featr(){
```

```
}

public String getFastaLabel() {
   fastaLabel = this.cdsGene + "<br/>"+
         formatLongStr(this.cdsSeq);
   return fastaLabel;
}



public String getGeneProd() {
   geneProd = this.cdsGene + "|"+
         this.cdsProduct;
   return geneProd;
}



public String getGeneProdTrans() {
   geneProdTrans = this.cdsGene + "|"+
         this.cdsProduct+"<br/>"+
         formatLongStr(this.cdsTransltn);
   return geneProdTrans;
}

public String getId() {
   return id;
}

public void setId(String id) {
   this.id = id;
}

public String getCdsGene() {
   return cdsGene;
}

public void setCdsGene(String cdsGene) {
   this.cdsGene = cdsGene;
}

public String getCdsLoc() {
   return cdsLoc;
```

```
}

public void setCdsLoc(String cdsLoc) {
   this.cdsLoc = cdsLoc;
}

public String getCdsNote() {
   return cdsNote;
}

public void setCdsNote(String cdsNote) {
   this.cdsNote = cdsNote;
}

public String getCdsProduct() {
   return cdsProduct;
}

public void setCdsProduct(String cdsProduct) {
   this.cdsProduct = cdsProduct;
}

public String getCdsProtID() {
   return cdsProtID;
}

public void setCdsProtID(String cdsProtID) {
   this.cdsProtID = cdsProtID;
}

public String getCdsSeq() {
   return cdsSeq;
}

public void setCdsSeq(String cdsSeq) {
   this.cdsSeq = cdsSeq;
}

public String getCdsTransltn() {
   return cdsTransltn;
}

public void setCdsTransltn(String cdsTransltn) {
```

```
    this.cdsTransltn = cdsTransltn;
}

public String getFtrGene() {
    return ftrGene;
}

public void setFtrGene(String ftrGene) {
    this.ftrGene = ftrGene;
}

public String getFtrLoc() {
    return ftrLoc;
}

public void setFtrLoc(String ftrLoc) {
    this.ftrLoc = ftrLoc;
}

public String getFtrName() {
    return ftrName;
}

public void setFtrName(String ftrName) {
    this.ftrName = ftrName;
}

public String getFtrNote() {
    return ftrNote;
}

public void setFtrNote(String ftrNote) {
    this.ftrNote = ftrNote;
}

public String getFtrProduct() {
    return ftrProduct;
}

public void setFtrProduct(String ftrProduct) {
    this.ftrProduct = ftrProduct;
}
```

```
    public String getFtrSeq() {
        return ftrSeq;
    }

    public void setFtrSeq(String ftrSeq) {
        this.ftrSeq = ftrSeq;
    }

    public int getFtrStart() {
        return ftrStart;
    }

    public void setFtrStart(int ftrStart) {
        this.ftrStart = ftrStart;
    }

    public String getSrcAccNum() {
        return srcAccNum;
    }

    public void setSrcAccNum(String srcAccNum) {
        this.srcAccNum = srcAccNum;
    }

    public String formatLongStr(String lonstr){
        String shrt = null;
        shrt = lonstr.substring(0,10) + ".....";
        shrt = shrt + lonstr.substring(lonstr.length() -  10);
        return shrt;
    }


}
```

SeqAlignService.java

```
/*
 * To change this template, choose Tools | Templates
 * and open the template in the editor.
 */

package service;

import java.io.BufferedReader;
```

```
import java.io.IOException;
import java.io.InputStreamReader;
import java.util.ArrayList;
import java.util.Iterator;
import java.util.NoSuchElementException;
import org.biojava.bio.BioException;
import org.biojava.bio.alignment.NeedlemanWunsch;
import org.biojava.bio.alignment.SequenceAlignment;
import org.biojava.bio.alignment.SubstitutionMatrix;
import org.biojava.bio.seq.FeatureFilter;
import org.biojava.bio.seq.FeatureHolder;
import org.biojava.bio.seq.ProteinTools;
import org.biojava.bio.seq.Sequence;
import org.biojava.bio.seq.SequenceTools;
import org.biojava.bio.seq.io.SubIntegerTokenization;
import org.biojava.bio.seq.io.SymbolTokenization;
import org.biojava.bio.symbol.Alignment;
import org.biojava.bio.symbol.AlphabetManager;
import org.biojava.bio.symbol.FiniteAlphabet;
import org.biojava.bio.symbol.IllegalAlphabetException;
import org.biojava.bio.symbol.IllegalSymbolException;
import org.biojava.bio.symbol.Symbol;
import org.biojavax.RichAnnotation;
import org.biojavax.bio.seq.RichFeature;
import org.biojavax.bio.seq.RichLocation;
import org.biojavax.bio.seq.RichSequence;
import org.biojavax.bio.seq.RichSequenceIterator;
import org.springframework.core.io.Resource;
import org.springframework.web.context.WebApplicationContext;
import org.springframework.web.multipart.MultipartFile;

/**
 *
 * @author Michael
 */
public class SeqAlignService {

    public SeqAlignService(Resource protMat) throws IOException {
                setProtMat(protMat);
       System.out.println("file:::::"+getProtMat().getFile().length());
                //resource.getInputStream();
                //resource.exists();
```

```
        }

    private Resource protMat;

    public Resource getProtMat() {
        return protMat;
    }

    public void setProtMat(Resource protMat) {
        this.protMat = protMat;
    }

    public ArrayList<Featr> makeFeatureListFromFastaFile(MultipartFile file) throws
IOException, NoSuchElementException, BioException{
        BufferedReader b = new BufferedReader ( new InputStreamReader (
file.getInputStream() ) );
        RichSequenceIterator seqIt =
org.biojavax.bio.seq.RichSequence.IOTools.readFastaProtein(b, null);

        ArrayList<Featr> list = new ArrayList<Featr>();
        int id = 0;
        while(seqIt.hasNext()){
            Featr f = new Featr();
            RichSequence seq = seqIt.nextRichSequence();
            f.setCdsGene(seq.getURN());
            f.setCdsSeq(seq.seqString());
            f.setId(Integer.toString(id));
            id = id + 1;
            list.add(f);
        }
        return list;
    }


    public ArrayList<Featr> makeCodingFeatureListFromFile(MultipartFile file)
            throws IllegalAlphabetException, IllegalSymbolException, IOException,
BioException{
        //make a holder for the coding filter
        RichSequenceIterator seqIt  = getSeqIterFromMultipartFile(file);
        //FeatureHolder ftrHld = getCdFtrHld(seqIt);
        FtrHldAndSeq fHS = getCdFtrHldandSeq(seqIt);

//        RichSequence seq = seqIt.nextRichSequence();
```

```
      //FeatureFilter src = new FeatureFilter.ByType("source");
      //FeatureHolder s = seq.filter(src);
      ArrayList<Featr> l = new ArrayList<Featr>();
      //iterate over the Features in fh
      int id = 0;
      for (Iterator i = fHS.getHld().features(); i.hasNext(); ){
         Featr ftr = populateCdFeatr(fHS.getSeq(), (RichFeature)i.next());
         ftr.setId(Integer.toString(id));
         id = id + 1;
         l.add(ftr);
      }
      return l;
   }

   public RichSequenceIterator getSeqIterFromMultipartFile(MultipartFile file) throws
IOException{
      BufferedReader b = new BufferedReader ( new InputStreamReader (
file.getInputStream() ) );
      RichSequenceIterator seqIt =
org.biojavax.bio.seq.RichSequence.IOTools.readGenbankDNA(b, null);
      return seqIt;
   }

   public FeatureFilter getCodingFilter(){
      FeatureFilter fCode = new FeatureFilter.ByType("CDS");
      return fCode;
   }

   public FtrHldAndSeq getCdFtrHldandSeq(RichSequenceIterator seqIt)
         throws BioException, NoSuchElementException{
      FtrHldAndSeq ftrHldSeq = new FtrHldAndSeq();
      FeatureFilter codeFltr = getCodingFilter();
      RichSequence seq = seqIt.nextRichSequence();
      ftrHldSeq.setSeq(seq);
      ftrHldSeq.setHld(seq.filter(codeFltr));
      return ftrHldSeq;
   }

   public Featr populateCdFeatr(RichSequence seq, RichFeature f
         ) throws IllegalAlphabetException, IllegalSymbolException{
      Featr ftr = new Featr();
      ftr.setSrcAccNum(seq.getAccession());
      //parseSrcInfo(fh, ftr);
```

```
    String gn = "";
    String prd = "";
    //ftr.setCdsSeq(getSequenceMotifNoSplice(f));
    ftr.setCdsSeq(getSequenceMotif(f));
    ftr.setCdsLoc(formatLocation(f, ftr));
    RichAnnotation an = (RichAnnotation) f.getAnnotation();
    if(an.containsProperty("protein_id")){
       ftr.setCdsProtID(an.getProperty("protein_id").toString());
    }
    if(an.containsProperty("gene")){
       //gn = an.getProperty("gene").toString();
       ftr.setCdsGene(an.getProperty("gene").toString());
    }
    if(an.containsProperty("product")){
       //prd = an.getProperty("product").toString();
       ftr.setCdsProduct(an.getProperty("product").toString());
    }
    if(an.containsProperty("translation")){
       ftr.setCdsTransltn(an.getProperty("translation").toString());
    }
    if(an.containsProperty("note")){
       ftr.setCdsNote(an.getProperty("note").toString());
    }
    //ftr.setCdsGeneProduct(gn+" "+prd);
    return ftr;
}

private String formatLocation(RichFeature ftr, Featr ft){
    String loc = null;
    ft.setFtrStart(ftr.getLocation().getMin());
    if (ftr.getLocation().toString().contains(":")){
       loc = ftr.getLocation().toString();
       loc = loc.substring(loc.lastIndexOf(":")+2,loc.lastIndexOf("]") );
       }
    else{
       loc = ftr.getLocation().toString();
       }
    if (ftr.getStrand().getValue() < 0){
       loc = "(" + loc + ")c";
    }
    if (loc.contains(",")){
       loc = loc.replaceAll(",", ", ");
    }
```

```
        return loc;
    }

    private String getSequenceMotif(RichFeature rf) throws IllegalAlphabetException,
IllegalSymbolException{
        Sequence sq = rf.getSequence();
        RichLocation l = (RichLocation) rf.getLocation();
        Sequence s = null;
        String sb = "";
        for (Iterator i = l.blockIterator(); i.hasNext();){
            RichLocation blk = (RichLocation) i.next();
            if (rf.getStrand().getValue() < 0){
                s = SequenceTools.subSequence(sq, blk.getMin(), blk.getMax());
                sb = sb + SequenceTools.reverseComplement(s).seqString();
            }
            else{
                s = SequenceTools.subSequence(sq, blk.getMin(), blk.getMax());
                sb = sb + s.seqString();
            }
        }
        //System.out.println(sb);
        return sb;
    }

    public Align getProtAlignmentsAndIds(String queryStr, String targetStr) throws
BioException, NumberFormatException, IOException, Exception{
        Align protAlign = new Align();
        // The alphabet of the sequences. For this example DNA is choosen.
        FiniteAlphabet alphabet =
                (FiniteAlphabet) AlphabetManager.alphabetForName("PROTEIN-TERM");
        // Read the substitution matrix file.
        // For this example the matrix NUC.4.4 is good.
        SubstitutionMatrix matrix =
                new SubstitutionMatrix(alphabet, protMat.getFile());
        // Define the default costs for sequence manipulation for the global alignment.
        SequenceAlignment aligner = new NeedlemanWunsch(
            (short)0,          // match
            (short)3,          // replace
            (short)2,      // insert
            (short)2,          // delete
            (short)1,      // gapExtend
            matrix  // SubstitutionMatrix
        );
```

```
        Sequence query  = ProteinTools.createProteinSequence(queryStr, "query");
        Sequence target = ProteinTools.createProteinSequence(targetStr, "target");
//          Sequence query  = DNATools.createDNASequence(queryStr, "query");
//          Sequence target = DNATools.createDNASequence(targetStr, "target");
//          // Perform an alignment and save the results.
        aligner.pairwiseAlignment(
           query, // first sequence
           target // second one
        );
//      // Print the alignment to the screen
//          System.out.println("Global alignment with Needleman-Wunsch:\n" +
//                aligner.getAlignmentString()+
//                aligner.getAlignment(query,
target).symbolListForLabel(target.getName()).seqString());
          //Save the info to the Aign object
          protAlign.setQueryAligned(aligner.getAlignment(
                query, target).symbolListForLabel(
                query.getName()).seqString());
          protAlign.setTargetAligned(aligner.getAlignment(
                query, target).symbolListForLabel(
                target.getName()).seqString());
          Alignment alignment = aligner.getAlignment(query, target);
          int matches = 0;
          for (int i = 1; i <= alignment.length(); i++) {
             Symbol querySym = alignment.symbolAt(query.getName(), i);
             Symbol subjectSym = alignment.symbolAt(target.getName(), i);
             if (querySym!=null && querySym.equals(subjectSym)) matches++;
          }
          protAlign.setPercentId((double)matches/(double)alignment.length());
          //identity = (double)matches/(double)alignment.length();
          //System.out.println(protAlign.getPercentId());
          return protAlign;
    }

    public ArrayList<Align> getProtAlignsFromFeatrLists(ArrayList<Featr> queryList,
          ArrayList<Featr> targetList) throws BioException, NumberFormatException,
IOException, Exception{
        ArrayList<Align> alignList = new ArrayList<Align>();
        int i = 0;
        for(Featr qF: queryList){
          Featr tF = targetList.get(i);
          i = i+1;
          Align al = getProtAlignmentsAndIds(qF.getCdsTransltn(), tF.getCdsTransltn());
```

```
            al.setQueryLabel(qF.getGeneProd()); al.setTargetLabel(tF.getGeneProd());
            alignList.add(al);
        }

        return alignList;
    }

    public ArrayList<Align> getProtAlignsFromFastaFeatrLists(ArrayList<Featr>
queryList,
            ArrayList<Featr> targetList) throws BioException, NumberFormatException,
IOException, Exception{
        ArrayList<Align> alignList = new ArrayList<Align>();
        int i = 0;
        for(Featr qF: queryList){
            Featr tF = targetList.get(i);
            i = i+1;
            Align al = getProtAlignmentsAndIds(qF.getCdsSeq(), tF.getCdsSeq());
            al.setQueryLabel(qF.getCdsGene()); al.setTargetLabel(tF.getCdsGene());
            alignList.add(al);
        }

        return alignList;
    }


    public class FtrHldAndSeq{
        private FeatureHolder hld;
        private RichSequence seq;

        public FeatureHolder getHld() {
            return hld;
        }

        public void setHld(FeatureHolder hld) {
            this.hld = hld;
        }

        public RichSequence getSeq() {
            return seq;
        }

        public void setSeq(RichSequence seq) {
            this.seq = seq;
```

```
        }

    }
}
```

REFERENCES

REFERENCES

1. Rowe, W.P., Huebner, R.J., Gilmore, L.K., Parrott, R.H., and Ward, T.G. Isolation of a cytopathogenic agent from human adenoids undergoing spontaneous degeneration in tissue culture. *Proceedings of the Society for Experimental Biology and Medicine. Society for Experimental Biology and Medicine (New York, N.Y 84*, 3 (1953), 570-3.
2. Hilleman, M.R. and Werner, J.H. Recovery of new agent from patients with acute respiratory illness. *Proceedings of the Society for Experimental Biology and Medicine. Society for Experimental Biology and Medicine (New York, N.Y 85*, 1 (1954), 183-8.
3. Davison, A.J., Benko, M., and Harrach, B. Genetic content and evolution of adenoviruses. *The Journal of general virology 84*, Pt 11 (2003), 2895-908.
4. Purkayastha, A., Ditty, S.E., Su, J., et al. Genomic and bioinformatics analysis of HAdV-4, a human adenovirus causing acute respiratory disease: implications for gene therapy and vaccine vector development. *Journal of virology 79*, 4 (2005), 2559-72.
5. Adenoviruses. http://www.microbiologybytes.com/virology/Adenoviruses.html.
6. Adenoviridae - Wikipedia, the free encyclopedia. http://en.wikipedia.org/wiki/Adenovirus.
7. Shenk, T.E. Adenoviridae: The Viruses and Thier Replication. In B.N. Fields, D.M. Knipe and P.M. Howley, eds., *Fields' Virology*. Wolters Kluwer Health/Lippincott Williams & Wilkins, Philadelphia, 2007.
8. Dán, Á., Élő, P., Harrach, B., Zádori, Z., and Benkő, M. Four New Inverted Terminal Repeat Sequences from Bovine Adenoviruses Reveal Striking Differences in the Length and Content of the ITRs. *Virus Genes 22*, 2 (2001), 175-179.
9. Ishiko, H., Shimada, Y., Konno, T., et al. Novel human adenovirus causing nosocomial epidemic keratoconjunctivitis. *Journal of Clinical Microbiology 46*, 6 (2008), 2002-2008.
10. Walsh, M.P., Chintakuntlawar, A., Robinson, C.M., et al. Evidence of molecular evolution driven by recombination events influencing tropism in a novel human adenovirus that causes epidemic keratoconjunctivitis. *PloS one 4*, 6 (2009), e5635.
11. Walsh, M.P., Seto, J., Jones, M.S., Chodosh, J., Xu, W., and Seto, D. Computational Analysis Identifies Human Adenovirus Type 55 as a Re-emergent Acute Respiratory Disease Pathogen. *Journal of Clinical Microbiology*, (2009).
12. Jones, M.S., Harrach, B., Ganac, R.D., et al. New adenovirus species found in a patient presenting with gastroenteritis. *Journal of Virology 81*, 11 (2007), 5978-5984.
13. Harrach. http://www.vmri.hu/~harrach/ADENOSEQ.HTM.

14. Walsh, M.P., Seto, J., Tirado, D., et al. Computational analysis of human adenovirus serotype 18. *Virology*, (2010).
15. Zock, C., Iselt, A., and Doerfler, W. A unique mitigator sequence determines the species specificity of the major late promoter in adenovirus type 12 DNA. *Journal of virology 67*, 2 (1993), 682-93.
16. Mei, Y., Skog, J., Lindman, K., and Wadell, G. Comparative analysis of the genome organization of human adenovirus 11, a member of the human adenovirus species B, and the commonly used human adenovirus 5 vector, a member of species C. *J Gen Virol 84*, 8 (2003), 2061-2071.
17. Seto, J., Walsh, M.P., Mahadevan, P., et al. Genomic and bioinformatics analyses of HAdV-14p, reference strain of a re-emerging respiratory pathogen and analysis of B1/B2. *Virus Research 143*, 1 (2009), 94-105.
18. Garnett, C.T., Erdman, D., Xu, W., and Gooding, L.R. Prevalence and Quantitation of Species C Adenovirus DNA in Human Mucosal Lymphocytes. *J. Virol. 76*, 21 (2002), 10608-10616.
19. Seto, J., Walsh, M.P., Metzgar, D., and Seto, D. Computational analysis of adenovirus serotype 5 (HAdV-C5) from an HAdV coinfection shows genome stability after 45 years of circulation. *Virology 404*, 2 (2010), 180-186.
20. Atkinson, R.L., Dhurandhar, N.V., Allison, D.B., et al. Human adenovirus-36 is associated with increased body weight and paradoxical reduction of serum lipids. *International journal of obesity (2005) 29*, 3 (2005), 281-6.
21. Echavarria, M. Adenovirus. In A.J. Zuckerman, J.E. Banatvala, B.D. Schoub, P.D. Griffiths and Mortimer, eds., *Principles and Practice of Clinical Virology*. John Wiley and Sons, San Diego, CA, USA, 2009, 463-488.
22. Lemckert, A.A.C., Grimbergen, J., Smits, S., et al. Generation of a novel replication-incompetent adenoviral vector derived from human adenovirus type 49: manufacture on PER.C6 cells, tropism and immunogenicity. *J Gen Virol 87*, 10 (2006), 2891-2899.
23. Kajon, A.E., Moseley, J.M., Metzgar, D., et al. Molecular epidemiology of adenovirus type 4 infections in US military recruits in the postvaccination era (1997-2003). *The Journal of Infectious Diseases 196*, 1 (2007), 67-75.
24. Steinthorsdottir, V. and Mautner, V. Enteric adenovirus type 40:E1B transcription map and identification of novel E1A-E1B cotranscripts in lytically infected cells. *Virology 181*, 1 (1991), 139-49.
25. Han, B., Guo, L., Qu, J., et al. Improved replication of enteric adenovirus type 41 in Hep2 cell line expressing E1B55K. *Bing du xue bao = Chinese journal of virology / [bian ji, Bing du xue bao bian ji wei yuan hui 23*, 4 (2007), 258-64.
26. Gustafsson, D.J., Segerman, A., Lindman, K., Mei, Y., and Wadell, G. The Arg279Glu Substitution in the Adenovirus Type 11p (Ad11p) Fiber Knob Abolishes EDTA-Resistant Binding to A549 and CHO-CD46 Cells, Converting the Phenotype to That of Ad7p. *J. Virol. 80*, 4 (2006), 1897-1905.
27. Robinson, C.M., Rajaiya, J., Walsh, M.P., et al. Computational analysis of human adenovirus type 22 provides evidence for recombination between human

adenoviruses species D in the penton base gene. *Journal of Virology*, (2009).
28. Hierholzer, C.J. Adenoviruses. In N.J. Schmidt, D.A. Lennette, E.T. Lennette, E.H. Lennette and R.W. Emmons, eds., *Diagnostic Procedures for Viral, Rickettsial and Chlamydial Infections*. American Public Health Association, 1995.
29. Crawford-Miksza, L. and Schnurr, D.P. Analysis of 15 adenovirus hexon proteins reveals the location and structure of seven hypervariable regions containing serotype-specific residues. *Journal of virology 70*, 3 (1996), 1836-44.
30. Madisch, I., Harste, G., Pommer, H., and Heim, A. Phylogenetic Analysis of the Main Neutralization and Hemagglutination Determinants of All Human Adenovirus Prototypes as a Basis for Molecular Classification and Taxonomy. *J. Virol. 79*, 24 (2005), 15265-15276.
31. Li, Q.G., Hambraeus, J., and Wadell, G. Genetic relationship between thirteen genome types of adenovirus 11, 34, and 35 with different tropisms. *Intervirology 32*, 6 (1991), 338-50.
32. Acute Respiratory Disease Associated with Adenovirus Serotype 14 --- Four States, 2006--2007. http://www.cdc.gov/mmwr/preview/mmwrhtml/mm5645a1.htm.
33. Metzgar, D., Osuna, M., Kajon, A.E., Hawksworth, A.W., Irvine, M., and Russell, K.L. Abrupt emergence of diverse species B adenoviruses at US military recruit training centers. *The Journal of infectious diseases 196*, 10 (2007), 1465-73.
34. Shott, J.P., McGrath, S.M., Pau, M.G., et al. Adenovirus 5 and 35 vectors expressing Plasmodium falciparum circumsporozoite surface protein elicit potent antigen-specific cellular IFN-gamma and antibody responses in mice. *Vaccine 26*, 23 (2008), 2818-23.
35. Wesley, R.D. and Lager, K.M. Evaluation of a recombinant human adenovirus-5 vaccine administered via needle-free device and intramuscular injection for vaccination of pigs against swine influenza virus. *American Journal of Veterinary Research 66*, 11 (2005), 1943-1947.
36. Hierholzer, J.C. and Pumarola, A. Antigenic characterization of intermediate adenovirus 14-11 strains associated with upper respiratory illness in a military camp. *Infection and immunity 13*, 2 (1976), 354-9.
37. Hierholzer, J.C., Pumarola, A., Rodriguez-Torres, A., and Beltran, M. Occurrence of respiratory illness due to an atypical strain of adenovirus type 11 during a large outbreak in Spanish military recruits. *American journal of epidemiology 99*, 6 (1974), 434-42.
38. Lukashev, A.N., Ivanova, O.E., Eremeeva, T.P., and Iggo, R.D. Evidence of frequent recombination among human adenoviruses. *J Gen Virol 89*, 2 (2008), 380-388.
39. Engelmann, I., Madisch, I., Pommer, H., and Heim, A. An outbreak of epidemic keratoconjunctivitis caused by a new intermediate adenovirus 22/H8 identified by molecular typing. *Clin Infect Dis 43*, 7 (2006), e64-6.
40. Yang, Z., Zhu, Z., Tang, L., et al. Genomic analyses of recombinant adenovirus type 11a in China. *Journal of Clinical Microbiology 47*, 10 (2009), 3082-3090.
41. Zhu, Z., Zhang, Y., Xu, S., et al. Outbreak of acute respiratory disease in China caused by B2 species of adenovirus type 11. *Journal of clinical microbiology 47*, 3

(2009), 697-703.
42. Delsuc, F., Brinkmann, H., and Philippe, H. Phylogenomics and the reconstruction of the tree of life. *Nature Reviews. Genetics 6*, 5 (2005), 361-375.
43. Xu, W., McDonough, M.C., and Erdman, D.D. Species-specific identification of human adenoviruses by a multiplex PCR assay. *Journal of Clinical Microbiology 38*, 11 (2000), 4114-4120.
44. Ovcharenko, I., Loots, G.G., Hardison, R.C., Miller, W., and Stubbs, L. zPicture: Dynamic Alignment and Visualization Tool for Analyzing Conservation Profiles. *Genome Res. 14*, 3 (2004), 472-477.
45. Lole, K.S., Bollinger, R.C., Paranjape, R.S., et al. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *Journal of Virology 73*, 1 (1999), 152-60.
46. Katoh, K. and Toh, H. Recent developments in the MAFFT multiple sequence alignment program. *Brief Bioinform 9*, 4 (2008), 286-298.
47. Pettersen, E.F., Goddard, T.D., Huang, C.C., et al. UCSF Chimera--a visualization system for exploratory research and analysis. *Journal of Computational Chemistry 25*, 13 (2004), 1605-1612.
48. Nuin, P., Wang, Z., and Tillier, E. The accuracy of several multiple sequence alignment programs for proteins. *BMC Bioinformatics 7*, 1 (2006), 471.
49. Rice, P., Longden, I., and Bleasby, A. EMBOSS: The European Molecular Biology Open Software Suite. *Trends in Genetics 16*, 6 (2000), 276-277.
50. Tamura, K., Dudley, J., Nei, M., and Kumar, S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution 24*, 8 (2007), 1596-9.
51. Tamura, K., Nei, M., and Kumar, S. Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proceedings of the National Academy of Sciences of the United States of America 101*, 30 (2004), 11030-11035.
52. Brudno, M., Do, C.B., Cooper, G.M., et al. LAGAN and Multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA. *Genome Research 13*, 4 (2003), 721-731.
53. Scott Schwartz, Zheng Zhang, Kelly A. Frazer, Arian Smit, Cathy Riemer, John Bouck, Richard Gibbs, Ross Hardison, and Webb Miller. PipMaker---A Web Server for Aligning Two Genomic DNA Sequences. *Genome Res. 10*, 4 (2000), 577-586.
54. Holland, R.C.G., Down, T.A., Pocock, M., et al. BioJava: an open-source framework for bioinformatics. *24*, 18 (2008), 2096–2097.
55. hibernate.org - Hibernate. http://www.hibernate.org/.
56. Springframework.org. http://www.springframework.org/.
57. Martin, D. and Rybicki, E. RDP: detection of recombination amongst aligned sequences. *Bioinformatics 16*, 6 (2000), 562-563.
58. Kendall, E.J., Riddle, R.W., Tuck, H.A., Rodan, K.S., Andrews, B.E., and McDonald, J.C. Pharyngo-conjunctival fever; school outbreaks in England during the summer of 1955 associated with adenovirus types 3, 7, and 14. *British medical*

journal 2, 5037 (1957), 131-6.
59. Van Der Veen, J. and Kok, G. Isolation and typing of adenoviruses recovered from military recruits with acute respiratory disease in The Netherlands. *American journal of hygiene 65*, 2 (1957), 119-29.
60. Seto, J., Walsh, M.P., Mahadevan, P., Zhang, Q., and Seto, D. Applying Genomic and Bioinformatic 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.
61. Graham, F.L., van der Eb, A.J., and Heijneker, H.L. Size and location of the transforming region in human adenovirus type 5 DNA. *Nature 251*, 5477 (1974), 687-691.
62. Huebner, R.J., Rowe, W.P., and Lane, W.T. Oncogenic effects in hamsters of human adenovirus types 12 and 18. *Proceedings of the National Academy of Sciences of the United States of America 48*, 12 (1962), 2051–2058.
63. Crawford-Miksza, L.K. and Schnurr, D.P. Adenovirus serotype evolution is driven by illegitimate recombination in the hypervariable regions of the hexon protein. *Virology 224*, 2 (1996), 357-67.
64. Robinson, C.M., Rajaiya, J., Walsh, M.P., et al. Computational analysis of human adenovirus type 22 provides evidence for recombination between human adenoviruses species D in the penton base gene. *Journal of virology 83*, (2009), 8980-85.
65. Robinson, C.M., Shariati, F., Gillaspy, A.F., Dyer, D.W., and Chodosh, J. Genomic and bioinformatics analysis of human adenovirus type 37: new insights into corneal tropism. *BMC genomics 9*, (2008), 213.
66. Madisch, I., Wolfel, R., Harste, G., Pommer, H., and Heim, A. Molecular identification of adenovirus sequences: a rapid scheme for early typing of human adenoviruses in diagnostic samples of immunocompetent and immunodeficient patients. *Journal of medical virology 78*, 9 (2006), 1210-7.
67. Miura-Ochiai, R., Shimada, Y., Konno, T., et al. Quantitative Detection and Rapid Identification of Human Adenoviruses. *J. Clin. Microbiol. 45*, 3 (2007), 958-967.

CURRICULUM VITAE

Michael Patrick Walsh is the grateful son of Irene Kearney Walsh, the proud brother of Major Kevin Collin Walsh USMC and the beloved nephew of both Mary Kathleen Kearney and John Francis Kearney. He received a Bachelor of Science degree from Virginia Polytechnic Institute and State University in 2001, a Master of Science degree from George Mason University in 2006 and a Doctor of Philosophy degree from George Mason University in 2010. He is godfather to John Richmond Long and Alexander Mathew Cetta and he is a practicing Roman Catholic.